国家出版基金项目
NATIONAL PUBLICATION FOUNDATION

Elsevier Handbook of the Philosophy of Science

爱思唯尔
科学哲学手册

物理学哲学

Philosophy of Physics

英文本丛书主编
[以色列]道·加比(Dov Gabbay)
[加拿大]保罗·撒加德(Paul Thagard)
[加拿大]约翰·伍兹(John Woods)

中译本丛书主编
郭贵春　殷　杰

本卷主编
[美　国]约翰·厄尔曼(John Earman)
[英　国]杰里米·巴特菲尔德(Jeremy Butterfield)

本卷译者
程　瑞　赵　丹　王凯宁　李继堂

北京师范大学出版集团
BEIJING NORMAL UNIVERSITY PUBLISHING GROUP
北京师范大学出版社

爱思唯尔科学哲学手册
物理学哲学

杰里米·巴特菲尔德
约翰·厄尔曼(编)

程　瑞
赵　丹
王凯宁
李继堂译

北京师范大学出版社

图书在版编目(CIP)数据

物理学哲学/郭贵春，殷杰主编．程瑞，赵丹，王凯宁，李继堂 译．—北京：北京师范大学出版社，2015.12

(爱思唯尔科学哲学手册)

ISBN 978-7-303-19177-2

Ⅰ.①物… Ⅱ.①郭… ②殷… Ⅲ.①物理学哲学 Ⅳ.①O4

中国版本图书馆CIP数据核字(2015)第164895号

营 销 中 心 电 话 010-58805072 58807651
北师大出版社学术著作与大众读物分社 http://xueda.bnup.com

WULIXUE ZHEXUE

出版发行：北京师范大学出版社 www.bnupg.com
北京市海淀区新街口外大街19号
邮政编码：100875

印 刷：北京盛通印刷股份有限公司
经 销：全国新华书店
开 本：170 mm×240 mm
印 张：109
字 数：1640千字
版 次：2015年12月第1版
印 次：2015年12月第1次印刷
定 价：350.00元

策划编辑：饶 涛 责任编辑：刘文平
美术编辑：王齐云 装帧设计：王齐云
责任校对：陈 民 责任印制：马 洁

Philosophy of Physics (ISBN: 9780444515605)
Edited by Andrew Irvine, Dov M. Gabbay, Paul Thagard and John Hayden Woods

This edition of Philosophy of Physics by Jeremy Butterfield, John Earman, Dov M. Gabbay, Paul Thagard and John Hayden Woods is published by arrangement with ELSEVIER BV of Radarweg 29, 1043 NX Amsterdam, Netherlands.

中文简体版 ISBN: 978-7-303-19177-2

中文简体版由 ELSEVIER BV 授权北京师范大学出版社出版发行
北京市版权局著作权合同登记号: 图字 01-2015-4371

译者序

一

整个20世纪，伴随着一系列科学上的革命性进展，因其而产生的科学哲学也经历着从初期到兴盛再到转型的过程。在刚刚步入21世纪之初，面对科学哲学未来走向的发问，一批科学哲学家以及科学家被召集起来共同完成了这部迄今门类规划最为全面的科学哲学丛书——《爱思唯尔科学哲学手册》。它以宏大的视角来展现步入新世纪的科学哲学研究面貌，通过对一般科学哲学以及各具体科学哲学研究的梳理与阐释，试图为未来科学哲学开启一幅远景。正如这套科学哲学手册的三位英文版主编——道·加比(Dov Gabbay)、保罗·撒加德(Paul Thagard)、约翰·伍兹(John Woods)所共同认为的，在已知的任何时期，科学在其前沿的运行总是遭遇到有关知识与实在的本质的哲学议题。科学论战会引发诸如理论与实验的关系，解释的本质以及科学接近于真实的程度的问题。在具体科学中，关于存在是什么以及如何知晓它的问题会引发特殊的关注，例如，物理学中的时空本质问题、心理学中的意识本质问题。因此，科学哲学是对世界进行科学调查的必要部分，并且总的来

说，科学哲学正日益成为哲学的核心。尽管仍有哲学家认为关于知识与实在的理论可以通过纯粹的沉思而发展出来，但大多数的哲学研究表明，重视相关科学发现才是必要的和有价值的。例如，心灵哲学已经很明显地与经验心理学绑定在一起，而政治理论经常与经济学产生交叉。这些科学哲学研究为哲学探究与科学研究之间架起了一座宝贵的桥梁。科学哲学本身也越来越不再局限于关注一般的科学本质与科学合法性议题，而是格外关注在具体科学哲学内部所引发的特有议题。

正是出于这个原因，该丛书的英文版主编们规划了目前最为齐全的科学哲学子学科群，并且力图确立一种偏向于具体科学哲学问题的研究模式。这一举动也在一定程度上表明，具体科学哲学已经成为未来科学哲学研究的重要方面，而造成这种趋向的原因，一方面在于当代科学自身的发展以及研究模式发生了巨大变化，科学哲学研究也有必要顺应这种变化来对自身的研究目标和方式进行重新定位和调整，以便保持连接哲学与科学之间的这座桥梁的通畅；另一方面，科学与其他人文社会学科的关系已经愈发紧密，存在着普遍的交叉，而产生的影响又多体现在科学中的子学科甚至一些研究前沿与人文、社会领域的交互上，这要求我们不能简单地将科学作为一个简单的整体，而是应该在各个具体科学领域中来探讨科学之于人文、社会领域的交叉与影响。基于这些判断，可以认为，未来科学哲学研究必将伴随着各个领域上的科学发现而不断走向前沿和深化，并在一些传统议题上走向新的理解与探讨，将科学哲学引向新的发展阶段。

为了实现最初的意图，丛书邀请了众多熟谙各门具体科学的分卷编者，他们从通晓科学的哲学家们以及(少部分)通晓哲学的科学家们那里征求来的宝贵稿件令人欣喜。这16卷系列丛书为当代科学哲学研究门类提供了一种目前最为完善和齐全的纵览。这套丛书由目前世界上最大的医学与其他科学文献出版机构之一的爱思唯尔(Elsevier)出版集团推出，具体目录如下：

《一般科学哲学：焦点主题》(2007年7月出版)

《物理学哲学》(2006年10月出版)

《生物学哲学》(2007年2月出版)

《数学哲学》(2009年6月出版)

《逻辑哲学》(2006年10月出版)

《信息哲学》(2008年12月出版)

《技术与工程科学哲学》(2009年8月出版)

《心理学与认知科学哲学》(2006年10月出版)

《人类学与社会学哲学》(2006年10月出版)

《复杂系统哲学》(2011年5月出版)

《统计学哲学》(2011年5月出版)

《经济学哲学》(2012年4月出版)

《医学哲学》(2011年7月出版)

《地球系统科学哲学》(2011年4月出版)

《化学与药物学哲学》(2011年11月出版)

《语言学哲学》(2012年1月出版)

这些分卷广泛考察了基础科学以及应用科学中业已提出的哲学议题。目前，与这套丛书类似的其他著名丛书有:

——《剑桥指南系列丛书》(*The Series of Cambridge Companion*)，这套丛书十分宏大，不过主要是从整个哲学体系的角度来分卷探讨不同的人物哲学思想、哲学领域或命题，其中自然包含了科学哲学，但依然主要是以某种思想或领域的形式来进行撰写(比如，《逻辑经验主义》、《生物学哲学》)。整体提纲偏重于哲学的历史性脉络，而内容同样是采取收录相关专家供稿的方式。

——《牛津哲学手册》(*The Oxford Handbooks in Philosophy*)，与上一部相似，在这套丛书中，科学哲学同样是作为整个哲学体系下的一个分支或衍生命题，甚至其学科定位在丛书规划中没有专门的体现。

——《劳特利奇指南系列丛书》(*The Series of Routledge Companion*)也是同样的风格，与前面介绍的丛书的最大共同点在于，直接将科学哲学纳入到哲学框架之下，注重科学哲学研究中的主要议题，而没有专门突出其学科性质。

——《哲学与科学》(*Philosophy and Science*)系列丛书由欧酷曼(Acumen)出版社出版，亚历山大·伯德(Alexander Bird)作为总主编，丛书主要偏向于有关科学知识的理论与方法论中的核心论题(比如，《科学方法的理论》、《经验主义》、《模型与理论》)，重在梳理有关这些论题的最新探讨，介绍一些新出现

的子学科领域。该丛书同时也期望尽力覆盖到近年来从科学理论以及一般哲学层面涌现出的重大基础性议题(比如，《生物学哲学》、《精神病学与科学哲学》)。总之，虽然专门针对科学哲学，但该套丛书的内容规划范围有限，依然偏重于一些传统领域和问题。

——《波士顿科学哲学与科学史研究》(*Boston Studies in Philosophy and History of Science*)这部由斯普林格(Springer)出版社出版的丛书则属于十分浩大与深入的科学哲学与科学史研究系列。这套丛书当属目前科学哲学界最为浩瀚夺目的系列成果，面向各个领域的具体论题，而不是偏重于学科领域来划分，这也就导致了该丛书的专业性质以及宏大论题规划，目前已滚动出版数十年，出版300余册，依照分属不同学科、层面以及角度的论题来进行收录，几乎涵盖了近半个世纪以来各个时期的科学哲学研究内容。

——《西安大略科学哲学丛书》(*The Western Ontario Series in Philosophy of Science*)则表现为类似杂志的期卷形式，每一卷都会设定一个具体的主题，而其中收录的文章都是围绕这一主题的研究，因而某种程度上每一卷更像是一部专题文集，它们共同表现出科学哲学丛书的特点。目前该丛书已经更新至78卷。

——《匹兹堡科学哲学与科学史系列丛书》(*Pittsburgh Series in Philosophy and History of Science*)由科学哲学研究重镇——匹兹堡科学哲学中心完成。该丛书目前只出版了五部，虽然内容规划性不强，主题也较为松散(包括《精神分析基础》、《演绎主义的局限》、《科学的界限》、《科学与价值》、《推理、解释以及其他的挫折》)，彼此之间不存在内容上的联系，但学术水平以及思想深度毋庸置疑。

目前国内已出版的较为知名的类似丛书有：山西大学科学技术哲学研究中心的《科学哲学文库》系列，该文库目前已经收录了50余部著作，其中主要是中心研究人员的研究成果，体现了该中心在科学技术哲学研究领域的研究实力。丛书内容规划全面，采取滚动收录的方式，突出前沿性，通过持续出版，保持着与国际该领域研究的同步，很受国内相关研究领域人士的重视。另一部为中国人民大学出版社出版的《科学哲学基本著作》丛书，这套丛书偏向于经典导读以及基础性介绍，主要收录了一些科学哲学领域代表性人物的经典译著

(例如，卡尔纳普的《科学哲学导论》、亨普尔的《自然科学的哲学》，等等)，以及一些国内科学哲学领域的学者所著的科学哲学基础性读物。

二

通过以上同类丛书的比较，可以发现爱思唯尔出版的这部《爱思唯尔科学哲学手册》最大特点在于其所规划的严密学科体系，并且面向基础性的学科导论，与具体科学发展的历史联系紧密(甚至邀请许多具体科学领域的学者撰写一些重要科学人物的传记以及某些学科或命题的发展简史)，科学色彩浓厚，所涉及的哲学论题十分前沿，不仅受到许多哲学学者的关注，也受到了来自于各相关领域科学家们的好评。在内容上兼顾了基础性及前沿性，是一部特点鲜明、不可多得的科学哲学丛书。下面我们对首批出版的9部译著做一简要的概览:

——《一般科学哲学: 焦点主题》由荷兰格罗宁根大学理论哲学系教授西奥·A. F. 库珀斯主编，是爱思唯尔科学哲学手册中具有纲领性意义的一卷。本书精心挑选了一般科学哲学研究领域中若干核心问题进行了深度阐释，提纲挈领地勾勒出一幅当代科学哲学研究的全景画卷。本分册包含十个系统论述: 第一章“定律、理论和研究纲领”，对经验科学的这三大主要单元进行了综合研究; 第二章“对解释的以往和当代观点”，全面描述了历史上若干最具代表性的哲学家，包括从亚里士多德、笛卡尔、莱布尼茨、牛顿、休谟、康德到穆勒对“解释”的理解，以此为基础对最重要和有争议的当代解释模型提供了一种系统性考察; 第三章“理论评价”考察了一致性、真、先天和后天的可能性、信息内容、经验内容、解释力和预言力、解决问题的能力、简单性、精确性、近似为真和似真性等理论评价的核心标准，并进一步讨论了如何从比较的以及量的方面来阐释理论确证; 第四章“自然科学中实验的功能”，以物理学与生物学为例，提出了一种实验的认识论，即在实验结果中提供合理信念的一套战略; 第五章“社会科学中实验的功能”，以经济学为案例，集中在社会科学中实验的发展上，尤其考虑了莱因哈德·泽尔腾的贡献; 第六章“本体论、认识论和方法论的立场”，指出在科学史中本体论课题具有最高的重要性，认识论难题主要涉及归纳推理，最具基础性的是阐释理论与证据之间的关系，方法论问题中最

关键的是科学方法的单一性或确定性问题；第七章“还原、整合与科学的统一”，考察了替代统一的支持整合的论证以及基于机械论的支持还原的论证，提供了热物理学、分子与发展生物学、考古学以及语言学中关于推定的还原与整合的案例研究；第八章“逻辑的、历史的和计算的方法”，指出对机器学习的研究为诸如归纳和溯因提供了新的洞见，推进了新的阐释，同时也使得科学发现的问题被提到了一个更加形式化的层面；第九章“科学与非科学的划界”探讨了划界问题的标准，提出了描述性与规范性科学指标的综合名单；第十章“科学哲学的历史”，创造性地重建了一种自20世纪初以来科学哲学从中欧到盎格鲁—撒克逊世界的长期转换、变形和互动发展。全书各章节主题各异，但都从不同的方面揭示和印证了“阐释”这一在科学哲学中极为重要，却由于其隐含性而长期被忽视的科学哲学方法。在这一意义上，“阐释”也是贯穿整套科学哲学手册的一条重要逻辑线索。

——《物理学哲学》在整套丛书中拥有最长的篇幅。物理学哲学是科学哲学研究中的一个传统领域，同时，从某种程度上说，它又是所有哲学中最古老的分支。之所以这么说，是因为与物理学相关的哲学探讨在古希腊时期就已经出现了。但相对来说，物理学哲学依然是一个新兴学科，正如其编者巴特菲尔德所说，物理学哲学作为一个专门的研究领域，并成为一个强大而且充满生机的哲学分支，还只是过去近半个世纪的事。

现代物理学的三大支柱——统计物理、量子理论与相对论，对当代科学哲学的发展影响甚大，它们不只为哲学理论提供了具体的研究案例，更促使科学哲学家们转向了对实在论与工具主义、决定论与非决定论等问题的争论。自20世纪60年代以来，物理学自身内部的一些基本问题，也引起了科学家和哲学家们的关注，并且逐渐成为物理学哲学研究的主要方向。《物理学哲学》的两位编者正是在这样的背景下，选取了14篇论题相对集中的代表性文章结集成册，来为我们展现近代物理学的这些基本问题。

《物理学哲学》的主要内容可以概括为对当代物理学的五个主题中基本问题的探讨，即热物理学、量子理论、相对论、量子场论和量子引力，它们一共占据了前十二章内容。对热物理学基本内容的介绍主要由尤菲克和埃姆什给出，他们分别从经典与量子两方面阐述了热力学与统计物理中基本概念的演变。关

于量子理论的内容主要有三章，分别是：迪克森对非相对论量子力学的形式体系及其中涉及的基本问题的回顾，兰兹曼从数学物理学的视角对量子与经典理论关系的探讨，以及巴布对量子信息和量子计算中一些核心思想和结果的阐释。与相对论有关的论题由马拉蒙特、贝洛特和埃利斯完成，其中马拉蒙特讨论了三个特定论题：狭义相对论中同时性的定义、牛顿引力的几何化以及因果结构如何决定时空几何的问题；贝洛特则主要对发生在经典广义相对论中的“时间难题”做了一个详细的陈述；埃利斯通过回顾相对论宇宙学理论的研究现状，探讨了包括人择原理和多重宇宙的存在可能性等九个哲学主题。在量子场论方面，特霍夫特从粒子物理学的视角对量子场论进行了权威的纵览，霍尔沃森和缪格则借助代数量子场论工具，讨论了粒子和定域化的本质、非定域性、量的赋值和单个时空点上量子场的可定义性问题。在量子引力方面，主要有两篇文章，其中罗韦利的论文介绍了该学科的历史、当前两种主要理论(弦理论和圈量子引力)以及量子宇宙学的现状；埃利斯在其撰写的章节提出关于“终极”物理理论观念等基本问题，可以看作是对罗韦利一文的自然补充。最后的两章则分别讨论了经典物理学和近代物理学中普遍存在的两个基本问题，即决定论与对称性。对这两个问题的探讨也为彼此相互分割的各物理学分支之间建立起了一种密切的联系。各章与各类物理学哲学期刊中的论文相比，在研究方法上，更偏重于对问题的历史性回顾而不是逻辑性论证；在探讨的问题方面，则尽量回避了偏向形而上学的问题，而集中对物理学哲学中的具体问题进行阐释。

现代物理学哲学与物理学自身之间并没有一条泾渭分明的界限，而且，物理学哲学领域的一些非常突出的成果甚至都是由物理学家做出的。这一点从各章作者的身份就可以看出，因为在他们之中除了专门从事物理学哲学研究的学者之外，还有几位是理论物理学家，即埃利斯、埃姆什、兰兹曼、罗韦利、缪格和特霍夫特，其中特霍夫特还是诺贝尔物理学奖获得者。从另一方面看，完成其他各章的科学哲学家们，包括贝洛特、布拉丁、巴布、巴特菲尔德、卡斯特拉尼、迪克森、厄尔曼、霍尔沃森、马拉蒙特和尤菲克，都接受过完整的哲学训练，也掌握了充足的物理学知识。因此每一章都非常专业且完整地介绍了它们所涉及的论题，明确展示了现代物理学哲学研究偏重物理学前沿问题的

特点。

——《生物学哲学》由多伦多大学的莫汉・马修与英属哥伦比亚大学的克里斯托弗・斯蒂芬斯两位加拿大学者共同主编。生物学哲学作为近几十年迅速崛起的科学哲学子学科，在各方面都取得了令人瞩目的成果。它形成于逻辑经验主义晚期不断走向衰落的历史背景之下，将生物学及其理论作为一种另类科学形式展开全方位的讨论，覆盖了从科学到社会人文领域的广阔论题，使人们看到了一种充满活力并多姿多彩的科学哲学研究。

本书较之其他的同类著作，最大的特点在于将研究的重点聚焦在进化理论上。特别是在第一部分，收录了达尔文、费希尔、赖特、霍尔丹、木村资生这五位历史上极为重要的进化理论家的传记，充分体现了这本书的主编强调进化理论的意图，同时也隐晦地表明一种态度，即自达尔文以来的进化理论研究很大程度上是一项充满哲学思辨性的议题，特别是关于物种起源与进化的理论化本身需要哲学的参与。第二部分直接将进化作为该部分的标题，在其中探讨自然选择、中性论、选择层级、可演化性、发育论、进化规范性、进化伦理学议题，从本质、机制、层次以及社会性议题等方面全方位阐述有关进化所引发的哲学问题，特别是在理论层面上的问题。这一部分充分体现了哲学分析方法在进化理论研究中不可或缺的地位。第三部分聚焦遗传问题，这一议题在某种程度上依然是进化论题的补充，而在讨论内容上则更加偏向于生物学研究的方法论。例如，遗传分析、群体遗传的理论构建问题、生物学中的还原论问题、遗传代码的定义与分析等。这一部分主要还是继承了生物学哲学中的一些传统议题，只不过是在现代生物学语境下对它们提出了一些新的解读。第四部分探讨分类学问题，就生物学命名体系所面临的概念问题进行深入的分析，而其中的哲学问题主要还是围绕传统分类学理论与当前进化理论观念之间的概念冲突与结构分析，充分体现了分析哲学方法在生物学中的应用。最后一个部分专门设为特殊话题，包括生物学理论的形式化议题、生物学功能概念的分析、生物学路径下的心灵问题研究、先天性的概念分析以及对于人工生命问题的哲学探讨。应该说，这一部分既囊括了像功能分析这样的传统哲学话题，也涵盖了人工生命这样的新兴话题，它们分别代表了生物学哲学研究开枝散叶的不同方向，很有可能发展为极具潜力的交叉研究。

从本书的章节作者上看，这部书不但请来了迈克尔·鲁斯、萨赫托·萨卡尔、拉斐尔·福尔克等老一代科学哲学家作为撰稿人，同时也邀请来了亚历山大·罗森伯格、金·斯特林等如日中天的一线学者，甚至包括詹姆斯·F. 克罗这样的遗传学家。可见这本书的主编们用心良苦，尽力为读者阐述当下的生物学哲学研究究竟是什么人在做，怎么做，有哪些议题，从而最大限度地将生物学哲学研究的面貌展现出来，形成系统的研究体系。

——《数学哲学》的写作立脚点在于，提供一种尽可能全面并具有一定深度的数学哲学研究回顾与前瞻。当前，数学哲学的根本任务在于为数学的本质及其数学实践提供一种令人满意的、连贯的、普遍的和整体的哲学说明。由此数学哲学的核心论域包括三个方面：第一，数学的本体论。数学的本质是什么？即，数学的研究对象是个别的作为个体的数学对象还是数学结构？数、函数、集合、群等这样的数学实体存在吗？它们是物质实体、心理实体抑或是柏拉图世界中的抽象实体？它们的存在独立于我们的物质世界以及我们人类的大脑吗？由此延伸出一系列有关数学实体的认识论问题和数学真理问题。对这类问题的探讨引发了各种数学实在论和反实在论的立场及争论。典型的有：数学柏拉图主义、数学对象柏拉图主义、结构主义、虚构主义、不可或缺性论证、数学的自然主义实在论、新逻辑主义或新弗雷格主义，等等。第二，数学知识的本质及其确定性。数学知识的本质是先验的还是经验的？数学知识确实有一个确定的基础吗？对这类问题的回答引发了数学知识的先验论和经验主义解释，继而激发了康德对数学知识所做的著名的先验综合说明，寻求数学基础的努力则造就了20世纪前半叶的三大基础主义学派：逻辑主义、直觉主义和形式主义。第三，数学实践中的哲学问题。这又包括两方面：(1)与一般科学哲学问题相类似的一般的数学哲学问题：数学说明的本质是什么？数学中有哪些推理？数学与实在世界的关系是什么？选择新的数学公理依赖的根据是什么？等等。(2)具体数学分支领域中的哲学问题，比如集合论、概率论、可计算性理论、范畴论等引发的哲学问题。

而本书从历史的脉络围绕数学哲学探讨的主题为我们提供了一幅内容较丰富且较全面的数学哲学图景：传统理论(柏拉图主义、亚里士多德主义、经验主义、康德主义)；三大基础主义学派(逻辑主义、形式主义、构造主义)；当

代的观点(虚构主义、全面柏拉图主义、不可或缺性论证、结构主义、演绎主义、约定主义、自然主义实在论、模态主义解释等)；具体数学分支领域中哲学问题的前沿研究(集合论、可计算性理论、概率论、弗协调性)。由于篇幅所限，本书尽管没有面面俱到地涵盖数学哲学的所有发展和观点，然而它仍不失为一本兼具前沿视域宽广和信息量丰厚之作。本书既可以为专业研究人员提供研究资料，又可以为想了解数学哲学发展的读者提供一种源自历史视角的全面概览。

——《逻辑哲学》主编戴尔·杰凯特，是当今最著名的逻辑学家之一。他的学术研究涉及逻辑、内涵性、形而上学、认识论、心灵哲学及哲学史等领域。在他主编的《逻辑哲学》中，收录了当代最著名的逻辑学家、数学家和集合论学家在逻辑哲学研究方面的最新成果。这些成果运用示范性的专业技术，在广度和深度上囊括了数理逻辑和哲学逻辑中的广泛主题。本书共收录 29 篇论文，这些论文代表了当代逻辑和逻辑哲学的研究现状。本书的目的之一是为学生和专业研究人员了解逻辑哲学的主要发展提供基本背景；二是收集了逻辑及其哲学中的新视角、新挑战和新进展，可作为当前逻辑实践的典范和逻辑未来发展的风向标。本书每章都具有特定主题选择和问题及方法说明，较为全面且深入地呈现了逻辑哲学这一领域的丰富性，既鼓励读者去探索数理逻辑最新发展的技术进步，同时保持了哲学对逻辑本质的敏锐关注。这些论文包括目前形式符号逻辑某些最重要的子学科的主要研究内容，以及围绕逻辑的新理论发现和应用的哲学本质及其影响，涉及形而上学基本概念、知识论、科学哲学、语言哲学以及价值观，特别是美学规范的基础，反映了以逻辑系统构造和评价这些概念的基本方式。在我们看来，这些论文均是以史论结合的方式对具体问题进行详尽阐释，旨在呈现研究的现状和成果，同时保留了问题的讨论空间。因此，通过学习和参考这些论文，读者不仅可以直接了解当前的论题、各位作者所提出的解决方案及逻辑技巧，而且能够直接进入对开放问题的深入探讨。

——《信息哲学》分卷的设立体现了丛书主编们的前沿视野。信息哲学作为一个独立的领域，其研究对象为计算机科学、信息技术以及哲学等学科交叉后出现的各种问题。这些问题包括了信息的概念本质与基本性质，例如其动态分析、使用方式以及相关科学等。同时，也包括对信息理论的细化与应用以及对

各种哲学问题所使用的算式方法。信息哲学源自人工智能哲学、信息逻辑、控制论、社会学、伦理学，以及语言与信息研究。近来这一领域逐渐以“信息哲学”的形式为人们所知。

《信息哲学》虽然内容纷繁浩杂，但是编纂主线明晰，由哲学入手，与信息结合，从整体思想到具体学科，因而从中不难看出学科间的相互交织与促进。第一章至第五章为信息与哲学的基本关系介绍，弗雷德·德雷斯克作为将信息论引入哲学研究的先驱，对信息概念在认识论中的作用，以及如何与其他为人们所热议的论题相结合等问题进行了讨论。随后，彼得· 阿德里安斯从计算学的角度，对物理学宇宙的可知性进行了重点介绍。值得一提的是，阿德里安斯本人原先从事的是古典哲学研究，后来成为研究机器学习的专家，这足以说明信息科学与哲学之间不可分离的紧密关系。第六章至第八章为技术性讨论，对相关手段、方法予以了介绍。弗莱明·托普索和彼得·哈瑞莫斯这两位数学家对香农信息论及其与数学的结合应用进行了阐释。彼得·格伦沃尔德和保罗·威塔涅对算法复杂性理论的发展现状进行了介绍，同时也对该理论与概率以及香农信息论的关联进行了讨论。而逻辑学家约翰·范·本瑟姆和马瑞卡门·马丁内斯通过对认知逻辑以及情境论研究传统的介绍，研究讨论了信息在逻辑中所起的作用。第九章至第十三章讨论了若干基于“信息学”的重要论题。其涉及范围变得更为广泛，凯文·凯利围绕奥卡姆剃刀原理撰写了学习、简化以及信念修正等方面的相关内容。逻辑学家亚历山德鲁·巴尔塔格、汉斯·范·迪特玛施以及劳伦斯·莫斯描述了“动态算法逻辑”中知识与信息更新问题。汉斯·罗特则对信念修正结构及其相应体系结构进行了讨论。而致力于研究“信息动态”的重要学者阿布拉姆斯基使用基于博弈论的交互处理模型，对计算过程中的信息流进行了讨论，并将其与量子信息流联系起来。经济学家伯纳德·瓦里泽也讨论了博弈中的信息和理性代理人本身等问题。第十四章至第十九章列举了许多科学与人文学科中的信息案例。逻辑学家、哲学家、计算机专家迈克尔·邓恩对信息在计算机研究中的各种运用做了调查研究。著名物理学家拜斯和法默则把物理中的信息与香农信息论和柯尔莫哥洛夫复杂性联系在了一起。基思·德夫林和杜斯卡·罗森伯格以情境论作为工具，建立起用于语言交际的深度处理模型。约翰·麦卡锡作为人工智能理论创始人之一，在其撰写的章节中研究了信

息在人工智能中的应用，同时也给哲学研究者们提出了一系列亟待解决的问题。此外，还有两章讨论了生命科学的相关内容。玛格丽特·博登讨论了信息在认知心理学中的作用，并对近期的一些神经学观点有所涉及。约翰·柯利尔则对当前信息以及编码技术在生物学中的应用进行了批判性研究，并产生了广泛反响。所讨论的论题涉及多个学科领域，并且多为前沿性研究，因此不少章节都以开放的方式结尾，这也充分表明信息哲学研究的多样性以及巨大的潜在发展空间。

——《技术与工程科学哲学》分卷的设立显现出丛书的别出心裁。技术哲学是长期以来不被科学哲学家认可的一个领域，甚至存在论者声称技术哲学是不存在的。由荷兰技术哲学家安东尼·梅杰斯主编的《技术与工程科学哲学》就是针对这一传统科学哲学观点而提出的挑战，该书不仅论证了“技术与工程科学哲学”存在的可能性和必要性，而且拓展和深化了传统技术哲学研究领域，极大地推进了技术哲学从批判传统向经验实践的转向，强化了技术哲学的工程传统。

本书除“总导言”之外，共有六个部分。第一部分“技术、工程与科学”，从哲学的角度阐述了科学与技术的关系，并且辅之以同样有差别的社会性与规范性问题的描述；指出 20 世纪以来科学和技术间的区别逐渐模糊，而工程科学也已转化成为一门技术性科学。第二部分“人工物的本体论与认识论”，提出了人工物是心智依赖存在的主张；论证了工程中功能性的部分—整体关系不能被标准的部分论所理解；探讨了在技术与工程中极具重要性的隐性知识的问题；讨论了应用于工程设计中的工具主义的可替代方式规范论。第三部分“工程设计哲学”，讨论了如何创造设计实践类型论和如何将客户需要转变成技术说明；阐明了工程设计在本质上不同于科学研究，工程设计可以被解释成一个要将功能结构转译成物质结构的过程。第四部分“工程科学中的模型化问题”，认为工程中模型化的终极目的是实现可靠的人工物或技术流程；认为基于模型的推理与形式演绎推理有着巨大的差异，因为前者还取决于内容，并非单纯的形式推理；认为与科学解释相比，技术解释包含的内容更加丰富。第五部分“技术与工程中的规范和价值”，阐述了技术与工程是价值负荷观点的几种论证策略；认为规范性既内在于作为一种实践的工程中，也内在于技术人工物和技

术系统中，规范标准不仅体现在技术守则和技术标准中，而且也体现在工程道德守则的基本准则中。价值冲突实际上是设计过程中的核心，认为技术评价越来越被发现其本身就是一项有价值的事业。第六部分“工程学科中的哲学问题”，认为目前致力于技术哲学中具体工程学科的研究比例，远低于致力于科学哲学中具体科学学科的研究比例。本书对技术与工程科学哲学领域相关问题的分析具有相当的广度和深度，引用文献资料丰富，为我们更好地了解西方技术与工程科学哲学的研究现状及其研究内容，开阔了视野，弥补了国内该领域资料的欠缺，这对于促进我国技术与工程哲学的发展将产生重要的影响。

——《心理学与认知科学哲学》的编者保罗·撒加德教授是一位哲学家和认知科学家，现为加拿大滑铁卢大学哲学系教授，兼任心理学和计算机科学系教授。该书的题目很有特色，十分少见地将认知与心理学糅合在一起。心理学是一种关于思想的研究，认知科学是一种关于心灵、智力，并且包括哲学、人工智能、神经科学、语言学以及人类学的跨学科研究。在这些研究中，许多涉及方法和核心概念的论题也随之而产生。本书共包括当前领军的科学哲学家撰写的16篇论文，这些论文阐明了心灵研究中理论与解释的本质，所讨论的论题包括表征、机制、还原、知觉、意识、语言、情绪、神经科学以及进化心理学等，涉及心理学与认知科学哲学的广泛领域，这些文章都与科学研究紧密相关。

当代西方哲学的一个重要特征，就是哲学与科学之间的密切结合，以至于哲学家们讨论的所有话题都只有通过科学的论证才能得到更多的认同。而心理学哲学和认知科学哲学更是集中体现了这个特征。本书向我们很好地展现了当代哲学家们是如何利用心理学和认知科学的研究成果，对传统哲学中提出的身心问题、意识问题和当代哲学中提出的知觉问题、认知问题等给出哲学上的论述。纵观全书，我们可以看到其中体现了三个明显特征：第一，通过对心理学和认知科学的哲学研究，突出这种哲学的科学主义倾向。本书特别强调要把科学心理学和认知科学研究与传统的扶手椅上的概念研究区分开来，要把心理学哲学和认知科学哲学与当下活跃的心灵哲学研究区别开来。这两个区别清楚地表明了本书讨论的问题的确与当代西方心灵哲学研究有了很大差别。如果说心理学哲学和认知科学哲学属于科学研究的范畴，那么，心灵哲学研究则更接近于传统哲学研究的范畴。第二，在科学与哲学的关系上明确采用了自然主义的

策略，强调哲学研究的普遍性和规范性必须结合科学研究的描述性。从科学的观点讨论哲学问题，这原本是科学哲学的研究方法，但自然主义则更重视哲学对科学的不可还原性。这就凸显了哲学研究对科学研究的重要意义。第三，通过对科学活动的解释，强调了心理学和认知科学为我们的认识活动提供了机械论的策略。无论这样的解释是认识论的还是形而上学的，心理学哲学和认知科学哲学似乎都可以帮助我们更好地理解理论与实验结果的关系，由此可以产生关于科学本质的实在论观点。虽然参与本书写作的作者都是职业哲学家，但他们的确都在科学研究领域独领风骚，从他们的论述中，我们更加感觉到机械论的实在论的确是当代心理学哲学和认知科学的主流倾向。

——《人类学与社会学哲学》是一部综合的、覆盖多问题的论文集，它收录了当代社会科学哲学领域内众多知名学者的研究成果，深入地探查了这两门社会科学中的哲学问题。本书共收录了23篇学术论文，可以认为每篇论文都包含着作者深邃的思想和精妙的阐述。按照研究主题的不同，该书将这些文章划分成五个部分，分别是社会学与量化，个体主义和整体主义，人类学、文化与解释，理性与规范性以及批判方法。

第一部分主要关注了社会科学中的量化、测量以及因果模型等问题，它以定量化、模型化的视角重新审视了传统社会科学的研究领域，通过方程和公式的运用，以及模型的建构，独辟蹊径地为我们展现出当代社会科学的定量化的另类研究路径。第二部分主要关注了社会科学中的个体主义和整体主义，既阐述了社会的层次问题，又包含了对整体主义和随附性的研究，当然对理性的选择以及社会科学的进化性解释和功能性解释也进行了深入的探讨，从而对社会科学本身的解释机制做出了很好的说明。第三部分主要关注了人类学中的若干议题，如常人方法学、民族志，这些独特的研究方法深化了当代人类学研究。除此之外，该部分还重点突出了解释的方法研究，如诠释学研究、现象学研究，这些研究方法本身就是当代社会科学方法论的重要组成部分，因而具有重要的研究价值。第四部分各章的共同特征在于对理性和规范性的强调，当然它们的出发点和切入视角也各不相同，如相对主义和历史主义、非理性问题、实践理论、语言翻译、自然主义，相对来说前两篇更加关注理性，而后三篇则更突出规范性本身。第五部分的四篇论文都基于一种批判的认识态度和方法，其

观察视角分别聚焦于性别、种族、阶级等特殊群体上，表征为女性主义的批判观、族群批判观、社会建构的批判观，所有这些批判方法从不同的视角对当代社会科学进行了细致而深入的剖析。

从讨论的内容来看，该书主要关注生成于人类学和社会学内部实践的哲学问题，既包括了社会科学哲学中的传统问题，也包括这些学科中的具体问题。很显然，两位编撰者既关注到目前正处于争论中的问题构成及其过往历史，也为这些议题的未来探讨搭建了良好的讨论平台。

三

整套《爱思唯尔科学哲学手册》始终在强调科学与哲学的紧密联系，其中涉及的案例大都聚焦于自然科学和社会科学研究的最前沿，丛书的宗旨便是为这些前沿问题提供一种最新、最深入的哲学分析，从而在整体上描绘当前科学哲学的研究面貌及其未来走向。因此，学科专业性与问题前沿性便是本丛书的基本特点，它在为读者带来一场思想盛宴的同时，也给翻译造成了相当大的难度。山西大学科学技术哲学研究中心作为丛书翻译工作的独家承担者，在时间短、工作量大的情况下，为翻译好这套丛书，起用了中心几乎全部的人力资源，将翻译工程列为山西大学科学技术哲学研究中心工作的第一要务。鉴于丛书涉及学科面十分广泛，问题艰深、前沿，容量巨大，需要许多熟悉这些领域的研究人员协力完成，因此，在整个翻译期间，中心的大部分专职研究人员以及部分博士后、博士研究生全程投入到这项艰巨的任务当中，形成老中青搭配的全方位人员架构，其中青年研究人员无疑占到了多数，通过在工作上取长补短，力求将这套巨著准确、流畅地呈现出来。丛书各分卷主要分工如下:

《一般科学哲学: 焦点主题》: 郭贵春等译，蔺雅婷、郑文琦、孙静静校译。

《物理学哲学》: 程瑞、赵丹、王凯宁、李继堂译，黄玉梅校译。

《生物学哲学》: 赵斌译，于晓皖校译。

《数学哲学》: 康仕慧译，梁芳校译。

《逻辑哲学》: 刘杰、郭建萍译，曹洪亮校译。

《信息哲学》：殷杰、原志宏、刘扬弃译，刘小雯、隋先凯、肖侃、孙兴国校译。

《技术与工程科学哲学》：张培富等译，刘小雯、梅洁、张舒婷、胡雅楠校译。

《心理学与认知科学哲学》：王姝彦译，郅明刚、杨平达校译。

《人类学与社会学哲学》：尤洋译，陈晃、李睿一、雷婷、刘兴驰、李炎新、任其然、张经纬校译。

自2010年1月山西大学科学技术哲学研究中心正式启动首批《爱思唯尔科学哲学手册》9卷的翻译工作以来，历时五年多，各位翻译人员克服种种困难，兢兢业业地履行着自己的职责，在此对他们所付出的努力表示谢意。此外，负责中心行政工作的诸位老师对本项工作也给予了充分的后勤支持，尽最大可能地为翻译顺利进行提供便利，给翻译人员们创造了一个良好的工作环境，在此专门致以感谢。在山西大学科学技术哲学研究中心全体同仁的共同努力之下，我们很高兴丛书16卷中的9卷能够如期顺利出版。如果条件成熟，我们将筹备展开后续7卷的翻译工作，使之早日与广大读者见面。由于本次翻译工程时间紧迫，翻译和协调难度大，难免在一些方面会不尽如人意，辜负广大读者的殷切之心，如有不当之处，还望众位专家学者不吝指教。最后，这里还要特别对北京师范大学出版社表示衷心的感谢，尤其要感谢学术著作与大众读物分社社长饶涛先生，若不是他的远见卓识、对于新兴思想进行引介的魄力以及对学术求真务实的态度，便不会有此次翻译工程，《爱思唯尔科学哲学手册》也就没有机会被列入“十二五”时期(2011—2015年)国家重点图书出版规划和国家出版基金资助项目(2015年)。北京师范大学出版社对此次翻译工程异常重视，专门选派了精良的编辑团队，他们的认真敬业与专业高效给我们留下了深刻的印象，我们谨代表全体翻译人员在这里向他们表达最诚挚的敬意。总之，希望我们的努力工作最终能够换来广大读者的肯定，以绵薄之力推动国内科学技术哲学事业蓬勃向前。

郭贵春　　殷　杰

2015年10月

总　序

道・加比
保罗・撒加德
约翰・伍兹

当科学沿着已知世界的前沿发展时，不可避免地会触及关于知识和实在性质的哲学问题。科学争论提出，诸如理论与实验的关系、解释的性质以及科学能够接近真理到何种程度等此类问题。在具体科学中，总是有何物存在及其如何被认识的特殊问题被提出。例如，物理学中关于时空的性质，以及心理学中关于意识的性质。因此，科学哲学是对世界进行科学探究的基础。

在最近几十年，科学哲学越来越成为一般哲学中居于核心地位的学科。尽管仍有哲学家认为关于知识和实在的理论能够通过纯粹反思得以发展，但更多的当下哲学工作发现，将相关的科学成果纳入考虑范围是必要和有价值的。例如，心灵哲学现在更紧密地与经验心理学联系起来，而政治理论则经常与经济学相互交叉。科学哲学就是这样在哲学和科学探索之间架设了有价值的桥梁。

科学哲学不仅关注科学的性质及其有效性的一般问题，也越来越关注具体科学中出现的一些特殊问题。因此，我们组织编撰了这套多卷本的手册，以反映当代科学哲学研究的全貌。我们邀请了对具体科学深有研究的分册编辑，非常高兴他们征集到了由熟悉科学的哲学家们和(若干)了解哲学的科学家们的

稿件。其结果就是这一科学哲学目前所能提供的最具综合性的研究。

本手册的各分卷如下:

《一般科学哲学: 焦点主题》(*General Philosophy of Science: Focal Issues*) ,西奥·A. F. 库珀斯(Theo A. F. Kuipers) 编;

《物理学哲学》(*Philosophy of Physics*) ,杰里米·巴特菲尔德(Jeremy Butterfield) 、约翰·厄尔曼(John Earman) 编;

《生物学哲学》(*Philosophy of Biology*) ,莫汉·马修(Mohar Matthen) 、克里斯托弗·斯蒂芬斯(Christopher Stephens) 编;

《数学哲学》(*Philosophy of Mathematics*) ,安德鲁·欧文(Andrew Irvine) 编;

《逻辑哲学》(*Philosophy of Logic*) ,戴尔·杰凯特(Dale Jacquatte) 编;

《信息哲学》(*Philosophy of Information*) ,彼得·阿德里安斯(Pieter Adriaans) 、约翰·范·本瑟姆(Johan van Benthem) 编;

《技术与工程科学哲学》(*Philosophy of Technology and Engineering Sciences*) ,安东尼·梅杰斯(Anthonie Meijers) 编;

《心理学与认知科学哲学》(*Philosophy of Psychology and Cognitive Science*) ,保罗·撒加德(Paul Thagand) 编;

《人类学与社会学哲学》(*Philosophy of Anthropology and Sociology*) ,斯蒂芬·特纳(Stephen Turner) 、马克·瑞斯乔德(Mark Risjord) 编;

《复杂系统哲学》(*Philosophy of Complex Systems*) ,克利夫·胡克(Cliff Hooker) 、约翰·科利尔(John Collier) 编;

《统计学哲学》(*Philosophy of Statistics*) ,普拉桑塔·S. 班迪奥帕德西亚(Prasanta S. Bandyopadhyay) 、马尔科姆·福斯特(Malcolm Forster) 编;

《经济学哲学》(*Philosophy of Economics*) ,乌斯凯利·玛吉(Uskali Mäki) 编;

《医学哲学》(*Philosophy of Medicine*) ,弗雷德·吉福德(Fred Gifford) 编。

《地球系统科学哲学》(*Philosophy of Earth Systems Science*) ,布劳森·布朗(Bryson Brown) 、肯特·皮考克(Kent Peacock) 编;

《化学与药物学哲学》(*Philosophy of Chemistry and Pharmacology*) ,安德里

亚·伍迪(Andrea Woody)、罗宾·亨德里(Robin Hendry)编;

《语言学哲学》(*Philosophy of Linguistics*),马丁·斯托克霍夫(Martin Stokhof)、杰伦·格罗尼迪杰克(Jeroen Groenendijk)编。

各分册内容及出版计划的细节见:http://www.johnwoods.ca/HPS/。

作为主编,我们非常感谢各分册编辑组织如此杰出的写作队伍,并编辑出版了其稿件。这些分册的诞生是一项宏大的事业,我们衷心地感谢简·斯珀尔(Jane Spurr)和卡罗尔·伍兹(Carol Woods)。同样感谢爱思唯尔的安迪·迪林(Andy Deelen)和阿尔扬·赛文斯特(Arjen Sevenster)的支持和指导。

引言

杰里米·巴特菲尔德
约翰·厄尔曼

1. 当代物理学哲学

在过去的40年里，物理学哲学已经成为哲学的一个庞大且富有生机的分支，足以在一系列科学哲学手册中赢得其位置。不难理解物理学哲学为何充满活力，正如我们所看到的，有两个主要的原因——一是与所分析的科学哲学的形成期相关，二则与最近40年的研究有关。

首先，物理学对哲学分析运动的早期阶段有着巨大的影响。不论是对于逻辑实证主义者、逻辑经验主义者还是罗素(Russell)等人而言，该影响不仅反映了物理学表征了经验知识的一种范式这样一个事实，而且还存在一些更加具体的影响。现代物理学的三个主要支柱——热物理学(thermal physics)、量子理论(quantum theory)与相对论(relativity)——每一个都为哲学争论贡献了具体的理念与依据。在这些影响之中，较为显著的影响如下所述。

热物理学与关于原子存在性的科学争论，涉及了实在论与工具主义间的哲学争论，统计力学的兴起则催生了概率哲学。至于量子理论，它在哲学中最为普遍的影响无疑是使得哲学家

们承认基本的物理学理论可以是非决定论的。但这一影响仍然有待商榷，因为正如科学哲学家们都知道的(或应该知道的!)那样，非决定论仅仅触及了量子理论最具争议的那部分，即所谓的“波包塌缩”。无论如何，量子理论解释的模糊性不仅使哲学家们、也使像爱因斯坦(Einstein)和玻尔(Bohr)这样的物理学巨人陷入了激烈的争论。这些争论不仅与决定论有关，同时也涉及别的哲学基础，如客观性的本性。最后，相对论(包括狭义的和广义的)革新了时空哲学，特别体现在对新康德主义关于几何本性的学说的威胁。

这些影响意味着，当分析运动成为英美哲学的主流时，对现代物理学的解释就作为其分支学科——科学哲学——的突出主题被确定下来。相应地，随着哲学的发展，物理学哲学也在发展。

但自1960年以来，物理学哲学的发展受到了某种哲学之外的因素的影响。
xiv 也就是说，在物理学自身内部也存在着相当多的对基础问题的兴趣，其结果是给哲学带来了许多启发性的反响。再者，物理学内部的诸多进展对哲学有很多不同的影响。我们相信，最终的结果是：当前基本物理学理论中的基础问题会为物理学哲学提供最为有趣且最为重要的问题。本书论题的选择遵从了这一信念。下一节中，我们将澄清这些基础问题中的一部分，同时介绍本书中的各章。

2. 当前物理学中的基础问题

我们将首先在五个标题下讨论这些问题。前三个标题对应于第1节中提到的现代物理学的三大支柱：热物理学、量子理论和相对论；第四和第五个标题关注这些理论的结合，并引向对物理学未来的思索。这五个标题介绍了本书中的绝大多数章节——尽管没有按照章节的顺序进行。在这五个标题之后，我们将介绍本书中的其他两章。

2.1 热物理学

对热物理学基础的争论，特别是对平衡途径的描述，自该领域的创立者——如麦克斯韦(Maxwell)与玻耳兹曼(Boltzmann)——那时起，就未曾衰退

过。最初争论的一些方面在现代的讨论中仍然可以看到。但是，由于某些科学领域的发展，尤其是1960年以来下面三个领域的巨大进展，争论也发生了转换:

(1) 经典力学及其分支理论，如各态历经理论与混沌理论;

(2) 量子热物理学;

(3) 宇宙学，它为当前推进与评价玻耳兹曼的大胆具体设想——“时间之箭”的最终源头是宇宙学的——提供了一个非常具体且富有成效的背景。

在本书中，尤菲克(Uffink)和埃姆什(Emch)撰写的章节描绘了热物理学的基础，他们分别涵盖了热物理学基础的经典和量子两方面。在尤菲克所讨论的论题中，有两个论题受到了特别关注: 玻耳兹曼观点的演变和随机动力学的数学框架。埃姆什采用了代数的量子统计力学形式体系，并评论了关于此体系中平衡概念(即KMS态)的许多结果。另外两章——巴特菲尔德(Butterfield)关于经典力学的一章和埃利斯(Ellis)关于宇宙学的一章——为尤菲克和埃姆什提供了一些背景，尽管他们没有进一步探讨这些理论与热物理学的关系。

2.2 量子理论 xv

自20世纪60年代以来，物理学家团体见证了关于量子理论解释争论的复苏，这些争论曾一度在量子理论的创立者中风行。在一般的物理学家团体中，最有影响力的一个学者无疑是约翰·贝尔(John Bell)，这不仅是因为他的非定域性定理及其引发的诸多实验，也因为他对“哥本哈根正统学说”的批判以及他对波导与动力学塌缩非正统学说的同情。但是在更多的专家团体看来，还存在一些激活争论的其他关键性因素。数学物理学家们已经对量子与经典理论间的各种关系有了深刻的理解。自20世纪70年代以来，对退相干(Decoherence)的理解取得了进展，这使得目前几乎所有人都承认退相干在从量子理论中突现经典世界起着至关重要的作用。并且自20世纪90年代以来，已经从解释性的争论，尤其是对量子非定域性的分析中产生出了量子信息与量子计算这些发展迅速的领域。

在本书中，迪克森(Dickson)、兰兹曼(Landsman)和巴布(Bub)讨论了这些论题。迪克森回顾了非相对论的量子理论形式体系和一些主要的解释问题，

包括经验内容、量子不确定性、测量问题以及非定域性。大体而言，兰兹曼从数学物理学的视角回顾了量子与经典理论之间的关系。特别是，他讨论了量子化的不同途径和对经典极限 $\hbar \to 0$ 与 $N \to \infty$ 的严格处理，然而，兰兹曼也讨论了哥本哈根解释和退相干。巴布介绍了关于量子信息和量子计算的一些核心思想和结果，作为背景，他也简要回顾了经典的信息和计算，并在最后给出了一些关于量子理论解释的有挑战性的准则。

2.3 相对论

在 20 世纪 60 年代以来的几十年里，广义相对论和宇宙学在理论与实验两方面都取得了惊人的发展。而在基础性问题与哲学问题上这种复兴也卓有成效。数学的相对论者在不断深化我们对广义相对论基础的理解。正如在第 1 节中看到的，早在 20 世纪 20 年代就已被意识到广义相对论的基础对时空哲学至关重要。近年来，宇宙学在很大程度上从具有猜测性的事业向真正的科学转变，这不仅带来了与科学革命相近的多种哲学问题，也使得其他哲学问题更加紧迫，例如关于宇宙学中的方法和说明等问题。

在本书中，马拉蒙特(Malament)、贝洛特(Belot) 和埃利斯所撰写的章节阐述了这些论题。马拉蒙特首先解释了经典相对论，然后他讨论了三个具体论题：狭义相对论中同时性的定义、牛顿引力的几何化以及因果结构决定时空几
xvi 何的程度。贝洛特的主要目的是对经典广义相对论中出现的“时间难题”给出一个清晰的陈述。为了做到这一点，他首先回顾了在更简单的经典理论(包括力学) 中时间的表征方式。(因而贝洛特的这章补充了巴特菲尔德的那章：不仅解释了经典哈密顿理论的各个方面，也强调了其中的某些方面是如何在量子理论中再现的。) 埃利斯首先回顾了相对论宇宙理论的现状及其观测基础，然后研究了九个哲学主题——包括人择原理(anthropic principle) 和多重宇宙存在的可能性。

对三大物理学支柱(热物理学、量子理论和相对论) 中产生的一些基础性问题和本书中相应章节的介绍就到这里。我们转向由这些支柱——或不如说由其部分——结合而产生的问题！我们已经勾勒出前两个支柱的结合，即由埃姆什回顾的量子热物理学。在这里我们必须提及的是后两个支柱——量子理论和相

对论——的结合。我们将在分别对应于狭义相对论和广义相对论的两个标题下讨论该问题。当然，前一个标题对应于量子场论，它构成了粒子物理学深刻且成熟的理论框架；后一个标题对应于引力的量子理论，不幸的是它仍然只是一个希望和目标。①

2.4 量子场论

尽管有确定粒子数的相对论量子力学理论，但到目前为止，结合了量子理论与狭义相对论的最重要的理论框架仍然是量子场论。泛泛而言，量子场论的基础问题不同于量子理论的传统解释问题，如测量与非定域性(参见上文2.2节，量子理论)。这里有两点区别:

(1)尽管量子场论与基础量子理论一样阐明了传统的解释问题，但它显然不能为这些问题提供解决方案。测量问题和非定域性的谜题直接来源于量子理论的幺正性(unitarity)和张量积(tensor-product)特征，这并不受量子场论中引入的外部数学结构与物理理念的影响。② 相应地，对大多数人来说，在非相对 xvii
论量子理论中追问传统的解释问题似乎是明智的：如果你在一个简单的背景中确定了一个问题，又确信它不是由于背景的简单性而产生的，当然明智的选择就是在那里解决它。(正如本书中迪克森和兰兹曼的章节中所表明的，那个背

① 我们关于三大支柱的图像引起了这样的问题：热物理学和相对论的结合是什么？在爱因斯坦的狭义相对论理论获得了认可之后，当务之急是修改经典物理学的不同分支，以使得它们能够正确地相对论化。在热动力学情形中，该方案产生了关于热、温度和熵这些热动力学量洛伦兹变换特性的争论，这种争论一直持续到20世纪70年代，关于该争论的综述见[Liu(刘)，1994]。对于经典的广义相对论理论，现在不存在包含“宇宙的引力熵”的统计力学，且似乎也不可能有这样的理论。但众所皆知，热物理学的理念可能会在所期望的引力量子理论中扮演一个关键的角色。例如，在黑洞热动力学、盎鲁效应(Unruh effect)和霍金辐射(Hawking radiation)中有一些迹象。这些论题在罗韦利(Rovelli)的章节中有简短的讨论。

② 从某些视角来看，相对论的量子场论使得测量问题变得更棘手。在非相对论量子力学中，态向量的塌缩被假定是瞬间发生的；因而，在相对论情形中，人们也需要提出某种恰当的类似情形。另一方面，普通量子力学的模态解释——它被论证为最有希望为量子测量的非塌缩提供说明——在相对论量子场论中面对着无法克服的障碍，见[Clifton(克里夫顿)，2000]和本书中霍尔沃森(Halvorson)和缪格(Müger)所撰第八章的第5节。

景绝不是“真正简单的”：非相对论的量子理论以及它与经典理论的关系，提供了大量用以研究的错综复杂的结构。)

(2) 另一方面，存在许多量子场论特有的基础性问题。或许最显著的问题包括：粒子的本性(包括局域化论题)，重整化(renormalization)的解释，规范结构的解释，正则对易关系(canonical commutation relation)的幺正等价表示的存在性。

在本书中，特霍夫特('t Hooft)和霍尔沃森与缪格讨论了这些论题。首先，特霍夫特从粒子物理学的视角对量子场论进行了权威性的纵览。论题中阐述了以下问题：标量场和旋量场的量子化、费曼路径积分(Feynman path integrals)、规范场与希格斯机制(Higgs mechanism)的理念、重整化、渐近自由和禁闭。

霍尔沃森和缪格借助代数量子场论工具，讨论了一组小而独特的基础性问题(所以他们对代数方法的使用与埃姆什和兰兹曼二人的应用相辅相成)。他们讨论了粒子和定域化的本性、非定域性、量的赋值(即测量问题)和单个时空点上量子场的可定义性。但他们把更多的精力都投入到了超选的多普利克尔—哈格—罗伯茨理论(Doplicher-Haag-Roberts theory)中。该理论为量子场论的关键结构——尤其是表征集、场与可观测量代数间的关系以及规范群——提供了深刻的认识。

2.5 量子引力

最后，我们转向量子理论与广义相对论的结合，即寻找引力的量子理论。这里当然不存在成熟的理论，甚至关于建构一个成熟理论的最好方式也没有达成一致意见。相反，存在着多种研究方案，这些方案在技术目标、动机和概念
xviii 框架上各不相同。在这种情形下，关于“组分”理论的各种基本问题就有了新的理解。例如，或许量子引力会否定正统量子理论的幺正性，因而顺便也解决量子测量问题。广义相对论中的广义协变性，即微分同胚不变性(diffeomorphism invariance)是否提供了时空的终极量子本性方面的重要线索?

在本书中，罗韦利讨论了上述的相关问题。他同样也介绍了其他论题的具体情况：例如该主题的历史、目前的两种主要方案(弦理论和圈量子引力)以及量子宇宙学。埃利斯的那一章仍然是对量子宇宙学的讨论。这样一来，再加上

他确实提出了关于“终极”物理理论观念的其他基础问题，故他这一章很自然是对罗韦利那一章的补充。

上述介绍了对应于现代物理学三大基础支柱或者它们互相结合的几个章节。下面我们将转向对本书余下两章的介绍。我们的意图是提供这样的章节：其讨论的内容能架起连接物理学理论间(甚至三大支柱间)鸿沟的桥梁。在我们看来，对于这样一个跨领域的讨论而言，这些连接各种可能的主题中最恰当的两个就是决定论和对称性。①

相应地，厄尔曼(Earman)讨论了决定论在多种理论中的情形，他的案例涵盖范围包括从经典力学到量子引力的提法。他也提到了决定论和其他问题的关系：尤其是可预言性、时空的本性以及对称性。在经典物理学中的对称性方面，布拉丁(Brading)和卡斯特拉尼(Castellani)进行了全面的回顾。在其他论题方面，他们讨论了以下内容：居里原理(Curie's principle)、群理论进入物理学的过程、正则变换理论、广义相对论的广义协变性以及诺特定理。迪克森、埃姆什、霍尔沃森和兰兹曼这些人的章节中也讨论了量子物理学中对称性与不变性的各个方面。尽管如此，关于这个复杂的论题仍缺少一个概要性的综述，我们希望将这个任务留给读者。

我们用两段评论来总结一下对各章的介绍，从而给那些可能因本书的巨大篇幅望而却步的读者一些勇气。

首先，在我们看来，物理学哲学与物理学之间很显然并不存在清晰的界线。所以，不必惊诧于物理学哲学中一些最高水平的工作是由物理学家们完成的(这可以从本书的部分内容得到印证)。对此我们不仅不必感到惊奇，反而应该乐于接受这一事实。相反地，对于受过传统训练的哲学家们来说，物理学哲学家们的工作看起来更像是物理学，而不是哲学。但对我们来说，也不需要为学科间模糊的界线而担忧。相反，它给哲学一个充实自身的机会。并且反过来看，哲学家们希望物理学的基础、甚至物理学哲学能够成为物理学启发性思想 xix
的源头，或至少是，当前物理学家们在基础问题上的兴趣为物理学哲学家们与

① 其他好的论题包括时间之箭或不可逆性(irreversibility)和物质的构成。但加上有关这些或其他跨领域论题的章节会使得本书过长。

物理学家们进行卓有成效的讨论提供了机会。

但是我们承认对哲学的丰富没那么容易。专业资料的掌握通常很困难，于是对此的需求就会成为进入物理学哲学的屏障。在规划本书时，我们对该问题的回应当然不是试图以牺牲学术性并给出虚幻的希望为代价降低门槛，恰恰相反，我们的策略是让每章都尽可能地专业并且完整地囊括他们所选择的论题。所以，对于读者，我们只能说：用心！一旦你越过屏障，就会开启科学哲学新的图景。

3. 展望：丛林穿越的中途

最后，我们想以对基础物理学现状做出两个相关评论的方式来为本书的展开作铺垫。虽然对于哲学家而言，做这样的评论可能显得有些天真或自负，但我们相信值得冒这个险。因为我们认为，当前基础物理学的结合对来自哲学反思的贡献异常地开放，从我们的评论中可以很明显地看到，它们在共同邀请读者做出这样的贡献！第一个评论是关于当代物理学的惊人成功，第二个评论则指出仍存在许多需要理解的地方这一事实。

3.1 成功之处

首先，我们要庆贺近代物理学非同寻常的成功，特别是量子理论和相对论。我们这样做是要强调：相对论和量子理论的基本假设在其适用范围中被证明是如此地成功，这远远超越了他们最初的预期，这是多么偶然却也确实令人惊诧的事实。

有许多例子，我们随机挑选两个。为什么 1905 年爱因斯坦的狭义相对论为电磁学引入的新时间几何学可以推广到力学、热力学和其他物理学领域中？为什么为原子尺度(10^{-8}cm)系统创立的量子理论可以同样出色地用于更小的(相当于钙原子核半径 10^{-12}cm)和更大的(与超导体与超流体尺寸相当，大到 10^{-1}cm 量级)尺度？事实上，20 世纪的物理学史基本上就是相对论与量子革命的合并史，即它们的基本假设被成功且更广泛地应用的历史。

当我们考虑超出地球物理学的范围时，这一点也运用得同样出色。首先我

们来看广义相对论。它成就了一个精彩的故事：主要由一个人创立、受概念性考虑(其中部分是纯哲学的考虑)而激发的理论在各种各样的天文学情形中被实验精确证实。这些情形从太阳系中的弱引力场——这里广义相对论很好地解释 xx
了水星近日点的微小进动(每世纪43角秒)，这在牛顿理论中没有得到解释；到远距离双脉冲星中10 000倍强的场——近二十年来这为被广义相对论所预言并长期寻找的现象(引力辐射)提供了有力的证据(虽然是间接的)，再到奇异天体，比如黑洞。但并非只有广义相对论是这样，将量子理论应用到天文学中也非常成功，显著的例子就是，运用核物理学提出精确而详尽的关于宇宙早期的核合成、恒星结构及其演化的理论。

事实上，在相对论和量子理论的成功之外，也存在着更为一般的情形，即我们尝试去习惯科学所揭示的自然中的各种各样的统一性——以至于忘了科学理论是多么的偶然和令人吃惊。当然，并不只是我们这个时代才具有这种倾向性。例如，19世纪的物理学确证了牛顿引力定律也适用于太阳系以外，并(通过光谱学)发现地球上的元素在星系中也存在。这些发现当时看来很不可思议，但很快就被接受并纳入受教育人群的“常识”中。现在的情形也类似，近几十年物理学的许多不同的成就，都揭示了自然的统一性。这些成就为(i)时间和空间上遥远的距离尺度，与/或(ii)迥异于我们通常实验室尺度(如能量、温度和压力等量的特征量)上的现象建立了非常准确的模型。这些统一性的普遍理论的案例，即那些长达200年的案例有场概念的普遍有效性，更具体的有最小作用原理的普遍有效性。至于更为现代与具体的(也确实引人入胜的)案例，可以考虑我们关于超新星的精确而详尽的模型，因为可借助于现代望远镜技术出色的功能对单个(甚至是其他星系的)超新星做观测和分析来确证模型。

3.2 天边的乌云

然而，我们暂且不必自鸣得意，更不用说信心满满了！如今物理学充满着未竟之事——在人类的探究中总是如此。但更重要的一点是，天边尚有乌云。正如在19世纪末经典物理学面对的反常一样，这些乌云很有可能成为20世纪物理学持续成功的巨大威胁。

当然，对何种问题感到不安因人而异，并且在那些令人不安的问题中，人

们对于哪些是目前足以解决的或至少是值得解决的也有分歧。作为哲学家，我们是多面手，所以我们很自然地觉得上述提到的种种基本问题都令人不安。但作为多面手，我们当然会尝试避免去判断哪些最易于求解，或哪些最值得求
xxi 解！不论如何，这些判断很难裁定，因为一个人的知性气质以及知识背景或兴趣的偶然性，在这些裁定的形成中起着主要作用。

现在我们回到第2节中的一朵"乌云"——一朵明显需要哲学反思并且可能是哲学贡献的乌云——并以此结束引言。这朵"乌云"就是量子引力问题，或者说是广义相对论与量子理论需要重新调和这一事实。正如在2.5节中提到的，罗韦利讨论了"组分"理论截然不同的概念结构和对它们解释的持续争论是如何促成量子引力相互矛盾的基本方法的。

但在这里我们想强调在付出巨大努力和智慧之后却仍缺乏一个成功理论的另一个原因。简而言之，是相对论和量子理论的成功(已在上述评论3.1中展示过)共同剥夺了我们获取相关实验数据的机会。

因而，预期量子引力的特征性数据仅产生于能量如此之高(相应地，距离和时间如此之短)以至于对我们来说完全是不可实现的区域，这是合理的。用长度来表达则是：普朗克长度值大约是10^{-33} cm，我们希望这是量子引力的特征尺度。这确实是非常小的，原子、核子、质子和夸克的直径分别约为10^{-8} cm、10^{-12}cm、10^{-13} cm和10^{-16} cm。所以，普朗克长度到夸克直径(它的上限)的数量级与夸克直径到我们熟悉的厘米的数量级相同！

我们现在可以看出，量子引力研究是如何在某种意义上受到了相对论和量子理论成功的限制。因为那些成功提示我们，即使在我们可以达到的最高能量下也不会看到任何意味着量子引力的"新物理学"。最显著的例子就是类星体：它们的直径通常为几个光天，却有着我们银河系(本身跨越100 000光年，包含一千亿颗恒星)1 000倍的光度。它们是我们所观测到的能量最大的、距离最远的(因而是最早的)天体，现在认为它们的能量是由它们内核处大质量的黑洞提供的。30年前曾认为，它们惊人的能量和其他我们能够观测到的特性只能用全新的物理学来解释，而现在却做出了让步并开始承认描述黑洞视界之外情况的传统物理学也可以担当此任。(完全同意！我们期待物理学在黑洞内部的深处——其奇点附近——展现出量子引力效应。但若存在一个可以称为"不可及"

的区域，无疑就是此处!）因而该情形颇具讽刺意味，甚至令人沮丧：量子引力研究受到了其“组分”理论成功的限制。

无论如何，量子引力探究是完全开放的。在最后，我们想要赞成与罗韦利[1997]相类似的观点。他说我们现在的研究就好像是17世纪早期机械论哲学家如伽利略(Galileo)和开普勒(Kepler)的研究。正如他们与哥白尼(Copernicus)和布拉赫(Brahe)给出的线索作斗争，走在通往由牛顿综合的途中一样，我们也在“丛林穿越的中途”。当然我们应该避免对科学革命和不同历史情形间表观相似性的过于简单化和周期化。然而，从后来综合的视角来看，很显然伽利略和开普勒等人的学说原来就是个“大杂烩”。至于天才的他们，在我们(受益于跨越历史的后见之明)看来都是“过渡性人物”。人们禁不住猜测，对于20世纪物理学的某个未来读者而言，由于他受到未来某种将广义相对论和量子理论综合起来的理论的启发，我们现阶段在量子引力方面的努力似乎也很怪异：从我们作者的视角看这是有价值和有意义的(也是人们希望的)，但从那些今后的读者的角度看来这却是真知和谬误的大杂烩!

参考文献

[Clifton, 2000] R. Clifton. The modal interpretation of algebraic quantum field theory. *Physics Letters A*, *271*, 167-177, 2000.

[Liu, 1994] C. Liu. Is there a relativistic thermodynamics? A case study in the meaning of special relativity. *Studies in the History and Philosophy of Science*, *25*, 983-1004, 1994.

[Rovelli, 1997] C. Rovelli. Halfway through the woods. In J. Earman and J. Norton (Eds.), *The Cosmos of Science* (pp. 180-223). Pittsburgh: University of Pittsburgh Press and Konstanz: Universitäts Verlag, 1997.

撰稿人名录

戈登·贝洛特(Gordon Belot)

Department of Philosophy, University of Pittsburgh, 1001 Cathedral of Learning, Pittsburgh, PA 15260, USA.

gbelot@pitt.edu

凯瑟琳·布拉丁(Katherine Brading)

Department of Philosophy, University of Notre Dame Notre Dame, 100 Malloy Hall, Indiana 46556, USA.

kbrading@nd.edu

杰弗瑞·巴布(Jeffrey Bub)

Department of Philosophy, University of Maryland, College Park, MD 20742, USA.

jbub@umd.edu

杰里米·巴特菲尔德(Jeremy Butterfield)

Trinity College, Cambridge, CB2 1TQ, UK.

jb56@cus.cam.ac.uk

埃琳娜·卡斯特拉尼(Elena Castellani)

Department of Philosophy, University of Florence, via Bolognese 52, 50139, Firenze, Italy.

elena.castellani@unifi.it

迈克·迪克森(Michael Dickson)

Department of Philosophy, University of South Carolina, Columbia, SC 29208, USA.

dickson@sc.edu

约翰·厄尔曼(John Earman)

Department of History and Philosophy of Science, 1017 Cathedral of Learning, University of Pittsburgh, Pittsburgh, PA 15260, USA.

jearman+@pitt.edu

乔治·F.R. 埃利斯(George F. R. Ellis)

Mathematics Department, University of Cape Town, Rondebosch, Cape Town 8001, South Africa.

ellis@maths.uct.ac.za

杰拉德·埃姆什(Gérard Emch)

Department of Mathematics, University of Florida, 358 Little Hall, PO Box 118105, Gainesville, FL 32611-8105, USA.

gge@math.ufl.edu

汉斯·霍尔沃森(Hans Halvorson)

Department of Philosophy, Princeton University, Princeton, NJ 08544, USA.

hhalvors@princeton.edu

杰拉德·特霍夫特(Gerard 't Hooft)

Institute for Theoretical Physics, Utrecht University, Leuvenlaan 4, 3584 CC Utrecht, The Netherlands, and Spinoza Institute, Postbox 80.195, 3508 TD Utrecht, The Netherlands.

g.thooft@phys.uu.nl

N.P. 兰兹曼(N.P. Landsman)

Institute for Mathematics, Astrophysics, and Particle Physics, Radboud Universiteit Nijmegen, Toernooiveld 1, 6525 ED Nijmegen, The Netherlands.

landsman@math.ru.nl

大卫·B. 马拉蒙特(David B. Malament)

Department of Logic and Philosophy of Science, University of California at Irvine, 3151 Social Science Plaza, Irvine, CA 92697-5100, USA.

dmalamen@uci.edu

迈克·缪格(Michael Müger)

Institute for Mathematics, Astrophysics, and Particle Physics, Radboud Universiteit Nijmegen, Toernooiveld 1, 6525 ED Nijmegen, The Netherlands.

mueger@math.ru.nl

卡罗尔·罗韦利(Carlo Rovelli)

Centre de Physique Théorique de Luminy, Université de la Méditerranée, case 907, F-13288 Marseille, France.

rovelli@cpt.univ-mrs.fr

乔斯·尤菲克(Jos Uffink)

Institute for History and Foundations of Science, Utrecht Univesrsity, PO Box 80000, 3508 TA Utrecht, The Netherlands.

uffink@phys.uu.nl

目　录

第一章　经典力学中的辛约化

杰里米·巴特菲尔德

题词

经过了很长的历史和发展，现在经典力学(包括对基本问题的探索)的生命力是非常引人注目的。这种生命力产生于与纯数学理论(从拓扑和几何到群表示理论)之间丰富的相互影响，以及在诸如控制论这样的领域中新的和令人振奋的应用。而一些绝对基本点则更加值得注意，比如李代数对偶中李—泊松(Lie-Poisson)括号与力学中诸如刚体(rigid body)和理想流体等最基本的例子之间的清晰而又明确的联系。它们的完成用了将近一个世纪。

——马斯登和拉蒂乌(Marsden and Ratiu)[1999, pp. 431—432]

在刚体的一般理论中，可鉴别六种不同的三维空间：$\mathbb{R}^3$，$\mathbb{R}^{3*}$，$\mathfrak{g}$，$\mathfrak{g}^*$，TG_g，T^*G_g。

——阿诺德(Arnold)[1989, 324]

1. 引言

1.1 为什么会有经典力学?

所有人都为当代物理学的兴起而欢呼!在1890年到1930年之间，量子力学和相对论的革命以及在原子的发现中得以巩固的统计物理学，彻底改变了我们对于自然的理解，并且对哲学产生了巨大的影响，例如[Kragh(克拉格)，1999; Ryckman(里克曼)，2005]。相应地，本书专注于当代物理学中的三大支柱——量子理论、时空理论以及热物理学。因此关于经典理论，实际上是关于有限维系统的经典力学一章的结论的基本解释已经就绪了。

首先要说明的一点是，经典物理学的各种领域(比如力学和光学)非常之丰富且深刻。这不仅体现在它们的技术上，而且体现在它们对哲学和物理学基础的启示上。从牛顿时代开始，经典力学和光学已经引起了大量的哲学反思。至于力学，通常认为其核心哲学论题有空间、时间、决定论以及牛顿引力的超距作用本质。虽然它们很重要，但我不打算讨论这些话题，而是在其他一些章节中会提到(至少是部分地提到，有些时候是与除经典力学以外的其他理论相联系而提到的)。我将关注辛约化理论，它使连续对称和守恒量之间产生联系，这在诺特(Noether)“第一定理”中得到概括。我选择这个关注点，一方面是为其他一些章节做准备;另一方面是因为，正如我们立刻就会看到的，在经典力学目前的复兴及其与量子物理学的联系中，辛约化具有核心的作用。

我说过，经典物理学引发了许多哲学反思。有两个相互关联的原因值得强调，今天的哲学对量子和相对论革命的重视很容易使我们忘记这两个原因。

第一，在牛顿之后的两个世纪，经典力学的这些领域已经改变得让我们认不出来了，因此，对它们的哲学反思也发生了改变。回顾一下，在19世纪经典力学和光学是如何引起经典场论，特别是电磁场理论的。在本章的特定领域，有限维系统的经典力学中，考虑它的核心理论原理究竟是如何被诸如欧拉(Euler)、拉格朗日(Lagrange)、哈密顿(Hamilton)和雅可比(Jacobi)这些人物改写的。

第二，各种困扰着尝试严格形式化经典力学和光学的难题，其中的一些问

题带有可观的哲学层面的含义。并不是说只要我们不理会熟悉的话题——空间、时间、决定论和超距作用，经典力学的世界图景就是简单而直接的：仅仅是“运动中的物质”。相反地，即使我们只考虑有限维系统，我们也可以提出以下问题。

(i)对于质点(material points)：它们能否拥有不同的质量？如果可以，如何拥有？它们碰撞时会发生什么？实际上，对仅仅在牛顿引力作用下的质点相互作用而言，碰撞包含了无限的动能。

(ii)对于因其是刚体而被处理为有限维的延展物体：它们碰撞的时候会发生什么？刚性就意味着力和位移，且力在物体之间是“无限快”地传递的。这些难道不应该按字面意义理解吗？但如果是这样的话，用什么来证明这种理想的物体呢？而它的范围和界限又是什么呢？

关于无限维系统(弹性固体、流体和场)，它们理论的很多部分，特别是关于其严谨的形式和结果的部分，仍然是活跃的研究领域。对于目前关于弹性固体的工作，可参见[Marsden and Hughes(马斯登和休斯)，1982]。关于流体，其主要基本方程纳维叶—斯托克斯方程(Navier-Stokes equations)精确解的存在性和唯一性，仍然是一个未决问题。这个问题不只是与决定论明显相关，它还被认为在科学上具有重要的意义，如果有人解决了它，将会得到一百万美元的克雷千禧年大奖(Clay Millennium)奖金。

这两个原因——经典力学的连续重新形式化及其哲学问题——当然是相关的。经典力学丰碑式的人物认可并且争论这些问题，他们很多技术性工作的目标就是为了解决这些问题。结果是，牛顿的《原理》(*Principia*)[1687]问世之后两个世纪内出现了大量关于经典物理学、特别是关于力学基础的争论。一个著名的例子是迪昂(Duhem)的工具论科学哲学，这多半源于他认识到在微观水平下想要为经典力学的严格基础做担保是多么的困难。类似的例子是，受当时持续争论的力学基础工作所驱使，希尔伯特(Hilbert)将力学和概率的公理化问题选为其著名数学问题清单中的第六个。对这个清单的历史，参见[Grattan-Guinness(格拉顿—吉尼斯)，2000]。第三个例子贯穿两个世纪，关系到变分原理(variational principles)：最早由莫佩蒂(Maupertuis)，然后由欧拉和后人提出的最小作用量的各种原理——首先是对有限经典力学系统，然后针对无限系

统——引起了关于目的论的许多讨论。甚至，这个讨论吸引了逻辑经验论者[Stöltzner(斯托尔兹纳)，2003]，它还与当代模态哲学有关[Butterfield(巴特菲尔德)，2004]。

在20世纪上半叶，量子和相对论革命倾向于使物理学家进而使哲学家分散对这些以及类似问题的注意力。对发展新理论和争论它们对于自然哲学的含义的狂热，使得忽略经典力学基本问题的理论成为可理解的甚至是必然的。

而且，这种趋向由于教学的要求而被加强了，教学要求在物理学本科教育中必须要包含新理论。在20世纪中叶，学时对于物理学课程的限制导致许多物理学本科生的经典力学教育结束于分析力学，特别是有限维系统的分析力学的基本部分，例如，戈德斯坦(Goldstein)著名的教科书[1950]。这样的限制是可以理解的，因为：(i)拉格朗日和哈密顿方程的基本理论要求常微分方程的知识；(ii)基本哈密顿力学成为学习基本正则量子化(正如哈密顿—雅可比理论那样，从另一个角度看)的跳板。此外，正如我提到的，这个关于理论的约束甚至为哲学分析提供了丰富的材料——我上面的例子可以证明，而且诸如欧拉、拉格朗日、哈密顿和雅可比等的讨论也可以证明。

然而，20世纪下半叶可以看到经典力学研究的复兴——因此就出现了我第一个题词，有四个明显的原因：前两个是“学术的”，后两个是“实践的”。

(i)要部分感谢希尔伯特提出他的问题清单之后几十年里力学的发展。基本问题被重新捡起，在这件事上，数学家和具有数学思维的工程师与物理学家做得一样多。大多数相关的发展都在如拓扑学、微分几何学、测量理论和泛函分析等领域。在这次复兴中，一向以力学和概率见长的苏联学校的贡献无出其右。力学发展还与以下问题相关。

(ii)对加深量子理论特别是量子场论的形式化的诉求，促进了对经典力学结构以及量子化的探索。对于二者来说，特殊的兴趣都伴随着无限系统通常更加难懂的情形。

(iii)航天领域的出现，刺激了天体力学的发展。它与下一问题相关。

(iv)因计算机的发明而受到促进的非线性动力学(“混沌理论”)的研究。

由于这些不同的原因，这次复兴持续繁荣——并且因此我将避免对它进行更多的预示！甚至我将避免去探索那些因为从牛顿到雅可比和庞加莱

(Poincaré)的力学的多种多样形式而出现的哲学问题。可以这么说，对于上面提到的各种论题，我们可以加上如下两条。总的来说，第一条是本体论的，第二条是认识论的。

(a)对诸如质量和力的观念(包括它们如何随着时间而变化)的分析。对于这个话题，老一点的书，包括[Jammer(雅默)，1957；1961]和[McMullin(麦克马林)，1978]；新近出版的包括[Boudri(博瑞)，2002；Jammer，2000；Lutzen(吕琛)，2005]以及[Slovik(斯洛维克)，2002]；[Grattan-Guinness，2006]中有着很多参考文献，是一本关于这段历史的很好的纲要。

(b)关于对一个力学问题具有精确解是怎么回事的分析(包括精确解的观念是如何逐渐普及的)，这个话题是多方面的。它不仅与函数观念的逐渐普及相关，即数学史上的一个重要主题——[Lutzen，2003]有详尽的研究，还与现代非线性动力学相关，与探索一种对称性以便减少需要的变量数这一问题的约化相关——这就是辛约化的核心思想。我现在开始介绍它。

1.2　内容说明

通过寻找一种对称性来减少变量数目，从而简化力学问题，这种策略是经典力学的重大主题之一。它在理论上是深刻的，在实践上是重要的，且在这一主题的历史中经常出现。关于这一策略最著名的一般性定理无疑是诺特定理，它描述了连续对称和守恒量之间的一种对应。这个定理同时拥有拉格朗日和哈密顿两种版本，虽然由于历史原因，“诺特定理”这一名字更多地以拉格朗日版本出现。然而，我们只需要这一定理的哈密顿版本，它是我们探索辛约化的

出发点。①

因此，我将在 2.1 节中开始简短地回顾哈密顿版本。现在，我来做出四条注解（下述四条重要性按渐强顺序排列）。

（i）两个版本都由在基本的拉格朗日和哈密顿力学中关于循环（可忽略的）坐标及其对应的守恒动量的定律所支撑。②

（ii）事实上，这个定理的哈密顿版本更强一些。这反映了一个从“更大的”群出发的正则变换而不是点变换。更加确切地说：虽然位形空间 Q 上的点变换 $q \to q'$ 包括了 qs 和 ps 的相空间 Γ 上的正则变换 $q \to q'$ 和 $p \to p'$，但也存在其他可以用点变换所不包括的变换方式来“混合”qs 和 ps 的正则变换。

（iii）我将把我们的讨论限制在（a）时间独立的哈密顿量和（b）时间独立的变换中。大家都同意，分析力学在拉格朗日和哈密顿框架中都可以得到发展，同时允许时间独立的动力学和变换。例如，在拉格朗日框架中，允许速度独立的势和/或者时间独立的约束会提示我们运用通常被叫作“延伸的位形空间”的 $Q \times \mathbb{R}$。而在哈密顿框架中，时间独立会提示我们运用“延伸相空间”$\Gamma \times \mathbb{R}$。而且从哲学的观点看，考虑时间独立的变换是很重要的，因为它们包含了加速（boosts），这对于时空对称群，特别是对于相对性原理哲学的讨论非常关键。但是请小心，关于对称大致的陈述，例如哈密顿量在对称变换下必须是不变的，有可能无意中发现这些变换。举一个最简单的例子：一个自由粒子的哈密顿量只是它的动能，它可以通过变换到粒子的静止框架（rest frame）而成为零，也就是说，它在加速下并非不变。

---

① 对于拉格朗日版本的讨论，参见布拉丁（Brading）和卡斯特拉尼（Castellani）所撰写的本书第十三章或者（限制在有限维系统）[Butterfield，2004a，Section 4.7]。对于两个版本的阐述，参见[Butterfield，2006]。布拉丁和卡斯特拉尼也指出，即使在数学的分支中没有诺特定理，仍然存在其他的“诺特定理”出现在经典动力学的对称中；所以，现在的定理有时被称为诺特“第一定理”。我们注意到，辛结构也能在拉格朗日方程的经典解中看到，因此，辛结构可以在拉格朗日框架内发展，参见[Marsden and Ratiu，1999，10，Sections 7.2—7.5，and 13.5]。

② 这里我们看到我们主题的漫长历史：这些定理对于主题的发现者当然是明确的。在 1687 年，开发对称性来减少变量的数目就已经出现在开普勒问题（Kepler problem）的牛顿解中。对称性是平动和转动，对应的守恒量是线动量和角动量。在下文中，这些对称性和守恒量将为我们提供一些例子。

因此在哈密顿力学中完全处理对称，也就是辛约化，需要我们处理时间独立的变换—— 并且要小心！但是我不会去管这些复杂的因素。这里，必须不加任何细节地断言，辛约化的现代理论确实要处理加速，通常要处理时间独立的动力学和变换。

(iv)正如我们将要更详尽地看到的，辛约化理论推广诺特定理有三种主要方式。正如可能有人预期的那样，这三种方式之间紧密联系。

(a)在对每一种对称(相空间上的一类特殊的矢量场)提供一个守恒量，也就是其值在实践中保持常数的相空间上的一个实值函数的意义上，诺特定理是“一维”的。因此特别地，一个守恒矢量的不同分量，比如总线性动量，是分开处理的(在这个例子中，对应的矢量场在三个不同的空间方向上产生了平动)。但是在辛约化中，动量映射(momentum map)的观念提供了一种对这些不同分量的“统一”描述。

(b)给定一种对称，诺特定理使我们能够把注意力限制于守恒量的表面层次，也就是相空间的子流形(sub-manifold)，在其上守恒量获得其初值：因为系统的时间演化受限于这个表面。在这个意义上，我们考虑的变量数目减少了。但是在辛约化中，我们走得更远并且从相空间出发形成了一个商空间(quotient space)。用逻辑的语言来说，我们在相空间上定义了一个等价关系(通常来讲并不像拥有一个守恒量的公共值那样简单)并且形成了等价类的集合。用群作用量的语言来说就是，我们形成了轨道的集合。通往这个商空间的途径可以是出于各种各样技术良好的、甚至是哲学的动机。而且，在良好的条件下，这本身就是一个低维流形。

(c)哈密顿力学和所谓的诺特定理，通常是根据辛流形、特别是位形空间 Q 的余切丛 T^*Q 而形式化的(2.1 小节会详细展示)。但在辛约化中，我们常常会需要对一类叫作泊松流形(Poisson manifold)的辛流形思想进行(适度的)推广。在泊松流形中被当作原始概念的是一个具有泊松括号的某些特性的括号。而且这与(b)相关，因为我们经常通过采用由对称群的行为而产生的辛流形(即通常类型的一个相空间)的商来实现泊松流形及其之上的动力学。

正如注解(iv)提示的，辛约化是一个很大的课题。因此，可以找到多个研究它的动机。至于物理学，它的许多思想和结果都可以在有限维经典系统(我

自己将在此范围内讨论问题)中提出，但是之后就会推广到无限维系统。而不论是在原始的还是在推广的形式中，它们都支持了量子理论的发展。因此，这些思想和结果部分地形成了目前经典力学的复兴。参见 1.1 节最后的(i)和(ii)。

至于哲学，对称性既是哲学讨论中早就形成的，也是目前活跃的一个焦点，参见[Brading and Castellani，2003]。但是关于辛约化的哲学讨论似乎直到最近，特别是在贝洛特(Belot)和厄尔曼(Earman)的论文中才开始进行。可以假定这样一种滞后是因为技术资料太复杂，事实上，辛约化的理论只是到了 20 世纪 70 年代才具有了目前的一般形式。但是，正如贝洛特和厄尔曼强调的，为获得哲学上的收益学习这些技术是物有所值的。最明显的论点是，辛约化用来商化(quotienting)一个态空间的设备，使两个明显不同但又全然不可辨别的可能性是否应当被看作是同一的这一问题，成为受关注的哲学问题。在第 2 节中，我将跟随贝洛特的脚步去阐述“关系主义(relationist)”力学的问题。事实上，我已经为我的阐述选择了主题，这个阐述着眼于给哲学读者一些由贝洛特论述的背景。他的论文(我将在第 2 节引用)有许多明智的哲学观点，但又不让读者有学术性论述的重担。要充分重视这个问题中优秀的成果，当然必须努力去了解细节。

最后，在本书的上下文中，辛约化为这些章节提供了一些关于力学中时间表述的背景[Belot，本书，第二章]，还有关于经典和量子物理学关系的背景[Landsman(兰兹曼)，本书，第五章](特别是 4.3～4.5 节和 6.5 节)以及[Dickson(迪克森)，本书，第四章]。

该章的计划如下：在 2.1 节中，我首先回顾通常表述的哈密顿力学中的诺特定理；然后介绍有关上文提到的(b)和(c)的主题，商化一个相空间和泊松流形(2.2 节)；之后阐述对于“相对论”力学的主题(2.3 节)。

此后，我阐述辛约化的基础(只限于有限维度的哈密顿力学)。这个计划将逐步如下所示。第 3 和第 4 节回顾所需的现代几何，第 3 节主要是关于弗罗宾

尼斯定理，李代数和李群[1]，第4节阐述了李群的行为。结尾的中心思想是关于李代数的对偶$\mathfrak{g}^*$的李群G的余伴随表示。这一回顾使我们能够更好地理解泊松流形的动机(5.1节)，然后展示例子，再然后证明一些主要属性(5.2节开始)。第6节在力学体系的对称和守恒中应用这些材料。特别是，它给出了作为动量图景的守恒量，并证明关于泊松流形的哈密顿力学适用于诺特定理。最后，在第7节，我们证明关于辛约化的一些主要定理中的一个。它关注的情况是，对于一个系统的自然位形空间本身就是一个李群G：这发生在刚体和理想流体中。在这种情况下，商化自然相空间(关于G的余切丛)给出了一个泊松流形，这个泊松流形是G的李代数的对偶$\mathfrak{g}^*$。[2]

总之，我希望这个综述的所有作用能阐明这一章节的题词：经典力学是充满活力的，尤其是彻底深化我们对由来已久的诸如刚体系统的理解——而传统教科书中对此的分析实在令人困惑！

2. 辛约化：综述

我们首先在2.1节[3]中简要回顾哈密顿力学和诺特定理。2.2节为我们辛约化的观点作了准备。然后在2.3节中，我们阐明对“相对论”力学的使用。

2.1　哈密顿力学和诺特定理：回顾

2.1.1　辛流形，作为辛流形的余切丛

在流形M上的辛结构或辛形式是定义在闭合的(即其表面微商$\mathrm{d}\omega$消失)且非简并的M上的微分2形式ω。即对于任何$x \in M$，任何两个x上的切矢量σ，

---

① 它的前两小节也为马拉蒙特(Malament)提供了一些必要的内容。

② 在这个尝试下，我们的来源有四本书：[Abraham and Marsden(亚伯拉罕和马斯登)，1978；Arnold，1989；Marsden and Ratiu，1999；Olver(奥尔弗)，2000]。

③ 关于微分几何的更多细节，参见3.1和3.2节。关于力学的几何形式的更多细节，参见[Arnold，1989；Marsden and Ratiu，1999]，或[Singer(辛格)，2001](要比这个解释更基本)，或[Abraham and Marsden，1978]，或[Butterfield，2006](相同水平的)。对于力学参考书，我尤其推崇[Desloge(德洛热)，1982；(Johns)约翰斯，2005]。

$\tau \in T_x$：

$$\mathbf{d}\omega = 0 \quad 和 \quad \forall \tau \neq 0,\ \exists \sigma:\ \omega(\tau,\ \sigma) \neq 0 \tag{1}$$

这样一对$(M,\ \omega)$被称为辛流形。辛流形理论异常丰富，但我们只需要它的一个小片段。特别是，我们大多数时间避免正则变换理论的事实意味着我们不需要拉格朗日子流形理论。

首先，它是由M是偶维的ω的非简并产生的。原因在于线性代数原理，后来应用在M的每个点的切空间。即对于任何双线性形式ω：$V \times V \to \mathbb{R}$，如果$\omega$是反对称的，秩$r \leqslant m \equiv \dim(V)$，$r$是偶数。即：$r = 2n$，$n$为整数，并且有一系列$V$的基$e_1$，…，$e_i$，…，$e_m$，对于$V$，$\omega$作为楔形积(wedge-products)可简单地展开为：

$$\omega = \sum_{i=1}^{n} e^i \wedge e^{i+n} \tag{2}$$

对应地，ω是一个$m \times m$矩阵：

$$\omega = \begin{pmatrix} 0 & \mathbf{1} & 0 \\ -\mathbf{1} & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{3}$$

$\mathbf{1}$是$n \times n$的单位矩阵，类似于各种大小的零矩阵。这个反对称双线性规范化形式是格拉姆—施密特定理(Gram-Schmidt theorem)的类似物。这个定理认为一个内积空间有一个正则正交基得到了其他类似证据的证明。

所以，如果一个反对称双线性形式是非简并的，则$r \equiv 2n = m$。即式(3)失去由零矩阵组成的底行和右列，并减少到$2n \times 2n$的辛矩阵ω，由下式给出：

$$\omega := \begin{pmatrix} 0 & \mathbf{1} \\ -\mathbf{1} & 0 \end{pmatrix} \tag{4}$$

第二，非简并的ω意味着对所有$x \in M$，有一个独立于从切空间T_x到其共轭T_x^*的同构$\omega^\flat$的基。即：对于任意$x \in M$和$\tau \in T_x$，1形式的$\omega^\flat(\tau) \in T_x^*$由以下定义：

$$\omega^\flat(\tau)(\sigma) := \omega(\sigma,\ \tau) \quad \forall \sigma \in T_x \tag{5}$$

这也意味着辛结构能使一个余矢量场，即一个微分1形式，确定一个矢量场。因此对于任意函数H：$M \to \mathbb{R}$，dH是M的一个微分1形式，$\omega^\flat$的逆(我们可以写作$\omega^\sharp$)将dH带入关于M的矢量场，写作X_H。哈密顿力学的关键想法是

标量函数 H 决定动力学，参见 2.1.2 节。

到目前为止，我们已经注意到一些非简并的 ω 的蕴含式。辛形式（对于一个流形）定义的其他部分，也就是闭合的 ω（$\mathrm{d}\omega=0$）也很重要。我们将在 2.1.3 节看到，这意味着当且仅当 X 是在某种情况下的哈密顿量时，关于辛流形 M 的一个矢量场保留了辛形式 ω，即用更加物理的术语来说，生成（单参数族）正则变化，这种情况就是存在一个标量函数 f，使得 $X=X_f\equiv\omega^\sharp(df)$。或者对于一个标量函数，根据泊松括号，用 · 表示争论的地方：$X(\cdot)=X_f(\cdot)\equiv\{\cdot, f\}$。

用这么多的方式引入辛流形后，我转向说明任何余切丛 T^*Q 都是这样一个流形，即它有一个不依赖于坐标或者基的选择的辛结构。

给定一个我们认为的系统位形空间的流形 $Q(\dim(Q)=n)$，选择 Q 上任意局部坐标系 q 和在 T^*Q 上产生的自然局部坐标 q，p。我们定义 2 形式：

$$dp\wedge dq:=dp_i\wedge dq^i:=\sum_{i=1}^{n}dp_i\wedge dq^i \tag{6}$$

事实上，关于 Q 上的 q，不管我们做出任何选择，式(6)都定义了相同的 2 形式。因为 $dp\wedge dq$ 是关于 T^*Q 的 1 形式外导数，这个外导数自然而然地（独立于坐标或基）从投射的导数（也称为正切）映射定义而来：

$$\pi:\ (q,\ p)\in T^*Q\to q\in Q \tag{7}$$

因此，我们认为切矢量 τ，不是正切于 Q，而是正切于点 $\eta=(q,\ p)\in T^*Q$，即 $q\in Q$ 且 $p\in T_q^*$ 处的余切丛 T^*Q。让我们写为：$\tau\in T_\eta(T^*Q)\equiv T_{(q,p)}(T^*Q)$。自然投影 π 的外导数，如 $D\pi$，适用于 τ：

$$D\pi:\ \tau\in T_{(q,p)}(T^*Q)\mapsto(D\pi(\tau))\in T_q \tag{8}$$

现在，我们通过下式定义关于 T^*Q 的 1 形式 θ_H：

$$\theta_H:\ \tau\in T_{(q,p)}(T^*Q)\mapsto p(D\pi(\tau))\in\mathbb{R} \tag{9}$$

在 θ_H 的这个定义下，p 被定义为 τ 的基点 $(q,\ p)\in T^*Q$ 的第二分量，即 $\tau\in T_{(q,p)}(T^*Q)$ 且 $p\in T_q^*$。

这个 1 形式被称为关于 T^*Q 的正则 1 形式。对于任何自然局部坐标 q，p，θ_H 由下式给出：

$$\theta_H=p_i dq^i \tag{10}$$

最后，我们通过外导数 θ_H 定义了一个 2 形式：

$$\mathbf{d}(\theta_H) \coloneqq \mathbf{d}(p_i dq^i) \equiv dp_i \wedge dq^i \tag{11}$$

这个2形式是闭合的(因为 $\mathbf{d}^2=0$)且是非简并的。所以(T^*Q，$\mathbf{d}(\theta_H)$)是一种辛流形。因此，$\mathbf{d}(\theta_H)$或$-\mathbf{d}(\theta_H)$被称为正则辛形式，或正则2形式。

关于这一作用存在达布定理(Darboux's theorem)，即任何辛流形局部地“看起来像”一个余切丛，换句话说，余切丛局部地是辛结构的“通用”例子。我们不会详细地描述，但在5.3.4节，我们将讨论关于泊松流形的推广定理。首先我们要在接下来的两个小节回顾哈密顿方程和诺特定理。

2.1.2 哈密顿方程的几何形式

正如我们已经强调的，哈密顿方程背后的主要几何想法是，一个梯度场(也就是余矢量场)dH 确定矢量场 X_H。所以，在余切丛 T^*Q 的点 $x=(q, p)$ 处给出一个哈密顿方程的几何形式，假设我们把从余切空间到切空间的(有独立基的)同构写作 $\omega^\sharp$，通过 $\omega \coloneqq -\mathbf{d}(\theta_H)=dq^i \wedge dp_i$ 得到 $T_x^* \to T_x$(参见方程5)。然后哈密顿方程可以写成：

$$\dot{x}=X_H(x)=\omega^\sharp(\mathbf{d}H(x))=\omega^\sharp(dH(x)) \tag{12}$$

还有其他不同的形式。将 $\omega^\flat$，即反同构(inverse isomorphism)$T_x \to T_x^*$ 应用到两边，我们得到：

$$\omega^\flat X_H(x)=dH(x) \tag{13}$$

按照在 x 处的辛形式 ω，即：对于所有矢量 $\tau \in T_x$，有：

$$\omega(X_H(x), \tau)=dH(x)\cdot\tau \tag{14}$$

或者按照具有矢量场 X 的微分形式 α 的缩写式(也称为内积) $\mathbf{i}_X\alpha$，用“·”表示 $\tau \in T_x$ 有争议的地方：

$$\mathbf{i}_{X_H}\omega \coloneqq \omega(X_H(x), \cdot)=dH(x)(\cdot) \tag{15}$$

更简单地说，对于任意函数 f，它是：

$$\mathbf{i}_{X_f}\omega=df \tag{16}$$

最后，记得函数的泊松括号和方向导数之间的关系(或李导数 $\mathcal{L}$)，即：

$$\mathcal{L}_{X_f}g=dg(X_f)=X_f(g)=\{g, f\} \tag{17}$$

结合式(16)，我们可以定义辛形式和泊松括号之间的关系为下面的形式：

$$\{g, f\}=dg(X_f)=\mathbf{i}_{X_f}dg=\mathbf{i}_{X_f}(\mathbf{i}_{X_g}\omega)=\omega(X_g, X_f) \tag{18}$$

2.1.3 诺特定理

在拉格朗日和哈密顿框架内，诺特定理的核心理念是系统的每一个连续对称都对应一个守恒量(第一积分，运动的常数)。连续对称的理念精确地按照以下思路：对称是态空间的矢量场，它分别保留了(i)拉格朗日量(哈密顿量)和(ii)关于态空间的结构。

在哈密顿框架内，证明的核心是基于泊松括号是反对称的这样一个事实的“单行方式(one-liner)”。因此对于任何标量函数f和在辛流形(M, ω)上的H(且因伴随方程18给出的泊松括号)，我们在任意点$x \in M$有：

$$X_f(H)(x) \equiv \{H, f\}(x) = 0\text{，当且仅当 } 0 = \{f, H\}(x) \equiv X_H(f)(x) \quad (19)$$

用文字表述：在x附近，当且仅当f是在流(flow)X_H之下的常量时，H是在矢量场X_f的流(即如果f是哈密顿函数，流形成将会如何)之下的常量。考虑到H是物理哈密顿量，所以，X_H代表着真实的时间演化，有时称为动力流(dynamical flow)，这意味着在x附近，X_f保留了哈密顿函数当且仅当f是时间演化之下的常量，即f是一个守恒量(一个运动的常量)。

但我们需要注意上面的条款(ii)“关于”态空间结构的矢量场的观点。在哈密顿框架内，这是精确的辛形式。因此我们定义在辛流形(M, ω)中的矢量场X是辛的(也称为正则的)，当且仅当沿着辛形式的X的李导数等于零，也就是说$\mathcal{L}_X\omega = 0$。这个定义相当于X的生成(有效)正则变换，也相当于保留了泊松括号。但我不会详细地讨论这些等价形式，因为它们属于正则变换理论，正如前面提到的，我不需要再提及它们。

我们还将哈密顿系统定义为三元数组(M, ω, H)，其中(M, ω)是一种辛流形且$H: M \to \mathbb{R}$，即$M \in \mathcal{F}(M)$。然后我们将哈密顿系统的(连续的)对称定义为在M上的矢量场X：

(i)保留了哈密顿函数，$\mathcal{L}_X H = 0$；

(ii)保留了辛形式，$\mathcal{L}_X\omega = 0$。

这些定义意味着从式(19)证明诺特定理，足以证明仅当它是X_f局部的形式时矢量场X是辛的。这样的一个矢量场被叫作定域哈密顿量。(如果存在一个全局标量f: $M \to \mathbb{R}$使得$X = X_f$，那么这个矢量场就称为哈密顿量。)事实上，微分形式的两个结果，即庞加莱引理和嘉当(Cartan)公式，使得上述理论容易

被证明。对于在任意辛流形(M, ω)中的矢量场不必是余切丛。

我们把外导数写作 $\mathbf{d}$，我们记得，一个 k 形式 α 被称为：

(i)完整的，如果有$(k-1)$形式β，使得 $\alpha=\mathbf{d}\beta$(参见全微分的基本定义)；

(ii)闭合的，如果 $\mathbf{d}\alpha=0$。

庞加莱引理认为，每一个闭合的形式是局域精确的。更准确的说法是：对于 M 的任何开集 U，我们定义了 U 上的 k 形式场的矢量空间 $\Omega^k(U)$。然后，庞加莱引理认为，如果 $\alpha\in\Omega^k(M)$是闭合的，且在每一个 $x\in M$ 时都会有一个 U 临近它，以至于 $\alpha|_U\in\Omega^k(U)$是正确的。

嘉当公式是一个涉及李导数、缩写和外导数的很有用的公式(通过直接的计算证明)。如果 X 是一个矢量场且 α 是在流形 M 上的 k 形式，那么关于 X 的 α 的李导数(即沿着 X 流的)是：

$$\mathcal{L}_X\alpha=\mathbf{di}_X\alpha+\mathbf{i}_X\mathbf{d}\alpha \tag{20}$$

我们现在的看法如下。因为 ω 是闭合的，即 $\mathrm{d}\omega=0$，嘉当公式即式(20)应用到 ω 变成了：

$$\mathcal{L}_X\omega\equiv\mathbf{di}_X\omega+\mathbf{i}_X\mathbf{d}\omega=\mathbf{di}_X\omega \tag{21}$$

所以辛的 X 对于 $\mathbf{i}_X\omega$ 是闭合的。但通过庞加莱引理，如果 $\mathbf{i}_X\omega$ 是闭合的，它也是局域确切的。即局域地存在一个标量函数f：$M\to\mathbb{R}$，导致：

$$\mathbf{i}_X\omega=df,\ \text{即}\ X=X_f \tag{22}$$

所以，如果 X 是辛的相当于 X 是定域哈密顿系统。因此，有如下定理。

对于哈密顿系统的诺特定理 如果 X 是一个对称的哈密顿系统(M, ω, H)，对于哈密顿系统我们有诺特定理，且局域地 $X=X_f$，所以通过非对称的泊松括号，式(19)中的f是运动学常量。而对应地，如果f：$M\to\mathbb{R}$是运动学常量，那么 X_f 是对称的。

我们将在6.2节中看到，大多数类似诺特定理的方法，尤其是“单行方式”要求泊松括号是反对称性的，式(19)继续传到泊松流形的更一般的框架中。现在，我们来看一个例子(后面我们还会再次谈到)。

对于欧几里得(Euclid)空间$\mathbb{R}^3$ 的大多数哈密顿系统，空间平动和转动都是(连续的)对称的。尤其我们考虑到这样一个将在2.3节更详细讨论的系统：N 个点粒子通过牛顿万有引力相互作用。哈密顿量有两项，每一项在平动和转动

之下都是独立的不变量：

(i)动能项 K，尽管我不详细论述，但它实际上是通过欧氏度量$\mathbb{R}^3$定义的，因此它是不变量；

(ii)势能项 V，它只取决于粒子的相对距离，因此也是不变量。

对应的守恒量完全是线动量和角动量①。

2.2 前方的道路

在本节，我们将通过四段评论详述 1.2 节的介绍性评论(iv)，也将给出一些有关辛约化的信息和一些关键例子的信息。

(1) 来自诺特定理的推广泊松流形

诺特定理告诉我们，连续对称(即单参数对称群)确定第一个积分(即运动常数)。所以一个较大的对称群，即带有一些参数的群，意味着数个第一积分。因此，相流(phase flow)只限于这些积分的等位面的交叉点，一个交叉点通常就是一个流形。换句话说，这些积分的同位流形是相流的不变流形。

事实证明，在许多有用的情况下，这个流形在对称群精选的子群下也是一个不变量。商空间，即在子群作用下的轨道集合，是通过最初的哈密顿系统产生的自然结构，这个最初的哈密顿系统足够以哈密顿形式处理力学问题。商空间因此被称为“约化的相空间”。

但在某些情况下，这种自然结构不是一个辛形式，而是一个(温和的)概括，在这种概括下这个形式可以是简并的，即更像式(3)而不是式(4)。配置如此结构的流形不需要是商流形，反而它可以根据通常的泊松括号的概括来定义，正如根据式(18)的辛形式定义的那样。

核心的理念是假定一个括号，作用于标量函数 $F: M\to\mathbb{R}$在任意流形 M 上都成立，并假定这个括号通过通常的泊松括号拥有四个属性。其中的一个属性是反对称性，这个反对称性在 2.1.3 节的诺特定理的证明中得到强调。其他三

① 顺便说一下，这个哈密顿量在加速下并不是守恒量。但正如我在 1.2 节的(iii)中所说的，这局限于不随时间变化的转动中；“表示运动的相对性”的对称性的处理需要单独地讨论。

个是，被再次写作{,}的公设括号(postulated bracket)是双线性的，并且对于在M上任意的实函数F，G，H都遵循雅可比恒等式。即：

$$\{\{F, H\}, G\} + \{\{G, F\}, H\} + \{\{H, G\}, F\} = 0 \tag{23}$$

并且遵循对于乘积的莱布尼茨规则(Leibniz' rule)，即：

$$\{F, H \cdot G\} = \{F, H\} \cdot G + H \cdot \{F, G\} \tag{24}$$

我们将会在第5节看到，这样一个括号，也被称为"泊松括号"，为哈密顿形式的力学提供了充足的框架。特别是，它引起了一个形如式(3)中的反对称双线性形式，有如此括号的流形M被称为泊松流形。

简并意味着一个泊松流形可以有奇数(odd)维，虽然我们在2.1.1节看到任何辛流形都是偶维的。另一方面，这个广义的哈密顿力学将和2.1节中的一般构想密切相关。主要的关联在于如下结果：任意泊松流形M都是偶维流形的不相交并集，在这个并集上，M的简并反对称双线性形式只限于非简并的。①

(2)历史根源

这个辛约化理论在经典力学丰碑式人物的工作中具有深刻的历史根源。在某种程度上，这是意料之中的。正如在1.2节的(i)中提到的，循环坐标巩固了力学中对称的作用，尤其是诺特定理。并且，开普勒问题的牛顿解决方案提供了一个例子，见证了转移到质心坐标和伴随有循环角的极坐标(产生角动量守恒)的教科书般的解释。所以，各种辛约化的结果和想法能够在欧拉、拉格朗日、哈密顿、雅可比、李和庞加莱的工作中看到也就不足为奇了。例如(正如我们将看到的)刚体的欧拉理论。

但是历史也具有令人惊讶的地方。事实证明，李关于李群的划时代工作已经包含许多基本理论和现代理论的详细发展。② 让人伤心且具有讽刺意味的是，李的大多数见解并不被接受，且之后被多次反复发现。这是不学习历史的人注定要重蹈覆辙的另外一个例子(数学史上有许多!)。令人欣慰的是，通常重新发现某事物要比从其他人那里学更容易、更有趣……

---

① 由于这些清晰的联系，把更一般的框架称为"哈密顿量"是很自然的事情，就像通常做的那样。当然这只是口头上的事情。

② 主要的来源是他的作品[1890]。此外，阿诺德[1989，456]表示，泊松流形的标准例子，即有限维李代数的对偶，已经被雅可比所理解。

因此这一工作是从20世纪60年代中期才开始的，本质上是李提出的形式被重新找到并且形成了通过现代力学所采纳的几何学语言，也就是说，被诸如阿诺德、科斯坦特(Kostant)、马斯登、迈耶(Meyer)、斯梅尔(Smale)、苏里奥(Sourian)和温斯坦(Weinstein)这样的当代大师(参见本章的题词)所接纳。令人高兴的是，这些现代作者都是历史学者，他们的课本甚至给出了一些历史细节，参见[Marsden and Ratiu，1999，pp. 336—338；369—370；430—432]及奥尔弗作品每一章的注释[2000，pp. 427—428]。[Hawkins(霍金斯)，2000]给出从1869年到1926年的李群的完整历史，对于李群，特别参见其中第1.3节、2.5节和第3章，尤其是3.2节。

无论如何，撇去历史不谈，辛约化自20世纪70年代以来在现代力学中一直都是一个活跃的研究领域，并且联合了诸如辛几何的场。所以它现在作为当代经典力学复兴的主要部分已经得到了应有的位置，正如后文所言。

(3)两个例子：刚体和理想流体

两个例子生动阐述了辛约化可如何给出新的物理解释。这两个例子由来已久，即刚体和理想流体——像在这一章的题词里证实的那样。(在2.3节将展开第三个例子，该例和哲学更密切相关。)

对于刚体，我们将看到(尤其是在第5节)，辛约化主要阐明了刚体的初等理论(欧拉方程、欧拉角等)。众所周知，这曾是令人困惑的！为简单起见，我将假定刚体是转动的，以便忽略平动运动。这将意味着要定义商过程(quotienting procedure)的对称群将是转动群。它还将意味着，位形空间的刚体通过转动群给出，因为任何位形都可能被贴上转动的标签，转动可以从参考位形(reference-configuration)中得到这些位形。所以，在辛约化的这个应用中，对称群(也就是转动群)作为位形空间将对自身起到作用。这个例子还将给我们一些泊松流形的典型范例。

对于理想流体，即一种流体是不可压缩的且是不计黏性的(黏度为零)，这当然是一个无限维系统，但这(正如我在1.2节宣称的)在本章的讨论范围之外。所以我不会研究任何细节，只是说明主要思想。

这个理想流体的运动方程，欧拉方程，通常是应用牛顿第二定律$F=ma$到一个小的流体元素或者通过拉格朗日或哈密顿方法启发式的应用而得到的（如

启发式经典场论)。但在20世纪60年代中期，阿诺德认为后者的导数可以通过碰撞来理解，和上述的刚体的处理进行类比发现这种方法甚至是非常优美的。也就是说，这个类比表明，流体的位形空间是一个如下无限维群。受限于一些体积为$V\subset\mathbb{R}^3$的容器的理想流体的结构被分配到无限小流体元的每个空间位置$x\in V$。鉴于作为参考位形的分配，任意其他位形都可以用体积守恒的从V到V的微分同胚映射(diffeomorpnism)d标记，V将参考位形带到给定的那个位形，上述的标记是通过沿着d拖动每个流体元实现的。所以给定一个参考位形的选择，流体的位形空间由微分同胚映射d：$V\to V$的无限维群$\mathcal{D}$给出的，正如转动群是(转动)刚体的流形空间。$\mathcal{D}$为理想流体的严格拉格朗日和哈密顿理论打下了基础。

这些理论都和刚体的拉格朗日和哈密顿理论相当类似，很幸运有这样的事实：不管在哪种情况下，对称群都形成了位形空间。特别是，理想流体的欧拉方程和刚体的欧拉方程类似。此外，这些严格的流体理论(辛约化适用于它们)在科学上是重要的，它们得到了各种一般定理，解决了以前棘手的问题。更多细节，参见亚伯拉罕和马斯登[1978]讨论刚体的4.4和4.6节及理想流体的5.5.8小节，Arnold[1989]附录中关于刚体的2：C到2：F及关于理想流体的2：G到2：L，马斯登和拉蒂乌[1999]中关于刚体的1.4节和15章及关于理想流体的1.5节。

(4)哲学重要性

我认为辛约化在哲学上也是很重要的，至少体现在两个方面。第一个方面比较特殊，它阐明了一些经典力学分析问题的方法论信条(methodological morals)，我在[Butterfield，2005]中展开了这个主题。第二个方面是较一般的，这个理论，或者说它的各种应用与时间、空间和力学的哲学争论直接相关，这种关联是被当代物理哲学所公认的。据我所知，较为详细地展开这些关系的学者有贝洛特和厄尔曼。他们讨论和如下主题相关的辛约化：

(i)关于对称性的处理，包括规范对称性；

(ii)关于时间和空间绝对概念和关系论概念之间的争论；

(iii)经典广义相对论的解释，即一个连接(i)和(ii)的主题，和量子引力的启发相关。

因此，贝洛特[1999；2000；2001；2003；2003a]和厄尔曼[2003]主要讨论(i)与/或(ii)，而贝洛特和厄尔曼[2001]则讨论(iii)。对于(i)和(ii)，我也推荐华莱士(Wallace)[2003]。

但这些论文有一个严苛的先决条件，他们借助但不陈述辛约化理论。他们也讨论无限维系统(尤其是经典电磁学和广义相对论)，而没有提出有限维的诸如刚体的例子。的确，据我所知，理论的冗长阐述无不使哲学家望而生畏。所以我的目标是给出这样一个阐述来帮助那些被引用的论文的读者。①

作为这个阐述的引子，我将首先(在2.3节)遵循贝洛特给出的具有鲜明哲学关系的有限维辛约化的一般特征，即绝对事物和关系论争论对比。这个例子关注欧氏空间中的点粒子系统，要么自由移动要么受牛顿引力相互作用(定义商过程的对称性由平动和转动的欧几里得群给定)。对于哲学家来说，这将是辛约化的一个很好的引子，因为采用了大量篇幅来阐明关于这类莱布尼茨和马赫提倡的空间关系论。

2.3 引子：贝洛特的相对论力学

2.3.1 两种商过程的对比

在一些论文中，贝洛特讨论了辛约化如何就空间问题使得绝对的和相对的争论建立联系。我将挑选一个讨论的主题：相对的经典力学理论和通过商化正统绝对论者(也可称为“实体论者”)经典力学及通过正确的对称群得到的信息之间的比较。我所支持的他的主要论点是，这比较清楚地阐明了关系论，关于它的动机和它的优缺点。②

---

① 正如我在1.2小节所说的那样，我的资料是从亚伯拉罕和马斯登、阿诺德、马斯登和拉蒂乌以及奥尔弗的书中得到的。更确切地说，我们主要利用[Abraham and Marsden, 1978, Sections 3.1—3.3, Sections 4.1—4.3]、[Arnold, 1989, App.2, 5, and 14]、[Marsden and Ratiu, 1999, Chapters 9—13]和奥尔弗[2000, Chapter 6]。作为一个对称约化(非哲学)的介绍，我也推荐[Singer, 2001]。它是在一个比下列所述内容更基本的水平上，例如，它省略了泊松流形和复共轭(co-adjoint)表示。

② 主要的参考文献是[Belot, 1999; 2001; 2003, Sections 3.5, 5]，也参见[2000, Sections 4—5.3; 2003a, Section 6]。尽管我推荐所有的这些论文，但是最接近的模板是[Belot, 2001]。

贝洛特的总体想法如下：关系论者承认作为所有物质部分之间距离参数的可能位形；绝对论者(或实体论者)看到无限种可能，关系论的位形(一个相对的位形)可以被嵌入绝对空间。这当然使关系论者正视力学，这个力学仿佛是某种绝对论力学的“商”。

特别是，在欧氏几何($\mathbb{R}^3$ 样式)空间的传统概念上，每个关系论的相对位形相当于绝对论位形(即将物质的排列嵌入$\mathbb{R}^3$)的等价类，这个位形伴有通过空间平动和转动而类相关的成员，即欧几里得元素。用群作用量的术语说，在第4节中有所发展的有：作用于所有绝对论位形的欧几里得群和相当于作用量轨道的相对位形。所以很自然地使关系论者正视商化欧几里得群作用量的力学，以获得相对位形空间。拉格朗日或哈密顿类型的相对论力学在相对位形空间中被逐渐建立。

但正如贝洛特强调的那样，我们反而可以考虑商化绝对论的态空间——即哈密顿框架内的相空间——而不是它们的位形空间。事实上，这正是我们在辛约化中所做的事情。特别是，作用于绝对位形空间的欧几里得群作用量可以被提升到作用于余切丛 T^*Q 上，因此被称为“余切提升”。我们作商，即认为 T^*Q可以被余切提升划分为多个轨道。

因此，我们有两种要比较的理论：(i)相对论理论，从相对位形空间逐渐建立起来的，为了与辛约化作比较，我们把这个位形空间看作是哈密顿量，而非拉格朗日量；(ii)通过“后来的”商得到的理论，即商化绝对论的余切丛。

现在，我将阐明这种比较，但我不会试图总结贝洛特的更详细的结论，这个结论是关于这个比较所揭示出的关于关系论优缺点的东西。它们都是极其复杂的，因此与上面的总结相对：它们主要是在他的[2000，pp. 573—574，582；2001，Sections VIII to X]中被发现。罗韦利在本书中也讨论了相对论。

我也将(像贝洛特那样)集中于尽可能简单的情况：假定欧几里得空间几何中 N 个点粒子的力学。当然，绝对论者通过假设一个欧氏空间来做这样的假定，但对于关系论者，这个假定是在关系到各种粒子间的距离来编码的。这样一个系统的主要相对论力学的例子，起因于巴伯(Barbour)和贝尔托蒂(Bertotti)(1982)，尽管他们在拉格朗日框架下发展了它(准确而言是根据雅可比原

理)。贝洛特还讨论了其他相关理论，包括场论，即无限系统的理论，它们中的一些也源于巴伯，也在拉格朗日框架内。但在这一节中我只考虑 N 个点粒子。

同时，我也不讨论加速，尽管关系论者在传统上提出了确定任何两个和加速相关的绝对论的运动态。根据群作用量，这意味着我将通过欧几里得群而不是伽利略群的作用量来考虑商，参见在 1.2 节的(iii)中，我是如何驳回时间相关变换的。即使我们先前的讨论使它们易于理解，我也将把技术细节推迟到后面的小节。

最后，为避免读者之后的失望，以后的小节将不包括欧几里得群作用量的完整分析，这个作用量是作用于位形空间、相空间和它们的商的。这将涉及引子以外的术语。相反(正如前面 1.2 节最后提到的)，挑选出以后小节的材料使得我们得到第 7 节的定理，李—泊松辛定理，关于商化位形空间是李群的系统的相空间。这个定理之所以呈现出来的更进一步的原因将在 5.1 节给出。

2.3.2 引入空间和群作用量

让我们首先确立 N 个点粒子通过牛顿引力相互作用的绝对论力学，伴有欧几里得群的作用量。

每个点粒子都占据了$\mathbb{R}^3$ 中的一个点，这样的位形空间 Q 是在$\mathbb{R}^{3N}$：$\dim(Q)=3N$。所以，对于哈密顿力学的相空间将是余切丛 $T^*Q\ni(q,p)$：$\dim(T^*Q)=6N$。

哈密顿量是动能和势能(K 和 V)的总和。K 仅仅依赖于 p，V 只依赖于 q。在笛卡尔坐标中，用 $i=1,\cdots,N$ 来标记粒子而不是自由度，我们可以得到熟悉的表达式：

$$H(q,p)=K(p)+V(q)，\text{其中 } K=\sum_i\frac{\mathbf{p}_i^2}{2m_i},\ V(q)=G\sum_{i<j}\frac{m_im_j}{\|\mathbf{q}_i-\mathbf{q}_j\|}\tag{25}$$

其中 m_i 是质量，G 是引力常量。①

欧几里得群 E，又名 $E(3)$，是$\mathbb{R}^3$ 中的平动、转动和反射(组成)的群。但自从我们对连续对称感兴趣，我们就忽略了反射，所以我们会考虑保留了平动和转动的定向的子群，即连接恒等变换(我将写作 E)群的分量。这是一个李群，即一个也是流形的群，且关于流形结构，这个群运算是平滑的(smooth)。第 3 节将给出适当的细节。这里我们只注意我们需要三个实数来指定一个平动($\mathbf{x}=(x, y, z)$)，也需要三个实数来指定一个转动(两个为坐标轴，一个为用来转动的角坐标)，因此，作为一个流形，E 的维度是 6，即 $\dim(E)=6$，这不足为奇。

E 很明显作用在$\mathbb{R}^3$ 上。例如，如果 $g\in E$ 通过 $\mathbf{x}\in\mathbb{R}^3$ 平动，g 引起了映射 $\mathbf{q}\in\mathbb{R}^3\longmapsto\mathbf{q}+\mathbf{x}$。同样，对于一个转动的产生，第 3 节将再一次给出一个正式定义。

现在，假设 E 以这样的方式作用于我们的系统位形流形 $Q=\mathbb{R}^{3N}$的 N 个商空间$\mathbb{R}^3$ 中的每一个。这定义了一个在 Q 上的作用量 Φ：即对于所有 $g\in E$，都有一个映射 Φ_g：$Q\to Q$。例如，通过 $\mathbf{x}\in\mathbb{R}^3$，$g=$一个平动，我们有

$$\Phi_g:(\mathbf{q}_j)=(\mathbf{q}_1,\cdots,\mathbf{q}_N)\in Q\longmapsto(\mathbf{q}_1+\mathbf{x},\cdots,\mathbf{q}_N+\mathbf{x})\in Q \tag{27}$$

和转动一致。因为式(25)的势函数 V：$Q\to\mathbb{R}$只取决于粒子间距离，每个映射 Φ_g：$Q\to Q$ 都势能对称；也就是说，我们有 $V(\Phi_g(q))=V(q)$。

作用量 Φ(即设定 $g\in E\longmapsto\Phi_g$)在 $T^*Q=T^*\mathbb{R}^{3N}$上对 E 产生了作用，被称为 Φ 到 T^*Q 的余切提升，通常被写成 Φ^*；所以，对于每个 $g\in E$，我们都有一个提升映射 $\Phi_g{}^*$：$T^*Q\to T^*Q$。同样，细节一直等到以后(第 4 节)再阐明。但是按照这个想法，Q 上的每个映射 Φ_g 都是平滑的，所以映射曲线对曲线，

① 从更深远的哲学观点看，式(25)最重要的特征毫无疑问是：势能是对于位形 $q\in Q$ 的所有两体势能的总和，不存在多体相互作用。备注：事实上，动能可以通过位形空间上的矩阵 g 来表示。对于哈密顿力学，这意味着，在余切从 T^*Q 上的动能标量 K 可以通过将 Q 的矩阵 g 应用到动量 p 的投射来定义，其中，在每个点$(q, p)\in T^*Q$ 处，投射是优先得到的。

同构 $\omega^\sharp$：$T_q^*\to T_q$，参见方程(12)，即：

$$K:(q, p)\in T^*Q\longmapsto g_q(\omega^\sharp(p), \omega^\sharp(q)) \tag{26}$$

矢量对矢量，余矢量对余矢量，等等。

不出所料，每个提升映射 $\Phi_g^*: T^*Q \to T^*Q$ 使势能 V(现在被看作是在 T^*Q 上的一个标量)为一个常量，即我们有 $V(\Phi_g^*(q, p)) = V(q, p) \equiv V(q)$。但除此之外，每个提升映射 Φ_g^* 在哈密顿系统中都是对称的(第 2.1.3 节)。即 Φ_g^* 保留了哈密顿量(事实上动能和势能项都是独立不变量)，它保留了辛结构。这意味着动力学在作用量 $g \in G$ 的作用下是不变的：通过(q, p)和 $\Phi_g^*(q, p)$的动力学历史系统是完全匹配在每个时间点的。在定性上它们是不可区分事物，用当代形而上学的术语来说，它们是复制品(duplicates)。

当然，在这一点上我们遇到了关于空间是绝对的还是相对的的讨论。绝对论者断言有这样两种不可区分的可能性是讲得通的，① 相对论者否认了这一点。所以，正如上述理论呈现的，相对论者认为我们应该消减概率空间。正如我在 2.3.1 节中所说的，很自然地会使在商化欧几里得群的作用量下非常精确：通过欧几里得群中元素相互关联的绝对论的可能性的集合形成一个等价类(一个轨道)，这个等价类表示了一个相对论的可能性。

但在这里我们需要区分两种不同的求商过程。我将称它们为关系论(Relationism)和约化论(Reductionism)(首字母 R 要大写)，因为前者接近于传统的和当代的相对论建议，后者是辛约化正统观点的例子。正如我在 2.3.1 小节所说的那样，主要的区别是：

(i)关系论执行了在 E 的作用量上的商，这个作用量是作用在位形空间 Q 上的，这组轨道形成一个相对的位形空间，在这个位形空间上关系论者提出建立一个动力学，无论是拉格朗日的或哈密顿的——在后者的情况下，产生了一个相对相位空间。

(ii)约化论执行了在 E 的作用量上的商，这个作用量是作用在通常的相空

① 对于这个争论的权威性章节当然是莱布尼茨—克拉克(Leibniz-Clarke)对应，尽管主角的论证有时是神学上的。克拉克这个绝对主义者认为，存在许多空间中和相对距离相一致的物质的可能安排，论说"如果(上帝纯粹的意志)在没有提前决定的情况下是不会起作用的……这倾向于拿走所有选择的力量来引进宿命"。莱布尼茨认为，仅有一种这样的安排："那两个状态……彼此之间没有一点不同。它们的不同在本质上仅仅能在我们的空间实在的荒唐猜测中发现。"

间 T^*Q 上，这组轨道形成了一个约化的相空间。

自从我们的讨论采用了哈密顿框架，对于下列叙述它将不受影响，关系论可以采用拉格朗日框架，而约化论是遵从哈密顿框架。重要的是，(i)和(ii)导致了不同维度的相空间，约化的相空间比相对相空间多了 6 个维度，即"维度差"是6。

我们将看到，描述维度的 6 个变量中，4 个是运动的常量，其他两个随时间变化。对于确定的运动(大致上没有转动)常量值的选取，随时间变化的变量就等于零，遵循约化论理论的动力学得到了简化，以符合相对论的理论。换句话说，如果没有转动，那么非正统的相对论动力学和传统的辛动力学相匹配。

2.3.3 关系论过程

关系论者寻求一种基于相对位形空间(RCS)的力学。RCS 中的一个元素将是成为几何上可能的粒子间距离和角度的一个模型，即和被嵌入在$\mathbb{R}^3$ 中的 N 粒子相兼容。因此，粗略地讲，RCS 中的元素是一个欧几里得位形，RCS 将被设置为轨道组$\mathbb{R}^{3N}/E$。

即便在给定一个更精确的陈述之前，我们可以陈述关于维度的"点睛之笔"：因为 $\dim(E)=6$，用 E 的商化减去 6 个维度，即 RCS 的维度将是 $3N-6$。

但是，对于 RCS，我们需要更加精确。对于轨道和存在流形的商空间，且以这种简单的方式在维度上加或减，在我们作商之前，我们需要删除两类来自$\mathbb{R}^{3N}$的"特殊"点(第 4 节讲述有关这些删除法的技术原理)。

假设 $\delta_Q \subset \mathbb{R}^{3N}$是对称的位形集：即每一个都被 E 中的元素(除了单位元素)固定。所有点粒子都是共线的任意位形提供的一个粒子：位形通过作为轴的线的任意转动来固定。假设 Δ_Q 为碰撞位形的集合，即两个或多个粒子在通常的位形空间$\mathbb{R}^{3N}$中一致(Q 的下标之后将会提醒我们，这些集合是位形集合)。δ_Q 和 Δ_Q 在$\mathbb{R}^{3N}$都是零测度(measure zero)。除去它们两个，并称为结果空间(resulting space)，$Q := \mathbb{R}^{3N} - (\delta_Q \cup \Delta_Q)$，它的维度是 $3N$。

δ_Q 和 Δ_Q 在 E 的作用量下都是闭合的，即每一个都是轨道的并集；对称位形的欧几里得变换也是对称的。所以 E 作用于 Q。现在通过 E 商化 Q。Q/E 是相对论的 RCS。因为 $\dim(E)=6$，我们有 $\dim(Q/E)=3N-6$。

这些 $3N-6$ 个变量编码所有的(相对)位形的粒子对相对距离，$r_{ij} \in \mathbb{R}$(用

i，j 标记粒子)。请注意，有 $N(N-1)/2$ 个这样的相对距离；且对于 $N>4$，这将大于 $3N-6$(对于 $N\gg4$，结果将更大)。所以，虽然是一种物理直觉，但相对距离仍给出了在 Q/E 上超完备(over-complete)的坐标系。(所以它们不能被自由选择，它们之间有约束关系。)

所以关系论者寻求一种使用这个 RCS 的力学。在时间上二阶的牛顿第二定律意味着她也将需要像速度(在拉格朗日框架)或者像动量(在哈密顿框架)的量。对于前者，她自然会考虑 $N(N-1)/2$ 个相对速度，$\dot{r}_{ij}:=\frac{d}{dt}r_{ij}$；对于后者，对应动量 $p_{ij}:=\frac{\partial L}{\partial \dot{r}_{ij}}$。并且，她必须注意约束条件。建立在她的 RCS 力学 Q/E 上的正切丛和余切丛维度将分别为 $2(3N-6)=6N-12$。所以，对于 $N>4$，$N(N-1)/2$ 个相对速度 $\dot{r}_{ij}$，或者相对动量 p_{ij} 大于相关的自由度的数目；且对于 $N\gg4$，结果将变得更大。所以，相对速度或相对动量是超完备的：有约束。

另一方面，如果关系论者仅仅使用这些相对数量，r_{ij}、$\dot{r}_{ij}$ 或者 p_{ij}，或 $T(Q/E)$ 或 $T^*(Q/E)$ 上坐标的“等价物”，这个坐标不是超完备的，不管她的理论其他细节是什么，她将面临一个传统的问题。至少，如果她期待一个在经验上与正统的绝对论理论等价的理论，她就会面临着一个问题。我将遵循传统，并依据相对速度而不是动量陈述问题。

这个问题关系到转动(在此呈现出和莱布尼茨辩论中牛顿和克拉克位置的强度)。按照绝对论理论，点粒子的两个系统可以匹配所有的相对距离和相对速度，但是未来的演变却不同，然而仅仅使用这些相对量(或“等价的”变量)，允许作为绝对论理论的相同可能性的理论必定是非决定论的。

最简单的例子是对牛顿双球体思想实验的点粒子模拟。因此，这些系统每个都包含两个相对速度为零的点粒子。一个系统可以是非转动的，因此这样的点粒子在重力的作用下对彼此向下倾斜，而其他系统关于垂直于粒子间连线的轴转动，并以一个平衡引力的速度转动。

关系论者在传统上已经回应，他们不期待经验上相当于绝对论的理论。相反，他们设想了一个在其中只可能有非转动演变的力学，即更一般地，在其中作为一个整体的宇宙必须有零角动量的力学。最初，在诸如莱布尼茨和马赫的

作者看来，这个回答是一张承诺函。但诸如巴伯和贝尔托蒂[1982]的现代相对论理论已取得很好的认可，他们已经远远超出通过牛顿引力相互作用的点粒子。此外，因为宇宙似乎是非转动的，这些理论甚至要求在经验上是充足的，至少对于这个原则上不同于绝对论理论的理论而言是如此。①

我并不是要在概要中加入这些理论细节，除了和被商化的版本绝对论理论(quotiented version)作比较的方法，参见2.3.4小节。

2.3.4 约化论过程

约化论者的主要想法是只在跳转到N个点粒子的正统相空间之后作商，即$\mathbb{R}^{3N}$上的余切丛$T^*\mathbb{R}^{3N}$。所以这个想法是考虑$(T^*\mathbb{R}^{3N})/E$，即通过欧几里得群E的余切提升作用量Φ^*对$T^*\mathbb{R}^{3N}$作商。

更准确地说，我们再次通过删除引起技术麻烦的特殊点继续前行。但现在，被删除的点是在余切丛$T^*\mathbb{R}^{3N}$中而不在$\mathbb{R}^{3N}$中。所以假设$\delta\subset T^*\mathbb{R}^{3N}$是位形对称（在2.3.3节的$\delta_Q$的意义上）的相空间态的集合。假设$\Delta\subset T^*\mathbb{R}^{3N}$是碰撞点的集合，即在其中两个或多个粒子在位形空间$\mathbb{R}^{3N}$中是一致的态。$\delta$和$\Delta$都是零测度的。删除它们两个，且称$6N$维的结果相空间为$M := T^*\mathbb{R}^{3N}-(\delta\cup\Delta)$。

δ和Δ在$T^*\mathbb{R}^{3N}$上的E的余切提升作用量下都是闭合的。即每一个都是轨道的并集，作用在具有对称位形的相空间态的欧几里得变换的余切提升将产生一个也有对称位形的态。所以E作用于M。现在，用M除以E，得到$\bar{M} := M/E$，这就是所谓的约化相空间。我们有$\dim(\bar{M})=\dim(M)-\dim(E)=6N-6$。

正如2.3.2小节最后强调的，$\bar{M}$比对应的相对性相空间(不管是速度相空间(正切丛)或动量相空间(余切丛))多了6个维度。这些相空间的维度是$2(3N-6)=6N-12$。事实上，通过考虑这些“维度差”，我们可以更好地理解约化的相空间$\bar{M}$和相对性相空间。有以下两个扩展的论述。

① 绝对论的提倡者可能会说，宇宙的表面偶然特征是非常奇怪的，如非转动性，力学原理；关系论者可能答复他们，他们的观点对预测宇宙不转动有一定的价值！我担心对于这些方法论的争论，不存在非常清楚的判定标准。

(1)获得关系论相空间

我们可以从最初的相空间 M 中获得关系论的动量相空间。因此假设 M_0 为 M 的子空间，在这个子空间中，系统使得总线动量和角动量都等于零。因为这些都是运动的常量，M_0 是动态闭合的，所以支持仅仅通过最初动力学的约束而给定的哈密顿动力学。线动量和角动量每个都贡献三个实数，$\dim(M_0)=\dim(M)-6=6N-6$。此外，M_0 在 E 的余切提升算符下是闭合的，写为 $\overline{M}_0 := M_0/E$。那么 $\dim(\overline{M}_0)=6N-6-6=6N-12$。

现在回想一下，这是建立在 RCS Q/E 上设想的关系论理论的相空间的维度。事实上，正如我们希望的 $\overline{M}_0$ 是巴伯和贝尔托蒂 1982 年的相对论理论的哈密顿版本(回想一下，他们的工作是在拉格朗日框架内)。

即 $\overline{M}_0$ 是辛流形，且在 $\overline{M}_0$ 的点用所有粒子间的相对距离和相对速度的参数表示。一旦最初的动力学被投射到 2.3.3 节的相对位形空间 Q/E，就有一个匹配最初的绝对论理论的确定动力学。

简而言之，总线动量和总角动量等于零意味着仅包含相对量的初态就足以确定所有未来的相对量。

(2)分解约化论的约化相空间

让我们回到还原的相空间 $\overline{M}$。第一点是，自从 M 上或者 $T^*\mathbb{R}^{3N}$ 上的哈密顿量 H 在 E 的余切提升作用量下是不变量，M 上通常的动力学被投射到 $\overline{M}=M/E$。即，还原的相空间动力学拥有所有一般动力学的 E 不变的特性。

事实上，$\overline{M}$ 是一个泊松流形。所以这是对于 2.2 小节的(1)中提出的哈密顿力学更普遍的框架内的第一个例子。我再一次将技术细节放到后面(尤其是 5.1 小节和 5.2.4 小节)论述。但这个想法是，泊松流形有一个简并的反对称双线性映射，这意味着这个流形是辛流形的不相交并集。每个辛流形都被称为泊松流形的叶(leaf)。叶的辛结构彼此之间的“网格”，在每个叶内有一个常见的哈密顿动力学。

即使没有泊松流形的精确定义，我们仍可以描述如何将 M 分解为辛流形，每一个都有一个哈密顿动力学。回想一下，我们有 $\dim(\overline{M})=\dim(M)-\dim(E)=6N-6$。这种分解为：

$$6N-6=(6N-12)+3+3=2(3N-6)+3+3=:\alpha+\beta+\gamma \tag{28}$$

在右边分别定义了 α、β、γ 为 $2(3N-6)$、3 和 3。根据 $\overline{M}$，其意义如下。

(i) α 对应于(1)中的 $\overline{M}_0$，即对应于 $T^*(Q/E)$。如前所述，$3N-6$ 个变量编码所有粒子对的相对距离；其他 $3N-6$ 个变量编码所有粒子对的相对动量。6 个额外的变量附加到这些 $6N-12$ 个的相对量，包括 4 个运动的常量和另外两个动力学变量，即随着时间改变的量。

(ii) β 代表 4 个运动的常量中的 3 个，也就是说，这 3 个变量编码了系统中所有的线动量，即质心的动量。这些运动的常量是在以下意义上“仅仅是参数”：(a)不仅为它们中的 3 个指定一个值以确定一个表面，即一个在 $\overline{M}$ 中 $(6N-9)$ 维的超曲面，在这个超曲面中存在哈密顿动力学；(b)这种哈密顿量和辛结构都独立于我们指定的数值。[①]

(iii) γ 代表 3 个编码系统总角动量的变量。其中之一是第 4 个运动的常量，也就是总角动量 L 的大小。其他两个随时间变化的量确定在半径为 L 的球体上的点，对这些绕着其转动的框架内的系统角动量方向进行编码。这种情况就好像是在刚体的初等理论中，尽管相对于固定在空间内坐标的总角动量是一个运动的常量(3 个恒定的实数)，但是总角动量相对于刚体只在大小(一个实数 L)上是一个常量，而在方向上不是。这将在第 5 节详细描述，到那时我们会在刚体理论中描述泊松流形结构。目前，有以下两个重要的评论(a)和(b)，来估计 N 粒子系统。

(a) 除了系统质心的动量之外，如果我们详细说明 L，我们将得到一个在 $\overline{M}$ 上的 $(6N-10)$ 维超曲面，在这个超曲面上，如(ii)中陈述的那样，存在哈密顿动力学。所以我们可以认为 $\overline{M}$ 包含了这些超曲面的 4 个实参数族，每一个超曲面的每个点都配备半径为 L 的球体受限于下面(b)中的条件。注意，这里的“被配备的每个点”并不意味着球体给了额外的作为纤维丛的 $\overline{M}$ 的维度(将有两个维度是缺失的)。恰恰相反，点的表示是通过 $6N-10$ 个实数进行的，其中的两个数可以用来表示

---

① 在 2.3.1 节中提到的，关系论者在传统上提议确定绝对论者的运动特性，这个运动特性仅因为总动量的值而不同。事实上，这个提议可以通过考虑一个作用在绝对相空间 M 上的伽利略群的作用量来实现。对于具体讨论和参考文献，参见[Belot，2000，Section 5.3]。

球体上的点。

(b)但是不像上述的(ii)中β的情况，在这样一个超曲面上的哈密顿动力学取决于L的值。尤其是，如果$L=0$，表示刚体角动量的球体是简并的，即半径为零，其他两个随时间变化的量也等于零。超曲面上的每个点通过$6N-12$个实数来表示，即超曲面是$6N-12$维的。现在回想一下2.3.3节或上述的(1)，$6N-12$是建立在RCS Q/E上的直观相对论理论的相空间维度。事实上，正如我们希望的那样，$L=0$且线动量等于零的超曲面是辛流形，也是巴伯和贝尔托蒂[1982]提出的N点粒子相对理论的哈密顿版本的动力学，这个超曲面就是$\overline{M}_0$。

我们可以总结该对比如下：在这个超曲面$\overline{M}_0$上，在约化的相空间里的动力学和适用于相对变量的动力学是一致的，如果我们任意地将它们的初始值嵌入到一般的绝对论的相空间$T^*\mathbb{R}^{3N}$中，由于约束的作用，总的角动量和线动量等于零，然后从$T^*\mathbb{R}^{3N}$里的一般演变中获取（仅仅通过投射）它们的演变。

2.3.5　关系论和约化论过程的对比

在关系论和约化论过程的对比中，我将只作两条推广的评论，并参考贝洛特作进一步的探讨。这两条评论的要点是，约化论就足够了，可以不需要关系论。第一条是大众的观点，第二条归功于贝洛特。

2.3.5.1　考虑到转动的约化论

第一条评论重申了约化论者的所能和约化论者的所不能，以支持牛顿的地球(或水桶)思想实验。约化论可以在下列情况下成立：

(i) $(6N-6)$维相空间$\overline{M}=M/E$;

(ii) $(6N-9)$维超曲面可通过指定质心的线动量从(i)中得到;

(iii) $(6N-10)$维超曲面可通过指定L的非零值从(ii)中得到。

在所有的三种情况下，约化论者可以以一种带有他们的$(6N-12)$维空间的关系论者所不能的方式描述转动。因为约化论者必须给出描述作为整体的系统转动的三个额外非相关变量(L和其他两个)。(顺便提一下，将系统作为整体描述是在只有其中之三时提出的，而不管N的值是多少。)特别是，在支持地球和水桶思想实验所主张的区别的意义上，约化论者可以区分转动态和非转动态($L=0$)。

约化论者也可以满足传统关系论的动机，这涉及一般哲学，而不是运动理论。尤其与莱布尼茨相联系，即我们的理论(或我们的形而上学)不应该承认有区别但非常难以识别的可能性。我们有理由这样问，为什么我们应该支持这个“不可观测事物的同一性原理”作为可能性而不是客观事实。对于莱布尼茨本人，正如贝洛特[2001]提出的，答案体现在他的理由充分原则(principle of sufficient reason)，并最终表现在神学上。

但在任何情况下，约化论者都可以满足这个需求。一致认为，通常嵌入在$T^*\mathbb{R}^{3N}$——或者，如果你喜欢，$M = T^*\mathbb{R}^{3N} - (\delta \cup \Delta)$——中的绝对论理论有九个变量，这些变量描述包括：(i)质心的位置；(ii)围绕其质心的系统的方向；(iii)系统的总线动量，即对于(i)~(iii)中的每一个都有三个变量和$\mathbb{R}^3$中的一个矢量。所以通常的绝对论者的理论有九个维度的截然不同的但又难以觉察其可能性的“浪费”。但正如我们所看到的，约化论者通过欧几里得群E的作用量求商，在$\overline{M} = M/E$中也是如此，这去除了关于(i)和(ii)的“浪费”。

对于(iii)，我同意我前面所述的观点，即还有工作有待完成。由$(6N-9)$维超曲面的三个实参量族表示且通过系统的总线动量来标记的$\overline{M}$的叶状结构(foliation)整合了上述的浪费——但不能消除它。但正如我上面所提到的，约化论者实际上可以进一步求商，通过考虑伽利略加速的作用量以及通过识别相差一个加速的相空间的点，即把轨道横断面定义到超曲面。

2.3.5.2 其他理论中类似的约化

通过总结贝洛特[2001, Sections VIII—IX]的总论，并参见他的[2003a]中的12和13节，我结束了对于辛约化的哲学家的引子。这些都关乎我们对关系力学的讨论如何成为许多情况的代表，以及辛约化在物理上为何重要。我标记为(1)~(3)。

(1)一般的对比：何时作商

N个点粒子通过牛顿引力相互作用的这个例子是典型的一大类情况（既可以是有限维的也可以是无限维的)。有一个通过对称连续群G起作用的位形空间Q，Q提升到余切丛T^*Q，其中，余切提升保持了不变的哈密顿量和动力学。所以我们可以通过G商化T^*Q来给出一个约化理论。(存在拉格朗日模拟，但正如上面所说的，我们不讨论这个。)但对于商化作用G在Q上的作用

量，不论我们如何去构造动力学，也存在一些动机。不管用什么方式促进商比位形空间，我们都用“关系论”作为标签。然后，当我们假设了适当的技术条件（回忆我们对于δ和Δ的处理），我们将有：

（i）对于约化的哈密顿理论，有 $\dim((T^*Q)/G)=2\dim Q-\dim G$；

（ii）对于在哈密顿和拉格朗日框架内的关系论理论，有 $\dim(T(Q/G))=\dim(T^*(Q/G))=2(\dim Q-\dim G)$。

因此，在约化理论中我们有 $\dim G$ 个变量，这些变量在相对论理论中不存在，我们称它们为“非相关变量”。

（2）非相关变量

典型地，这些非相关变量表示系统全局的（集体的）特征。因为这些变量的数量 $\dim G$ 独立于系统的自由度数量（$\dim Q$，或者如果你分别计算一下自由度的变化率，那就是 $2\dim Q$）。

这些变量中的一些是守恒量，这些守恒量起因于对称性。此外，可能还有守恒量的比值，比如在 2.3.4 小节中为零的角动量，对于这些守恒量的比值，约化理论退化成相对论理论。即不仅仅是等维的相关态空间，而且它们的动力学也是统一的。

（3）约化理论

通常，约化相空间$(T^*Q)/G$的拓扑学和几何学，以及在相空间上的哈密顿函数，例如$\bar{H}$：$(T^*Q)/G\to\mathbb{R}$，要比在T^*Q上的非约化理论的相应特征更复杂。特别是，约化的哈密顿量$\bar{H}$通常有和不在非约化理论中的力相一致的势能项。但这个不应该被当作必然的缺点，原因有如下两点。

第一，有著名的案例可说明这一点。在这些案例下约化理论有一个特殊的动机。其中一个例子是力学中赫兹（Hertz）的处理方法，即“解释”我们宏观经验（例如重力）的表现力，宏观经验起因于具有适当对称性理论的约化。该处理方法设想我们所不知的微观自由度的循环变量，参见［Lutzen，1995；2005］。另一个著名的例子是通过电磁场施加于带电粒子上的卡鲁扎—克莱因（Kaluza-Klein）处理。那就是常见的在四维时空中描述带电粒子运动的洛伦兹力定律（Lorentz force-law）可以被显示出来，具体是通过辛约化从假定时空由第五个（微小以及闭合）空间维度的理论产生出来，在第五维度上，粒子做直线运动。

值得注意的是，相关的守恒量，即沿着第五维的动量可能被认为和带电粒子是等同的，因此这个理论可以用来解释电荷守恒。这个例子概括到其他场，在细节和引用方面，参见[Marsden and Ratiu，1999，Section 7.6]。

第二，约化理论不需要如此复杂，以至于极难处理。事实上，这两个例子证明了这一点，因为约化理论在它们中完全是容易处理的。这是熟悉的理论——为了虚构的非约化理论，我们可能不会放弃这一理论。① 此外，即使当约化理论看似是复杂的(而不只是因为它是陌生的)时候，贝洛特仍描述了辛约化的一般理论是如何使得我们能在约化的相空间里经常“做物理”。如在卡鲁扎—克莱因的例子中，约化的相空间中的物理学可能具有启发式和解释的价值。

3. 一些几何工具

介绍了这么多的引子！包含这一节和后面四节的本章剩余部分是五道“大餐”！本节阐述现代微分几何，尤其是关于李代数和李群。第4节采用李群来研究作用量。然后第5节描述作为哈密顿力学的一个广义框架内的泊松流形。正如我在2.2节中的(2)所提到的，李自己发展了这个框架，因而事实上，他知道在这两节中的一切——因此用真实的双关语说(尽管痛苦)，就是这三节给了我们“彼岸的方向”。在任何情况下，这两节将为我们准备依据动量映射的对称性和守恒的描述。最后，第7节将提出一个主要的关于辛约化的定理。它将关注这种情形：对于一个系统的自然位形空间本身就是一个李群 G。参见2.2节的(3)。商化自然相空间(在 G 上的余切丛)将会给出一个泊松流形，这个流形是 G 的李代数的对偶。

在本节中，我首先概述了微分几何中的一些概念，并定义符号(3.1节)。然后我介绍李代数和李括号的矢量场(3.2节)。尽管这一节的大部分是关于微

① 我们应该抵制由于熟悉带来的偏见。为什么不让牛顿力学产生于一个微观循环自由度呢？为什么不让洛伦兹力产生于在一个包含第五个维度的五维空间中的短程线运动，因此，根据诺特定理，电荷守恒是通过对称解释的？

分概念而不是积分概念，我仍需要在 3.3 节提出弗罗宾尼斯(Frobenius' theorem)定理。然后，关于李群和李代数，我将给出一些基本的信息(3.4 节)。

3.1　流形上的矢量场

3.1.1　流形、矢量、曲线和导数

通过固定的想法和符号，我首先给出一些关于微分几何(已经在 2.1 节使用)中的观点的细节，并为它们引入了一些新的符号。

一个流形 M 将是有限维的，除了明显的和明确的例外，比如流形(仍是有限维的)的微分同胚映射的有限维群。我不会关注在流形的定义中的可微性程度，或任何相关的几何对象的定义中的可微性程度，"平滑"将遍及下列所述，并意味着 C^∞。我经常不会关心整体的(对应于局域的)结构和结果(尽管约化结果在本质上是整体的)。例如，我不会关心曲线是否是非延伸性的或流(flows)是完备的。

一般而言，我将点 $x \in M$ 上的矢量写作 X；或依据局域坐标 x^i 将其写作$X = X^i \dfrac{\partial}{\partial x^i}$(求和约定)。从现在起，我将给出在点 $x \in M$ 时的切空间，记作 T_xM(而不仅仅是 T_x)，从而用来明确表明流形 M 是切向的。像以前一样，我将由切空间的"网格系列"组成的切丛写作 TM；同样，我将在点 $x \in M$ 上的 1 形式(余矢量)写作 α，所以在 $x \in M$ 的余切空间为 $T_x^* M$，像以前一样，余切丛为 $T^* M$。

在流形 M 和 N 之间的平滑映射 f: $M \to N$(也许 $N = M$)将平滑曲线映射到平滑曲线，从而由切矢量到切矢量；对于 1 形式或更高的张量以此类推。对于切丛上归纳的映射，我们方便地写成 Tf，称为 f 的导数或切线(也写为 f_* 或 df 或 Df)。

更详细地，假设我们在 M 中取曲线 c 使其为从间隔 $I \subset \mathbb{R}$到 M 的平滑映射，如果在 $x \in M$，$X \in T_xM$ 上的切矢量是通过 x 的曲线的等价类$[c]_x$。等价关系是，对于在 x 的每一个局域图(local chart)，曲线在 x 处是切线，但是在这里我省略了细节。然后我们可以通过下式定义 Tf: $TM \to TN$(也写为 f_*: $TM \to TN$)：

$$对于所有的\ x \in M,\ f_*([c]_x) \equiv Tf([c]_x) := [f \circ c]_{f(x)} \tag{29}$$

如果我们把 Tf 仅限制于 x 点的切空间 T_xM，我们就把 Tf 写作 T_xf，即：

$$T_xf:\ [c]_x \in T_xM \longmapsto [f \circ c]_{f(x)} \in T_{f(x)}N \tag{30}$$

在3.1.2.2小节中，我们将讨论如何将切矢量定义为在争议点的邻域上定义的所有标量函数集合上的微分算符，而不是曲线的等价类。然后我们可以以一种等价于前面的方式来定义切映射 $f_* \equiv Tf$。

3.1.2 矢量场、积分曲线和流

我们尤其关注在 M 上的矢量场，即 X：$x \in M \longmapsto X(x) \in T_xM$，或者在子集 U⊂M 上的矢量场。因此，假如 X 是在 M 上的矢量场，且 f：$M \to N$ 是平滑的映射，那么 T_xf：$T_xM \to T_{f(x)}N$。

3.1.2.1 前推和回拉

我们应该注意到，$(T_xf)(X(x))$ 没有在 N 上定义矢量场。因为 $f(M)$ 可能不是所有的 N，所以，对于 $y \in (N - \mathrm{ran}(f))$ $(T_xf)(X(x))$ 没有分配 T_yN 中的元素。而且，f 可能不是内射的，这样我们有 x，$x' \in M$ 且 $f(x) = f(x')$ 以及 $(T_xf)(X(x)) \neq (T_{x'}f)(X(x'))$。因此我们说，矢量场并没有推前(push forward)。

另一方面，假设 f：$M \to N$ 是一个到 N 上的微分同胚映射，即平滑映射 f 是一个双射(bijection)，它的逆 f^{-1} 也是平滑的。那么，对于任何在 M 上的矢量场 X，$Tf(X)$ 是在 N 上的矢量场。所以，在这种情况下，矢量场没有推前。因此，$Tf(X)$ 称为 X 的前推(push-forward)，通常写作 $f_*(X)$。因此，对于任意 $x \in M$，在像点 $f(x)$ 处的前推矢量场通过下式给定：

$$(f_*(X))(f(x)) := T_xf \cdot X(x) \tag{31}$$

这将提出另外三个注释。

(1) 更一般地，如果在所有的 $x \in M$，都有 $(Tf)(X) = Y$（分别地 $x \in S$），我们就说 M 上的 X 以及 N 上的 Y 这两个矢量场在 M 上都是 f 相关的（分别地在 $S \subset M$ 上）。

(2) 我们可以推广这一理念，那就是，微分同胚映射意味着一个矢量场可以以两种方式推前。首先，微分同胚映射只需要在一些 $x \in M$ 的兴趣点的邻域局域地定义。第二，一个微分同胚映射建立一一对应，不只是在定义在定义域

和值域的矢量场之间，而且也在定义在定义域和值域的所有微分几何对象之间：特别是1形式场和更高阶的张量。

(3)评论(2)的后续：尽管矢量场并不总是推前，但1形式场却总是拉回(pull back)，用上标的星号表示。即对于任意平滑的f: $M \to N$，未必是一个微分同胚映射(甚至局域的微分同胚映射)，或在N上任意的1形式场(微分的1形式) α，我们定义回拉$f^*(\alpha)$是在M上的1形式，对于每个$x \in M$和每个$X \in T_xM$，M的作用量由下式给出：

$$(f^*(\alpha))(X) := \alpha|_{f(x)}(Tf(X)) \tag{32}$$

同样地，如果映射f仅局域地定义在M的一个子集上，定义在f的值域内的1形式回拉到一个在f的定义域上的1形式。

3.1.2.2　矢量场和流之间的对应

关于矢量场的主导理念是，对于任何流形，关于常微分方程系统的解的局域存在性定理、唯一性定理以及可微性定理[Arnold，1973，pp. 48—49，77—78，249—250；Olver，2000，Prop 1.29]，将获得四个概念中两两之间的一一对应。

(i) X为子集$U \subset M$上的矢量场，在这个矢量场上，它们是非零的，X: $x \in U \longmapsto X(x) \in T_xM$，$X(x) \neq 0$。

(ii)矢量$X(x)$的方向上，在每个点$x \in U$上的有非零的方向导数。根据坐标$\mathbf{x} = x^1$，…，x^n，这些都是一阶线性微分算符$X^1(\mathbf{x})\frac{\partial}{\partial x^1} + \cdots + X^n(\mathbf{x})\frac{\partial}{\partial x^n}$，$X^i(\mathbf{x})$是矢量$X(x)$在坐标系上分解的第$i$分量。这样一个算符通常是完全作为导数被引进的，定义在x的邻域的平滑实值函数集合上的映射，即线性的且遵守莱布尼茨规则。

(iii) 在U里的场X的积分曲线(也叫作解曲线)，即来自$I \subset \mathbb{R}$到U的实开区间的平滑映射ϕ: $I \to M$，其中$0 \in I$，$\phi(0) = x \in U$，且当$\tau \in I$，在每个$\phi(\tau)$处它的切矢量是$X(\phi(\tau))$。

(iv) 对于每个场X和每个$x \in U$，流X^τ将U的一些适当子集映射到另一个：X^τ: $U \to M$。这个流保证只存在于给定点x的邻域里，且对于τ，那将只存在于$0 \in \mathbb{R}$的邻域里，但对于我们，这已经足够了。这样的流是所有局域微分

同胚映射的“有限维群”的单参数子群。

我将更详细地讲清楚这个对应。在局部坐标 x^1，x^2，…，x^n 中，任何平滑曲线 ϕ：$I \to M$ 通过 n 个平滑函数 $\phi(\tau)=(\phi^1(\tau),\ \phi^2(\tau),\ \cdots,\ \phi^n(\tau))$ 给定，且在 $\phi(\tau)\in M$ 上，正切于 ϕ 的切矢量是：

$$\dot{\phi}(\tau)=\dot{\phi}^1(\tau)\frac{\partial}{\partial x^1}+\cdots+\dot{\phi}^n(\tau)\frac{\partial}{\partial x^n} \tag{33}$$

所以，对于 ϕ 作为 X 的积分曲线，我们要求对于所有的 $i=1$，…，n 和所有的 $\tau\in I$，有：

$$\dot{\phi}^i(\tau)=X^i(\tau) \tag{34}$$

对于给定的矢量场 X 和 $x\in M$，通过 X 的积分曲线 $\phi_{X,x}$ 的局域存在性和唯一性 $(\phi(0)=x)$ 确保流可以写作 X^τ 或 $\phi_X(\tau)$：

$$X^\tau:\ x\in M \longmapsto X^\tau(x)\equiv\phi_{X,x}(\tau)\in M \tag{35}$$

这是(至少局域地)定义明确的。流是 M 的一个单参数变换群，X 被称为它的无穷小算子。

指数记号：

$$\exp(\tau X)(x):=X^\tau(x)\equiv\phi_{X,x}(\tau) \tag{36}$$

是启发性的。例如，在流中的群运算，即：

$$X^{\tau+\sigma}(x)=X^\tau(X^\sigma(x)) \tag{37}$$

是在启发性符号的记法：

$$\exp((\tau+\sigma)X)(x)=\exp(\tau X)(\exp(\sigma X)(x)) \tag{38}$$

下写成的。

所以，对给定的 X 计算流(即求解 n 个一阶微分方程的系统)被称为矢量场 X 的取幂。

备注：上述的对应可能与微分同胚映射和前推矢量场有关。特别是，如果两个矢量场，M 上的 X 和 N 上的 Y 通过 f：$M\to N$ 是 f 相关的，所以 $(Tf)(X(x))=Y(f(x))$，那么 f 引起了从 X 的积分曲线到 Y 的积分曲线的映射。我们可以根据 X 和 $Y=(Tf)(X)$ 的取幂来表达这一点：

$$f(\exp(\tau X)x)=\exp(\tau(Tf)(X))(f(x)) \tag{39}$$

备注：我强调，上述的(i)(ii)(iii)和(iv)之间的两两对应在单个点处是不

正确的。

(a) 一方面，(i)和(ii)之间的对应关系在一个点上适用，也适用于零矢量。即单一矢量 $X \in T_xM$ 对应于一个在 x 处的方向导数算符(导数)；$X=0$ 对应于映射在所有局域标量为 0 的零导数算符。(事实上，正如我所提到的，矢量通常定义为这样的算符/导数。)

(b) 另一方面，(i)和(iii)之间的对应关系，或者(i)和(iv)之间的对应关系都需要一个邻域。对于一个单一矢量 $X \in T_xM$ 对应一个通过 x 的曲线整体类，而不是对应一个单一曲线。也就是说，它对应于作为它们的切矢量 X 的所有曲线(流)。

然而，我们应当看到(3.4 节的开始)，对于一个有合适外部结构的流形，单一矢量确定一条曲线(我们将再次谈论取幂)。

我们需要概括上述对应的(i)—(iv)的某些方面，即矢量和方向导数之间的(i)—(ii)对应，这种概括就是李微商。

3.1.3　李导数

前面的小节已经简单地使用了李导数(微商)。因为我在后面还会大量用到，所以我们现在要详细介绍它。

我们已经看到，给定一个流形 M 上的矢量场 X，一个点 $x \in M$ 以及定义在 x 邻域上的任意标量函数，自然而然将有一个确定的沿着在 x 点处的 X 的 f 的变化率，即方向导数 $X(x)(f)$。

现在我们将沿着 X 方向的李导数定义为算符 $\mathcal{L}_X$，算符 $\mathcal{L}_X$ 定义沿着 X 方向的变化率：不仅针对局域定义的函数——这个定义和我们先前的是一致的，即我们有 $\mathcal{L}_X(f)=X(f)$，而且针对矢量场和微分 1 形式。[①] 我们按照三个阶段进行。

(1) 我们首先根据在 M 上的矢量场 X 来定义作为在标量函数上的算符的李导数。我们沿着场 X(又名：在 X 方向上的导数)定义李导数 $\mathcal{L}_X$，作为标量函数 $f: M \to \mathbb{R}$ 的算符被定义为：

---

① 事实上，这个定义可以被扩展到更高阶的张量上。但是，我将不会阐述这些细节，因为——除了 2.1.3 节提到的辛形式 $\mathcal{L}_X\omega$ 的李导数(即必要条件是，如果 X 是对称的，$\mathcal{L}_X\omega=0$)——我们将不需要它们。

$$\mathcal{L}_X: f \longmapsto \mathcal{L}_X f: M \to \mathbb{R} \text{ 有：} \forall x \in M: (\mathcal{L}_X f)(x) := \frac{d}{d\tau}\Big|_{\tau=0} f(X^\tau(x)) \equiv X(x)(f) \tag{40}$$

尽管这个定义假设 X 和 f 都被整体地定义，即在所有的 M 上，它能被限定在一个邻域内，$\mathcal{L}_X$是线性的且服从莱布尼茨规则，即：

$$\mathcal{L}_X(fg) = f\,\mathcal{L}_X(g) + g\,\mathcal{L}_X(f) \tag{41}$$

在坐标 $\mathbf{x} = x^1, \cdots, x^n$ 中，$\mathcal{L}_X f$ 通过下式给定：

$$\mathcal{L}_X f = X^1(\mathbf{x})\frac{\partial f}{\partial x^1} + \cdots + X^n(\mathbf{x})\frac{\partial f}{\partial x^n} \tag{42}$$

$X^i(\mathbf{x})$是矢量 $X(x)$的第 i 分量。式(42)意味着尽管式(40)提及到流 X^τ，标量的李导数和先前方向导数的概念是一致的，即对于所有的 f，$\mathcal{L}_X(f) = X(f)$。

(2) 在(1)中，矢量场 X 确定算符$\mathcal{L}_X$，依据 3.1.2.2 所对应的，我们从(i)转移到(ii)。但是我们可以反过来依据李导数来定义一个矢量场，在 3.2.2 小节关于李括号的讨论中，我们将确切地做到这一点。更加详细地，我们注意到，所有 M 上的标量场的集合$\mathcal{F}(M)$，f: $M \to \mathbb{R}$形成了在逐点增加下的实矢量空间，所以也做了在 M 上矢量场的集合$\mathcal{X}(M)$，X: $x \in M \longmapsto X(x) \in T_x M$。此外，$\mathcal{X}(M)$作为实矢量空间且作为标量场上的模与算符$\mathcal{L}_X$的集合是同构的，此同构通过定义在(1)上的映射 θ: $X \longmapsto \mathcal{L}_X$给定。

(3) 我们现在延伸$\mathcal{L}_X$的定义以便将它定义在矢量场 Y 和 1 形式 α 上。我暂时用 θ 作为对于矢量场或者微分 1 形式的符号。给定一个矢量场 X 和流 $X^\tau \equiv \phi_X(\tau)$，我们需要比较点 $x \in M$ 处的 θ 与在附近的点 $X^\tau(x) \equiv \phi_{X,x}(\tau)$处的 θ。但在 $X^\tau(x)$处 θ 的值在切空间或者余切空间内分别为 $X^\tau(x)$：$T_{X^\tau(x)}M$ 或 $T^*_{X^\tau(x)}M$。所以通过对比，我们需要以某种方式将这个值传回 $T_x M$ 或 $T^*_x M$。

幸运的是，矢量场 X 提供了一个自然的方式来定义这样一个传递(transport)。对于矢量场 Y，我们使用逆流(inverse flow)的微分(即前推)从 $X^\tau(x)$ "返回"到 x。

对于这个 $\phi_{X,x}(\tau)$的"回拉"的使用，我们定义：

$$\phi^*(\tau) := T(\exp(-\tau X)) \equiv d\exp(-\tau X): T_{X^\tau(x)}M \equiv T_{\exp(\tau X)(x)}M \to T_x M \tag{43}$$

对于我们已经在式(32)定义的 1 形式 α，我们通过"回拉"定义这个传递：

$$\phi^*(\tau) := (\exp(-\tau X))^* : T^*_{X^\tau(x)}M \equiv T^*_{\exp(\tau X)(x)}M \to T^*_x M \tag{44}$$

由于$\phi^*(\tau)$的定义，当θ是矢量场Y或者微分1形式α，我们将李导数$\mathcal{L}_X\theta$定义为矢量场或者微分1形式，在x处的值由下式给出：

$$\lim_{\tau\to 0}\frac{\phi^*(\tau)(\theta|_{X^\tau(x)}) - \theta|_x}{\tau} = \frac{d}{d\tau}|_{\tau=0}\phi^*(\tau)(\theta|_{X^\tau(x)}) \tag{45}$$

最后，一个说明本章“幕后花絮”的附带结论。从2.1.3小节开始，它连接了诺特定理和这一节关于李导数的详情，还连接了诺特定理和叙述常微分方程解的局域存在性和唯一性的定理(参见3.1.2.2小节的开始部分)。后面的定理意味着在任意流形上的任意矢量场X可能被“拉直”，在这个意义上，在任意点(在这个点处X是非零的)的周围，有一个局域坐标系，在这个坐标系内，X除了一个不为零，其他所有的分量都等于零，最后一个等于1。通过使用这个定理，我们明确地表示，对于一些辛形式，在任意偶维流形上的任意矢量场X都是局域的哈密顿量，就某些辛形式而言，在某一点附近的X是非零的。可仅仅通过从一个横断面到X的积分曲线来定义辛形式。

3.2 李代数和李括号

现在我介绍两个矢量场的李代数和李括号。

3.2.1 李代数

李代数是一个具有双线性反对称算符的矢量空间V，通常通过方括号表示(称为“括号”或“对易子”)$[\,,]: V\times V\to V$，它将满足雅可比恒等式，即：

$$[[X, Y], Z] + [[Y, Z], X] + [[Z, X], Y] = 0 \tag{46}$$

3.2.1.1 实例：引入转动

这里有三个例子。

(i)$n\times n$矩阵具有通常对易子，即$[X, Y] := XY - YX$(因此，矩阵乘法有助于对易，而不是潜在的矢量场空间结构)；

(ii)具有通常对易子的3×3反对称矩阵；

(iii) 具有矢量乘法的$\mathbb{R}^3$。

事实上，实例(iii)本质上和实例(ii)是相同的，这个实例将在下文中反复出现，与刚体和转动有关。我们也可以看到实例(ii)在某种意义上将是更基

础的。

为了解释这一点，我首先回顾，每一个在三维定向欧氏空间上的反对称算符 A 都是用确定的矢量(如 ω)的矢量乘法的算符。即对于所有的 $\mathbf{q}$，$A\mathbf{q}=[\omega, \mathbf{q}]\equiv\omega\wedge\mathbf{q}$。(证明：对于一个三维矢量空间的$\mathbb{R}^3$ 上的反对称算符，因为一个反对称 3×3 矩阵有三个独立的分量。用矢量 ω 的矢量乘法是线性的且是反对称的算符，对于不同的 ω 我们将得到在$\mathbb{R}^3$ 上的所有反对称算符的空间的子空间，但是这个子空间的维度是 3，因此，它与所有反对称算符的空间相一致。)

对于所掌握的结果，下列三点将很容易证实。

(1) 笛卡尔坐标系中的 A 的矩阵表示是：

$$A=\begin{pmatrix} 0 & -\omega_3 & \omega_2 \\ \omega_3 & 0 & -\omega_1 \\ -\omega_2 & \omega_1 & 0 \end{pmatrix} \tag{47}$$

我们可以写作：

$$A\leftrightarrow\omega \text{ 或 } A_{ij}=-\epsilon_{ijk}\omega_k \text{ 或 } \omega_i=-\frac{1}{2}\epsilon_{ijk}A_{jk} \tag{48}$$

(2) 垂直于矢量 ω 的平面 Π 是对于 A 的不变子空间，即 $A(\Pi)=\Pi$。ω 是关于 A 的本征值为零的本征矢量。这让我们想起了常见的以后被证实(3.4 小节)的基本解释，也就是，任意 3×3 反对称矩阵 A 表示一个无穷小转动，ω 表示瞬时角速度。即对于所有的 $\mathbf{q}\in\mathbb{R}^3$：有 $\dot{\mathbf{q}}=A\mathbf{q}=[\omega, \mathbf{q}]$。

(3) 任意两个 3×3 反对称矩阵 A，B 的对易子，即$[A, B]:=AB-BA$，通过式(48)和转动轴的矢量乘法相一致。即式(48)的从矢量到矩阵的双射被写作 Θ：$\omega\longmapsto A=:\Theta(\omega)$，对于矢量 $\mathbf{q}$，$\mathbf{r}$，$\mathbf{s}$，我们有：

$$(\Theta(\mathbf{q})\Theta(\mathbf{r})-\Theta(\mathbf{r})\Theta(\mathbf{q}))\mathbf{s}=\Theta(\mathbf{q})[\mathbf{r}, \mathbf{s}]-\Theta(\mathbf{r})[\mathbf{q}, \mathbf{s}] \tag{49}$$

$$=[\mathbf{q}, [\mathbf{r}, \mathbf{s}]]-[\mathbf{r}, [\mathbf{q}, \mathbf{s}]] \tag{50}$$

$$=[[\mathbf{q}, \mathbf{r}], \mathbf{s}]=\Theta([\mathbf{q}, \mathbf{r}])\cdot\mathbf{s} \tag{51}$$

$[,]$表示矢量乘积，即$[\mathbf{q}, \mathbf{r}]\equiv\mathbf{q}\wedge\mathbf{r}$。

式(51)意味着 Θ 给出了李代数同构，因此我们的实例(iii)本质上和实例(ii)相同。

此外，我们已经能初窥为什么实例(ii)在某种意义上是更基础的。因为式

(48)的反对称矢量(或矩阵)和矢量之间的这个对应是针对三维空间的。在 n 维空间里，反对称矩阵的独立分量的数目是 $n(n-1)/2$，仅当 $n=3$ 时，这将等于 n，然而我们将在后面(3.4.4 小节)看到在任意 n 维欧氏空间 $\mathbb{R}^n$ 上的转动通过 $n\times n$ 反对称矩阵的李代数而产生。因此，仅仅对于 $n=3$，通过在 $\mathbb{R}^n$ 上的矢量，存在一个转动的对应表示。

在下两个小节中，我们将看到李代数的其他例子如它的矢量是矢量场(3.2.2 小节)或者在李群(3.4 节)的单位元素中的切矢量。第一个例子将是无限维的李代数，第二个是有限维的(因为我们将仅仅考虑有限维李群)。此外，上述的(i)和(ii)的例子每一个都将是在李群的单位元素中的切矢量的矢量空间。

3.2.1.2　结构常量

相对于基，有限维李代数的特点是一组具有称为结构常量的数，这个结构常量定义括号算符。因此，如果 $\{v_1, \cdots, v_n\}$ 是李代数 V 的基，那么我们通过延伸定义这个结构常量为 c_{ij}^k，其中 $i, j, k=1, \cdots, n$，依据这个基，任意两个基本元素的括号为：

$$[v_i, v_j] = \sum_k c_{ij}^k v_k \tag{52}$$

括号的双线性意味着式(52)决定了所有矢量对 $v, w\in V$ 的对易。对于任意的基，这个括号的反对称性和雅可比恒等式意味着这个结构常量服从：

$$c_{ij}^k = -c_{ji}^k;\quad \sum_k (c_{ij}^k c_{kl}^m + c_{li}^k c_{kj}^m + c_{jl}^k c_{ki}^m) = 0 \tag{53}$$

相反地，服从式(53)的常量 c_{ij}^k 的任意集合是 n 维李代数的结构常量。

3.2.2　两个矢量场的李括号

给定流形 M 上的两个矢量场 X，Y，对应流不存在基本的对易：$X^t Y^s \neq Y^s X^t$。非对易性通过 X 和 Y 的李导数的对易子(也就是 $\mathcal{L}_X\mathcal{L}_Y - \mathcal{L}_Y\mathcal{L}_X$)进行判断，参见对于李导数定义的式(40)和式(45)。在这里，"判断"可以通过考虑泰勒展开确切地做到。

对于我们来说，重要的是这个初看起来似乎是二阶矢量的对易子事实上是一阶矢量。这通过坐标系里的计算证实，以及由于异号二阶导数会发生两次：

$$(\mathcal{L}_X\mathcal{L}_Y-\mathcal{L}_Y\mathcal{L}_X)f=\Sigma_i X^i\frac{\partial}{\partial x^i}\left(\Sigma_j Y^j\frac{\partial f}{\partial x^j}\right)-\Sigma_j Y^j\frac{\partial}{\partial x^j}\left(\Sigma_i X^i\frac{\partial f}{\partial x^i}\right) \tag{54}$$

$$=\cdots=\Sigma_{i,j}\left(X^i\frac{\partial Y^j}{\partial x^i}-Y^i\frac{\partial X^j}{\partial x^i}\right)\frac{\partial f}{\partial x^j} \tag{55}$$

因此，$\mathcal{L}_X\mathcal{L}_Y-\mathcal{L}_Y\mathcal{L}_X$对应于一个矢量场，可回想 3.1.3 中的(2)，从它的李导数定义矢量场。我们称这个场 Z 为场 X 和 Y 的李括号(又称泊松括号、对易子和雅可比—李括号)，将它写为$[X, Y]$。也可写成$\mathcal{L}_XY$以及将其称为关于 X 的 Y 的李导数(注意有的书中采用相反的符号法则)。

因此 $Z\equiv[X, Y]\equiv\mathcal{L}_XY$ 被定义为如下的矢量场：

$$\mathcal{L}_Z\equiv\mathcal{L}_{[X,Y]}=\mathcal{L}_X\mathcal{L}_Y-\mathcal{L}_Y\mathcal{L}_X \tag{56}$$

结果就是，$Z\equiv[X, Y]$在坐标系的分量通过式(55)给定。这个公式可将它(用求和缩写法，即省略了 Σ)写为下式：

$$\left[X^i\frac{\partial}{\partial x^i},\ Y^j\frac{\partial}{\partial x^j}\right]=X^i\frac{\partial Y^j}{\partial x^i}\frac{\partial}{\partial x^j}-Y^j\frac{\partial X^i}{\partial x^j}\frac{\partial}{\partial x^i} \tag{57}$$

式(55)的另一种表达方式为：

$$[X,\ Y]^j=(X\cdot\nabla)Y^j-(Y\cdot\nabla)X^j \tag{58}$$

或者不用坐标，对于通过雅可比矩阵给定的微分映射 $\mathbf{D}$，可写为：

$$[X,\ Y]=\mathbf{D}Y\cdot X-\mathbf{D}X\cdot Y \tag{59}$$

此外，矢量场 $Z\equiv[X, Y]$可以检测流 X^t 和 Y^s 的非对易性，尤其当且仅当$[X, Y]=0$ 时，这些流是否对易。

我们需要三个关于李括号的结论，它们分别与李代数、泊松括号和弗罗宾尼斯定理相关。

(1) 李括号明显是在流形 M 上所有矢量场的(无限维)矢量空间上的双线性算符和反对称算符，即$[\,,]:\mathcal{X}(M)\times\mathcal{X}(M)\to\mathcal{X}(M)$。已证明它满足雅可比恒等式(展开$\mathcal{L}_{[[X,Y],Z]}=\mathcal{L}_{[X,Y]}\mathcal{L}_Z-\mathcal{L}_Z\mathcal{L}_{[X,Y]}$等)，因此，$\mathcal{X}(M)$是(无限维)李代数。

(2) 返回到哈密顿力学(2.1 节)，通过哈密顿矢量场的概念，存在一个李括号和泊松括号之间的简单和基本的关系。即在辛流形 M 上的两个标量函数 f 和 g 的泊松括号的哈密顿矢量场是哈密顿矢量场的李括号，对应 f 的是 X_f，对

应 g 的是 X_g：

$$X_{\{f,g\}} = -[X_f,\ X_g] = [X_g,\ X_f] \tag{60}$$

证明：将 rhs 应用于任意标量 h：$M\to\mathbb{R}$。我们很容易通过使用下列内容得到 $X_{\{f,g\}}(h)$：

(i)哈密顿矢量场的定义；

(ii)函数的李导数等于它的方向导数的分量，参见式(40)；

(iii)泊松括号是反对称的且遵守雅可比恒等式。

这个结果意味着在辛流形 M 上且具有泊松括号的哈密顿矢量场形成了一个在辛流形 M 上的所有矢量场的李代数 $\mathcal{X}(M)$ 的李子代数。后来，这个结果从辛流形延伸到泊松流形(5.2.2 小节详细说明)，这点很重要。

(3) 对于弗罗宾尼斯定理(3.3 小节)，我们需要将李括号和 3.1.2 小节的矢量场的观点联系起来，这个矢量场通过流形 M 和 N 之间的映射 f: $M\to N$ 而与 f 相关。简而言之，如果两对矢量场是 f 相关的，因此就是它们的李括号。更明确地，如果 X，Y 是流形 M 上的矢量场，且 f: $M\to N$ 是一个映射，那么 $(Tf)(X)$ 和 $(Tf)(Y)$ 是在 N 上定义明确的矢量场，然后 Tf 和李括号对易：

$$(Tf)[X,\ Y] = [(Tf)X,\ (Tf)Y] \tag{61}$$

3.3　子流形和弗罗宾尼斯定理

这个小节不同于先前的以三种方式论述的小节。第一，它强调了积分的而不是微分的概念。

第二，3.1.2.2 节已经强调矢量场的积分曲线对应于常微分方程的积分系统。因为这样的曲线是给定流形的一维子流形。我们现在的话题，也就是高维子流形自然地给出了偏微分方程。因为，给定一个切空间 T_xM 的子空间 S_x 中流形 M 上的每个点 x 的赋值，它们的积分包括找到一个积分曲面，即 M 的子

流形 S，其中 M 在每一个点上的切空间是 S_x。①

然而，我们将不再关注偏微分方程。对于我们而言，当矢量场的一个集合在 x 处的切矢量的生成空间组合在一起形成一个子流形的时候，高于一维的子流形就出现了。因此，弗罗宾尼斯定理认为，粗略地说，当且仅当矢量场处于对合时，矢量场的有限集才是可积的。即，当且仅当它们成对的李括号关于场是可展开的，也就是说这个矢量场形成了矢量场全部的李代数的李子代数。我们不需要去证明这个定理，但是我们需要陈述它并且运用它——尤其对于泊松流形的叶状结构(foliation)。

第三，前车之鉴是有道理的。一个子集 $S \subset M$ 的直觉观点认为一个平滑流形本身就可以以不同的方式确定。因此，存在关于"子流形"的定义的精妙之处，且术语在解释之间存在差异——在某种程度上，它不是为了之前小节则提供的材料。我将采取看起来似乎普遍的(也许不是大多数采用的)术语。②

3.3.1 子流形

基本的定义是：

给定一个流形 $M(\dim(M)=n)$，k 维的 M 的子流形是一个子集 $N \subset M$，那么对于任意 $y \in N$，都存在一个可接受的局域图表(local chart)(U，ϕ)(即在 M 上的最大图表集内的图表)，其中 $y \in U$ 伴随有子流形特征(submanifold property)，也就是：

$$(\mathrm{SM}).\ \phi: U \to \mathbb{R}^k \times \mathbb{R}^{n-k} \text{和}\ \phi(U \cap N) = \phi(U) \cap (\mathbb{R}^k \times \{0\}) \qquad (62)$$

集合 N 变成了一个流形，它通过形式为($U \cap N$，$\phi \mid (U \cap N)$)的所有图表

---

① 注意：对于定域存在的偏微分方程和常微分方程的唯一性定理，不存在相似的地方。即使一个在三维空间的二维平面的场通常是不可积的，例如，通过方程 $dz = ydx$ 给出的平面场，因此，平面的可积场，或者流形上的其他子空间都是一个例外，从而对于偏微分方程的积分理论要比对于常微分方程的积分理论更加不统一，且更加复杂。

② 对于3.3.1小节我的处理是基于[Marsden and Ratiu，1999，pp. 124—127，140]，对于3.3.2小节则是基于[Olver，2000，pp. 38—40]。关于不同的术语，奥尔弗[2000，9]将"子流形"定义为我们所称的浸入子流形(immersed submanifold)，对于我们而言，以后没必要是子流形，因为，浸入不需要是一个嵌入(embedding)。毕肖普和戈登伯格(Bishop and Goldberg)[1980，pp. 40—41]提供了一个相似的例子。对于子流形的不同观念的详细介绍，参见达令(Darling)[1994，Chapters 3 and 5]。我将省略一些细节，尤其是关于提供规则浸入的弗罗宾尼斯定理。

的图表集(atlas)产生，其中(U, ϕ)是具有子流形属性的M上的图表(这使得N的拓扑为相对拓扑)。

我们需要注意两种方式，在这两种方式下，子流形可能根据流形之间的平滑函数被确定。

(1) 子流形可以被确定为集合，在这个集合上，流形之间的平滑函数$f: M \to P$具有某一特定值。实际上，这将是式(62)的必要条件的一个概括，这个必要条件是图表ϕ的$(n-k)$个坐标分量都等于零。这将涉及这样一个观点：切映射Tf是满射的，在这种情况下的f被称为一个浸没(submersion)。对于李群的作用量的商，我们需要这一方法。

(2) 一个子流形可以参量地被确定，它作为局域参量的值集，即作为平滑函数f的值域，其中M为定义域。这就包含了这样一个观点：切映射Tf是内射的，在这种情况下的f被称为浸入(immersion)。对于弗罗宾尼斯定理，我们需要这一方法。

①浸没

如果$f: M \to P$是流形间的平滑映射，且切映射$T_x f$是满射的，点$x \in M$被称为正则点；否则x就是f的临界点。如果$C \subset M$是M的临界点的集合，我们就说$f(C)$是f的临界值的集合，$P - f(C)$是f的临界值的集合。所以，如果$p \in P$是f的正则值，那么，在任意点$x \in M$且$f(x) = p$，$T_x f$就是满射的。

浸没定理(submersion theorem)认为，如果$p \in P$是f的正则值，那么，(i)$f^{-1}(p)$是$\dim(M) - \dim(P)$维的子流形；(ii)在任意点$x \in f^{-1}(p)$的这个子流形的切空间是f的切映射的核(kernel)：

$$T_x(f^{-1}(p)) = \ker T_x f \tag{63}$$

如果$T_x f$对于所有的$x \in M$都是满射的，f被称为一个浸没。

②浸入

当$T_x f$在所有的点$x \in M$处都是内射的，流形间的平滑映射$f: M \to P$被称为一个浸入。浸入定理认为，当且仅当存在一个M内的x的邻域U使得$f(U)$是P的子流形，且$f|_U: U \to f(U)$是微分同胚映射，那么$T_x f$是内射的。

注意：这并不是说，$f(M)$是P的子流形。因为f可能不是内射的，以致$f(M)$有自相交。即使f是内射的，f可能不是M和$f(M)$之间的微分同胚映射。

一个简单的标准例子是将$\mathbb{R}$的开区间内射到$\mathbb{R}^2$中的“几乎闭合的”八字形。

然而，当f: $M \to P$是一个浸入，且是内射的，则我们称$f(M)$是内射地浸入的子流形(简短地说是一个浸入子流形)。

我们将一个嵌入认为是浸入，浸入是M和$f(M)$($f(M)$是从P产生的相对拓扑)之间的同胚(因此也是内射的)。如果f是一个嵌入，$f(M)$就是N的子流形，且f是微分同胚映射f: $M \to f(M)$。

事实上，弗罗宾尼斯定理将提供内射地浸入的子流形，这个内射地浸入的子流形不需要是被嵌入的，因此不需要是子流形(它们也要遵守其他的条件，称为“正则性”，我将不继续深入说明)。

3.3.2 定理

在3.2.2节的最后我们看到，如果两对矢量场是f相关的，它们的李括号可参见式(61)的形式。对于正切于嵌入的子流形的两个矢量场，这个结果立即形成了一个必要条件，即：

如果X_1，X_2是M上的矢量场，M正切于一个嵌入的子流形S(即在所有点$x \in S$，$X_i(x) \in T_xS < T_xM$)，那么它们的李括号$[X_1, X_2]$也正切于S。

这将通过考虑M里产生的一个S的嵌入的微分同胚映射f: $\tilde{S} \to S$继续进行。在式(61)中，Tf和李括号对易，即在$\tilde{S}$上且正切于$\tilde{S}$的f相关的矢量场$\tilde{X}_1$和$\tilde{X}_2$的李括号通过Tf被带到X_1和X_2的泊松括号$[X_1, X_2]$。因此，$[X_1, X_2]$正切于S。

弗罗宾尼斯定理的理念是正切于子流形的两个矢量场的必要条件也是充分条件。更具体地说，我们需要如下定义：

流形M上的分布函数(distribution)D是正切从TM的子集，在所有的点$x \in M$上，$D_x := D \cap T_xM$有一个矢量空间。D_x的维度是在x上的D的秩。如果D的秩是M上的常量，我们就说广义函数是正则的。

如果对于所有的$x \in M$，且$X_0 \in D_x$，有广义函数是平滑的，则存在x的邻域$U \subset M$，且U上的平滑矢量X使得：(i)$X(x) = X_0$；(ii)对于所有的$y \in U$，$X(y) \in D_y$。这样的矢量场X被称为D的局部剖面。例如：定义在M上的r个矢量场X_1，…，X_r的集合共同定义了一个秩(这个秩最大是r)的平滑广义函数。

如果对于局部剖面的任意对 X_1，X_2，在这两个剖面共同定义域的李括号 $[X_1,\ X_2](y)\in D_y$，那么我们就说广义函数是对合的。

类似地，我们说，如果它们在 M 的任何地方都生成李括号，那么在流形 M 上的 r 个平滑矢量场 X_1，…，X_r 的集合就是对合的。即，存在平滑实函数 h_{ij}^k：$M\to R$，i，j，$k=1$，…，r 使得在所有的点 $x\in M$，有：

$$[X_i,\ X_j](x)=\sum\nolimits_k h_{ij}^k(x)X_k(x) \tag{64}$$

注意：对合被应用在和刘维尔定理（Liouville's theorem）相关的不同情况下，即当所有它们的成对的泊松括号都等于零的时候，在相空间上的实函数集被认为就是对合的。

如果对于任意 $x\in M$，存在 M 上的局域子流形，其中 M 的正切丛等于 D 到 $N(x)$ 的限制，那么分布函数 D 就是可积的。如果 D 是可积的，通过所有的 $x\in M$，变量 $N(x)$ 可能被展开以得到唯一的最大连通集；其中，它在每个分量下的切空间是 D_y。这样的一个集合被称为（最大的）积分流形。

注意：一般来说，每个积分流形都内射地嵌入在 M 里，因此通过 3.3.1 小节中的（2）的讨论，一个积分流形可能不是 M 的子流形。但是，和大多数处理一样，我将忽略这一点，且认为它们是子流形和积分流形。

如果 D 的秩是 M 上的常量，所有的积分流形都具有一个共同的维度，即 D 的秩。但是，大体上说，D 的秩随着 M 变化，因此，积分流形的维度也是如此变化。

同样地，我们说，r 个矢量场 X_1，…，X_r 的集合是可积的，即如果对于每个点 $x\in M$，那么将出现 M 的局域子流形 $N(x)$，其中在它的每一个点处 M 的切空间将由 X_1，…，X_r 生成。此外，我们允许，在一些点 x，$X_1(x)$，…，$X_r(x)$ 可能是线性相关的，所以子流形的维度随着发生了变化。

我们可以说（对于分布函数和矢量场集），积分流形的集合是 M 的叶状结构，它的分量是叶（leaves）。此外，如果叶的维度是 M 上的常量，我们说这个叶状结构是正则的。

根据这些已有的定义，我们现在可以陈述弗罗宾尼斯定理，即涉及常秩情形的两种一般形式，即正则分布函数的情形和处处都是线性独立矢量场的情形，还有一种广义的形式。一般形式是：

弗罗宾尼斯定理(一般形式)：当且仅当平滑正则分布函数(regular distribution)是对合的，那么它就是可积的。

或者依据矢量场当且仅当在流形 M 上的 r 个矢量场 X_1，…，X_r 的集合是对合的，且处处都是线性独立的，那么就说它是可积的。

这个概括来自两个阶段。第一阶段涉及变化的秩，但是也假定了矢量场的有限集。简言之，这是由同一个陈述得出的。也就是说，当且仅当在流形 M(也许并不是处处都线性独立的)上的 X_1，…，X_r 个矢量场 X_1，…，X_r 的集合是对合的，那么就说它是可积的。

但是，对于泊松流形的叶状结构(5.3.3 节)，我们需要考虑一个矢量场的有限集，或许带有不同的秩；且对于这样一个集合，这个陈述不成立。值得庆幸的是，存在一个如下所示有用的推广。

假设$\mathcal{X}$是流形 M 上的形成矢量场的矢量场集。所以，在上面 r 个矢量场的讨论中，$\mathcal{X}$被认为是所有的线性组合 $\sum_{i=1}^{r} f_i(x)X_i(x)$，$x \in M$，其中 f_i 是任意的平滑函数 f：$M \to \mathbb{R}$，这样的$\mathcal{X}$被称为有限生成。

对于任意形成矢量空间的$\mathcal{X}$，如果无论何时都有 X，$Y \in \mathcal{X}$时$[X, Y] \in \mathcal{X}$，那么我们就说(如之前提到的)$\mathcal{X}$是对合的。假设对于所有 $X \in \mathcal{X}$，$\mathcal{X}_x$是通过 $X(x)$生成的 T_xM 的子空间。如以前一样，我们定义$\mathcal{X}$的积分流形是 $N \subset M$ 的子流形，使得对于所有的 $y \in N$，有 $T_yN = \mathcal{X}_y$，并且，当且仅当对于任意 $x \in M$，都能传递一个积分流形，那么$\mathcal{X}$就被称为可积的。

如前所述，如果$\mathcal{X}$是可积的，那它就是对合的，但是反过来就不成立。需要一个额外的条件，如下所述。

如果对于任意矢量场 $X \in \mathcal{X}$，沿着 X 方向产生的流的子空间$\mathcal{X}_{\exp(\tau X)(x)}$的维度是独立于 τ(但是它可能依赖于点 x)的一个常量，那么我们就说$\mathcal{X}$是秩不变量。

因为通过 x 的积分曲线 $\exp(\tau X)(x)$ 应该被包含在任意的积分子流形，所以，秩不变量当然是可积性的必要条件。事实上，我们有如下定理。

弗罗宾尼斯定理(广义形式)：当且仅当它是秩不变量且是对合的，那么 M 上的矢量场系统$\mathcal{X}$就是可积的。

证明的方法就是去直接地建构积分子流形。通过 x 的子流形从下列形式获得：

$$N=\{\exp(X_1)\exp(X_2)\cdots\exp(X_p)(x):\ p\geqslant 1,\ X_i\in\mathcal{X}\} \tag{65}$$

对于任意的 $y\in N$，秩不变性确保$\mathcal{X}_y$的维度是 $\dim(N)$。

3.4　李群和它们的李代数

我将介绍李群和它们的李代数。通过上两小节（3.4.3 和 3.4.4 小节），我们将有足够的理论来有效地计算重要的李群和转动群的李代数。

3.4.1　李群和矩阵李群

李群是一个群 G，群 G 也是一个流形，对于李群，乘积 $G\times G\to G$ 和逆运算 $G\to G$ 都是平滑的。

实例：

(i)加法下的$\mathbb{R}^n$。

(ii)$\mathbb{R}^n$ 到$\mathbb{R}^n$ 的线性同构群，由 $GL(n,\ \mathbb{R})$ 表示并被称为一般线性群，用实可逆矩阵 $n\times n$ 表示。这是$\mathbb{R}^{n^2}$的一个开子集，因此是一个 n^2 维的流形，对于矩阵的乘积和逆，这个公式在矩阵分量内是平滑的。

(iii)关于$\mathbb{R}^3$ 的原点的转动群，通过行列式为 1 的 3×3 矩阵来表示，写作 $SO(3)$，其中，S 代表“特别的”（即行列式为 1），O 代表“正交的”。

事实上，这三个实例被认为是矩阵李群，其中矩阵乘法就是算符。在实例(i)中，考虑到加法下的$\mathbb{R}^n$ 和对角线元素都等于 1 的 $(n+1)\times(n+1)$ 矩阵之间的同构 θ，最右列的元素等于$\mathbb{R}^n$ 中给定的矢量，其他元素都等于 0，因此，对于 $n=3$，我们有：

$$\theta:\ \begin{pmatrix} x\\ y\\ z\end{pmatrix}\longmapsto\begin{pmatrix}1&0&0&x\\0&1&0&y\\0&0&1&z\\0&0&0&1\end{pmatrix} \tag{66}$$

这意味着，我们定义了一个矩阵李群为可逆实矩阵的任意集合，在矩阵乘法下，这个可逆实矩阵在乘法、求逆和取极限下都是封闭的。一个矩阵李群是一个李群，这将由作为李群的 $GL(n,\ \mathbb{R})$ 和下面的定理（在 3.4.3 节）推出。这个定理认为任意李群的封闭子群本身就是李群。

对于矩阵李群，具有如下简化理论。例如，这个群的李代数的一个分量的

幂的定义归纳为矩阵的幂。但是我们将提出这个基本理论的一部分理论，因为它既是发人深省的，也是强有力的。

3.4.2 李群的李代数

本小节的主要结论是，对于任意李群 G，在单位元素 $e \in G$ 上的切空间 T_eG 很自然有一个李代数结构，其中这个李代数结构是通过 G 上特定的矢量场产生的，如下所示。

3.4.2.1 左不变矢量场定义李代数

假设 G 是李群，任意 $g \in G$ 通过左平移将 G 的微分同胚映射定义到它自身，通过右平移也是相似的：

$$L_g:\ h \in G \longmapsto gh \in G;\ R_g:\ h \in G \longmapsto hg \in G \tag{67}$$

注意：在第4节，考虑到在式(67)中，G 作用在它自身上的左和右平移，我们将在群作用量的语言下描述这些。

现在考虑这个在切空间上的归纳映射，即切(又叫作微分)映射，参见式(29)和式(30)。它们是 $(L_g)_* =: L_{g*}$，$(R_g)_* =: R_{g*}$，其中对于任意 $h \in G$：

$$L_{g*}:\ T_hG \to T_{gh}G \text{ 且 } R_{g*}:\ T_hG \to T_{hg}G \tag{68}$$

尤其是在单位元素 $e \in G$ 上的导数 $(R_g)_*$ 将 T_eG 映射到 T_gG。这意味着，每一个矢量 $\xi \in T_eG$ 定义了 G 上的矢量：在任意 $g \in G$ 上的值是 ξ 在 $(R_g)_*$ 下的像 $(R_g)_*\xi$。这样的矢量场被称为右不变矢量场，它通过在单位元素 $e \in G$ 上的值唯一地定义。

更详细地，现在定义左不变矢量场。如果对于任意的 $g \in G$，都有 $(L_g)_*X = X$，那么 G 上的矢量场 X 就被称为左不变矢量场，对于 L_g 在 h 处的导数或切线，让我们将其写作 T_hL_g，即对于 L_{g*}：$T_hG \to T_{gh}G$。左不变式要求对于每一个 g 和 $h \in G$，都有：

$$(T_hL_g)X(h) = X(gh) \tag{69}$$

因此，每一个矢量场 $\xi \in T_eG$ 都定义一个左不变矢量场，写作 G 上的 X_ξ，在任意 $g \in G$ 上，X_ξ 的值就是 ξ 在 $(L_g)_*$ 下的像 $(L_g)_*\xi$。换句话说，$X_\xi(g) := (T_eL_g)\xi$。

不仅仅是左不变矢量场通过在单位元素 $e \in G$ 上的值唯一地定义。而且，G

上的左不变矢量场集$\mathcal{X}_L(G)$作为矢量空间与在单位元素 e 上的切空间 T_eG 同构。线性矢量 α，β 通过下式定义：

$$\alpha: X \in \mathcal{X}_L(G) \longmapsto X(e) \in T_eG \text{ 和 } \beta: \xi \in T_eG \longmapsto \{g \longmapsto X_\xi(g) := (T_eL_g)\xi\} \in \mathcal{X}_L(G) \tag{70}$$

给出单位映射：

$$\beta \circ \alpha = id_{\mathcal{X}_L(G)}; \quad \alpha \circ \beta = id_{T_eG} \tag{71}$$

$\mathcal{X}_L(G)$是 G 上的所有矢量场的李代数的李子代数，因为它在李括号下是闭合的。即左不变矢量场 X 和 Y 的李括号自身就是左不变的，因为，对于每一个 $g \in G$，我们有：

$$L_{g*}[X, Y] = [L_{g*}X, L_{g*}Y] = [X, Y] \tag{72}$$

上式中，L 意思是“左(Left)”而不是“李(Lie)”。如果我们现在通过下式定义 T_eG 上的括号：

$$[\xi, \eta] := [X_\xi, X_\eta](e) \tag{73}$$

那么，T_eG 成为一个李代数。被称为 G 的李代数，写作$\mathfrak{g}$，或为了避免对所考虑李群的歧义，写作$\mathfrak{g}(G)$。它由式(72)得出，那么：

$$[X_\xi, X_\eta] = X_{[\xi,\eta]} \tag{74}$$

也就是说，映射 α，β 是李代数同构的。

T_eG 有一个自然的李代数结构，这个结论是非常重要的。因为，正如我们将在 3.4 节的剩余部分看到的那样，李群的结构主要通过这个李代数的结构确定。因此，正如我们将在 4 节和 5 节以及下列等式看到的那样，这个李代数巩固了大部分由李群构成的结构，例如在李群作用量下的结构。奥尔弗写道，这个结论“是李群理论的基石……李群应用于微分方程的几乎所有的领域最终停留在这个构造中”[Olver, 2000, 42]。

在进入下一小节的例子以及子群和子代数的话题之前，我将以(1)—(4)四个结论结束本节，它们在后面也是重要的结论，需要引起重视。

3.4.2.2 四个结论

(1) 在下文中，李群结构决定了李代数结构。如果 G，H 是李群，且 f: $G \to H$是平滑的同态映射(homomorphism)，那么 f 在恒等式 T_ef: $\mathfrak{g}(G) \to$

$\mathfrak{g}(H)$上的导数是李代数同态映射。尤其是，对于所有的ξ，$\eta\in\mathfrak{g}(G)$，可以得到，$(T_e f)[\xi,\eta]=[T_e f(\xi),T_e f(\eta)]$，参见式(61)。

(2) 再次取幂，左不变矢量场和一维子群之间的对应。回想到3.1节，尤其是式(36)，在流形G上的矢量场X决定了包含单位元素e的G内的积分曲线ϕ_X，其中：$\phi_X(0)=e$。我们现在将这个积分曲线内的点写成g_τ(x与e是可理解的)：

$$\exp(\tau X)(e)\equiv X^\tau(e)\equiv\phi_{X,e}(\tau)=:g_\tau \tag{75}$$

它明确地表示，如果X是左不变的，这个曲线就是G的一个单参数子群，即不仅仅如式(35)以及下列等式那样的流形G上的微分同胚映射群的单参数子群。事实上：

$$g_{\tau+\sigma}=g_\tau g_\sigma,\ g_0=e,\ g_\tau^{-1}=g_{-\tau} \tag{76}$$

此外，这个群是在$\tau\in\mathbb{R}$下定义的，所以它是和$\mathbb{R}$或圆群$SO(2)$同构的。相反地，G的任意相关单参子群都是以这种方式通过一个左不变矢量场产生的。

于是，我们通过参考式(70)和式(71)的同构定义$\mathfrak{g}$的元素ξ的幂。我们也方便地将这定义为G上取值的映射。因此，对于$\xi\in\mathfrak{g}$以及它对应的在$g\in G$上取值的左不变矢量场X_ξ，我们将得到$X_\xi(g):=(T_e L_g)(\xi)$，我们将这个经过e($\tau=0$时有e值)的积分曲线X_ξ写为：

$$\phi_\xi:\ \tau\in\mathbb{R}\longmapsto\exp(\tau X_\xi)(e)\in G \tag{77}$$

然后，我们将$\mathfrak{g}$的指数映射定义到G成为映射：

$$\exp:\ \xi\in\mathfrak{g}\longmapsto\phi_\xi(1)\in G \tag{78}$$

使用了式(70)所定义的β的线性，式(77)和式(78)这两个方程非常简单地相关：

$$\exp(\tau\xi):=\phi_{\tau\xi}(1):=\exp(1.X_{\tau\xi})(e)=\exp(\tau X_\xi) \tag{79}$$

当上下文暗示了除了G之外的李群，我们将其写作$\exp_G$而不是exp。

映射exp是从$0\in\mathfrak{g}$的邻域到$e\in G$的邻域的局域微分同胚映射，但不是映射到G上的整体的微分同胚映射。用现代术语来说，这个结论随着将反函数定理应用于上述讨论而产生。它也表示了关于这个主题历史的有趣例子，参见对于这个结论的李版本的讨论[Hawkins，2000，pp. 82—83]，并没有明确提及它

的局域性。

(3)关于幂的同态映射

如果f: $G \to H$是李群上的平滑同态，那么对于所有的$\xi \in \mathfrak{g}$:

$$f(\exp_G \xi) = \exp_H((T_e f)(\xi)) \tag{80}$$

(4)作为变换方法的右不变矢量场

我们已经继承了依据左不变矢量场定义$\mathfrak{g}$的惯例。我们可以换成右不变矢量场。在这个过程中，我们在符号方面做了一些改变，也在某些被定义的算符是否遵守或者违反用于它们定义中的两个元素的规则方面做了改变。我将不会深入研究关于这些内容的细节。但是当我们考虑下列内容时，我们还是需要一些细节的。

(i)李群作用量，尤其是它们的无穷小算子(4.4 节和 4.5 节)；

(ii)李群余切丛上的约化——和发生在刚体理论中的一样(6.5 和 7.3.3 节)。

目前，我将只记录两条基本结论，(A)和(B)，其他的结论推迟到 4.4 节及以下的章节。

(A) 对应于 g 和左不变矢量场之间的矢量空间同构，如式(70)表示的那样，即：

$$\xi \in T_e G \longmapsto X_\xi \in \mathcal{X}_L(G)\text{，其中 } X_\xi(g) := (T_e L_g)\xi \tag{81}$$

这是同构于右不变矢量场集的矢量空间同构：

$$\xi \in T_e G \longmapsto Y_\xi \in \mathcal{X}_R(G)\text{，其中 } Y_\xi(g) := (T_e R_g)\xi \tag{82}$$

此外，右不变矢量场的李括号本身就是右不变的。因此，对应于我们先前式(73)给出的$T_e G$上李括号的定义以及它的推论$[X_\xi, X_\eta] = X_{[\xi,\eta]}$，使$T_e G \cong \mathcal{X}_L(G)$成为李代数同构，我们也可以通过下式定义$T_e G$上的李括号：

$$[\xi, \eta]_R := [Y_\xi, Y_\eta](e) \tag{83}$$

然后得到一个李代数同构$T_e G \cong \mathcal{X}_R(G)$。

(B) 但是，式(73)和式(83)的在G上的两个李括号是不同的。事实上，我们能够表明：

(i) X_ξ 和 Y_ξ 通过下式相关：

$$I_* X_\xi = -Y_\xi \tag{84}$$

其中 I: $G \to G$ 是逆映射 $I(g) := g^{-1}$，I_* 是通过 I 产生的矢量场上的前推，参见式(31)，即：

$$(I_* X_\xi)(g) := (TI \circ X_\xi \circ I^{-1})(g) \tag{85}$$

此外，因为 I 是一个微分同胚映射，式(84)使 I_* 成为一个矢量空间同构。

(ii) 从式(84)可以得出：

$$[X_\xi, X_\eta](e) = -[Y_\xi, Y_\eta](e)\text{，因而}[\xi, \eta] = -[\xi, \eta]_R \tag{86}$$

最终，关于物理的备注：在应用物理学中，G 通常是物理系统的对称群，因此，G 上的矢量场是单参数对称群的无穷小算子。关于力学，我将在第 2 节反复提到，尤其对于在物理空间$\mathbb{R}^3$ 上关于原点的平动群和转动群。这一小节 G 上的李代数$\mathfrak{g}$和左不变矢量场之间的同构意味着，我们可以认为$\mathfrak{g}$也是无穷小对称系统的分量($\xi \in \mathfrak{g}$也被称为群 G 的算子)。

3.4.3 实例，子群和子代数

我将以 3.4.1 的三个例子的前两个开始。我将多讲些理论，这能使我们在下一小节有效地处理第三个例子，即转动群。

(1)实例

(i) 加法下的 $G := \mathbb{R}^n$，G 是阿贝尔的(Abelian，即可交换的)，因此，左右平移是一致的。这个不变矢量场恰恰是常矢量，所以$\mathcal{X}_L(G) = \mathcal{X}_R(G) \cong \mathbb{R}^n$。因此在 T_eG 下的切空间，即李代数$\mathfrak{g}$自身就是$\mathbb{R}^n$。这个括号结构完全是简并的，对于所有不变矢量场 X，Y，有$[X, Y] = 0$；且对于所有的 ξ，$\eta \in \mathfrak{g}$，得出$[\xi, \eta] = 0$。

(ii) 一般线性群 $G := GL(n, \mathbb{R})$。因为 G 在 $End(\mathbb{R}^n, \mathbb{R}^n)$ 内是开区间，"*End*"表示"同构(endomorphism)"，在$\mathbb{R}^n$ 上的所有线性映射(G 的李代数)的矢量场是 $End(\mathbb{R}^n, \mathbb{R}^n)$，参见实例(i)。为了计算李括号，我首先指出，任意 $\xi \in End(\mathbb{R}^n, \mathbb{R}^n)$ 通过下式来定义 $GL(n, \mathbb{R})$ 上的对应矢量场：

$$X_\xi: A \in GL(n, \mathbb{R}) \longmapsto A\xi \in End(\mathbb{R}^n, \mathbb{R}^n) \tag{87}$$

此外，X_ξ 是左不变的，因为对于每一个 $B \in GL(n, \mathbb{R})$，左平移：

$$L_B: A \in GL(n, \mathbb{R}) \longmapsto BA \in GL(n, \mathbb{R}) \tag{88}$$

是线性的，因此：

$$X_\xi(L_B A) = BA\xi = T_A L_B X_\xi(A) \tag{89}$$

将 $I \in GL(n, \mathbb{R})$ 上的式(59)应用于式(73)中李代数的括号的定义中，我们有：

$$[\xi, \eta] := [X_\xi, X_\eta](I) = \mathbf{D}X_\eta(I) \cdot X_\xi(I) - \mathbf{D}X_\xi(I) \cdot X_\eta(I) \tag{90}$$

但 $X_\eta A = A\eta$ 在 A 上是线性的，因此 $\mathbf{D}X_\eta(I) \cdot B = B\eta$。这意味着：

$$\mathbf{D}X_\eta(I) \cdot X_\xi(I) = \xi\eta \tag{91}$$

类似地，有：

$$\mathbf{D}X_\xi(I) \cdot X_\eta(I) = \eta\xi \tag{92}$$

因此，李代数 $End(\mathbb{R}^n, \mathbb{R}^n)$ 有作为它的括号的矩阵对易子：$[\xi, \eta] = \xi\eta - \eta\xi$。这个李代数经常被写成 $\mathfrak{gl}(n, \mathbb{R})$。

让我们将 3.4.2.2 小节的结论(2)应用于这个例子。简言之，这个结论认为左不变矢量场对应于 G 的相关单参数子群。为了找到 $GL(n, \mathbb{R})$ 上的单参数子群 $\exp(\tau X_\xi)(e)$，我们将矩阵元素 $x_{ij}(i, j = 1, \cdots, n)$ 看作是 $GL(n, \mathbb{R})$ 上的 n^2 个坐标。所以，在单位矩阵上的切空间是矢量集：

$$\sum_{ij} \xi_{ij} \frac{\partial}{\partial x_{ij}} \Big|_I \tag{93}$$

其中 $\xi = (\xi_{ij})$ 是任意矩阵。对于给定的 ξ，$\exp(\tau X_\xi)e$ 通过积分下面的 n^2 个常微分方程得到：

$$\frac{dx_{ij}}{d\tau} = \sum_k \xi_{ik} x_{kj};\ x_{ij}(0) = \delta_{ij} \tag{94}$$

解恰恰就是矩阵指数：

$$X(\tau) = \exp(\tau\xi) \tag{95}$$

一般来说，让我们返回到 3.4.1 小节中矩阵李群的观点。对于矩阵李群 G，它的李代数可以被定义为：

$$\mathfrak{g} = \{\text{矩阵组 } \xi = \phi'(0):\ \phi \text{ 是微分映射：} \mathbb{R} \to G,\ \phi(0) = e_G\} \tag{96}$$

关于李代数结构的推论将继续进行。尤其是我们得到，$\xi \in \mathfrak{g}$ 产生的单参数子群通过矩阵指数给出，和式(9)的一样，即这个群是 $\{\exp(\tau\xi):\ \tau \in \mathbb{R}\}$。

这个结论将帮助我们计算出我们的第三个例子，即得到转动群的李代数。但是，它需要首先发展 3.4.2.2 小节的结论(2)，即左不变矢量场和 G 的相关

的单参数子群之间的对应。

（2）更多理论

首先，给出需要警醒之处。我们在后面将需要注意，李群 G 的子群，甚至单参数子群，不需要是 G 的子流形。让我们回想一下 3.3.1 小节浸入和嵌入的定义。因此，如果包含映射 $i: H\to G$ 是一个内射的浸入，我们就将李群 G 的子群 H 定义为 G 的李群。

正如我们在 3.3.1 小节看到的那样，不是所有内射的浸入都是嵌入，因此，也存在不是子流形的李子群的例子。

实例：圆环面 T^2 可以很自然地成为李群（习题：证明一下这个！）；圆环面 T^2 上的单参数子群就是李子群，其中，这个李子群不是流形。这个例子的更详细内容，参见［Arnold，1973，pp. 160—167；1989，pp. 72—73］或者［Butterfield，2004a，2.1.3.B］。

但结果是，闭合是更进一步的充分必要条件。即：

如果 H 是李群 G 的闭子群，那么 H 是 G 的子流形，尤其是李子群的子流形。而相反，如果 H 是李子群（这个李子群也是子流形），那么 H 就是闭合的。

来自 3.4.2.2 的结论（2），即 G 的一维子群和$\mathfrak{g}$的一维子空间之间的对应，将推广到更高维的子群和子代数。即：

如果 $H\subset G$ 是 G 的李子群，那么它的李代数$\mathfrak{h}:=\mathfrak{g}(H)$是$\mathfrak{g}\equiv\mathfrak{g}(G)$的子代数。事实上：

$$\mathfrak{h}=\{\xi\in\mathfrak{g}:\ \exp(\tau X_{\xi})(e)\in H,\ \text{对于所有的}\ \tau\in\mathbb{R}\} \tag{97}$$

而相反，如果$\mathfrak{h}$是$\mathfrak{g}$的任意 m 维子代数，那么，存在唯一相关的 G 的 m 维李子群 H，其中，G 的李代数为$\mathfrak{h}$。

前两个陈述的证明用了 3.4.2.2 小节的结论（1）。对于第三个，即主要的观点是$\mathfrak{h}$定义了 G 上的 m 个矢量场，其中的 G 是线性独立的也是对合的，因此，我们能将弗罗宾尼斯定理应用于推断积分子流形。然后我们不得不证明 H 是李子群，奥尔弗［2009，Theorem1.51］、马斯登和拉蒂乌［1999，pp. 279—280］给出了细节和参考文献。历史备忘录：这个结论有时被称为李的“第三基本定理”，李本人认为这是他的群论的主要定理，参见［Hawkins，2008，83］。

李子群和李子代数之间的基本对应引出一个问题，即是否每一个有限维李代数$\mathfrak{g}$是李群的李代数。答案是肯定的。此外，这个问题被归纳为矩阵李群的情况，即 $GL(n,\mathbb{R})$ 的李群，在这个意义下，每一个有限维李代数$\mathfrak{g}$同构于$\mathfrak{gl}(n,\mathbb{R})$的子代数。但是需要注意的是，这并不意味着(且也不是真实的)，每一个李群作为矩阵李群是可实现为矩阵李群，即每个李群都同构于 $GL(n,\mathbb{R})$的李子群。

这个基本对应在很大程度上也简化了李群的李代数的计算，例如 $H \coloneqq SO(3)$ 是 $GL(n,\mathbb{R})$的李子群。我们只需将它和上面的例子(ii)合并，$\mathfrak{gl}(n,\mathbb{R})$是 $End(\mathbb{R}^n,\mathbb{R}^n)$，其中作为它的括号的矩阵对易子是：$[\xi,\eta]=\xi\eta-\eta\xi$。

因此，我们推断 $SO(3)$的李代数——写作$\mathfrak{so}(3)$——是 $End(\mathbb{R}^n,\mathbb{R}^n)$的子代数。此外，我们能通过考虑所有被包含其中的 G 的单参子群来识别$\mathfrak{so}(3)$。结合式(95)和式(97)，我们有：

$$\mathfrak{so}(3)=\{\xi\in\mathfrak{gl}(n,\mathbb{R}):,\ \text{矩阵指数}\exp(\tau\xi)\in SO(3),\ \forall\tau\in\mathbb{R}\} \tag{98}$$

有了这样的结论，我们现在可以计算$\mathfrak{so}(3)$。

3.4.4 *转动群的李代数*

我们的第一个目的就是计算 $H \coloneqq SO(3)$的李代数$\mathfrak{so}(3)$，也可写作 $so(3)$，即转动群。我们将追溯到3.2.1.1 小节中$\mathbb{R}^3$ 中的矢量和反对称群之间的对应。

$SO(3)$通过行列式为 1 的 3×3 正交矩阵表示。因此，式(98)的必要条件写成 e 而不是 exp：

$$(e^{\tau\xi})(e^{\tau\xi})^T=I\ \text{且}\ \det(e^{\tau\xi})=1 \tag{99}$$

第一个方程对 τ 求微分，再令 $\tau=0$，就得到：

$$\xi+\xi^T=0 \tag{100}$$

因此 ξ 必须是反对称的，即通过反对称矩阵表示；相反地，对于任意的反对称矩阵 ξ，我们可以得到 $\det(e^{\tau\xi})=1$，因此，事实上：

$$\mathfrak{so}(3)=\{3\times3\ \text{反对称矩阵}\} \tag{101}$$

注意：自变量是不依赖于 $n=3$ 的选择，相似地我们可以计算对于任意整数 n 的$\mathfrak{so}(n)$：

$$\mathfrak{so}(n)=\{n\times n\ \text{反对称矩阵}\} \tag{102}$$

因此，在任意 n 维的欧氏空间$\mathbb{R}^n$ 中的转动可以通过 $n\times n$ 反对称矩阵的李代数得到。

这证明了我们在 3.2.1.1 小节中的断言，即三维的转动群通过空间$\mathbb{R}^3$ 中的矢量专门表示。正如我们刚刚看到的那样，转动的无穷小算子通常是反对称矩阵，在 n 维中将有 $n(n-1)/2$ 个独立分量；但是仅当 $n=3$ 时才等于 n。

备注：由于 $e^{\tau\xi}$的高阶项可能被忽略(参见物理观点：ξ 表示无穷小转动)，$\mathfrak{so}(3)$的非正式计算如下：

对于作为转动的$(I+\tau\xi)$，要求：

$$(I+\tau\xi)(I+\tau\xi)^T=I,\ \text{且}\ \det(I+\xi\tau)=1 \tag{103}$$

抛去高阶项，第一个方程将变成：

$$I+\tau(\xi+\xi^T)=I,\ \text{即}\ \xi+\xi^T=0 \tag{104}$$

此外，式(103)的第二个方程没有其他的约束条件了，因为任意反对称矩阵 ξ 写为如下形式，可参见式(47)：

$$\xi=\begin{pmatrix} 0 & -\xi_3 & \xi_2 \\ \xi_3 & 0 & -\xi_1 \\ -\xi_2 & \xi_1 & 0 \end{pmatrix} \tag{105}$$

我们马上可以计算出 $\det(I+\xi\tau)=1+\tau^2(\xi_1^2+\xi_2^2+\xi_3^2)$。因此，抛去高阶项，$\det(I+\xi\tau)=1$。简而言之，我们将再次得到：

$$\mathfrak{so}(3)=\{3\times3\ \text{反对称矩阵}\} \tag{106}$$

为了以后的使用(例如，4.4 和 4.5.1 小节)，我们得到如下三个矩阵：

$$A^x=\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix},\ A^y=\begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix},\ A^z=\begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{107}$$

生成空间$\mathfrak{so}(3)$，然后产生了单参数子群：

$$R_\theta^x=\begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{pmatrix},\ R_\theta^y=\begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix},\ R_\theta^z=\begin{pmatrix} \cos\theta & -\sin\theta & 0 \\ \sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{108}$$

表示围绕物理空间$\mathbb{R}^3$中的各自坐标轴的逆时针转动。

已经计算了由反对称矩阵组成的$\mathfrak{so}(3)$，我们可以使用3.2.1.1小节$\mathbb{R}^3$中的矢量和这些矩阵的对应来实现$\mathfrak{so}(3)$，因为带有李括号的矢量和矢量乘积是一样的。基于已经得到的这种认识，我们能够很容易地获得关于转动的更深入的结论，已经不需要其他的说明了。但是我们可以通过式(48)中从$\omega \in \mathbb{R}^3$到矩阵$A \in \mathfrak{so}(3)$的同构Θ得到如下所述的一个好的例子：$\exp(\tau\Theta(\omega))$是以角$\tau\|\omega\|$绕着轴ω转动。

我们现在开始领会这一章的第二个题词(来自阿诺德)的观点，即刚体的初等理论是六个概念不同的三维空间。因为我们的讨论已经区分了阿诺德列出的六个空间中的三个。换句话说，我们已经辨别出：

(i) $\mathbb{R}^3$，尤其是被看作物理空间时；来自于(ii)$\mathfrak{so}(3) \equiv T_e(SO(3))$，转动的算子；它们作为李代数是同构的，是通过式(48)从矢量$\omega \in \mathbb{R}^3$到矩阵$A \in \mathfrak{so}(3)$的双射Θ得到的；

(ii)从左平移关于g导数之下——即在$(L_g)_*$之下——的同构副本的$\mathfrak{so}(3) \equiv T_e(SO(3))$，即$T_g(SO(3))$，参见式(69)。在题词中，阿诺德对应于$\mathfrak{so}(3)$写为$g$，对应于$SO(3)$为$G$。

在5.2.4小节中，我们将理解(即使没有发展刚体理论)其余的题词。即我们将理解为什么阿诺德也提到了三种对应的对偶空间，$\mathbb{R}^{3*}$，$\mathfrak{so}(3)^*$和$T_g^*(SO(3))$。但是我们已经说过很多关于两个切空间$\mathfrak{so}(3) \equiv T_e(SO(3))$和$T_g(SO(3))$的内容，关系到对于转动刚体的观点，位形空间被看作是$SO(3)$，参见2.2小节的(3)。我们将指出，存在两种从$T_g(SO(3))$到$T_e(SO(3))$的同构，它们不仅在基独立的数学意义上是自然的，而且在物理解释的意义上也是自然的。即它们代表了来自拉格朗日广义速度的计算，即q。实际上，一个同构可以计算出关于固定在空间上的正则正交框架的角速度分量(称作空间坐标)；另一个计算出关于固定在刚体的正则正交框架的角速度分量(称作体坐标)。事实上，这些同构分别是右平移和左平移的导数，可参见式(67)和式(68)。

因此，假设一个转动刚体有一个右旋的正则正交框架$\{a, b, c\}$。我们可以认为这三个单位矢量a，b，c就是$\mathbb{R}^3$的列矢量。将它们设置在一个3×3的矩阵$g := (a\ b\ c) \in GL(3, \mathbb{R})$中，我们将得到一个从$e_1$映射到$a$的单位$x$矢量，

e_2映射到 b 的单位 y 矢量等的矩阵。即 g 映射正则框架 e_1，e_2，e_3 到 a，b，c 且 g 是一个正交矩阵：$g\in SO(3)=\{g\in GL(3,\mathbb{R})\mid \tilde{g}g=I\}$。因此 g 代表刚体的位形，且位形空间是 $SO(3)$。

通过微分条件 $\tilde{g}g=I$，我们得出，在特定的 g，$T_g(SO(3))$中的切空间(即速度空间 $\dot{g}$)是 $GL(3,\mathbb{R})$的三维矢量子空间：

$$T_g(SO(3))=\{\dot{g}\in GL(3,\mathbb{R})\mid \dot{\tilde{g}}g+\tilde{g}\dot{g}=0\} \tag{109}$$

现在回想3.2.1.1小节的例子(ii)和(iii)。我们说，尽管刚体的角速度通常被认为是矢量 ω，因此对“体矢量”a，b，c，有：

$$\dot{a}=\omega\wedge a,\ \dot{b}=\omega\wedge b,\ \dot{c}=\omega\wedge c \tag{110}$$

我们反而能通过反对称矩阵 $A:=\Theta(\omega)\in\mathfrak{g}\equiv T_e(SO(3))$的角速度来编码。正如我们看到的，式(110)变成了：

$$\dot{a}=\Theta(\omega)a,\ \dot{b}=\Theta(\omega)b,\ \dot{c}=\Theta(\omega)c \tag{111}$$

或者相当于关于位形 $g=(a\ b\ c)$的矩阵方程：

$$\dot{g}\equiv(\dot{a}\ \dot{b}\ \dot{c})=\Theta(\omega)g;\ \text{即}\ \Theta(\omega)=\dot{g}g^{-1} \tag{112}$$

然后我们看到从 $T_g(SO(3))$到$\mathfrak{g}=T_e(SO(3))$的映射：

$$\dot{g}\in T_g(SO(3))\longmapsto \dot{g}g^{-1}\equiv\dot{g}\tilde{g}\in\mathfrak{g} \tag{113}$$

映射广义速度 $\dot{g}$ 到角速度 $\Theta(\omega)$上。这是一个以基本方式表示的关于固定在空间的坐标角速度。我们可以立刻得出，它是一个同构(练习之!)。

另一方面，我们考虑到 $\Theta(\omega)$是一个线性变换 $\Theta(\omega):\mathbb{R}^3\to\mathbb{R}^3$，并且我们在体坐标 a，b，c 中表示它。我们可以得出：$g^{-1}\Theta(\omega)g\equiv g^{-1}\dot{g}$。因此这个映射：

$$\dot{g}\in T_g(SO(3))\longmapsto g^{-1}\dot{g}\equiv\tilde{g}\dot{g}\in\mathfrak{g} \tag{114}$$

将广义速度 $\dot{g}$ 映射到在体坐标内表示的角速度。它也是一个同构。

总结：我们有两种自然同构可以从广义速度 $\dot{g}$ 计算角速度，尤其是在空间坐标和体坐标中。

顺便说一句，我们可以直接证实式(113)和式(114)的同构的像 $\dot{g}\tilde{g}$ 和 $\tilde{g}\dot{g}$ 位于$\mathfrak{g}$内，即它们是反对称矩阵。因此用·表示点积，得出：

$$g^{-1}\dot{g}\equiv\tilde{g}\dot{g}=\begin{pmatrix}\tilde{a}\\ \tilde{b}\\ \tilde{c}\end{pmatrix}(\dot{a}\ \dot{b}\ \dot{c})=\begin{pmatrix}0 & a\cdot\dot{b} & a\cdot\dot{c}\\ b\cdot\dot{a} & 0 & b\cdot\dot{c}\\ c\cdot\dot{a} & c\cdot\dot{b} & 0\end{pmatrix}\tag{115}$$

这是一个反对称矩阵，因为对 $a\cdot b=b\cdot c=a\cdot c=0$ 关于时间求微分，我们可以得出 $a\cdot\dot{b}+\dot{a}\cdot b=0$ 等结果。最终，我们从 $\dot{g}\tilde{g}=g(g^{-1}\dot{g})g^{-1}$ 的事实得出 $\dot{g}\tilde{g}$ 是反对称的，且反对称性通过 g 的共轭保持。

我们通过两个附带的备注来结束这个小节（它们不会被运用到以后的内容）。

(1)在 2.1.1 小节，我们本可以详细说明从辛流形到辛矢量空间的讨论，即一个具有非简并的反对称双线性形式 ω: $Z\times Z\to\mathbb{R}$的（实数的，有限维）矢量空间。由此得出的结论是，Z 是偶数维的。问题是线性映射 A: $Z\to Z$ 通过式(4)保持正则形式 ω。很明显地，这和保持哈密顿方程形式的 A 是等价的，因此映射 A 被称作辛的（或正则的，或泊松的）。所有此类映射的集合形成了李群，即辛群，写作 $\mathrm{Sp}(Z,\ \omega)$。但是因为这一章不需要正则变换理论，我将 $\mathrm{Sp}(Z,\ \omega)$ 的结构的研究作为练习！更多细节参见亚伯拉罕和马斯登[1978, pp. 167—174]，马斯登和拉蒂乌[1999, pp. 69—72, 293—299]。

(2)最后，看一看这一章已经放弃的无限维流形。考虑到所有 M 上的微分同胚映射的无限维李群为 $Diff(M)$。李代数的元素，即矢量 $A\in T_e(Diff(M))$ 是 M 上的一个矢量场或者相当于 M 上的流。此外，在式(73)中的这个李代数中的李括号 $T_e(Diff(M))$ 是 3.2.2 小节中 M 上的矢量场的李括号。

4. 李群的作用量

我们来研究李群在流形上的作用量。本小节中的一些概念（结论和实例）对于 5.4 节及其往后的内容都是至关重要的。幸运的是，前文中给出了关于我们所需的概念和结论的若干实例。4.1 节将给出包含决定性的余切提升的概念的基本材料。4.2 和 4.3 节描述了轨道和商空间成为流形的条件。4.4 节极小地描述了作用量，即关于它们的无穷小算子。4.5 节给出了分别基于其李代数和对

偶$\mathfrak{g}^*$的两个重要李群表示，即其伴随的和余伴随的表示。最后，4.6 节收集了若干思路来考虑我们核心的并反复用到的例子，即转动群。

4.1 基本定义和实例

流形 M 之上的李群 G 的左作用量是一个平滑映射 Φ：$G \times M \to M$，其满足：

(i) 对于所有的 $x \in M$，都有 $\Phi(e, x) = x$；

(ii)对于所有的 g，$h \in G$ 和所有的 $x \in M$，都有 $\Phi(g, \Phi(h, x)) = \Phi(gh, x)$。

我们通常把 $g \cdot x$ 写作 $\Phi(g, x)$。

同样地，在流形 M 之上的李群 G 的右作用量是一个平滑映射 Ψ：$M \times G \to M$，其满足：(i)$\Psi(x, e) = x$；(ii)$\Psi(\Psi(x, g), h) = \Psi(x, gh)$。我们通常把 $x \cdot g$ 写作 $\Psi(x, g)$。

我们同样很容易就得出一个下标符号。对于所有 $g \in G$，我们定义：

$$\Phi_g: M \to M: x \longmapsto \Phi(g, x) \tag{116}$$

按照这个符号，(i) 变成了 $\Phi_e = id_M$，且 (ii) 变成了 $\Phi_{gh} = \Phi_g \circ \Phi_h$。对于右作用量而言，(ii) 变成了 $\Psi_{gh} = \Psi_h \circ \Psi_g$。

我们立即证实了，流形 M 之上的任意关于 G 的左作用量 Φ，$g \longmapsto \Phi_g$：$M \to M$，根据下式定义了右作用量 Ψ：

$$g \longmapsto \Psi_g := \Phi_{g^{-1}}: M \to M;\ \text{即：}\Psi: (x, g) \in M \times G \longmapsto \Phi(g^{-1}, x) \in M \tag{117}$$

使用这样一个事实，即在 G 中，$(gh)^{-1} = h^{-1}g^{-1}$。同样地，通过在 G 中进行倒转，右作用量也能定义左作用量。我们将偶尔使用这种左—右“替换”。

对左作用量进行定义就等于认为映射 $g \longmapsto \Phi_g$ 就是进入 Diff(M)的 G 的同态，即 M 的微分同胚映射群。M 是巴拿赫(Banach)空间 V，且所有 Φ_g：$V \to V$ 是一种连续线性变换的特例，在 V 之上的 G 的作用量被称为 V 上的 G 的表示。$x \in M$ 的轨道(在作用量 Φ 之下)就是集合：

$$\mathrm{Orb}(x) = \{\Phi_g(x): g \in G\} \subset M \tag{118}$$

如果只有一个轨道，即对于所有 x，$y \in M$，存在一个 $g \in G$ 满足 $g \cdot x = y$，

那么这个作用量就被称为可传递的。如果 $\Phi_g = \mathrm{id}_M$ 意味着 $g=e$，即如果 $g\longmapsto\Phi_g$ 是一一对应的，那么这个作用量被称为有效的(或者可靠的)。如果这个作用量对于所有 $g\neq e$ 都没有固定点，即 $\Phi_g(x)=x$ 意味着 $g=e$，那么这个作用量被称为自由的。换言之，如果对于所有 $x\in M$，$g\longmapsto\Phi_g(x)$ 都是一一对应的，那么这个作用量是自由的(因此，所有自由的作用量都是可靠的)。

4.1.0.1 实例：余切提升

我们从几何学的实例着手研究，然后赋予其某种一般理论，我们就回到了带有几个实例的力学的研究。

(1) 几何学的实例

(i)根据$(A, x)\longmapsto Ax$，$SO(3)$对$\mathbb{R}^3$ 起作用。这个作用量是可靠的，但是它既不是自由的(所有转动都固定在其轴线上的点)，也不是可传递的(轨迹是以起点为中心的球面)。

(ii) 根据$(A, x)\longmapsto Ax$，$GL(n, \mathbb{R})$对$\mathbb{R}^n$ 起作用。这个作用量是可靠的，不是自由的，并且“几乎是可传递的”，零的子空间$\{0\}$是一个轨迹，因此是$\mathbb{R}^n-\{0\}$。

(iii) 假设X是M之上的矢量场，对所有$\tau\in\mathbb{R}$，我们定义式(35)中的流$\Phi_X(\tau)$，这个意义上X是完备的。于是，这个流在M上定义了$\mathbb{R}$的作用量。

我们来研究关于辛约化讨论中主要的，并且反复用到的两个实例。

(iv)参见式(67)，每个$g\in G$的左平移L_g：$h\in G\longmapsto gh\in G$定义了$G$的左作用量。因为$G$是一个群，它是可传递的并且是自由的(并因此是可靠的)。同样地，右平移($g\longmapsto R_g$ 具有 R_g：$h\in G\longmapsto hg\in G$)定义了一种右作用量。并且$g\longmapsto R_{g^{-1}}$定义了一种左作用量，参照式(117)。我们能够很轻易地证明左平移提升对于切丛TG而言就如同左作用量一样。即根据链式法则，我们确证了：

$$\Phi_g:\ TG\to TG:\ v\equiv v_h\in T_hG\longmapsto(T_hL_g)(v)\in T_{gh}G \tag{119}$$

在TG之上定义了左作用量。同样地，右平移提升对于切丛而言就如同右作用量作用在TG上一样。然而，我们对于哈密顿力学的兴趣必然使得我们对余切提升更加感兴趣。参见下面的(2)的一般定义，以及下面(3)中的左平移的余切提升的实例(viii)。

(v) G通过共轭(内自同构)对自身起作用：$g\longmapsto K_g:=R_{g^{-1}}\circ L_g$，即 K_g：

$h \in G \longmapsto ghg^{-1} \in G$。每个 K_g 都是 G 的同构，轨道都是共轭类的。4.5 节将根据共轭（即伴随和余伴随的作用量）介绍两个作用量的“微分版本”，其在辛约化中是十分重要的。

（2）哈密顿对称性和余切提升

我们来研究哈密顿力学。根据 2.1.3 节中的讨论，我们说：已知具有辛流形(M, ω)和对称的哈密顿群 H：$M \to \mathbb{R}$的哈密顿系统(M, ω, H)是一个在 M 之上起作用的李群 G，这使得所有 Φ_g：$M \to M$ 都维持 ω 和 H。于是，最简单的可能的例子就是空间平动和（或）对自由粒子起作用的转动。如果我们首先制定了某种一般理论，那么这些例子的细节如下面的（vi）和（vii），将会更加清晰。

这个理论将阐明作用量的左—右对比之间的相互作用，以及丛的正切和余切之间的相互作用。除此之外，一般理论和实例的细节都直接继续成为在 2.3.2 小节讨论的牛顿引力感兴趣的 N 粒子情况，即定义在单个粒子上的作用量只对每个 N 粒子是重复的。

于是，我们将采用 $M := (\mathbb{R}^3) \times (\mathbb{R}^3)^*$，$\omega := dq^i \wedge dp^i$，$H := p^2/2m$。首先，平动（根据 $\mathbf{x} \in \mathbb{R}^3$）和转动（根据 $A \in SO(3)$）都作用于位形空间 $Q = \mathbb{R}^3$。根据下式，我们将得到$\mathbb{R}^3$ 和 $SO(3)$在$\mathbb{R}^3$ 之上的作用量：

$$\Phi_{\mathbf{x}}(\mathbf{q}) = \mathbf{q} + \mathbf{x}; \quad \Phi_A(\mathbf{q}) = A\mathbf{q} \tag{120}$$

然而，这些作用量提升到余切丛 $T^*Q = (\mathbb{R}^3) \times (\mathbb{R}^3)^* \cong \mathbb{R}^6$（如 2.3.2 小节中提到的）。我们用没有用到作用量这一概念的结论定义了这些作用量的提升。也就是：

任意微分同胚映射f：$Q_1 \to Q_2$ 归纳出了辛余切提升 T^*f：$T^*Q_2 \to T^*Q_1$（即在相反方向），即把 T^*Q_2 上的正则 1 形式（以及辛形式）映射到 T^*Q_1 上的 1 形式（以及辛形式）。

要定义作用量的提升，我们就要详细描述 T^*f 的定义。但是，我不会去证明刚才陈述的结论；欲知详情，请参见［Marsden and Ratiu，1999，Section 6.3］。

这个想法是这样的：T^*f 是正切映射 Tf：$TQ_1 \to TQ_2$ 的“逐点伴随矩阵（pointwise adjoint）”式（29）。换言之，我们依据其值的缩写定义 T^*f，因为对

于任意 $\alpha \in T^*_{q_2}Q_2$，有任意正切矢量 $v \in T_{f-1(q_2)}Q_1$。这里，按照很多表述把 T^*Q_2 中的一个点——即严格地说一个对组(q_2, α)，$q_2 \in Q_2$，$\alpha \in T^*_{q_2}Q_2$——与其形式 α 合并起来，似乎是没有坏处的。同样地，把 TQ_1 中的一个点(q_1, v)与其矢量$v \in T_{q_1}Q_1$ 合并起来，似乎也是没有坏处的。

我们回想一下，任意有限维的矢量空间都自然地(即与依据无关地)与其双对偶(double dual)$(V^*)^* \cong V$是同构的；并且，我们将使用一些尖角括号 $<;>$表示 V 和 V^*之间自然成对(natural pairing)。于是，我们通过要求下式，定义了T^*f: $T^*Q_2 \to T^*Q_1$:

$$<(T^*f)(\alpha);\ v> := <\alpha;\ (Tf)(v)>,\ \forall \alpha \in T^*_{q_2}Q_2,\ v \in T_{f^{-1}(q_2)}Q_1 \tag{121}$$

注意：因为 T^*f“去往相反的方向”，具有复合函数的提升的复合就包含了这个顺序的倒数。即如果 $Q_1 = Q_2 \equiv Q$，并且f，g 是 Q 的两个微分同胚映射，那么：

$$T^*(f \circ g) = T^*g \circ T^*f \tag{122}$$

鉴于这个 T^*f的定义，对于每个 $g \in G$，在流形 Q 之上的 G 的左作用量 Φ 归纳出 Φ_g：$Q \to Q$ 的余切提升。也就是说，我们有映射：

$$T^*\Phi_g \equiv T^*(\Phi_g): T^*Q \to T^*Q,\ \text{其中}\ \alpha \in T^*_qQ \longmapsto (T^*\Phi_g)(\alpha) \in T^*_{g^{-1}\cdot q}Q \tag{123}$$

现在，考虑分配到每个 $g \in G$ 的映射，$T^*\Phi_g$：

$$g \in G \longmapsto T^*\Phi_g: T^*Q \to T^*Q \tag{124}$$

为了验证这确实是 T^*Q 之上的 G 的作用量，我们首先要核对，因为 $\Phi_e = id_Q$，$T\Phi_e$: $TQ \to TQ$ 是 id_{TQ}，且 $T^*(\Phi_e)$是 id_{T^*Q}。然而注意，式(122)推出：

$$T^*\Phi_{gh} = T^*(\Phi_g \circ \Phi_h) = T^*\Phi_h \circ T^*\Phi_g \tag{125}$$

于是，式(124)定义了右作用量。

但是，这里我们回想起使用任意左作用量的逆就定义了右作用量，参照式(117)。把这个与 Q 之上的关于作用量的余切提升概念相结合，我们得到：

Q 之上的左作用量 Φ 不仅定义了 T^*Q 之上的右作用量，见式(124)，而且定义了 T^*Q 之上的左作用量，也就是通过：

$$g \in G \longmapsto \Psi_g := T^*(\Phi_{g^{-1}}): T^*Q \to T^*Q \tag{126}$$

因为$(gh)^{-1} = h^{-1}g^{-1}$，有：

$$\Psi_{gh} \equiv T^*(\Phi_{(gh)^{-1}}) = T^*(\Phi_{h^{-1}g^{-1}}) = T^*(\Phi_{h^{-1}} \circ \Phi_{g^{-1}}) = T^*\Phi_{g^{-1}} \circ T^*\Phi_{h^{-1}} \equiv \Psi_g \circ \Psi_h \tag{127}$$

简而言之，这两个方向的逆相互抵消了。在下文中的某些重要情境下会出现这种左—右替换，尤其是在6.5节和7节中。

(3)力学实例

通过大量使用广义坐标的方式，现在我们把它们用于自由粒子的平动和转动，用于转动刚体的转动，以及用于N个点粒子。

(vi)依据下式，令平动群$G=(\mathbb{R}^3, +)$通过下式作用于自由粒子的位形空间$Q=\mathbb{R}^3$：

$$\Phi_{\mathbf{x}}(\mathbf{q}) = \mathbf{q}+\mathbf{x} \tag{128}$$

因为G是阿贝尔的，于是G的左和右作用量之间的区域便失效了。并且，如果我们认为G和Q是一样的，这就是基于其自身的经由$\mathbb{R}^3$的左=右平移，即实例(iv)，因此是可传递的和自由的。然而可想而知，我们分别在式(124)和式(126)中用“g”和“g^{-1}”定义的提升作用量(the lifted actions)仍然是不同的作用量。于是，写出$\alpha=(\mathbf{q}, \mathbf{p})\in T^*_{\mathbf{q}}Q$，并使用这样的事实：$T\Phi_{\mathbf{x}}(\mathbf{q}-\mathbf{x}, \dot{\mathbf{q}}) = (\mathbf{q}, \dot{\mathbf{q}})$，我们注意到，式(121)意味着如下内容，首先：

$$T^*(\Phi_{\mathbf{x}})(\mathbf{q}, \mathbf{p}) \in T^*_{\mathbf{q}-\mathbf{x}}Q \tag{129}$$

第二，对于所有$\dot{\mathbf{q}}\in T_{\mathbf{q}-\mathbf{x}}Q$：

$$< T^*(\Phi_{\mathbf{x}})(\mathbf{q}, \mathbf{p}); (\mathbf{q}-\mathbf{x}, \dot{\mathbf{q}}) > = < (\mathbf{q}, \mathbf{p}); (\mathbf{q}, \dot{\mathbf{q}}) > \equiv \mathbf{p}(\dot{\mathbf{q}}) \tag{130}$$

如果式(130)对于所有$\dot{\mathbf{q}}\in T_{\mathbf{q}-\mathbf{x}}Q$都成立，需要$T^*(\Phi_{\mathbf{x}})(\mathbf{q}, \mathbf{p})$不会影响$\mathbf{p}$，即：

$$T^*(\Phi_{\mathbf{x}})(\mathbf{q}, \mathbf{p}) = (\mathbf{q}-\mathbf{x}, \mathbf{p}) \tag{131}$$

于是，这就是用“g”定义的与式(124)相一致的提升作用量。相同地，用“g^{-1}”定义的与式(126)相一致的提升作用量是：$\Psi_{\mathbf{x}}(\mathbf{q}, \mathbf{p}) := T^*(\Phi_{-\mathbf{x}})(\mathbf{q}, \mathbf{p}) = (\mathbf{q}+\mathbf{x}, \mathbf{p})$。

我们已经发现，这些提升作用量保证了$\omega = dq^i \wedge dp^i$(操控外导数的练习)和$H := p^2/2m$。于是，我们有一个哈密顿对称群。这个作用量是不可传递的，它们用$\mathbf{p}\in(\mathbb{R}^3)^*$的值标记了其轨道，但该作用量是自由的。

（vii）依据下式，令 $SO(3)$ 作用在 $Q=\mathbb{R}^3$ 的左边：

$$\Phi_A(\mathbf{q}) = A\mathbf{q} \tag{132}$$

这又是实例（i），我们用“g”——即式（124）——来提升这个作用量，以便得到 T^*Q 之上的右作用量。

正如在实例（vi）中，我们写出 $\alpha=(\mathbf{q},\ \mathbf{p})\in T^*_{\mathbf{q}}Q$。使用这个等式 $T\Phi_A(\mathbf{q},\ \dot{\mathbf{q}})=(A\mathbf{q},\ A\dot{\mathbf{q}})$，于是式（121）意味着下列内容，首先：

$$T^*(\Phi_A)(\mathbf{q},\ \mathbf{p})\in T^*_{A^{-1}\mathbf{q}}Q \tag{133}$$

第二，对于所有 $\dot{\mathbf{q}}\in T_{A^{-1}\mathbf{q}}Q$：

$$<T^*(\Phi_A)(\mathbf{q},\ \mathbf{p});\ (A^{-1}\mathbf{q},\ \dot{\mathbf{q}})> = <(\mathbf{q},\ \mathbf{p});\ (\mathbf{q},\ A\dot{\mathbf{q}})> \equiv \mathbf{p}(A\dot{\mathbf{q}}) \equiv p_iA^i_j\dot{q}^j \tag{134}$$

如果对于所有 $\dot{\mathbf{q}}\in T_{A^{-1}\mathbf{q}}Q$，式（134）都成立，需要：

$$T^*(\Phi_A)(\mathbf{q},\ \mathbf{p})=(A^{-1}\mathbf{q},\ \mathbf{p}A) \tag{135}$$

其中 $\mathbf{p}A$ 是一个行矢量。或者如果我们把 $\mathbf{p}$ 分量看成是列矢量，它要求：

$$T^*(\Phi_A)(\mathbf{q},\ \mathbf{p})=(A^{-1}\mathbf{q},\ \tilde{A}\mathbf{p})=(A^{-1}\mathbf{q},\ A^{-1}\mathbf{p}) \tag{136}$$

其中 ~ 表示矩阵的转置，并且最后的等式成立，因为 A 是一个正交矩阵。

于是，这是用“g”定义的提升作用量。同样地，用“g^{-1}”定义的提升作用量是：$\Psi_A(\mathbf{q},\ \mathbf{p}) := T^*(\Phi_{A^{-1}})(\mathbf{q},\ \mathbf{p})=(A\mathbf{q},\ A\mathbf{p})$。

此外，我们很轻易就能检查出，这些提升作用量不仅维持了 $\omega=dq^i\wedge dp^i$，（操作外导数的另一个练习！）而且维持了 $H := p^2/2m$。于是 $SO(3)$ 是一个哈密顿对称组。

同 Q 上的 $SO(3)$ 的原始作用量一样，这些作用量是可信的。但是这些作用量是不可传递的，以 $\mathbb{R}^3$ 和 $(\mathbb{R}^3)^*$ 的原点为中心点的两个球体的半径来标识轨道，并且这些作用量是不自由的，假设 $\mathbf{q}$ 和 $\mathbf{p}$ 是平行的并且都在 A 的转动轴上。

（viii）现在我们来考虑转动刚体。但是不同于实例（vi）和（vii），我们将只考虑运动学，而不是动力学，即便是对于自由体。也就是说，我们将不谈论 ω 和 H 的定义和不变量。就它们的细节而言，参见[Abraham and Marsden, 1978, Sections 4.4 and 4.6]以及 2.2 节的（3）中所给出的其他参考。在 6.5.3 节和第 7

节中，我们无论如何都会用更加一般的术语(用动量映射)来思考这个例子的动力学。

我们回想起3.4.4小节末尾的讨论，转动刚体的位形空间是$SO(3) =: G$。我们同样发现了，角速度$v = \dot{g} \in T_g G$的空间和形体表示是由右平移和左平移给出的。因而，式(113)和式(114)给出：

$$v^S \equiv \dot{g}^S := T_g R_{g^{-1}}(\dot{g}),\ 且\ v^B \equiv \dot{g}^B := T_g L_{g^{-1}}(\dot{g}) \tag{137}$$

然而，我们现在关注于左(右)平动的余切提升。于是令$SO(3)$通过左平移$\Phi_g h \equiv L_g h = gh$对其自身起作用。我们用"$g$"，即式(124)提升这个作用量，从而得到$T^* G$之上的右作用量。于是，令$\alpha \in T_h^* G$且$(TL_g)(h, \dot{h}) = (gh, g\dot{h})$。于是，式(121)意味着如下内容，首先：

$$(T^* L_g)(\alpha) \in T^*_{g^{-1}h} G \tag{138}$$

第二，对于所有$v \in T_{g^{-1}h} G$，都有：

$$< T^* (L_g)(\alpha);\ v > = < \alpha;\ gv > \tag{139}$$

换言之，就式(131)和式(135)而言，依此类推，如果要式(139)对所有$v \in T_{g^{-1}h} G$都成立，需要$gv \in T_h G$：

$$T^* (L_g)(\alpha):\ v \in T_{g^{-1}h} G \longmapsto \alpha(gv) \tag{140}$$

同样地，与式(126)相一致，用"g^{-1}"提升作用量，即$T^* G$之上的左作用量是：

$$< T^* (L_{g^{-1}})(\alpha);\ v > = < \alpha;\ g^{-1} v >,\ \forall \alpha \in T_h^* G,\ v \in T_{gh} G \tag{141}$$

在较多地研究了李群作用量的理论之后，我们将在4.6小节中继续研究这个实例。

最后，我们来描绘出另一个力学实例的梗概，我们在2.3.2小节中讨论了通过牛顿引力N粒子和位形空间$Q := \mathbb{R}^{3N}$相互作用。这些将把实例(vi)和(vii)结合起来，并把它们推广，因而，就进入到下个小节中关于轨道和商的讨论。

(ix) 如上所述，即在式(120)之前，在单个粒子之上的平动和转动余切提升作用量直接转移到了N粒子的情况，即在单个粒子之上定义的作用量只是重复的，每个分量都是如此，对于N粒子的每个粒子给出$T^* Q \cong \mathbb{R}^{3N} \times (\mathbb{R}^{3N})^*$之上的作用量。

此外，平动和转动的群是单群(即欧氏群)的子群。我没有精确地定义 E。这里，令它足以说明：

(a) E 在位形空间 $Q := \mathbb{R}^{3N}$ 之上的按分量逐个进行的作用量具有余切提升，这个余切提升对每个分量都是如此。

(b) E 的余切提升作用量既是不可传递的，也是不自由的，但却是可靠的。

(c) 如果我们把式(25)的 H 看作是哈密顿函数，根据牛顿引力把粒子描述成相互作用的，那么 E 就是哈密顿对称群。事实上，动能和势能分别是不变量，本质上是由于粒子的相互作用仅仅取决于粒子内部的距离，而不是取决于它们的位置或者方向，参见 2.3.2 节中的讨论。

关于实例(ix) 的最终评论，详见下面的内容。

回想在 2.3.3 节和 2.3.4 节中，我们把这个实例看作是讨论关系论和还原论步骤的跳板，其求得了位形空间和相位空间的商。然而，为了商空间(和轨道)能成为流形，尤其是为了能够用简单的方法增加和减少维度，我们需要在求商之前，切掉两类"特殊"点。也就是对称位形或态的类别(即通过 E 的某种元素设定的位形或者态)，以及碰撞位形或者态的类别。对于约化论提倡的相位空间的商而言，态的类就是 $\delta \subset T^* \mathbb{R}^{3N}$ 和 $\Delta \subset T^* \mathbb{R}^{3N}$ (参见 2.3.4 小节的定义)。

通过分析实例(vi)到(ix)，我们现在能够发现：

(a) δ 和 Δ 各自在 $T^* \mathbb{R}^{3N}$ 之下的 E 的余切提升作用量之下是闭合的，即各自为轨道的并集。于是 E 作用于 $M := T^* \mathbb{R}^{3N} - (\delta \cup \Delta)$ 上。

(b) E 自由地作用于 M。我们后来发现(尤其是在 4.3B 和 5.5 节中)，自由作用量是使轨道和商空间成为流形的重要充分条件的一半(共轭的一个)。共轭的另一半将是正确的作用量，我们将在 4.3 小节中对其进行定义。

4.2 来自于群作用量的商结构

在有限的维度中，任意轨道 Orb(x)是关于 M 的浸入子流形(immersed submanifold)。我们能够直接证明这点，参见亚伯拉罕和马斯登[1978, Ex. 1.6F (b), p. 51, and 4.1.22, p. 265]。但是，出于我们的目的，我们最好把这看成是某些条件的推论，在这些条件下商结构是流形，如下所示。

如果有 $g \in G$ 满足 $g \cdot x = y$，那么 $x \cong y$，这个关系是等价关系，并将轨道作为等价类。我们用 M/G(有时称作是轨道空间)表示商空间，即轨道的集合。我们把正则映射看作是：

$$\pi: M \to M/G, \ x \longmapsto \mathrm{Orb}(x) \tag{142}$$

并且，我们把 $U \subset M/G$ 定义为开放的当且仅当 $\pi^{-1}(U)$ 在 M 中是开放的，我们赋予了 M/G 以商拓扑。

简单的实例如 4.1.0.1 中的例(ii)，说明了这个商拓扑不需要是豪斯道夫(Hausdorff)拓扑。然而，我们很容易就能说明，如果集合

$$R := \{(x, \Phi_g x) \in M \times M: (g, x) \in G \times M\} \tag{143}$$

是 $M \times M$ 的一个封闭的子集，那么 M/G 之上的商拓扑就是豪斯道夫拓扑。

但是，为了保证 M/G 具有流形结构，我们需要进一步的条件。主要的条件(以及更加困难的定理)就是：

当且仅当 M/G 是具有 $\pi: M \to M/G$ 浸没的流形，R 是 $M \times M$ 的闭合的子流形。

这个定理的两个推论对我们而言是重要的。

(1) 对于 $\pi: M \to M/G$ 是一个浸没，从流形 M/G 到流形 N 的映射 $h: M/G \to N$ 是平滑的，当且仅当 $h \circ \pi: M \to N$ 是平滑的。

这个推论具有有用的含义，我们把它叫作通往商的途径，即关于等变性(equivariance)的概念——其在辛约化中是重要的。

如果平滑的映射 $f: M \to N$ 关系到李群 G 在流形之上的作用量，那么我们把它称为等变化的。也就是说，分别通过 $\Phi_g: M \to M$ 和 $\Psi_g: N \to N$，令 G 对 M 和 N 起作用。如果对于所有 $g \in G$，下式成立，那么对于这些作用量而言，$f: M \to N$ 被称为等变化的。

$$f \circ \Phi_g = \Psi_g \circ f \tag{144}$$

也就是说，f 是等变化的，当且仅当对于所有 g，下述图表互换：

$$\begin{tikzcd} M \arrow[r, "f"] & N \\ M \arrow[u, "\Phi_g"] \arrow[r, "f"] & N \arrow[u, "\Psi_g"'] \end{tikzcd} \tag{145}$$

等变性直接意味着，在商之上，f 自然引起一个映射，例如$\hat{f}$。也就是说，这个映射：

$$\hat{f}: \mathrm{Orb}(x) \in M/G \longmapsto \mathrm{Orb}(f(x)) \in N/G \tag{146}$$

是定义明确的，即不依赖于为轨道选择的表示 x。

通过应用这个推论，我们知道，如果f: $M \to N$ 是等变化的，商 M/G 和 M/N是具有正则投影的流形，那么f是平滑的就意味着$\hat{f}$是平滑的。这叫作通往商的途径。

(2)令 H 为李群 G 的封闭子群。依据 3.4.3 节的(2)，这相当于 H 为 G 的子流形的子群。令 H 依据左平移对 G 起作用：$(h, g) \in H \times G \longmapsto hg \in G$，那么，轨道就是右陪集(coset)$Hg$。于是，$G/H$ 是流形并且 π：$G \to G/H$ 是一个浸没。

4.3 正确的作用量

通过对 4.2 小节主要定理的补充(即当且仅当 M/G 是具有 π：$M \to M/G$ 浸没的流形，R 是 $M \times M$ 的闭合子流形)，我们可以为正确的作用量给出充分条件：

(A)轨道成为子流形；

(B)M/G 成为流形。

作用量 Φ：$G \times M \to M$ 被称为适当的，如果映射

$$\tilde{\Phi}: (g, x) \in G \times M \longmapsto (x, \Phi(g, x)) \in M \times M \tag{147}$$

是适当的。鉴于此，我们说如果$\{x_n\}$是 M 中的收敛序列，并且$\{\Phi_{g_n}(x_n)\}$是 M 中的收敛序列，那么$\{g_n\}$在 G 中就有一个收敛序列。在有限的维度中，这意味着紧致集(compact set)使逆像(inverse image)紧致，即如果 $K \subset M \times M$ 是紧致的，那么 $\tilde{\Phi}^{-1}(K)$是紧致的。

如果 G 是紧致的，这个条件就自动满足了。此外，通过左(或者右)平动作用于自身的群作用量，如 4.1.0.1 节实例(iv)就永远是正确的。此外，左(或者右)平动的余切提升，如 4.1.0.1 节的(2)和实例 (viii)，总是正确的。我们可以不证明这些，但这在接下来的内容中将是很重要的。

4.3.0.1 各向同性群，作为流形的轨道

对于 $x \in M$，Φ 在 x 的各向同性（或稳定或对称）群是：

$$G_x := \{g \in G: \Phi_g(x) \equiv \Phi(g, x) = x\} \subset G \tag{148}$$

于是，当且仅当对于所有 $x \in M$，$G_x = \{e\}$，这一个作用量是自由的。

因而，如果我们定义：

$$\Phi^x: G \to M: \Phi^x(g) := \Phi(g, x) \tag{149}$$

我们有：$G_x = (\Phi^x)^{-1}(x)$。符号 Φ^x 是我们在式(116)中定义的符号 Φ_g 的“近亲”。

因为 Φ^x 是连续的，G_x 是 G 的封闭子群。于是根据 3.4.3 节中(2)的结论即在式(97)之前的结论，G_x 是 G 的子流形(以及李子群)。并且如果这个作用量是正确的，那么 G_x 是紧致的。

再者，对于所有 $h \in G_x$，我们有 $\Phi^x(gh) = \Phi_g \circ \Phi_h(x) = \Phi_g(x)$，这意味着 Φ^x 自然地归纳出映射：

$$\tilde{\Phi}^x: [g] = gG_x \in G/G_x \longmapsto \Phi_g x \in \mathrm{Orb}(x) \subset M \tag{150}$$

也就是说，这个映射是定义明确的。$\tilde{\Phi}^x$ 是内射的，因为如果 $\Phi_g x = \Phi_h x$，那么 $g^{-1}h \in G_x$，于是 $gG_x = hG_x$。

我们从 4.2 小节的主要定理(即当且仅当 M/G 是具有浸没 $\pi: M \to M/G$ 的流形，那么 R 是 $M \times M$ 的封闭子流形)推断出以下内容。

(a)如果 $\Phi: G \times M \to M$ 是一个作用量并且 $x \in M$，于是式(150)定义的 $\tilde{\Phi}^x$ 是一个单射浸入。

这里我们回想起 3.3.1 节中的结论：单射浸入不需要嵌入。

(b)但是，如果 Φ 也是正确的，那么轨道 $\mathrm{Orb}(x)$ 是 M 的封闭子流形并且 $\tilde{\Phi}^x$ 是微分同胚映射。换言之：$\mathrm{Orb}(x)$ 的流形结构是由微分同胚映射的双射映射 $[g] \in G/G_x \longmapsto g \cdot x \in \mathrm{Orb}(x)$ 给出的。

实例：我们使用 4.1.0.1 节中相一致的实例的编号方式。

(i) $G = SO(3)$ 通过 $(A, x) \longmapsto Ax$ 作用于 $M = \mathbb{R}^3$。因为 $\mathrm{Orb}(x)$ 是一个以半径的初始值 $\|x\|$ 为中心的球体，$M/G \cong \mathbb{R}^+$，它不是流形。但是举例说明了结论(a)和(b)，在 x 的各向同性群 G_x 是轴心为 x 的转动群；作用量是正确的(因

为 G 是紧致的)；轨道 Orb(x)是 M 的封闭流形；并且各向同性群的陪集$[g] \in G/G_x$ 是由 $\tilde{\Phi}^x$ 微分同胚映射在球体 Orb(x)的点上。

(iii)令 X 为 $M=\mathbb{R}^3$ 之上的常数矢量场∂_x，X 是完备的。M 之上的$\mathbb{R}$的作用量有(当轨道通过点 $\mathbf{x}=(x,\ y,\ z)\in\mathbb{R}^3$)$y=$常量，$z\ =$常量。这个作用量是自由的，因而是可信的，且各向同性群的所有变元都为零。于是 $G/G_x\equiv G$。此作用量是正确的。这再次说明了结论(a)和(b)：轨道 Orb($\mathbf{x}$)是 M 的封闭子流形，即实线$\mathbb{R}=G\equiv G/G_x$ 的副本通过 $\tilde{\Phi}^x$ 与$\mathbb{R}$是微分同胚的。

4.3.0.2　轨道空间 M/G 成为流形的充分条件

根据 4.3.0.1 小节末尾的结论(b)，我们可以证明：

如果 Φ：$G\times M\to M$ 是正确的自由作用量，那么轨道空间 M/G 是一个具有 π：$M\to M/G$ 的流形。

实例：再次使用 4.1.0.1 节的编号方式。

(i)$G=SO(3)$通过$(A,\ x)\longmapsto Ax$ 作用于 $M=\mathbb{R}^3$。因为 Orb(x) 是一个以半径初始值 $\|x\|$ 为中心的球体，$M/G\cong\mathbb{R}^+$，这不是流形，且实际上这个作用量也不是自由的。

(iii)令 X 为 $M=\mathbb{R}^3$ 之上的常数矢量场∂_x。X 是完备的，并且 M 之上的$\mathbb{R}$的作用量像轨道一样具有 $y=$常数，$z=$常数。这个作用量是可信的、自由的且正确的，于是轨道空间 M/G 是一个流形：$M/G\cong\mathbb{R}^2$。

(iv)左(或者右)平动显然是群 G 作用于自身的自由作用量，并且我们注意到，如上文所述这个作用量是正确的。但是，因为这是可传递的，轨道空间 G/G 就是所有变元都为零的 0 维流形(G 的单元素集)。

(viii)依据 $SO(3)$，或者更加一般地说，依据李群 G 的左(或者右)平动的余切提升，这个作用量是正确的是在式(147)之后标注的，并且显然是自由的。

(ix)欧氏群 E 自由地作用于 $M:=T^*\mathbb{R}^{3N}-(\delta\cup\Delta)$之上，这个作用量也是正确的，这是一个给读者的(更难的)练习。

4.4　作用量的无穷小算子

现在我们把这个小节的主题(群作用量)与涉及 3.4 小节(尤其是 3.4.2)主

题的李群的李代数相结合。

令 Φ: $G\times M\to M$ 为李群 G 在流形 M 之上的(左)作用量。于是，每个$\xi\in\mathfrak{g}$定义了 M 之上的$\mathbb{R}$的作用量，我们在下面写作 Φ^{ξ}。

我们可以依据与 ξ 对应的左不变矢量场 X_{ξ} 的幂或是根据对 ξ 自身求幂认为：

$$\Phi^{\xi}:\ \mathbb{R}\times M\to M:\ \Phi^{\xi}(\tau,\ x) := \Phi(\exp(\tau X_{\xi}),\ x) \equiv \Phi(\exp(\tau\xi),\ x) \tag{151}$$

也就是说，根据我们对于原始作用量 Φ 的下标符号，参见式(116)：$\Phi_{\exp(\tau X_{\xi})} \equiv \Phi_{\exp(\tau\xi)}$: $M\to M$ 是 M 上的流。

于是，这个流是完备的，即我们再次按照 3.4.2 小节的(2)的再次求幂，尤其在式(76)之后，定义了$\mathbb{R}$的作用量。参见 4.1 小节的实例(iii)。

我们认为，M 之上对应的矢量场写作 ξ_M，即根据下式在 $x\in M$ 上定义的矢量场：

$$\xi_M(x) := \frac{d}{d\tau}\Big|_{\tau=0}\Phi_{\exp(\tau X_{\xi})}(x) \equiv \frac{d}{d\tau}\Big|_{\tau=0}\Phi_{\exp(\tau\xi)}(x) \tag{152}$$

是与 ξ 相一致的作用量的无穷小算子。

按照在式(149)中定义的映射 Φ^{x}，对于所有 $\xi\in\mathfrak{g}$，我们有：

$$\xi_M(x) = (T_e\Phi^{x})(\xi) \tag{153}$$

注意：我们用不同的方式(尽管是相关联的)使用“无穷小算子”这个词。在 3.4.2 节末尾的评论 (2) 中，群 G 之上的矢量场，或者元素 $\xi\in\mathfrak{g}$，都被称为“无穷小算子”。这里的无穷小算子就是作用量空间 M 之上的矢量场。同样地，我们要注意：ξ_M 是 M 之上的矢量场，而 X_{ξ} 是 G 之上的矢量场。

作为一个例子，我们再次让转动群 $SO(3)$作用于$\mathbb{R}^3$：$(A,\ \mathbf{x})\in SO(3)\times\mathbb{R}\longmapsto A\mathbf{x}$。我们很容易就能发现，鉴于 $\omega\in\mathbb{R}^3$，于是 $\Theta(\omega)\in\mathfrak{so}(3)$，与 $\xi\equiv\Theta(\omega)$相一致的作用量的无穷小算子就是$\mathbb{R}^3$ 之上的矢量场：

$$\xi_{\mathbb{R}^3}(\mathbf{x}) \equiv (\Theta(\omega))_{\mathbb{R}^3}(\mathbf{x}) = \omega\wedge\mathbf{x} \tag{154}$$

尤其是，$\mathbb{R}^3$ 之上代表了无穷小的 x 轴的逆时针方向的转动的矢量场是 $e_1 := y\partial_z - z\partial_y$，参见式(107)。同样地，关于 y 轴和 z 轴转动的作用量的无穷小算子分别是 $e_2 := z\partial_x - x\partial_z$ 和 $e_3 := x\partial_y - y\partial_x$。下式给出了李括号：

$$[e_1,\ e_2] = -e_3 \quad [e_3,\ e_1] = -e_2 \quad [e_2,\ e_3] = -e_1 \tag{155}$$

这里的负号就是平动$\xi \in \mathfrak{g} \longmapsto \xi_M \in \mathcal{X}(M)$的一般特征，参见下面的结果(4)。

如另一个例子所示，我们采用群G之上的左和右平移的无穷小算子(我们关于辛约化的定理需要这个实例，参见6.5.3，7.2和7.3.3节)。注意：这里会有一个“左—右替换”，这继续了3.4.2.2小节的(4)中的讨论，我们把左不变量矢量场和右不变量矢量场相比较来定义李群的李代数。

因为左平移$\Phi(g, h) \equiv L_g h := gh$，对于所有$\xi \in \mathfrak{g}$，我们有：

$$\Phi^{\xi}(\tau, h) = (\exp \tau\xi) h = R_h(\exp \tau\xi) \tag{156}$$

于是，无穷小算子是：

$$\xi_G(g) = (T_e R_g)\xi \tag{157}$$

因而，ξ_G是一个右不变量矢量场，并且除非G是阿贝尔的，否则它就不等于左不变量矢量场$g \longmapsto X_{\xi}(g) :=(T_e L_g)\xi$，参见式(68)和式(70)。同样地，对于右平移而言，其是右作用量，参见4.1.0.1小节中的(1)(iv)，无穷小算子就是左不变量矢量场：

$$g \longmapsto X_{\xi}(g) :=(T_e L_g)\xi \tag{158}$$

三个直接的结论把无穷小算子的概念和先前的观点联系在一起。我将不会给出证明，而是按照先前观点的顺序呈现出它们。

(1)回想在3.4.3节的末尾，尤其是式(97)中的李子群和李子代数之间的一致性。这些意味着，当$x \in M$(被称为各向同性代数)，各向同性群G_x的李代数是：

$$\mathfrak{g}_x = \{\xi \in \mathfrak{g}: \xi_M(x) = 0\} \tag{159}$$

(2)在4.2节的(1)的讨论中，无穷小算子ξ_M给出了等价概念的微分版本，这个版本被称为无穷小等变性。在式(144)中，我们设定了$g = \exp(\tau\xi)$，且关于τ在$\tau = 0$处求微分。这就给出了$Tf \circ \xi_M = \xi_N \circ f$。也就是说，$\xi_M$和$\xi_N$是$f$相关的。按照$f$的回拉$f^*$，我们有$f^* \xi_N = \xi_M$。

(3)假设作用量Φ是正确的，于是根据4.3.0.1节末尾的结论(b)，任意点$x \in M$的轨道$\mathrm{Orb}(x)$就是M的(闭合的)子流形。因而，在$\mathrm{Orb}(x)$中点y的对于$\mathrm{Orb}(x)$的余切空间是：

$$T\mathrm{Orb}(x)_y = \{\xi_M(y): \xi \in \mathfrak{g}\} \tag{160}$$

最终，如下所示，第四个结论把无穷小算子 ξ_M 和先前的概念联系在了一起。然而，这不如先前的(1)—(3)那样直接明了，因为其证据需要下个小节描述的伴随表示的概念。

(4)无穷小算子映射 $\xi \longmapsto \xi_M$ 设立了$\mathfrak{g}$和 M 之上的所有矢量场的李代数 $\mathcal{X}_M$之间的反同态的李代数。对比$\mathfrak{g}$和群 G 之上的左不变量矢量场的集合$\mathcal{X}_L(G)$之间同构的李代数；见3.4.2节，尤其是式(70)。也就是说：

$$(a\xi + b\eta)_M = a\xi_M + b\eta_M;\ [\xi_M,\ \eta_M] = -[\xi,\ \eta]_M \quad \forall \xi,\ \eta \in \mathfrak{g},\ 和\ a,\ b \in \mathbb{R} \tag{161}$$

顺便说一下，回到3.4.2.2小节的(4)，我们深思熟虑地按照右不变量矢量场而不是左不变量矢量场定义了李群的李代数，如果我们这样做了，与之对应的映射 $\xi \longmapsto \xi_M$ 就会成为李代数同态。

4.5 伴随和余伴随表示

随后的某些章节(尤其是5.4节，6.4节和第7节)的主要观点认为在任意李群的特定自然表示的轨道上，存在固有的辛结构。也就是说，在其自身李代数的对偶之上的群的表示，被称为余伴随表示。我们在这里引入这个表示，然而我们首先描述了它自己在李代数之上的李群表示。尽管与辛结构不同(并且与力学中的应用不同)，这两个表示都阐述了前面的小节中的观点。我将再次使用 $SO(3)$ 和$\mathfrak{so}(3)$作为例证。

4.5.1 伴随表示

我们继续第四个阶段。首先定义表达形式，随后讨论无穷小算子，然后讨论矩阵李群，最后讨论转动群。

(1)定义表达

令 G 为一个李群且$\mathfrak{g}$是它的李代数，即单位元素 $e \in G$ 的群的正切空间，具有换位子括号运算[,]。

我们回想起(例如，在3.4.2节一开始的时候)G 通过左和右平移作用于其自身，所有 $g \in G$ 通过下式定义了 G 到其自身之上的微分同胚映射：

$$L_g:\ h \in G \longmapsto gh \in G;\ R_g:\ h \in G \longmapsto hg \in G \tag{162}$$

对于每个 $h\in G$，归纳出的正切空间映射是：

$$L_{g*}:\ TG_h\to TG_{gh}\text{和}\ R_{g*}:\ TG_h\to TG_{hg} \tag{163}$$

微分同胚映射 $K_g:=R_{g^{-1}}\circ L_g$（即通过 g 的共轭：K_g：$h\longmapsto ghg^{-1}$）就是 G 的内自同构，参见 4.1 小节末尾的实例（v）。其在单位元素 $e\in G$ 的导数是从李代数$\mathfrak{g}$到它自身的线性映射，这表示为：

$$Ad_g:=(R_{g^{-1}}\circ L_g)_{*e}:\ \mathfrak{g}\to\mathfrak{g} \tag{164}$$

于是，令 g 随着通过 G 变化，映射 Ad：$g\longmapsto Ad_g$ 为每个 g 分配了一个数字 End($\mathfrak{g}$)，即（自同态）$\mathfrak{g}$上的线性映射空间。链式法则意味着 $Ad_{gh}=Ad_gAd_h$。于是：

$$Ad:\ g\longmapsto Ad_g \tag{165}$$

是在$\mathfrak{g}$：$G\times\mathfrak{g}\to\mathfrak{g}$之上的一个关于 G 的左作用量，即一个表示。这被称为伴随表示。

我们从 3.4.2.2 节的（1）和（3）的结论推断出三个关于 Ad 有用的结论，参见式（80）中关于幂的同态。

①如果 $\xi\in\mathfrak{g}$生成了单参数的子集 $H=\{\exp(\tau X_\xi):\ \tau\in\mathbb{R}\}$，那么 $Ad_g(\xi)$ 生成了共轭子集 $K_g(H)=gHg^{-1}$。

$$\exp(Ad_g(\xi))=K_g(\exp\xi):=g(\exp\xi)g^{-1} \tag{166}$$

顺便说一下，式（166）有一个多参量的推广。令 H 和 H'为和对应的李代数 $\mathfrak{g}=\mathfrak{g}(G)$ 的李子代数$\mathfrak{h}$和$\mathfrak{h}'$相关的两个李群 G 的 r 维李子群。那么 H 和 H'是共轭子群，$H'=gHg^{-1}$，当且仅当$\mathfrak{h}$和$\mathfrak{h}'$是对应的共轭子代数，即$\mathfrak{h}'=Ad_g(\mathfrak{h})$。

②式（166）也意味着另一个结论，这个结论是 6.5.2 节中关于辛约化的重要结论所必需的（我们刚刚提到的多参量的推广是不必要的）。它把 Ad 与任意作用量 Φ 的回拉相联系。

因而，令 Φ 为 M 上 G 的左作用量。那么，对于所有 $g\in G$ 和 $\xi\in\mathfrak{g}$而言：

$$(Ad_g\xi)_M=\Phi^*_{g^{-1}}\xi_M \tag{167}$$

其中 Φ^* 表示了矢量场的回拉。因为我们有：

$$(Ad_g\xi)_M(x):=\frac{d}{d\tau}\Big|_{\tau=0}\Phi(\exp(\tau Ad_g\xi),\ x) \tag{168}$$

根据式（166）可得：

$$=\frac{d}{d\tau}|_{\tau=0}\Phi(g(\exp\tau\xi)g^{-1},\ x) \tag{169}$$

$$=\frac{d}{d\tau}|_{\tau=0}(\Phi_g\circ\Phi_{\exp\tau\xi}\circ\Phi_{g^{-1}},\ (x)) \tag{170}$$

根据链式规则和式(152)得到：

$$=T_{\Phi_{g^{-1}}(x)}\Phi_g(\xi_M(\Phi_{g^{-1}}(x))) \tag{171}$$

根据回拉的定义得到：

$$=(\Phi^*_{g^{-1}}\xi_M)(x) \tag{172}$$

这个结论不仅仅是稍后需要用到的。此外，这是4.4节末尾关于结论(4)的证明的主要部分，即 $\xi\longmapsto\xi_M$ 是反同态的李代数。

③Ad_g 是同态代数，即：

$$Ad_g[\xi,\ \eta]=[Ad_g\xi,\ Ad_g\eta],\ \xi,\ \eta\in\mathfrak{g} \tag{173}$$

(2)无穷小算子：映射 ad

映射 Ad 是可微的。它在 $e\in G$ 的导数是从李代数到$\mathfrak{g}$之上的线性映射空间的线性映射。这个映射被称为 ad，并且它对于自变量 $\xi\in\mathfrak{g}$的值被写作 ad_ξ。也就是说：

$$ad:=Ad_{*e}:\ \mathfrak{g}\rightarrow\mathrm{End}\ \mathfrak{g};\ ad_\xi=\frac{d}{d\tau}|_{\tau=0}Ad_{\exp(\tau\xi)} \tag{174}$$

其中 $\exp(\tau\xi)$是在单位元素具有正切矢量 ξ 的单参数子群。但是，如果我们把这个关于作用量的无穷小算子的定义应用于伴随作用量 Ad 参见式(152)，对于所有 $\xi\in\mathfrak{g}$，算子 $\xi_{\mathfrak{g}}$，即在$\mathfrak{g}$之上的矢量场，我们得到：

$$\xi_{\mathfrak{g}}:\ \eta\in\mathfrak{g}\longmapsto\xi_{\mathfrak{g}}(\eta)\in\mathfrak{g},\ \text{其中}\ \xi_{\mathfrak{g}}(\eta):=\frac{d}{d\tau}|_{\tau=0}Ad_{\exp(\tau\xi)}(\eta) \tag{175}$$

与式(174)相比，我们发现 ad_ξ 只是与 ξ 相对应的伴随作用量的无穷小算子 $\xi_{\mathfrak{g}}$：

$$ad_\xi=\xi_{\mathfrak{g}} \tag{176}$$

我们现在来计算伴随作用量的无穷小算子。这对于$\mathfrak{g}$中的李括号所给出的后面的内容(尤其是5.4节)将是至关重要的。

我们首先来考虑函数 $Ad_{\exp(\tau\xi)}(\eta)$是如何微分的。根据式(164)，我们有：

$$\begin{aligned}Ad_{\exp(\tau\xi)}(\eta)&=T_e(R_{\exp(-\tau\xi)}\circ L_{\exp(\tau\xi)})(\eta)\\&=(T_{\exp(\tau\xi)}(R_{\exp(-\tau\xi)})\circ T_eL_{\exp(\tau\xi)})(\eta)\end{aligned}$$

$$= (T_{\exp(\tau\xi)}(R_{\exp(-\tau\xi)}) \cdot X_\eta(\exp(\tau\xi)) \tag{177}$$

其中，第二行依据链式规则，第三行依据左不变量矢量场的定义。我们把 X_ξ 的流写作 $\phi_\tau(g) = g\exp\tau\xi = R_{\exp(\tau\xi)}g$，并且应用这个关于李导数的定义，参见式(45)，于是我们有：

$$\xi_{\mathfrak{g}}(\eta) := \frac{d}{d\tau}|_{\tau=0}Ad_{\exp(\tau\xi)}(\eta) = \frac{d}{d\tau}[T_{\phi\tau(e)}\phi_\tau^{-1} \cdot X_\eta(\phi_\tau(e))]|_{\tau=0} = [X_\xi, X_\eta](e)$$

$$= [\xi, \eta] \tag{178}$$

其中，最后的等式是李代数中李括号的定义，参见式(73)。

于是，对于伴随作用量而言，与 ξ 相一致的无穷小算子采取了李括号：$\eta \longmapsto [\xi, \eta]$。总之，现在式(174)和式(175)成为：

$$ad = Ad_{*e}: \mathfrak{g} \to \mathrm{End}\,\mathfrak{g};\ ad_\xi = \frac{d}{d\tau}|_{\tau=0}Ad_{\exp(\tau\xi)} = \xi_{\mathfrak{g}}: \eta \in \mathfrak{g} \longmapsto [\xi, \eta] \in \mathfrak{g} \tag{179}$$

（3）实例：矩阵李群

在 $G \subset GL(n, \mathbb{R})$ 是一个具有李代数 $\mathfrak{g} \subset \mathfrak{gl}(n)$ 的矩阵李群的情况中，我们很容易就能证实这些结论。我们把矩阵 $n \times n$ 写作 A，$B \in G$，其共轭是 $K_A(B) = ABA^{-1}$，根据下式的共轭，我们也得到了伴随映射：

$$Ad_A(X) = AXA^{-1},\ A \in G,\ X \in \mathfrak{g} \tag{180}$$

因为 $A(\tau) = \exp(\tau X)$，于是有 $A(0) = I$ 以及 $A'(0) = X$，对于 $Y \in \mathfrak{g}$，我们有：

$$\frac{d}{d\tau}|_{\tau=0}Ad_{\exp\tau X}Y = \frac{d}{d\tau}|_{\tau=0}[A(\tau)YA(\tau)^{-1}] \tag{181}$$

$$= A'(0)YA^{-1}(0) + A(0)YA^{-1'}(0)$$

然而对 $A(\tau)A^{-1}(\tau) = I$ 进行微分，得到：

$$\frac{d}{d\tau}(A^{-1}(\tau)) = -A^{-1}(\tau)A'(\tau)A^{-1}(\tau)，因此 A^{-1'}(0) = -A'(0) = -X \tag{182}$$

于是，我们实际上有：

$$\frac{d}{d\tau}|_{\tau=0}Ad_{\exp\tau X}Y = XY - YX = [X, Y] \tag{183}$$

（4）实例：转动群

我们给予 $G=SO(3)$ [即 $\mathfrak{g}=\mathfrak{so}(3)$]的情况以详细说明是值得的。我们在3.4.4小节中式(107)看到了三个矩阵

$$A^x=\begin{pmatrix}0&0&0\\0&0&-1\\0&1&0\end{pmatrix},\ A^y=\begin{pmatrix}0&0&1\\0&0&0\\-1&0&0\end{pmatrix},\ A^z=\begin{pmatrix}0&-1&0\\1&0&0\\0&0&0\end{pmatrix}\tag{184}$$

生成空间 $\mathfrak{so}(3)$，并且生成了单参数的子群

$$R_\theta^x=\begin{pmatrix}1&0&0\\0&\cos\theta&-\sin\theta\\0&\sin\theta&\cos\theta\end{pmatrix},\ R_\theta^y=\begin{pmatrix}\cos\theta&0&\sin\theta\\0&1&0\\-\sin\theta&0&\cos\theta\end{pmatrix},\ R_\theta^z=\begin{pmatrix}\cos\theta&-\sin\theta&0\\\sin\theta&\cos\theta&0\\0&0&1\end{pmatrix}\tag{185}$$

表示了在物理空间 $\mathbb{R}^3$ 中围绕各自坐标轴系的逆时针转动。为了在算子 A^y 之上计算出 R_θ^x 的伴随作用量，我们对 $R_\theta^x R_\tau^y R_{-\theta}^x$ 求关于 τ 的微分，并令 $\tau=0$。即我们发现：

$$Ad_{R_\theta^x}(A^y)=R_\theta^x(A^y)R_\theta^x=\begin{pmatrix}0&-\sin\theta&\cos\theta\\\sin\theta&0&0\\-\cos\theta&0&0\end{pmatrix}=\cos\theta\cdot A^y+\sin\theta\cdot A^z\tag{186}$$

我们同样发现：

$$Ad_{R_\theta^x}(A^x)=A^x,\ Ad_{R_\theta^x}(A^z)=-\sin\theta\cdot A^y+\cos\theta\cdot A^z\tag{187}$$

于是，通过在李代数空间 $\mathfrak{so}(3)$ 中围绕 A^x 轴的转动，给定了表示围绕物理空间 x 轴转动的子群 R_θ^x 的伴随作用量。这同样适用于表示围绕 y 或者 z 轴转动的其他子群。而且，相对于 $\mathbb{R}^3$ 的已知轴心 x，y，z 的任意转动矩阵 $R\in SO(3)$，作用于 $\mathfrak{so}(3)\cong\mathbb{R}^3$ 的伴随映射 Ad_R 具有和 $\mathfrak{so}(3)$ 的归纳基(induced basis) $\{A^x, A^y, A^z\}$ 相同的矩阵表示。注意：这个在 $SO(3)$ 的表示和它的自然物理解释之间的论证对 $SO(3)$ 而言是特殊的，因为它不适用于其他矩阵李群。

最终，通过微分法给出了伴随作用量的无穷小算子。例如使用式(108)，我们发现：

$$ad_{A^x}(A^y):=\frac{d}{d\theta}\Big|_{\theta=0}Ad_{R_\theta^x}A^y=A^z\tag{188}$$

与对易子(commutator)$A^z=[A^x, A^y]$相一致。

4.5.2 余伴随表示

现在，让我们更进一步。首先我们定义了这个表示，随后讨论了无穷小算子，然后把转动群当成是一个实例。

(1) 定义表达

我们回想起线性映射 $A: V\to W$ 独立归纳了对偶空间上的转置(对偶)映射，写作 A^* 或 $\tilde{A}$ 或 A^T，$A^*: W^*\to V^*$，其中 $V^*:=\{\alpha: V\to\mathbb{R} \mid \alpha$ 线性$\}$，这同样地适用于 W^*，其依据是：

$$\forall\alpha\in W^*,\ \forall v\in V:\ A^*(\alpha)(v)\equiv\langle A^*(\alpha);\ v\rangle:=\alpha(A(v))\equiv(\alpha\circ A)(v) \quad (189)$$

于是，矢量空间 V 之上的群 G 的任意表示(比方说$\mathcal{R}$)，$\mathcal{R}: G\to\mathrm{End}(V)$，通过转置，归纳出了对偶空间 V^* 之上的 G 的表示$\mathcal{R}^*$。我们应该把$\mathcal{R}^*$叫作$\mathcal{R}$的对偶或者转置，有时也叫作“逆步表示(contragredient representation)”。也就是说，对于$\mathcal{R}(g): V\to V$，我们根据下式，定义$\mathcal{R}^*(g): V^*\to V^*$：

$$\mathcal{R}^*(g):\ \alpha\in V^*\to\mathcal{R}^*(g)(\alpha):=\alpha(\mathcal{R}(g))\in V^* \quad (190)$$

因而，在$\mathfrak{g}$之上的 G 的表示归纳出在其李代数$\mathfrak{g}$的对偶$\mathfrak{g}^*$之上(即在恒等式 $\mathfrak{g}^*=T_e^*G$ 处的群 G 的余切空间上)的 G 的余伴随表示。余伴随表示将在辛约化中起到核心的作用(从5.4小节开始)。

于是，令 $Ad_g^*: \mathfrak{g}^*\to\mathfrak{g}^*$ 为 Ad_g 的对偶(亦称转置)，根据下式定义：

$$\forall\alpha\in\mathfrak{g}^*,\ \xi\in\mathfrak{g}:\ \langle Ad_g^*\alpha;\ \xi\rangle:=\langle\alpha;\ Ad_g\xi\rangle \quad (191)$$

因为 $Ad: g\longmapsto Ad_g$ 是一个左作用量 $(Ad_{gh}=Ad_gAd_h)$，于是 $g\longmapsto Ad_g^*$ 是一个右作用量。所以，为了定义左作用量，我们使用了逆 g^{-1}，参见式(117)和式(126)。也就是说，我们定义了左作用量：

$$(g,\ \alpha)\in G\times\mathfrak{g}^*\longmapsto Ad_{g^{-1}}^*\alpha\in\mathfrak{g}^* \quad (192)$$

这称之为 g^* 之上 G 的余伴随作用量。并且$\mathfrak{g}^*$之上与之对应的 G 的余伴随表示，由下式表示：

$$Ad^*:\ G\to\mathrm{End}(\mathfrak{g}^*);\ Ad_{g^{-1}}^*=(T_e(R_g\circ L_{g^{-1}}))^* \quad (193)$$

(2) 映射 ad^*；无穷小算子

映射 Ad^* 是可微分的。它在 $e\in G$ 的导数是从李代数$\mathfrak{g}$到$\mathfrak{g}^*$之上的线性映射

空间的线性映射。这个映射被称为 ad^*，并且其对于自变量 $\xi \in \mathfrak{g}$的值被写作 ad_ξ^*。因而，ad_ξ^* 是$\mathfrak{g}^*$的自同态，并且我们有：

$$ad^* = Ad^*_{*e}: \xi \in \mathfrak{g} \to Ad_\xi^* \in \mathrm{End}\, \mathfrak{g}^* \tag{194}$$

现在回想我们来自式(174)和式(175)的推论 $ad_\xi = \xi_\mathfrak{g}$，即式(176)。在这里，我们以同样的方式从等式推论出余伴随作用量的无穷小算子：

$$ad_\xi^* = \xi_{\mathfrak{g}^*} \tag{195}$$

事实上，在通常以矢量空间和其对偶为自然对的意义上，正如我们现在所说明的，ad_ξ^* 是(以负号为模的)ad_ξ 的伴随。因而符号 ad^*是正确的，模为负。

我们来计算对于这个作用量的无穷小算子 $\xi_{\mathfrak{g}^*}$(由 $\xi \in \mathfrak{g}$产生的在$\mathfrak{g}^*$之上的矢量场)在点 $\alpha \in \mathfrak{g}^*$ 的值。也就是说，我们将计算出值 $\xi_{\mathfrak{g}^*}(\alpha)$。像通常一样，我们把这个值存在的切空间$(T\mathfrak{g}^*)_\alpha$ 和$\mathfrak{g}^*$本身看成是一样的，这也同样适用于$\mathfrak{g}$。于是，$\xi_{\mathfrak{g}^*}$作用于 $\eta \in \mathfrak{g}$，我们计算出：

$$< ad_\xi^*(\alpha);\ \eta > \equiv < \xi_{\mathfrak{g}^*}(\alpha);\ \eta > = \langle \frac{d}{d\tau}|_{\tau=0} Ad^*_{\exp -\tau\xi}(\alpha);\ \eta\rangle \tag{196}$$

$$= \frac{d}{d\tau}|_{\tau=0}\langle Ad^*_{\exp -\tau\xi}(\alpha);\ \eta\rangle = \frac{d}{d\tau}|_{\tau=0}\langle \alpha;\ Ad_{\exp -\tau\xi}\eta\rangle \tag{197}$$

$$= \langle \alpha;\ \frac{d}{d\tau}|_{\tau=0} Ad_{\exp -\tau\xi}\eta\rangle = < \alpha;\ -[\xi,\ \eta] > = - < \alpha;\ ad_\xi(\eta) > \tag{198}$$

于是，被定义为 Ad_ξ^* 的导数的 ad^*是由一个符号决定的，即 ad_ξ 的伴随。

(3) 实例：转动群

我们现在把$\mathbb{R}^3$ 中的基本矢量积写作$\wedge$，并且来确定$\mathfrak{so}(3) \cong (\mathbb{R}^3,\ \wedge)$和$\mathfrak{so}(3)^* \cong \mathbb{R}^{3^*}$。并且让我们确定通过基本欧氏内积·而给定的自然对。于是，我们的结论为(现在用•标记争论的地方)：

$$< \xi_{\mathfrak{g}^*}(\alpha);\ \bullet > = - < \alpha;\ [\xi,\ \bullet] > \tag{199}$$

对于 $\alpha \in \mathfrak{so}(3)^*$和 $\xi \in SO(3)$，这变成：

$$\xi_{\mathfrak{so}(3)^*}(\alpha) \cdot \bullet = -\alpha \cdot (\xi \wedge \bullet) \tag{200}$$

于是，对于 $\eta \in \mathfrak{so}(3)$，我们有：

$$<\xi_{\mathfrak{so}(3)^*}(\alpha);\ \eta> = \xi_{\mathfrak{so}(3)^*}(\alpha)\cdot\eta = -\alpha\cdot(\xi\wedge\eta)$$
$$= -(\alpha\wedge\xi)\cdot\eta = -<\alpha\wedge\xi;\ \eta> \tag{201}$$

简言之：

$$\xi_{\mathfrak{so}(3)^*}(\alpha) = -\alpha\wedge\xi = \xi\wedge\alpha \tag{202}$$

现在，因为 $SO(3)$ 是紧致的，我们知道余伴随作用量是正确的，于是 $\mathrm{Orb}(\alpha)$ 是 $\mathfrak{so}(3)^*$ 的闭合子流形，并且应用于4.4小节的式(160)。于是，如果我们确定 α，并且令 ξ 随着 $\mathfrak{so}(3)\cong\mathbb{R}^3$ 发生变化，我们就得到了经过 α 的轨道的全部切空间 $T_\alpha\mathrm{Orb}(\alpha)$。于是，式(202)意味着切空间垂直于 α，并且经过 α 的终点。令 α 随着 $\mathfrak{so}(3)^*$ 变化，于是我们总结出余伴随的轨道是以初始点为中心的球体。

在下文中，我们将发现任意李群 G 的余伴随表示的轨道具有自然的辛结构。于是轨道总是偶数维的，并且通过考虑所有李群和所有可能的轨道，我们得到了关于辛流形的一系列的实例。

除此之外，这个事实将在哈密顿力学的广义构造和辛约化中发挥核心的作用。并且我们已经在5.1节(幸运地)得到对这个作用充分的理解。我们应当收集关于我们经常使用的实例的一些思路，并把它们推广到其他的李群……

4.6 李群上的运动学

为了总结本小节的某些观点，并且为了使我们随后关于约化的讨论更加清楚明了，我们需要收集和归纳关于 $SO(3)$ 的结论以及它所提供的刚体的描述。更加清楚地说，现在我们要结合：

(i) 在3.4.4节的末尾，我们依据左和右平移描述空间和立体的坐标系；

(ii) 平动的余切提升，参见4.1.0.1小节的例(viii)；

(iii) $SO(3)$ 的伴随和余伴随表示，参见如4.5.1节的(4)和4.5.2节的(3)所述。

我们也将概括出，即我们将考虑把(i)和(iii)结合在一起，使之适用于任意李群 G，而不仅仅只适用于 $SO(3)$。我们这样做的理由将会在5.1节的(3)中变得清晰。这已经在4.6.1小节中出现了。于是，在4.6.2节中，我们将说明这些材料是如何产生固有微分同胚映射的：

$$TG \to G \times \mathfrak{g} \text{和} T^* G \to G \times \mathfrak{g}^* \tag{203}$$

因此，如果 $\dim G = n$，那么四个流形全是 $2n$ 维的。我们同样也会发现：通过应用 4.2 节的等变性的概念，我们可以“转而讨论商”，并且从式(203)得到固有微分同胚映射的：

$$TG/G \to \mathfrak{g} \text{和} T^* G/G \to \mathfrak{g}^* \tag{204}$$

其中，左边(那个域)的商依据的是左平移的作用量(更加准确地说，这是依据其导数对于 TG 的作用量，以及其对于 $T^* Q$ 的余切提升)。

4.6.1 推广到 G 的空间和立体坐标系

于是，令(有限维的)李群 G 通过左和右平移(L_g 和 R_g)对其自身起作用。对于任意 $g \in G$，我们通过 $v \in T_g G \longmapsto (T_e L_g)^{-1}(v) \equiv (T_g L_{g^{-1}})(v) \in \mathfrak{g}$定义：

$$\lambda_g: T_g G \to \mathfrak{g} \tag{205}$$

类似地，我们定义：

$$\rho_g: v \in T_g G \longmapsto (T_e R_g)^{-1}(v) \equiv (T_g R_{g^{-1}})(v) \in \mathfrak{g} \tag{206}$$

在与转动刚体的例子类比时，参见式(113)和式(114)，或者式(137)，我们认为 λ_g 在立体坐标系中表示 $v \in T_g G$，并且 ρ_g 在空间坐标系中表示 v。我们也谈到了立体和空间的表示。于是，从立体坐标系到空间坐标系的平动是$\mathfrak{g}$的同构，即通过式(164)，我们得到：

$$\forall \xi \in \mathfrak{g}, \ (\rho_g \circ \lambda_g^{-1})(\xi) = \rho_g(T_e L_g(\xi)) \equiv Ad_g \xi \tag{207}$$

于是，我们能够把式(137)的 S 和 B 的上标符号与 4.5.1 小节的表示的概念相结合，并且写作：

$$v^S = Ad_g v^B \tag{208}$$

用相同的方式，左和右平移的余切提升提供了对偶空间 $T_g^* G$，$g \in G$ 和$\mathfrak{g}^*$之间的同构。因而，对于任意 $g \in G$，我们通过 $\alpha \in T_g^* G \longmapsto \alpha \circ T_e L_g \equiv (T_e L_g)^*(\alpha) \equiv (T_e^* L_g)(\alpha) \in \mathfrak{g}^*$ 定义：

$$\bar{\lambda}_g: T_g^* G \to \mathfrak{g}^* \tag{209}$$

并且，同样可得：

$$\bar{\rho}_g: \alpha \in T_g^* G \longmapsto \alpha \circ T_e R_g \equiv (T_e^* R_g)(\alpha) \in \mathfrak{g}^* \tag{210}$$

并且，我们再次使用了式(137)的 S 和 B 的上标符号，且对于 $\alpha \in T_g^* G$ 我们定

义：

$$\alpha^S := (T_e^* R_g)(\alpha) \equiv \bar{\rho}_g(\alpha) \text{和} \alpha^B := (T_e^* L_g)(\alpha) \equiv \bar{\lambda}_g(\alpha) \tag{211}$$

我们把它们分别叫作 α 的空间（或者“空间的”）表示和立体表示。现在，从立体表示到空间表示的平动是$\mathfrak{g}^*$的一个同构，即：

$$\forall \alpha \in \mathfrak{g}^*,\ (\bar{\rho}_g \circ \bar{\lambda}_g^{-1})(\alpha) = Ad^*_{g^{-1}}(\alpha)，\text{即 } \alpha^S = Ad^*_{g^{-1}}(\alpha^B) \tag{212}$$

4.6.2 通往商的途径

出于今后的目标，我们需要研究出单位元素 $g \in G$ 从式(205)“一直贯穿到”式(212)的细节。更加准确地说，我们有两个同构：

$$TG \cong G \times \mathfrak{g} \text{和} T^*G \cong G \times \mathfrak{g}^* \tag{213}$$

这些是矢量丛的同构，但是我们不会提及纤维丛的语言。对我们而言，重要的是：一旦我们展示了这些同构，我们就将发现，按照式(144)和式(145)，我们拥有与两个群作用量等变化的映射。并且这将意味着，我们可以转向商从而推断出 TG/G 与$\mathfrak{g}$微分同胚，并且对应地，T^*G/G 与$\mathfrak{g}^*$微分同胚。

最后的这个微分同胚映射将构成第 7 小节的主要定理的第一个部分，即李—泊松约化定理，其说明了 T^*G/G 和$\mathfrak{g}^*$是与泊松流形同构的。从第 5 小节开始，我们将提出泊松流形的概念，以及这个同构对于力学问题的还原的重要性。

这里，我需要注明，关于第一个微分同胚映射（即关于和$\mathfrak{g}$是微分同胚的 TG/G），有一个相同的经历。它构成了另一个约化定理的第一个部分，也就是第 7 小节的李—泊松定理的拉格朗日算符的类比。但是，由于本章采用了哈密顿的方法，我将不会详细叙述。详见马斯登和拉蒂乌[1999, Sections 1.2, 13.5, 13.6]，标题为《欧拉—庞加莱约化》(*Euler-Poincaré reduction*)的文章。

因为，与式(205)相一致，我们通过 $\lambda(v) := (g, (T_e L_g)^{-1}(v)) \equiv (g, (T_g L_{g^{-1}})(v))$定义同构：

$$\lambda: TG \to G \times \mathfrak{g} \tag{214}$$

其中 $v \in T_g G$，即 $g = \pi G(v)$ 且 π_G：$TG \to G$ 是正则投影(canonical projection)。正如式(121)提到的，（按照一些投影）合并 TG 中的一个点是无妨的，严格地说，一个具有矢量 v 的对 (g, v)，其中 $g \in G$，$v \in T_g G$。并且与式(206)相一

致，我们通过$\rho(v) := (g, (T_eR_g)^{-1}(v)) \equiv (g, (T_gR_{g^{-1}})(v))$定义同构：

$$\rho: TG \to G \times \mathfrak{g} \tag{215}$$

现在，由式(207)给出的从立体到空间表示的平动意味着：

$$(\rho \circ \lambda^{-1})(g, \xi) = \rho(g, T_eL_g(\xi)) = (g, (T_eR_g)^{-1} \circ T_eL_g(\xi)) = (g, Ad_g\xi) \tag{216}$$

按照同样的方式，余切从T^*G按照两种方式与$G \times \mathfrak{g}^*$是同构的，也就是通过：

$$\bar{\lambda}(\alpha) := (g, \alpha \circ T_eL_g) \equiv (g, (T_e^*L_g)\alpha) \in G \times \mathfrak{g}^* \tag{217}$$

或通过：

$$\bar{\rho}(\alpha): (g, \alpha \circ T_eR_g) \equiv (g, (T_e^*R_g)\alpha) \in G \times \mathfrak{g}^* \tag{218}$$

其中，$\alpha \in T_g^*G$，即$g = \pi_G^*(\alpha)$并且π_G^*：$T^*G \to G$是正则投影。同样，我们把T^*G中的点(g, α)和它的形式$\alpha \in T_g^*G$合并在一起是没有坏处的。

现在，我们来计算立体表示，即作用量：(i)左平移映射的导数TL_g；(ii)对应的余切提升T^*L_g。这说明了，对于确定的群作用量，λ和$\bar{\lambda}$是等变化的映射。

(i) 我们计算：

$$\begin{aligned}(\lambda \circ TL_g \circ \lambda^{-1})(h, \xi) &= (\lambda \circ TL_g)(h, TL_h(\xi)) \\ &= \lambda(gh, (TL_g \circ TL_h)(\xi)) \\ &= (gh, ((TL_{gh}{}^{-1} \circ TL_{gh})(\xi)) = (gh, \xi)\end{aligned} \tag{219}$$

于是在立体表示中，左平移没有作用于矢量分量(这是直观的，因为矢量ξ是“附属于立体图形的”，因而，不会根据它的坐标系变化)。式(219)意味着λ是等变化映射，这个映射涉及从TG之上的左平移TL_g到仅通过第一个分量上的左作用量给定的$G \times \mathfrak{g}$之上的G作用量的映射：

$$\Phi_g((h, \xi)) \equiv g \cdot (h, \xi) := (gh, \xi) \tag{220}$$

等变性意味着λ归纳出了商之上的映射λ。也就是说，同式(146)中的一样，我们定义了这个映射为：

$$\hat{\lambda}: TG/G \to (G \times \mathfrak{g})/G \tag{221}$$

对于任意g，把任意$v \in T_gG$的轨道映射到$\lambda(v)$的轨道，即：

$$\hat{\lambda}: \mathrm{Orb}(v) \equiv \{u \in TG \mid T_gL_h(v) = u, \text{某些 } h \in G\} \longmapsto \mathrm{Orb}(\lambda(v)) \tag{222}$$

$$\equiv \{(hg,\ (T_e L_g)^{-1}(v)) \mid 某些\ h \in G\}$$

是定义明确的，即不依赖于轨道选择的表示 v。除此之外，因为正则投影 $v \in TG \longmapsto \mathrm{Orb}(v) \in TG/G$ 和 $(g,\ \xi) \longmapsto \mathrm{Orb}((g,\ \xi)) \in (G \times \mathfrak{g})/G$ 都是浸没，我们能够应用4.2节的结论(1)并且推断 $\hat{\lambda}$ 是平滑的。

最后，我们注意到，因为左平移的作用量是可传递的，我们能够把确定式(220)的 Φ 的每个轨道和其轨道的右分量 $\xi \in \mathfrak{g}$ 看成一样的，并因此我们能够把轨道 $(G \times \mathfrak{g})/G$ 的集合和 $\mathfrak{g}$ 看成是一样的。总而言之，我们已经说明了 TG/G 和 $(G \times \mathfrak{g})/G$ 是微分同胚的：

$$\hat{\lambda}:\ TG/G \to (G \times \mathfrak{g})/G \equiv \mathfrak{g} \tag{223}$$

(ii)关于余切丛的结论类似于我们在(i)中的那些结论。类比式(219)，我们通过应用式(217)在立体表示中给出了左平移 $T^* L_g$ 的余切提升作用量，从而得到：

$$(\hat{\lambda} \circ (T^* L_g) \circ \bar{\lambda}^{-1})(h,\ \alpha) = (g^{-1}h,\ \alpha) \tag{224}$$

或者同样地，现在我们采用左平移的余切提升来定义左作用量，参见式(216)：

$$(\bar{\lambda} \circ (T^* L_{g^{-1}}) \circ \bar{\lambda}^{-1})(h,\ \alpha) = (gh,\ \alpha) \tag{225}$$

于是，在立体表示中，左平移没有作用于余矢量分量(再一次，就 α 而言，直观的结论是“附属于立体图形”)。于是，仅是依据第一个分量的左平移，我们得出，式(225)意味着 $\hat{\lambda}$ 是把 $T^* G$ 之上的左平移的余切提升左作用量映射到 $G \times \mathfrak{g}^*$ 之上的 G 作用量的等变化映射：

$$\Phi_g((h,\ \alpha)) = g \cdot (h,\ \alpha) := (gh,\ \alpha) \tag{226}$$

于是，类比式(221)和式(222)，我们转而研究商并定义映射：

$$\dot{\bar{\lambda}}:\ T^* G/G \to (G \times \mathfrak{g}^*)/G \tag{227}$$

这要求对于所有的 $\alpha \in T_g^* G$，$T^* L_{h^{-1}} \alpha \in T_{hg}^* G$：

$$\dot{\bar{\lambda}}:\ \mathrm{Orb}(\alpha) \equiv \{\beta \in T^* G \mid \beta = T^* L_{h^{-1}}(\alpha) \tag{228}$$

$$某些\ h \in G\} \longmapsto \mathrm{Orb}(\bar{\lambda}(\alpha)) \equiv \{(hg,\ (T_e^* L_g)(\alpha)) \mid 某些\ h \in G\} \equiv$$

$$\{(h,\ (T_e^* L_g)\alpha) \mid 某些\ h \in G\}$$

最终，我们把轨道 $(G \times \mathfrak{g}^*)/G$ 的集合和 $\mathfrak{g}^*$ 看成是一样的，于是我们推断

出 T^*G/G 和 $\mathfrak{g}^*$ 是微分同胚的。也就是说，我们认为微分同胚映射 $\dot{\bar{\lambda}}$ 就如同把 T^*G/G 映射到 $\mathfrak{g}^*$：

$$\dot{\bar{\lambda}}: \mathrm{Orb}(\alpha) \equiv \{\beta \in T^*G \mid \beta = T^*L_{h^{-1}}(\alpha), \text{某些 } h \in G\} \in T^*G/G \longmapsto (T_e^*L_g)(\alpha) \in \mathfrak{g}^* \tag{229}$$

如上所述，微分同胚映射主要是第 7 节约化定理的重要部分，然而我们将在 5.1 节的(3)中见到它的作用。

最后，一个以后不再必要的结论是，为了计算空间表示中左平移的导数和余切提升，我们用式(215)和式(218)定义的 ρ 和 $\bar{\rho}$ 替代了 λ 和 $\bar{\lambda}$。如式(219)和式(224)的类比，我们分别得到：

$$(\rho \circ TL_g \circ \rho^{-1})(h, \xi) = (gh, Ad_g(\xi)) \tag{230}$$

以及

$$(\bar{\rho} \circ T^*L_g \circ \bar{\rho}^{-1})(h, \alpha) = (g^{-1}h, Ad_g^*(\alpha)) \tag{231}$$

尽管这些结论在下文中并不需要，它们也是下文中某些结论的类比，参见式(399)和式(400)，我们会需要那些结论。需要注意的是：与式(191)和式(192)之间的讨论一致，式(231)涉及了右作用量。

5. 泊松流形

5.1 导言：泊松流形的三个理由

现在，我们装备了第 3 节和第 4 节的近代几何学的工具箱，在本小节和接下来的两个小节中，我们能够发展关于辛约化的理论。本小节发展了泊松流形的一般理论，将其作为把哈密顿力学推广的框架。其主要结论涉及了泊松流形的叶状结构和商。于是，第 6 节把我们带回了对称性和守恒量，即一些自 2.1.3 节开始就为我们熟知的论题，但第 6 节我们将探讨在推广的框架中使用这个动量映射的概念。最后，在第 7 节中，我们的拼图都会组合在我们的辛约化定理中。

我们已经在 2.2 节的(1)中把泊松流形的概念看成了辛流形的推广，这为哈密顿力学的推广提供了正确的框架。这是具有括号(叫作“泊松括号”)的流

形，这个括号在本质上具有和在辛几何学中一样的形式的定义特性，除了它可以是“简并的”。尤其是，泊松流形 M 的维 m 可以是偶数或是奇数。正如我们即将看到的，我们能够在泊松力学之上设立哈密顿力学，用通常的形式体系的自然推广，即在局域坐标 x^1，x^2，…，x^m 演化时，有 m 个一阶常微分方程，并且任意动力学变量的时间导数(泊松流形 M 之上的标量函数)是由其泊松括号和哈密顿函数给出的。除此之外，这个推广在下文中被归纳为通常的形式体系。任意泊松流形 M 被分成辛流形，我们在 M 上定义的一般哈密顿力学通过归纳的辛形式将每个辛叶(symplectic leaf)限制在一个常见的哈密顿力学中。

最后一点，辛叶对动力学之下的不变性提出质疑：“既然动力学可被写在每个叶上，为何还要对泊松流形用心?”对此有三个理由。我这里只谈第一个，第 5 小节的其余部分将会研究第二个，随后的两个小节将会研究第三个。

(1) 参量和稳定性

前两个理由关注于这样的事实：对于哈密顿力学中的很多问题，我们很自然地考虑奇数维的态空间。其发生的一个主要方式是，如果系统具有参量为某个奇数的特征，例如说 s(也许 $s=1$)。那么，即使对于参量(s)的固定值，仍然存在辛流形(如 $2n$ 维的辛流形)之上的哈密顿力学，我们设想 $2n+s$ 维空间从而记录系统的行为是如何依赖于参量的，这是非常有意义的。例如，对分析稳定性是很有帮助的，尤其是如果我们能够用某种方式控制参数的值的话。稳定性理论(以及相关领域，例如分歧理论)是至关重要的且非常大的论题——我对此不予详述。①

(2) 奇数维空间：再次应用刚体

其次，即便是在没有这些可控制的参量的时候，也有一些力学系统，关于这些力学系统的描述自然地导向了奇数维的态空间。这个范式的基本例子就是以某个点为轴心的刚体，参见在 2.2 节的(3)提到的。在所有教科书中重复的初步分析导致了用角动量的三个分量对立体图形的描述(相对于体坐标，即固

① 除了谈到的一般的哲学观点。这些参量阐明了力学的模态或者虚拟事实介入。s 维的状态空间以及基于它的数学构造说明了这些介入是多么的丰富和结构化。关于力学的模态介入详细论述，请参见[Butterfield，2004]。

定于立体图形的坐标)：这些分量对应于三个一阶欧拉方程。

这个情况引起了两个基本问题(当然是大多数教科书忽视的问题!)。第一，我们注意到，立体图形的位形是由三个实数给出的，即为了详述转动需要把刚体从基准的位形转动到给定的位形中。于是，刚体的传统哈密顿描述将使用六个一阶方程(实际上，类似于拉格朗日算符的描述，如果我们把三个 $\dot{q}s$ 看成变量)。那么，欧拉方程的描述是如何与六维哈密顿(或拉格朗日算符)描述相联系的?

第二，我们能够以某种方式认为依据欧拉方程的描述本身就是哈密顿函数，或者拉格朗日算符吗?

本章中将不会探寻那些关于刚体的问题，如果想要知道详情，请参见2.2节(3)末尾的参考文献。对于我们而言，重要的是关于辛约化的理论说明了对于第二个问题的答案是肯定的。事实上，是"完全的肯定"。因为我们很快就会看到(在5.2.4.1小节)，立体坐标系中角动量的分量的三维空间就是我们关于泊松流形的标准实例，并且，依据欧拉方程的发展是这个流形的每个辛叶上的哈密顿力学。简言之，在我们推广的框架中，欧拉方程已经在哈密顿形式中了。

此外，这个泊松流形已经为我们所熟知：它就是$\mathfrak{so}(3)^*$，即转动群的李代数对偶。这里我们与上述若干讨论相联系(也与本章的第二个题词相联系)。

首先，我们将关系论中关于转动的讨论和约化论中关于转动的讨论联系起来(2.3.3至2.3.5节)。尤其是，参见关于γ的评论(iii)，即在2.3.4节末尾，编码了系统的全部角动量的三个变量。于是，当考虑到适当的参量常数来给辛叶加添标记的(1)的想法时，在这个例子中，整个立体图形角动量的大小L就是这个参量。

其次，我们与3.4.4节关于$\mathfrak{so}(3)$的讨论相联系，4.5.2节关于$\mathfrak{so}(3)^*$上余伴随表示的讨论相联系，并与4.6节关于任意李群上的运动学的讨论相联系。就刚体而言，主要的物理概念就是，我们按照空间和立体坐标系统之间的坐标变换(即转动)解释了依据左平移在其本身之上的$SO(3)$的作用量。

但是，除了刚体之外，回想在4.5.2节中，我们发现对于$\mathfrak{so}(3)^*$而言，余伴随轨道是以原点为中心的球体。我同样认为，它们具有自然的辛结构——并

且这点对于任意李群的余伴随表示的轨道而言同样如此。现在，我们拥有泊松流形的概念，我们能够再多说一点，尽管总是会给出证明过程。

对于任意李群 G，其李代数$\mathfrak{g}^*$的对偶是一个泊松流形，并且 G 在$\mathfrak{g}^*$上拥有一个余伴随表示，它们的轨道是$\mathfrak{g}^*$的辛叶，这同泊松流形一样。

特别地，我们注意到刚刚描绘的刚体理论是独立于三维物理空间维度的，对于任意 n，它都是$\mathfrak{so}(n)^*$。于是，我们很容易就能在任意维中做刚体的哈密顿力学。这听起来多少是带着学术性的！但是，这把我们带到了一个更加一般的想法中，这个想法明显具有极大的实用价值。

在工程学中，我们通常需要分析或者设计由两个刚体或多个刚体结合在一起的立体图形，例如，在万向接头(universal joint)上。通常情况下，这种结合的立体图形的位形空间可以是由转动的序列(尤其是关于连接的)和(或者)来自基准位形的平动给出的，于是，我们能够把正确的李群 G 看成是空间图形的位形空间。如果是这样的话，我们能够设法模拟我们适用于刚体的策略，即应用刚刚提到的结论。并且，实际上，对于这些立体图形，左平移的作用量，以及在$\mathfrak{g}$和$\mathfrak{g}^*$上 G 的伴随和余伴随表示，通常在物理上都是十分重要的。

但是，把工程学放在一边，我们把泊松流形的第二个原因总结如下。对于某些力学系统而言，适用于哈密顿力学的自然态空间是泊松流形。并且在刚体的情形中，我们把泊松流形的叶解释成由转动群 $SO(3)$的余伴随表示的轨道。

(3) 约化

我们的前两个理由没有提到约化。然而，它们应该与这个概念有着一些联系。在这里，我将仅仅阐述一个主要关系，这个关系把4.6节中关于李群的动力学与我们主要的约化定理联系在一起，这将是我研究泊松流形的第三个动机。

简而言之，这个联系就是：

(i)对于不同的系统而言，我们很自然地把位形空间看成是李群 G(正如我们刚刚用刚体论述的那样)。

(ii)于是，我们很自然地在余切丛 T^*G 上设立了关于这个系统的正统哈密顿力学。但是(正如2.3.4节的约化论步骤所示)，我们同样很自然地就能依据

提升到余切丛求商，这个余切丛是通过左平移作用于自身的 G 的作用量。

(iii)当我们这样做时，结果产生的还原相空间 $T^* G/G$ 是一个泊松流形。事实上，它是$\mathfrak{g}^*$的同构副本。也就是说，我们拥有泊松流形的同构$\mathfrak{g}^* \cong T^* G/G$。这就是李—泊松约化定理。

我将对(i)~(iii)的每一条动机给出详细的说明。

(i)对于不同的系统而言，把李群 G 的元素(例如单位元素 $e \in G$)作用于某些相关位形(这些位形本身就被 G 的元素贴上了标签)能够获得任何位形。于是，我们把李群 G 看成是位形空间。如 2.2 节的(3)所提到的，这甚至有一个有限维的例子，即理想流体。

(ii)于是，$T^* G$ 就是这个系统的传统哈密顿相空间。但是，G 通过左平移作用于自身。于是，我们可以根据左平移的余切提升求得 $T^* G$ 的商。直观地讲，这是关于"去除"$T^* G$ 编码(i)的参照位形方式的问题。通过像 4.6 节中那样讨论商，我们推断出，$T^* G/G$ 是一个流形。然而，它在一般意义上当然不是偶数维的。因为它的维度是$\frac{1}{2}\dim(T^* G) = \dim(G)$。于是考虑任意奇数维 G，例如我们的老朋友，即三维转动群 $SO(3)$。

(iii)然而，$T^* G/G$ 总是一个泊松流形，并且作为一个泊松流形，它总是与$\mathfrak{g}^*$同构，它的辛叶是$\mathfrak{g}^*$：$\mathfrak{g}^* \cong T^* G/G$ 的余伴随轨道。

我在此不讨论第三个理由是为了研究具有关于实例的两个泊松流形备注。

第一个备注在 4.5.2 节末尾得到了回应，其中，我认为通过考虑所有可能的李群以及其余伴随表示的所有轨道，我们得到了一系列关于辛流形的实例。现在，我们能够把它们和泊松流形的概念放在一起，并且把它们和 3.4.3 小节末尾的评论放在一起，即所有(有限维的)李代数都是李群的李代数。简而言之，我们得到了一系列关于泊松流形的实例，以两种等价方式的任意一种进行：从任意(有限维)李代数$\mathfrak{g}$的对偶$\mathfrak{g}^*$，或者从左平移的余切提升的商 $T^* G/G$。不论是以上哪种方式，实例都是余伴随表示。

第二个说明就是认为有其他的一些关于泊松流形和约化的实例。实际上，我们在 2.3.4 节中提到了一个，即约化论的约化相空间 $\bar{M} := M/E$，这是依据 $\mathbb{R}^{3N}$上的欧氏群 E 的作用量(余切提升)求得相空间 $M := T^* \mathbb{R}^{3N} - (\delta \cup \Delta)$ 的商而

得到的。但是，我没有详细描述这个实例(因为其参考了 2.3.1 节列出的贝洛特的论文，以及其中的一些参考文献)。在这里我们有足够的理由注意到这个实例不是具有上述形式的：$\mathbb{R}^{3N}$不是 E，并且在$\mathbb{R}^{3N}$上的 E 的作用量不是左平移。这些当然与在 1.2 节结尾的说明产生了共鸣，即对于这章而言，辛约化的理论太大也太复杂，于是只能作为一个“引子”。

本小节的其余部分将覆盖理由(1)和(2)，但是关于约化的理由(3)将留在第 6 节和第 7 节予以讨论。我们为泊松流形给出了一些基础，主要是 5.2 节中的与坐标有关的语言。在 5.3 节中，我们转移到更加依赖于坐标的语言，并且说明了泊松流形叶片状地进入辛流形之中。在 5.4 节中，我们说明了有限维李代数$\mathfrak{g}$的叶状结构的叶就是$\mathfrak{g}^*$上的 G 的余伴随表示的轨道。最后，在 5.5 节中，我们通过李群的作用量证明了关于商化泊松流形的一般定理，这对于第 7 节的主要定理而言是至关重要的。

5.2 基础

从 5.2.1 节到 5.2.3 节，我们提出了一些基本的定义以及关于泊松流形的结论。这使我们进入了 5.2.4 节，我们在其中说明了任意有限维的李代数的对偶拥有自然的(即与基础无关的)泊松流形结构。自始至终，都有一些有关反对称形式、泊松括号、哈密顿矢量场、李括号(2.1 和 3.2 节)的之前讨论的共鸣。但是，我在大多数情况下没有清晰地说明这些共鸣。

5.2.1 泊松括号

如果一个流形 M 具有泊松括号(又被看作是泊松结构)，那么这个流形 M 被叫作泊松流形。泊松括号就是分配给每对其他函数的平滑实值函数 F，H：$M\to\mathbb{R}$，表示为$\{F, H\}$，受下列四个条件支配：

(a) 双线性：

$$\forall a, b\in\mathbb{R},\ \{aF+bG, H\} = a\{F, H\} + b\{G, H\},\ \{F, aG+bH\} = a\{F, G\} + b\{F, H\} \tag{232}$$

(b) 反对称性：

$$\{F, H\} = -\{H, F\} \tag{233}$$

(c)雅可比恒等式：

$$\{\{F, H\}, G\}+\{\{G, F\}, H\}+\{\{H, G\}, F\}=0 \tag{234}$$

(d)莱布尼茨规则：

$$\{F, H\cdot G\}=\{F, H\}\cdot G+H\cdot\{F, G\} \tag{235}$$

换句话说，M 是泊松流形，当且仅当（i）拥有括号$\{,\}$的 M 之上的平滑标量函数的集合$\mathcal{F}(M)$是李代数，和(ii)括号$\{,\}$在所有因数中都是一个导数。

任意的辛流形都是泊松流形。泊松括号是由流形的辛形式定义的，参见式(18)。

下面是一个“规范的”实例。

令 $M=\mathbb{R}^m$，$m=2n+l$，具有正则坐标$(q, p, z)=(q^1, \cdots, q^n, p^1, \cdots, p^n, z^1, \cdots, z^l)$。用下式定义任意两个函数 $F(q, p, z)$，$H(q, p, z)$的泊松括号：

$$\{F, H\}:=\sum_i^n\left(\frac{\partial F}{\partial q^i}\frac{\partial H}{\partial p^i}-\frac{\partial F}{\partial p^i}\frac{\partial H}{\partial q^i}\right) \tag{236}$$

因而，这个括号忽视了 z 坐标；并且如果 l 等于零，那么对于$\mathbb{R}^{2n}$的正则泊松括号就如同辛流形。我们立即能够归纳出适用于这个坐标函数的泊松括号。这些对于 qs 以及 ps 和对于通常的辛例子是一样的：

$$\{q^i, q^j\}=0 \quad \{p^i, p^j\}=0 \quad \{q^i, p^j\}=\delta_{ij} \tag{237}$$

另一方面，所有这些都包括 zs 等于零：

$$\{q^i, z^j\}=\{p^i, z^j\}=\{z^i, z^j\}\equiv 0 \tag{238}$$

除此之外，任意只依赖于 a 的函数 F，$F\equiv F(z)$将得到，由于所有函数 H：$\{F, H\}=0$，这个泊松括号也就等于零。

这个实例似乎是特殊的，因为 M 被叶化(foliated)到了 $2n$ 维的辛流形中，每个都由 zs 的 l 常数值标记。但是，5.3.4 节将给出达布定理的泊松流形的概括(在 2.1.1 节末尾提出的)。粗略地说，这个概括说明，所有泊松流形“在局部上看来都是如此”。

对于任意泊松流形而言，如果这个伴随所有平滑函数 H：$M\to\mathbb{R}$的泊松括号等于零，即$\{F, H\}=0$，那么我们认为函数 F：$M\to\mathbb{R}$是可区分的(distinguished)或者是卡西米尔(Casimir)的。

5.2.2 哈密顿矢量场

已知平滑函数 $H: M\to\mathbb{R}$，考虑在平滑函数上的这个映射 $F\longmapsto\{F, H\}$。泊松括号是双线性的并且服从莱布尼茨规则的事实意味着这个映射 $F\longmapsto\{F, H\}$ 在平滑函数的空间上是一个导数，于是决定了 M 上的矢量场，参见 3.1.2.2 小节的(ii)。我们把这个与(也看作是 H 产生的) H 相关的矢量场叫作哈密顿矢量场，表示为 X_H。

然而，与泊松结构无关，任何在平滑函数 F 之上的矢量场的作用量，即，$X_H(F)$，也等价于 $L_{X_H}(F)\equiv dF(X_H)$，参见式(40)。于是，对于所有平滑的 F，我们有：

$$L_{X_H}(F)\equiv dF(X_H)\equiv X_H(F)=\{F, H\} \tag{239}$$

X_H 的流的方程被称作哈密顿方程，因为把 H 选择为哈密顿函数。在之前具有 $M=\mathbb{R}^{2n+l}$ 的实例中，我们有：

$$X_H=\Sigma_i^n\left(\frac{\partial H}{\partial p^i}\frac{\partial}{\partial q^i}-\frac{\partial H}{\partial q^i}\frac{\partial}{\partial p^i}\right) \tag{240}$$

并且这个流由常微分方程给出：

$$\frac{dq^i}{dt}=\frac{\partial H}{\partial p^i}\quad\frac{dp^i}{dt}=-\frac{\partial H}{\partial q^i}\quad\frac{dz^j}{dt}=0.\qquad i=1,\ 2,\ \cdots,\ n;\ j=1,\ 2,\ \cdots,\ l \tag{241}$$

z 和任意单独的它们的函数 $F(z)$ 都是可区分的，并且哈密顿矢量场等于零。另一方面，坐标函数 q^i 和 p^i 分别生成了哈密顿矢量场 $-\frac{\partial}{\partial p^i}$和$\frac{\partial}{\partial q^i}$。

关于式(239)的两个更深层的说明：

(1)我们得到的结论认为，函数 H 是可区分的(即具备所有函数的泊松括号)，当且仅当它的哈密顿矢量场 X_H 在所有地方都等于零，并且，因为泊松括号是反对称的，于是，当且仅当沿着所有哈密顿矢量场的流 H 都是常数的时候，可区分的函数 H 才成立；

(2)对于泊松流形而言，这个方程是运动(第一积分)常数理论和诺特定理的开始，正如对应的方程从辛的情形下开始一样，我们将在第 6 部分讨论这些。

关于泊松括号和李括号：

根据式(239)的定义，我们很容易就能建立起在泊松流形和第 3 节的李结

构之间的第一个重要关系。即3.2.2节末尾的结论(2)和式(60)对于泊松流形也是有效的。

也就是说，泊松流形 M 上的标量 F 和 H 的泊松括号的哈密顿矢量场是关于 F 和 H 的哈密顿矢量场的李括号，即 X_F 和 X_H:

$$X_{\{F,H\}} = -[X_F, X_H] = [X_H, X_F] \tag{242}$$

这个证据正是关于式(60)的。

于是，根据泊松括号，哈密顿矢量场构成了泊松流形 M 之上的所有矢量场的李代数 $\mathcal{X}_M$ 的李子代数。在5.3.3节中关于所有泊松流形都是辛流形的不相交并集的证明中，这个结论将是非常重要的。

5.2.3 结构函数

已知某些局域坐标 $\mathbf{x} = x^1, x^2, \cdots, x^m$，在计算任意两个函数的泊松括号时，我们完全能够知道这个坐标的泊松括号。对于任意函数 H: $M \to \mathbb{R}$，令坐标系 $\mathbf{x}$ 中的这个函数的哈密顿矢量场的分量为 $h^i(x)$。于是 $X_H = \sum_i^m h^i(x)\frac{\partial}{\partial x^i}$。于是，对于任意其他函数 F，我们有:

$$\{F, H\} = X_H(F) = \sum_i^m h^i(x)\frac{\partial F}{\partial x^i} \tag{243}$$

我们把 x^i 看作函数 F，我们得到: $\{x^i, H\} = X_H(x^i) = h^i(x)$。于是式(243)变成:

$$\{F, H\} = \sum_i^m \{x^i, H\}\frac{\partial F}{\partial x^i} \tag{244}$$

如果现在我们在式(244)中为 H 放置 x^i，并且为 F 放置 H，我们得到:

$$\{x^i, H\} = -\{H, x^i\} = -X_{x^i}(H) = -\sum_j^m \{x^j, x^i\}\frac{\partial H}{\partial x^j} \tag{245}$$

把式(244)和式(245)结合在一起，我们按照局域坐标的泊松括号得到了任意两个函数的泊松括号的基本公式:

$$\{F, H\} = \sum_i^m \sum_j^m \{x^i, x^j\}\frac{\partial F}{\partial x^i}\frac{\partial H}{\partial x^j} \tag{246}$$

我们把这些基本括号(我们叫作泊松流形的结构函数)

$$J^{ij}(x) := \{x^i, x^j\} \quad i, j = 1, 2, \cdots, m \tag{247}$$

放置到一个 $m\times m$ 反对称函数矩阵，即 $J(x)$ 中，并把这称作 M 的结构矩阵。更加准确地说，M 的结构矩阵与我们的坐标系 x 相关。当然，坐标变换 $x'^i:=x'^i(x^1,\cdots,x^m)$ 之下的 J 的变化式是通过在基本公式(246)中设置 $F:=x'^i$ 和 $H:=x'^j$ 决定的。

于是我们把 H 的(列)梯度矢量写作 ∇H，式(246)变成：

$$\{F,\ H\}=\nabla F\cdot J\nabla H \tag{248}$$

例如，$\mathbb{R}^{2n+1}$ 上的正则括号，即写在 $(q,\ p,\ z)$ 坐标中的式(236)，具有简单形式：

$$J=\begin{pmatrix}0 & I & 0\\ -I & 0 & 0\\ 0 & 0 & 0\end{pmatrix} \tag{249}$$

其中 I 是 $n\times n$ 单位矩阵。

我们能够写出按照 J 与函数 H 相联系的哈密顿矢量场，以及哈密顿方程。因为：

$$\{x^i,\ H\}=\sum_j^m\{x^i,\ x^j\}\frac{\partial H}{\partial x^j} \tag{250}$$

我们得到：

$$X_H=\sum_i^m\left(\sum_j^m J^{ij}(x)\frac{\partial H}{\partial x^j}\frac{\partial}{\partial x^i}\right) \tag{251}$$

或者用矩阵符号表示：$X_H=(J\nabla H)\cdot\partial_x$。同样地，由哈密顿方程：

$$\frac{dx^i}{dt}=\{x^i,\ H\} \tag{252}$$

得到矩阵形式：

$$\frac{dx}{dt}=J(x)\nabla H(x),\ 即\frac{dx^i}{dt}=\sum_j^m J^{ij}(x)\frac{\partial H}{\partial x^j} \tag{253}$$

为了总结如何从哈密顿方程归纳出一般形式，我们分别把式(253)、式(248)和式(249)与式(12)、式(18)和式(3)相比较。

需要注意的是，不是所有 n 维流形上的关于函数的 $m\times m$ 反对称矩阵(或者 $\mathbb{R}^m$ 的开放子集)都是泊松流形的结构矩阵，因为雅可比恒等式约束这个函数。事实上，我们很容易就能说明雅可比恒等式相当于下面的确定 $J^{ij}(x)$ 的 m^3

偏微分方程，这个方程通常是非线性的。通常把$\partial/\partial x^l$写作∂_l：

$$\forall x\in M,\ \sum_{l=1}^{m}(J^{il}\partial_l J^{jk}+J^{kl}\partial_l J^{ij}+J^{jl}\partial_l J^{ki})=0\quad i,\ j,\ k=1,\ \cdots,\ m \tag{254}$$

尤其是，任意恒定的反对称矩阵J定义了泊松结构。

5.2.4 $\mathfrak{g}^*$上的泊松结构

现在，我们能够说明任意m维的李代数$\mathfrak{g}$在任意m维矢量空间V上定义了泊松结构（常称之为李—泊松括号）。我们分两个阶段研究。

（1）我们首先提出定义似乎依赖于在$\mathfrak{g}$和空间V中的基的选择方式，不管在$\mathfrak{g}$中（此处定义选择了结构常数）还是在空间V中。

（2）于是我们将看到，如果将V选择为$\mathfrak{g}^*$，那么实际上，这个定义与基无关的。

这种$\mathfrak{g}^*$上的泊松结构从现在开始将具有核心的重要性。正如马斯登和拉蒂乌写的那样："除了辛流形上的泊松结构之外，$\mathfrak{g}^*$上的李—泊松括号，即李代数的对偶，也许是关于泊松结构的最基础的实例"[1999, 415]。这里我们回到我们关于泊松流形的激烈讨论，尤其是5.1节的理由（2）和（3）：其分别涉及了刚体和约化。确实，我们已经在本小节（5.2.4.1）的末尾的实例中看到特例$\mathfrak{g}^*:=\mathfrak{so}(3)^*$之上的李—泊松括号是如何澄清刚体理论的。并且我们将在7.2和7.3.3小节中看到，对于任意$\mathfrak{g}$而言，约化是如何从余切从T^*G上的正则泊松（即辛）结构$\mathfrak{g}^*$上的李—泊松括号归纳出$\mathfrak{g}^*$上的李—泊松括号的。这将是我们的约化理论，$T^*G/G\cong\mathfrak{g}^*$。

在（2）之后，我们将说明$\mathfrak{g}^*$上的李—泊松括号意味着$\mathfrak{g}^*$上的哈密顿方程可以用ad^*表示：一个我们将在稍后需要的形式，这将是下面的（3）。于是，我们将在5.2.4.1节中讨论实例$\mathfrak{g}^*:=\mathfrak{so}(3)^*$。

（1）任意矢量空间V上的泊松括号

在$\mathfrak{g}$，采用所说的基e_1，e_2，…，e_m，以及结构常量c_{ij}^k，参见式（52）。把空间V当作流形，并且通过决定坐标x^1，x^2，…，x^m的基（例如ϵ_1，ϵ_2，…，ϵ_m）确定它的坐标。我们现在把平滑函数F，H：$V\to\mathbb{R}$之间的泊松括号（本例中，通常称之为李—泊松括号）定义为：

$$\{F,\ H\}:=\sum_{i,j,k=1}^{m}c_{ij}^k x^k\frac{\partial F}{\partial x^i}\frac{\partial H}{\partial x^j} \tag{255}$$

这表现为式(246)的形式，其具有线性结构函数 $J^{ij}(x)=\sum_k^m c_{ij}^k x^k$。我们就能轻易地检查出：对于结构常数，即式(53)而言，反对称和雅可比恒等式说明这些 J^{ij}是反对称的并且遵从它们的雅可比恒等式(254)。于是，式(255)定义了 V 上的泊松括号。

尤其是，相关的哈密顿方程，式(252)和式(253)，表现为如下形式：

$$\frac{dx^i}{dt}=\sum_{j,k=1}^m c_{ij}^k x^k\frac{\partial H}{\partial x^j} \tag{256}$$

(2) $\mathfrak{g}^*$ 上的李—泊松括号

为了给出关于李—泊松括号与基无关的特性描述，我们首先回顾：

(i) F: $V\to\mathbb{R}$在任一点 $x\in V$ 的梯度$\nabla F(x)$是在 V 上的(连续)线性泛函的对偶空间 V^* 中；

(ii)任意有限维矢量空间都是正则地(即与基无关地)与它的双对偶$(V^*)^*\cong V$是同构的。

于是，对于任意 $y\in V$，我们把 V 和 V^* 之间的自然对写作 $<;\ >$，我们有：

$$<\nabla F(x);\ y>:=\lim_{\tau\to 0}\frac{F(x+\tau y)-F(x)}{\tau} \tag{257}$$

现在，我们把李—泊松括号定义中的 V 看作$\mathfrak{g}$使得$\mathfrak{g}^*$以一种与基无关的方式成为了泊松流形。并且令基 ϵ_1，ϵ_2，…，ϵ_m和对于$\mathfrak{g}$的基 e_1，e_2，…，e_m 是对偶的。如果 F: $\mathfrak{g}^*\to\mathbb{R}$是任意的平滑函数，它在任意点 $x\in\mathfrak{g}^*$ 的梯度$\nabla F(x)$就是$(\mathfrak{g}^*)^*\cong\mathfrak{g}$的元素。现在我们证实，式(255)定义的李—泊松括号具有与基无关的表达式：

$$\{F,\ H\}(x)=<x;\ [\nabla F(x),\ \nabla H(x)]>,\ x\in\mathfrak{g}^* \tag{258}$$

其中，$[\ ,]$是李代数$\mathfrak{g}$自身之上的普通李括号。

(3) $\mathfrak{g}^*$之上的哈密顿方程

我们也可给出哈密顿方程式(256)与基无关的表达式，即按照 ad 表示式(258)中的李括号，如式(179)所示。

因而，令 $F\in\mathcal{F}(\mathfrak{g}^*)$为$\mathfrak{g}^*$上的任意平滑标量函数。通过链式规则：

$$\frac{dF}{dt}=\mathrm{D}F(x)\cdot\dot{x}=<\dot{x};\ \nabla F(x)> \tag{259}$$

然而，把式(179)和式(198)应用于式(258)意味着：

$$\{F, H\}(x) = <x; [\nabla F(x), \nabla H(x)]> =$$

$$-<x; ad_{\nabla H(x)}(\nabla F(x))> = <ad^{*}_{\nabla H(x)}(x); \nabla F(x)> \tag{260}$$

因为 F 是任意的，并且这个对(pairing)是非简并的，我们推导出哈密顿方程形式表现为：

$$\frac{dx}{dt} = ad^{*}_{\nabla H(x)}(x) \tag{261}$$

5.2.4.1 实例$\mathfrak{so}(3)$和$\mathfrak{so}(3)^*$

如同关于泊松流形的李代数对偶的实例一样，我们再次考虑我们的标准实例$\mathfrak{so}(3)^*$。因而，我们将兑现我们在5.1节的(2)中的承诺，说明适用于刚体的欧拉方程已经以哈密顿函数的形式出现——在一般的意义上。我们也说明了为什么在本章的第二条题词中，阿诺德提到了三个对偶空间：$\mathbb{R}^{3*}$、$\mathfrak{so}(3)^*$和$T^*(SO(3))_g$，可参见3.4.4节末尾的讨论。

$SO(3)$的李代数$\mathfrak{so}(3)$具有表示围绕着$\mathbb{R}^3$的 x，y 和 z 轴的无穷小转动基 e_1，e_2，e_3。正如我们看到的，我们能够把这些基元素看作是具有[,]的$\mathbb{R}^3$ 中的矢量，作为基本的矢量乘法；或者看作是具有作为矩阵交换子(matrix commutator)的[,]的反对称矩阵；或者看作是 $SO(3)$上具有作为矢量场换位子(即李括号)的[,]的左不变矢量场。

令 ϵ_1，ϵ_2，ϵ_3 为适用于$\mathfrak{so}(3)^*$的对偶基，其中 $x = x^1\epsilon_1 + x^2\epsilon_2 + x^3\epsilon_3$ 这一标准点在其中。如果 F：$\mathbb{R}(3)^* \to \mathbb{R}$，它在 x 的梯度是矢量：

$$\nabla F = \frac{\partial F}{\partial x^1}e_1 + \frac{\partial F}{\partial x^2}e_2 + \frac{\partial F}{\partial x^3}e_3 \quad \in \mathfrak{so}(3) \tag{262}$$

于是，式(258)告诉我们，对于基本矢量乘法而言，如果我们把$\mathfrak{so}(3)$写作具有$\mathbb{R}^3$，$\mathfrak{so}(3)^*$上的李—泊松括号就是：

$$\{F, H\}(x) = x^1\left(\frac{\partial F}{\partial x^3}\frac{\partial H}{\partial x^2} - \frac{\partial F}{\partial x^2}\frac{\partial H}{\partial x^3}\right) + \cdots + x^3\left(\frac{\partial F}{\partial x^2}\frac{\partial H}{\partial x^1} - \frac{\partial F}{\partial x^1}\frac{\partial H}{\partial x^2}\right) \tag{263}$$

$$= -x \cdot (\nabla F \times \nabla H) \tag{264}$$

于是，结构矩阵 $J(x)$是：

$$J(x)=\begin{pmatrix} 0 & -x_3 & x_2 \\ x_3 & 0 & -x_1 \\ -x_2 & x_1 & 0 \end{pmatrix},\ x\in \mathfrak{so}(3)^* \tag{265}$$

因此，对应于哈密顿函数 $H(x)$ 的哈密顿方程是：

$$\frac{dx}{dt}=x\times\nabla H(x) \tag{266}$$

现在我们来考虑哈密顿表示的自由转动刚体的动能：

$$H(x)=\frac{1}{2}\left(\frac{(x^1)^2}{I_1}+\frac{(x^2)^2}{I_2}+\frac{(x^3)^2}{I_2}\right) \tag{267}$$

其中，I_i 是关于三个坐标轴的转动惯量，且 x^i 是相对应的立体角动量的分量。对于这个哈密顿函数，哈密顿方程式(266)变成：

$$\frac{dx^1}{dt}=\frac{I_2-I_3}{I_2I_3}x^2x^3,\ \frac{dx^2}{dt}=\frac{I_3-I_1}{I_3I_1}x^3x^1,\ \frac{dx^3}{dt}=\frac{I_1-I_2}{I_1I_2}x^1x^2 \tag{268}$$

显然，这些都是适用于自由转动刚体的欧拉方程。我不会去论述关于刚体的细节，我只是要说明：

(i) 在关于这个刚体的基础理论中，角动量的大小 L 是守恒的，并且式(268)描述了在以原点为中心半径为 L 的球体上的 x^i 的运动；

(ii) 在5.4节中，我们将返回去把这些球体看成是 $\mathfrak{so}(3)^*$ 上的 $SO(3)$ 的余伴随表示(参见4.5.2节)；

(iii) 我们来总结这个主题，我们认为根据马斯登和拉蒂乌的说法[1999, 11]，可以得到“一个适用于刚体方程的简洁而优美的哈密顿结构”。

5.3 泊松流形的辛叶状结构

我们首先用更加独立于坐标的语言重新表述了5.2节的某些观点，从5.2.3节的结构矩阵 $J(x)$ 的观点开始(5.3.1节)。然后，我们讨论了泊松流形上的正则变换(5.3.2节)。这可以说明，任何泊松流形都被辛叶分解成叶状(5.3.3节)。最后，我们陈述了达布定理的推广，并且再次把 $\mathfrak{so}(3)$ 看作一个例子(5.3.4节)。

5.3.1 泊松结构及其秩

现在，我们从结构矩阵 J，即式(247)，转移到独立于坐标的对象，即泊松

结构(也被看作是余辛结构)，写作B。然而，J 与梯度矢量相乘，如式(248)和(253)所示，B把1形式 dH 映射到了其哈密顿矢量场中，如下所示。

在泊松流形 M 上的每个点 x，都有一个唯一的线性映射 B_x，我们同样把它写作B：

$$\mathrm{B} \equiv \mathrm{B}_x: T_x^* M \to T_x M \tag{269}$$

以至于：

$$\mathrm{B}_x(dH(x)) = X_H(x) \tag{270}$$

根据式(251)式(270)的必要条件意味着，对于每个 $j=1, 2, \cdots, m$，有：

$$\mathrm{B}_x(dx^j) = \sum_i J^{ij}(x) \frac{\partial}{\partial x^i}\Big|_x \tag{271}$$

因为微分 dx^i 生成 $T_x^* M$，则通过线性确定了 B_x。B_x 在任意1形上的作用量 $\alpha = \sum a_j dx^j$ 是：

$$\mathrm{B}_x(\alpha) = \sum_{i,j} J^{ij}(x) a_j \frac{\partial}{\partial x^i}\Big|_x \tag{272}$$

于是 B_x 本质上就是依据 $J(x)$ 的矩阵乘法。这里，我们再次比较式(252)和式(253)。

这里我们回顾了(真实有限维的)矢量空间之间的任意线性映射，即 $B: V \to W^*$，拥有 $V \times W^{**} \cong V \times W$ 之上的相关双线性型 $B^\sharp$，即：

$$B^\sharp(v, w) := < B(v); w > \tag{273}$$

于是，一些作者提出了泊松结构作为双线性型 $\mathrm{B}_x^\sharp$：$T_x^* M \times T_x^* M \to \mathbb{R}$，这通常被称为泊松张量。因而，对于 α，$\beta \in T_x^* M$，式(273)给出：

$$\mathrm{B}_x^\sharp(\alpha, \beta) := < \mathrm{B}(\alpha), \beta > \tag{274}$$

$\mathrm{B}_x^\sharp$ 是非对称的，因为矩阵 $J(x)$ 就是非对称的。于是，如果现在我们令 x 在 M 之上变化，那么，我们能够用传统的张量分析的术语总结出 $\mathrm{B}^\sharp$ 是非对称的反变式的两个张量场。

考虑我们第一个实例，在5.2.1节一开始的式(236)显示的具有“普通括号”的 $M = \mathbb{R}^{2n+l}$。对于任何1形式，有：

$$\alpha = \sum_{i=1}^n (a_i dq^i + b_i dp^i) + \sum_{j=1}^l c_j dz^j \tag{275}$$

我们有：

$$\mathsf{B}(\alpha)=\sum_{i=1}^{n}\left(b_i\frac{\partial}{\partial q^i}-a_i\frac{\partial}{\partial p^i}\right) \tag{276}$$

在这个实例中，B的形式从头到尾都是一样的。尤其是，B的核(kernel)在所有地方都具有相同的维，即可区分的坐标数 l。

我们现在把泊松流形 M 在 x 上的秩定义为它的泊松结构B在 x 上的秩，即 B_x值域的维。这个值域也是 x 处关于 M 的所有哈密顿矢量场的范围：

$$\mathrm{ran}(\mathsf{B}_x):=\{X\in T_xM:\ X=\mathsf{B}_x(\alpha);\ 某些\ \alpha\in T_x^*M\}=\{X_H(x):\ H:\ M\to\mathbb{R}平滑\} \tag{277}$$

于是，x 处 M 的秩也等于B_x 域的维，即维 $\dim(T_x^*M)=\dim(M)$减去核的维 $\dim(\mathsf{B}_x)$。

因为在局域坐标中，B_x 是由结构矩阵 $J(x)$ 通过乘法给出的，于是，在 x 点的 M 的秩就是矩阵 $J(x)$ 的秩(在任何坐标中都一样)。$J(x)$ 是反对称的，这意味着 M 的秩是偶数，参见反对称双线性型的正则形式，式(2)和式(3)。

辛流形 M 当然处处相当于最大的B的秩，即等于 $\dim(M)$。

在这种情况下，B的核是不重要的，任意可区分的函数 H 在 M 上都是常数。当且仅当 $X_H=0$ 时，H 是可区分的，并且如果秩是最大的，那么 $dH=0$，于是 H 是常数。

除此之外，M 上的每个泊松结构和辛形式都是相互决定的。尤其是式(274)的泊松张量$\mathsf{B}^\sharp$是 M 的辛形式 ω 的“逆变近亲(contravariant cousin)”。回顾一下：(i)辛流形的泊松括号及其形式之间的关系，如式(18)，也就是说：

$$\{F,\ H\}=dF(X_H)=\omega(X_F,\ X_H) \tag{278}$$

和 (ii) 对于泊松流形上的哈密顿矢量场的式(239)，也就是：

$$X_H(F)=\{F,\ H\} \tag{279}$$

如果我们开始于式(274)和式(270)，那么应用这些方程可得：

$$\mathsf{B}^\sharp(dH,\ dF):=\langle B(dH),\ dF\rangle=dF(X_H)=X_H(F)=\{F,\ H\}=\omega(X_F,\ X_H) \tag{280}$$

我们也看到了一些实例，在这些例子中，泊松结构B是非最大的秩。

(i) 在我们起初的“正则”实例中，式(236)中 $M=\mathbb{R}^{2n+l}$上的泊松括号在任何情形下都具有秩 $2n$。

（ii） 在$\mathfrak{so}(3)^*$上的李—泊松结构中，其秩随着流形变化：它在各处都是2，除了在原点$x=0$处它是0，参见式(265)中矩阵J的秩。

5.3.2 泊松映射

在研究泊松流形的初期，我们已经认为标量函数H：$M\to\mathbb{R}$定义了类似泊松括号类型的运动方程，对于所有其他函数F：$M\to\mathbb{R}$，我们把H看作“哈密顿函数”：

$$\dot{F}=\{F,\ H\} \tag{281}$$

参见5.2.2小节，尤其是关于式(239)的说明。现在，我们对泊松流形的一些相关概念和结论进行了推广。

我们认为，泊松流形$(M_1,\ \{,\}_1)$和$(M_2,\ \{,\}_2)$之间的平滑映射f：$M_1\to M_2$是泊松的或者正则的，当且仅当它保留了泊松括号。更加精确地说，我们首先需要函数的回撤的概念，参见3.1.2.1小节。在这个情境中，函数F：$M_2\to\mathbb{R}$的回撤f^*由下式给出：

$$f^*F:=F\circ f;\ 即f^*F:\ x\in M_1\longmapsto F(f(x))\in\mathbb{R} \tag{282}$$

于是，我们认为f：$M_1\to M_2$是泊松的，当且仅当对于所有平滑函数F，G：$M_2\to\mathbb{R}(F,\ G\in\mathcal{F}(M_2))$，有：

$$f^*\{F,\ G\}_2=\{f^*F,\ f^*G\}_1 \tag{283}$$

根据式(282)的定义，lhs$\equiv\{F,\ G\}_2\circ f$，并且rhs$\equiv\{F\circ f,\ G\circ g\}_1$。

我们注意到这个特例，其中$M_1=M_2=:M$，且M是辛的，即泊松括号是最大秩的，因此定义了M上的辛形式，同式(280)中一样。在这种情况下，我们返回到2.1.3节关于哈密顿力学的一般构造中在保留泊松括号和保留辛形式之间等价性。也就是说，辛流形M上的映射f：$M\to M$是泊松的当且仅当它是辛的。

除此之外，对于辛流形，我们已经拥有了关于泊松映射或者辛映射想法的无穷小版本，即局部的哈密顿矢量场的想法，参见2.1.3节。同样地适用于泊松流形，在6.1.1节，我们将需要对应的泊松映射的无穷小版本。

我们能够说明(尤其是使用雅可比恒等式)哈密顿矢量场的流是泊松的，当然，这里$(M_1,\ \{,\}_1)=(M_2,\ \{,\}_2)$，也就是说，如果$\phi_\tau$上$X_H$的流，即$\phi_\tau=$

$\exp(\tau X_H)$，那么：

$$\phi_\tau^*\{F, G\} = \{\phi_\tau^* F, \phi_\tau^* G\} \text{ 即 } \{F, G\} \circ \phi_\tau = \{F \circ \phi_\tau, G \circ \phi_\tau\} \tag{284}$$

同样地，我们能够轻易地说明等价的命题，即沿着哈密顿矢量场的流，泊松张量$\mathrm{B}^\sharp$的李导数等于零。也就是说，对于任何平滑函数 H: $M \to \mathbb{R}$，我们有：

$$\mathcal{L}_{X_H} \mathrm{B}^\sharp = 0 \tag{285}$$

因为保留泊松括号尤其意味着保留其秩，我们从式(284)推断出：

如果 X_H 是泊松流形 M 上的哈密顿矢量场，那么对于任何 $\tau \in \mathbb{R}$和 $x \in M$，M 在 $\exp(\tau X_H)(x)$的秩与在 x 的秩相等。换言之，在用于弗罗宾尼斯定理的一般形式的意义上，哈密顿矢量场是秩不变的(3.3.2 节)。

这个结论对于适用于泊松流形的叶状结构定理是很重要的。

我们同样需要这个结论(也已经给出了)，即泊松映射把哈密顿流推进到哈密顿流。更精确地说，令f: $M_1 \to M_2$ 为泊松映射，于是在每个 $x \in M_1$ 处，我们都在切空间上拥有导数映射，Tf: $(TM_1)_x \to (TM_2)_{f(x)}$。并且，令 H: $M_2 \to \mathbb{R}$为平滑函数。如果 ϕ_τ 是 X_H 的流且 ψ_τ 是 M_1 上 $X_{H\circ f}$的流，那么：

$$\phi_\tau \circ f = f \circ \psi_\tau, \text{ 且 } Tf \circ X_{H\circ f} = X_H \circ f \tag{286}$$

尤其是，这个正方形对易如下：

$$\begin{tikzcd} M_1 \arrow[r, "f"] & M_2 \\ M_1 \arrow[u, "\psi_\tau"] \arrow[r, "f"] & M_2 \arrow[u, "\phi_\tau"'] \end{tikzcd} \tag{287}$$

5.3.3 泊松子流形：叶状结构定理

为了说明泊松流形的叶状结构定理，我们需要泊松浸入的概念，这个概念把我们带到与泊松子流形密切相关的概念。实际上，这些想法把泊松映射的想法和 3.3.1 节的(2)中的内射浸入的想法结合在一起。回顾之前的一个讨论，即对于内射浸入，f: $N \to M$，其值域$f(N)$不一定是 M 的子流形，然而，$f(N)$仍会被叫作 M 的"内射地浸入的子流形"。但是如 3.3.2 节所示，很多论述忽视了这一点，它们实际上假定了内射浸入f也是嵌入的，即在 N 和$f(N)$之间的同胚映射，于是$f(N)$实际上是 M 的子流形且f是微分同胚映射。

如果任何定义在包含$f(N)$的 M 的开子集的哈密顿矢量场都在 $y \in N$ 的f的

导数映射的值域内，即对于 $y \in N$，在所有点 $f(y)$ 处的 $\mathrm{ran}(T_y f)$ 内，我们就把泊松流形 M 的内射浸入 $f: N \to M$ 叫作泊松浸入。

成为一个泊松浸入等同于下列条件，但不是技术条件。

泊松内射的特征 当且仅当如果 $F, G: V \subset N \to \mathbb{R}$，其中 V 在 N 中是开区间的，并且如果 $\bar{F}, \bar{G}: U \to \mathbb{R}$是 $F \circ f^{-1}$，$G \circ f^{-1}: f(V) \to \mathbb{R}$在 M 里延伸到 $f(V)$ 的一个开邻区间的伸延集，那么，$\{\bar{F}, \bar{G}\}\mid_{f(V)}$是定义明确的，并且与这个伸延集无关。我们就说，具有 M 泊松流形的内射浸入 $f: N \to M$ 是一个泊松浸入。

这个等价说法的要点是，它确保如果 $f: N \to M$ 是泊松浸入，那么 N 拥有泊松结构，并且 $f: N \to M$ 是泊松映射。通过证明这个等价说法去探讨一下这些是如何发生的是值得的。

证明：令 $f: N \to M$ 为泊松浸入，并且令 $F, G: V \subset N \to \mathbb{R}$和 $\bar{F}, \bar{G}: U \supset f(V) \to \mathbb{R}$为 $F \circ f^{-1}$的伸延集 $G \circ f^{-1}: f(V) \to \mathbb{R}$。于是，对于 $y \in V$，存在一个唯一的矢量 $v \in TN_y$，使得：

$$X_{\bar{G}}(f(y)) = (T_y f)(v) \tag{288}$$

于是，用式(239)在 $f(y)$ 求得 $\bar{F}$ 和 $\bar{G}$ 的泊松括号的值产生：

$$\{\bar{F}, \bar{G}\}(f(y)) = d\bar{F}(f(y)) \cdot X_{\bar{G}}(f(y)) = d\bar{F}(f(y)) \cdot (T_y f)(v) = d(\bar{F} \circ f)(y) \cdot v \equiv dF(y) \cdot v \tag{289}$$

因此 $\{\bar{F}, \bar{G}\}(f(y))$ 与 $F \circ f^{-1}$的伸延集 $\bar{F}$ 无关。因为泊松括号是非对称的，于是它也与 $G \circ f^{-1}$的伸延集 $\bar{G}$ 无关。于是，我们能够通过将任意的 y 定义在开区间 $V \subset N$ 里来定义 N 上的泊松结构：

$$\{F, G\}_N(y) := \{\bar{F}, \bar{G}\}_M(f(y)) \tag{290}$$

这使得 $f: N \to M$ 成为泊松映射，因为对于 M 上的任意 $\bar{F}, \bar{G}$ 以及任意 $y \in N$，我们有：

$$[f^*\{\bar{F}, \bar{G}\}_M](y) \equiv [\{\bar{F}, \bar{G}\}_M \circ f](y) = \{F, G\}_N(y) \equiv \{f^*\bar{F}, f^*\bar{G}\}_N(y); \tag{291}$$

其中，中间的等式使用了式(290)。

对于逆蕴含(converse implication)，我们假设式(289)成立，并且令 $H: U \to \mathbb{R}$为一个定义在与 $f(N)$ 相交的 M 的开子集 U 之上的哈密顿函数。那么，正

如我们刚刚看到的，N 是一个泊松流形，并且 f：$N \to M$ 是一个泊松映射。因为 f 是泊松的，它把 $X_{H\circ f}$ 推进到 X_H。也就是说，式(286)意味着，如果 $y \in N$ 满足 $f(y) \in U$，那么：

$$X_H(f(y)) = (T_y f)(X_{H\circ f}(y)) \tag{292}$$

于是，$X_H(f(y))$ 在 $T_y f$ 的范围里，f：$N \to M$ 是一个泊松浸入。

现在，我们假设这个包含作用 id：$N \to M$ 是一个泊松浸入。于是，我们把 N 叫作关于 M 的泊松子流形。根据我们在 3.3.1 节的(2)中的警示，我们需要强调的是：N 不需要是 M 的子流形；然而它仍然被称作 M 的"内射地浸入的子流形"。

我们从泊松浸入的定义推断出，任何哈密顿矢量场都必须与泊松子流形正切。换言之，把 M 上的哈密顿矢量场的系统写作 $\mathcal{X}$，并把它们在 $x \in M$ 时的值写作 $\mathcal{X}|_x$，我们有：如果 N 是 M 的泊松子流形，并且 $x \in N$，那么 $\mathcal{X}|_x \subset TN_x$。

对于 M 在其中是辛流形的特例而言，我们有 $\mathcal{X}|_x = T_x M$，并且 M 的唯一泊松子流形就是它的开放集。

最后，我们在泊松流形 M 上定义了如下等价关系。如果在 M 中存在分段的平滑曲线把两个点 x_1，$x_2 \in M$ 联结在一起，它们的每一段都是局部定义的哈密顿矢量场的积分曲线，那么这两个点在相同的辛叶上。这个等价关系的等价类别是辛叶。

我们现在能够阐明并且证明泊松流形是叶状的。

5.3.3.1 泊松流形的叶状结构定理结论

泊松流形 M 是它的辛叶的不相交并集。每个辛叶都是一个内射地浸入的泊松子流形，并且在这个叶上引入的泊松结构是辛的。穿过点 x 的叶(例如 N_x)拥有的维等于 x 上的泊松结构的秩，x 上的叶的切空间等于：

$$\begin{aligned} TN_x = \mathrm{ran}(\mathsf{B}_x) &:= \{X \in T_x M\text{：} X = \mathsf{B}_x(\alpha)\text{，某些 } \alpha \in T_x^* M\} \\ &= \{X_H(x)\text{：} H \in \mathcal{F}(U)\text{，} U \text{ 在 } M \text{ 内是开区间的}\} \end{aligned} \tag{293}$$

证明：我们把弗罗宾尼斯定理的一般形式(3.3.2 节)应用于 M 上的哈密顿矢量场的系统 $\mathcal{X}$。从式(242)(5.2.2 节)中，我们知道了 $\mathcal{X}$ 是对合的，并且从上述式(284)知道了它是秩不变的。于是，根据弗罗宾尼斯定理，$\mathcal{X}$ 是可积的。式(293)的 rhs 根据定理给出了积分子流形。

我们也可说明：

(i) 我们可以把 F 和 G 限制在辛叶 N_x 上，以及求得叶 N_x 上的辛形式定义的泊松括号的值，从而求在 $x \in M$ 时 F，G：$M \to \mathbb{R}$的泊松括号的值，即在式(18)中定义的泊松括号。

(ii) 在 M 的任意辛叶 N_x 上，可区分函数都是常量。

我们以两个备注结束。第一个是数学方面的警示，第二个涉及了物理的解释。

(1) 回想我们认为辛叶不必是子流形的警示。这也意味着，作为常量的所有可区分的函数并不意味着泊松结构是非简并的。实际上，我们能够轻易地构造出一个例子。其中，辛叶不是流形，所有可区分的函数都是常量，泊松结构是简并的。也就是说，我们适应了先前在 3.4.3 节提到的例子，即在微风轻轻环绕的圆环面 T^2 上的流。关于这个例子的更多详情，参见[Arnold，1973，pp. 160—167；1989，pp. 72—74]或者[Butterfield，2004a，Section 2.1.3.B]。就如何应用它，参见[Marsden and Ratiu，1999，347]。

(2) 正如我们已经看到的，任何哈密顿矢量场 X_H 的积分曲线都被限制在其中的一个辛叶中。于是，如果我们只对穿过点 $x \in M$ 的单一解的行为感兴趣，我们就能够把我们的注意力限制在穿过 x 的辛叶 N_x 上，因为解决方案总是停留在 N_x 里。但是正如我们在 5.1 节中强调的，至少有三个很好的理由说明不应该忽略更为一般的泊松结构！

5.3.4 达布定理

在 2.1.1 节的末尾，我们提到了达布定理，它说明了任意辛流形“在局部上都看起来像”一个余切丛。鉴于 5.2.1 节一开始的内容，对于泊松流形的推广说明了任意泊松流形“在局部上都看起来像”我们关于$\mathbb{R}^m$($m = 2n + l$)的正则实例。

令 M 为 m 维的泊松流形，并且令 $x \in M$ 为具有开区间邻域 $U \subset M$ 的点，相交它的秩是常数 $2n \leqslant m$。于是，定义 $l := m - 2n$，我们有一个可能更小的关于 x 的邻域 $U' \subset U$，在其上存在局部坐标$(q, p, z) = (q^1, \cdots, q^n, p^1, \cdots, p^n, z^1, \cdots, z^l)$。对于这些局部坐标，泊松括号采取了形式：

$$\{F, H\} := \sum_i^n \left(\frac{\partial F}{\partial q^i} \frac{\partial H}{\partial p^i} - \frac{\partial F}{\partial p^i} \frac{\partial H}{\partial q^i} \right) \tag{294}$$

于是，坐标函数的泊松括号采用了式(237)和式(238)给出的目前非常熟悉的形式。M 的辛叶相交于由可区分的坐标 z 的常数值给定的片段 $\{z^1 = c_1, \cdots, z^l = c_l\}$ 中的坐标图。

我们将不会给出这个证明。只要说明如下的观点就够了。

(i) 像辛流形的达布定理一样，它依据"半秩"n 上的归纳继续，并且它首先把任意函数 F 看作是"动量"p^1 并且构造出使 $\{q^1, p^1\} = 1$ 的正则共轭坐标 q^1。

(ii) 归纳的步骤给出了弗罗宾尼斯定理的一个版本，其中的内容是，贯穿 U 的秩 $2n$ 是常数，这确保了一个坐标系，在这个坐标系中，剩下的 l 坐标的常数值定义的片段给出了 $2n$ 维的积分流形。于是，泊松结构确保了这些剩余的坐标是可区分的。

5.3.4.1 例子：再次讨论$\mathfrak{so}(3)^*$

我们用$\mathfrak{so}(3)^*$阐明了(1)叶状结构定理以及(2)达布定理，我们在5.2.4.1小节中描述了它们的李—泊松结构。

(1) 在 $x \in \mathfrak{so}(3)^*$ 时，局部哈密顿矢量场的值的子空间 $\mathcal{X}|_x := \{X_{H(x)}: H \in \mathcal{F}(U), U$ 在 M 内是开区间的$\}$ 是通过用 $e_1 := y\partial_z - z\partial_y$ 表示围绕 x 轴的无穷小转动生成的，参见式(48)、式(107)、式(154)；$e_2 := z\partial_x - x\partial_z$ 表示围绕 y 轴的转动；$e_3 := x\partial_y - y\partial_x$ 表示围绕 z 轴的转动。如果 $x \neq 0$，那么，这些矢量生成了关于 $T\mathfrak{so}(3)_x^*$ 的2维子空间：即以原点为中心半径为 $|x|$ 的球体 $S_{|x|}$ 的切平面。于是，叶状结构定理意味着$\mathfrak{so}(3)^*$的辛叶是这些球面以及原点。

我们能够把 F 和 G 延伸到 $S_{|x|}$ 的邻域从而计算 $F, G: S_{|x|} \to \mathbb{R}$的泊松括号，参见式(290)。也就是说，我们能够考虑延伸 $\bar{F}, \bar{G}: U \supset S_{|x|} \to \mathbb{R}$，并且计算$\mathfrak{so}(3)^*$中的泊松括号，我们已经在式(264)中计算过它们的泊松结构。

通过采用球极坐标 $r = |x|$，即 $x^1 = r\cos\theta\sin\phi$，$x^2 = r\sin\theta\sin\phi$，$x^3 = r\cos\phi$，我们能够仅用 $\bar{F}(r, \theta, \phi) := F(\theta, \phi)$ 及 $\bar{G}(r, \theta, \phi) := G(\theta, \phi)$ 定义 $\bar{F}, \bar{G}$，

于是，关于球面角 θ，ϕ 的偏导数是相等的，即 $\bar{F}_\theta = F_\theta$，$\bar{F}_\phi = F_\phi$，$\bar{G}_\theta = G_\theta$，$\bar{G}_\phi = G_\phi$。

除此之外，式(246)意味着我们仅需要在$\mathfrak{so}(3)^*$中计算出球面角 θ 和 ϕ 的泊松括号。于是式(264)给出了：

$$\{\theta, \phi\} = -x \cdot (\nabla\theta \times \nabla\phi) = \frac{-1}{r\sin\phi} \tag{295}$$

并且式(290)和式(246)给出了：

$$\{F, G\} = \{\bar{F}, \bar{G}\} = \frac{-1}{r\sin\phi}(F_\theta G_\phi - F_\phi G_\theta) \tag{296}$$

(2) $z := x^3$ 定义了产生关于 $z \equiv x^3$ 轴逆时针转动的哈密顿矢量场 $X_z = x^2\partial_{x^1} - x^1\partial_{x^2}$。于是，除去原点，极角 $\theta := \arctan(x^2/x^1)$ 拥有一个泊松括号，这个极角和 z 的这个泊松括号等于：$\{\theta, z\} = X_z(\theta) = -1$。按照坐标 z，θ 和 $r := |x|$，表示为 F，H：$\mathfrak{so}(3)^* \to \mathbb{R}$，我们发现李—泊松括号是：$\{F, H\} = F_z H_\theta - F_\theta H_z$。于是，$(z, \theta, r)$是正则坐标。

5.4 余伴随表示的辛结构

5.2.4 节描述了李群 G 的有限维的李代数的对偶$\mathfrak{g}^*$如何具有泊松流形的结构。在这种情况下，我们在前面小节构建的叶状结构就拥有了极其简洁的解释。也就是说：叶是$\mathfrak{g}^*$上的关于 G 的余伴随表示的轨道。

这种余伴随表示中的辛结构总结了来自 4.5 节(尤其是 4.5.2 节)和 5.2.4 节以及 5.3 节的主题。尤其是结合了李括号在$\mathfrak{g}$的两个性质，正如我们已经看到的，即：

(i) $\mathfrak{g}$中的李括号给出了关于伴随作用量的无穷小算子，参见式(179)；

(ii) $\mathfrak{g}$中的李括号(以与基础无关的方式)定义了$\mathfrak{g}^*$上的李—泊松括号，因而，使得$\mathfrak{g}^*$成为一个泊松流形，参见式(255)中的定义，这个定义通过式(258)被看作是与基础无关的。

事实上，我们有很多关于余伴随表示结构的大量建设性结论和实例，我们只做一些浅显的研究——正如我们在其他小节中所做的一样！在简化假设的条件下，我们将只给出一个关于主要结论的证明，然后给出一些关于其他结论的

评论。

这个结论是：

这个余伴随表示的轨道是$\mathfrak{g}^*$的叶。

令G为一个李群，它在$\mathfrak{g}^*$上的余伴随表示为Ad^*。也就是说，结合式(193)，我们有：

$$Ad^*: G \to \mathrm{End}(\mathfrak{g}^*),\ Ad^*_{g^{-1}} = (T_e(R_g \circ L_{g^{-1}}))^* \tag{297}$$

这个表示的轨道是$\mathfrak{g}^*$的辛叶，它被认为具有它的固有泊松结构，即式(258)的李—泊松括号。

证明：我们将通过简化假设来证明，这个简化假设是：$\mathfrak{g}^*$上的G的余伴随作用量是正确的。我们从正确作用量的定义中回想到式(147)，对于任意紧李群，如$SO(3)$，这个条件自动被满足。于是，在4.4小节的结尾，我们从结论(3)和式(160)了解到，这些意味着所有$\alpha \in \mathfrak{g}^*$的余伴随轨道$\mathrm{Orb}(\alpha)$是关于$\mathfrak{g}^*$的封闭子流形，并且在点$\beta \in \mathrm{Orb}(\alpha)$的$\mathrm{Orb}(\alpha)$的正切空间是：

$$T\mathrm{Orb}(\alpha)_\beta = \{\xi_{\mathfrak{g}^*}(\beta): \xi \in \mathfrak{g}\} \tag{298}$$

我们马上就会看到这些假设为何意味着$\mathfrak{g}^*$的辛叶是子流形。①

目前我们的论辩如下。就$\xi \in \mathfrak{g}$而言，考虑$\mathfrak{g}^*$上的标量函数，K_ξ：$\alpha \in \mathfrak{g}^* \longmapsto K_\xi(\alpha) := \langle \alpha, \xi \rangle \in \mathbb{R}$以及其哈密顿矢量场$X_{K_\xi}$。在每个$\alpha \in \mathfrak{g}^*$，被看作是$(T^*\mathfrak{g}^*)_\alpha \cong \mathfrak{g}$的一个元素梯度$\nabla K_\xi(\alpha) \equiv dK_\xi(\alpha)$就是$\xi$本身。现在，我们将按以下顺序计算适用于任意$F$：$\mathfrak{g}^* \to \mathbb{R}$以及任意$\alpha \in \mathfrak{g}^*$的$X_{K_\xi}(F)(\alpha)$：

(i)　在$\mathfrak{g}^*$上的李—泊松括号的本质定义，即式(258)；

(ii)　这样一个事实，即伴随作用量的无穷小算子是$\mathfrak{g}$中的李括号，即式(179)；

(iii)　这样一个事实，即余伴随作用量Ad^*的导数ad^*是ad_ξ的伴随，即式(198)。

① 为了确认我们的条件确实是简化的——即一般而言，$\mathfrak{g}^*$中的余伴随轨迹不是子流形——考虑马斯登和拉蒂乌[1999，14.1.(f)，p.449]和其中的例子，摘自基里洛夫(Kirillov)[1976，293]。

因此，对于所有 F：$\mathfrak{g}^* \to \mathbb{R}$以及 $\alpha \in \mathfrak{g}^*$，我们得到：

$$X_{K_\xi}(F)(\alpha) \equiv \{F, K_\xi\}(\alpha) = <\alpha; [\nabla F(\alpha), \nabla K_\xi(\alpha)]> \tag{299}$$

$$= <\alpha; [\nabla F(\alpha), \xi]> = -<\alpha; [\xi, \nabla F(\alpha)]> \tag{300}$$

$$= -<\alpha; ad_\xi(\nabla F(\alpha))> \tag{301}$$

$$= <ad_\xi^*(\alpha): \nabla(\alpha)> \tag{302}$$

但是，另一方面，矢量场 X_{K_ξ}是由其在所有 $\alpha \in \mathfrak{g}^*$ 时所有此类函数 F 上的作用量唯一确定的：

$$X_{K_\xi}(F)(\alpha) \equiv <X_{K_\xi}(\alpha); \nabla F(\alpha)> \tag{303}$$

于是，我们在每个 $\alpha \in \mathfrak{g}^*$得到：

$$X_{K_\xi} = ad_\xi^* \tag{304}$$

然而，哈密顿矢量场在 α 的值的子空间$\mathcal{X}|_\alpha$ 是由 $X_{K_\xi}(\alpha)$生成的，并由 ξ 通过$\mathfrak{g}$发生变化。并且，正如 ξ 通过$\mathfrak{g}$发生变化一样，$ad_\xi^*(\alpha)$是经由的 G 的余伴随轨道 Orb(α)的正切空间 TOrb$(\alpha)_\alpha$。于是：

$$\mathcal{X}|_\alpha = T\mathrm{Orb}(\alpha)_\alpha \tag{305}$$

因而哈密顿矢量场的系统$\mathcal{X}$的积分子流形是余伴随轨道，根据 5.3.3.1 小节的叶状结构定理，这个积分子流形是$\mathfrak{g}^*$的辛叶。

根据我们的标准实例$\mathfrak{so}(3)^*$对这个定理的说明，参见我们先前关于它的讨论：在 4.5.2 小节我们讨论了它的余伴随结构；在 5.2.4.1 小节我们讨论了它的李—泊松结构；在 5.3.4.1 小节我们讨论了它的辛叶结构。

我们用对另外两个结论的陈述来结束这个小节。稍后，将不需要讨论它们，但是，它们确实启示了余伴随轨道的理论是多么的迷人。

(1)对于任意 $g \in G$，余伴随映射 Ad_g^*：$\mathfrak{g}^* \to \mathfrak{g}^*$ 是维持了$\mathfrak{g}^*$ 的辛叶的泊松映射。

(2)与我们刚刚证明的定理相似的定理是，$\mathfrak{g}$上的李括号定义了辛形式，即通过其在$\mathfrak{g}^*$上的李—泊松括号的定义，参见式(258)，即一个非简并的在每个余伴随轨道上的闭合 2 形式，根据下式给定：

$$\omega(\alpha)(ad_\xi^*(\alpha), ad_\eta^*(\alpha)) := <\alpha; [\xi, \eta]_\mathfrak{g}>, \quad \forall \alpha \in \mathfrak{g}^*, \ \forall \xi, \eta \in \mathfrak{g} \tag{306}$$

马斯登和拉蒂乌［1999，Thm 14.3.1，pp.453—456］详细证明了这个定理（没有使用我们简化的假设，G 的作用量是正确的）；阿诺德［1989，321，pp.376—377，457］相对简单地证明了这个定理；［Abraham and Marsden，1978，pp.302—303］相当标新立异地证明了这个定理（虽然没有使用泊松流形的概念）。

5.5　泊松流形的商

现在，我们用关于商化泊松流形上的李群作用量的最简单的一般定理来结束第5节，以便得到一个本身就是泊松流形的商空间（轨道集合）。于是这个定理结合了来自第4节的定理——尤其是，来自4.3.0.2节的思想，对于一个自由的且正确的群作用量而言，其轨道和商空间都是流形，具有来自5.2节的泊松流形的信息（不需要5.3和5.4节中的资料）。这个定理在第7节中将会是很重要的。

泊松约化定理：假设李群 G 作用于泊松流形 M 上是这样一个方法：Φ_g：$M\to M$ 是一个泊松映射。并假设商空间 M/G 是一个流形并且投影 π：$M\to M/G$ 是一个平滑浸没（smooth submersion）（因为 G 在 M 上的作用量是自由的且是正确的，参见4.3.0.3小节）。于是，在 M/G 上存在一个唯一的泊松结构使得 π 是一个泊松映射。M/G 上的泊松括号被叫作约化的泊松括号。

证明：我们首先假设 M/G 是一个泊松流形并且 π 是一个泊松映射，并且表现出唯一性。我们首先注意到，对于任意 f：$M/G\to\mathbb{R}$，函数 $\bar{f}:=f\circ\pi$：$M\to\mathbb{R}$ 显然是 M 上唯一的 G 不变函数，这个 G 不变函数通过 π 投射到 f。也就是说，如果 $[x]\equiv\mathrm{Orb}(x)\equiv G\cdot x$ 是 $x\in M$ 的轨道，于是，$\bar{f}$ 为轨道 $[x]$ 的所有元素分配了相同的值 $f([x])$。除此之外，根据回撤，参见式（282），$\bar{f}=\pi^{*}f$。

于是，π 是泊松即式（283）的条件是：对于任意两个平滑的标量 f，h：$M/G\to\mathbb{R}$，我们有一个 M 上的平滑标量的方程：

$$\{f,\ h\}_{M/G}\circ\pi=\{f\circ\pi,\ h\circ\pi\}_{M}=\{\bar{f},\ \bar{h}\}_{M} \tag{307}$$

其中，下标表明泊松括号被定义在哪个空间。因为 π 是满射的，式（307）唯一地决定了值 $\{f,\ h\}_{M/G}$。

但是式(307)同样把$\{f, h\}_{M/G}$定义为泊松括号：

①Φ_g是泊松的事实以及$\bar{f}$和$\bar{h}$是轨道上的常量，这意味着：

$$\{\bar{f}, \bar{h}\}(g \cdot x) = (\{\bar{f}, \bar{h}\} \circ \Phi_g)(x) = \{\bar{f} \circ \Phi_g, \bar{h} \circ \Phi_g\}(x) = \{\bar{f}, \bar{h}\}(x) \tag{308}$$

也就是说，$\{\bar{f}, \bar{h}\}$同样是轨道上的常量，于是唯一地定义了$\{f, h\}$。

②通过检查从M上的泊松结构$\{,\}_M$得到的诸如雅可比恒等式这样的必要特性，我们说明了$\{f, h\}$是M/G上的泊松结构。

这个定理是后面将得到的资料的“雏形”。我们用两个简单的备注来说明一下，它们涉及之后的两个小节。

(1)其他定理

这个定理是许多通过商化旧的泊松流形和辛流形而产生新的泊松流形和辛流形的定理之一。特别是，正如我们在第7节的详细描述中所看到的一样，这个定理是$M = T^* G$时的情况下的例证(因而，这里M是辛的，既然它是余切丛)，并且G通过左平移作用于它本身，通过余切提升作用于$T^* G$。在这种情况下，我们将有$M/G \cong \mathfrak{g}^*$，并且刚刚通过式(307)定义的约化的泊松括号将成为我们已经在5.2.4节中见到过的李—泊松括号。

(2) 动力学约化

使用这个定理，我们已经能够补充一些涉及约化动力学的东西，我们仅在2.3节和5.1节的导引性讨论中就能看到。我们提出两个基本观点，如下所示。

(A) 如果H是M上的G不变哈密顿函数，那么它通过$H = h \circ \pi$定义了M/G上对应的函数h。泊松映射把哈密顿流推向哈密顿流的事实意味着，因为π是泊松的，所以π把M上的X_H转化为M/G上的X_h。也就是说：

$$T\pi \circ X_H = X_h \circ \pi \tag{309}$$

即X_H和X_h是π相关的。对应地，我们说M上的哈密顿系统X_H约化为M/G上的哈密顿系统X_h。

(B) 我们在6.2节中将会看到H的G不变式与一组守恒量相联系（运动常量，第一积分)，也就是对于每个$\xi \in \mathfrak{g}$，给出一个运动常量$J(\xi)$：$M \to \mathbb{R}$。这里，J是守恒的意味着$\{J, H\} = 0$；正如在我们用一般哈密顿力学的(2.1.3

节)关于诺特定理的讨论中。除此之外，如果 J 也是 G 不变量，那么，根据 X_h，M/G 之上对应的函数 j 是守恒的，因为：

$$(j,\ h)\circ\pi=\{J,\ H\}=0 \text{ 意味着 } \{j,\ h\}=0 \tag{310}$$

6. 重温对称性和守恒性：动量映射

现在，我们在泊松流形的情况下给出几个关于守恒量和对称性的主题(以及诺特定理)。在这些主题的中心矗立着的是泊松流形上关于李群作用量的动量映射的思想，我们在6.1小节引入这些。这是关于守恒量(如欧几里得群的线动量和角动量)的近代几何学的推广，由此得名。形式上，它将是一个从泊松流形 M 到对称群 G 的李代数的对偶$\mathfrak{g}^*$的映射 $\mathbf{J}$。因为，它的值存在于矢量空间，具有一些分量。于是，我们关于守恒量的描述将不再是“单向度(一维)的”，即关注于态空间中的一个单一矢量场，如2.1.3节中所示。映射 $\mathbf{J}$ 将与从$\mathfrak{g}$到$\mathcal{F}(M)$的线性映射 J 有关，即流形 M 上的标量函数。也就是说，对于每个 $\xi\in\mathfrak{g}$，如果哈密顿函数 H 是无穷小算子 ξ_M 之下的不变量，即如果 $\xi_M(H)=0$，那么 $J(\xi)$将会是一个守恒量。

诺特定理的泊松流形版本(6.2节)和线动量和角动量的常见实例(6.3节)分别表现和阐述了动量映射的守恒，于是，我们讨论了动量映射的等变性，以及关于 G 在$\mathfrak{g}^*$上的余伴随表示，见6.4节。最后在6.5节，我们讨论了涉及余切丛的动量映射至关重要的特例，这同样是举例说明的。

6.1　正则作用和动量映射

我们首先把泊松映射的定义(5.3.2节)应用于群作用(6.1.1节)。这将引入动量映射的概念(6.1.2节)。

6.1.1　正则作用和无穷小算子

令 G 为通过平滑左作用量 Φ：$G\times M\to M$ 作用于泊松流形 M 的李群，于是，像往常一样，我们写作 Φ_g：$x\in M\longmapsto\Phi_g(x)\coloneqq g\cdot x\in M$。正如泊松映射的定义，参见式(283)，如果对于任意 F_1，$F_2\in\mathcal{F}(M)$以及任意 $g\in G$，都有：

$$\Phi_g^*\{F_1, F_2\} = \{\Phi_g^* F_1, \Phi_g^* F_2\} \tag{311}$$

于是我们就说这个作用量是正则的。如果是辛的并具有辛形式 ω，那么，当且仅当它是辛的，这个作用量是正则的，即 $\Phi_g^*\omega = \omega$ 适用于所有 $g \in G$。

我们对这个概念的无穷小版本特别感兴趣，并且对作用量的无穷小算子也很感兴趣。我们从式(152)回想起，与李代数元素 $\xi \in \mathfrak{g}$—致的作用量的无穷小算子是通过在方向 ξ 的单位元素上求关于 g 的作用量的微分从而获得的 M 上的矢量场 ξ_M：

$$\xi_M(x) = \frac{d}{d\tau}[\exp(\tau\xi) \cdot x]\mid_{\tau=0} \tag{312}$$

于是我们在方向 ξ 上对式(311)求关于 g 的微分，从而给出：

$$\xi_M(\{F_1, F_2\}) = \{\xi_M(F_1), F_2\} + \{F_1, \xi_M(F_2)\} \tag{313}$$

这个矢量场 ξ_M 被称为无穷小泊松自同构。

旁注：我们稍后将会看到式(311)中以及相应在式(313)和之后的式(315)中涵盖 $g \in G$ 的通用量化，这意味着关于守恒量的描述不再关注于单一的矢量场，尤其是，动量映射代表的守恒量具有分量。

在辛的情况中，微分 $\Phi_g^*\omega = \omega$ 意味着 ω 的李导数$\mathcal{L}_{\xi_M}\omega$ 对于 ξ 而言等于零，即$\mathcal{L}_{\xi_M}\omega = 0$。我们在 2.1.3 节中发现这等同于 ξ_M 是局部的哈密顿函数，即存在局部标量 J：$U \subset M \to \mathbb{R}$使得 $\xi_M = X_J$。这就是 2.1.3 节是如何从证明式(19)的“单行方式(one-liner)”推出诺特定理是正确的。因为矢量场 X_f 是局部的哈密顿函数，它保持了辛结构，即李导出的辛结构$\mathcal{L}_{X_f}\omega = 0$——对称性应具备的条件。

我们同样在 3.2.2 节结尾的结论(2)中发现了这个具有矢量场上的李括号关于标量的泊松括号的“网格(meshing)”意味着局部的哈密顿函数矢量场构成了所有矢量场的李代数$\mathcal{X}(M)$的李子代数。

回到泊松流形的背景中，我们需要指出两点：第一点就是与辛的情况的相似性；第二点就是对比(不同之处)。

(1) 我们只需应用式(313)就能轻易地检验出无穷小泊松自同构在李括号下是闭合的。于是我们把这些矢量场的李代数写作$\mathcal{P}(M)$：$\mathcal{P}(M) \subset \mathcal{X}(M)$。

(2) 另一方面，2.1.3 节中局部的哈密顿函数的矢量场和保持态空间的几何结构之间的等效性失效了。

显然，第一点意味着第二点即局部哈密顿函数矢量场保持了泊松括号。我们已经在5.3.2节中提到了这一点。这个微分的陈述就是说，对于这样一个场 X_H 而言，李导出的泊松张量：$\mathcal{L}_{X_H}B^\sharp=0$，见式(285)。有限的陈述就是说，这个场的流是泊松映射：$\phi_\tau^*\{F, G\}=\{\phi_\tau^*F, \phi_\tau^*G\}$，见式(284)。

但是逆向的蕴含是错误的，即泊松流形上的无穷小泊松自同构不必是局部的哈密顿函数。例如，通过定义泊松结构使得$\mathbb{R}^2$ 成为一个泊松流形：

$$\{F, H\}=x\left(\frac{\partial F}{\partial x}\frac{\partial H}{\partial y}-\frac{\partial H}{\partial x}\frac{\partial F}{\partial y}\right) \tag{314}$$

于是，y 轴上的点周围的矢量场 $X=\partial/\partial y$ 是非哈密顿函数的无穷小泊松自同构。这点将影响6.2小节中适用于泊松流形的诺特定理的构想。

然而，从现在开始我们应该对对于所有 ξ，ξ_M 是整体的哈密顿函数的情况感兴趣。这意味着存在映射 J：$\mathfrak{g}\to\mathcal{F}(M)$使得：

$$X_{J(\xi)}=\xi_M \tag{315}$$

适用于所有 $\xi\in\mathfrak{g}$。关于这个条件，我们需要注意的有以下三点。

(1) 因为在 ξ 中式(315)的右边是线性的，我们可以要求 J 是线性映射。为了使得任意 J 符合式(315)，我们能够依据$\mathfrak{g}$的 e_1，e_2，…，e_m 的基并且通过设置对于任意 $\xi=\xi^ie_i$ 有$\bar{J}(\xi):=\xi^iJ(e_i)$来定义一个新的线性的$\bar{J}$。

(2) 式(315)不能确定 $J(\xi)$。通过映射B：$dJ(\xi)\longmapsto X_{J(\xi)}$ 的线性度，我们能够给这个 $J(\xi)$附加任意的可区分函数，也就是使得 $X_F=0$ 的 F：$M\to\mathbb{R}$。也就是说，$X_{J(\xi)+F}\equiv X_{J(\xi)}$。(当然，在辛的案例中，只有可区分的函数是常数。)

(3) 在泊松括号方面，值得一提的是式(315)。这让我们回想起，对于任意 F，$H\in\mathcal{F}(M)$，我们有 $X_H(F)=\{F, H\}$，这个方程变成：

$$\{F, J(\xi)\}=\xi_M(F),\ \forall F\in\mathcal{F}(M),\ \forall\xi\in\mathfrak{g} \tag{316}$$

我们同样需要下面的结论：

$$X_{J([\xi,\eta])}=X_{\{J(\xi),J(\eta)\}M} \tag{317}$$

为了证明这些，我们只是应用了两个先前的结论分别给出了李代数反同态。

(i) 4.4节末尾的结论(4)。对于任意流形 M 上的任意李群 G 的左作用量，映射 $\xi\longmapsto\xi_M$ 是$\mathfrak{g}$和 M 上所有矢量场的李代数$\mathcal{X}_M$之间的李代数反同态：

$$(a\xi+b\eta)_M=a\xi_M+b\eta_M;\ [\xi_M,\ \eta_M]=-[\xi,\ \eta]_M \quad \forall\xi,\ \eta\in\mathfrak{g},\ 且\ a,\ b\in\mathbb{R} \tag{318}$$

（ii）正如在辛的例子中所示的，“网格”取决于标量上具有矢量场的李括号的泊松括号的符号，同 5.2.2 小节末尾的式(242)相似：

$$X_{\{F,H\}}=-[X_F,\ X_H]=[X_H,\ X_F] \tag{319}$$

于是，对于泊松流形 M 而言，映射 $F\in\mathcal{F}(M)\longmapsto X_F\in\mathcal{X}(M)$ 是一个李代数反同态。

应用(i)和(ii)，我们用下式推论出式(317)：

$$X_{J([\xi,\eta])}=[\xi,\ \eta]_M=-[\xi_M,\ \eta_M]=-[X_{J(\xi)},\ X_{J(\eta)}]=X_{\{J(\xi),J(\eta)\}M} \tag{320}$$

6.1.2 引入动量映射

于是，假如在泊松流形 M 上存在 G 的正则左作用量，并且假设存在线性映射 J：$\mathfrak{g}\to\mathcal{F}(M)$ 对于所有 $\xi\in\mathfrak{g}$ 满足：

$$X_{J(\xi)}=\xi_M \tag{321}$$

两个要求——无穷小的正则作用量，即每个 $\xi_M\in\mathcal{P}(M)$，以及每个整体哈密顿函数 ξ_M——能够被表示为需要一个 J：$\mathfrak{g}\to\mathcal{F}(M)$，使得存在交换图表(commutative diagram)。也就是说，映射 $\xi\in\mathfrak{g}\mapsto\xi_M\in\mathcal{P}(M)$ 也等于下面的映射：

$$\mathfrak{g}\xrightarrow{J}\mathcal{F}(M)\xrightarrow{F\longmapsto X_F}\mathcal{P}(M) \tag{322}$$

于是映射 $\mathbf{J}$：$M\to\mathfrak{g}^*$ 由下式定义：

$$<\mathbf{J}(x);\ \xi> := J(\xi)(x) \tag{323}$$

对于所有 $\xi\in\mathfrak{g}$ 和 $x\in M$，式(323)被称为作用量的动量映射。

陈述这个定义的另一种方法如下所示：任意平滑函数 $\mathbf{J}$：$M\to\mathfrak{g}^*$ 在任意 $\xi\in\mathfrak{g}$ 定义了标量 $J(\xi)$：$x\in M\mapsto(\mathbf{J}(x))(\xi)\in\mathbb{R}$。通过把 $J(\xi)$ 看作是哈密顿函数，我们定义了哈密顿矢量场 $X_{J(\xi)}$。但是，因为 G 作用于 M，每个 $\xi\in\mathfrak{g}$ 定义了 M 上的矢量场，即 ξ_M。于是，我们说，如果对于每个 $\xi\in\mathfrak{g}$，这两个矢量场都是相等的，即 $X_{J(\xi)}=\xi_M$，那么 $\mathbf{J}$ 是关于作用量的动量映射。

通过阐述这个定义，给出三个进一步的备注。

(1) 同构

我们很轻易地就能检验出式(323)定义了从 M 到 $\mathfrak{g}^*$ 的平滑映射 $\mathbf{J}$ 空间以及从$\mathfrak{g}$到标量函数$\mathcal{F}(M)$的线性映射 J 空间之间的同构。如果在 $x\in M$ 时，$\mathbf{J}(x):\xi\in\mathfrak{g}\longmapsto\mathbf{J}(x)(\xi)\in\mathbb{R}$是由如下的映射给出的，我们可以用 J 定义 $\mathbf{J}$：

$$\mathfrak{g}\xrightarrow{J}\mathcal{F}(M)\xrightarrow{|_x}\mathbb{R} \tag{324}$$

其中$|_x$ 意味着 $x\in M$ 时的估值。或者我们可以说明在每个 $\xi\in\mathfrak{g}$，$J(\xi)$：$x\in M\longmapsto J(\xi)(x)\in\mathbb{R}$是由下面的映射给出的，从而用 $\mathbf{J}$ 定义 J:

$$M\xrightarrow{\mathbf{J}}\mathfrak{g}^*\xrightarrow{|_\xi}\mathbb{R} \tag{325}$$

其中$|_\xi$ 意味着 $\xi\in\mathfrak{g}$时的估值。

(2) 动量映射的微分方程

通过使用哈密顿方程，我们很容易就能把动量映射的定义表示为微分方程的集合。我们回想起在泊松流形上，哈密顿方程是由式(270)决定的，即在 $x\in M$时，有：

$$\mathsf{B}_x(dH(x))=X_H(x) \tag{326}$$

或者在局部坐标系 x^i 中，$i=1,2,\cdots,m\equiv\dim(M)$，其中 $J^{ij}(x)\equiv\{x^i,x^j\}$，结构矩阵为：

$$\mathsf{B}_x\left(\frac{\partial H}{\partial x^j}dx^j\right)=\sum_{i,j}J^{ij}(x)\frac{\partial H}{\partial x^j}\frac{\partial}{\partial x^i}|_x \tag{327}$$

可参照式(272)，于是在局部坐标系中，哈密顿方程是由式(253)给出的，即：

$$\frac{dx^i}{dt}=\sum_j^m J^{ij}(x)\frac{\partial H}{\partial x^j} \tag{328}$$

因而适用于动量映射 $X_{J(\xi)}=\xi_M$ 的条件说明了对于所有 $\xi\in\mathfrak{g}$和所有 $x\in M$，有：

$$\mathsf{B}_x(d(J(\xi))(x))=\xi_M(x) \tag{329}$$

在坐标系中，这里的要求是对于所有 $i=1,2,\cdots,m$，有：

$$\sum_j^m J^{ij}(x)\frac{\partial J(\xi)}{\partial x^j}=(\xi_M)^i(x) \tag{330}$$

其中(很抱歉)左手边的两个 J 具有截然不同的涵义。

在辛情况中，$\dim(M)\equiv m=2n$，并且我们把哈密顿方程看作式(15)，即

$$\mathbf{i}_{X_H}\omega := \omega(X_H, \ \cdot\ ) = dH(\cdot) \tag{331}$$

于是，适用于动量映射的条件说明，对于所有 ξ：

$$\omega(\xi_{M,\cdot}) = d(J(\xi))(\ \cdot\ ) \tag{332}$$

在哈密顿力学中，我们一般把 $2n$ 个局部坐标 q，p 写为 ξ，即：

$$\xi^{\alpha} := q^{\alpha},\ \alpha = 1,\ 2,\ \cdots,\ n;\ \xi^{\alpha} := p_{\alpha - n},\ \alpha = n+1,\ \cdots,\ 2n \tag{333}$$

因此，为了表示局部坐标系中的式(332)，我们暂时把$\mathfrak{g}$的任意元素写作 η，从而通过写出 $\eta_M = (\eta_M)^{\alpha}\dfrac{\partial}{\partial \xi^{\alpha}}$和 $\omega_{\alpha\beta} := \omega(\dfrac{\partial}{\partial \xi^{\alpha}},\ \dfrac{\partial}{\partial \xi^{\beta}})$，式(332)变成了：

$$\omega_{\alpha\beta}(\eta_M)^{\alpha} = \frac{\partial J(\eta)}{\partial \xi^{\beta}} \tag{334}$$

(3) 分量：一个实例

正如式(313)之后的讨论，当 ξ 随$\mathfrak{g}$变化时我们把函数 $J(\xi)$ 的集合看作是 $\mathbf{J}$ 的分量。

拿我们的标准实例来说：在欧氏空间中，一个粒子在态 $x = (\mathbf{q},\ \mathbf{p})$ 中的角动量是 $\mathbf{J}(x) := \mathbf{q} \wedge \mathbf{p}$。用$\mathbb{R}^3$ 确定$\mathfrak{so}(3)^*$，于是点积给出了自然对，参见4.5.2节末尾的(3)，我们得知围绕轴 $\xi \in \mathbb{R}^3$ 的 $\mathbf{J}(x)$ 的分量 $<\mathbf{J}(x);\ \xi> = \xi \cdot (\mathbf{q} \wedge \mathbf{p})$。这个哈密顿函数 $x = (\mathbf{q},\ \mathbf{p}) \longmapsto \xi \cdot (\mathbf{q} \wedge \mathbf{p})$ 决定的哈密顿矢量场当然是围绕 ξ 轴转动的无穷小算子。在6.3节中，我们将看到关于动量映射的更多实例。

6.2 动量映射的守恒：诺特定理

在一般的哈密顿力学中，我们发现诺特定理具有一个简单的表达式，如基于泊松括号的反对称性的“单行方式”一样，也就是说，在式(19)中，对于任意标量函数 F，H：

$$X_F(H) = \{H,\ F\} = 0,\ \text{当且仅当}\ 0 = \{F,\ H\} = X_H(F) \tag{335}$$

换言之，哈密顿函数 H 是 F 产生的流之下的常量，当且仅当 F 是动态流 X_H 之下的运动常量。

更确切地说，2.1.3节已经证明了这个单行方式作为表示诺特定理的正确性。因为这个单行方式遵守对称性应该保持辛形式(相等于泊松括号)的要求，并不仅仅是哈密顿函数 H，同式(335)左手边所示，而且，根据嘉当魔术公式

(Cartan's magic formula)，一个保持了辛形式的矢量场就等同于它是局部的哈密顿函数。

对于泊松流形而言，与最后的陈述相对应的等效性并不成立。也就是说，同我们在 6.1.1 节的(2)中指出的一样，无穷小的泊松自同构不必是定域哈密顿函数。

虽然如此，对诺特定理的大多数“单行”方法都可使泊松流形的框架保持下去。实际上，我们仅把我们的讨论限制在相关哈密顿矢量场存在的情况中。回想我们在 6.1.1 节的(2)之后所说的，我们将集中精力研究所有 ξ_M 都是整体哈密顿函数的情况。

然而，我们直接说明，对于泊松流形 M，正如对于辛流形一样，如果 F，$H\in\mathcal{F}(M)$，当且仅当 $\{H, F\}=0$ 时，H 是沿着 X_F 的积分曲线的常数，且 F 也是沿着 X_H 的积分曲线的常数(我们在 5.2.2 小节中已经证明了这一点，但是直到现在，我们才用到它)。

以这个结论作为定理，我们立即得到以下内容。

适用于泊松流形的诺特定理 假设 G 正则地作用于泊松流形 M 并且具有动量映射 $\mathbf{J}$: $M\to\mathfrak{g}^*$；并且 H 是 ξ_M 之下的不变量，其中 $\xi\in\mathfrak{g}$，即 $\{H, J(\xi)\}=\xi_M(H)=0$，$\forall\xi\in\mathfrak{g}$；参见式(316)。于是 $\mathbf{J}$ 是由 H 决定的运动常量。也就是说：

$$\mathbf{J}\circ\phi_\tau=\mathbf{J} \tag{336}$$

其中 ϕ_τ 是 X_H 的流。

证明：根据这个引理，$\{H, J(\xi)\}=\xi_M(H)=0$ 的事实意味着 $J(\xi)$ 是沿着 X_H 的流的常量。于是，根据动量映射的定义，即式(323)，其对应的 $\mathfrak{g}^*$ 值映射 $\mathbf{J}$ 也是运动常量。

它立即导出，H 自身(以及任意可区分的函数)是运动常量。除此之外，正如 6.1.1 节结尾的(2)中标注的：运动常量 $J(\xi)$ 仅是由可区分的函数的任意选择决定的。实际上，尽管本章从 1.2 节的(iii)以来并不关心依赖时间的函数，但如果我们考虑它们，那么这里就存在对于依赖于时间的可区分函数的任意选择。

6.3 例子

我们给出两个常见的例子。作为对这个理论的一般简介，我们讨论两个抽象的例子(我们稍后将不再需要)。

(1) N个粒子的完全线动量

在4.1.0.1小节末尾的(3)中，我们说明了提升到$M=T^*\mathbb{R}^{3N}$的$Q=\mathbb{R}^{3N}$上的平动群$\mathbb{R}^3$的作用量的左余切提升，即与式(126)一致的左作用量：

$$\psi_{\mathbf{x}}(\mathbf{q}_i,\ \mathbf{p}^i) := T^*(\Phi_{-\mathbf{x}})(\mathbf{q}_i,\ \mathbf{p}^i) = (\mathbf{q}_i+\mathbf{x},\ \mathbf{p}^i),\ i=1,\ \cdots,\ N \tag{337}$$

这里，我们结合了4.1.0.1小节中关于实例(vi)和(ix)的讨论。

为了找到动量映射，我们(a)计算了适用于$\mathfrak{g}=\mathbb{R}^3$的任意元素$\xi$的无穷小算子$\xi_M$，然后(b)解决了式(332)，或者式(334)中的坐标。

(a)就$\mathbf{x}$而言，我们在ξ方向上微分了式(337)，从而得到：

$$\xi_M(\mathbf{q}_i,\ \mathbf{p}^i) = (\xi,\ \cdots,\ \xi,\ 0,\ \cdots,\ 0) \tag{338}$$

(b)任意函数$J(\xi)$具有哈密顿矢量场：

$$X_{J(\xi)}(\mathbf{q}_i,\ \mathbf{p}^i) = \left(\frac{\partial J(\xi)}{\partial \mathbf{p}^i},\ -\frac{\partial J(\xi)}{\partial \mathbf{q}_i}\right) \tag{339}$$

于是，我们想要的具有$X_{J(\xi)}=\xi_M$的$J(\xi)$解决了如下形式：

$$\frac{\partial J(\xi)}{\partial \mathbf{p}^i} = \xi \text{ 和 } \frac{\partial J(\xi)}{\partial \mathbf{q}_i} = 0,\ 1 \leqslant i \leqslant N \tag{340}$$

我们选择常数，于是J是线性的，这个解是：

$$J(\xi)(\mathbf{q}_i,\ \mathbf{p}^i) = \left(\textstyle\sum_{i=1}^{N}\mathbf{p}^i\right)\cdot\xi,\ \text{即 } \mathbf{J}(\mathbf{q}_i,\ \mathbf{p}^i) = \textstyle\sum_{i=1}^{N}\mathbf{p}^i \tag{341}$$

即常见的总线动量。

(2) 单个粒子的角动量

$SO(3)$通过$\Phi_A(\mathbf{q})=A\mathbf{q}$作用于$Q=\mathbb{R}^3$。于是，正切(导数)映射是：

$$T_{\mathbf{q}}\Phi_A:\ (\mathbf{q},\ \mathbf{v}) \in T\mathbb{R}^3_{\mathbf{q}} \longmapsto (A\mathbf{q},\ A\mathbf{v}) \in T\mathbb{R}^3_{A\mathbf{q}} \tag{342}$$

正如我们在4.1.0.1小节的实例(vii)中所看到的，这个提升到$M=T^*\mathbb{R}^3$的作用量的左余切提升如下，其中用“g^{-1}”表示提升作用量，与式(126)相一致：

$$T^*_{A\mathbf{q}}(\Phi_{A^{-1}})(\mathbf{q},\ \mathbf{p}) = (A\mathbf{q},\ A\mathbf{p}) \tag{343}$$

为了寻找动量映射，我们进行了两个步骤(a)和(b)，如例(1)中所示。

(a)　我们在方向 $\xi=\Theta(\omega)\in\mathfrak{so}(3)$ 上对式(343)作关于 A 的微分，其中 $\omega\in\mathbb{R}^3$，且 Θ 同在式(48)和式(51)中一样。我们得到：

$$\xi_M(\mathbf{q},\ \mathbf{p})=(\xi\mathbf{q},\ \xi\mathbf{p})=(\omega\wedge\mathbf{q},\ \omega\wedge\mathbf{v}) \tag{344}$$

(b)　于是，期望值 $J(\xi)$ 是哈密顿方程的解，在 ξ 中是线性的：

$$\frac{\partial J(\xi)}{\partial\mathbf{p}}=\xi\mathbf{q}\text{ 和 }\frac{\partial J(\xi)}{\partial\mathbf{q}}=-\xi\mathbf{p} \tag{345}$$

于是根据下式，我们给出解：

$$J(\xi)(\mathbf{q},\ \mathbf{p})=(\xi\mathbf{q})\cdot\mathbf{p}=(\omega\wedge\mathbf{q})\cdot\mathbf{p}=(\mathbf{q}\wedge\mathbf{p})\cdot\omega \tag{346}$$

那么：

$$\mathbf{J}(\mathbf{q},\ \mathbf{p})=\mathbf{q}\wedge\mathbf{p} \tag{347}$$

即常见的角动量。

(3) 李代数同态的对偶

我们从叙述一个后文中没有证明的引理开始，详见马斯登和拉蒂乌 [1999, 10.7.2, 372]。也就是说，令 G, H 为李群并令 α: $\mathfrak{g}\to\mathfrak{h}$ 为它们的李代数之间的线性映射。于是，α 是李代数同态当且仅当它的对偶 α^*：$\mathfrak{h}^*\to\mathfrak{g}^*$ 是(线性)泊松映射(其中 $\mathfrak{h}^*$，$\mathfrak{g}^*$ 具有其自然李—泊松括号，同 5.2.4 小节中一样)。

现在，令 G, H 为李群，令 A: $H\to G$ 为李群同态，并令 α: $\mathfrak{h}\to\mathfrak{g}$ 为归纳的李代数同构，于是，根据这个引理，α^*：$\mathfrak{g}^*\to\mathfrak{h}^*$ 是泊松映射。我们将证明对于 $\mathfrak{g}^*$ 上的 H 的作用量而言，α^* 也是动量映射，$h\in H$, $x\in\mathfrak{g}^*$：

$$\Phi(h,\ x)\equiv h\cdot x:=Ad^*_{A(h)^{-1}}x \tag{348}$$

证明：我们首先回想伴随和余伴随作用量 Ad_g: $\mathfrak{g}\to\mathfrak{g}$ 和 Ad^*_g：$\mathfrak{g}^*\to\mathfrak{g}^*$，尤其是式(191)。于是式(348)中的作用量是：

$$\forall x\in\mathfrak{g}^*,\ \forall\xi\in\mathfrak{g}:\ <h\cdot x;\ \xi>=<x;\ Ad_{A(h)^{-1}}\xi> \tag{349}$$

和往常一样，对于 $\eta\in\mathfrak{h}$，我们通过在方向 $\eta\in\mathfrak{h}$ 上 e 处对式(349)求关于 h 的微分来计算在 $x\in\mathfrak{g}^*$ 的无穷小算子 $\eta_{g}{}^*$ 的值。参见式(198)我们得到：

$$<\eta_{g^*}(x);\ \xi>=-<x;\ ad_{\alpha(\eta)}\xi>=<ad^*_{\alpha(\eta)}(x);\ \xi> \tag{350}$$

我们定义 $\mathbf{J}(x):=\alpha^*(x)$，即：

$$J(\eta)(x) \equiv <\mathbf{J}(x);\ \eta> := <\alpha^*(x);\ \eta> \equiv <x;\ \alpha(\eta)> \tag{351}$$

这意味着：

$$\nabla_x J(\eta) = \alpha(\eta) \tag{352}$$

现在，我们回想一下，作为哈密顿函数的 $J(\eta)$ 的哈密顿方程如下，参见式(261)：

$$\dot{x} \equiv X_{J(\eta)}(x) = ad^*_{\nabla J(\eta)}(x) \tag{353}$$

结合式(350)和式(353)，我们得到：

$$X_{J(\eta)}(x) = ad^*_{\alpha(\eta)}(x) = \eta_{\mathfrak{g}^*}(x) \tag{354}$$

这证明了 $\mathbf{J}(x) := \alpha^*(x)$ 是动量映射。

(4) 适用于子群的动量映射

假设 $\mathbf{J}$：$M \to \mathfrak{g}^*$ 是关于 G 在 M 上的正则左作用量的动量映射，并且令 $H < G$ 为 G 的子群。于是，H 也正则地作用于 M，并且这个作用量像动量映射一样把 $\mathbf{J}$ 的值限制于 $\mathfrak{h} \subset \mathfrak{g}$。也就是说，通过 $\mathbf{J}_H(x) := \mathbf{J}(x) \mid_{\mathfrak{h}}$ 出映射：

$$\mathbf{J}_H:\ M \to \mathfrak{h}^* \tag{355}$$

G 的正则作用量确保：如果 $\eta \in \mathfrak{h} \subset \mathfrak{g}$，那么 $\eta_M = X_{J(\eta)}$。于是，$J_H(\eta) := J(\eta)\ \forall \eta \in \mathfrak{h}$ 定义了 H 的作用量的动量映射。即：

$$\forall x \in M,\ \forall \eta \in \mathfrak{h}:\ <\mathbf{J}_H(x);\ \eta> = <\mathbf{J}(x);\ \eta> \tag{356}$$

6.4 动量映射的等变性

在 4.2 小节的(1)中，我们定义了流形之间的等变映射 f：$M \to N$ 的一般概念作为涉及 M 和 N 上的群 G 的作用量的概念，即式(144)。现在我们提出了这个概念非常重要的一种情形：当 $\mathfrak{g}^*$ 上的作用量是余伴随作用量时，动量映射的等变性 $\mathbf{J}$：$M \to \mathfrak{g}^*$，见式(192)。

对于我们，这个概念将有两个重要的意义：

(i) 发生在例子中的大量动量映射在这种意义上都是等变化的；

(ii) 等变性有许多理论结果，尤其，对于余切提升作用量的动量映射往往是等变化的(6.5 节)，等变性在关于约化的定理中是非常重要的。在这个小节，我们将借助下列内容思考这些观点：

(i) 定义概念，以及讨论关于概念的稍微不同的版本(6.4.1 节)；

(ii) 证明等变动量映射都是泊松的(6.4.2 节)。

6.4.1　等变性和无穷小等变性

假设 Φ 是 M 上 G 的正则左作用量，且假设 $\mathbf{J}$: $M \to \mathfrak{g}^*$ 是对于它的一个映射，我们就说 $\mathbf{J}$ 对于所有 $g \in G$ 都是等变化的。

$$\mathbf{J} \circ \Phi_g = Ad^*_{g^{-1}} \circ \mathbf{J} \tag{357}$$

参见式(144)以及式(193)对于余伴随的定义：

$$\begin{array}{ccc} M & \xrightarrow{\mathbf{J}} & \mathfrak{g}^* \\ \uparrow \Phi_g & & \uparrow Ad^*_{g^{-1}} \\ M & \xrightarrow{\mathbf{J}} & \mathfrak{g}^* \end{array} \tag{358}$$

一个等效的表达式通过考虑我们添加式(358)中的可交换正方形和两个可交换三角形而出现：

$$M \xrightarrow{J(\xi)} \mathbb{R} \text{是} M \xrightarrow{\mathbf{J}} \mathfrak{g}^* \xrightarrow{|_\xi} \mathbb{R} \tag{359}$$

表示 $J(\xi)(x) = \mathbf{J}(x)(\xi)$，且：

$$\mathfrak{g}^* \xrightarrow{|_\xi} \mathbb{R} \text{是} \mathfrak{g}^* \xrightarrow{Ad^*_{g^{-1}}} \mathfrak{g}^* \xrightarrow{|_{Ad_g\xi}} \mathbb{R} \tag{360}$$

表示对于所有的 $\eta \in g^*$，有：

$$< Ad^*_{g^{-1}}(\eta);\ Ad_g(\xi) > = < \eta;\ Ad_{g^{-1}} Ad_g(\xi) > \equiv < \eta;\ \xi > \tag{361}$$

式(359)和式(360)意味着，等变性的等效表达式是，对于所有的 $x \in M$，$g \in G$ 和 $\xi \in \mathfrak{g}(g \cdot x \equiv \Phi_g(x))$，我们有：

$$\mathbf{J}(g \cdot x)(Ad_g \xi) \equiv J(Ad_g \xi)(g \cdot x) = J(\xi)(x) \equiv \mathbf{J}(x)(\xi) \tag{362}$$

在4.4 小节中，我们区分了等变映射的一般概念，并得到了稍微不同的概念，即作用在 M 和 N 上 G 的作用量的无穷小算子 ξ_M 和 ξ_N 是 f 相关的。

我们也区分了等变性，并得到了无穷小等变性(infinitesimal equivariance)。但是我不会逐一细说，因为：

(i) 我们不需要这个概念，不只是因为之前提到的大量的动量映射是等变的；

(ii) 在特定的一般条件下(如一个紧凑或者连贯的群 G)，一个无穷小等变

动量映射往往能被一个等变动量映射取代。

因此，可以说，无穷小等变性在理论上是重要的。尤其是式(317)的结论，即：

$$X_{J([\xi,\eta])} = X_{\{J(\xi),J(\eta)\}_M} \tag{363}$$

这意味着：

$$\Sigma(\xi,\ \eta) := J([\xi,\ \eta]) - \{J(\xi),\ J(\eta)\}_M \tag{364}$$

是泊松流形 M 上非常不同的函数，因此在每一个辛叶上都是常量。

这自然要问什么时候 $\Sigma \equiv 0$，可参见式(322)。$\xi \longmapsto \xi_M$ 和 $F \longmapsto X_F$ 是李代数的反同态。因此，常常要问 J 是否是李代数的同态，即是否 $\Sigma = 0$，结果无穷小等变性等价于 $\Sigma = 0$。

6.4.2 等变动量映射是泊松的

接下来的结论很重要，作为寻找泊松流形之间的正则映射的一般方法，以及对于第 7 节的李—泊松约化定理。

等变动量映射是泊松的 假设 $\mathbf{J}$：$\mathbf{M} \to \mathfrak{g}^*$ 对于泊松流形 M 上 G 的正则左作用量是一个等变动量映射，那么对于所有的 F_1，$F_2 \in \mathcal{F}(\mathfrak{g}^*)$，$\mathbf{J}$ 是泊松映射：

$$\mathbf{J}^*\{F_1,\ F_2\}_{\mathfrak{g}^*} = \{\mathbf{J}^* F_1,\ \mathbf{J}^* F_2\}_M;\ 即\{F_1,\ F_2\}_{\mathfrak{g}^*} \circ \mathbf{J} = \{F_1 \circ \mathbf{J},\ F_2 \circ \mathbf{J}\}_M \tag{365}$$

证明：我们将(i)式(365)的左边和 J 联系起来，然后再将(ii)式(365)的右边和 J 联系起来，最终我们将使用这个事实，即 M 上的泊松括号仅依赖于一阶导数的值。

(i) 假设 $x \in M$，$\alpha = \mathbf{J}(x) \in \mathfrak{g}^*$，且假设在 α 处的值 $\xi = \nabla F_1$ 和 $\eta = \nabla F_2$，使得 ξ，$\eta \in \mathfrak{g}^{**} = \mathfrak{g}$。那么：

$$\{F_1,\ F_2\}_{\mathfrak{g}^*}(\mathbf{J}(x)) \equiv <\alpha;\ [\nabla F_1,\ \nabla F_2]> = <\alpha;\ [\xi,\ \eta]> = J([\xi,\ \eta])(x)$$
$$= \{J(\xi),\ J(\eta)\}(x) \tag{366}$$

这里第三个方程仅适用于式(323)中 $\mathbf{J}$ 的定义，第四个方程使用了(无穷小)等变性。

(ii) 我们认为 $(F_1 \circ \mathbf{J})(x)$ 和 $J(\xi)(x)$ 和都有相同的 x 导数。对于任意的 $x \in M$ 和 $v_x \in T_xM$，有：

$$\mathrm{d}(F_1 \circ \mathbf{J})(x) \cdot v_x = \mathrm{d}F_1(\alpha) \cdot T_x\mathbf{J}(v_x) = <T_x\mathbf{J}(v_x);\ \nabla F_1> = \mathrm{d}\mathbf{J}(\xi)(x) \cdot v_x \tag{367}$$

当第一个方程使用链式法则(chain rule)时，且最后一个方程使用式(323)中 $\mathbf{J}$ 的定义，那么 $\xi=\nabla F_1$。

最终，因为 M 上的泊松括号仅依赖于一阶导数的值，那么，我们可以从式(367)中推断出：

$$\{F_1\circ J,\ F_2\circ J\}(x)=\{J(\xi),\ J(\eta)\}(x) \tag{368}$$

结合这个式子和(i)，可以得出结论。

6.5　余切丛上的动量映射

假设李群 G 作用在流形("位形空间")Q 上。我们在4.1.0.1 小节看到，这个作用量可以被提升到余切丛 T^*Q，参见式(121)、式(124)和式(126)。在这个小节中，我们关注这些余切提升的动量映射。我们将看到，任意这样的作用量都有一个等变动量映射，对于这个等变动量映射，存在一个明确的通式。一般理论(6.5.1 节和 6.5.2 节)将只需一个主要的新概念，即动量函数。我将在最后举一些例子(6.5.3 节)。

6.5.1　动量函数

给定一个流形 Q 和它的矢量场$\mathcal{X}(Q)$，对于 $q\in Q$，$X\in\mathcal{X}(Q)$和 $\alpha_q\in T_q^*Q$，我们通过$(\mathcal{P}(X))(\alpha_q):=<\alpha_q;\ X(q)>$定义映射：

$$\mathcal{P}:\ \mathcal{X}(Q)\rightarrow\mathcal{F}(T^*Q) \tag{369}$$

从严格意义上说，α_q 是基点 $q\in Q$ 之上的余切丛的点，因此，α_q 可以被写作$(q,\ \alpha)$，其中 α 是在 q 点的余矢量，即 $\alpha\in T_q^*Q$。但是，就在定义余切提升之前，参见式(121)我们提到合并 T^*Q 中的点是可以的，即$(q,\ \alpha)$中的 $q\in Q$，$\alpha\in T_q^*Q$ 带有它的形式 α，条件是我们通过将其写为形式 α_q 和 q 保持联系。

式(369)定义的$\mathcal{P}(X)$被称作 X 的动量函数。在坐标系中，$\mathcal{P}(X)$通过下式给定：

$$\mathcal{P}(X)(q^i,\ p_i)=X^j(q^i)p_j \tag{370}$$

其中我们在 $j=1,\ 2,\ \cdots,\ n$ 上的求和等于 $\dim Q$。备注：这个$\mathcal{P}$不同于$\mathcal{P}(M)$中的$\mathcal{P}$，它是6.1.1 节讨论过的 M 的无穷小泊松自同构。

我们也可以用$\mathcal{L}(T^*Q)$来表示在 T^*Q 的纤维上是线性的平滑函数 F：T^*Q

$\to\mathbb{R}$的空间，也就是把丛点(bundle points)α_q，$\beta_q\in T_q^*Q$写为(q,α)和(q,β)，对λ，$\mu\in\mathbb{R}$可得：

$$F(q,(\lambda\alpha+\mu\beta))=\lambda F((q,\alpha))+\mu F((q,\beta)) \tag{371}$$

因此，在$\mathcal{L}(T^*Q)$里的函数F，H对于函数X^i和Y^i可以在坐标系中被写作：

$$F(q,p)=X^i(q)p_i\text{ 和 }H(q,p)=Y^i(q)p_i \tag{372}$$

从而得到，任意动量函数$\mathcal{P}(X)$都是在$\mathcal{L}(T^*Q)$里的。

容易得知，这样来自小节2.1.1中的T^*Q的辛结构的F和H的正则泊松括号在T^*Q的结构上也是线性的。事实上，式(372)意味着：

$$\{F,H\}(q,p):=\frac{\partial F}{\partial q^j}\frac{\partial H}{\partial p_j}-\frac{\partial H}{\partial q^j}\frac{\partial F}{\partial p_j}=\left(\frac{\partial X^i}{\partial q^j}Y^j-\frac{\partial Y^i}{\partial q^j}X^j\right) \tag{373}$$

因此，$\mathcal{L}(T^*Q)$是$\mathcal{F}(T^*Q)$的一个李子代数。

下面的结论概述了动量函数是如何将$\mathcal{X}(Q)$以及T^*Q上的哈密顿矢量场与$\mathcal{L}(T^*Q)$相联系起来的。

三个(反)同构李代数　两个李代数：

(i)　Q上矢量场的$(\mathcal{X}(Q),[\,,])$；

(ii)　T^*Q上的哈密顿矢量场$X_F(F\in\mathcal{L}(T^*Q))$是同构的，每一个都和(iii)是反同构的；

(iii)　$(\mathcal{L}(T^*Q),\{\,,\})$。

尤其是，映射$\mathcal{P}$是一个从(i)到(iii)的同构，这使得我们有：

$$\{\mathcal{P}(X),\mathcal{P}(Y)\}_{T^*Q}=-\mathcal{P}([X,Y]) \tag{374}$$

证明：因为$\mathcal{P}(X):T^*Q\to\mathbb{R}$在结构上是线性的，$\mathcal{P}$将$\mathcal{X}(Q)$映射入$\mathcal{L}(T^*Q)$内。$\mathcal{P}$也映射到$\mathcal{L}(T^*Q)$上：给定$F\in\mathcal{L}(T^*Q)$，我们通过下式定义$X(F)\in\mathcal{X}(Q)$：

$$<\alpha_q;X(F)(q)>:=F(\alpha_q)\quad\forall\alpha_q\in T_q^*Q \tag{375}$$

因此，$\mathcal{P}(X(F))=F$。$\mathcal{P}$是线性的，且$\mathcal{P}(X)=0$意味着$X=0$。并且，通过对比式(373)和X，$Y\in\mathcal{X}(Q)$的李括号立刻可得式(374)，参见式(55)。因此$\mathcal{P}$是从$(\mathcal{X}Q,[\,,])$到$(\mathcal{L}(T^*Q),\{\,,\})$的反同构。

映射：

$$F\in(\mathcal{L}(T^{*}Q),\ \{,\})\longmapsto X_{F}\in(\{X_{F}\mid F\in\mathcal{L}(T^{*}Q)\},\ [\,,])\tag{376}$$

根据定义是满射的。通过式(60)，它是一个李代数的反同态。如果 $X_F=0$，那么 F 在 T^*Q 上是常量，因此 $F\equiv 0$，因为 F 在结构上是线性的，参见式(371)。

6.5.2 余切提升作用的动量映射

我们将以一个结论开始这一小节，这个结论把通过动量函数$\mathcal{P}(X)$推导出的 T^*Q 上的哈密顿流和通过 X 推导出的 X 的哈密顿流联系起来。从这个结论，我们的主要结论，即对于余切提升作用量的等变动量映射的保证以及对于它的显式公式，可直接得出。

动量的哈密顿流 假设 $X\in\mathcal{X}(Q)$ 且有 Q 上的流 ϕ_τ，参见 3.1.2.2 小节。那么，T^*Q 上的流 $X_{\mathcal{P}(X)}$ 是 $T^*\phi_{-\tau}$，即流 $X_{\mathcal{P}(X)}$ 是 $\phi_{-\tau}$ 的余切提升(4.1.0.1 小节)，π_Q 是正则投射(canonical projection)：

$$\begin{tikzcd} Q \arrow[r, "\phi_\tau"] & Q \\ T^{*}Q \arrow[u, "\pi_Q"] \arrow[r, "T^{*}\phi_{-\tau}"] & T^{*}Q \arrow[u, "\pi_Q"'] \end{tikzcd}\tag{377}$$

证明：我们对式(377)中的关系求微分，即：

$$\pi_Q\circ T^{*}\phi_{-\tau}=\phi_\tau\circ\pi_Q\tag{378}$$

当 $\tau=0$ 时，我们得到：

$$T\pi_Q\circ Y=X\circ\pi_Q\text{，其中}\ \forall\,\alpha_q\in T_q^{*}Q,\ Y(\alpha_q)=\frac{d}{d\tau}\Big|_{\tau=0}T^{*}Q_{-\tau}(\alpha_q)\tag{379}$$

即 $T^*\phi_{-\tau}$是 Y 的流。

现在，我们将表明 $Y=X_{\mathcal{P}(X)}$，这个结论是通过使用式(379)和2.1 节中哈密顿力学的几何公式，尤其是式(20)中应用于正则 1 形式 $\theta\equiv\theta_H$ 的嘉当公式得到的。我们认为，在4.1.0.1 小节中(2)的开始部分，余切提升 $T^*\phi_{-\tau}$使得在 T^*Q 上，$\theta\equiv\theta_H$，因此，$\mathcal{L}_Y\theta=0$。作为 θ 的负外导数 ω 的定义和嘉当公式使得：

$$\mathbf{i}_Y\omega=-\mathbf{i}_Y d\theta=\mathbf{d}\mathbf{i}_Y\theta\tag{380}$$

另一方面，我们有：

$$\mathbf{i}_Y\theta(\alpha_q)\equiv\langle\theta(\alpha_q);\ Y(\alpha_q)\rangle=\langle\alpha_q;\ T\pi_Q(Y(\alpha_q))\rangle=\langle\alpha_q;\ X(q)\rangle=\mathcal{P}(X)(\alpha_q)\tag{381}$$

这个结论是当第二个方程应用正则1形式，见式(8)、第三个应用式(379)以及第四个应用动量函数的式(369)的定义的时候得到的。

结合式(380)和式(381)，我们得到：

$$\mathbf{i}_Y\omega = \mathbf{d}\,\mathcal{P}(X) \tag{382}$$

哈密顿方程即式(15)告诉我们，$Y = X_{\mathcal{P}(X)}$。因此，T^*Q 上的哈密顿矢量场 $X_{\mathcal{P}(X)}$ 被称为 $X \in Q$ 到 T^*Q 的余切提升。在局域坐标中，结合式(15)和式(370)，我们可以写作：

$$X_{\mathcal{P}(X)} = \frac{\partial \mathcal{P}(X)}{\partial p_i}\frac{\partial}{\partial q^i} - \frac{\partial \mathcal{P}(X)}{\partial q^i}\frac{\partial}{\partial p_i} = X^i\frac{\partial}{\partial q^i} - \frac{\partial X^i}{\partial q^j}p_i\frac{\partial}{\partial p_j} \tag{383}$$

尤其要注意，结合通常的泊松括号和李代数之间的符号变更以及关于动量函数的符号变更，我们有：

$$[X_{\mathcal{P}(X)},\ X_{\mathcal{P}(Y)}] = -X_{\{\mathcal{P}(X),\mathcal{P}(Y)\}} = -X_{-\mathcal{P}([X,Y])} = X_{\mathcal{P}([X,Y])} \tag{384}$$

现在可以很容易证明对于等变动量映射我们给出的一个公式的主要结论。

等变动量映射 假设 G 作用在 Q 上的左边，因而它是通过 T^*Q 上的余切提升作用在 Q 上的左边。那么，这个余切提升作用量具有一个等变映射 $\mathbf{J}$: $T^*Q \rightarrow \mathfrak{g}^*$，这个映射通过下式给定：

$$<\mathbf{J}(\alpha_q);\ \xi> = <\alpha_q;\ \xi_Q(q)> \equiv \mathcal{P}(\xi_Q)(\alpha_q) \tag{385}$$

在 T^*Q 上的坐标 q^i，p_i 和 $\mathfrak{g}$ 上的坐标 ξ^a，其中 ξ_Q 的分量 $\xi^i_Q = \xi^a A^i_a$，这写为：

$$J_a\xi^a = p_i\xi^i_Q = p_iA^i_a\xi^a \tag{386}$$

因此，$J_a(q,\ p) = p_iA^i_a(q)$。

证明：前面的结论告诉我们，对于任意的 $\xi \in \mathfrak{g}$，作用在 T^*Q 上的余切提升作用量的无穷小算子是 $\xi_{T^*Q} \equiv X_{\mathcal{P}(\xi_Q)}$。因此，对于这个作用量的动量映射通过下式给定：

$$J(\xi) = \mathcal{P}(\xi_Q) \tag{387}$$

这仅仅通过应用动量映射 $\mathbf{J}$ 即式(323)和动量函数的定义即式(369)得到式(385)。

为了证明等变性，我们讨论如下内容。

$$<\mathbf{J}(g \cdot \alpha_q); \xi> = <(g \cdot \alpha_q); \xi_Q(g \cdot q) \tag{388}$$

$$= <\alpha_q; (T\Phi_{g^{-1}})\xi_Q(g \cdot q)> \equiv <\alpha_q; (T_{g \cdot q}\Phi_{g^{-1}} \circ \xi_Q \circ \Phi_g)(q) \tag{389}$$

$$= <\alpha_q; (\Phi_g^* \xi_Q)(q)> \tag{390}$$

$$= <\alpha_q; (Ad_{g^{-1}}\xi)_Q(q)> \tag{391}$$

$$= <\mathbf{J}(\alpha_q); Ad_{g^{-1}}\xi> = <Ad^*_{g^{-1}}(\mathbf{J}(\alpha_q)); \xi> \tag{392}$$

这里，我们将连续地应用：(i)式(385)；(ii)$g \cdot \alpha_q$ 是 $T^*(\Phi_{g^{-1}})(\alpha_q)$ 的简写，参见式(126)和式(121)；(iii)回撤的定义，参见式(172)；(iv)4.5.1 中式(167)的结论[2]；(v)式(385)；(vi)Ad^* 是 Ad 的共轭，参见式(191)。

6.5.3　例子

我们首先讨论一个相似的例子：线动量和角动量，即 6.3 小节中的(1)和(2)；G 上的左和右平移的余切提升——通过 4.6 小节在李群 G 上的运动学的描述产生的例子。

(1)　N 粒子总线动量

因为平动群$\mathbb{R}^3$ 通过 $\Phi(\mathbf{x}, (\mathbf{q}_i)) = (\mathbf{q}_i + \mathbf{x})$ 作用在 $Q := \mathbb{R}^{3N}$ 上，那么，Q 上的无穷小算子是：

$$\xi_{\mathbb{R}^{3N}}(\mathbf{q}_i) = (\xi, \cdots, \xi)(N\text{个}\,\xi) \tag{393}$$

应用式(385)，等变动量映射通过下式给出：

$$J(\xi)(\mathbf{q}_i, \mathbf{p}^i) = (\textstyle\sum_{i=1}^N \mathbf{p}^i) \cdot \xi, \text{ 即 } \mathbf{J}(\mathbf{q}_i, \mathbf{p}^i) = \sum_{i=1}^N \mathbf{p}^i \tag{394}$$

这和以前式(341)基于式(334)的微分方程的结论是一致的。

(2)　单粒子的角动量

$SO(3)$通过 $\Phi(A, \mathbf{q}) = A\mathbf{q}$ 作用在$\mathbb{R}^3$ 上。将 $\xi \in \mathfrak{so}(3)$写作 $\xi = \Theta\omega$，参见式(47)、式(51)和式(105)，无穷小算子是：

$$\xi_{\mathbb{R}^3}(\mathbf{q}) = \xi\mathbf{q} = \omega \wedge \mathbf{q} \tag{395}$$

因此，通过应用式(385)，等变动量映射 $\mathbf{J}: T^*\mathbb{R}^3 \to \mathfrak{so}(3) \cong \mathbb{R}^3$ 通过下式给出：

$$<\mathbf{J}(\mathbf{q}, \mathbf{p}); \omega> = <\mathbf{p}; \omega \wedge \mathbf{q}> = \mathbf{p} \cdot (\omega \wedge \mathbf{q})$$

$$= \omega \cdot (\mathbf{q} \wedge \mathbf{p}), \text{ 即 } \mathbf{J}(\mathbf{q}, \mathbf{p}) = \mathbf{q} \wedge \mathbf{p} \tag{396}$$

这和以前式(347)的基于微分方程式(334)的结论是一致的。

(3) G 上左和右平移的余切提升

回想起式(157)，即左平移无穷小算子是：

$$\xi_G(g) = (T_e R_g)\xi \tag{397}$$

这是一个不变的标量场，我们应用式(385)可以得到，对于左平移的余切提升，动量映射 $\mathbf{J}_L$：$T^* G \to \mathfrak{g}^*$ 通过下式给出：

$$<\mathbf{J}_L(\alpha_g);\ \xi> = <\alpha_g;\ \xi_G(g)> = <\alpha_g;\ (T_e R_g)\xi> = <(T_e^* R_g)(\alpha_g);\ \xi> \tag{398}$$

这里最后一个方程应用了式(121)中余切提升的定义，即等变动量映射是：

$$\mathbf{J}_L(\alpha_g) = T_e^* R_g(\alpha_g) \tag{399}$$

用文字表达，则为：左平移的余切提升的动量映射 $\mathbf{J}_L$ 是右平移的余切提升。

按相同的方式，我们可以认为，右平移 R_g：$h \longmapsto hg$。右平移定义了 G 上的一个右作用量，$\xi_G(g) = (T_e L_g)\xi$ 作为它的无穷小算子。因此，我们将：

$$\mathbf{J}_R:\ T^* G \to \mathfrak{g}^*;\ \mathbf{J}_R(\alpha_g) := T_e^* L_g(\alpha_g) \tag{400}$$

作为它的余切提升的动量映射。这个动量映射是关于 Ad_g^* 等变的，正如式(191)之后讨论的那样，是右作用量。

7. 约化

7.1 导言

在最后这一节中，论题终于和之前在第 2 节中的主题结合起来了。正如 5.1 节所说的，我们将集中精力证明什么是现今被称为的李—泊松约化定理，即泊松流形的同构：

$$T^* G/G \cong \mathfrak{g}^* \tag{401}$$

这里 $T^* G$ 的商是通过被左平移作用在 G 自身的作用量的余切提升所除所得到的。

恰巧，这一章的主要来源，即[Abraham and Marsden，1978；Arnold，1989；Olver，2000；Marsden and Ratiu，1999]不包括这个结论的最直接证明。

因此，我把它放在7.2节。这个结论是由先前的四个主要结论产生的，其中的一个来自第5节，其他三个来自第6节。

有一个问题关系到左平移和右平移之间的“转换”以及它们的各种提升。这四个结论表明T^*G/G作为泊松流形是同构的，不是和带有式(255)和式(258)的李—泊松括号的$\mathfrak{g}^*$同构，而是和带有这个括号的负值的$\mathfrak{g}^*$同构，即满足：

$$\{F, H\}_-(x) := -\langle x; [\nabla F(x), \nabla H(x)]\rangle, \quad x \in \mathfrak{g}^* \tag{402}$$

但是，由于适当地增加负号，我们将(幸运地)不需要再现随着式(255)而来的李—泊松括号的全部讨论(读者自行练习)。

为防止产生歧义，式(258)中带有正的李—泊松括号的$\mathfrak{g}^*$，我们有时将其写作$\mathfrak{g}^*_+$；对于式(402)中带有负的李—泊松括号的$\mathfrak{g}^*$，我们有时将其写作$\mathfrak{g}^*_-$。

事实上，从现在开始我们对待右作用量和左作用量是相同的，尽管我们先强调后者。这将意味着，我们对待右不变矢量场(另一个右不变的概念将在7.3.1节中定义)和左不变矢量场(7.3.1节中定义对应的左不变的新概念)是同等重要的。事实上，我们已经看到下列内容将会是必要的：

(i) 4.4节的结论，即左平移无穷小算子是一个右不变矢量场，反之亦然，如式(157)，式(158)；

(ii) 在6.5.3节中的例子(3)的推论是：左平移的余切提升的动量映射$\mathbf{J}_L$是右平移的余切提升，见式(399)；

(iii) 右平移的余切提升的动量映射$\mathbf{J}_R$是左平移的余切提升，如式(400)。

因此，在7.2节的最后，我们将给出一个李—泊松约化定理的简短证明。但是(正如常见的那样)，最直接的证明不会给出关于这种情形的更多信息。因此，在7.3节中，我们将给出更多的信息，见马斯顿和拉蒂乌(1999)。然后在7.4节中，我们将讨论从T^*G到$\mathfrak{g}^*$的动力学约化(与泊松结构对应)。

最终，在7.5节中，我们将陈述另一个约化定理，呈现在辛流形而不是泊松流形的术语中——但是将使用几个来自第3节的概念，如自由作用量、适当作用量和各向同性群。但是我们不证明这个定理，我们提及它是为了强调我们先前的评论，这一章(不管它的篇幅)对这一主题只做了肤浅的研究。我们也讨

论了它和李—泊松约化定理的关系。

7.2 李—泊松约化定理

首先让我们从4.6.2节式(225)开始回顾，$\bar{\lambda}$：$T^*G \to G\times\mathfrak{g}^*$是等变映射，这个等变映射将作用在$T^*G$上的左平移的余切提升左作用量和作用在$G\times\mathfrak{g}^*$上的$G$作用量联系在一起，$G\times\mathfrak{g}^*$仅仅通过在第一个分量上的左平移给出。因此我们转而讨论商，并通过式(228)定义$\dot{\bar{\lambda}}$：$T^*G/G \to (G\times\mathfrak{g}^*)/G$，即：

$$\dot{\bar{\lambda}}: \mathrm{Orb}(\alpha) \equiv \{\beta \in T^*G \mid \beta = T^*L_{h^{-1}}(\alpha),\ 一些\ h\in G\} \longmapsto \tag{403}$$

$$\mathrm{Orb}(\bar{\lambda}(\alpha)) \equiv \{(hg,\ (T_e^*L_g)(\alpha)) \mid 一些\ h\in G\} \equiv$$

$$\{(h,\ (T_e^*L_g)\alpha) \mid 一些\ h\in G\}$$

当$\alpha\in T_g^*G$时，$T^*L_{h^{-1}}\alpha \in T_{hg}^*G$。最终，我们用$\mathfrak{g}^*$定义$(G\times\mathfrak{g}^*)/G$，使得微分同胚映射$\dot{\bar{\lambda}}$将$T^*G/G$映射到$\mathfrak{g}^*$，正如在式(229)中的那样：

$$\dot{\bar{\lambda}}: \mathrm{Orb}(\alpha) \equiv \{\beta \in T^*G \mid \beta = T^*L_{h^{-1}}(\alpha), \tag{404}$$

$$某些\ h\in G\} \in T^*G/G \longmapsto (T_e^*L_g)(\alpha) \in \mathfrak{g}^*$$

因此，现在我们将说明，微分同胚映射$\dot{\bar{\lambda}}$：$T^*G/G\to\mathfrak{g}^*$在式(283)(5.3.2节)的意义上是一个泊松映射，因此我们需要指出：

(i) T^*G/G是一个泊松流形；

(ii) $\dot{\bar{\lambda}}$将(i)中T^*G/G上的泊松结构映射到$\mathfrak{g}^*$上。事实上，正如我们在7.1节中所说的那样，$\dot{\bar{\lambda}}$映射到$\mathfrak{g}^*_-$上的泊松结构，即如式(402)给出的那样。

表面上看，可能存在一个关于(i)的明智选择，即关于如何定义T^*G/G上的泊松结构，以便确保(ii)成立，即使得$\dot{\bar{\lambda}}$符合泊松结构。但事实上，我们先前的工作给出了一个最重要的且显而易见的选择，即我们通过5.5节的泊松约化定理来使用T^*G/G上的泊松结构。结论可通过结合这个定理直接地显现出来，即第6节的三个结论：

(i) 等变动量映射是泊松的，见6.4.2节的式(365)；

（ii）　余切提升左作用量是一个等变动量映射，见6.5.2节的式(385)；

（iii）　G上的左和右平移的余切提升是$\mathbf{J}_L = T_e^* R_g$和$\mathbf{J}_R = T_e^* L_g$，见6.5.3节的式(399)和式(400)。

特别是，结合(i)—(iii)，我们推论出$\mathbf{J}_R = T_e^* L_g$对于Ad_g^*是等变化的，因此它关于$\mathfrak{g}^*$上的正李—泊松括号是泊松的。即在值域$\mathfrak{g}_-^*$，它是泊松的。

因此，我们有如下定理。

李—泊松约化定理　微分同胚映射$\dot{\bar{\lambda}}$：$T^* G/G \to \mathfrak{g}^*$：

$$\dot{\bar{\lambda}}: \mathrm{Orb}(\alpha) \equiv \{\beta \in T^* G \mid \beta = T^* L_{h^{-1}}(\alpha)$$
$$\text{某些 } h \in G\} \in T^* G/G \longmapsto (T_e^* L_g)(\alpha) \in \mathfrak{g}^* \tag{405}$$

是泊松的。

证明：首先，式(405)意味着我们有一个可交换的三角形。因为，对于π：$T^* G \to T^* G/G$，动量映射$\mathbf{J}_R$：$T^* G \to \mathfrak{g}^*$，$\alpha_g \longmapsto (T_e^* L_g)\alpha_g$等于$\dot{\bar{\lambda}} \circ \pi$：

$$T^* G \xrightarrow{\pi} T^* G/G \xrightarrow{\dot{\bar{\lambda}}} \mathfrak{g}^* \tag{406}$$

因为左平移是G的微分同胚映射，流形的任意微分同胚映射的余切提升到它的余切丛都是辛的，参见4.1.0.1小节的式(120)，5.5节的泊松约化定理适用于这一结论。即存在$T^* G/G$上的唯一的泊松结构，使得π是泊松的。我们也从式(385)、式(365)和式(400)得知，$\mathbf{J}_R = T_e^* L_g$对于$\mathfrak{g}^*$上式(402)的括号是泊松的。

我们能推导出$\dot{\bar{\lambda}}$是泊松的，即对于所有的$x \in T^* G/G$和所有的F以及所有F，$H \in \mathcal{F}(\mathfrak{g}_-^*)$，我们有：

$$(\{F, H\}_{\mathfrak{g}_-^*} \circ \dot{\bar{\lambda}})(x) = \{F \circ \dot{\bar{\lambda}}, H \circ \dot{\bar{\lambda}}\}_{T^*G/G}(x) \tag{407}$$

我们仅(按顺序)使用如下事实。

（i）　π是满射的，使得对于所有的$x \in T^* G/G$，都存在$\alpha_g \in T^* G$，且$x = \pi(\alpha_g) \equiv \mathrm{Orb}(\alpha_g)$；

（ii）　$\mathbf{J}_R = \dot{\bar{\lambda}} \circ \pi$；

（iii）　$\mathbf{J}_R$是泊松的；

(iv)π 是泊松的：

$$(\{F,\ H\}_{\mathfrak{g}^*_-}\circ\hat{\bar{\lambda}})(x)=\{F,\ H\}_{\mathfrak{g}^*_-}\circ(\hat{\bar{\lambda}}\circ\pi)(\alpha_g) \tag{408}$$

$$=\{F,\ H\}_{\mathfrak{g}^*_-}\circ\mathbf{J}_R(\alpha_g)=\{F\circ\mathbf{J}_R,\ H\circ\mathbf{J}_R\}_{T^*G}(\alpha_g) \tag{409}$$

$$=\{F\circ\hat{\bar{\lambda}},\ H\circ\hat{\bar{\lambda}}\}_{T^*G/G}(\pi(\alpha_g))\equiv\{F\circ\hat{\bar{\lambda}},\ H\circ\hat{\bar{\lambda}}\}_{T^*G/G}(x) \tag{410}$$

7.3 与 T^*G 上的辛结构结合：不变函数

我们转而给出关于通过李—泊松约化定理描述的情形的更多信息。基本观点是：$\mathfrak{g}^*$ 上的李—泊松括号和 T^*G 上的辛结构一致。这将通过两种方式确切地表达，第一种在前两小节中讨论，第二种在第三小节中讨论。第一种讨论将分为以下三步。

(i) 我们表明，$\mathfrak{g}^*$ 上的标量，即 $F\in\mathcal{F}(\mathfrak{g}^*)$ 是和 T^*G 上的标量是一一对应关系，T^*G 上的标量在左平移余切提升的轨道上是常量，被称为左不变函数，相似地，对于右平移余切提升也是如此(对应于右不变函数)；

(ii) 我们将通常的正则泊松括号用于这些左不变或右不变标量的 T^*G 上，并将这个括号限制在被认为是在 $e\in G$ 上的余切空间 T^*_eG 的$\mathfrak{g}^*$ 的范围内；

(iii) 我们表明，这个限制是$\mathfrak{g}^*$ 上的李—泊松括号：对于右不变函数的熟悉的正的李—泊松括号，和对于左不变函数的式(402)的新的负的李—泊松括号。

我们在 7.3.1 节中讨论步骤(i)和(ii)，这些步骤不涉及正和负的李—泊松括号之间的选择。但是 7.3.2 节的步骤(iii)将提及这个选择。它将是 6.4.2 节的结论的必然推论，即式(365)的等变动量映射是泊松映射(这是意料之中的，而且我们在 7.2 节中证明约化定理时使用了这个结论)。

在第三部分，我们使用不变函数来展现一个不同的意义，在这个意义下，$\mathfrak{g}^*$ 上的李—泊松括号和 T^*G 上的辛结构是一致的。也就是说，我们通过应用不变函数和动量函数的思想从 5.5 节的泊松约化定理衍生出$\mathfrak{g}^*$ 上的李—泊松括号。

7.3.1　T^*G 上的左不变和右不变函数

我们认为，如果对所有的 $g\in G$ 和所有的 $\alpha_g\in T_g^*G$，那么函数 F：$T^*G\rightarrow\mathbb{R}$ 就是左不变的：

$$(F\circ T^*L_g)(\alpha_g)=F(\alpha_g) \tag{411}$$

当 T^*L_g 是 L_g：$G\rightarrow G$ 的余切提升。相似地，对于所有的 $g\in G$，F：$T^*G\rightarrow\mathbb{R}$被称作右不变的：

$$(F\circ T^*R_g)=F \tag{412}$$

因此如果 F：$T^*G\rightarrow\mathbb{R}$是左不变或右不变，那么它就由在 $T_e^*G=\mathfrak{g}^*$ 中讨论的参数的值所决定。

因为任意的 $\alpha\in\mathfrak{g}^*$ 通过 $T^*L_{g^{-1}}\equiv(T^*L_g)^{-1}$映射到 T_g^*G 中的元素，当且仅当对于所有的 $g\in G$，它在不同的 T^*L_g 的轨道上是常量，这个函数才是左常量，即在左平移的余切提升的轨道上是常量。相似地，当且仅当这个函数在右平移的余切提升的轨道上是常量，那么它就是右不变的。

因此，左不变函数引起了商空间 T^*G/G 上完全定义的函数，由 7.2 节可知，在它的微分同胚副本$\mathfrak{g}^*$ 上也是完全定义的函数，对于右不变函数亦是如此。

但是，让我们暂时考虑 T^*G 上平滑的左不变函数而不是商空间上的归纳映射。我们将通过$\mathcal{F}_L(T^*G)$表示 T^*G 上所有的平滑左不变函数空间，相似地，我们也将通过$\mathcal{F}_R(T^*G)$表示 T^*G 上所有的平滑右不变函数空间。

从式(120)以后的讨论回忆起余切提升都是辛映射，即 T^*L_g 和 T^*R_g 都是 T^*G 上的辛映射，那么，$\mathcal{F}_L(T^*G)$和$\mathcal{F}_R(T^*G)$在 T^*G 上的正则泊松括号下是闭合的。因此，它们都是具有这个括号的李代数。

现在我们可以使用6.5.3 节中实例(3)中的动量映射 $\mathbf{J}_L$ 和 $\mathbf{J}_R$ 来延伸任意标量 F：$\mathfrak{g}^*\rightarrow\mathbb{R}$，即 $F\in\mathcal{F}(\mathfrak{g}^*)$到 T^*G 上的左不变或右不变标量。

因此，给定 F：$\mathfrak{g}^*\rightarrow\mathbb{R}$和 $\alpha_g\in T_g^*G$，我们通过下式定义 $F_L\in\mathcal{F}_L(T^*G)$：

$$F_L(\alpha_g):=(F\circ\mathbf{J}_R)(\alpha_g)\equiv(F\circ T_e^*L_g)(\alpha_g) \tag{413}$$

因此 F_L 是左不变的，而且被称作从$\mathfrak{g}^*$ 到 T^*G 的 F 的左不变延伸(left-invariant extension)：

我们类似地通过下式定义对于任意的 $F\in\mathcal{F}(\mathfrak{g}^*)$，$F_R\in\mathcal{F}_R(T^*G)$ 的右不变延伸(right-invariant extension)：

$$F_R(\alpha_g) := (F\circ \mathbf{J}_L)(\alpha_g) \equiv (F\circ T_e^* R_g)(\alpha_g) \tag{414}$$

然后下列映射：

$$F\in\mathcal{F}(\mathfrak{g}^*)\longmapsto F_L\in\mathcal{F}_L(T^*G)\text{和 }F\in\mathcal{F}(\mathfrak{g}^*)\longmapsto F_R\in\mathcal{F}_R(T^*G) \tag{415}$$

是矢量空间同构的(读者自行练习!)，它的逆恰恰限制在 $T_e^*G=\mathfrak{g}^*$ 上。

这完成了我们所说的步骤(i)和(ii)：描述 $\mathfrak{g}^*$ 上的标量和 T^*G 上的标量的对应，这些标量在左和右平移的余切提升的轨道上都是常量，且也考虑到这些标量的正则泊松括号，即李代数 $\mathcal{F}_L(T^*G)$ 和 $\mathcal{F}_R(T^*G)$。

7.3.2 恢复李—泊松括号

我们现在讨论步骤(iii)。我们表明在 T^*G 上的右/左不变函数正则泊松括号的限制，对被认为是在单位元素 $e\in G$ 上的余切空间 T^*G 的 $\mathfrak{g}^*$，这个限制就是正/负李—泊松括号。

因为式(415)的倒数是被限制到 $T_e^*G=\mathfrak{g}^*$，这足以表明，式(415)的映射是李代数同构。

恢复李—泊松括号 使用 $\mathfrak{g}^*$ 上的正李—泊松括号(写作 $\mathfrak{g}_+^*$)：$F\longmapsto F_R$ 是一个李代数同构。相似地，使用 $\mathfrak{g}^*$ 上的负李—泊松括号(写作 $\mathfrak{g}_-^*$)：$F\mapsto F_L$ 也是一个李代数同构。

即对于所有的 F，$H\in\mathcal{F}(\mathfrak{g}^*)$，有：

$$\{F, H\}_+ = \{F_R, H_R\}_{T^*G}|_{\mathfrak{g}^*};\ \{F, H\}_- = \{F_L, H_L\}_{T^*G}|_{\mathfrak{g}^*} \tag{416}$$

证明：考虑到 $\mathbf{J}_L$：$T^*G\to\mathfrak{g}^*\equiv\mathfrak{g}_+^*$，$\mathbf{J}_L=T_e^*R_g$。$\mathbf{J}_L$ 是一个等变动量映射。因此，根据6.4.2节的式(365)的结论，它是泊松的。即：

$$\{F, H\}_+\circ\mathbf{J}_L = \{F\circ\mathbf{J}_L, H\circ\mathbf{J}_L\}_{T^*G} = \{F_R, H_R\}_{T^*G} \tag{417}$$

式(417)在 $\mathfrak{g}^*$ 的限制给出了式(416)的第一个方程。

相似地可以证明，通过使用 $\mathbf{J}_R$：$T^*G\to\mathfrak{g}^*\equiv\mathfrak{g}_-^*$，$\mathbf{J}_R=T_e^*L_g$，第二个方程式是一个等变动量映射，而且也是泊松的。即：

$$\{F, H\}_-\circ\mathbf{J}_R = \{F\circ\mathbf{J}_R, H\circ\mathbf{J}_R\}_{T^*G} = \{F_L, H_L\}_{T^*G} \tag{418}$$

然后我们将式(418)限制在$\mathfrak{g}^*$上。

7.3.3 导出李—泊松括号

在7.2小节和前面两小节的讨论已经给出了李—泊松括号。我们现在表明，使用不变函数和6.5.1小节中动量函数的观点是如何导出$\mathfrak{g}^*$上的李—泊松括号的。

因此，这个导出相当于李—泊松约化定理的另一个(更“有益”的)证明。正如7.2小节证明的那样，两个主要的要素是：

(a)T^*G/G和$\mathfrak{g}^*$之间的微分同胚映射$\bar{\lambda}$，参见式(229)或式(404)或式(405)；

(b)5.5节中被应用于T^*G上G的作用量的泊松约化定理。

但是，7.2节的证明采用了这样的事实：(i)动量映射$\mathbf{J}_R \equiv T_e^*L_g$和$\mathbf{J}_L \equiv T_e^*R_g$都是等变化的；(ii)等变动量映射是泊松的，我们现在将应用不变函数和动量函数的观点。

我们首先回想(因为左平移是G的微分同胚映射，且流形的任意微分同胚映射的余切提升到它的余切丛都是辛的)，泊松约化定理意味着存在T^*G/G上唯一的泊松结构，使得π：$T^*G \to T^*G/G$是泊松的。我们现在使用微分同胚映射$\bar{\lambda}$：$T^*G/G \to \mathfrak{g}^*$将这个泊松结构转移到$\mathfrak{g}^*$，让我们将这个结果称为$\{,\}_-$。尽管它不被叫作负李—泊松括号，但我们的目的是核实它就是这个括号。

首先，注意因为动量映射$\mathbf{J}_R$：$T^*G \to \mathfrak{g}^*$，所以$\alpha_g \mapsto (T_e^*L_g)\alpha_g$等于$\bar{\lambda} \circ \pi$(式406)，我们知道$\mathbf{J}_R$是关于$\mathfrak{g}^*$上的这个括号，是泊松的。即：

$$\{F,\ H\}_- \circ \mathbf{J}_R(\alpha_g) = \{F \circ \mathbf{J}_R,\ H \circ \mathbf{J}_R\}_{T^*G}(\alpha_g) = \{F_L,\ H_L\}_{T^*G}(\alpha_g) \quad (419)$$

为了计算式子的右边，我们将不变函数和动量函数的观点应用到括号的每个论证，尤其是第一个：

$$F_L(\alpha_g) = F(T_e^*L_g \cdot \alpha_g) \quad (420)$$

我们注意到，因为泊松括号仅仅依赖于一阶导数的值，我们可以用它的线性化来取代$F \in \mathcal{F}(\mathfrak{g}^*)$。即我们可以假设$F$是线性的，因此在任意点$\alpha \in \mathfrak{g}^*$，都有$F(\alpha) = \langle \alpha;\ \nabla F \rangle$，其中$\nabla F$在$\mathfrak{g} = \mathfrak{g}^{**}$是常量。我们将这个结论和式

(369)中的动量函数的定义应用于式(420)，得到：

$$F(T_e^* L_g \cdot \alpha_g) = <T_e^* L_g \cdot \alpha_g;\ \nabla F> = <\alpha_g;\ T_e L_g \cdot \nabla F> = \mathcal{P}(X_{\nabla F})(\alpha_g) \tag{421}$$

当最后的方程将动量函数的定义应用于 G 上的左不变矢量场时，对于 $\xi = \nabla F$，我们有 $X_\xi(g) \equiv T_e L_g(\xi)$。

现在依次将式(374)的李代数括号的定义和式(369)的左不变矢量场的定义应用于式(421)，我们将得到：

$$\{F_L,\ H_L\}_{T^*G}(\alpha_g) = \{\mathcal{P}(X_{\nabla F}),\ \mathcal{P}(X_{\nabla H})\}_{T^*G}(\alpha_g) = -\mathcal{P}([X_{\nabla F},\ X_{\nabla H}])(\alpha_g) \tag{422}$$

$$= -\mathcal{P}(X_{[\nabla F, \nabla H]})(\alpha_g) = -<\alpha_g;\ X_{[\nabla F, \nabla H]}> \tag{423}$$

$$= -<\alpha_g;\ T_e L_g([\nabla F,\ \nabla H])> = -<T_e^* L_g(\alpha_g);\ [\nabla F,\ \nabla H]> \tag{424}$$

结合式(419)和式(424)，并对 $(T_e^* L_g)\alpha_g \equiv \mathbf{J}_R(\alpha_g)$ 写出 $\alpha \in \mathfrak{g}^*$，我们得到结论：

$$\{F,\ H\}_-(\alpha) = -<\alpha;\ [\nabla F,\ \nabla H]> \tag{425}$$

类似地，我们通过考虑线性函数的右不变延伸来导出正李—泊松括号。来自式(374)的负号通过在右不变矢量场的李括号的符号反转而取消了，即它被来自式(86)的负号取消了。

7.4 动力学约化

我们以讨论从 $T^* G$ 到 $\mathfrak{g}^*$ 的动力学约化来结束李—泊松约化定理。

当讨论泊松约化定理的时候，我们可以简要概述，因为之前已经陈述了它的主要观点，参见 5.5 节的(2)(A)。因此，在该实现条件下回顾时，泊松流形 M 上 G 不变的哈密顿函数 H: $M \to \mathbb{R}$ 通过 $H = h \circ \pi$ 定义了 M/G 上的对应函数 h，其中 π 是投射 π: $M \to M/G$，且因为 π 是泊松的，因此 π 将哈密顿流推向哈密顿流，π 将 M 上的 X_H 推向 M/G 上的 X_h：

$$T_\pi \circ X_H = X_h \circ \pi \tag{426}$$

尤其是式(426)。将其应用于李—泊松约化定理，我们得到下面的内容。

动力学约化 假设 H: $T^* G \to \mathbb{R}$ 是左不变的。即，$\mathfrak{g}^*$ 上的函数 $H^- := H|_{\mathfrak{g}^*}$ 满

足：

$$H(\alpha_g)=H^-(\mathbf{J}_R(\alpha_g))\equiv H^-(T_e^*L_g\cdot\alpha_g),\ \alpha_g\in T_g^*G \tag{427}$$

那么 $\mathbf{J}_R$ 将 X_H 推向 X_{H^-}。或者分别依照 X_H 和 X_{H^-} 的流 $\phi(t)$ 和 $\phi^-(t)$：

$$\mathbf{J}_R(\phi(t)(\alpha_g))=\phi^-(t)(\mathbf{J}_R(\alpha_g)) \tag{428}$$

相似的陈述适用于右不变函数 H: $T^*G\to\mathbb{R}$，它的限制为 $H^+:=H|_{\mathfrak{g}^*}$ 和 $\mathbf{J}_L\equiv T_e^*R_g$。

此外，我们已经知道$\mathfrak{g}^*$上的矢量场 H^-。因为5.2.4小节中(3)的式(261)根据 ad^* 给出了$\mathfrak{g}^*$上哈密顿方程的基独立表示。需要注意，因为我们现在使用$\mathfrak{g}^*$上的负李—泊松括号，除了式(261)的右边，所有的项都得到一个负号。因此，写作 $\alpha\in\mathfrak{g}^*$，对于矢量场 X_{H^-}，式(261)变为：

$$\frac{d\alpha}{dt}=-ad^*_{H^-(\alpha)}(\alpha) \tag{429}$$

另一方面，我们可以转向另一方向。从式(429)中$\mathfrak{g}^*$上的动力学重建 T^*G 上的动力学，下列主要结论是直观的陈述，在这个陈述中，对于 $g(t)\in T^*G$ 的“重建方程”是：

$$g^{-1}\dot{g}=\nabla H^- \tag{430}$$

这是直观的，因为它将我们回归到$\mathfrak{g}$和$\mathfrak{g}^*$上的力学基本观点，即映射：

$$\lambda_g:\ \dot{g}\in T_gG\mapsto\lambda_g(\dot{g}):=(T_gL_{g^{-1}})\dot{g}\in\mathfrak{g} \tag{431}$$

将广义速度映射到它的体表示，参见式(205)。然而，这个结果的证明也包括在内[Marsden and Ratiu，1999，Thm. 13.4.3，Thm. 13.4.4，pp. 423—426]，因此，我们只陈述结果。

动力学重建　假设给定一个李群 G，一个左不变量 H: $T^*G\to\mathbb{R}$，它的限制为 $H^-:=H|_{\mathfrak{g}^*}$，且给定一下$\mathfrak{g}^*$上的式(429)的李—泊松哈密顿函数的积分曲线 $\alpha(t)$，并伴随有初始条件 $\alpha(0)=T_e^*L_{g_0}(\alpha_{g_0})$。那么在 X_H 的 T^*G 的积分曲线通过下式给定：

$$T^*_{g(t)}L_{g(t)^{-1}}(\alpha(t)) \tag{432}$$

其中 $g(t)$ 是重建方程：

$$g^{-1}\dot{g}=\nabla H^- \tag{433}$$

的解，初始条件为 $g(0)=g_0$。

7.5 结束语：马斯登—温斯坦—迈耶定理

我强调，我们对于约化的讨论只做了肤浅的研究，毕竟这小节相当短！但是现在读者已经有了前面第 3 节较长的阐述，给出这些已经准备好继续约化的主题，例如，这一章的主要来源[Abraham and Marsden，1978；Arnold，1989；Olver，2000；Marsden and Ratiu，1999]。

尤其是，读者现在可以将李—泊松约化定理和其他的关于辛约化的主要定理联系起来，通常称作马斯登—温斯坦—迈耶定理或者马斯登—温斯坦定理（见这些作者在 1973 年和 1974 年的论文）。

这个定理关系到辛流形(M, ω)上李群 G 的辛作用量。为了完备性起见，以及使读者适应这个定理（见本卷第五章，尤其是 4.5 小节）的讨论，很值得结合用来证明它的引理和接着出现的动力学约化一起陈述它（像往常一样，仅仅对于有限维情况）。这些陈述也将通过阐释前面节 3 的一些概念是如何解释的完成我们的讨论。但是，这将不应用在这小节，不过，这仍然是有作用的——例如在陈述这个定理的假设时。

因此，假设李群 G 辛地作用即式(311)在辛流形(M, ω)上；$\mathbf{J}: M \to \mathfrak{g}^*$ 是对于这个作用量的 Ad^* 等变动量映射，见式(357)和式(362)。假设 $\alpha \in \mathfrak{g}^*$ 是 $\mathbf{J}$ 的正则值，即在任意点 $x \in \mathbf{J}^{-1}(\alpha)$处，$T_x\mathbf{J}$ 都是满射的。因此，3.3.1 节中(1)的浸没定理是适用的，尤其，$\mathbf{J}^{-1}(\alpha)$是维度为 $\dim(M)-\dim(\mathfrak{g}^*)\equiv\dim(M)-\dim(G)$的 M 上的子流形。

假设 G_α 是 α 在伴随作用量下的各向同性群，见式(148)，即：

$$G_\alpha := \{g \in G \mid Ad^*_{g^{-1}}\alpha = \alpha\} \tag{434}$$

因此，如果 $\mathbf{J}$ 在 G_α 下是 Ad^* 等变化的，那么商空间 $M_\alpha := \mathbf{J}^{-1}(\alpha)/G_\alpha$ 就是定义明确的。

现在假设 G_α 自由准确地作用在 $\mathbf{J}^{-1}(\alpha)$上，使得商空间 $M_\alpha = \mathbf{J}^{-1}(\alpha)/G_\alpha$ 是流形，M_α 是约化的相空间（对应于动量值 α）。

马斯登—温斯坦—迈耶定理 M_α 有一个从如下所述的(M, ω)产生的固有

辛形式 ω_α。假设 u，v 是在某些点 $p\in M_\alpha$ 处正切于 M_α 的两个矢量，因此 p 是 G_α 作用在 $\mathbf{J}^{-1}(\alpha)$ 上的作用量的轨道，且 u，$v\in T_pM_\alpha$。那么 u 和 v 是通过投射 π_α：$\mathbf{J}^{-1}(\alpha)\to M_\alpha$ 分别从正切于在轨道 p 的某些点 $x\in\mathbf{J}^{-1}(\alpha)$ 的 $\mathbf{J}^{-1}(\alpha)$ 的矢量 u' 和 v' 得到的。即：

$$T\pi_\alpha(u')=u;\ T\pi_\alpha(v')=v \tag{435}$$

结论是，不论 x 的选择是什么，通过 M 的辛形式 ω 指定的 u' 和 v' 的值都是相同的。因此，我们定义 M_α 上的辛形式 ω_α 作为指定的值。换句话说，将投射写作 π_α，将包含(inclusion)写作 i_α：$\mathbf{J}^{-1}(\alpha)\to M$，将拉回(pull back)写作 *：

$$\pi_\alpha^*\omega_\alpha=i_\alpha^*\omega \tag{436}$$

这个定理的证明使用了下面的引理。假设将所有 G 的作用量下 x 的轨道 Orb(x)写作 $G\cdot x$，且相似地，G_α 下的轨道写作 $G_\alpha\cdot x$，即 $\{\Phi(g,x)\mid g\in G_\alpha\}$。

对于任意的 $x\in\mathbf{J}^{-1}(\alpha)$：

(i) $T_x(G_\alpha\cdot x)=T_x(G\cdot x)\cap T_x(\mathbf{J}^{-1}(\alpha))$；

(ii) $T_x(G\cdot x)$ 和 $T_x(\mathbf{J}^{-1}(\alpha))$ 是在 TM 里的相互正交补(orthogonal complement)。即，当且仅当对于所有的 $v'\in T_x(G\cdot x)$ 我们都能得到 $\omega(u',v')=0$，那么对于所有的 $u'\in T_xM$：$u'\in T_x(\mathbf{J}^{-1}(\alpha))$。

引理和定理都已被大量证明。详细内容参见[Abraham and Marsden，1978，Theorems 4.3.1—2，pp. 299—300；Arnold，1989，Appendix 5.B，pp. 374—376]。

最后两条备注：

(1)对于泊松约化定理和李—泊松约化定理，通过马斯登—温斯坦—迈耶定理得到的动力学约化和我们之前看到的一样。我们通过下列内容来证明，见[Abraham and Marsden，1978，Theorems 4.3.5，304]。

马斯登—温斯坦—迈耶动力学约化 假设 H：$M\to\mathbb{R}$在作用于 M 上 G 的作用量下是不变量。因此，通过诺特定理，动量映射 $\mathbf{J}$ 是守恒的，即 $\mathbf{J}^{-1}(\alpha)$ 在 M 上 X_H 的流 $\phi(t)$ 是不变量。$\phi(t)$ 和 $\mathbf{J}^{-1}(\alpha)$ 上 G_α 的作用量是对易的，即对于 $g\in G_\alpha$，$\phi(t)\circ\Phi_g=\Phi_g\circ\phi(t)$，因此，定义在 M_α 上的流 $\hat{\phi}(t)$，使得 $\pi_\alpha\circ\phi(t)$

$=\hat{\phi}(t)\circ\pi_\alpha$，即：

$$
\begin{array}{ccc}
\mathbf{J}^{-1}(\alpha) & \xrightarrow{\pi_\alpha} & M_\alpha \\
\downarrow{\scriptstyle \phi(t)} & & \downarrow{\scriptstyle \hat{\phi}(t)} \\
\mathbf{J}^{-1}(\alpha) & \xrightarrow{\pi_\alpha} & M_\alpha
\end{array}
\tag{437}
$$

其中流$\hat{\phi}(t)$是通过$H_\alpha\circ\pi_\alpha=H\circ i_\alpha$定义的哈密顿量为$H_\alpha$的哈密顿函数。

（2）我在这一小节的开始说过，读者可以将李—泊松约化定理和马斯登—温斯坦—迈耶定理联系起来。不难表明，前者是后者的例子。通过左平移的余切提升作用于辛流形M，我们得到T^*G。因此，我们知道，$\mathbf{J}_L := T_e^* R_g$是$Ad^*$等变动量映射……我将这个作为留给读者的习题！答案见[Arnold，1989，377，321]和[Abraham and Marsden，1978，302]，亚伯拉罕和马斯登称之为“基列洛夫—科斯坦特—苏里奥”(Kirillov-Kostant-Souriau)定理。

简单地说，这个习题给出了我们的中心思想其中之一的另一种说明，$\mathfrak{g}^*$的辛叶是余伴随表示的轨道。因为，约化相空间M_α和$\alpha\in\mathfrak{g}^*$的余伴随轨道Orb(α)是同一的，和辛形式也是同一的，参见5.4小节最后的结果(2)。

致谢

感谢尔湾市、牛津、普利斯顿和圣巴巴拉的读者，以及一些同仁的鼓励，他们是：戈登·贝洛特、克拉斯·兰兹曼(Klaas Landsman)、巴里·汤金森(Barrie Tonkinson)、大卫·华莱士(David Wallace)，尤其是格雷姆·西格尔(Graeme Segal)的帮助和耐心！以及和他们的交谈、通信及他们的修改。

参考文献

[Abraham and Marsden，1978] R. Abraham and J. Marsden. *Foundations of Mechanics*, second edition：Addison-Wesley，1978.

[Arnold，1973] V. Arnold. *Ordinary Differential Equations*, MIT Press，1973.

[Arnold，1989] V. Arnold. *Mathematical Methods of Classical Mechanics*, Springer，1989 (second edition).

[Barbour and Bertotti, 1982] J. Barbour and B. Bertotti. Mach's principle and the structure of dynamical theories, *Proceedings of the Royal Society of London* A 382, p. 295-306, 1982.

[Belot, 1999] G. Belot. Relationism rehabilitated, *International Studies in Philosophy of Science* 13, p. 35-52, 1999.

[Belot, 2000] G. Belot. Geometry and motion, *British Journal for the Philosophy of Science* 51, p. 561-596, 2000.

[Belot, 2001] G. Belot. The principle of sufficient reason', *Journal of Philosophy* 98, p. 55-74, 2001.

[Belot, 2003] G. Belot. Notes on symmetries, in [Brading and Castellani, 2003, pp. 393-412].

[Belot, 2003a] G. Belot. Symmetry and gauge freedom, *Studies in the History and Philosophy of Modern Physics* 34, p. 189-225, 2003.

[Belot, 2006] G. Belot (this volume), 2006.

[Belot and Earman, 2001] G. Belot and J. Earman. Pre-Socratic quantum gravity, in C. Callender and N. Huggett, eds. *Physics meets Philosophy at the Planck Scale*, Cambridge University Press, pp. 213-255, 2001.

[Bishop and Goldberg, 1980] R. Bishop and S. Goldberg. *Tensor Analysis on Manifolds*, New York: Dover, 1980.

[Boudri, 2002] J. Boudri. *What was Mechanical about Mechanics: the Concept of Force between Metaphysics and Mechanics from Newton to Lagrange*, Dordrecht: Kluwer Academic, 2002.

[Brading and Castellani, 2003] K. Brading and E. Castellani, eds. *Symmetry in Physics*, Cambridge University Press, 2003.

[Brading and Castellani, 2006] K. Brading and E. Castellani (this volume), 1006.

[Butterfield, 2004] J. Butterfield. Some Aspects of Modality in Analytical mechanics', in *Formal Teleology and Causality*, ed. M. Stöltzner, P. Weingartner, Paderborn: Mentis, 2004. Available at Los Alamos arXive: http://arxiv.org/abs/physics/0210081 ; and at Pittsburgh archive: http://philsci-archive.pitt.edu/archive/00001192.

[Butterfield, 2004a] J. Butterfield. Between Laws and Models: Some Philosophical Morals of Lagrangian Mechanics, 2004; available at Los Alamos arXive: http: //arxiv. org/abs/physics/0409030 ; and at Pittsburgh archive: http: //philsciarchive. pitt. edu/archive/00001937/.

[Butterfield, 2005] J. Butterfield. Between Laws and Models: Some Philosophical Morals of Hamiltonian Mechanics, in preparation, 2005.

[Butterfield, 2006] J. Butterfield. On Symmetry and Conserved Quantities in Classical Mechanics, forthcoming in a *Festschrift* for Jeffrey Bub, ed. W. Demopoulos and I. Pitowsky, Kluwer: University of Western Ontario Series in Philosophy of Science, 2006. available at Los Alamos arXive: http: //arxiv. org/abs/physics/ ; and at Pittsburgh archive: http: //philsciarchive. pitt. edu/archive/00002362/

[Darling, 1994] R. Darling. *Differential Forms and Connections*, Cambridge University Press, 1994.

[Dickson, 2006] M. Dickson (this volume), 2006.

[Desloge, 1982] E. Desloge. *Classical Mechanics*, John Wiley, 1982.

[Earman, 2003] J. Earman. Tracking down gauge: an ode to the constrained Hamiltonian formalism, in [Brading and Castellani, 2003, pp. 140-162].

[Goldstein, 1950] H. Goldstein. *Classical Mechanics*, Addison-Wesley, 1950; (1966 third printing).

[Grattan-Guinness, 2000] I. Grattan-Guinness. A sideways look at Hilbert's twenty-three problems of 1900, *Notices of the American Mathematical Society* 47, p. 752-757, 2000.

[Grattan-Guinness, 2006] I. Grattan-Guinness. Classical mechanics as a formal(ised) science, in B. Loewe (ed.), *Foundations of the Formal Sciences*, Kluwer, 2006.

[Hawkins, 2000] T. Hawkins. *Emergence of the Theory of Lie Groups: an essay in the history of mathematics 1869—1926*, New York: Springer, 2000.

[Jammer, 1957] M. Jammer. *Concepts of Force*, Harvard University Press, 1957.

[Jammer, 1961] M. Jammer. *Concepts of Mass in Classical and Modern Physics*, Harvard University Press, 1961; republished by Dover in 1997.

[Jammer, 2000] M. Jammer. *Concepts of Mass in Contemporary Physics and Philosophy*, Princeton University Press, 2000.

[Johns, 2005] O. Johns. *Analytical Mechanics for Relativity and Quantum Mechanics*, Oxford University Press, 2005.

[Kirillov, 1976] A. Kirillov. *Elements of the Theory of Representations*, Grunlehren Math. Wiss., Springer-Verlag, 1976.

[Kragh, 1999] H. Kragh. *Quantum Generations*, Princeton University Press, 1999.

[Landsman, 2006] N. Landsman (this volume), Between Classical and Quantum.

[Lie, 1890] S. Lie. *Theorie der Transformationsgruppen: zweiter abschnitt*, Leipzig: B. G. Teubner, 1890.

[Lutzen, 1995] J. Lutzen. *Denouncing Forces; Geometrizing Mechanics: Hertz's Principles of Mechanics*, Copenhagen University Mathematical Institute Preprint Series No 22, 1995

[Lutzen, 2003] J. Lutzen. Between rigor and applications: developments in the concept of function in mathematical analysis, in *Cambridge History of Science, vol. 5: The modern physical and mathematical sciences*, ed. M. J. Nye, p. 468-487, 2003.

[Lutzen, 2005] J. Lutzen. *Mechanistic Images in Geometric Form: Heinrich Hertz's 'Principles of Mechanics'*, Oxford University Press, 2005.

[Marsden and Hughes, 1982] J. Marsden and T. Hughes. *Mathematical Foundations of Elasticity*, Prentice-Hall; Dover, (1982) 1994.

[Marsden and Ratiu, 1999] J. Marsden and T. Ratiu. *Introduction to Mechanics and Symmetry*, second edition: Springer-Verlag, 1999.

[McMullin, 1978] E. McMullin. *Newton on Matter and Activity*, University of Notre Dame Press, 1978.

[Newton, 1687] I. Newton. *Principia: Mathematical Principles of Natural Philosophy*, 1687. Trans. I. B. Cohen and A. Whitman, Cambridge University Press, 1999.

[Olver, 2000] P. Olver. *Applications of Lie Groups to Differential Equations*, second edition: Springer-Verlag, 2000.

[Rovelli, 2006] C. Rovelli (this volume), Quantum Gravity.

[Ryckman, 2005] T. Ryckman. *The Reign of Relativity: Philosophy in Physics 1915—1925* Oxford University Press, 2005.

[Singer, 2001] S. Singer. *Symmetry in Mechanics: a Gentle Modern Introduction*, Boston: Birkhauser, 2001.

[Slovik, 2002] E. Slovik. *Cartesian Spacetime: Descartes' Physics and the Relational Theo-*

ry of Space and Motion, Dordrecht: Kluwer Academic, 2002.

[Stöltzner, 2003] M. Stöltzner. The Principle of Least Action as the Logical Empiricist's Shibboleth, *Studies in History and Philosophy of Modern Physics* 34B, p. 285-318, 2003.

[Wallace, 2003] D. Wallace. Time-dependent Symmetries: the link between gauge symmetries and indeterminism, in [Brading and Castellani, 2003, pp. 163-173].

第二章　力学中时间和变化的表征

戈登·贝洛特

如果时间是客观的，那么物理学家们一定已经发现了这个事实；如果时间有一个起点，那么物理学家一定了解它；但如果时间仅仅是主观的，而存在是永恒的，那么物理学家在对现实世界进行建构时一定能忽略时间，在描述世界时也不需要时间的帮助……如果关于时间的哲学问题有一个解答，那么它是由数学物理方程记录下来的。

也许一种更好的说法是，这一解答隐藏在物理学家著作的字里行间。物理方程表达出了特定的规律……但哲学分析关心的是对方程的陈述，而不是方程的内容本身。

——莱辛巴赫(Reichenbach)①

多年来，我一直深受一个事实的折磨，那就是：最惊人的发现在时间的疆域内等待着我们。关于时间我们所知道的比其他任何事物都要少。

——塔尔科夫斯基(Tarkovsky)②

① [Reichenbach，1991，16f.]。
② [Tarkovsky，1991，53]。

1. 引言

本文关注的是经典(非量子的)物理理论中时间和变化的表征。本章的主要目标之一是试图澄清所谓时间问题的性质和范围，当我们试图量子化广义相对论时，就会出现一些技术性和解释性问题，这些都源于这一理论的广义协变性。

解决这些问题的最常见方法是考虑一些更清晰的例子。本章的大部分内容是用来讨论其他更好理解的理论中时间和变化的表征时，一开始探讨最简单的例子，然后考虑其他例子，这些例子分别在不同的意义上为我们展示了广义相对论的特征，而正是广义相对论提出了关于时间的问题。

请允许我首先说一下，当谈到物理理论中时间和变化的表征时，出现在我脑海中的是，应该从最容易处理的情况(即牛顿物理学)开始讨论。

作为一个完美的一般性理论，关于物理理论内容的许多问题和断言需要接受两方面的问题——即接受关于理论中运动方程解的结构特性的问题以及关于这些方程的结构特性的问题。举例来说，一方面时间在物理学中表现为时空的一个维度——也就是说，作为为求得理论方程的解而设定的背景的一个方面；另一方面，时间通过它在物理定律中的作用得以体现——尤其是通过它在表示这些定律的微分方程中的作用。因此物理理论关于时间性质的问题和断言需要两类解读。

比如，考虑这样一种说法：时间在牛顿物理学中是均匀的(或者如牛顿的设定，时间是均匀流逝的)。基于这个断言我们将面对两方面的事实。

1. 在一种意义上，时间是牛顿物理学时空中一个可分离的部分；在另一种意义上，正如我们所考虑的，时间是均匀的。①

2. 经典力学基础理论(比如牛顿的重力理论)的定律是时间平移不变

① (新)牛顿时空以自然的方式由具有绝对同时性的点时间来分割，时间可以等同于这些瞬间集从时空结构中延承下来的结构，因此时间具有可以被实数模型化的仿射空间的结构——所以对于任意两个时刻，存在由此映射到彼的时间对称性。

的——当时空坐标原点改变后，这些理论微分方程的形式不会改变——这类理论的定律不关注时间瞬间的同一性。

在牛顿的设定中，这两类观点配合得很好并且相互提供支持，即定律的对称性和时空的对称性之间具有一致性。但在原理上，这两种观点不必然导致同样的回答，我们可以认为牛顿时空中的系统是受随时间变化的力所支配的；或者我们可以将牛顿 n 体问题设定在一个具有瞬时优先特征但在其他方面又具有牛顿时空结构的时空中。而且，随着我们远离熟悉的牛顿物理学假设，一种更重要的区分这两种观点的方法就变为：在广义相对论中，定律具有一个巨大（事实上是无穷维）的对称性群，而通解则没有任何对称性。

在讨论时间的表征和变化时，本章将重点放在物理理论定律的结构特点而不是特解的特征上。为了强调这一点，我会说，可以将这个或那个理论的结构看作是动力学理论。

我将以经典理论的拉格朗日（Lagrangian）方法和哈密顿（Hamiltonian）方法来开始我的主题，它们是经典理论的两个支柱性的且密切相关的框架，这些主题在这个框架下可以很自然地开展。①粗略地说，在这两种方法中，一个理论方程的内容被编码为某些特定结构，这些结构建立在与理论相关的概率空间上。②在拉格朗日方法中，特征空间是理论方程的解空间，为了实现启发式的目的，我们可以将其与理论允许的可能世界的空间等同起来。③在哈密顿方法中，特征空间是理论方程的初始数据空间。出于同样的目的，我们可以将其等同为理论

① 为什么我们要将问题放在拉格朗日和哈密顿力学领域内讨论，而不是直接探讨理论的微分方程呢？因为这带来的好处是巨大的：这些基础性的方法为比较理论提供了强大的数学框架。另一个原因是这样做的成本是微乎其微的，因为几乎我们感兴趣的每一个理论都可以表示为拉格朗日或哈密顿形式，其内容没有任何明显的变化。而且，还因为它能引导我们实现我们的目的，即理解经典物理理论内容的尝试必然会受到构造和理解更深入的量子理论的努力的影响，现在看来经典理论必须在拉格朗日和哈密顿形式下才能被量子化。

② ［Tarkovsky，1991，53］中谈到了这里涉及的概率。

③ 在通常的经典力学背景下，人们往往认为拉格朗日力学是构造在速度相空间中的——因此它更紧密地与初始数据空间而不是解空间相关。然而，这个熟悉的方法需要一个绝对的同时性概念，基于这个原因，当人们转向探讨相对论时，通常会采用时空协变拉格朗日方法（在这种方法下初始数据空间不起任何作用）。这是接下来将会采用的观点。

允许的可能瞬时状态的空间。

在牛顿力学中，定律的时间平移不变性在拉格朗日框架下表现为解空间自身在时间平移下是不变的。给定时空中一个服从牛顿运动定律的粒子运动轨迹的集，所有事件的时间变量的变化为 t，那么我们就可以构造这个结果的粒子运动轨迹集；后一个集为一个解(即是符合运动定律的)，当且仅当前一个集为一个解。此外，从一个解到其自身时间平移的映射保持编码动力学理论的解空间结构不变。另一方面，在哈密顿框架下，定律的时间平移不变性还反映在存在这样的一个映射，它将一个初始数据集赋予在时间单位 t 内会演化成的那个态；然后这个映射保持编码动力学理论空间上的结构不变。所以动力学理论的时间对称性在拉格朗日这边由时间平移的概念反映出来，而在哈密顿那边则由时间演化的概念反映出来。

牛顿物理学中变化的表达在拉格朗日框架和哈密尔顿框架内也分别采用了不同(但密切相关)的形式。变化存在于一个在不同时间具有不同且不相容属性的系统中。我们想要说，比如在一个两体系统的可观测属性中，存在一个变化，当且仅当粒子之间的相对距离随时间的不同而不同。

哈密顿方法。指定这样一个系统的瞬时动力学态足以确定粒子之间的瞬时相对距离。因此对应于该量，存在一个初始数据空间上的函数。该系统的历程是穿过初始数据空间的一条轨迹。在我们的简单例子中，可观测的变化发生在一个给定的历程中，当且仅当对应于粒子之间相对距离的函数在论及的轨迹的不同点上取不同的值。更一般地，在任意牛顿系统中，如果相应的函数在轨迹上的不同点取不同值的话，任何物理上感兴趣的(可观测的或不可观测的)量由初始数据空间上的函数表示，该空间上的轨迹就将这些量表示为变化。

拉格朗日方法。显然，解空间上没有函数能像初始数据空间上的函数那样用直接的方式表示变量。但是对于每个 t，在我们两体问题的解空间上存在一个函数，能在时刻 t 根据解来为每个解赋予粒子之间的相对距离。令 t 取不同值，我们在解空间上构造了一个函数的单参数族。运动方程的一个解将函数间的相对距离表示为变化，当且仅当函数的这个单参数族的不同元素在求一个给定解的值时取不同的值。以此类推，可以更普遍地说，任何物理变量都对应于这样一个解空间上的单参数族，而变化就如在上述简单两体例子中那样理解。

谈到物理理论中时间和变化的表征，我脑海里会出现很多诸如此类的想法。在介绍本章关于这些主题的研究路线之前，多介绍一下其最终目标——澄清所谓时间问题的本质——也许会对读者有所帮助。关于时间问题的讨论通常会集中于广义相对论的哈密顿版本，其中的重点是可能的瞬时几何结构的空间，即柯西(Cauchy)面上的度规和第二基本形式。多少有点不幸的是，这种方法从一开始就要求将时空划分为一个类空超曲面族，这似乎与通常理解的理论的广义协变性相悖。鉴于这一事实，我们有可能会担心通常表现出的时间问题是以非常难操作的方式处理问题导致的结果。这里我采取多少有些不同的方法，我总是用拉格朗日方法进行研究，它讨论的是系统的基本完整历程而不是瞬时状态。

粗略地讲，下面展开的观点是：时间问题的核心在于，如果在广义相对论中从动力学的角度理解，我们没有办法将时间演化或时间平移看作理论的对称性，并且相关地，在拉格朗日和哈密顿的方法中，没有一种自然的方式来通过空间的函数将变化模型化。①这标志着在这个方面广义相对论与之前的基本理论是非常不同的，而它也正是这样被构想出来的。

关于时间的问题听起来可能并不十分紧迫，但可以肯定的是，这里确实存在难题。广义相对论在这方面为什么与之前的理论如此不同呢？在广义相对论之前的理论中，时间的表征和变化的表征是紧密联系在一起的——那么广义相对论是如何替代这种联系的呢？这些都是有意义的问题。但后来当然谁也不会期望在广义相对论中时间的表征和在之前理论中的一样——它提出的关于时间

① 上述公式只给出了一级近似，主要基于以下几个原因(每个原因都会在下面的各节中更充分地讨论)。(i) 时间问题只出现在最合适的宇宙学设定的广义相对论的那些版本中，在其他理论中，时间的表示方式与它在狭义相对论中的表示方式非常相似。(ii) 在一般的与时间相关的系统中，时间演化和时间平移不是理论的守恒量——但这并不会导致在这些理论中会出现关于时间和变化的表征的任何实质性问题，因为我们仍然有群操作来实现时间演化和时间平移，即使它们并不是定律的守恒量，这足以保证这些变化会与在时间无关的理论中的变化非常相似。(iii) 在那些解不是全域性地定义在时间之上的理论中，时间演化和时间平移不能由群操作来实现，而仅由局域流(可以被看作是无穷小的群操作)来实现。这足以满足在这类理论中构造关于变化的熟悉的图景，然而甚至这些在广义相对论中也是不存在的。

和空间的全新图景是该理论的成就之一。我们可能也会认为，既然在广义相对论中时空结构随解的不同而不同，那么如果我们想要明白在关于我们世界中时间的性质上理论能告诉我们什么的话，从这样或那样的物理现实解的角度而不是从理论方程的角度来理解时间的表征肯定会更合适。

但是，当考虑到广义相对论的量子化(或相关意义上的任何其他的一般协变理论)时，时间问题假定了一个更加紧迫的方面。构建后续理论这项事业，自然将我们的注意力集中到当前理论的结构特征上——在构造后续理论时，人们需要做出选择，当前理论的哪些特征能够沿用(也许会换一种新的形式)，哪些需要被抛弃。并且已知的量子化技术不只是要求将微分方程作为输入，而且要求将理论表示为哈密顿或拉格朗日形式。因此对于那些对广义相对论的量子化有兴趣的人，有关作为动力学理论的那些理论结构的问题，当然会显得尤为突出。由于上述难题缺乏解决方案，人们预计在构造广义相对论的量子化(或由其进行预测)时会遇到概念性困难。因此，从这个角度看，时间问题其实是相当紧迫的。

本文关于时间问题探讨得很多。我首先在第 2 节中简单介绍哈密顿和拉格朗日力学并将此作为引出后面内容的方式。在第 3 节中，介绍了辛几何(Symplectic geometry)，经典力学领域数学基础的一些重要概念和结果。这里所介绍的概念对接下来的内容是至关重要的：对于好的理论，解空间(拉格朗日方面)和初始数据空间(哈密顿方面)都具有辛结构。而且我们将看到，当我们偏离理想情况时，各种辛(近似辛)空间就会出现。在第 4 节中，我将会概述现代拉格朗日力学的非常强大的框架及其局域守恒定律的结构。

在第 5 节中，我将概述满足下列条件的理想理论的拉格朗日和哈密顿图景：(i) 时空几何体系的背景承认时间变换群和理论的拉格朗日形式在这个群的作用下(在适当意义上)是不变的；(ii) 为理论方程指定初始数据足以确定一个唯一的最大解；(iii) 这个最大解由时间参数的所有值来定义，当这些条件不变时，我们发现拉格朗日方法存在一个解空间上的时间平移对称操作群，而哈密顿方法则存在一个初始数据空间上实现时间演化的群。这两个空间是同构的，且两个群操作以令人满意的方式紧密地联系在一起。我们能够给出一个简单的且有吸引力的方法，用以在这两个基本空间中表示变化。

在第 6 节中我转向讨论一些难题。当我们去掉上面条件(i)—(iii)中的任一条时，这些难题就必然会出现。最后在第 7 节，我将会讨论广义相对论中时间和变化的表征，这直接导致了时间问题。

正如这个提纲明确指出的，本章大部分内容着墨于技术性资料的阐释。为了保证文章在合理的篇幅之内，我不得不假定本章的读者具有相当多的技术背景。我试图为一个理想的读者进行阐述，他需要研究过广义相对论或规范理论，并因此熟悉微分几何的基本概念、结论和结构(虽然在一些关键点上我所做的讨论对唤醒这类读者的记忆有很重要的作用)。

本章是建立在拉格朗日力学的现代几何学方法之上的，这种方法在第 4 节的概述中会提到。最近由数学家们相应地发展的这种方法为思考物理理论提供了一个高度抽象的框架，而不是完全严格地论述那些给定的理论。它存在于形式化的、微分几何的层面，关注各种空间的几何结构以及方程和结构的几何内容，暂且搁置泛函分析的细节。其他节中概述的大部分内容与此处于同一层次。

在内容上，本章与[Malament(马拉蒙特)，本书]、[Rovelli(罗韦利)，本书]以及[Brading and Castellani(布拉丁和卡斯特拉尼)，本书]等其他章节有一些重叠。但与本章最密切相关的是[Butterfield(巴特菲尔德)，本书]。巴特菲尔德的那章提供了一个关于现代力学几何方法的哲学概述，本章是将这种方法应用于哲学问题的一个例子。然而，本章的内容是自成体系的。事实上需要强调的是，本章与巴特菲尔德那章存在一个相当大的差异：后者被限制于有限维系统中，重点在于哈密顿方面；本章主要关注场论，更大程度上重点在于拉格朗日方法。

注释 1(符号和术语)。场论解空间上的元素和结构始终由大写字母(希腊文或拉丁文)表示，而场论初始数据空间上的元素和结构始终由小写字母(希腊文或拉丁文)表示。黑体字表示三维向量或三维向量值函数。在本章中，曲线是由实数区间到作为流形的空间或可稍微推广为流形的空间的映射——有时为了强调，我有些多余地将曲线称为参数化曲线。仿射参数化曲线是这类曲线的等价类，当两条曲线具有相同的图像且它们的参数化在最初选择是相同的时，

那么这两条曲线可算作等价的。[①]伪参数化曲线也是一类等价曲线，它们的等价关系是如果它们具有相同的图像，那么它们就是等价的。我有时不会区分曲线和它的图像。

注释 2(可能世界的讨论)。下面，尤其是在第 7 节，我有时用解(或初始数据)空间上的点来表示理论所允许的可能世界(可能的瞬时态)——即使在这里我并不会假装涉及解释的精细化问题，而仅仅采用一种粗略的、启发式的方式。这里的想法是试图理解理论，我们只是部分涉及寻找理论的一个明晰的公式化。合理的是，我们希望如果公式是明晰的，那么将存在一个“表面上”最有吸引力的对理论的解释，根据该解释存在一个解(或初始数据)空间和在该解释下理论所允许的可能世界(可能的瞬时态)之间的双射。不可否认的是，可能存在一些原因导致这种解释最终会被拒斥，比如莱布尼茨主义者(Leibnizean)可能选择经典力学的标准公式，即使这意味着需要将解与可能世界之间的表征关系看作是多对一的，因为与时间变换相关的解必须被看作对应于相同的可能世界。

2. 哈密顿和拉格朗日力学

这一节包含了牛顿 n 体问题[②]的哈密顿和拉格朗日方法的非常简短的概述，希望达到的目的是为我们后面要讨论的内容提供一些导引。

2.1 n 体问题

我们考虑 n 个受引力作用的点粒子，令第 i 个粒子质量为 m_i，相对一个静止惯性系，我们写出粒子位置 $q := (\mathbf{q}_1, \cdots, \mathbf{q}_n) = (q_1, \cdots, q_{3n})$，它们的速度

---

① 即仿射参数化曲线是等价关系下曲线的一个等价类，由此可知曲线 $\gamma_1(t) = \gamma_2(t+s)$ 和 γ_1: $[a, b] \to M$ 是等价的，当且仅当存在 $s \in \mathbb{R}$，使得 $\gamma_1(t) = \gamma_2(t+s)$ 对所有 $t \in [a, b]$ 成立。

② 各种不同风格的经典力学教科书有[Goldstein(戈德斯坦)，1953]，[Lanczos(兰索斯)，1986]，[Singer(辛格)，2001]，[Marsden and Ratiu(马斯登和拉蒂乌)，1994]，[Arnold(阿诺德)，1989]，[Arnold *et al.*，1997]，以及[Abraham and Marsden(亚伯拉罕和马斯登)，1978]。

为$\dot{q} := (\dot{\mathbf{q}}_1, \cdots, \dot{\mathbf{q}}_n) = (\dot{q}_1, \cdots, \dot{q}_{3n})$，加速度为$\ddot{q} := (\ddot{\mathbf{q}}_1, \cdots, \ddot{\mathbf{q}}_n) = (\ddot{q}_1, \cdots, \ddot{q}_{3n})$。在这一章，黑体字均表示三维向量。第$j$个粒子作用于第$i$个粒子的引力为：

$$\mathbf{F}_{ij} = \frac{m_i m_j}{r_{ij}^2}\mathbf{u}_{ij} \tag{1}$$

这里r_{ij}代表第i和第j个粒子之间的距离，$\mathbf{u}_{ij}$是从i粒子到j粒子的单位向量，选定单位以后牛顿常数就可以得到统一。当然，$r_{ij}=0$时方程(1)没有明确定义。所以从现在开始我们假设$q \in Q := \mathbb{R}^{3n}/\Delta$，其中$\Delta$是碰撞集$\{q \in \mathbb{R}^{3n}, \mathbf{q}_i = \mathbf{q}_j$，对某些$i \neq j\}$。

作用在i粒子上的合力为：

$$\mathbf{F}_i = \sum_{j \neq i} \mathbf{F}_{ij}$$

因此我们理论的运动方程为$\mathbf{F}_i = m_i \ddot{\mathbf{q}}_i$。① 把每一个力和加速度进行分解，我们得出$3n$个二阶微分方程。大致来说，这些方程式都存在一个确定初始值的问题，即要确定决定运动方程的唯一解析解的$3n$个初始位置以及初始速度(动量)的值。这个解析解告诉我们其他时候这些粒子的位置和速度(动量)如何，这些量就是如此定义的。②

2.2　哈密顿方法

哈密顿方法的基本变量是粒子的位置及相应的动量$p := (m_1\dot{\mathbf{q}}_1, \cdots, m_n\dot{\mathbf{q}}_n) = (m_1 q_1, \cdots, m_n q_{3n})$，系统的态$(q, p)$通过每个粒子的具体位置和动量来确定。对于每一个态我们可以指定其动能：

$$T(q,p) := \sum_{i=1}^{n} \frac{1}{2m_i} |\mathbf{p}_i|^2$$

和势能：

$$V(q,p) := \sum_{i<j} \frac{m_i m_j}{r_{ij}}$$

---

① 这应当被解读为一个偏微分方程，它限制着容许的轨迹$q(t)$，这一节中出现的其他偏微分方程类似于此。

② 注意，有些解并不能对t的所有取值定义，见下面例子 33 的讨论。

注意 $\mathbf{F}_i = -\nabla_i V(q)$，这里$\nabla_i$ 是和第 i 个粒子的配置变量对应的梯度算符$\left(\frac{\partial}{\partial q_{3i-2}}, \frac{\partial}{\partial q_{3i-1}}, \frac{\partial}{\partial q_{3i}}\right)$。因此势能编码了引力的信息，而动能可以被认为是编码了牛顿时空惯性结构的信息。人们可能期望这些量共同编码了 n 体问题的所有物理学，事实也的确如此。

我们为理论引入初始数据的空间$\mathcal{I} := \{(q, p): q \in Q\}$和哈密顿量 $H: \mathcal{I} \to \mathbb{R}$，$H(q, p) := T(q, p) + V(q)$，因此哈密顿量就是总能量。

原始运动方程 $m_i \ddot{\mathbf{q}}_i = \mathbf{F}_i$ 可重写为 $\dot{\mathbf{p}}_i = -\nabla_i V(q)$，或者由于$\nabla_i T = 0$，重写为 $\dot{\mathbf{p}}_i = -\nabla_i H$，换一种标记方法变为$\dot{p}_i = -\frac{\partial H}{\partial q_i}$。此外，由于在 H 中依赖于 p_i 的唯一项为$\frac{1}{2m}p_i^2$，我们可以发现$\frac{\partial H}{\partial p_i} = \dot{q}_i$。

这样我们从原来的牛顿方程得到哈密顿方程：

$$\dot{q}_i = \frac{\partial H}{\partial p_i} \text{ 和 } \dot{p}_i = -\frac{\partial H}{\partial q_i} \quad (i = 1, \cdots, 3n)$$

事实上，通常的牛顿方程和哈密顿方程是等价的。所以我们可以看到函数 $H = T + V$ 编码了所有 n 体问题的动力学内容。

目前我们感兴趣的是哈密顿方程的几何内涵。哈密顿方程给出每一点$(q, p) \in \mathcal{I}$的 $\dot{q}_i(q, p)$和$\dot{p}_i(q, p)$的值。也就是说，哈密顿方程在每个点$(q, p) \in \mathcal{I}$给出一个正切矢量 $X_H(q, p)$的分量表达式。在$\mathcal{I}$上的矢量场 X_H 编码了我们理论的动力学，通过每个点$(q_0, p_0) \in \mathcal{I}$确切地存在一条曲线$(q(t), p(t)): \mathbb{R} \to \mathcal{I}$，使得：(i)$(q(0), p(0)) = (q_0, p_0)$；(ii)对于每个 s，在 $t = s$ 时曲线的切向量$(q(t), p(t))$由 $X_H(q(s), p(s))$给出。这条曲线告诉我们，如果系统在 $t = 0$ 时的态为(q_0, p_0)，那么当 $t = s$ 时其态为$(q(s), p(s))$。

我们可以重新将哈密顿公式写为：

$$(\dot{q}_1, \cdots, \dot{q}_{3n}, \dot{p}_1, \cdots, \dot{p}_{3n}) \begin{vmatrix} 0 & I \\ -I & 0 \end{vmatrix} = \left(\frac{\partial H}{\partial q_1}, \cdots, \frac{\partial H}{\partial q_{3n}}, \frac{\partial H}{\partial p_1}, \cdots, \frac{\partial H}{\partial p_{3n}}\right)$$

这里 I 是一个 $3n \times 3n$ 的单位矩阵。左边是一个矢量和矩阵的乘积，右边是另一个矢量。把$\mathcal{I}$看作一个流形，我们可以认出公式背后独立于坐标的对象。左边是一个与 2 形式相关的正切向量场 X_H；右边是微分 dH（即 H 的外导数），因此

我们可以把哈密顿方程重新改写为一个协调独立的形式：

$$\omega(X_H, \cdot) = dH$$

这里 ω 是$\mathcal{I}$上的 2 形式，它假定了我们坐标中的形式 $\sum_i dq_i \wedge dp_i$。ω 是$\mathcal{I}$上的一个辛形式：闭合的、非简并的 2 形式。[①] ω 可以看作有些类似于一个$\mathcal{I}$上的反对称度规（比如，两种对象都可以在矢量场与 1 形式之间建立优越的同构）。但是鉴于两种对象之间的以下几种明显差别，这种类比不能被看得过于认真。

1. 有限维黎曼（Riemannian）流形等距群的维度也总是有限的。但我们的辛形式在一组从$\mathcal{I}$到自身的无限维微分同胚下不变，具体如下：我们把$\mathcal{I}$作为 Q 的一个余切丛，即把(q, p)看作由一点 $q \in Q$ 和余向量 $p \in T_q^* Q$ 组成。[②] 一个$\mathcal{I} = T^* Q$ 上的余切系统出现了：在 Q 上任选坐标 $\{q_i\}$，把 $p \in T_q^* Q$ 写为 $p = \sum p_i dq^i$，那么 $\{q_i, p_j\}$ 就形成了 $T^* Q$ 上的一组坐标。在任何余切坐标系中都有：

$$\omega = \begin{vmatrix} 0 & I \\ -I & 0 \end{vmatrix} \tag{2}$$

因此在把我们从$\mathcal{I}$上的一组余切坐标引向另一组余切坐标的变换中，ω 是不变的。而且这种变换的集合是无限维的，因为任何微分同胚映射 d：$Q \to Q$ 都会产生这样的变换。

2. 人们没有期望任何流形和丛会带有自然的黎曼度规。但如果 M 是任意的有限维流形，则余切丛 $T^* M$ 就带有正则辛形式 ω，这就引出 M 上相对于任何一组局域余切坐标的形式 $\omega = \sum_i dq_i \wedge dp_i$。[③]

3. 如果(M, g)和(M', g')是 n 维空间黎曼流形，那么对于任何的 $x \in M$ 和 $x' \in M'$，我们知道 g 和 g'赋予了正切空间 $T_x M$ 和 $T_{x'} M'$以同样的几何。但一般来说我们认为没有微分同胚映射 d：$M \to M'$会在 x 和 x'的邻域中给出一个等

① 更多讨论和对这个定义的分解详见 3.2 节。

② 为什么把 p 看作余向量而不是正切向量？因为一般来讲有着拉格朗日量的 L 系统中的动量 p 被定义为$\frac{\partial L}{\partial \dot{q}}$，它是 Q 上的坐标变换下的协变量。

③ 如上，M 上的一组局域坐标引出了 $T^* M$ 上一组自然的余切坐标，在下面例子 7 中我们将看到这一构造的无坐标版本，其继续存在于无限维情况中。

距。但是达布(Darboux)定理告诉我们如果(M, ω)是一个有着辛形式的有限维流形，那么(M, ω)与某个具有其余切丛辛形式的$T^*\mathbb{R}^n$是局域同构的。直接的推论就是，每一个有限维辛流形的维度都是偶数。

当然，就当前的意图来说，识别n体问题哈密顿方法背后的辛结构的兴趣变得普遍化了。(1)注意，如果我们对如上n粒子通过从势能函数V产生的力的相互作用感兴趣，那么我们就可以再一次把$\mathcal{I}$作为原始数据空间，给它加以上述辛形式ω，并定义哈密顿量H: $\mathcal{I}\to\mathbb{R}$作为动能和势能的总和，把我们的动力学轨迹当作是解出了$\omega(X_H, \cdot) = dH$的$\mathcal{I}$上的矢量场X_H的积分曲线，从而构造一个与牛顿方法等同的哈密顿方法。(2)更普遍的是，我们可以如下构建大量的经典力学系统：令初始数据空间为辛流形(M, ω)(不一定是余切丛)，给出哈密顿量H: $M\to\mathbb{R}$，然后令动力学由可解$\omega(X_H, \cdot) = dH$的矢量场X_H给出。

2.3 拉格朗日方法

以某种方式间接接近n体问题的拉格朗日方法是有用的。①

计算临界点

对于f: $\mathbb{R}^n\to\mathbb{R}$，$f$的微分由$df = \left(\frac{\partial f}{\partial x_1}, \cdots, \frac{\partial f}{\partial x_n}\right)$给出。我们说，如果$df(x_0) = 0$，那么$f$在$x_0\in\mathbb{R}^n$点有一临界点，即如果对每一个$e\in\mathbb{R}^n$有$df(x_0)\cdot e = 0$，那么$f$在$x_0$就有一个临界点（由于$\mathbb{R}^n$是线性空间，我们便可以认为$T_{x_0}\mathbb{R}^n$等同于$\mathbb{R}^n$自身，这里让$e\in\mathbb{R}^n$)。有很多解决$df(x_0)\cdot e$的有效方法：

(i)这些量在e方向上与f在x_0点的方向导数：

$$df(x_0)\cdot e = \lim_{t\to 0}\frac{f(x_0+te) - f(x_0)}{t}$$

① 运用变换的微积分的拉格朗日方法的介绍，见[Dubrovin *et al.*(杜布罗温等)，1992，Chapter 6]、[Lanczos，1986，Chapters Ⅱ and Ⅴ]和[van Brunt(范·布伦特)，2004]。这些变换的微积分的严格基础，见[Choquet-Bruhat *et al.*(肖盖—布鲁阿等)，1977，§§Ⅱ.A and Ⅱ.B]以及[Choquet-Bruhat and Dewitt-Morrette(肖盖—布鲁阿和德维特—梅里特)，1989，§Ⅱ.3]。

是一致的；

(ii) 如果我们有曲线 γ：$\mathbb{R}\to\mathbb{R}^n$，$\gamma(0)=x_0$、$\dot{\gamma}(0)=e$，那么 $df(x_0)\cdot e=\frac{d}{dt}f(\gamma(t))|_{t=0}$。

计算微积分和欧拉—拉格朗日(Euler-Lagrange)方程的变分

现在假设一个无限维模拟，我们寻找一个定义在欧几里得(Euclidean)空间的曲线空间上的函数的临界点，这是粒子力学拉格朗日方法的基础。

令 $Q=\mathbb{R}^n$，$[a, b]\subset\mathbb{R}$ 为一个闭区间，$x, y\in Q$。令 $\Gamma(a, b; x, y)$ 为满足 $\gamma(a)=x$ 和 $\gamma(b)=y$ 的 C^2 曲线 γ：$[a, b]\to Q$ 的集合。并且令 $\Gamma(a, b; 0, 0)$ 为 $\gamma(a)=(0, \cdots, 0)$ 和 $\gamma(b)=(0, \cdots, 0)$ 的 C^2 曲线 γ：$[a, b]\to Q=\mathbb{R}^n$ 的空间。$\Gamma(a, b; x, y)$ 和 $\Gamma(a, b; 0, 0)$ 在无限维空间中都状态良好。① 对于 $\gamma\in\Gamma(a, b; x, y)$，我们可以把 $\Gamma(a, b; 0, 0)$ 看作 $T_\gamma\Gamma(a, b; x, y)$，把 $h\in\Gamma(a, b; 0, 0)$ 看作描述了沿 γ 的矢量场。②

Q 的正切丛为 $TQ=\mathbb{R}^{2n}$。令 L：$TQ\to\mathbb{R}$ 为光滑函数。这允许我们通过 $I_{a,b}(\gamma):=\int_a^b L(\gamma(t), \dot{\gamma}(t))dt$ 定义函数 $I_{a,b}$：$\Gamma(a, b; x, y)\to\mathbb{R}$。我们感兴趣的是发现 $I_{a,b}$ 的临界点：这些点将是 $\Gamma(a, b; x, y)$ 即曲线 γ：$[a, b]\to Q$ 中特别有意义的点。就像一个状态良好的空间中的任何函数一样，$I_{a,b}$ 有一微分，标记为 $\delta I_{a,b}$；这可以被看作 $\Gamma(a, b; x, y)$ 上的 1 形式。

定义 3(平稳曲线)。我们说，如果 $\delta I_{a,b}(\gamma)=0$，那么 γ：$[a, b]\to Q$ 对于 $[a, b]$ 上的 L 是平稳的。如果对于所有的封闭区间 $[a, b]$ 它对 $[a, b]$ 的限制

---

① 令 $\Gamma(a, b)$ 为曲线 γ：$[a, b]\to Q$ 的空间 C^2，这是一个逐点增加即 $(\gamma+\gamma')(x)=\gamma(x)+\gamma'(x)$ 下的线性空间，能够以各种各样的方式转化为一个巴拿赫(Banach)空间。$\Gamma(a, b; 0, 0)$ 是 $\Gamma(a, b)$ 的一个线性子空间，而 $\Gamma(a, b; x, y)$ 是 $\Gamma(a, b; 0, 0)$ 上模型化的仿射子空间。

② 我们可以把 $T_\gamma\Gamma(a, b; x, y)$ 看作是按如下方式建立的：考虑有着 $\gamma[0]=\gamma$ 的曲线的单参数族 $\gamma[\varepsilon]$：$\varepsilon\in\mathbb{R}\mapsto\gamma[\varepsilon]\in\Gamma(a, b; x, y)$，断言如果 $\frac{d}{d\varepsilon}\gamma[\varepsilon]|_{\varepsilon=0}=\frac{d}{d\varepsilon}\gamma'[\varepsilon]|_{\varepsilon=0}$，那么这样的单参数族 $\gamma[\varepsilon]$ 和 $\gamma'[\varepsilon]$ 是等价的。$T_\gamma\Gamma(a, b; x, y)$ 是作为结果的等价类的空间。我们在 $\Gamma(a, b; 0, 0)$ 和 $T_\gamma\Gamma(a, b; x, y)$ 之间构造一个双射，把它看作是这种等价类通过与约束 $\gamma[\varepsilon]$：$\varepsilon\mapsto\gamma+\varepsilon\cdot h$ 的等价类 $h\in\Gamma(a, b; 0, 0)$ 相联系而得到的空间。

在$[a, b]$上平稳的，那么γ: $\mathbb{R}\to Q$对于L就是平稳的。

与$\mathbb{R}^n$上的普通函数情况相同，当且仅当对于所有的$h\in T_\gamma\Gamma(a, b; x, y)=\Gamma(a, b; 0, 0)$且$\delta I_{a,b}(\gamma)\cdot h=0$时，$\delta I_{a,b}(\gamma)=0$。然后我们可以通过寻找一条$\gamma[0]=\gamma$、$h=\frac{d}{d\varepsilon}\gamma[\varepsilon]|_{\varepsilon=0}$的曲线$\gamma[\cdot]$: $\varepsilon\in\mathbb{R}\mapsto\gamma[\varepsilon]\in\Gamma(a, b; 0, 0)$的$\frac{d}{d\varepsilon}I_{a,b}(\gamma[\varepsilon])|_{\varepsilon=0}$来计算$\delta I_{a,b}(\gamma)\cdot h$。

让我们来计算：固定L和$[a, b]$，令$\gamma\in\Gamma(a, b; x, y)$和$h\in\Gamma(a, b; 0, 0)$。对于每个趋于0的$\varepsilon$，我们借助$\gamma[\varepsilon](t):=r(t)+\varepsilon h(t)$来定义曲线$\gamma[\varepsilon]$: $\mathbb{R}\to Q$。因此$\gamma[\varepsilon]$是一条$\gamma[0]=\gamma$的$\Gamma(a, b; x, y)$上的曲线，其正切量为$h=\frac{d}{d\varepsilon}\gamma[\varepsilon]|_{\varepsilon=0}$。那么：

$$
\begin{aligned}
\delta I_{a,b}(\gamma)\cdot h &= \frac{d}{d\varepsilon}I_{a,b}(\gamma[\varepsilon])|_{\varepsilon=0} \\
&= \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\int_a^b[L(\gamma[\varepsilon](t),\dot{\gamma}[\varepsilon](t))-L(\gamma(t),\dot{\gamma}(t))]dt \\
&= \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\Big(\int_a^b\varepsilon\Big[\frac{\partial L}{\partial x}(\gamma(t),\dot{\gamma}(t)\cdot h(\gamma(t))) \\
&\quad +\frac{\partial L}{\partial\dot{x}}(\gamma(t),\dot{\gamma}(t)\cdot\dot{h}(\gamma(t))\Big]+O(\varepsilon^2)dt\Big) \\
&= \int_a^b\frac{\partial L}{\partial x}\cdot hdt-\int_a^b\frac{d}{dt}\Big(\frac{\partial L}{\partial\dot{x}}\Big)\cdot hdt+\Big(h\frac{\partial L}{\partial\dot{x}}\Big)\Big|_a^b
\end{aligned}
$$

第一个等式产生于函数微分的基本事实；第二个等式遵循定义；第三个源于泰勒定理(Taylor's theorem)；第四个则源自对部分的整合。现在我们注意，由于在$\gamma(a)=x$和$\gamma(b)=y$时h为零，最后一行第三个项也就为零。因此：

$$\delta I_{a,b}(\gamma_0)\cdot h=\int_a^b\Big[\frac{\partial L}{\partial x}-\frac{d}{dt}\Big(\frac{\partial L}{\partial\dot{x}}\Big)\Big]hdt$$

如果说γ: $[a, b]\to Q$在$[a, b]$是固定的，就是说这个表达式每一个h都为零。因此γ在$[a, b]$上对于L是平稳的条件是，拉格朗日方程：

$$\frac{\partial L}{\partial x}-\frac{d}{dt}\frac{\partial L}{\partial\dot{x}}=0 \tag{3}$$

在$t\in[a, b]$时沿$\gamma(t)$成立。γ: $[a, b]\to\mathbb{R}$对于L平稳的条件就是方程(3)在

整个 γ 上都被满足。

注释4(解析欧拉—拉格朗日方程)。这里讲如何分解方程(3)。[①] 重写 L 的表达式，处处用 ξ 代替 $\dot{x}$，然后把方程(3)解释为一个可接受的轨迹 $x(t)$ 的微分方程，理解 $\frac{\partial L}{\partial \dot{x}}$ 意味着 $\left.\frac{\partial L(x,\ \xi)}{\partial \xi_i}\right|_{\xi=\dot{x}(t)}$，理解 $\frac{d}{dt}\frac{\partial L}{\partial \dot{x}}$ 意味着：

$$\left.\left(\frac{\partial^2 L}{\partial \xi_i \partial \xi_j}\ddot{x}_j+\frac{\partial^2 L}{\partial \xi_i \partial x_j}\dot{x}_j\right)\right|_{\xi=\dot{x}(t)}$$

哈密顿原理

考虑一个有着位形空间的物理系统 Q(即考虑一个系统，其可能的空间位置被 Q 参数化)。令动能为一函数 T: $TQ\to\mathbb{R}$，它是通过 $T(x,\ v) := g_x(v,\ v)$ 从 Q 上的黎曼度规 g 中得出的。令 V: $Q\to\mathbb{R}$ 为一个只依赖于位形自由度的力的势，那么系统的拉格朗日量为 $L(x,\ v) := T(x,\ v) - V(x)$。哈密顿原理阐明了 L 平稳曲线是物理上可能的轨迹。许多物理上有意义的系统可以塑造为这种形式——例如上述 n 体问题。对于欧拉—拉格朗日方程所起源于的这种系统，哈密顿原理与一般的牛顿运动方程是等价的。

欧拉—拉格朗日方程解空间的辛结构

令 Q 为流形，TQ 是它的正切丛。令 L: $TQ\to\mathbf{R}$ 为光滑函数。L 的平稳空间 $\mathcal{S}$ 有着自然的流形结构，对于那些定义在 $t=0$ 的 $\gamma\in\mathcal{S}$，我们可以相对于 Q 上的坐标 $\{x_i\}$ 取 $x=\gamma(0)$ 和 $v=\dot{\gamma}(0)$ 作为 $\mathcal{S}$ 上的坐标。对于 t 的每个值都这样做，就可以给出一个 $\mathcal{S}$ 的可微图解集。由此可得 $\dim\ \mathcal{S}=\dim\ TQ$。我们可以赋予 $\mathcal{S}$ 一个几何结构：因为边界项 $h\frac{\partial L}{\partial \dot{x}}$ 在上面导出欧拉—拉格朗日方程时被消除了。由于 h 可被看作空间 $\Gamma(a,\ b;\ x,\ y)$ 的一个正切向量，我们必须把 $\alpha=\frac{\partial L}{\partial \dot{x}}$ 看作那个空间中的一个 1 形式。采用其外导数会带给我们一个 2 形式，$\Gamma(a,\ b;\ x,\ y)$ 上的 $\omega=:\delta\alpha$。把这种形式约束于 $\mathcal{S}$ 上就是我们所要寻求的结构。我们在 $\mathcal{S}$ 上已经引入的坐标中，ω 所采用的形式为：

$$\omega=\frac{\partial^2 L}{\partial x^a \partial v^b}dx^a\wedge dx^b+\frac{\partial^2 L}{\partial v^a \partial v^b}dv^a\wedge dx^b$$

---

① 参见[Dubrovin *et al.*，1992，318]。

对于任何 L，这都是一个封闭的 2 形式。它是非简并的，因而也就是辛的，只要 $\left[\frac{\partial^2 L}{\partial v^a \partial v^b}\right] \neq 0$。① 对于上面考虑的形式的拉格朗日量，该条件永远成立，那么我们就会发现 $(\mathcal{S}, \omega)$ 与那些对应的来自哈密顿方法的初始数据空间（局域地）辛同构。②

3. 辛事物

整个这一章，我们将通过探寻时间和变化在物理学理论的拉格朗日和哈密顿形式中的表示来研究其在物理学理论中的表示。在拉格朗日形式中，焦点一直是我们理论方程的解空间，而在哈密顿形式中，焦点则一直是那些方程的初始数据空间。对于那些表现好的理论，初始数据空间和解空间拥有共同的几何结构——这些空间作为辛流形是同构的。这是最重要的事实，因此辛流形的观念及其一般化在我们的研究中将会起到重要的作用。

以辛流形本质的一般讨论开始会比较有利：接下来的 3.1 介绍一些基本事项；3.2 会提供一些基本概念、结构以及辛几何在力学中的结果的概述；3.3 提供对预辛几何的同类处理方法（将起到重要作用的辛几何的推广见 6.2 和 6.7）；3.4 讨论辛结构在何种意义上是量子化必不可少的条件。

3.1 初步准备

我们接下来遇到的空间将会对 n 维流形从三个方面进行推广：(i) 它们被允许是非豪斯道夫（non-Hausdorff）的；③ (ii) 它们被允许是无限维的，一个局

---

① 关于拉格朗日理论解空间的辛结构的讨论，[Woodhouse（伍德豪斯），1991，§§2.3 and 2.4]中有详细讨论。

② 这是从以下结果得出的：上述考虑的来自诸如动能和势能项的拉格朗日量总是超规则的，见[Abraham and Marsden，1978，226]。

③ 回顾一下，拓扑空间 X 是豪斯道夫的，如果对于所有不同的 x，$y \in X$，存在不相连的开集 U 和 V，其中 $x \in U$，$y \in V$，虽然大部分教科书要求流形是豪斯道夫的，但所有的基本构造和结构都没有提到这个假设，见[Lang（朗），1990]。在下面的例子 32 和例子 33 中，即便是最简单的物理系统的解空间都可以是非豪斯道夫的。

部地被构建在矢量空间上的流形，我们允许我们自己的流形构建在$\mathbb{R}^n$或者一个无限维巴拿赫空间上；①（iii）它们被允许有温和奇点——大体说来，我们的空间将按照一个普通圆锥体构成其顶点（一个零维流形）和锥面（一个二维流形）的方式组成流形——但我们的空间仍然有着光滑的结构并在很大程度上以和流形一样的方式支撑着张量。②

为了避免陷于技术性困难，我将粗略地解释流形背景中辛和预辛几何的概念和结构，但是在接下来的部分中，我会提到"空间"而不是流形，这应当理解为我允许讨论的空间中具有上面所提到的温和奇点。

接下来我们常常会对李群在流形上的作用量，以及对作为这些作用量的无限小生成元的矢量场中的作用量感兴趣。让我通过回顾一些恰当的定义和结构来结束这个讨论。

回想一下，李群是一个流形，同时也是一个有着群的乘法算子$(g, h) \in G \times G \mapsto g \cdot h \in G$，并把其反演$g \in G \mapsto g^{-1} \in G$当作光滑映射的群。流形$M$上一个李群$G$的行为是一个光滑映射$\Phi$：$G \times M \to M$，使得：（i）对于$G$的同一元素$e$以及对所有的$x \in M$，有$\Phi(e, x) = x$；（ii）对于所有的$g$，$h \in G$和$x \in M$，有$\Phi(g, \Phi(h, x)) = \Phi(gh, x)$。常常也会把$g \cdot x$或$\Phi_g(x)$写成$\Phi(g, x)$。③ 作用在$x \in M$上的轨道是集合$[x] := \{g \cdot x: g \in G\}$。李群的作用把一个

① [Abraham *et al.*, 1988]和[Lang, 1999]介绍了涵盖无限维巴拿赫流形的情况的微分几何。[Milnor（米尔诺），1984, §§2—4]对一个更加普遍的方法进行了介绍。在这种方法中，流形是在一个局域凸拓扑向量空间中模型化的。注意，常微分方程的反函数定理及存在性和唯一性定理在这种更加普遍的方法中失效了。

② 我们考虑的空间是惠特尼（Whitney）分层空间。正如文中指出的，每一个这样的空间都容许被正则分解到流形。这种分解允许我们把这样的空间中的每一个点都看作是位于流形之中，这又允许我们在每一点上构造一个正切向量空间和余切向量空间。分层空间上流形（以及正切和余切空间）的维度一般来讲会在每一点上不同。有限维情况下对这种空间的处理，见[Pflaum（普夫劳姆），2001]或者[Ortega and Ratiu（奥尔特加和拉蒂乌），2004, §§1.5-1.7]。这种图景看起来非常类似于物理学中出现的有限维的例子，广义相对论的例子见[Andersson（安德森），1989]和[Marsden, 1981, Lecture 10]；杨—米尔斯（Yang-Mills）理论的例子，见[Arms（阿姆斯），1981]和[Kondracki and Rogulski（康达基和罗古斯基），1986]。

③ 等价的，M上G的行为是一个从G到$\mathcal{D}(M)$（从M到自身的微分同胚群）的群微分同胚$g \mapsto \phi_g$计划，这样映射$(g, x) \in G \times M \mapsto \phi_g(x) \in M$就是光滑的。

流形分割为轨道。

虽然下面还会提到其他李群，我们大多数情况下还是对最简单的李群即加法群$\mathbb{R}$感兴趣。在流形M上的一个流是一个从M到M的微分同胚的单参数群。因此，如果$\{\Phi_t\}_{t\in\mathbb{R}}$是$M$上的一个流，那么对于所有的$x\in M$和$s$，$t\in\mathbb{R}$，有$\Phi_0(x)=x$，且$\Phi_t\circ\Phi_s(x)=\Phi_{t+s}(x)$。$M$上的流$\{\Phi_t\}$以及在$M$上的作用量$\Phi$：$\mathbb{R}\times M\to M$基本上是同一个东西，即给定一个$\mathbb{R}$作用量$\Phi$：$\mathbb{R}\times M\to M$，可以通过对于所有$t\in\mathbb{R}$和$x\in M$的$\Phi_t(x)=\Phi(t,\ x)$定义一个流$\{\Phi_t\}$，如果给定一个流而想要定义一个$\mathbb{R}$作用量，也是同样的方式。

M上任意的$\mathbb{R}$作用量都会引起其上的一个矢量场X。令$x\in M$，考虑通过γ_x：$t\mapsto\Phi_t(x)$定义的曲线$\gamma_x(t)$：$\mathbb{R}\to M$。γ_x在M中的图像就是轨道$[x]$。现在假定$y\in[x]$——即存在$t\in\mathbb{R}$使得$y=\Phi_t(x)$。从群$\{\Phi_t\}$的特性立刻会得到一些事实。我们发现γ_y的图像同样是$[x]$——所以$[x]=[y]$。我们发现，事实上，对于所有的$s\in\mathbb{R}$，有$\gamma_y(s)=\gamma_x(s+t)$。即每一条对应于点$y\in[x]$的曲线$\gamma_y$与决定它们的参数化的原点的选择一致。因此，$\mathbb{R}$作用量的每一条轨道$[x]$都会作为(最大)曲线族的共有图像出现，这个曲线族在参数化上与取决于原点的选择相一致。作为一种方便简易的方式，我们将把这样的曲线族称为仿射参数化曲线，把这种曲线看作是其参数的确定只取决于原点的选择。现在我们可以如下在M上建立一个矢量场X：对于$x\in M$，我们定义$X(x)=\dot{\gamma}_x(0)$。以上论述表明X是M上的一个光滑矢量场。

现在假设给定了流形M上的一个矢量场X，让我们看看是否可以认为X在M上产生了一个$\mathbb{R}$作用量。给定$x\in M$，存在唯一的曲线γ_x在$t=0$时经过x，使得对于曲线所定义的任何t的取值，其在$\gamma(t)\in M$点的正切矢量都由那一点的X的取值给出。① 把这条曲线叫作基于x的积分曲线。我们发现，如果y出现在其中，那么基于x和y的积分曲线具有同样的图像，而且它们在参数化时取决于原点的选择上一致。也可以用对应的仿射参数化曲线来替换这些曲线，我们把它们称为通过x（或y）的积分曲线。因此，矢量场X允许我们在M上定义一个积分曲线族，M上的每个点恰好位于一条这样的曲线之上。对于$x\in M$，

---

① 这只是对一阶常微分方程的存在性和唯一性定理的一种陈述。

$t \in \mathbb{R}$，让我们承认 $\Phi(t, x)$ 是沿着经过 x 的积分曲线通过追踪 t 单元所达到的点，条件是当这一做法是意义明确的（记住基于 x 的积分曲线可能只被定义在 $\mathbb{R}$ 的子区间上）。当且仅当每条积分曲线的定义域都是全部的 $\mathbb{R}$ 时，Φ 才能称为 $\mathbb{R}$ 作用量。在这种情况下，我们把 X 称为一个完备的向量场，把 Φ 称为由 X 产生的 $\mathbb{R}$ 作用量。

图景如下：$\mathbb{R}$ 称为作用量，引出了 M 上的一个矢量场 X，X 产生 Φ。我们把群 $\{\Phi_t\}$ 看作是极小变换 X（这里及后文中"极小"的意思都是指"位于正切空间中"）产生的有限变换组成的。当 X 不完备时，$\Phi(t, x)$ 不能对于所有的 (t, x) 进行定义。在这种情况下，Φ 就是一个局域流。对不少目的而言，局域流和流几乎一样好，认为它们具有矢量场作为其无限小生成元，这仍然是有益的。

3.2 辛流形

定义 5（辛流形）。令 M 为一流形，M 上的辛形式是一个封闭的非简并 2 形式 ω。这里非简并是指在每一点 $x \in M$ 上，映射 $\omega^\flat(x)$：$v \in T_xM \mapsto \omega(v, \cdot) \in T_x^*M$ 都是单射的。① (M, ω) 被称为辛流形②。

定义 6（辛对称）。令 (M, ω) 为辛流形。(M, ω) 的辛对称是一个微分同胚 Φ：$M \to M$，它在 $\Phi^*\omega = \omega$ 的意义上保留了 ω（即 ω 通过 Φ 的拉回就是 ω）。

例子 7（余切丛辛结构）。令 Q 为一个有限或无限维流形，且余切丛为 T^*Q。在 T^*Q 上如下定义一个典型的辛形式：令 π：$T^*Q \to Q$ 为一个正则投影 $(q, p) \in \mapsto q$，$T\pi$ 为相应的正切映射。在 T^*Q 上存在唯一的 1 形式 θ 使得对于所有的 $(q, p) \in T^*Q$ 和 $\omega \in T_{(q,p)}T^*Q$，有 $\theta(q, p) \cdot w = p(T\pi \cdot w)$。然后我

---

① 当然，对有限维的 M 来说，$\omega^\flat$ 是满射的，当且仅当它是单射的。

② 从辛观点出发，[Abraham and Marsden, 1978] 以及 [Arnold, 1989] 是对力学的标准处理方法。[Schmid（施密特），1987] 也涉及了无限维流形例子中的一些相同基础。[Ortega and Ratiu, 2004] 是有限维辛空间（包括奇异空间）的几何和对称的综合性参考文献。[Cannas da Silva（坎纳斯・达・席尔瓦），2006] 是对辛几何的有益探索。[Weinstein（温斯坦），1981] 和 [Gotay and Isenberg（戈塔伊和伊森伯格），1992] 提供辛几何在数学和物理学中的作用的概述。

们可以定义我们想要的辛形式为$\omega := -d\theta$，其中 d 是 T^*Q 上的外导数。[①]

令(M, ω)为一个辛形式，$C^\infty(M)$是 M 上的一组光滑函数。对于当前的目的，ω 基本的作用是允许我们把每个$f \in C^\infty(M)$与 M 上的一个光滑矢量场 X_f 联系起来。X_f 绝对地由方程 $\omega(X_f, \cdot) = df$ 定义，这里 df 是 f 的一个外导数（ω 的非简并性确保这个方程有唯一解）。[②] 我们说 f 产生了 X_f 或 X_f 由 f 产生。

这个基本结构主要有两个成果：

1. 通过映射$f \mapsto X_f$，ω 允许我们在 $C^\infty(M)$上定义一个新的代数算子：f 和 $g \in C^\infty(M)$之间的泊松(Poisson)括号是$\{f, g\} := \omega(X_f, X_g)$。[③] 这在量子化理论中起到重要作用，见 3.4 节。

2. 通过映射$f \mapsto X_f$，ω 通常允许我们把 M 上的光滑函数与(M, ω)的对称性的单参数群联系起来，反之亦然。(i)令$f \in C^\infty(M)$，且 X_f 是由f(通过 ω)产生的矢量场。假定 X_f 是完备的，使得我们能构造一个相应的流 $\xi = \{\Phi_t\}_{t \in \mathbb{R}}$，那么每个 Φ_t 在 $\Phi_t^* \omega = \omega$ 的意义上保留了 ω，[④] 且 f 自身在每个 Φ_t 下都是不变的。[⑤] (ii)令 $\xi = \{\Phi_t\}_{t \in \mathbb{R}}$ 是(M, ω)的对称单参数群，X 是 M 上的矢量场，是 ξ 的无限小生成元。我们自然会问是否能够找到一个产生 X 的 $f \in C^\infty(M)$。在某些情形中不可能找到。[⑥] 但是在物理学中发生的例子里，这常常是可以做到的。

---

① 在有限维情况下，相对于一组余切坐标，ω 由上述方程(2)给出。

② 在无限维情况下，X_f 可能不是定义在所有 M 之上。对于表现良好的 f，我们可以通过把 M 置换为一个子空间来解决，X_f 就定义在这个子空间上。下面我将会假定已经做到了这一点。讨论可参照[Marsden, 1981, 11ff.]和[Marsden and Ratiu, 1994, 106]。

③ 泊松括号是一个遵守莱布尼茨规则(Leibniz's rule)的李括号$\{fg, h\} = f\{g, h\} + g\{f, h\}$。

④ 事实上我们可以更进一步注意到，ω 沿 X_f 的李导数为零——并且即便 X_f 是不完备的时候，这也成立。这就表明，在某种意义上由一个不完备的矢量场产生的局域流保留了 ω。

⑤ 事实上，f 沿 X_f 的李导数为零。这在 X_f 不完备的时候也成立，因此在某种意义上由这样一种不完备的矢量场产生的局域流保留了 f。

⑥ 例子见[Ortega and Ratiu, 2004, §4.5.16]，进一步的讨论见巴特菲尔德在本书第 2.1.3 小节中的讨论。

并且由上述(i)，我们可以找到这样的f，它为流ξ所保留。[①]

如果牢记这种框架的哈密顿应用，那么掌握辛结构的函数就会简单许多。

定义 8(哈密顿系统)。哈密顿系统(M, ω, h)是由一个被称为相空间的辛流形(M, ω)和一个被称为哈密顿量的函数$h: M\to\mathbb{R}$构成的。

我们把(M, ω)看作是某个物理系统的相空间，比如粒子的位置和动量的空间；把h看作为系统的每一个态赋予那个态的总能量。h和ω共同决定M上的一个流$\{\Phi_t\}_{t\in\mathbb{R}}$，即每个$\Phi_t$把每个态映射到$t$时间单位后动态地紧随它的态上。在这个群下$h$将得到保留——这符合能量守恒。[②]

3.3 预辛流形

在后面的6.2节和第7节中我们将考虑那些解空间和初始数据空间都不是辛结构的理论。

定义 9(预辛流形)。令M为一个流形，M上的预辛结构是一个封闭且简并的2形式ω，我们称(M, ω)为一个预辛流形。[③] 这里简并的意思是：在每个点

---

① 更大胆些，令G为一个通过辛对称作用于M上的李群，有$\dim G>1$。这样一个群将会包含许多单参数子群——就像欧几里得空间的等距群包含了一个对应于欧几里得空间中每一个方向上的平移单参数群和对应于欧几里得空间每一个轴的转动单参数群。在这种情况下，我们能够期望对于G的每一个单参数子群，有可能找到一个M上的函数，它产生了那个子群。如果一切顺利——正如物理学中出现的很多例子那样——我们能够期望这些生成元之间的泊松括号的代数能够反映群的代数(即这里会存在一个李代数同构)。在这种情况下，我们谈一谈动量映射的存在(注意：这里术语发生了变化——许多作者称之为极小等价动量映射)。如果f和g是M上使泊松括号为零的函数，那么我们发现f在由g产生的单参数群中的每一个辛对称下保持不变。特别地，如果G是(M, ω)的一个辛对称群，f是M上的一个函数，并且任何产生G的单参数子群的函数f的泊松括号为零，那么这些生成元中的每一个在f产生的对称的单参数子群下不变。[Woodhouse，1991，§3.4]提供了一个关于动量映射在某种形势下存在或者不存在的有用引导。更进一步的讨论见[Butterfield，本书]。

② 通常，有可能支持一个(M, ω)的辛对称的更大群G，在这种对称下h不变(比如欧几里得对称群以明显的方式作用于牛顿粒子力学)。那么一个动量映射将会允许人们构造$\dim G$独立量，它们的代数将会反映G的代数，这在由h产生的动力学下将会守恒。

③ 术语变化了：通常(但不是此处)辛形式是作为预辛形式的特殊情形，有时候(但不是此处)预辛形式被要求有着常数阶或者有良定的规范轨道空间。关于预辛几何，见[Gotay and Nester(戈塔伊和内斯特)，1980]。

x上都存在一个非平凡的零空间$N_x \subset T_xM$，由正切矢量v，比如$\omega_x(v, \cdot)=0$组成。

一个流形M上的辛结构ω通过子流形$\{M_\alpha\}$给出了M的划分。我们宣称如果$x, y\in M$能够被一曲线γ: $\mathbb{R}\to M$相连接，而它们各自的正切矢量都为零——即对每一个$t\in\mathbb{R}$，$\dot{\gamma}(t)\in N_{\gamma(t)}$，那么它们就是等价的，从而在$M$上定义一个等价关系。这个关系的等价类$M_\alpha$被称为规范轨道。对$x\in M$，我们也表示规范轨道通过$[x]$包括了$x$。每个规范轨道是$M$的一个子流形。① 如果只要$x$和$y$属于$M$的同一个规范轨道就有$f(x)=f(y)$，那么我们可以认为$f\in C^\infty(M)$是规范不变的(即一个函数是规范不变的，当且仅当它在规范轨道上是不变的)。

我们把一个从ω保留了预辛结构的M向自身的微分同胚称为(M, ω)的预辛对称。如果我们说两个预辛对称Φ和Φ'，对于每一个$x\in M$，有$[\Phi(x)]=[\Phi'(x)]$(即对于每个$x\in M$，Φ和Φ'把x映射在同一个规范轨道上)，那么它们就是规范的。我们把这种与Φ一样规范的预辛对称的集合称为Φ的规范等价类。同样我们也可以说，如果Φ_t和Φ'_t对于每个t都是规范的，两个预辛对称的单参数群$\xi=\{\Phi_t\}$和$\xi'=\{\Phi'_t\}$就是规范的。$\xi=\{\Phi_t\}$的规范等价类包含了在这个意义上与它一样规范的所有ξ'。

如果一个预辛对称Φ: $M\to M$可以确定每一个M_α(即Φ把M_α中的点映射在M_α中的点上)，那么我们称Φ为一个规范变换。注意，规范变换是有着M上的恒等映射的规范。

在辛情况中，当一切顺利，方程$\omega(X_f, \cdot)=df$允许人们把辛流形(M, ω)上的每一个光滑函数与(M, ω)的辛对称单参数群联系起来——反之亦然。

在预辛情况中，当一切顺利，方程$\omega(X_f, \cdot)=df$允许人们把预辛流形(M, ω)上的每一个光滑规范不变函数与(M, ω)的预辛对称单参数群的规范等价类联系起来——反之亦然。所以在预辛情况中，如果f通过方程$\omega(X_f, \cdot)=df$产

---

① 如果X和Y是M上的零矢量场(即对于每一个$x\in M$)，有$X(x)$，$Y(x)\in N_x$，那么$[X, Y]$也是一个零矢量场。因此，由弗罗宾尼斯(Frobenius)定理，N_x形成了一个可积分布，M_α为积分流形。

生预辛对称的单参数群 $\xi=(\Phi_t)$，那么它也会在 ξ 的规范等价类中产生每一个 $\xi'=\{\Phi'_t\}$。

注意一个有趣且特殊的例子：f 为常值函数，方程 $\omega(X_f,\cdot)=df$ 的任何解 X_f 是 M 上的矢量场，由零向量组成，因此对应的 (M,ω) 的预辛对称的一维群由规范变换构成。相反，如果 $\xi=\{\Phi_t\}$ 是 (M,ω) 的规范变换的一维参数群，那么（通过 ω）产生 ξ 的任何函数都是常值函数。

给定一个预辛流形 (M,ω)，我们可以构造 M 的规范轨道空间 M'。M' 继承了 M 的拓扑结构。① 我们可以称从 M 到 M' 的过程为约化，把 M' 叫作约化空间。一般来说，M' 不一定是流形，也不需要是任何差不多这样良好表现的东西，就如我们接下要考虑到空间一样。② 但当一切顺利（下面的例子都会是这样）M' 将会继承 M 的光滑结构。并且只要 ω 上的一些更进一步的技术条件成立，M' 就会继承 (M,ω) 的非简并且封闭的 2 形式 ω'。③ 因此，在这个例子中，(M',ω') 是辛空间。注意每个规范不变的 $f\in C^{\infty}(M)$ 对应唯一的一个 $f'\in C^{\infty}(M')$。在 f 产生 (M,ω) 辛变换的单参数群的一个等价类的同时，f' 产生了 (M',ω') 辛变换的单一的单参数群。④

3.4　辛结构和量子化

量子化是构造给定经典理论的量子对应部分的过程。⑤ 就目前的理解，它

---

① 我们令 M' 具有商拓扑，据此，集合 $U'\subset M'$ 为开集，当且仅当 M 中 $\pi^{-1}(U')$ 为开集，这里，π 为投影 $x\in M\mapsto[x]\in M'$。

② 如果 (M,ω,H) 在上述定义 8 的意义上为一个哈密顿系统，那么 ω 对一个常能量面 E 的限制是预辛的——规范轨道 $(E,\omega|_E)$ 为哈密顿系统的一个动力学轨迹。如果动力学是各态历经的，那么一般的轨迹就对任意接近每一个 $x\in E$。因此，商空间 E' 有着平凡拓扑，因而唯一的开集是空集以及空间本身。

③ 见[Marsden，1981，6]和[Ortega and Ratiu，2004，§6.1.5]。

④ (M,ω) 的每种预辛对称对应于 (M',ω') 的一种辛对称，两种预辛对称对应于同一种辛对称，当且仅当它们都是规范的。因此，预辛对称的每一种规范等价类对应于一种单一的辛对称。预辛对称单参数群的每一种规范等价类对应于约化空间的单一辛对称单参数群。

⑤ 有关量子化的文献概述，见[Landsman（兰兹曼），本书]和[Ali and Engliš（阿里和英吉利），2005]。

是一种以哈密顿或者拉格朗日形式的理论(或者这样一种理论的离散时间类比)为起点的过程。人们不知道怎样才能不首先转化为哈密顿或者拉格朗日形式就可以直接以微分方程的形式对理论进行量子化。①

下面的发现丰富了辛结构是量子化的必要条件这一思想的合理性。

1. 量子化的核心观念包括以下步骤。人们从辛空间(经典解或初始数据的空间)出发，在这个由辛结构诱导产生的叠加和泊松括号下是闭合的空间上选择一组函数(经典物理量)。然后寻找一组作用于量子态空间上的算子(量子可观测量)，使得这些算子的代数可以反映(或者近似地反映，越接近经典极限就匹配得越好)所选择的经典量的代数(在叠加和泊松括号下)。人们可能还需要进一步增加一个实现量子动力学的哈密顿算子。

2. 有些经典理论具有如下令人遗憾的特征：当转换为拉格朗日或是哈密顿形式时，它们向我们显现的是其解空间或初始数据空间只有预辛结构。一般而言，众所周知的是，在即将发生的约化中存在一个辛空间，如上 3.3 部分中粗略地讲过。但可能很难构造这个空间，或者可能由于某种原因它看起来会比较容易采用理论的预辛形式。因此，已经有人提出了一些策略来在预辛形式中量子化理论，如规范固定、狄拉克(Dirac)约束量子化、BRST 量子化等。但把这些技巧的每一种看作是提供了对基本辛空间进行量子化的一种间接方法是很自然的想法。②

然而，存在一些似乎并不包含任何经典空间的辛结构的量子化方法——比如麦基(Mackey)量子化(其应用有限)和路径积分量子化(其应用非常广泛，但

---

① 表明为了使一个理论可以量子化，它必须可以从拉格朗日形式推出的文献，见[Hojman and Shepley(霍里曼和谢普利)，1991]。

② (i)规范修正只是要找到预辛空间的一个子流形，这个子流形辛同构于约化空间，见[Henneaux and Teitelboim(亨诺和泰特尔鲍姆)，1992，§1.4]。(ii)狄拉克方法，见[Dirac，2001]或[Henneaux and Teitelboim，1992]。当狄拉克的结果与直接量子化约化理论的结果不同的时候，人们觉得狄拉克的运算法则应当进行修正，参考[Duval *et al.*(杜瓦尔等)，1990]。(iii)在有限维系统的情况下，可以表明 BRST 运算规则的运用导致了对约化理论的量子化，见[Loll(劳)，1992]或者[Tuynman(图尔曼)，1992]。(iv)BRST 方法和狄拉克方法的修正之间的关系，见[Guillemin and Sternberg(吉耶曼和斯特伯格)，1990，§12]。

在应用于场论时会面临困境)。正如[Landsman,本书]所强调的,经典和量子之间的关系离被完全理解还差得远。

4. 拉格朗日场论

微分方程通常以下列方式给出:给定一组自变量和一组因变量,以及一个由函数 u 组成的函数空间$\mathcal{K}$,这个空间在自变量的值和因变量的值之间建立起映射关系。一个微分方程 Δ 可被看作是函数上的一个条件,它的微商只能为某些$u \in \mathcal{K}$所满足。我们称满足 Δ 的 u 为 Δ 的解,通过$\mathcal{S}$指示了这些解的空间。

在物理学应用中,自变量常常参数化空间、时间或者时空,而因变量则参数化某些量的可能值。我们可以把函数 $u \in \mathcal{K}$看作是描述了在某种意义上可能的状态,把解 $u \in \mathcal{S}$看作是描述了根据那些其定律编码在 Δ 中的理论来说真正物理可能的状态。虽然术语不是那么明白易懂,我还是要谈及与运动学可能性相一致的$\mathcal{K}$的元素以及与动力学可能性相一致的$\mathcal{S}$的元素。

例子 10(粒子的力学)。考虑粒子理论,一个欧几里得空间中的粒子受到依赖于位置的力的支配。自变量参数化时间,而因变量参数化粒子的可能位置。具有形式 $t \in \mathbb{R} \mapsto x(t) \in \mathbb{R}^3$ 的任意连续函数 $x(t)$ 描述了粒子的运动学可能的行为方式,$x(t)$ 如果满足牛顿方程 $\ddot{x}(t) = F(x(t))$,则描述了运动学可能的行为。

例子 11[克莱因—戈登场(Klein-Gordon Field)]。通常标量场的理论都有以下成分:作为自变量我们有闵可夫斯基(Minkowski)时空 V 上的惯性系 $\{t, x, y, z\}$,理论有单一的因变量,参数化了实数,因此运动学可能的场由闵可夫斯基时空上(适当光滑)的实值函数所给出,动力学可能的场是指那些满足克莱因—戈登方程的 Φ: $V \to \mathbb{R}$。

$$\frac{\partial^2\Phi}{\partial t^2} - \frac{\partial^2\Phi}{\partial x^2} - \frac{\partial^2\Phi}{\partial y^2} - \frac{\partial^2\Phi}{\partial z^2} - m^2\Phi = 0$$

下面我们主要关注的是场论——那些定律被编码在自变量确定了时空参数①的微分方程中的物理学理论。我们把这样一个场论看作是由以下成分构成的：时空 V、空间 W(场在其中取值)、一个动力学可能场的(即从 V 到 W 满足适当光滑性和边界条件的函数的)空间$\mathcal{K}$，以及一组微分方程 Δ。

这一部分结构如下。在下面第 1 节中，我讨论拉格朗日方法，在其中通过拉格朗日变分问题而不是通过直接强加一个微分方程选出运动学可能性空间中的一组动力学可能性。在第 2 节中，我讨论拉格朗日方法相对于直接方法的一个非常重要的优势，是前者而不是后者允许动力学可能性空间有(预)辛形式。在第 3 节中，我讨论拉格朗日方法中著名的守恒量和对称性之间的关系，这个关系最初是由诺特(Noether)阐明的。这些小节的讨论的基础有[Zuckerman(朱克曼)，1987]、[Deligne and Freed(德利涅和弗里德)，1999，Chapters 1 and 2]，以及[Woodhouse，1991，Chapters 2 and 7]。

在开始之前有必要对理论做出一些特定的假设，这些假设我们在下面都会讨论到。这些假设在本章的剩余部分都有效。

时空。我们的时空 V 将永远是一个 n 维豪斯道夫流形 V，有着某种$(n-1)$维流形 M 的拓扑 $M\times\mathbb{R}$ 。我们永远把时间看作带有$\mathbb{R}$ 拓扑，因此我们可以说，带有拓扑 $M\times\mathbb{R}$ 的时空的空间拓扑为 M。特别地，我们可以说，如果 M 是紧致的，那么 V 是空间紧致的。

在大多数的理论中，时空的几何在解与解之间是固定的。因此我们通常认为 V 携带了一个独立于解的几何结构(我比较懒，有时会用 V 来表示流形，有

---

① 乍一看，这种经典场论的本体满足刘易斯的休谟附随性(Lewis' Humean supervenienсs)——“对这个世界来说所有存在的是大量马赛克式的有着特别真相的局部事件，一件接着一件的小事……我们有几何：作为点之间的时空距离的外在关系系统……在那些点处，我们有局域量，完全固有的内在特性，它们不需要比它们例示于其上的点更大的任何东西……这就是全部”[1986，ix f.]。事实上，刘易斯说，这种图景是为“经典物理学所激励的”[1999，226]。关于对休谟附随性和经典物理学之间的适当性的质疑，见[Butterfield，Unpublished]。译者注：本文最后发表于 *British Journal for the philosophy of Science*，第 57 卷，NO.4(15 December 2006)，第 709—753 页，因此，这里应该改为“[Butterfield 2006]”。

时会把流形和几何都这样表示）。[①] 我不担心是否精确，现在就约定我们只对表现正常的时空感兴趣。比如牛顿时空、新牛顿时空、闵可夫斯基时空，或者其他的全域双曲广义相对论时空。

我们考虑的时空几何在 V 中选出了一类突出的超曲面对应于时间的某些瞬间。[②] 在前相对论时空中，瞬间只是绝对同时性的一个超曲面。一般而言，在相对时空中，瞬间只是柯西面。偶尔在高度对称的相对时空中，人们会要求瞬间有着很好的对称的柯西面——因此，在某些情况里人们可能要求闵可夫斯基时空中的瞬间作为相对于惯性观测者的同时性的超曲面出现。而且，在携带几何结构的时空中，把 V 中的某些曲线当作点粒子的可能世界线往往是有意义的。在前相对论时空中，一条曲线如果横穿同时性超曲面，那么它就可被看作点粒子的可能世界线。在相对论时空中，这样的可能世界线由类时曲线给出。

除了考虑固定的广义相对论时空背景中的场论之外，我们还希望把广义相对论自身作为一个拉格朗日场理论进行考虑。在这种情况中，时空度规 g 自身是动力学的，随不同的解发生变化。记住这个例子，即我们将允许一个有着任意几何、拓扑为 $M\times\mathbb{R}$ 的裸流形被看作是一个有利于当前目标的时空，使广义相对论能够与固定几何背景中的理论一起发展。注意，即便在一个类似于广义相对论的理论中，我们仍然能够把超曲面当作是相对于一个解 g 的瞬间来讨论。[③]

场值。我们把场值的空间 W 看作是一个有限维的矢量空间。但是，如果以下面的一些概念更加复杂为代价，我们也可以更加普遍。我们的 $\mathcal{K}$ 是 V 上的一个平凡向量丛的部分的空间，从而对于 $\Phi\in\mathcal{K}$，正切矢量 $\delta\Phi\in T_\Phi\mathcal{K}$ 也是从 V 到 W 的映射。我们可能允许 $\mathcal{K}$ 是一个任意纤维丛 $E\to V$ 的部分的空间。引起的最主要的困难在于，这样的话正切矢量 $\delta\Phi\in T_\Phi\mathcal{K}$ 就可能成为从 $\Phi^*T(E/V)$ 的

---

① 与外在场等相对应，我们也可以允许 V 携带非几何的且独立于解的结构。

② 这里的关键是，人们需要以以下方式选择边界条件和瞬间的概念：对某些 $(n-1)$ 形式 ω，对于任何瞬间 $\Sigma\subset V$，$\int_\Sigma\omega$ 都是收敛的，并且独立于瞬间的选择。在标准情况下，显而易见瞬间的概念得到满足。

③ 当然，在一个时空几何独立于解的理论中，$\Sigma\subset V$ 是一个相对于解 Φ 的瞬间，当且仅当相对于 V 的几何它是一个瞬间。

一个部分。

运动学可能的场。在构造一个严格的经典场论的时候，必须小心地选择运动学可能场的可微和边界条件。在这里我们可以忽略其细节，只去讲对于下面所考虑的每一个理论，$\mathcal{K}$都被看作是良好的函数 Φ：$V \to W$ 的空间，要求满足无限远处适当的可微条件和行为，而不是任意选取。注意当$\mathcal{K}$是一个流形时（往往甚至是一个仿射或者线性空间），一般来讲$\mathcal{S}$会是$\mathcal{K}$的一个有着温和奇点的非线性子空间。

微分方程。下面概述的拉格朗日框架非常普遍，并且不需要对微分方程的阶数做任何约束。但是，由于在后面的小节中我们总会对同一理论的哈密顿和拉格朗日形式进行对比，而且哈密顿框架采用二阶方程作为其出发点，因此，我们将从下面第 5 节开始重点关注这样的方程。

注释 12（作为场理论的有限维理论）。在一个有着有限自由度（粒子的有限系统、刚体等）的系统的经典理论中，位形空间 Q 是一个流形，决定着系统在物理空间中的可能分布参数。系统的历史是一条曲线 x：$t \in \mathbb{R} \mapsto x(t) \in Q$。我们可以通过令 $W = Q$ 和 $V = \mathbb{R}$（这样的话唯一的自变量是时间），从而把这样的理论放入当前的框架中。把这样一种理论当作场理论的简并情况来处理并没有任何坏处，只要人们没有忘记在这种情况中由理论自变量确定其参数的“时空” V 与系统所处的时空是不同的就行。

注释 13（标记）。因为 V 和 W 的选择隐含在$\mathcal{K}$的选择中，因此我们可以用（$\mathcal{K}$，Δ）标记一个场理论。

4.1 拉格朗日方法

理论微分方程 Δ 的作用是把运动学可能的$\mathcal{K}$的空间缩小到动力学可能的$\mathcal{S}$

的空间。[1] 拉格朗日方法的主要见解在于，对于经典物理学中出现的大多数方程，存在一个可替换的方式来选择其解的子空间。[2]

定义 14(拉格朗日量)。令$\mathcal{K}$为一个动力学可能场的空间。一个$\mathcal{K}$上的拉格朗日量 L，是一个从$\mathcal{K}$到 V 上的 n 形式空间的局域映射(说 L 是局域的，是说 $L(\Phi)$在 $x\in V$ 的值只取决于 Φ 及其有限个微商在 x 的值)。[3]

给定一个拉格朗日量 L，人们可以像在第 2.3 节中处理 n 体问题时那样，寻找那些动力学可能的 Φ，这些 Φ 有着特别的特性使得它那里的极小微扰与 $\int L(\Phi)$的值一样。

定义 15(变分问题)。注意，对于每一个紧致的 $U\subset V$，S_U：$\Phi\mapsto\int_U L(\Phi)$都是一个$\mathcal{K}$上的真值函数。让我们把 $U\mapsto S_U$ 称作 L 的变分问题。

定义 16(恒定场)。如果对于每一个紧致的 $U\subset V$，U 中的极小微扰 Φ 的效

---

① 这一节的内容是非正式的，下列术语和结果可能会有用。

空间 $V\times\mathcal{K}$是一个流形，因此携带了微分形式和外在的微商算子。对于 $0\leqslant p\leqslant n$ 和 $q\geqslant 0$，令 $\Omega^{p,q}(V\times\mathcal{K})$为$\mathcal{K}$上的 q 形式的空间，在 V 上的 p 形式空间中取值，因此，如果 $K\in\Omega^{p,q}(V\times\mathcal{K})$、$\Phi\in\mathcal{K}$并且 $\delta\Phi_1$，…，$\delta\Phi_q\in T_\Phi\mathcal{K}$，那么 $K(\Phi, \delta\Phi_1, \cdots, \delta\Phi_q)$是我们时空 V 上的一个 p 形式。$V\times\mathcal{K}$上的每一个微分形式都属于某个 $\Omega^{p,q}(V\times\mathcal{K})$。而且，我们可以写出 $V\times\mathcal{K}$上的外导数 d 为 $d=D+\partial$，这里 D 是 V 上的外导数，把 $\Omega^{p,q}(V\times\mathcal{K})$的元素映射在 $\Omega^{p+1,q}(V\times\mathcal{K})$的元素上(对于 $0\leqslant p<n$)，∂ 是$\mathcal{K}$上的外导数，把 $\Omega^{p,q}(V\times\mathcal{K})$的元素映射在 $\Omega^{p,q+1}(V\times\mathcal{K})$的元素上。我们有$\partial D=-D\partial$。

注意，如果 $\Phi\in\mathcal{K}$，那么一个正切矢量 $\delta\Phi\in T_\Phi\mathcal{K}$自身就是一个从 V 到 W 的映射，因此对于每一个可能的 p 和 q，我们都能考虑局域形式的子空间 $\Omega^{p,q}_{loc}(V\times\mathcal{K})\subset\Omega^{p,q}(V\times\mathcal{K})$，这个子空间由那些 K 组成，使得对于任何 $\Phi\in\mathcal{K}$和 $\delta\Phi_1$，…，$\delta\Phi_q\in T_\Phi\mathcal{K}$，$p$ 形式 $K(\Phi, \delta\Phi_1, \cdots, \delta\Phi_q)$在点 $x\in V$ 的值只取决于 Φ，$\delta\Phi_1$，…，$\delta\Phi_q$ 在 x 点的值以及它们的有限多个导数的值。

② 对拉格朗日方法适用范围的讨论，见[Bluman(贝尔曼)，2005，§2.1]。

③ 即 $L\in\Omega^{n,0}_{loc}(V\times\mathcal{K})$。

应不会影响到 S_U 的值,① 我们就称 $\Phi\in\mathcal{K}$对于 L 是恒定的。

定义 17(Δ 许可的拉格朗日量)。如果对于 L 恒定的 Φ 的集合与 Δ 的解空间$\mathcal{S}$一致，我们称 L 为($\mathcal{K}$，Δ)的拉格朗日量。在这种情形中，我们也说 Δ 容许拉格朗日量 L，并把$\mathcal{S}$称为($\mathcal{K}$，L)的解空间。

注释 18(欧拉—拉格朗日方程)。给定一个拉格朗日量，我们总是可以找到一组方程组 Δ(L 的欧拉—拉格朗日方程)使得 L 是 Δ 的拉格朗日量。也就是说，一个动力学可能的场 Φ：$V\to W$ 对于拉格朗日量 L 是恒定的，当且仅当 L 的欧拉—拉格朗日方程被满足。因为拉格朗日量只取决于场及其一阶导数，这些方程要求：

$$\frac{\partial L}{\partial \Phi^{\alpha}}(x_a)-\sum_{a=1}^{n}\frac{\partial}{\partial x_a}\left(\frac{\partial L}{\partial \Phi_a^{\alpha}}\right)(x_a)=0 \tag{4}$$

在每个点 $x\in V$ 上都成立(这里 a 指示的是 V 上的坐标，α 指示 W 上的坐标，Φ_a^{α} 表示$\frac{\partial}{\partial x_a}\Phi^{\alpha}$)。②

注释 19(平凡相异的拉格朗日量)。如果拉格朗日量 L 和 L'对每一个 $\Phi\in\mathcal{K}$都有着形式 $L'(\Phi)=L(\Phi)+\alpha_{\Phi}$，其中 α_{Φ} 是 V 上一个确切依赖于 Φ 的 n 形

---

① 如果对每一个紧致的 $U\subset V$ 及其每一个支撑都包含在 U 中的 $\delta\Phi\in T_{\Phi}\mathcal{K}$，我们发现$\partial S_U(\delta\Phi)=\int_U \partial L(\Phi,\delta\Phi)$为零，那么 Φ 对于 L 来讲是固定的。我们可以这样考虑：固定 Φ、U 和 $\delta\Phi$，我们找到一条曲线 $\Phi[\varepsilon]$：$[-1,1]\to\mathcal{K}$，有 $\Phi[0]=\Phi$ 和$\frac{d}{d\varepsilon}\Phi[\varepsilon]|_{\varepsilon=0}=\delta\Phi$；$\partial S_U(\delta\Phi)=0$ 的要求等同于$\frac{d}{d\varepsilon}\int_U L(\Phi[\varepsilon])|_{\varepsilon=0}=0$。

② 当然，存在一个独立于坐标的描述。有可能表明$\partial L=E+DM$，其中 $E\in\Omega_{loc}^{n,1}(V\times\mathcal{K})$，$M\in\Omega_{loc}^{n-1,1}(V\times\mathcal{K})$，$E$ 则唯一地取决于 L，M 取决于附加一个确切的形式 DN，其中 $N\in\Omega_{loc}^{n-2,1}(V\times\mathcal{K})$。对于所有其支撑都包含在 U 中的 $\delta\Phi$，$\partial S_U(\delta\Phi)=0$ 的条件变为$\int_U E(\Phi,\delta\Phi)+DM(\Phi,\delta\Phi)=0$。由于$\delta\Phi$ 沿 U 的边界为零，斯托克斯定理告诉我们第二被积函数不起任何作用。因此，Φ 是恒定的，当且仅当对于所有这样的 U 和可容许的 $\delta\Phi$，有$\int_U E(\Phi,\delta\Phi)=0$——这就等于说对于所有这样的 $\delta\Phi$，有 $E(\Phi,\delta\Phi)=0$。相对于坐标，最后的方程在拉格朗日情形中等价于方程(4)，只取决于一阶导数。

式，那我们就说L和L'是平凡相异的。① 如果L和L'为拉格朗日量，并且对于每个紧致的$U\subset V$和场$\Phi\in\mathcal{K}$，Φ的任何极小微扰都不会改变S_U的值，当且仅当它不改变S_U'的值，② 这样它们的变分问题$U\mapsto S_U$和$U\mapsto S_U'$是等价的。平凡相异的拉格朗日量有着等价的变分问题。③ 因此，平凡相异的拉格朗日量有着同样的解空间——事实上，它们有着相同的欧拉—拉格朗日方程。④

注释20(拉格朗日量的唯一性)。前一个注释表明，如果Δ确实容许一个拉格朗日量，那么它就容许无限多个平凡相异的拉格朗日量。有些Δ还容许多个并不平凡相异的拉格朗日量——比如，运动在三维欧几里得空间球面势中的粒子的牛顿方程。⑤

注释21(拉格朗日量的存在性)。注意，并非每一个方程组Δ都容许拉格朗日量。⑥ 运动在磁单极子的电磁场中的带电粒子就是系统的一个例子，它并不容许拉格朗日的处理方法。⑦

---

① 即$L'=L+DK$，其中$K\in\Omega^{n-1,0}_{loc}(V\times\mathcal{K})$。

② 即拉格朗日量L和L'的变分问题$U\mapsto S_U$和$U\mapsto S_U'$是等价的，如果对于每一个紧致的$U\subset V$，每一个场$\Phi\in\mathcal{K}$，以及每一个正切矢量$\delta\Phi\in T_\Phi\mathcal{K}$，其支集包含在$U$中，我们有$\partial S_U(\Phi)(\delta\Phi)=0$，当且仅当$\partial S_U'(\Phi)(\delta\Phi)=0$。

③ 令$L'=L+DK$，其中$K\in\Omega^{n-1,0}_{loc}(V\times\mathcal{K})$，那么对于任意紧致的$U\subset V$，$\Phi\in\mathcal{K}$和$\delta\Phi\in T_\Phi\mathcal{K}$，其支撑包含在$U$中，我们有$\partial S_U(\Phi)(\delta\Phi)-\partial S_U'(\Phi)(\delta\Phi)=\int_U\partial DK(\Phi)(\delta\Phi)$。但是$\partial D=-D\partial$，因此右边是$-\int_U D\partial K(\Phi)(\delta\Phi)$，它为零(因为斯托克斯定理和$\delta\Phi$在$U$的边界为零)。

④ 即如果拉格朗日量L和L'由于一个DK形式的项而不同，那么它们共有同一个欧拉—拉格朗日算子E。

⑤ 这个例子见[Crampin and Prince(克兰平和普利斯)，1988]和[Henneaux and Shepley，1982]。场论的例子见[Nutku and Pavlov(乌特库和帕夫洛夫)，2002]。一个能充分地保证Δ并不容许非平凡不同的拉格朗日量的$V\times W$上的拓扑条件，见[Anderson and Duchamp(安德森和杜尚)，1980，Theorem 4.3.ii]。

⑥ 决定一个给定微分方程集是否容许拉格朗日量的问题，是在数学家们中被称之为变分计算的反问题，也是物理学家所说的亥姆霍兹(Helmholtz)问题。[Prince，2000]是一个关于有限维系统结果的有价值的研究。[Anderson and Duchamp，1980，§5]包含了不承认拉格朗日形式的场论的例子。

⑦ 见[Anderson and Thompson(安德森和汤普森)，1992，4f.]。其他例子见[Prince，2000]。

4.2 解空间的结构

拉格朗日量 L 的选择允许$\mathcal{S}$具备一个封闭的 2 形式 Ω。[①] 因此，当 Ω 是非简并的，($\mathcal{S}$, Ω)就是一个辛空间，否则它将是预辛空间。[②] 粗略地讲，人们认为 Ω 是非简并的，当且仅当我们理论的方程 Δ 有如下特性：指定的初始数据决定唯一的不可延拓解。[③]

拉格朗日量的选择带来了构造量子理论所需要的一类结构。单独一组微分方程似乎并不能确定这样的结构，并且人们还不知道怎样才能不引入拉格朗日量或者哈密顿量就直接量子化一个微分方程。如果 Δ 容许拉格朗日量 L，那么它也就容许一系列与 L 平凡相异的拉格朗日量(见上述注释20)。自然地，平凡相异的拉格朗日量导致了$\mathcal{S}$上相同的 Ω。[④] 但是，当 Δ 容许非平凡相异的拉格朗日量 L 和 L'时，这些拉格朗日量就能够导致$\mathcal{S}$上不同的几何结构。并且，人们认为这些不同的(预)辛结构将会导致对给定的经典理论的不同量子化。[⑤] 在

---

① 回顾一下之前的注释，我们有分解$\partial L = E + DM$，其中 E 和 M 由于附加了一个 D 确切的形式而成为唯一的。我们现在定义 $Z := \partial M$。由于其中附加了一个 DY 形式的项 $Y \in \Omega_{loc}^{n-2,2}(V \times \mathcal{K})$，因此 $Z \in \Omega_{loc}^{n-1,2}(V \times \mathcal{K})$且唯一地取决于 L。

令 $\Phi \in \mathcal{S}$为一个解，$\delta\Phi_1$，$\delta\Phi_2 \in T_\Phi \mathcal{S}$，$\Sigma \subset V$ 为一个相对于 Φ 的瞬间。然后我们定义 $\Omega_\Sigma(\Phi, \delta\Phi_1, \delta\Phi_2) := \int_\Sigma Z(\Phi, \delta\Phi_1, \delta\Phi_2)$。我们假设在无限远处有很好的边界条件，使得 Ω_Σ是稳定的，因此用 $Z + DY$ 代替 Z 对 Ω_Σ 不会产生任何影响。我们发现，Ω_Σ 的值独立于瞬间的选择——因为 $Z(\Phi, \delta\Phi_1, \delta\Phi_2)$是一个 V 上的封闭$(n-1)$形式，而且我们在选择瞬间的概念的时候已经很小心了。因此，我们可以忽略注释，把 Ω 看作是一个$\mathcal{S}$上的 2 形式，由于 Z 是 ∂ 确切的因而 Ω 是封闭的。

② 拉格朗日似乎是第一个用辛结构解决动力学问题空间的人。详见[Weinstein, 1981, §2]、[Souriau(苏里奥), 1986]或[Iglesias(伊格莱西亚斯), 1998]。

③ 正如我们将会在6.2 中看到的，如果运动方程承认规范对称性(那么唯一性将会以某种戏剧性的方式失效)，那么 Ω 是预辛的。我相信，人们广泛认为这是 Ω 不能成为辛的的唯一解释——至少对于物理中这样的例子来说是这样。

④ 用 $L + DK$ 替代 L，改变了有着 DY 形式的项 Z，$Y \in \Omega_{loc}^{n-2,2}(V \times \mathcal{K})$。但由于它是 D 确切的，这个新项对于定义 Ω 的空间上和边界条件)的积分没有贡献，(见斯托克斯理论(stokes's theorem)。

⑤ 当(像在牛顿情形中一样)运动方程是二阶的，解空间是有限维的，那么拉格朗日量 L 和 L'会在解空间上得到同样的 2 形式，当且仅当它们平凡相异时。见[Crampin and Prince, 1988, §II]。这可能在更广泛的情况下也成立。

上述粒子运动于球面势的例子中，这些元素的每一种都存在，多个非平凡相异的拉格朗日量导致了解空间上的不同辛结构，这转而导致了物理上不同的量子化。①

4.3 对称性和守恒量

给定一组方程组 Δ 和一个 Δ 容许的拉格朗日量 L，我们可能会考虑三种不同的对称性概念。② 粗略地讲，Δ 的对称性是从$\mathcal{K}$到它自身的映射，此映射将S作为一个集合并且由场中的量和场的派生物生成。③ 然后我们可以考虑变分对称性的子集或者拉格朗日对称性的子集。变分对称性允许 L 的变分是不变的，而拉格朗日对称性则允许 L 自身是不变的。这三个概念是不同的：每个拉格朗日对称性都是变分对称性，但是有些理论拥有的变分对称性却并非拉格朗日对称性；类似地，每个变分对称性都是相关运动方程的对称性；但有些承认拉格朗日量的方程所拥有的对称性却并非理论的任何拉格朗日量的变分对称性。④

出于当前的目的，聚焦于物理学理论的变分对称性是很自然的。因为，一方面，拉格朗日对称性的类别并不包括某些物理上重要的对称性——并且在任何情况下，在拉格朗日方法中聚焦于拉格朗日量的对称性比聚焦于变分问题的对称性更加自然并不是一件明朗的事。另一方面，非变分对称方程的对称性种类看起来并不包括任何最重要的物理兴趣的对称性——正是在变分对称性(而不是在方程对称性)的层次上，诺特的加强变分对称的某些特定类型的单参数群和经典场论中的某些特定类型的守恒量之间的联系的权威结论，很自然地成

---

① 详见[Henneaux and Shepley，1982]。

② 相关的观点见[Olver(奥尔弗)，1993，Chapters 2，4 and 5]。注意：术语在变化——有时候我的拉格朗日对称被叫作变分对称，有时候我的变分对称被称作发散对称，等等。

③ 详细讨论见[Olver，1993，§5.1]。

④ (2+1)维的波动方程有着扩张的对称，它不是变分对称，也不是变化却非拉格朗日的反演对称，见[Olver，1993，Examples 2.43，4.15，4.36，and 5.63]。同一文献中的例子4.35 表明伽利略加速对于 n 体问题是变分对称的但不是拉格朗日对称的。事实上，没有牛顿粒子的拉格朗日量遭遇了来自一个在新牛顿时空的对称完全群之下不变的势的力。见[Souriau，1997，Remark 12.136]。

立了。[①]

这里是核心结论的陈述。如果一个从$\mathcal{K}$到它本身的微分同胚的单参数群$\xi=\{g_t\}$的无限小生成元不改变L的变分问题并且在适当意义上是局域的,[②] 就让我们称其为L的诺特群。给定一个关于($\mathcal{K}$, L)的诺特群$\xi=\{g_t\}$，存在映射J_ξ，称之为与ξ相关的诺特流，在解与V上的($n-1$)形式之间建立起映射关系。[③] 给定一个任意解$\Phi\in\mathcal{S}$，以及一个瞬间$\Sigma\subset V$，我们在Σ上对$J_\xi(\Phi)$进行积分来给出诺特荷，$Q_{\xi,\Sigma}(\Phi):=\int_\Sigma J_\xi(\Phi)$。我们注意到，$Q_{\xi,\Sigma}(\Phi)$是独立于$\Sigma$的选择的(只要积分是确切定义好的)。[④] 也就是说，$Q_{\xi,\Sigma}(\Phi)$这个量在解(V, Φ)中是不随时间变化的常量。这样我们也可以把它简单表示为$Q_\xi(\Phi)$，并把与ξ相联系的诺特荷Q_ξ看作是$\mathcal{S}$上的一个函数。

注释22(诺特荷生成对称性)。由于Ω是一个封闭的2形式，($\mathcal{S}$, Ω)是辛空间或预辛空间，因此前面3.2和3.3节的结果得以应用。正如有人认为的，Q_ξ事实上是单参数群ξ(现在把它看作是作用在$\mathcal{S}$上)的辛或预辛生成元。诺特结果的优美之处在于它表明了如何通过局域物体在时空中的整合来构造ξ的生成元。

---

① 但是注意，确实存在着在方程的对称性和守恒量之间建立起联系的结果，不需要绕过拉格朗日框架，见[Bluman，2005]。

② 更精确地，令ξ是一个从$\mathcal{K}$到自身的微分同胚的单参数群，令X是$\mathcal{K}$上对应的矢量场(即X是流为ξ的矢量场)。如果以下两个条件成立，ξ是诺特群，即(i)X是L的一个无限小的变分对称：存在$R\in\Omega_{loc}^{n-1,0}(V\times\mathcal{K})$，使得对于所有的$\Phi\in\mathcal{S}$，$\partial L(\Phi,(X(\Phi)))=DR(\Phi)$；(ii)$X$是局域的：对于任何$\Phi\in\mathcal{K}$，$X(\Phi)\in T_\Phi\mathcal{K}$在$V$上是局域的，在任意一点$x\in V$的意义上，我们发现$X(\Phi)(x)$只与$x$在$\Phi$的值以及其有限多的导数有关(记得$T_\Phi\mathcal{K}$的一个元素自身就是从$V$到$W$的映射)。

③ 与L和ξ相联系的诺特流是$J_\xi\in\Omega_{loc}^{n-1,0}(V\times\mathcal{K})$的元素，它由$J_\xi(\Phi):=R(\Phi)-M(\Phi,X(\Phi))$给出。这里$X$是$\xi$的无限小生成元，$R$是上一个脚注中引入的对象。

④ 由于$J_\xi(\Phi)$封闭为一个V上的($n-1$)形式，并且由于我们在选择瞬间观念的时候很小心(见函数47)。注意，事实上，只要Σ，$\Sigma'\subset V$是决定V中的相同同调类的紧致($n-1$)维流形，我们就会有$\int_\Sigma J_\xi(\Phi)=\int_{\Sigma'}J_\xi(\Phi)$，相关的观念和结果见[Lee(李)，2003，431]和[Lee，2000，300f.]等。因此我们得到了一种守恒定律，即便根据V的几何Σ和Σ'并不是类空的。对这种守恒定律的介绍，见托尔(Torre)的相关著作。

注释 23(平凡的守恒定律)。到目前为止，我们还没有说有任何东西可以保证 Q_ξ 是S上一个有意义的函数——比如，如果 $J_\xi(\Phi)$确切地是 V 上的一个$(n-1)$形式，那么它可能是一个零函数。当 Ω 是预辛的并且 ξ 是一组规范变换的时候，这样的平凡诺特荷事实上确实会发生。我们将在下面的 6.2 节看到相关的例子。

5. 良态场理论中的时间和变化

现在我们转向物理学理论中时间和变化的表示。在余下的章节中，理论的哈密顿公式将起到重要作用。因此，从现在开始我们主要关注那些含有二阶运动方程的理论。

在这一节中，我们讨论理想的良态理论。我们进一步提出三个假设，这些假设实际上只与这一节相关：(a)解的全域存在性；(b)解的唯一性；(c)我们的时空容许时间平移对称性，在这种对称性下我们的拉格朗日变分问题是不变的。

在这种背景下我们将会看到，我们有三种$\mathbb{R}$行为，基于时空的时间平移概念和基于理论解空间的时间平移概念和基于理论初始数据空间的时间演化概念。我们还发现，解空间和初始数据空间是作为辛空间同构的，在基于解空间的时间平移概念和基于初始数据空间的时间演化概念之间有着自然的相互交错。因此在这一领域人们可以简单地说，时间被表示为定律的一种对称性——而对它意味着时间平移还是时间演变则不作结论，因为最终它们两个在很大程度上成为了相同的东西。

这一节有 5 个小节。第 1 小节关注拉格朗日图景，第 2 小节关注哈密顿图景，第 3 小节关注这些图景之间的关系，第 4 小节关注一个关于时间和变化的表示的讨论，最后一小节进行总结回顾。

5.1 拉格朗日图景

我们在对待这一节中运用的特定假设的时候要更加精确。我们给时空 V、运动方程 Δ 和拉格朗日量 L 加上下列条件：

解的全域存在性。我们假定 Δ 的每一组可容许的初始数据都与所有 V 上定义的解一致。①

解的唯一性。如果 Φ 和 Φ'在它们瞬间 $\Sigma \subset V$ 产生的初始数据上有一致的解，那么它们在任何它们定义于其上的点 $x \in V$ 上一致。

拉格朗日量的时间平移不变性。我们要求我们的时空 V 具有非平凡的几何结构，足够确定一类$(n-1)$维子流形作为瞬间和确定一类一维子流形作为点粒子的可能世界线。令 $\bar{\xi} = \{\bar{g}_t\}$ 为 V 的时空对称性的一个单参数群，考虑 $\bar{\xi}$ 在 V 内的轨道($\bar{\xi}$ 经过一点 $x \in V$ 的轨道$[x]$是曲线 $x(t) := \bar{g}_t \cdot x$ 的图像)。如果根据 V 的几何，$\bar{\xi}$ 的轨道是点粒子的可能世界线，那么我们称 $\bar{\xi}$ 是 V 时间平移群。在这种情况下，我们称这些轨道为符合 $\bar{\xi}$ 的世界线。我们通常将时间平移群表示为 $\bar{\tau}$。

令 $\bar{G}$ 为 V 的时空对称群。给定 $\bar{g} \in G$，我们可通过 $g(\Phi(x)) = \Phi(\bar{g}^{-1} \cdot x)$ 定义一个微分同胚 g：$\mathcal{K} \to \mathcal{K}$。在正统的拉格朗日理论中，人们认为如果 $\bar{\xi} = \{\bar{g}_t\}$ 是时空对称性的一个单参数群，那么 $\xi = \{g_t\}$ 就是 L 的诺特群。在这种情况下，ξ 将在解和解之间建立起映射关系，使得每一个 $g_t \in \xi$ 都被限制在从$\mathcal{S}$到它本身的映射中，这些映射是$(\mathcal{S}, \Omega)$的辛自同构(我不会麻烦地引入符号来区分 ξ 在$\mathcal{K}$上的作用量和$\mathcal{S}$对这类作用量的限制)。在本节中我们假设 V 的每一时间平移群 $\bar{\tau}$ 以这种方式引出了 L 的诺特群 τ。我将把这样的 τ 叫作一个动力学时间平移群。

在物理学中出现的这类理论中，L 在解空间中引发的形式 Ω 是非简并的，因而也是辛的，这似乎是唯一性假设的直接结果。我们用 H 表示相应的由诺特定理保证的守恒量(在物理学的实际理论中，H 是通过在任意时间上对场的应力—能量进行积分而得到的)。② 当然，H 是通过 Ω 产生$\mathcal{S}$上 τ 的作用的。

例子 24(牛顿时空中的场论)。在牛顿时空中，每一种对称性都可能被写

① 由于我们把注意力限制在含有二阶运动方程的理论中，确定初始条件包括确定某个初始瞬间的场值及其时间变化率。

② 应力能量张量及其在下面例子中的作用，见[Choquet-Bruhat and DeWitt-Morrette, 1989, §II.7]和[Deligne and Freed, 1999, §2.9]。

作是绝对空间等距性中时间平移的产物。在采用的优越的绝对参照系的坐标系中，我们可以把时空点写作$(t, \mathbf{x})$。那么V的(保向)对称性的形式就是$(t, \mathbf{x}) \mapsto (t+s, R(\mathbf{x})+\mathbf{c})$，这里$s \in \mathbb{R}$实现时间平移，$R$是一个矩阵，实现绝对空间中的转动，$\mathbf{c} \in \mathbf{R}^3$实现空间平移。由于时间单位的选择，存在唯一的时间平移群$\bar{\tau}$：$(t, \mathbf{x}) \mapsto (t+s, \mathbf{x})$；绝对空间点的世界线适应于这个群。我们假定了作用于解空间的对应群τ是一个动力学时间平移群。与τ、H：$\mathcal{S} \to \mathbb{R}$相关的诺特荷在任意时刻为每一个解分配系统的总能量(由于我们正在考虑一个时间平移变换下不变的理论，因此沿着一个截面的总能量的值为常量)。①

例子25(闵可夫斯基时空中的场论)。闵可夫斯基时空的对称群是庞加莱(Poincaré)群。每个惯性系会选出一个同时性的概念以及一个时间平移群$\bar{\tau}$。在选出的坐标系中，静止观测者的世界线会适应于这个群(等价地，这样的群取决于时空中类时矢量的选择)。在庞加莱不变性场论中我们可以选择惯性坐标(t, x_1, x_2, x_3)，这使得我们的选择$\bar{\tau}$通过$(t, x_1, x_2, x_3) \mapsto (t+s, x_1, x_2, x_3)$起作用。在这样的坐标中，诺特流仅仅是场的应力—能量张量的组分T^{00}——和往常一样，诺特荷是通过诺特流在任意时间上的积分给定的。②

例子26(弯曲时空中的场论)。尽管一般的广义相对论时空不承认非平凡对称性，直观地在时间上受限的空间几何中的解却承认一个时间平移群。令V为拥有这样一个$\bar{\tau}$的全域双曲且时间定向的广义相对论时空。令X_a为一与$\bar{\tau}$的轨道相切的矢量场，因此X是一个类时(timelike)的基灵(Killing)场。令$T^{ab}(\Phi)$为场Φ的应力—能量张量，并假定$\nabla_a T^{ab}=0$，这常常在有物理意义的情况下成立。令$\Sigma \subset V$为一个瞬间(即一个柯西面)，令n_a为沿着Σ的单位法线向量场。我们可以定义相对于X_a的T^{ab}的能量—动量矢量为$P^b := X_a T^{ab}$，定义沿

---

① 产生了空间在一个给定方向(差不多是一个给定的轴线)上的平移(转动)的诺特荷为一个解分配了系统在某个瞬间对应的线性(角)动量的组分。事实上，我们得到了一个牛顿时空对称群行为的动量映射(见函数34)——诺特荷的泊松括号代数反映了对应的单参数群的无限小生成元之间的李括号关系。但是，不可能构造新牛顿时空对称群的动量映射。关于这和其他不可能构造动量映射的例子见[Woodhouse，1991，§3.4]。

② 人们可以重新构造一个动量映射(见函数34)——其类空变换由线性动量的组分等以熟悉的方式产生。

着Σ的能量为$\int_{\Sigma} P^{a} n_{a} dx$。这个最后的量实际上就是诺特荷，它独立于$\Sigma$。

5.2 哈密顿图景

哈密顿方法背后的基本思想是运用理论方程的初始数据空间而不是运用方程的解空间——粗略且启发性地讲，这意味着运用理论瞬间态的空间而不是运用它的可能世界的空间。

确定的运动方程将告诉我们，如果一个系统的态处于一个给定的初始态中，那么它在更早和更晚的时候一定是什么样子。因此，至少对于良定的运动方程，运动方程的动力学内容应当可以在初始数据空间上的一个流中编码，这个流的积分曲线应当是动力学通过瞬时态的空间的可能轨迹。

这一节的特定假设(全域存在性、解的唯一性和动力学时间平移群的存在性)意味着(至少对于物理学中出现的这种理论)初始数据空间携带辛结构，这种结构辅以在那个态中为初始数据组分配系统的总能量的函数，从而产生了理论的动力学。这种动力学可以被看作是在实现时间演化的初始数据空间上的一个$\mathbb{R}$行为中编码。正如我们将要看到的，这些初始数据空间上的结构——辛结构、哈密顿量和群作用量——都与拉格朗日方法中出现的解空间中相应的对象密切相关。

直观地讲，场的一个瞬时态是一种对场值空间的每一点及其变化时率的说明。在给定这种瞬时态的一个序列时，我们描述了这些变量的值是如何在空间的每一点上随时间演化的。因此，如果一个理论中系统的整个历史都是由初始数据空间中的轨迹描述的，那么为了构造这个理论的哈密顿公式，我们需要把时空观念性地分解为空间和时间。①

非正式地，我们可以想象什么是作为一族首选的观测者连同同时性的概念而被要求的。时空被这些观测者的世界线分割开来(这些观测者不需要相对于对方静止，但是根据V的几何，我们确实要求相关的世界线是点粒子的可能世

---

① 注意我们在上面第4节设置拉格朗日形式时没有要求任何这样的分解。当然，区分从这种形式中产生的解的辛空间和某些拉格朗日方法中出现的辛速度相空间是非常重要的——后者预设了时空可以分割为瞬间，但前者没有。

界线)。每个观测者都携带一个时钟，我们假定这些时钟的读数为$t=0$的一组点形成V中的一个瞬间。我们称这样一组具有这种同时性概念的观测者为一个参照系。我们说，当下面两个条件获得满足的时候，参照系适合于时间平移群$\mathcal{T}$：(i)每个观测者的世界线是作用在V上的群$\mathcal{T}$的一个轨道；(ii)$\mathcal{T}$给我们的V的瞬间组的参数取决于原点的选择和测量单位的选择。这个参数使我们可以确定时间间隔的比率，我们要求时钟读数遵守它。

仍然是非正式地，我们可以说，相对于参照系的选择，在时刻t场的态是场值和动量对于每个观测者的赋值(即在一个给定的瞬间，观测者占据的时空点上的场值及其变化时率)，而场的历史则是这些量沿每个观测者的世界线的赋值。因此我们可以把一个初始数据集看作一对定义在我们参照系中的观测者空间上的函数(对应于场值及其变化时率)——这个空间作为一种抽象的瞬间起作用，它与作为V的子集出现的具体瞬间具有相同的拓扑与几何结构。

我们可以如下对其进一步精确化。

定义27(片断)。令V为具有几何结构的时空，S为$(n-1)$维流形(可能具有黎曼度规)。那么V的S片断是一个微分同胚σ：$\mathbb{R}\times S\rightarrow V$，使得：(i)每一个$\Sigma_t := \sigma(\{t\}\times S)$，其中$t\in\mathbb{R}$，是$V$中的一个瞬间($\sigma$提供了$\Sigma_t$的几何和$S$的几何之间的同构性，如果这种同构性存在的话)；(ii)每个$X_x := \sigma(\mathbb{R}\times\{x\})$，其中$x\in S$，是按照$V$的几何得到的点粒子的一条可能世界线。我们称$S$为$\sigma$的抽象瞬间，每个$\Sigma_t$为片断中的一个瞬间。当$V$容许一个时间平移群$\mathcal{T}$时，如果下列条件被满足，我们就称$V$的一个片断$\sigma$适应于$\mathcal{T}$：(a)每一个$X_x$都是$\mathcal{T}$的一个轨迹；(b)片断的任意两个瞬间通过$\mathcal{T}$中的时间平移群相关联；(c)根据选择的单元和原点，由σ给出的每个X_x的参数和由$\mathcal{T}$给出的每一个X_x的参数相一致。

例子28(牛顿片断)。在牛顿时空中，当然存在唯一的时空分割，这种分割是通过瞬间和(取决于单元的选择)唯一的时间平移群$\mathcal{T}$进行的。而且，在这种背景下，有可能把S看作是绝对空间的点的世界线空间。[①] 因此在构造适应于

---

① 这个空间具备一个自然的欧几里得结构——由于绝对空间中的点之间的距离在时间中是不变的，我们可以把这种点组成的两条世界线之间的距离定义为点之间的距离。

$\bar{\tau}$ 的片断时唯一的自由度是通过现实的东西为瞬间的参数选择一个原点和单位。

例子 29(平直闵可夫斯基片断)。在闵可夫斯基时空背景下，有时候注意力很自然地被限制在那些作为惯性观测者的同时性的曲面而出现的瞬间上。在这种情况下，我们的抽象瞬间 S 将再一次具有欧几里得空间的结构。为了构造一个片断，我们必须选择一个与 $t=0$ 对应的瞬间 $\Sigma_0 \subset V$，一个从 S 到 Σ_0 的等距性，一个时间测量的单元，以及一个与惯性观测者相联系的时间平移的概念。

例子 30（一般的闵可夫斯基片断）。更一般地，在闵可夫斯基时空背景下，有可能允许把任意的柯西面看作是瞬间——在这种情况下，人们会选择拥有某种非平凡的黎曼几何的 S。现在有真正巨大的——事实上是无穷维的——一族瞬间，可以从中进行选择(当我们允许 S 的几何有所不同的时候)。令人高兴的是，一般的瞬间不容许任何非平凡的等距性——因此一旦选择了 S 和 Σ，在构造它们从一个到另一个的等距性的时候，将不存在任何自由。

让我们考虑一个满足所有当前条件的拉格朗日理论，确定一个 V 的片断，它要适应于时间平移 $\bar{\tau}$ 的概念。这将引出一个动力学时间平移群 τ，然后我们可以按照下面的步骤构造哈密顿版本。

1. 给定一个瞬间和一个解，构造瞬时场的结构和动量。令 Σ 为包含在给定的片断中的瞬间，令 Φ: $V \to W$ 为一个解。我们通过 $\phi := \Phi|_{\Sigma}$ 定义 ϕ: $\Sigma \to W$，它是 Σ 上的场结构。然后如下定义 Σ 上的场速度 $\dot{\phi}$: $\Sigma \to W$: 在每一点 $x \in \Sigma$ 上，$\dot{\phi}(x)$ 是场值沿 $\bar{\tau}$ 通过 x 的轨道在 x 点上的变化率。[①] 为了构造场的瞬时动量，我们运用构造正则动量变量的常规方法定义 $\pi := \dfrac{\partial L}{\partial \dot{\phi}}$($\pi$ 是一个从 Σ 到 W^* 的映射，是与 W 对偶的矢量空间)。

2. 给定瞬时场的结构和动量，构造相应的初始数据。这仅仅是关于运用 σ 拖曳 ϕ 和 π 的问题，因此我们可以把它们看作是 S 上而不是 Σ 上的函数。我将不那么精确地对 S 上定义的初始数据和相应地定义在 $\Sigma \subset V$ 上的函数使用相同

① 即令 $x_0 \in \Sigma$ 并发现 $y_0 \in S$，$t_0 \in \mathbb{R}$，因而 $\sigma(t_0, y_0) = x_0$，通过 $x(t) := \sigma(t, y_0)$ 并定义曲线 x: $\mathbb{R} \to V$，然后令 $\dot{\phi}(x) := \lim_{h\to 0} \frac{1}{h}(\phi(x(t_0+h)) - \phi(x_0))$。

的名称。

3. 构造初始数据的空间$\mathcal{I}$。令$\mathcal{Q}$为所有ϕ: $S\to W$的空间，当我们允许Φ在$\mathcal{S}$中变化时，它可以通过前面两步出现。① 所有能够通过这些步骤出现的(ϕ, π)的集合就是余切丛$T^{*}\mathcal{Q}$。这个空间是我们理论的初始数据空间$\mathcal{I}$，它具有正则辛结构ω(见例子7)。

4. 构造一个哈密顿量。我们如下在初始数据空间上定义哈密顿量h: $\mathcal{I}\to\mathbb{R}$：令(ϕ, π) $\in\mathcal{I}$为初始数据，$\Sigma\subset V$为一个瞬间(并不必然在我们的片断中)。令Φ为给出Σ上的(ϕ, π)的解，定义$h(\phi,\pi) := \int_{\Sigma}\pi(x)\dot{\phi}(x) - L(\Phi)(x)dx$(在当前情况中，结果并不取决于选择的瞬间)。②

5. 构造动力学。h和ω共同确定在$\mathcal{I}$上的矢量场$\mathcal{X}$，这一矢量场编码了我们理论的动力学。$\mathcal{X}$的积分曲线是可能的动力学轨迹——如果在$t=0$时的态为(ϕ_0, π_0)，那么t个单位时间之后的态可以通过沿着经过(ϕ_0, π_0)的积分曲线追踪时间的t个单位而得到。这给了我们一个$\mathcal{I}$上的流，它同时保留了h和ω(这个流是全域的而不是局域的，因为我们假定其解是对于t的所有值而定义的)。

5.3 拉格朗日图景和哈密顿图景之间的关系

对于我们的片断σ中的每个瞬间$\Sigma_t := \sigma(\{t\}\times S)$，我们定义$T_{\Sigma_t}$: $\mathcal{S}\to\mathcal{I}$为一个映射，它在当片断$\sigma$被用来把$\Sigma_t$上的$\Phi$诱导的初始数据拖曳到$S$的时候为初始数据集($\phi$, π) $\in\mathcal{I}$提供解Φ。因为我们假设了给定初始数据后解的全域存在性和唯一性，因此T_{Σ_t}实际上是一个双射。事实上，它是一个微分同胚。而且$T^{*}_{\Sigma_t}\omega=\Omega$，因此每个$T_{\Sigma_t}$事实上是($\mathcal{S}$, Ω)和($\mathcal{I}$, ω)之间的一个辛同构。

注意，在典型的理论中，片断中不同的瞬间导致不同的同构。如果Σ_t和$\Sigma_{t'}$是我们片断中的瞬间并且$T_{\Sigma_t}=T_{\Sigma_{t'}}$，那么，对于每一个解$\Phi$，$\Phi$将诱导出$\Sigma_t$和$\Sigma_{t'}$上相同的初始数据——即每个解的周期为$|t-t'|$。因此如果对于每

① 只要S及其几何是固定的，那么允许Σ和片断可以变化对目前的例子来说也同样不会造成任何不同。

② 这里我们运用以下事实：π在W^{*}中取值而ϕ在W中取值，我们依赖V的几何引入的自然测量来容许我们把$L(\Phi)$处理为一个函数而不是一个n形式。

个Σ_t和$\Sigma_{t'}$有$T_{\Sigma_t}=T_{\Sigma_{t'}}$，那么每个解都将必然是$V$上的一个常数函数。

映射T_{Σ_t}：$\mathcal{S}\to\mathcal{I}$在我们的哈密顿量$H$：$\mathcal{S}\to\mathbb{R}$和$h$：$\mathcal{I}\to\mathbb{R}$：$h=H\circ T_{\Sigma_t}^{-1}$之间建立了一个简单的关系（我们可能已经把这当作是我们对h的定义了）。

Ω和H共同决定$\mathcal{S}$上在解的层次上实现时间平移的流，而ω和h则决定$\mathcal{I}$上实现初始数据时间演化的流。由于任何T_Σ都一方面与Ω和ω相关，另一方面与H和h相关，因此人们可能会希望它也会和与这些流对应的群行为有所纠缠。事实确实如此。当我们通过t单位对解Φ进行时间平移的时候，让我们把这个解写作$t\cdot_{\mathcal{S}}\Phi$，并且让我们把t个单位时间以后初始数据集$(\phi,\ \pi)$的态写为$t\cdot_{\mathcal{I}}(\phi,\ \pi)$，那么我们就会发现$t\cdot_{\mathcal{I}}T_\Sigma(\Phi)=T_\Sigma(t\cdot_{\mathcal{S}}\Phi)$。①

相对于一个片断，V上的每个解对应于初始数据空间中的一条曲线$(\phi(t),\ \pi(t))$，这条曲线有着$(\phi(t),\ \pi(t)):=T_{\Sigma_t}(\Phi)$。这种形式的曲线永远是$\mathcal{I}$上的一个动力学轨迹（即流在$\mathcal{I}$上产生时间演化的一条积分曲线）。反过来，$\mathcal{I}$上的动力学轨迹$(\phi(t),\ \pi(t))$决定唯一解$\Phi:=T_{\Sigma_0}^{-1}(\phi(0),\ \pi(0))$——该解可被视为在片断中的瞬间$\Sigma_t$建立瞬时场构形$\phi(t)$的结果。

5.4 时间和变化

变化表现为个体对象在一个给定时刻有着某种既定特性，而在另一个不同时刻其特性却变得不同而且不相容。在拉格朗日方法中，很容易在表示变化和不表示变化的解之间做出区分，因为不变的解在时间平移群的作用下是不变的。相应地，我们会说当V上的对应解在某个时间平移群下不变的时候，初始数据空间中的一个动力学轨迹代表一个不变的现实。②

---

① 即是说，每个T_Σ对于$\mathbb{R}$作用量$\cdot_{\mathcal{S}}$和$\cdot_{\mathcal{I}}$来说都是等价的。

② 可能有人天真地认为初始数据空间中的动力学轨迹应当被看作是表示了一个不变的实在，仅当它是一个常数——即是说，如果系统被表示为在每一个时间都处于相同的瞬时态的时候。但是这是错的。考虑闵可夫斯基空间中一个良好的理论集，令Φ是一个解，这个解在与惯性系A相联系的时间平移观念下是不变的，但是在与惯性系B对应的情况下并不是不变的。想必这被看作是稳定的——并且应当是这样的，不管我们是通过适应于坐标系A的片断（这会导致一个动力学轨迹，根据它系统的态是一个常量）还是通过适应于坐标系B的片断（这导致的是一种图景，在其中态会经历非平凡的演化）去讨论其哈密顿图景。

这在很大程度上是完全直接的。但值得停下来思考一下，在物理量由$\mathcal{S}$和$\mathcal{I}$上的函数表示的基础上，变化是如何得到表示的；在数学量定义在初始数据空间上时，情况很简单：令$f \in C^{\infty}(\mathcal{I})$对应于瞬时态的某个可确定的物理特性，如果$(\phi_0, \pi_0)$演化为$(\phi_1, \pi_1)$且$f(\phi_0, \pi_0) \neq f(\phi_1, \pi_1)$，那么包括了这些态的解会显示关于$f$所表示的特性的变化。[①] 我们当然可以继续追问，比如说，关于f沿一个动力学轨迹的变化速度。

但是我们应该怎样用定义在解空间上的函数的术语来叙述它？

假设我们感兴趣的是测量那些给定场在其上取非零值的空间区域的体积时的量。虽然这样的量在哈密顿框架中是用函数f: $\mathcal{I} \to \mathbb{R}$ 表示的，但在解空间上没有函数可以与这个量等同——因为这样的函数是对整个物理可能的历史赋值的，因此不能表示在历史的不同瞬间取值不同的量（或者在一定程度上，它们不能用与初始数据空间上的函数相同的方式来做到这一点）。

然而，直观地，对于每个瞬间$\Sigma \subset V$都存在一个函数f_Σ：$\mathcal{S} \to \mathbb{R}$，使得$f_\Sigma(\Phi)$是我们的场在解$\Phi$中支持的$\Sigma$上的体积。因此，如果对于瞬间$\Sigma$、$\Sigma' \subset V$，有$f_\Sigma(\Phi) \neq f_{\Sigma'}(\Phi)$，那么说我们选择的量被表示为解$\Phi$中展现的变化就是一种很诱人的说法。而且为了讨论我们的量的变化速度，我们需要考虑瞬间的参数化族Σ_t，并计算$\frac{d}{dt} f_{\Sigma_t}(\Phi)$。[②]

当然，在当前情况下，在建立这一框架时运用我们首选的片断是合理的。[③] 对于片断中的每个瞬间Σ_t，我们有一个辛同构T_{Σ_t}：$\mathcal{S} \to \mathcal{I}$。如果$f$: $\mathcal{I} \to \mathbb{R}$ 是初始数据空间上表示我们感兴趣的量的函数，那么$f_t := f \circ T_{\Sigma_t}$是解空间上的期望函

---

① 即便Φ表示闵可夫斯基时空中的一个事件的态，由于在与惯性系A相联系的时间平移中不变而成为一个稳定量，它仍然可能表示某些经历变化的物理量——比如一个相对于惯性系B的质量系统的中心的位置。

② 见[Rovelli，1991]。

③ 否则我们就会遇到麻烦。考虑一个定义在闵可夫斯基时空上的Φ，对于每一个惯性观测者，场在其中非零区域的空间容积在时间上是恒定的。因为长度收缩，相对运动的惯性观测者将会为这个容积分配不同的数值。因此，如果我们选择的Σ和Σ'对应于不同的惯性系，那么我们会发现即使对于每一个惯性观测者Φ都是不变的，仍然有$f_\Sigma(\Phi) \neq f_{\Sigma'}(\Phi)$。

数，它在 Φ 于 Σ_t 上引发的初始数据上对解 Φ 赋值 f。因此，每个片断 σ 决定一族 $\mathcal{S}$ 上的单参数函数，它们编码了我们所选的相对于 σ 中的瞬间的物理量的瞬时值。因此，相对于片断的选择，追问这个量是否经历变化以及其变化速度如何等等，是合理的。

注释 31（构建 $\{f_t\}$ 的一种可选择的方法）。在当前设定下，我们宁愿运用 Σ_0 去构造 f_0，然后运用我们的动力学时间平移群去定义：

$$f_{-t}(\Phi) := f_0(t \cdot_{\mathcal{S}} \Phi)$$

而不是依赖整族单参数同构 $\{T_{\Sigma_t}\}$ 去建立单参数函数族 $\{f_t\}$。

5.5 回顾

我们已经看到，如果设置一些非常强大的假设，那么我们会换来一个非常清晰的关于时间和变化的表示的图景。这些假设是：我们的运动方程 Δ 是二阶的；这些方程有理想的存在性和唯一性，它们源于拉格朗日量 L，这个拉格朗日量有动力学对称群 τ，而它又来自于我们时空 V 上的时间平移群 $\bar{\tau}$；我们已经选择了一个符合于 $\bar{\tau}$ 的 V 的片断 σ。

拉格朗日图景。解空间 $(\mathcal{S}, \Omega)$ 是辛空间。对每个解赋予相对于 $\bar{\tau}$ 总的瞬间能量的值的函数 H：$\mathcal{S}\to\mathbb{R}$ 是 τ 的辛生成元（也是与它相关的诺特守恒量）。

哈密顿图景。我们能够构造我们理论的哈密顿版本，即一个初始数据的辛空间 $(\mathcal{I}, \omega)$，具有可以产生理论动力学的哈密顿量 h：$\mathcal{I}\to\mathbb{R}$。在一个 $\mathcal{I}$ 上实现时间演化的 $\mathbb{R}$ 行为中动力学被编码。

图景之间的关系。对于我们的片断中每个对应于辛同构 T_Σ：$\mathcal{S}\to\mathcal{I}$ 的瞬间，它把一个解 Φ 映射到它在 Σ 上引起的初始数据上。每个这样的 T_Σ，一方面把 H 和 h 关联起来，另一方面把 Ω 和 ω 关联起来——并且使得在 $\mathcal{S}$ 上实施时间平移的群的行为和在 $\mathcal{I}$ 上实现时间演化的群的作用之间发生相互纠缠。

时间。在这个方案中它的一个方面被三个 $\mathbb{R}$ 作用所表示：在 V 上通过对称性实现时间平移的作用；在 $\mathcal{S}$ 上实现时间平移的辛作用；在 $\mathcal{I}$ 上实现时间演化的辛作用。注意：在某些时空中，只会存在一个时间平移概念，在其他时空中则会有许多这样的概念。

变化。在拉格朗日图景中，不变性以一种直接的方式得到表示——有些解

在它们基本时空的时间平移群下是不变的。因此，变化可以被描述为缺乏不变性，那么其定义就可以被翻译为哈密顿方法的语言。当谈及通过初始数据空间和解空间上的函数行为来表示给定物理量的变化时，事情变得更有趣一些了。在这里，正是哈密顿图景担保了一种直接的方法：人们发现初始数据空间上的函数对应于我们感兴趣的量，并随着态的演变检验了其行为。就拉格朗日方法而言，事情要更加复杂一些。解空间上没有函数可以直接表示一个可变量。但是通过在两种图景之间采用一个依赖于片断的对应，人们可以在解空间上找到单参数函数族，它们中的每一个都描述沿片断上的不同瞬间的量的值。人们可以用这种单参数族来定义量的变化速度等。

6. 复杂情况

上一节的叙述是以几个很强的假定作为背景的，现在我要来看看，如果取消某个假定，上面发展的图景会是什么结果。我的策略是，不触及我们在前面依次建构哈密顿图像所必需的假设——运动方程是二阶的和时空具有足够的几何结构支持切分——然后看看取消下面假设的结果：(i)解最终都是在整体上确定的；(ii)存在一个跟任何初始数据集合一致的最大解；(iii)拉格朗日量承认一个源自于时空上时间变换群的动力学时间变换群。我在这一节只是考察依次取消(i)—(iii)中的某一个的结果——下一节我会转向广义相对论，那是一个(i)—(iii)都被取消的理论，就像时空具有足够无关解几何来支持切分的假定一样。

简单地说，我们发现有以下内容。

1. 如果取消解最终都是在整体上存在的这个假定，那么时间演化就不再依靠$\mathcal{I}$上的$\mathbb{R}$作用来实现，并且$\mathcal{S}$和$\mathcal{I}$也不再是辛同构的；反而是时间演化由$\mathbb{R}$作用的一种局域化无穷小相应作用来实现，并且$\mathcal{S}$和$\mathcal{I}$都是局域辛同构的。简而言之，在上面发展起来的时间和变化的表征图像里只要求小小的改变。

2. 如果我们通过考虑其拉格朗日量方案和哈密顿量方案展现了规范自由的这类(广泛而重要)理论，来放弃具体化初始数据就足以决定一个唯一解这个假定，即便解最终是局域性地决定的，那么解的空间和初始数据的空间(甚至在

局域上)都是不同构的预辛空间。而且，时间演化也不再由一个单参数群来实现，而是由这些群的一个规范等价类实现。困难也表现在拉格朗日量方案，问题在于这种类型的理论描述了非物理变量的特征，补救办法就是约化——解的约化空间和初始数据的约化空间都是辛的和同构的，很多时间和变化表象的图景在此约化水平都重新出现。

3. 如果我们取消我们的拉格朗日量是时间变换不变性的这一假定，那么我们不得不勉强对付时间依赖的拉格朗日量理论和哈密顿量理论。这里解的空间和初始数据的空间就会是辛空间，也会是同构的。但是我们不再具有跟拉格朗日量方案对称性一样的解的时间变换，也没有像哈密顿量方案上对称性那样的时间演化。不过，我们还能按照通常的方法构造解的空间跟初始数据空间之间同构的一个切分相关的单参数族，进而允许我们重构时间和变化表征的很多熟悉的图景。

6.1 奇异动力学

让我们假定解的整体存在条件对我们的运动方程不成立——存在不能够扩展到定义在全部 V 上解的初始数据集合。然而让我们继续假设我们的理论在以下方面运行良好：我们的时空 V 具有足够结构来支持片断；我们的方程 Δ 是二阶的并且具有唯一性解；以及我们的拉格朗日量 L，承认一个由在 V 上的时间变换群 $\bar{\tau}$ 约化的动力学时间变换群 τ。那么，至少对于物理中出现的各种情况，我们有望能够得到如下发现：

拉格朗日图景。解的空间($\mathcal{S}$, ω)是一个辛流形。动力学时间变换群 τ 按通常的方式作用在$\mathcal{S}$上，即每一个群元按照给定量来时间变换每一个解[①]。τ 是通过 Ω 由哈密顿函数生成的，哈密顿函数 H：$\mathcal{S}\rightarrow\mathbb{R}$ 赋予一个解相应的瞬态能量。

哈密顿图景。我们能够像上面那样构造一个哈密顿图景：给定一个时间变换群 $\bar{\tau}$、一个适当的片断 σ、一个解 Φ 以及一个瞬间 Σ，我们能够构造初始数据，即 Φ 在跟 σ 相关的 Σ 上约化，并且用 σ 把这拖回到我们抽象的瞬间 S。那

① 当然，如果一个解不是所有时间都有定义，那么其定义域就会以明显的方式不同于其时间变换的定义域。

么我们就能够构造初始数据$\mathcal{I}$的空间，具有其正规辛形式 ω，使用我们的拉格朗日量来定义一个哈密顿量h：$\mathcal{I}\to\mathbb{R}$，从而研究最终的动力学。根本创新之处在于，因为某些解具有定义的有限时间范围，人们发现由 h 生成的在$(\mathcal{I}, \omega)$上的矢量场是不完备的——它具有仅仅定义在$\mathbb{R}$的子集上的积分曲线。因此时间演化并不通过$\mathcal{I}$上的$\mathbb{R}$作用来表示，一般来说，说它会演化到下一个任意时刻这样的初始数据空间里的某一点根本没有意义。不过，由哈密顿量产生的矢量场，通常转换成动力学，能够看作局域定义在初始数据空间上$\mathbb{R}$作用的一种无穷小生成元——特别是，如果说数据集合 x 演化成时间的 t 个单位后的数据集合 y 具有意义的话，那么我们发现把一个态映射成时间的 t 个单位后的态，是一个在 x 的充分小领域跟 y 的充分小领域之间的辛(并且保留哈密顿量)映射。

图景之间的关系。跟上面一样，对于在我们的片断里的每个瞬间Σ，我们能够把$T_\Sigma(\Phi)$定义成拖回到解 Φ 在Σ上约化的初始数据的抽象瞬间 S。不过现在每一个 T_Σ 只是部分定义为从S到$\mathcal{I}$的函数，这是由于 $T_\Sigma(\Phi)$的值在 Φ 没有被定义的情况下在Σ时是不确定的。虽然如此，每个这样的 T_Σ 也是一个在S的定义域跟$\mathcal{I}$的定义域之间的辛同构。[①] 通常，我们获得一个为我们所选择的每一个瞬间各自不同的映射。

时间。时间的表示在目前的情境中更加复杂化：就每个时间概念在时空上的变换对应解空间上的一个好对称——并且是一个初始数据上的纯粹无穷小对称。

变化。我们还是能够通过$\mathcal{I}$上函数表示可变性质，并且决定是否给定动力学轨道表现了这些性质的变化，具体通过研究沿着轨道的相应函数行为来进行。尽管解的空间跟初始数据空间之间的整体同构没有了，我们还是发现片断的选择产生了在$\mathcal{I}$跟S之间的一个局域同构单参数族$\{T_{\Sigma_t}\}$，给定一个对应于相关量的初始数据空间上的函数f，族$\{T_{\Sigma_t}\}$就能够用来构造一个对应于所给可变物理量的S上部分确定函数$\{f_t\}$的单参数族。因而在此情况下，变换的表示跟我们拥有解整体性存在的情况是完全一样的。

---

① 因此，直观上讲，解的空间比初始数据的空间更大——我们能够发现初始数据的空间跟解空间的子空间之间的自然同构。

真正的创新之处在于缺少解空间跟初始数据空间之间的一个整体同构。这个现象能够通过经典力学的例子很好地说明。

例子32(开普勒问题)。考虑运动在平面 $x-y$ 里质量为 m 的一个点粒子，容易受到固定在原点的单位质量的一个点粒子的引力影响。① 这里我们的时空 V 是$\mathbb{R}$，而场值的空间 W 就会是运动粒子的可能位置的空间 $Q=\{(x, y)\}$，拉格朗日量是 $L=\frac{1}{2}(\dot{x}^2+\dot{y}^2)+\frac{m}{r}$，其中 $r^2:=x^2+y^2$；对应的哈密顿量是 $H=\frac{1}{2}(\dot{x}^2+\dot{y}^2)-\frac{1}{r}$，为了使 L 和 H 成为明确定义的，我们不得不把物体的位置限制在 $Q:=\mathbb{R}^2/(0, 0)$里的点，我们关心的是 $H<0$ 的情况，这是轨道在空间里有界的情况——因此，具体来说，我们就排除了抛物线运动和双曲线运动。

我们发现存在两种类型的解：(i)规则解，其中粒子具有非零角动量，是周期性的并且为 t 的所有值所定义，它们表示粒子沿着具有作为焦点的原点的椭圆运动。(ii)奇异解，其中粒子具有减少的角动量，是定义在 $t_0<t<t_0+2\varepsilon$ 范围内，它们表示粒子在 t_0(即当 $t\to t_0$ 时从上面 $|r(t)|\to 0$)逐渐逃离原点，沿着曲线按照递减速度飞离原点直到在时间 $t_0+\varepsilon$ 达到一个端点，然后沿着相同曲线落回到原点，当 $t\to t_0+2\varepsilon$ 时从下面回到 $|r(t)|\to 0$。

解的空间是拓扑学病态的，设 $\Phi(t)$是为 $t\in(t_0, t_0+2\varepsilon)$定义的奇异解，设 $\Lambda\subset Q$ 是沿着粒子按照运动的线段，有可能构造一个具有下述特征的正规解系列$\{\Phi_k\}$：每一个 Φ_k 具有跟 Φ 同样的能量，因此每个 Φ_k 表示粒子在周期性地沿着一个椭圆 E_k 按 2ε 周期运动；对于每一个 k，E_k 指向为连接其焦点的线段是包括在 Λ 里的；当 $k\to\infty$ 时，E_k 的离心率达到无穷大，使得 $k\to\infty$ 时 $E_k\to\Lambda$，因此 Φ 是系列$\{\Phi_k\}$的极限。② 但是我们现在考虑解 $\Phi'(t):=\Phi(t+2\varepsilon)$。这是一个奇异解，对于 $t\in(t_0-2\varepsilon, t_0)$有定义，表示在时间 $t_0-2\varepsilon$ 发散的粒子，

---

① 对于开普勒问题的解的空间结构，参见[Woodhouse, 1991, §2.3]和[Marco(马尔科), 1990b]。

② 在解空间上的拓扑学能够构造为，对于每一个 $t\in\mathbb{R}$，取某一给定时刻粒子的位置和速度作为解空间上坐标系，并且构造跟这些坐标系相关的开球集。现在取这些集合的并集为 t 变化的，结果就是我们寻求的拓扑学基的一个基。

沿着 Λ 运动，然后在时间 t_0 被吸收。Φ′等价于 $\{\Phi_k\}$ 的一个极限。确实，我们能够通过乘 2ε 的时间上变换 Φ 产生无穷多个 $\{\Phi_k\}$ 的极限。

由于我们能够找到 S 里的一个序列，它收敛于不止一个极限点，因此 S 不是豪斯道夫的。[①] 但是，显然开普勒问题的初始数据仅仅是 T^*Q——这是豪斯道夫的。因此这些空间肯定不是同构的！

非豪斯道夫流形可能很难驾驭，但是也存在相对驯服的例子，就像下面那样。设 X 是从实线删去原点并在其原来位置添加两个新对象 a 和 b 的结果，一个 X 的子集是一个开球，如果它碰巧是 $\mathbb{R}$ 里的并不包含原点的一个开区间，或者如果它出现在人们采取包含 0 并用 $\{a, b\}$ 之一代替 0 的实数开区间的时候。我们赋予 X 一个拓扑，断言 X 里的开球的任何并集是一个开集。按照我们目前的标准 X 是一个流形，但是它不是豪斯道夫的，因为 a 的任何邻域跟 b 的一个邻域重叠。并且，显然一个像 $\{\frac{1}{k}\}$ 的序列收敛到 a 和 b。

更一般地，我们能够构造一个非豪斯道夫流形 $X_j^{n,m}$，办法是通过取 $\mathbb{R}^n$ 的 m 复制，并且除了给定的过原点($1 < m \leqslant \infty$，$1 \leqslant n < \infty$ 以及 $0 \leqslant j < n$)的 j 维超平面之外，认为它们处处一样。[②] 对应于在平面开普勒问题里的能量任何固定的负值，解空间是由 $X_1^{3,\infty}$ 的复制拼装出来的。

例子 33(n 体问题的奇异性)。对于 $\mathbb{R}^3$ 里的 n 个粒子，可能的粒子位形空间是 $\mathbb{R}^{3n}$。但是这个空间包括碰撞——并且 n 体问题的势能在这些点上是奇异的。因此，就像前面一样，我们设 $\Delta \coloneqq \{q \in \mathbb{R}^{3n}: \mathbf{q}_i = \mathbf{q}_j \text{ 对于某些 } i \neq j\}$，并且设 Q 是 $\mathbb{R}^{3n}/\Delta$，然后取为我们初始数据空间 $T^*Q = \{(q, p) \in T^*\mathbb{R}^{3n}: q \notin \Delta\}$。

我们引出时间 $t = 0$ 时初始数据(q, p)。我们知道这决定了一个最大动力学轨迹 $t \mapsto (q(t), p(t))$，定义在区间$[0, t^*)$，其中 $0 < t^* \leqslant \infty$（对负时间时的相应情况是一样的）。显然，有可能选择(q, p)使得 t^* 是有限的——比如，如果我们认为对于 $n > 1$ 有 $p = 0$，系统就会塌缩并且产生碰撞。如果 $t^* < \infty$，

---

① 回顾一下，如果对于任何 x，$y \in X$，有可能找到具有 $x \in U$ 和 $y \in V$ 不相交开集 U，$V \subset X$，一个拓扑学空间 X 是豪斯道夫空间。一个豪斯道夫空间里的序列最多具有一个极限。

② 前一段的例子是 $X_0^{1,2}$。

让我们称我们的动力学轨迹是奇异的。这就有可能证明，如果 $t^* < \infty$，然后 $t \to t^*$，$q(t) \to \Delta$，当然是在 $\lim_{t\to t^*} \min_{1\leqslant i<j\leqslant n} r_{ij} = 0$ 的意义上。如果存在一个点 $q_0 \in \Delta$，使得 $\lim_{t\to t^*} q(t) = q_0$，我们就说奇异轨迹终止于一个碰撞；相反，我们说它终止于一个伪碰撞。

看看下面的情况。①

$\boldsymbol{n=1}$。这是单个自由粒子的情况，动力学是非奇异的。

$\boldsymbol{n=2}$。这是开普勒问题②，仅有的奇异性是碰撞奇异性。这些情况发生在系统的角动量消失的情况下。

引人关注的是，这些奇异性能够被规则化。③

这在物理上是十分清楚的，人们只要加上发生的任何碰撞都是弹性的这个条件就行。这就允许人们把终止于在 t_0 的一个碰撞的一个解，跟直观上开始于 t_0 具有交换了速度的粒子的解结合起来。继续按照这种方法，人们建构全部时段的连续和分段解析解。因为碰撞解在整个时段现在是无限的，在这一新的意义上，解的空间同构于包括处于 $T^*\Delta$ 的碰撞状态的初始数据的扩展空间（让我们把这些态解释为表征粒子下一步发射时具有的速度）。

数学上，存在大量能够完成这个过程的方法。④ 一个比较老的解析延拓方法——把原碰撞解理解为一个复数，人们会问是否存在任何这个函数在碰撞之后的解析拓展。在一个更现代的方法中，人们寻求一种连续奇异解的方法，保

---

① 对 n 体问题的概括，参见[Diacu（迪阿库），1992；2002]。就一般处理，见[Diacu and Holmes（迪阿库和霍姆斯），1996，Chapter 3]。

② 从二体问题开始，密切关注粒子运动平面；选择一个参照系，二体系统的质心静止在原点，并且二体的位置分别表示为 $\vec{q}_1$ 和 $\vec{q}_2$。明显地，如果我们知道 $\vec{r} = \vec{q}_2 - \vec{q}_1$，那么我们就知道两个粒子的位置（因为我们知道它们的质量及其质心的所在）。现在注意到对于 $\vec{r}$ 的运动方程是对于在原点周围的一个引力势里移动的单个粒子，如果我们认为原点具有单位质量，那么运动粒子具有质量 $m = \dfrac{m_1 m_2}{m_1 + m_2}$。

③ 对于二体碰撞规则化的各种方法，见[Souriau，1982]、[Marco，1990a]和[Cushman and Bates（库什曼和贝茨），1997，§II.3]。

④ 对此参见[McGehee（麦吉），1975]。

持对于初始数据上演化的连续依赖。① 在此二体问题的情况下，每种方法都能维护奇异解的拓展，通过弹性碰撞是这种拓展的唯一可行方法。

$n=3$。碰撞里的奇异轨迹，只涉及两体的碰撞能够被正规化为弹性碰撞；但是某些三体碰撞就是非正规化的（按照几个标准的任何一个）。② 这些三体碰撞是复杂的，估计很难决定解空间的拓扑学——因此跟 $n=2$ 的情况不一样，在这种情况下，人们根本没有一个解空间整体结构跟初始数据空间结构之间关系的清晰图像。

$n\geqslant 4$。对于 $n\geqslant 4$，跟通常情况一样，奇异轨迹能够终止于碰撞，二体碰撞是可以规则化的，但是无论如何有些牵涉大量粒子的碰撞是不可规则化的。而且，对于 $n>4$，已经知道伪碰撞也可能发生，因此要确定解空间拓扑学就比之前显得更加困难。③

注释 34（奇异系统的量子化）。当解的空间和初始数据是同构的，人们选择哪个空间作为量子化的起点当然无关紧要。当动力学是奇异的，并且这些空间不再同构，人们就会面临真正的选择，而这个选择差强人意——人们不得不在初始数据空间跟解空间之间选择，前者是动力学由一个不完备的矢量场得以补充的基础，而关于后者人们料想具有一个复杂的病态拓扑学。可能无法保证这两种方法在奇异动力学里总可以导致同样的量子化。④

---

① 更准确地，人们从 $T^*\mathbb{R}^{6n}$ 挖去一个具有包含碰撞的紧致闭包开集，这将集合的边界分成两片，对应于进入集合的轨迹的初始数据和离开集合的轨迹的产生数据，演化给出一个微分同胚，从前者对应于非奇异解的子集，到后者对应于非奇异解的子集，人们会问是否这能够被拓展到从一个整体片断到其余整体片断的微分同胚。

② 参见[McGehee，1975]。

③ 参见[Saari and Xia（萨里和夏），1995]。问题对于 $n=4$ 还没有解决；不过有一个可能的例子可以参见[Gerver（杰弗），2003]。伪碰撞要求至少有些粒子的位置在 $t\to t^*$ 时变成没有限制的——通过随意利用势能转化为动能的伟大守恒，这些粒子在有限时间逃到无限远。就像厄尔曼（Earman）在[1986，Chapter III]和本书所强调的，这意味着伪碰撞涉及决定论的彻底瓦解——最值得注意的是人们考虑这个过程的时间翻转时，在其中空间任何地方都没有的粒子从无穷远处突然涌现。

④ 对于奇异动力学系统的量子化方法，参见[Gotay and Demaret（戈塔伊和德玛雷特），1983]和[Landsman，1998]。

6.2 规范自由

我们接下来要考虑一旦我们放弃如下假设会发生什么，即把初始数据具体化到足以决定对我们运动方程的唯一最大解的假设。为了这个目的，我们要假设我们的运动方程不完全决定场的行为，而且在极端的下述意义上，即对所给的初始数据来说，一般解跟至少包含独立变量的完全集的一个任意函数的数据一致。① 存在一大类物理理论，其方程展示了这个表面上的病态情形——最重要的是，包括了电磁学的麦克斯韦(Maxwell)理论、广义相对论以及它们的推广。

在此小节里，我们首先概述这些理论的拉格朗日量处理的一些理论，而不作任何特殊的时间变换不变性、解的整体性存在或者时空结构的假设。这些进一步的假设在后面将会起作用，并且还会保证一个理论的哈密顿形式的想法，这个理论表现良好，除了具有一个不完全决定的动力学以及在这些理论里时间和变化的表征之外。这个讨论后面会给出三个例子。

让我们从引入拉格朗日量理论的规范对称族的概念开始，回顾作用在运动学上可能场的空间$\mathcal{K}$上的群 G，是一个定义在$\mathcal{K}$上拉格朗日量 L 的变分对称的群，如果 G 的作用是适当局域的，并且保留 L 不变量的变分问题。② 我们称($\mathcal{K}$, L)的变分对称的群 G 为一个规范对称的群，如果它能够以一个自然方式通过时空上一个任意函数族参数化。③ 简单说，每一个时空上函数产生($\mathcal{K}$,

① 关于不完全决定的运动方程的相关概念，参见[Olver，1993，pp. 170—172，175，342—346，377]。

② 一个更加明确的定义，见[Zuckerman，1987，274]。

③ 让我们更明确些。首先，设 Y 是一个矢量空间，并且设 Γ 是从 V 到 Y 的函数的空间(更一般地，设 Γ 是一些矢量从 $E\to V$ 的截面的空间)。我们假设 Γ 包括所有从 V 到 W 的光滑、紧致支撑映射，但是弄不清楚确切边界条件、光滑条件，等等，这些是为描述 Γ 所要求的(特别注意相关边界条件在 Γ 包含具有非紧致支集的元素时是必需的)。现在我们定义 Γ 参数化的规范对称的一个群为一对线性局域映射，$\varepsilon\mapsto X_\varepsilon$ 和 $\varepsilon\mapsto R_\varepsilon$，把 Γ 元素 ε 分别映射成$\mathcal{S}$上的局域矢量场和对应的 $\Omega^{n-1,0}_{loc}(V\times\mathcal{K})$ 的元素，使得对于所有 $\Phi\in\mathcal{S}$和 $\varepsilon\in\Gamma$ 都有 $\partial L(\Phi, X_\varepsilon(\Phi))=DR_\varepsilon(\Phi)$，因此每一个 $\varepsilon\in\Gamma$ 都跟 L 的诺特群的无穷小生成元结合在一起。

L)的对称的一个诺特群——一个 L 的变分问题的(适当局域的)对称的单参数群。① 由于时空上函数的集合是无穷维的，一个理论的局域对称的任何群是无穷维的。

大多数熟悉的物理学理论对称群——具有非平凡几何时空的等距群、通过由在每一个时空点上同样的因素改变单粒子波函数来作用的群，等等——都是有限维的，因此并不能算作是目前意义上的规范对称群。

容易理解，承认这样一个规范对称群的拉格朗日量理论的运动方程不完全决定该理论的解。设 ε 是一个时空上函数，除了在某些紧致集合 $U \subset V$ 之外任何地方都没有，如果我们允许相应的诺特群 $\xi = \{g_t\}$ 作用在一个解 Φ 上，那么对于 $t \neq 0$ 最终解 $\Phi_t = g_t \cdot \Phi$ 就会符合 U 之外的 Φ，但是一般不符合 U 内的 Φ。因此，如果我们选择一个跟 U 不交叉的瞬时 $\Sigma \subset V$，我们发现 Φ 和 Φ_t 约化 Σ 上同样的初始数据，但是在整体上不一样——因此唯一性对理论的运动方程不成立。

回顾3.3节，一个预辛形式是一个简并闭合2形式，并且这样一个形式加在一个空间上，通过称为规范轨道的子流形来划分空间。一个非常类似于上一段的论证证明，如果 L 承认规范对称群，那么 L 在解空间上约化的形式 Ω 是预辛形式的，并且相应的规范轨道使得两个解属于同样的规范轨道，当且仅当它们是通过 L 的规范对称群的元素关联起来的。② 因此 L 的规范对称是($\mathcal{S}$, Ω)的规范变换，在前面3.3节里规定的意义上它们保持了解空间的规范轨道。

根据关于预辛形式的一般事实可以得到，如果一个 $\mathcal{S}$ 上的函数产生规范对称的单参数群，那么那个函数是一个常数函数。特别是，诺特守恒量 Q_ξ：$\mathcal{S} \to$

① 更明确点，前面注释的讨论表明，映射 $\varepsilon \mapsto (X_\varepsilon, R_\varepsilon)$ 是一个从 Γ 到($\mathcal{K}$, L)的诺特群的生成元集合的一个映射；事实上，这个群的像在非平凡例子中就会是无穷维的，虽然它可能具有一个非平凡核心。在下面例子38里，不变函数产生相同的(平凡)诺特群。

② 回顾前面的函数61，Ω 被定义为在一任意瞬间 $\Sigma \subset V$ 上某个对象的整体。如果我们考虑一个无穷小局域对称 X_ε，它对沿着 Σ 的解没有影响，那么 Ω 就看不见 X_ε——即 $X_\varepsilon(\Phi)$ 将会是在每一个 $\Phi \in \mathcal{S}$ 上的一个零矢量。参见[Deligne and Freed，1999，§2.5]和[Woodhouse，1991，145]。

$\mathbb{R}$，结合一个规范对称 ξ 的单参数群，必定是常数函数。[1] 在这些守恒量不提供任何方式区分物理上不同的解的意义上，这样的守恒量都是平凡的。

我们能够对容许规范对称群的任何拉格朗日量理论说的就这么多。现在让我们集中看看下面的情况，我们的运动方程 Δ 是二阶的，我们的空间 V 具有足够的几何结构来承认切片，解在时间中是整体存在的，以及我们的拉格朗日量 L 承认动力学时间变换群 τ。借助这些适当的进一步假设，就其在前面第 5 节里发展起来的时间和变化图像而言，我们能够考察放弃解的局域唯一性的含义。我们有如下发现。

拉格朗日量图像。我们假设我们有超出我们背景时空 V 的结构之外的时间变换概念。按照通常方式，这个概念产生一个动力学时间变换群 τ，相应的守恒量是通常的哈密顿量 H，它赋予每一个解瞬时能量。到此一切都好，但是现在回顾前面 3.3 节的讨论，在预辛空间的设定里，如果一个函数产生一个给定的空间变换族的群，那么它也会产生等价于所给那个群的变换规范的全部单参数族。在当前情况下，这意味着，除了动力学时间变换群 τ，H 产生 (S, Ω) 变换的全部单参数群，它跟具有 τ 的规范一致。

哈密顿量图像。固定时空中时间变换的概念和成瞬间的相关时空切片后，我们就能够像通常那样进一步建构初始数据空间，它们出现在场的位形和动量变量被限制在我们的切片里的一个任意瞬间的时候。[2] 跟在情况良好的情况下一样，给定我们切片的一个瞬时 $\Sigma \subset V$ 和一个解 Φ，我们能够在我们抽象的瞬时 S 上建构一个相应的初始数据集合 (ϕ, π)，靠撤回到 S 在 Σ 上 Φ 引入的初始数据。在情况良好时，我们发现初始数据具有结构 $T^{*}Q$，其中 Q 是全部 ϕ 的空间，通过限制解到常数作为瞬时场位形而出现。在当前情况下，我们发现作为初始数据集合而出现的 (ϕ, π) 形成 $T^{*}Q$ 的一个子集（其中 Q 又是作为限制解

① 事实上，它将会是零函数，因为诺特流 J_ξ 将会精确为 V 上 $(n-1)$ 形式，参见 [Zuckerman, 1987, 274]。

② 关于建构对应于一个承认规范对称的给定拉格朗日量理论的约束哈密顿系统，参见 [Dirac, 2001]、[Gotay *et al.*, 1978] 以及 [Henneaux and Teitelboim, 1992]。对于哲学讨论，见 [Earman, 2003] 和 [Wallace（华莱士），2003]。

到瞬间而出现的所有 ϕ 的空间)。[①]另外，我们也可以发现，为了建构一致的动力学，我们需要进一步限制可能的初始数据。结果是取我们初始数据的空间为一个子空间$\mathcal{I}\subset T^{*}Q$。$\mathcal{I}$装配有一个自然几何结构；切从 $T^{*}Q$装配其正则辛形式(参见上面的例子 7)；对$\mathcal{I}$的这个形式的限制产生一个预辛形式 ω。一旦一切正常，由 ω 决定的规范轨道具有下面的结构：作为初始数据出现的初始数据集合(ϕ, π)和(ϕ', π')，是通过解 Φ 和 Φ'在一个所给定的瞬间$\Sigma\subset V$上引入的，都属于$(\mathcal{I}, \omega)$里同样的规范轨道，当且仅当 Φ 和 Φ'属于$(\mathcal{S}, \Omega)$里同样的轨道。[②]人们能够继续以通常方式定义一个在$(\mathcal{I}, \omega)$上的哈密顿量函数 h。当然，由于$(\mathcal{I}, \omega)$是一个纯粹预辛空间，h 产生动力学概念的整个规范等价类，即$(\mathcal{I},\omega)$对称的单参数群。假设按照一个这样的动力学概念，初始态 x_0 演化为时间 t 的态 $x(t)$，那么由 h 产生的其他时间演化概念一般也会跟 x_0 在 t 演化成的不一致，它们就会完全跟在 t 处于$[x(t)]$的态一致，即 $x(t)$的规范轨道。[③]

图像之间的关系。跟通常一样，对于我们的切片里的每一个常数Σ，我们能够定义 $T_{\Sigma}:\mathcal{S}\to\mathcal{I}$，把一个解映射到(拖回$\mathcal{S}$到)它在$\Sigma$上约化的初始数据。在第 5 节的背景里，这些映射赋予我们的解空间跟初始数据空间之间的同构。但是在规范对称的情况下，这些映射不是同构的，因为规范对称的存在意味着很多解约化了在任何给定Σ上同样的初始数据。一旦我们考虑一个只有有限多自由度且承认规范对称的理论时，这个情况非常奇怪，因为解空间会是无穷维的，而初始数据的空间会是有限维的(参见下面的例子 37)。[④]如果一切顺利的话，我们就得到解跟初始数据空间里动力学轨迹之间关系的下述图像(保持固有的时间变换概念和适合它的片断)。

1. 设 Φ 是一个解，　并且设 $x(t)=(\phi(t), \pi(t))$是$\mathcal{I}$里的曲线，它的出现

---

① 这是因为所谓第一类约束的产生：它根据动量的定义，$p_i:=\frac{\partial L}{\partial \dot{q}_i}$，得到某些动量的组元要求恒等于零。

② 警告：不难建构(非物理的)例子，其中没有好的图景，参见下面的例子 36。

③ 事实上，对于每一个点 $y\in[x(t)]$，存在一个 h 产生的时间演化的概念，其根据的是 x_0 在时间 t 演化成 y。

④ 因此在此情况下我们看到不存在从$(\mathcal{S}, \Omega)$到$(\mathcal{I}, \omega)$的映射是一个同构，直观上这对于任何承认局域对称性的理论都是对的。

是通过设定 $x(t)$ 为在瞬间 $\Sigma_t \subset V$ 上 Φ 引入的初始数据集合。那么 $x(t)$ 是理论的一个哈密顿量方案的动力学轨迹。

2. 给定理论的哈密顿量方案的一个动力学轨迹 $x(t)$，我们发现存在一个唯一解 $\Phi \in \mathcal{S}$，使得对应于 Φ 的$\mathcal{I}$里的曲线，在上述条款的意义上，正好是 $x(t)$。

3. 如果 Φ，$\Phi' \in \mathcal{S}$属于在解空间里同样的规范轨道，那么对应的良性轨道，$\mathcal{I}$里的 $x(t)$ 和 $x'(t)$ 跟规范一致。对于每一个 t 而言，$x(t)$ 和 $x'(t)$ 属于$\mathcal{I}$里的同样规范轨道意义上。

4. 如果$\mathcal{I}$里的动力学轨迹 $x(t)$ 和 $x'(t)$ 跟规范一致，那么它们对应的解 Φ，$\Phi' \in \mathcal{S}$属于$\mathcal{S}$里同样规范轨道。

时间。我们的时间变换概念，以一个良好的方式从我们的时空 V 提升到解的空间$\mathcal{S}$，在那里我们通过一个$\mathbb{R}$ 作用获得通常的时间表征。即便如此还是有点古怪，产生这个作用的哈密顿量，也产生每一个规范等价于它的$\mathbb{R}$ 作用。这个情况在初始数据空间里更糟糕：给定一个时空上时间变换的概念和适合那个概念的切片，我们能够建构一个哈密顿量的图像，但是在存在规范对称的情况下，我们发现存在很多通过初始数据空间里每一点的动力学轨迹。结果，我们在时空里的单个时间变换概念分成大量的初始数据空间上$\mathbb{R}$ 作用，每一个都同样声称实现了理论的动力学。

变化。在定义在初始数据空间上的动力学之下的任意量的演化是非决定性的：如果 $x_0=(\phi_0, \pi_0)$ 是一个初始数据集合，就会存在一个通过在 $t=0$ 的 x_0 点的不同动力学轨迹 $x(t)$ 和 $x'(t)$，对于任意函数f：$\mathcal{I}\to\mathbf{R}$，我们没有理由期望对 $t\neq0$ 有 $f(x(t))=f(x'(t))$，因此固定在时间 $t=0$ 的态并不能够决定由 f 表示的量的过去和未来值。但是由于在此情形下 $x(t)$ 和 $x'(t)$ 会跟处于每一个时间的态的那个规范轨道一致，我们发现规范不变量的演化(即那些由沿着规范轨道是常数的初始数据空间上的函数所表示的)是完全决定性的——给定初始态，人们就能够全时段预测这样一个量。而且，我们的切片允许我们结合每一个在初始数据空间上的规范不变性函数 f 跟一个解空间上的规范不变性函数的单参数族$\{f_t\}$：设 $f_t(\Phi)$ 是 f 在 Φ 约化于 Σ_t 上的初始数据集合上所取的值。按照这种方式，我们能够通过初始数据空间上函数，或者常规方法里解空间上的

函数，来表示规范不变量的变化。

尤其最后一点应该引起我们的疑惑，我们的理论在目前所考虑的形式下会包含多余的结构，因为理论具有一些很令人失望的特征——病态的初始值问题、平凡守恒定律、纯粹预辛几何结构，甚至在解空间跟初始数据空间之间偶局域同构的失败——人们发现存在一个巨大的物理量的子集，其行为就像表现良好的理论的量那样。人们自然会想是不是存在一个表现良好的理论，统治隐藏在背景中某些地方的这些量的行为。

这种疑惑推动前面 3.3 节里讨论的约化过程对($\mathcal{S}$, Ω)和($\mathcal{I}$, ω)的应用。当一切正常时，下面的图像就会出现：解的约化空间(即解空间的规范轨道的空间)和初始数据的约化空间(即初始数据空间的规范轨道)两者都是辛空间，并且这些约化空间都是同构的。① 对应于解原始空间上的时间变换的哈密顿量函数和在初始数据原始空间上的时间演化，都投影到约化空间。最终约化的哈密顿量产生时间变换并在它们各自空间上的时间演化。

出现在物理学中的典型案例里，人们看到原始理论在规范对称群下的不变性，事实上是一个物理上不必要的变量被包括在理论里的标志。的确，初始数据的原始空间预辛于一个辛约化空间的事实，表明在理论的原始哈密顿量形式体系里，人们能够把初始数据的原始空间参数化变量的集合分成两类，我们称其为物理上相关的变量类和物理上不必要的变量类。具体化所有变量的初始值足以决定物理上相关变量的整个时段的值，而物理上不必要的变量的演化完全留有任意性。② 至少局域上物理相关变量能够被用来参数化初始数据的约化空间。这就提供了一个很强的理由认为定义在初始数据约化空间的哈密顿量理论，给出了所考察的物理学的一个清楚表述，在其精确涉及由原始理论所决定其演化的那些量的时候。而这反过来提供了好的理由把解的约化空间当作表示

① 下面的例子 36 是一个(非物理的)情况，在那里这个同构不成立。

② 在满足适当技术条件的一个预辛流形里，任何点具有一个承认其坐标分成两类的坐标图的领域——参数化规范轨道的那些和参数化跟规范轨道相切方向的那些，参见[Abraham and Marsden, 1978, Theorem 5.1.3]。在初始数据的空间里，自然取第一类变量为物理上不必要的，而第二类为物理上相关的。

系统的可能历史，该系统的可能瞬时态由初始数据空间里的点表示。①

注释35(约化和决定论)。假设一个通过看上去非决定性的理论陈述的东西，其中态的很多未来序列跟给定初始态是一致的。那么人们总可以通过简单等同所有特征跟给定态一致来建构一个决定性理论。正如莫德林(Maudlin)已经注意到的，人们遇到一个非决定性理论的任何时候，运用这种策略都可能是愚蠢的：(i)这个策略的一般运用，会因命令而放弃决定论的真；(ii)人们可能常常最终接受平凡而无知的理论。② 例如：在牛顿物理学里，在其中空间里完全没有粒子的初始态跟空间仍然为空的未来是一致的，并且也跟其中粒子从无穷远处涌来的未来是一致的，那么对在全部未来时间里有引力上的相互作用。要想分清这些未来——视其为仅仅是单独物理可能性的重新描述——可能是荒谬的。

现在，约化是莫德林反对的一般策略的特殊情况。但是由于任何明智的作用方针都是一般无知的策略的特殊情况，因此这本身不是对约化的反对。我们应该检验是否是莫德林明确反对总策略的抱怨强化了对特殊情况的不信任。我断言情况并非如此，因为：(i)约化具有规范对称的理论，看上去把非决定性的理论转变成决定性的理论，但是这是无法反驳的，作为规范对称理论看上去的特征，这种非决定论(即其演化整体上不受系统初始态约束这种量的存在)显然是非物理的；(ii)对这种出现在物理学里的理论，人们没必要害怕约化会导致一个无关紧要或者荒谬的结果——在已经知道的情况下，约化把人们带入一个情况好的辛空间，它是对物理理论的一个适当设置。确实，在此情况下，它(几乎毫无异议地)同意最终辛空间参数化真正的自由度，并且提供了原始理论动力学的正确设置。③

---

① 由于这个空间是通过理论的规范对称群元联系起来的解而出现的，而初始数据的约化空间是靠理论的规范对称群元的联系起来的解，在给定常数上约化的初始数据而出现的。

② 参见[Maudlin，2002，pp. 6—8]。

③ 广义相对论提供了唯一的例子，在那里存在对一致观点的少许异议，参见[Kuchař(库查尔)，1986]和[Kuchař，1993]。这也是莫德林关心的情况——他跟库查尔一样，担心没有反思就把约化用到广义相对论，会导致关于时间的荒谬结论，并且阻碍了量子引力里的概念进步。后面第7节的主要任务就是要证明约化在那个情况下的运用没有导致错误。

例子 36(一个病态例子)。在开始之前，重要的是要强调一下，不难设想不遵从前面为规范对称理论勾画出来的模式的简单(但非物理的)例子。[①] 考虑一个在 $x-y$ 平面运动的粒子，其拉格朗日量 $L=\frac{1}{2}e^{y}\dot{x}^{2}$。相应的欧拉—拉格朗日方程告诉我们 x 是关于时间的常数，而 y 整体上是任意的。因此解空间由 $(x_0, y(t))$ 对构成，其中 $x_0\in\mathbb{R}$ 并且 $y(t)$：$\mathbb{R}\to\mathbb{R}$ 是一个任意光滑函数；两个解 $(x_0, y(t))$ 和 $(x'_0, y'(t))$，当且仅当 $x_0=x'_0$ 时属于同样的规范轨道。因此解的约化空间刚好是 $\mathbb{R}$ ——具有奇数维，不能承担一个辛结构。就哈密顿量看来，人们发现共轭于 x 动量和共轭于 y 动量都必须为零——这意味着初始数据空间是 $\mathbb{R}^2=\{(x, y)\}$，任何点都规范等价于任何其他点。[②]因此初始数据的约化空间是一单独点，这不同构于解的约化空间。

例子 37(线上的粒子)。我们考虑两个在一条线上运动的引力点粒子，简单起见，我们选择单位使得牛顿常数为 1，假设粒子具有单位质量，并且不管碰撞及其规则化。我们考虑关于这个系统的三个理论。

牛顿理论。我们将粒子位置表示为 q_1 和 q_2，且 $q_2>q_1$，我们解释这些都是在给定的相对于绝对空间静止的参照系里。这个系统的拉格朗日量是 $L=T-V$，其中动能是 $T:=\frac{1}{2}(\dot{q}_1^2+\dot{q}_2^2)$，而势能是 $V=-\frac{1}{q_2-q_1}$，遵循通常的牛顿运动方程。考虑这个理论的一个变形是有用的，我们定义新的位形变量，$r_0:=\frac{1}{2}(q_1+q_2)$ 和 $r_1:=q_2-r_0=\frac{1}{2}(q_2-q_1)$(因此 r_0 是系统质心位置，而 r_1 是粒子间相对距离的一半)。按照这些变量，我们的拉格朗日量是$L(r_0, r_1, \dot{r}_0, \dot{r}_1)=\frac{1}{2}(\dot{r}_0^2+\dot{r}_1^2)+\frac{1}{2r_1}$。运动方程告诉我们 r_0 是时间的线性函数(因为孤立的质心按惯性运动)，而$r_1(t)$解决 $\ddot{r}_1=-\frac{1}{2r_1^2}$，从而描述了粒子在引力相互作用时的相对

---

① 对下面的例子，参见[Henneaux and Teitelboim，1992，§1.2.2]，对这个例子的进一步讨论，参见[Gotay，1983]。

② 这是一个例子，其中一个约束直接从动量的定义产生，而要形式化一致的动力学还需要其他约束。

运动。

莱布尼茨理论。在此理论中，空间和运动都是相对的，从而粒子之间的相对距离 r_1 是仅有的位形变量(甚或 r_1 是相对距离的一半)。莱布尼茨理论的拉格朗日量是 $L'(r_1, \dot{r}_1) := \frac{1}{2}\dot{r}_1^2 + \frac{1}{2r_1}$，运动方程是 $\ddot{r}_1 = -\frac{1}{2r_1^2}$。因此莱布尼茨理论给出了跟牛顿理论一样对于粒子之间相对距离的动力学。

半莱布尼茨理论。我们取 r_0 和 r_1 为位形变量，并取拉格朗日量 $L''(r_0, r_1, \dot{r}_0, \dot{r}_1) := \frac{1}{2}\dot{r}_1^2 + \frac{1}{2r_1}$(因此 L''是 r_0、r_1、$\dot{r}_0$ 以及 $\dot{r}_1$ 的函数，恰好仅仅依赖于 r_1 和 $\dot{r}_1$)。我们使用变分算法，跟往常一样，它导致这样的结论，即一个曲线$x(t):=(r_0(t), r_1(t))$是运动方程的一个解，当且仅当$\frac{\partial L''}{\partial r_i} - \frac{d}{dt}\frac{\partial L''}{\partial \dot{r}_i} = 0$ 被 $i=0$ 或 1 的曲线上每一点所满足。对于 $i=1$，我们再次发现 $\ddot{r}_1 = -\frac{1}{2r_1^2}$，因此我们得到跟牛顿情况和莱布尼茨情况一样的相对距离的演化动力学，但是对于 $i=0$，我们关于曲线的条件是空的，因为 L''并不依赖于到底是 r_0 还是 $\dot{r}_0$。因此一条曲线 $x(t) :=(r_0(t)r_1(t))$算作一个对我们运动方程的解，只要 $r_1(t)$描述了牛顿或者莱布尼茨理论允许的运动，而 r_0 是任何(连续的和适当可微的)形式的函数。

让我们来对照一下这三个理论的结构。

对称。牛顿理论的变分群是三维的，由伽利略加速和空间与时间变换构成。莱布尼茨理论的变分对称群由时间变换构成，但是半莱布尼茨理论的变分对称群是无穷维的，除了时空变换，它包括作为规范对称群质心的时间依赖的空间变换。如果 $r_0(t)$和 $r_1(t)$是连续函数，那么 $x(t) = (r_0(t), r_1(t))$是运动学上的可能性。设 $\Lambda(t)$是从$\mathbb{R}$ 到自身的任何其他连续函数，那么 $x'(t) := (r_0(t) + \Lambda(t), r_1(t))$也是运动学上可能的，并且对于全部 t 有 $L''(x(t)) = L''(x'(t))$，因为 L''并不完全顾及 r_0。那就是说，Φ_Λ: $(r_0(t), r_1(t)) \mapsto (r_0(t) + \Lambda(t), r_1(t))$从运动学上的可能性到自身保留了拉格朗日量，因此是变分对称的。的确，对于每一个这样的 Λ，我们得到了一个 L''的不同变分对称。因此连续的 Λ: $\mathbb{R} \to \mathbb{R}$ 的空间参数化了 L''的规范对称群。

规范对称和初始值问题。我们能够探究这些对称以证明半莱布尼茨理论的初始值问题是如何彻底病态的。假设在 $t=0$ 我们固定 r_0、r_1、$\dot{r}_0$ 和 $\dot{r}_1$ 的值，设 $x(t)=(r_0(t), r_1(t))$ 是一个满足这些初始数据的一个解。现在选择 Λ：$\mathbb{R}\to\mathbb{R}$ 使得 $\Lambda(0)=0$ 和 $\dot{\Lambda}(0)=0$，由于 Φ_Λ 是一个拉格朗日对称，$\Phi_\Lambda(x(t))=(r_0(t)+\Lambda(t), r_1(t))$ 也是一个解，当然满足在 $t=0$ 的具体初始数据。按照这种方式，我们建构了对初始数据每一个具体集合解的一个无穷维族。

解空间的结构。牛顿理论和莱布尼茨理论解空间是维度分别为四和二的辛空间，就像我们已经看到的那样，半莱布尼茨理论的解空间是无限维的，并且 L'' 在这个空间上约化的形式是简并的，这个空间不是辛的。相关规范轨道具有下面的结构：$x(t)=(r_0(t), r_1(t))$ 和 $x'(t)=(r_0'(t), r_1'(t))$ 处于同样的规范轨道，当且仅当对于所有的 t 有 $r_1(t)=r_1'(t)$（即解处在同样的规范轨道当且仅当它们跟粒子间相对距离一致——它们关于质心运动的说法是不相关的）。

哈密顿绘景。写出 $p_i=\dot{r}_i$，我们发现我们理论如下的初始数据的空间。

1. 对于牛顿理论，初始数据的空间是 $T^*\mathbb{R}^2=\{(r_0, r_1, p_0, p_1): r_i, p_i\in\mathbb{R}\}$，携带其正则辛结构 $\omega=\sum_{i=0,1}dr_i\wedge dp_i$。哈密顿量是 $H(r_0, r_1, p_0, p_1)=\frac{1}{2}(p_0^2+p_1^2)-\frac{1}{2r_1}$，运动方程是通常的决定论的牛顿方程。

2. 对于莱布尼茨理论，初始数据空间是 $T^*\mathbb{R}=\{(r_1, p_1): r_1, p_1\in\mathbb{R}\}$，携带其正则辛结构 $\omega=dr_1\wedge dp_1$。哈密顿量是 $H'(r_1, p_1)\coloneqq\frac{1}{2}p_1^2-\frac{1}{2r_1}$，运动方程是通常的决定论的莱布尼茨方程。

3. 回顾在建构对应于一个给定拉格朗日量系统的哈密顿系统时，我们必须首先建构对应于拉格朗日量系统的位置变量的动量变量。半莱布尼茨理论具有两个位置变量 r_0 和 r_1，我们的方法告诉我们，对应的动量变量是 $p_i\coloneqq\frac{\partial L''}{\partial\dot{r}_i}$，这里 L'' 是半莱布尼茨拉格朗日量。像往常一样，$p_1\coloneqq\dot{r}_1$，但是因为 L'' 独立于 $\dot{r}_0$，我们发现 $p_0\equiv0$。因此该理论初始数据空间是空间 $\Gamma=\{(r_0, r_1, p_1): r_0, r_1, p_1\in\mathbb{R}\}$，它出现在我们关注 $p_0=0$ 的牛顿理论初始数据空间里的那些态；限定牛顿理论的辛结构到 Γ 产生一个预辛结构（指向 r_0 方向的矢量是零矢量）。规

范轨道具有下面的结构：$x=(r_0, r_1, p_1)$和$x'=(r_0', r_1', p_1')$处于同样的规范轨道，当且仅当$r_1=r_1'$和$p_1=p_1'$时，这个理论的哈密顿量是$H''(r_0, r_1, p_1):=\frac{1}{2}p_1^2-\frac{1}{2r_1}$，在让$r_0$的演化完全不受约束时，它决定了关于$r_1$和$p_1$的牛顿/莱布尼茨式的情况。那就是说，如果$x(t)$和$x'(t)$在对应于该哈密顿量问题解的初始数据空间里的曲线，则人们发现，一般来说对$t\neq 0$有$x(t)\neq x'(t)$，但是对于全部t有$[x(t)]=[x'(t)]$。注意每一个这样的曲线$x(t)$对应于解空间里的一个点，并且对于全部t条件，$[x(t)]=[x'(t)]$只是说明对应于曲线x和x'的解空间里的点，两曲线自身处在同样的规范轨道。

约化。正如我们所期望的，半莱布尼茨的约化初始数据空间同构于莱布尼茨理论的初始数据空间，并且半莱布尼茨理论的约化解空间同构于莱布尼茨的解空间——在两种情况下，这是因为等同相关规范轨道里的点意味着把r_0降为动力学变量。因此约化实现了我们的物理直觉，即r_0是应该被删去的一个额外变量，并且排除了半莱布尼茨理论的病态。而且，初始数据的约化空间（约化的解空间）从原始理论继承了一个真正是莱布尼茨理论的那种哈密顿量（拉格朗日量），因此这些约化空间支持了具有正确动力学和对称群的动力学理论。

当然，这是一个理想化例子——可能是最简单的一个。并且它在此已经建立起来，使得它从开始就清楚半莱布尼茨理论的变量能够被分成物理上相关的r_1（对拉格朗日量起作用，并且它的动力学是决定论的），和物理上不必要的r_0（对拉格朗日量不起作用，并且其演化完全不受动力学所约束）。因此从开始就很清楚r_0应该从理论里删去——已经没有必要保留它，也没有必要总结我们手上的有关非决定论的理论。

但是注意，如果我们保留我们原始的牛顿变量q_1和q_2（其中$q_1<q_2$），并且已写成$L'':=\frac{1}{2}(\dot{q}_2-\dot{q}_1)^2-\frac{1}{2(q_2-q_1)}$，那么事情就不可能如此清楚。$q_1$和$q_2$的演化方程可能会跟物理上相关的信息和不必要的信息混合在一起，而它对看清发生了什么起不到什么作用。

一旦我们面对承认局域对称群的拉格朗日量理论，我们知道（除非它们展示了我们在上面例子36看到的病态情况）存在某些方式把变量分离成物理上相

关的和不必要的(这在哈密顿量情况中最容易看到)。但是并不是总能发现这样的分离。这是为什么我们放弃这样的理论转向它们背后更有吸引力的约化理论的几个原因之一。

例子 38(麦克斯韦理论)。我们考虑电磁场，设 V 是闵可夫斯基时空，并且固定一个惯性参照系和一组相关的坐标系(x_0，x_1，x_2，x_3)，我们选择为我们场 $W=\mathbb{R}^4$ 目标空间，因此运动学上可能场是形式 A：$V\to\mathbb{R}^4$ 的(遵循一些不具体的可微性和边界条件)，A 是通常的四维势。

我们定义 $F_{uv}:=\frac{\partial A_v}{\partial x_\mu}-\frac{\partial A_\mu}{\partial x_v}(u,\ v=0,\ \cdots,\ 3)$，因此一个运动学上可能的场 $A(x)$ 决定一个矩阵值的函数 F。我们按照下面的方案来标记构成 F 的分函数，因此等同 F 的组分为电场分量和磁场分量，$\mathbf{E}(x)=(E^1(x),\ E^2(x),\ E^3(x))$ 和 $\mathbf{B}(x)=(B^1(x),\ B^2(x),\ B^3(x))$：

$$F_{\mu\nu}(x)=\begin{vmatrix}0 & -E^1(x) & -E^2(x) & -E^3(x)\\ E^1(x) & 0 & B^3(x) & -B^2(x)\\ E^2(x) & -B^3(x) & 0 & B^1(x)\\ E^3(x) & B^2(x) & -B^1(x) & 0\end{vmatrix}$$

我们取理论的拉格朗日量 $L:=-\frac{1}{2}(|\mathbf{B}(x)|^2+|\mathbf{E}(x)|^2)$，写出 $A(x)=(A_0(x),\ A_1(x),\ A_2(x),\ A_3(x))$ 和 $\mathbf{A}(x):=(A_1(x),\ A_2(x),\ A_3(x))$，我们发现拉格朗日量的运动方程为：

$$\nabla^2 A_0+\frac{\partial}{\partial x_0}(\nabla\cdot\mathbf{A})=0$$

$$\nabla^2\mathbf{A}-\frac{\partial^2\mathbf{A}}{\partial x_0^2}=0$$

其中 $\nabla:=(\frac{\partial}{\partial x_1},\ \frac{\partial}{\partial x_2},\ \frac{\partial}{\partial x_3})$ 是通常的三维梯度算符。这些方程等价于通常的电场和磁场真空麦克斯韦方程：$\dot{\mathbf{B}}=-\nabla\times\mathbf{E}$、$\nabla\cdot\mathbf{B}=0$、$\dot{\mathbf{E}}=\nabla\times\mathbf{B}$，以及 $\nabla\cdot\mathbf{E}=0$。

设 Λ：$V\to\mathbb{R}$ 是一个连续函数(适当可微并且满足适当边界条件)。那么映射 Φ_Λ：$A\mapsto A':=A+d\Lambda$ 是一个从运动学上可能的场空间到自身的映射。如果人们计算对应于 A 和 A' 的矩阵 F' 和 F，人们发现 $F'=F$。因此 $\mathbf{E}$ 和 $\mathbf{B}$ 在我们的

规范变换 $A \mapsto A'$下是不变量。因此 $L(\Phi_\Lambda(A)) - L(A) = 0$，所以 Φ_Λ 是拉格朗日量对称的——特别是，A'是一个解当且仅当 A 也是一个解。由于 Λ 是任意的，并且由于只要 $d\Lambda \neq d\Lambda'$就有 Λ 和 Λ'导致不同的对称，我们实际上已经发现一个巨大的我们理论对称性的族。的确，Φ_Λ 形成了在前面引入的官方意义上的我们理论的一个规范对称群。

当然，可见 A 的初始值问题是病态的：设 $A(x)$是一个在常数 $x_0 = 0$ 上提出的初始数据解，并且设 Λ 是一个消失在超曲面 $x_0 = 0$ 的邻域里的非常数式函数，那么 A 和 $A' = A + d\Lambda$ 是跟 $x_0 = 0$ 一致而非整体一致的解。

当然，还有就是在解空间上引入的拉格朗日量的形式是简并的。对应的规范轨道具有如下形式：解 A 和 A'属于同样的规范轨道，当且仅当存在一个 Λ：$V \to \mathbb{R}$ 使得 $A' = A + d\Lambda$。一个等价条件是 A 和 A'处于同样的规范轨道，当且仅当它们导致同样的 $\mathbf{E}$ 和 $\mathbf{B}$——这等于说解的约化空间是 $\mathbf{E}$ 和 $\mathbf{B}$ 的场方程解空间(注意，我们是处于一个固定的坐标系，因此这些都是明确定义的)。这个约化空间是一个辛流形。

我们能够建构对应于我们拉格朗日量理论的哈密顿量理论(我们所选的惯性坐标系给了我们一个片断)。为了方便起见，我们取我们拉格朗日量理论的位形变量为 $A_0(x)$和 $\mathbf{A}(x)$，设Q是可能的空间(A_0，$\mathbf{A}$)，并且 $T^* Q$是对应的余切丛，支撑其正则辛结构。在 $T^* Q$里一点由空间上的场的四元组($A_0(x)$，$\mathbf{A}(x)$，$\pi_0(x)$，$\boldsymbol{\pi}(x)$)构成，其中 A_0 和 π_0 在$\mathbb{R}$ 里取值，而 $\mathbf{A}$ 和 $\boldsymbol{\pi}$ 在$\mathbb{R}^3$ 里取值。我们通常的做法告诉我们，对应于 A_0 的动量 π_0 等同于零(我们的拉格朗日量并不依赖于 $\dot{A}_0$)；对应于 $\mathbf{A}$ 的动量 $\boldsymbol{\pi}$ 是 $\boldsymbol{\pi}(x) = -\mathbf{E}(x)$。因此，我们理论初始数据空间 Γ 是形式为($A_0(x)$，$\mathbf{A}(x)$，0，$\boldsymbol{\pi}(x)$)的点的子空间 $T^* Q$——所以我们能够在 Γ 里取点成形式为(A_0，$\mathbf{A}$，$\boldsymbol{\pi}$)的三元组。Γ 从其在 $T^* Q$的嵌入继承来的预辛形式产生下面形式的规范轨道：(A_0，$\mathbf{A}$，$\boldsymbol{\pi}$)和(A'_0，$\mathbf{A}'$，$\boldsymbol{\pi}'$)属于同样的规范轨道，当且仅当 $\boldsymbol{\pi} = \boldsymbol{\pi}'$和$\nabla \times \mathbf{A} = \nabla \times \mathbf{A}'$。由于 $\boldsymbol{\pi} = -\mathbf{E}$ 和 $\mathbf{B} = \nabla \times \mathbf{A}$，这告诉我们初始数据空间里的两个点处于同样规范轨道，当且仅当它们对应于同样的电场和磁场时。因此初始数据约化空间正好是电场和磁场的瞬时态空间。我们再次发现初始数据的约化空间是辛地同构于解的约化空间。

在目前的情况下，正如上面的半莱布尼茨例子一样，我们能够把给定的拉格朗日量理论视为包含多余的非物理变量，其演化取决于动力学且是非决定论的，与那些演化完全由动力学决定的物理可观测量并列。不过，在目前情况下，要做到明确区分还有点难：显然好的变量是电场和磁场，而差的是那些在 A 里表示额外信息的——我们所关心的是处于规范轨道里的，因此 A 的具体化给了我们剩余信息。约化允许我们永远回避提这种剩余信息。

我们能够在初始数据约化空间的背景下，形式化麦克斯韦理论的哈密顿量方案：在此空间的点具体化了在给定时间空间点上的电场和磁场的值；这个空间是辛的；并且有可能在空间上找到一个哈密顿量，它导致包含 $\mathbf{E}$ 和 $\mathbf{B}$ 的麦克斯韦方程动力学的演化。①

担心解的约化空间是否也支持麦克斯韦理论的拉格朗日量方案是很自然的。那就是说，存在一个 $\mathbf{E}$、$\mathbf{B}$ 及其导数的拉格朗日量，使其变分问题具有跟其欧拉—拉格朗日方程一样的 $\mathbf{E}$ 和 $\mathbf{B}$ 的麦克斯韦方程吗？

乍一看，似乎我们仅仅可以为此目的使用我们最初的拉格朗日量：

$$L := -\frac{1}{2}(|\mathbf{B}|^2 + |\mathbf{E}|^2)$$

但是这会导致错的运动方程，并且有理由担心 $\mathbf{E}$ 和 $\mathbf{B}$ 不适合于变分方法，因为它们的 6 个分量是不独立的——它们能够从有 3 个分量的矢量势导出——从而不能够独立变化。②因此解的约化空间不支持麦克斯韦理论的拉格朗日量方案。

不管这个问题是否在闵可夫斯基时空里的麦克斯韦理论情况下是不可克服的，其他问题就摆在面前。假设我们建构我们的时空 V，把闵可夫斯基时空的空间维度之一卷缩起来：V 是局域闵可夫斯基的但具有 $\mathbb{R}^3 \times S^1$ 的整体结构，这跟我们的理论形成巨大差异。在解空间里规范轨道是如下形式：对于所有适当的 Λ 都有 $[A] := \{A + d\Lambda\}$ 仍然是对的，而且具体化一个规范轨道 $[A]$ 决定时空上的电场和磁场这一点仍然是对的。但是我们能够在其他方向上更进一步，这一点将不再正确。比如为了具体化规范轨道 $[A]$，人们不得不具体化除了 $\mathbf{E}$

---

① 参见[Marsden and Weinstein，1982]。

② 对于这一点参见[Goldstein，1953，366]，对于这种担心见[Sudbery(萨德伯里)，1986]，只不过，其中要求目前的拉格朗日量理论概念的略微推广。

和 B 之外的单个复数，我们称其为绕异性(holonomy)。直观上，绕异性度量了相位改变，那是一个电子沿着所给包围空间闭合维度一圈平移的结果。因此在解的约化空间里的点能被视为由 E 和 B 的具体化加上绕异性构成。这个额外数破坏了一切：因为一旦 E 和 B 是适当的局域对象，赋予一个属性给时空的每一点，绕异性就是一个非局域的东西，这一点甚至在我们寻找一种描述约化解空间方法时更加清楚，这种空间并不具有包括两种截然不同变量的奇怪特征。最好的继续研究的办法显然是把解的约化空间里的点，描述为某种(高度约束的)对时空里每一个闭合曲线的复数赋值。因此在这样一种拓扑学非平凡时空里，为了具体化规范轨道[A]，我们需要具体化非局域信息。目前的框架要求一个拉格朗日场理论涵盖每一个时空点属性的赋值，从而不符合这个例子。①

注释 39(拉格朗日量和解的约化空间)。在例子 37 中考虑的任何简单粒子理论里，我们看到一个情况，承认规范对称的理论解的约化空间从原始理论继承了一个拉格朗日量，它包含原始动力学的规范不变性方面。但是在麦克斯韦理论更有趣的情况下，就像在例子 38 考虑的，似乎不太可能存在任何意义上的解的约化空间直接从局域拉格朗日量产生，而不经过一个承认规范对称的形式化体系。并且这似乎的确很有可能，如果我们选择我们的时空是拓扑学上非平凡的，因为在此情况下麦克斯韦场显然牵涉一个非局域的自由度。

要注意的是事情在非阿贝尔(non-Abelian)的杨—米尔斯理论里变得更糟糕，在这些理论里，场空间是在时空上适当主丛 $P\rightarrow V$ 上联络 1 形式的空间，拉格朗日量是麦克斯韦理论拉格朗日量的直接推广，并且规范对称群是 P 的垂直自同构的群，解的约化空间是 P 的联络模垂直自同构空间。即使当 V 是闵可夫斯基时空，最好的解的约化空间的参数化显然也是在处理围绕时空里闭合曲线的线异性。② 因此要想通过局域拉格朗日量的变分问题(或许有可能)抓住这

① 那就是说，我们看起来是在讨论要求比以点为例更大的某种东西，违反休谟式随附性(见前面的定义 45)。

② 当然，在哲学家中间关于经典杨—米尔斯理论的最好解释存在相当大的争议，参见[Belot(贝洛特)，2003，Chapter 12]。

个解的约化空间还是显得有些困难。① 的确，在物理理论里规范自由的普遍存在，是由于包含非物理的变量，人们有时能够把一个本质上非局域理论弄成局域形式，这一点似乎是合理的。②

6.3 时间—相关系统

让我们假设我们的时空 V 承认一个片断，以及我们的运动方程 Δ 是二阶的并且展示好的存在性和唯一性。③但是现在假设我们的拉格朗日量 L 是时间相关的，在其并不承认从一个 V 上的时间变换群 $\bar{\tau}$ 产生的一个动力学时间变换群 τ 的意义上。

产生于物理学里的时间相关拉格朗日量理论分成下面两个情况。

情况(A)：V 承认一个时间变换群 $\bar{\tau}$，但是这个群并不对应于运动方程的对称。例子：一个牛顿时空里的粒子系统，容易受到从时间相关的势产生的力。

情况(B)：V 并不承认一个时间变换群，比如，设(V, g)是一个弯曲的没有时间对称的广义相对论时空，并且把(V, g)上标量场的克莱因—戈登方程 $\nabla^a \nabla_a \Phi - m^2\Phi = 0$ 作为运动方程(注意 V 上度规在定义微分算子时起作用)，相应的拉格朗日量是 $L = \frac{1}{2}\sqrt{-g}\nabla_a \nabla^a\Phi + m^2\Phi^2$。

我们也需要时间相关的哈密顿系统的概念。

定义 40(时间相关的哈密顿系统)。一个时间相关的哈密顿系统(M, ω, h)，由一个叫相位空间的辛流形(M, ω)结合一个叫哈密顿量的光滑函数 h：

---

① 在跟目前这个不同的一个用法下，任何在速度相位空间上的哈密顿量(即切丛)算作是拉格朗日量理论，参见[Abraham *et al.*, 1988, Chapter 8]，在此替代用法下，拉格朗日量不再要求是局域的，并且变分原理起不了必要作用。最好是存在这种意义上的拉格朗日量理论的处理，而不是我在文中关注的那种意义上的。

② 关于这一点，参见[Belot, 2003, §13]，对于规范自由度重要性的进一步反思，参见[Redhead(雷德黑德), 2003]。

③ 回顾前面 5.2 节，一个时空片断是空间和时间的分解，不是任何片断都满足这个分解跟 V 上时间变换群一致的强条件，只有具有强到足以决定瞬间族和可能点粒子世界线族的几何才承认片断。

$\mathbb{R} \times M \to \mathbb{R}$ 构成，我们常常把 $h(t, \cdot): M \to \mathbb{R}$ 写成 $h(t)$。

通常哈密顿系统(参看前面定义 8)是时间相关的哈密顿系统，其中 $h(t)$ 是对于 t 的每一个值在 M 上的相同函数；我们也叫这个系统为时间无关的哈密顿系统。在时间无关的系统里，动力学轨迹可以被想象成是在相位空间里的曲线，由原点的选择来决定参数化，具有确切地穿过空间每一点的这样一条曲线。在时间相关的情况下，情况更加复杂化。对于 t 的每一个值，我们能够对由 $h(t)$ 产生的矢量场 $X_{h(t)}$ 解方程 $\omega(X_{h(t)}, \cdot) = dh(t)$。那时，如果对于每一个 $t \in \mathbb{R}$、$\dot{\gamma}(t) = X_{h(t)}(\gamma(t))$——即对于每一个 t，在 $x = \gamma(t)$ 对于 γ 的切矢量是由在 x 的矢量场 $X_{h(t)}$ 的值给定的，我们就能说曲线 $\gamma: \mathbb{R} \to \mathcal{I}$ 是 $(\mathcal{I}, \omega, h)$ 的一个动力学轨迹。注意，一旦在时间无关的哈密顿系统的情况下，存在通过相位空间每一点的单个动力学轨迹，在目前情况下一般也存在很多这样的通过每一点的轨迹(由于其态直接跟在 $x \in \mathcal{I}$ 的后面，依赖于通过 x 的动力学轨迹的切线；并且在时间相关的情况下，这个切线也会变化，就像我们考虑产生初始数据 x 在不同可能瞬间那样)。

给定我们所用的假设，我们期望在我们考察时间相关的拉格朗日量理论时有如下发现。

拉格朗日绘景。人们能够运用通常的变分程序得到从一个拉格朗日量到一个运动方程的集合，我们也能够按照通常办法，装备具有 2 形式的 Ω 对应的解空间 $\mathcal{S}$，并且，像通常一样，人们假设对于这种出现在物理学里的例子，运动方程解的唯一性意味着 Ω 是辛的。① 不过，要注意在所考虑的这种类型的时间相关理论里，$\mathcal{S}$ 并不支持一个实现时间变换的单参数群：在落入上述情况(A)的理论里，这样一个群虽然作用在运动学上可能场的空间上，但是(一般是)映射解到非解；在落入情况(B)的理论里，不存在适用的时间变换概念。我们通常能够应用场的应力—能量张量来定义沿着任何给定瞬间的能量——但是结构不再独立于瞬间的选择。

哈密顿绘景。我们时空 V 片断的选择导致哈密顿绘景，其中很多方法类似于出现在时间无关的情况下的情形。设 S 是一个同胚于任意常数 $\Sigma \subset V$ 的流形

① 对于$(\mathcal{S}, \Omega)$在时间无关的情况下构造的讨论，见[Woodhouse，1991，§2.4]。

(并且具有几何，无论如何都具有同样的瞬间)，并且设σ是采用S为一个抽象瞬间的V的片断，即把σ的选择想成具备同时性的最佳观测者族的选择。那么我们就能够开始建构我们理论的哈密顿量方案，尽可能地按照时间无关情况的方法。

1. 给定一个我们片断里的常数Σ和一个解Φ，我们定义对Σ场的限制ϕ；沿着Σ对Φ的时间变换率$\dot{\phi}$，相对于观测者和定义σ的时钟；并且$\pi := \frac{\partial L}{\partial \dot{\phi}}$，沿着$\Sigma$的场动量与片断$\sigma$相联系。

2. 给定一个解Φ和在我们片断里的瞬间Σ，我们把由Σ上的Φ产生的初始数据(ϕ, π)用σ拖回到S，因此，在方便的时候，把ϕ和π想成定义在S上的函数。

3. 设Q是按此方式出现的全部ϕ：$S \rightarrow W$，那么$T^* Q$是按此方式出现的全部(ϕ, π)对的空间。这就是我们的初始数据的空间$\mathcal{I}$，它支撑了一个正则辛形式ω(参见前面例子7)。

4. 哈密顿量的建构是我们进入任何新情况的第一个阶段。[①]设Σ_t是一个我们片断里的瞬间，并且定义$h(t)$：$\mathcal{I} \rightarrow \mathbb{R}$为$h(t)(\phi, \pi) := \int_{\Sigma_t} \pi\dot{\phi} - L(\Phi)dx$，其中$\Phi$是在$\Sigma_t$约化$(\phi, \pi)$的解，并且$\dot{\phi}$是$\Phi$在$\Sigma_t$上约化的场速度。一般情况下，这个建构产生一个对$t$的每一个时间值而言的$\mathcal{I}$上的不同真值函数，一旦初始数据$(\phi, \pi)$产生在瞬间$\Sigma_t$上，人们期望$h(t)(\phi, \pi)$给出总的瞬间的能量。但是在不同的时间施以同样的初始数据，一般导致具有不同总能量的态(简单说，是由于我们处理的是容易受到时间相关的势影响的系统)。

5. 现在考虑t为一个变量，我们看到我们已经定义了一个光滑的h：$\mathbb{R} \times \mathcal{I} \rightarrow \mathbb{R}$，因此$(\mathcal{I}, \omega, h)$是一个在前面定义40意义上的时间相关的哈密顿系统。得到的动力学能够按如下理解，假设我们对动力学感兴趣，其结果出现在一旦把我们的初始数据施加在我们的片断里的一个固定瞬间Σ_{t_0}上。那么，对于每一个$s \in \mathbb{R}$，我们可以问，在时间的s单位后，演化成发生在Σ_{t_0}上的什么样的

① 对于这个建构，见[Kay(凯)，1980，§1]。

态 $x \in \mathcal{I}$，我们称为结果 $g_s^{t_0}(x)$。这给了我们一个对于 s 的映射 $g_s^{t_0}$：$\mathcal{I} \to \mathcal{I}$，并且集合 $\{g_s^{t_0}\}_{s \in \mathbb{R}}$ 形成一个单参数群，$g_s^{t_0}$ 的每一个是$\mathcal{I}$的辛自同构，而并不保留 h 的不变性。因而在此我们具有并非$(\mathcal{I}, \omega, h)$对称实现的，而是$(\mathcal{I}, \omega)$的对称实现的动力学。t_0 是变化的这一设定，给我们这样一个单参数动力学实施群的单参数族。

绘景之间的关系。就像在时间无关设定里一样，对于我们片断里的每一个 Σ，我们能够定义映射 T_Σ：$\mathcal{S} \to \mathcal{I}$，把解发送到在 Σ 上引入的初始数据，因为我们假设对于运动方程的解有全局存在性和唯一性，每一个这样的映射都是双射，而且，正如在时间无关的情况下一样，每一个这样的 T_Σ 事实上都是$\mathcal{S}$和$\mathcal{I}$之间的一个辛同构。我们能够用这些映射证明，前面建构的时间相关的哈密顿系统包含我们运动方程的正确动力学：设 Φ 是一个解，设 x_0 是由 Φ 在 Σ_0 上约化的初始数据，设 $x_0(t)$是在初始数据空间里的相应动力学轨迹，那么对于每一个 $t \in \mathbb{R}$，$x_0(t)$是在Σ_t 上由 Φ 引入的初始数据。

在时间无关的情况下，我们也发现映射 T_Σ 在解空间上实现时间变换和在初始数据空间上的时间演化的群的作用交错在一起。在目前的情况下，我们至今仍没有什么东西对应于解空间上的时间变换；而在初始数据空间上，我们具有时间演化概念的整个族(由一个初始数据被施加上的瞬间选择所标记的)。现在注意到对于每一个我们片断里的常数Σ_t 和每一个 $s \in \mathbb{R}$，我们能够定义 $\hat{g}_s^t := T_{\Sigma_t}^{-1} \circ g_s^t \circ T_{\Sigma_t}$，族$\{\hat{g}_s^t\}_{s \in \mathbb{R}}$是一个$(\mathcal{S}, \Omega)$的辛自同构单参数群，这不是我们拉格朗日量的变分对称群。运用 $\hat{g}_s^t$ 到解 Φ 的结果是下面情况可能导致的解，如果由 Φ 在Σ_t 上引入的初始数据已经被产生在瞬间Σ_{t-s}上。①

时间。在当前的情境下，时间变换可能是或者不可能是一个我们时空的对称；但是如果它是，则不存在相应的动力学对称，并因此我们的图景是困难重重的——我们没有得到解空间上和初始数据空间上实数的良好作用，以实现时

① 当然，在时间无关的情况下，这约化了到解的时间变换——因此我们能够把变换 $\hat{g}_s^t$ 看作解的时间变换的普通概念的推广。

间变换和时间演化。在初始数据空间上，对于我们可能选择施加初始数据在其上的每一个瞬间，我们获得一个实现时间演化的单参数群，但是这并不是哈密顿量对称的。在解空间上，我们没有对应于时间变换的自然群作用。如果我们选择一个片断和瞬间，那么我们得到的$\mathbb{R}$作用给我们的信息不是解的时间变换，而是关于什么样的最终解，如果我们把所给解约化的初始数据取在所给瞬间上，并且重新施加在另一个瞬间上。

变化。有些物理量会用初始数据空间上的函数表示。比如，在受到时间相关的外力的两个牛顿粒子的理论里，粒子之间的相对距离会包含在初始数据空间上的函数里。但是一些量会通过初始数据空间上函数的单参数族来表示，比如能量也会是在任何时间相关的系统里这样一个量的例子。① 正如前面我们用哈密顿量所作的，让我们用符号$f(t)$来代表这样一个单参数族——我们能够把一个普通函数想成是简并情况，其中$f(t)$是对于每一个$t\in\mathbb{R}$而言在$\mathcal{I}$上同样的函数。设$x(t)$是在$\mathcal{I}$里的一个动力学轨迹，那么$x(t)$表示作为变化的$f(t)$模型化的量，当且仅当$\exists t_1$，$t_2\in\mathbb{R}$使得$f(t_1)(x(t_1))\neq f(t_2)(x(t_2))$。

在解空间上，我们期望，一旦我们选择了一个片断，每一个我们关心的量都会像往常一样表示为函数的单参数族——通常，我们用$\{f_t\}$代表这样一个$\mathcal{S}$上函数的族。假设我们所关心的量是由在初始数据空间上的$f(t)$表示的，并且设Σ_{t_0}是我们片断里的一个瞬间，那么我们定义$f_{t_0}:=f(t_0)\circ T_{\Sigma_{t_0}}$: $\mathcal{S}\to\mathbb{R}$，对每一个$t\in\mathbb{R}$实施这一点给了我们所期望的$\{f_t\}$。就像平常一样，如果$\exists t_1$，$t_2\in\mathbb{R}$使得$f_{t_1}(\Phi)\neq f_{t_2}(\Phi)$，我们把一个解$\Phi\in\mathcal{S}$视为表示我们作为变化的量。

注释41（人为时间相关理论）。如果我们具有时间相关的拉格朗日量理论，但是仍固执地选择一个不适合我们时间变换概念的片断，那么我们按照前面的过程得到的结果可能是一个时间相关的哈密顿系统。

注释42（时间相关系统的量子化）。在量子化一个具有有限多个自由度系

① 很难发现其他例子，在一个不稳定时空的场论集合里，抽象瞬间不会支撑黎曼度规（因为瞬间$\Sigma\subset V$不共享一个黎曼几何）。在此情况下，我们发现一个把场表示为具有两个尖峰的初始数据集合，将会对应于其峰为不同位置的瞬时态，它们依赖于在其中初始数据产生在片断里的常数Σ_t上。因此在这种例子里，即使是相对距离也由初始数据空间上函数族表示。

统的时间相关的哈密顿处理时，不存在特殊困难。但是一般不可能建构一个时间相关场论的好的量子哈密顿量。① 因为这个原因，在弯曲时空上自由量子场论的标准建构，作为出发点的是解空间而不是初始数据空间。②

7. 广义相对论里的时间问题

广义相对论不同于上述考虑的那些理论的地方就在于它是广义协变的。被广泛接受的一点是，这会导致某些专门的技术性和概念性问题，它们被纳入到时间问题这个题目之下。这一节给出一个关于广义相对论中时间问题的扩展性评论。接下来的第一小节致力于广义相对论广义协变性及其直接结果的讨论。下一小节包括时间问题自身的讨论——尤其是变化不能在理论里以一种前面第5节和第6节里熟悉的方式讨论。最后一小节讨论一个在广义相对论里发现时间和变化的策略(这个讨论倾向于以进一步澄清时间问题的方式进行，而不是提出一种解决方案)。

重要的是要强调目前的讨论集中在广义相对论，而且所讨论的问题在人们涉及某种意义上的广义协变性理论时也出现。

7.1 广义相对论的广义协变性

设 V 是一个时空流形，不管是否具有几何结构。回顾一个 $C^k(0<k\leqslant\infty)$ 微分同胚 d: $V\to V$，是一个具有 C^k 逆的 C^k 双射。③ 可微性程度尚不确定，我

---

① 参见[Kay，1980，§2.1]。

② 参见[Wald(瓦尔德)，1994，Chapter 4]。

③ 一个微分同胚 d: $V\to V$ 被称为小的，如果它同伦(homotopic)于单位，否则它就是大的。为了便于表述，我在下面关注的都是小的微分同胚。我也常常说用微分同胚对张量的拖曳。在大多数重要情况下，将会是微分同胚 d 对时空度规 g 的拖曳 d^*g。直观地看 $(V,\ d^*g)$ 是这样的时空几何，如果我们从 V 提升 g，然后用 d 重新排列 V 的点的单位，再让 g 回来。$(V,\ g)$ 和 $(V,\ d^*g)$ 共享一套时空点，并且具有同构的几何；它们仅仅就 V 中什么样的点在什么几何起作用方面不同——除非 d 是一个度规对称的，在这样的情况下它们连这样的不同也没有。

们用$\mathcal{D}(V)$代表从V到自身的微分同胚群。①

粗略地讲，我们要说一个理论，在其具有$\mathcal{D}(V)$为一个对称群时，一般是协变的。② 因此对于几个理论对称概念中的每一个，我们具有广义协变的相应概念。按照[Earman，2006]，我从我们的目的出发挑出下面两个最重要的：

弱广义协变性：$\mathcal{D}(V)$是理论的方程 Δ 的一个对称群；

强广义协变性：$\mathcal{D}(V)$是理论的拉格朗日量 L 的一个规范对称群。

当然，强广义协变性意味着弱广义协变性成立(因为每一个规范对称是一个变分对称，由此有运动方程的对称)，但是反过来不成立(一个理论可能是弱广义协变的，即便它并不承认一个拉格朗日量，因此没有资格成为强广义协变的)。

我们已经允许 V 拥有一个固定几何背景，包含在某些并不从解到解变化的张量里。我们可能允许 V 具备进一步的非几何解无关的结构。③ 另一方面，在任何非平凡理论里，由运动方程统治的场当然会随解而变化。因此我们已经在理论之间作了一个划分，一些理论装备了非平凡解无关结构，一些解则没有。④

直观上看，一个理论是弱广义协变的，当且仅当它的解不具有解无关张量(或者旋量、或者……)——因为它是确定的，一旦我们具有某些在 V 上固定的背景张量，其运动方程能够“关注”一个解 Φ 跟由一个微分同胚 d：$V\rightarrow V$ 得到

---

① 在处理微分同胚群时要特别小心。一方面，从紧致流形到自身的 C^k 微分同胚群具有一个很好的可微结构——它是一个巴拿赫流形——但不是一个巴拿赫李群，因为群的乘法运算不是光滑的；另一方面，从紧致李群到自身的 C^∞ 微分同胚群，具有不太满意的微分结构——它是一个纯粹的弗雷谢(Fréchet)流形——但是它是一个弗雷谢李群。详情见[Adams *et al.*(亚当斯等)，1985]和[Milnor，1984]。这个情形对于从非紧致流形 V 到自身的微分同胚群更糟糕：显然人们需要假定某些 V 上的几何结构，以便给群一个微分诺顿(Norton)结构，参见[Cantor(坎托尔)，1979]和[Eichhorn，1993]。参见[Isenberg and Marsden，1982]绕过这些困难的技巧。

② 有关广义协变性概念的纠葛史参见[Norton，1995]。

③ 比如，在研究强的外部电磁场中的带电物质的运动时，我们可能采用这样的理论，其中麦克斯韦场和时空几何是解无关的，并且只有物质的运动从解到解是不同的。

④ 注意在解无关跟解相关结构之间所作的区分，跟安德森—弗里德曼(Anderson-Friedman)所作的绝对对象跟动力学对象之间的区分并不一致，参见[Friedman，1983，§Ⅱ.2]：解无关对象被要求从解到解是一样的，而绝对对象仅仅要求从解到解只是关于微分同胚是一样的。

的拖回$d^*\Phi$之间的区分。

当然，广义相对论是弱的广义协变性——的确，在广义相对论的重要真空截面里，时空度规是理论的仅有基本量，并且它是解相关的。①

广义相对论是否满足强广义协变性的问题更加微妙。直观上看，它应该说是，因为在形式上V的微分同胚是理论拉格朗日量的变分对称，而这样的微分同胚群通过V上矢量场的集合在适当意义上参数化。但是，正如我们即将看到的，这是边界条件的技术不能忽略的一点。

但是我们能够勾画出这些技巧，主要是关注自子集$\mathcal{D}_C(V)\subset\mathcal{D}(V)$，它由从$V$到自身的紧致具有的微分同胚组成。② $\mathcal{D}_C(V)$变成广义相对论拉格朗日量的规范对称群，$\mathcal{D}_C(V)$是由V上紧致支撑的矢量场族所参数化了的，因此上述条件(2)相反的情况在$\mathcal{D}(V)$被$\mathcal{D}_C(V)$取代后是行得通的。

为了进一步着手处理有关时间问题的广义协变性意义的问题，我们转向下面两个具体情况：(i)空间上紧致区域的广义相对论；(ii)渐进平坦施加在类空无限远上的区域里的广义相对论。第一个情况主要集中在宇宙学，由于要求空间是紧致的，人们消除了对空间无限远上边界条件的担心，这就允许人们考察物质聚集的宇宙，同时保持对技术问题的控制。第二个情况是关注更加严格的数学上和概念上的利益(最大物理利益的渐进边界条件，把渐进平坦施加在空无穷远而不是空间无穷远，这允许人们考察引力辐射)。在讨论这些情况后，我简单转向是否每个问题都能够给出广义协变公式的疑问。

作为宇宙学理论的广义相对论

我们先看看时空度规g是唯一场的真空广义相对论。因此我们把我们运动学上可能场的空间，作为在一些固定n维时空流形V上的洛伦兹(Lorentz)记号度规的空间。③这个理论的运动方程是$R_{ab}-\frac{1}{2}Rg_{ab}=0$，其中$R_{ab}$是$g$的里奇曲

① 在此方法里，爱因斯坦场方程只是告诉我们，如果V上度规g算作一个解，当且仅当g的里奇曲率张量为零。并且显然g是里奇平坦的，当且仅当d^*g是里奇平坦的。因此$\mathcal{D}(V)$映射解到解。

② 那就是，一个微分同胚d: $V\to V$是在$\mathcal{D}_C(V)$里的，当且仅当存在一个紧致集合$U\subset V$，使得d作为单位作用在V/U上。

③ 因此一个运动学上可能场是V上洛伦兹记号的对称双线性形式丛的截面。

率，而R是g的曲率标量；在这里我们要求宇宙学常数为零。

回顾子集$\Sigma \subset V$被称为$(V,\ g)$的柯西面，如果每一个$(V,\ g)$里的不可延拓类时曲面与Σ相交一次，则柯西面是V的一个$(n-1)$维类空子流形。我们称$(V,\ g)$是全局双曲的，然后能够通过柯西面被叶形化，并且其全部柯西面同胚于另外一个。确实，如果$(V,\ g)$是全局双曲的，那么V是同胚于某些$(n-1)$维流形S的$S \times \mathbb{R}$形式的流形，而$(V,\ g)$的全部柯西面同胚于S。为了把广义相对论作为宇宙学理论讨论的这个目的，我们先看看具有紧致可定向柯西面的解。①

我们能够着手建构我们理论的拉格朗日量方案和哈密顿量方案。

拉格朗日绘景。广义相对论的拉格朗日量是爱因斯坦—希尔伯特拉格朗日量$L = \sqrt{-g}R$。自然，解空间是无穷维的，让我们称一个解是良好的，如果它通过具有常数平均曲率的柯西面承认一个叶形。②相信良好解集合形成完全解空间的一个庞大开子集，并且知道在良好解空间内发生的仅有奇异性是承认基灵场度规上温和的那个（这些都是矢量场，能够被想成时空对称的无穷小产生元）。③群$\mathcal{D}(V)$是爱因斯坦—希尔伯特拉格朗日量的规范对称群：每一个从V到自身的单参数群是这个拉格朗日量的变分对称群，并且群$\mathcal{D}(V)$能够通过任意V上矢量场被参数化。④因此，按照前面6.2节里发展起来的有关各种规范理论的理论，我们发现良好解的空间$\mathcal{S}$支撑一个预辛几何形式Ω（从此我不要修饰语而说$\mathcal{S}$是解空间）。⑤像通常一样，这个预辛几何形式通过规范轨道约化了解空间的分离。两个度规g和g'属于同样的规范轨道，当且仅当存在微分同胚d: $V \to V$使得$g' = d^{*}g$。当然，跟微分同胚的单参数群结合在一起的守恒量是平凡

① 全局双曲线解的限制不要求广义相对论拉格朗日量方案的建构，而是要求哈密顿量的处理，并且在下面引用的涉及解空间结构的某些结果里起作用。要求空间拓扑要有方向是哈密顿量处理所必需的。

② 平均曲率定义在下文中初始数据空间的讨论过程中。

③ 关于良好解空间的结构，参见[Isenberg and Marsden，1982]。

④ 参见[Crnković and Witten（申科维奇和威腾），1987]和[Woodhouse，1991，pp. 143—146]。后者提供一个论证，非紧致具有微分同胚属于拉格朗日量的规范对称群。

⑤ 关于不通过拉格朗日形式进行的解空间上的预辛几何形式的建构，也见[Frauendiener and sparling（弗朗迪纳和斯帕林），1992]。

的——每一个都是$\mathcal{S}$上零函数。确实，在此情况中，广义相对论没有非平凡诺特量——超出微分同胚，定律的唯一连续局域对称都是度规重标度，这些不是变分对称。①

解的约化空间。广义相对论解空间的规范轨道空间$\mathcal{S}'$，是在对应于具有基灵场解的点上温和奇异性的辛空间。② 让我们称解的约化空间里的点$[g]$为几何——因为$[g]$的不同表象代表 V，它具有同样时空几何，而对 V 的点的几何作用分配则不一样。就我所知，把这个约化空间说成是源自于局域拉格朗日量变分问题解空间是没有意义的。确实，一个几何$[g]$不可能赋予任何特定局域属性给任何点 $x \in V$。

哈密顿绘景。相应的哈密顿绘景的建构需要注意一下。③ 我们想尽量模拟前面第 5 节中提到的很多过程，假定我们并没有适当的一个片断(这要求时空具有非平凡解无关的几何)。我们按如下内容进行。④

1. 建构初始数据空间。迄今为止，我们能够做的是：(i)选择 V 的片断 σ 和 σ 里的瞬间 $\Sigma \subset V$，然后建构可能瞬时场位形空间$\mathcal{Q}$，办法是查看把解 Φ 限制到Σ产生的全部 ϕ：$\Sigma \rightarrow W$；(ii)建构初始数据空间$\mathcal{I} \subseteq T^{*}\mathcal{Q}$，办法是寻找所有的$(\phi, \pi)$对，它们被引入为所有$\Sigma$上初始数据(其中 π 是瞬时场动量，通过 $\pi := \frac{\partial L}{\partial \dot{\phi}}$定义，$\dot{\phi}$ 是场按照跟片断 σ 相关的观测量的时间变化率)。我们发现，只要在理论的拉格朗日量 L 承认规范对称群的时候，$\mathcal{I}$是 $T^{*}\mathcal{Q}$的一个真正子集。不依赖于片断的概念，我们也能够通过特别接近于通常的一个程序来建构初始数据空间。

如果$\Sigma \subset V$是超曲面，而 g 是爱因斯坦方程一个解，那么 $q := g|_{\Sigma}$是一个

---

① 参见[Torre and Anderson，1996，489]。

② 见[Isenberg and Marsden，1982]。

③ 对于下面的建构，参见[Wald，1984，Appendix E. 1]或者[Beig(贝格)，1994]。

④ 这里勾画的建构并不依赖于消失转移的场。固定这些非物理场的情况，允许人们从抽象瞬间 I 上的初始数据 S 到对于某些(可能小的)实数期间 I 来说 $S \times I$ 上的一个解。像这样它们允许人们在初始数据空间跟有限时间范围内解的集合之间建立双射。我在此避免了消失和转移，因为我想集中在全局结果和物理场上。对于消失和转移形式最好的介绍，见[Marsden *et al.*，1972，§Ⅲ]。

二阶的对称协变张量。但是在目前设置中，对任意超曲面解的限制不是场的一个瞬时位形的好的候选者。直观上讲，由于广义相对论引力场是空间几何，这个场的瞬时位形也应该是空间几何。不过，肯定 $q := g|_{\Sigma}$ 是在 Σ 上的黎曼度规，当且仅当 Σ 根据 g 是类空的。因此 $q = g|_{\Sigma}$ 表示场的瞬时态当且仅当 Σ 是类空的。① 因此，这似乎是合理的，即取作可能场位形的空间 Q，通过把每一个解限制到由它引起类空的超曲面而出现的黎曼度规 q 的空间。② 瞬时场动量的定义更加复杂。在所熟悉的情况下，片断 σ 起到重要作用。但在目前情况下找不到那样的情况：不好在广义相对论背景中引入片断的独立解概念，来作为动力学理论。③ 然而，对此困难有种办法，考虑一个解 g 和一个 g 表示为类空的超曲面 $\Sigma \subset V$，相对于 g 我们能够在通常意义上选择 V 的一个片断（因为相对于 g，我们能够选出瞬间和作为 V 子流形的点粒子世界线）。我们称这样一个片断为高斯的，如果它对应于一个在其通过 Σ 时其时钟全部为零，并且其世界线全部正交于 Σ 的自由下落观测者的集合。对于有效小的 t，常数 t 的超曲面，按照高斯观测者，会是支撑黎曼度规 $q(t) := g|_{\Sigma_t}$ 的柯西曲面。因此给定一个 Σ 的高斯曲面，我们能够定义 $\dot{q}_{ab} := \frac{\partial q_{ab}(t)}{\partial t}\big|_{t=0}$，这是在 Σ 上的二阶对称协变张量。事实上，$\dot{q}_{ab}$ 独立于高斯片断选择，并且能够被看作告诉我们有关 Σ 嵌入在 (V, g) 的几何。我们能够采取 Σ 在 (V, g) 里的外曲率相类似的观点，$k_{ab} := \frac{1}{2}\dot{q}_{ab}$，并且沿 Σ 在 (V, g) 里的平均曲率，$k := q^{ab}k_{ab}$。现在相对于我们的高斯片断，张

---

① 在空间紧致全局双曲线区域，一个具有对于 (V, g) 柯西曲面拓扑学的 (V, g) 的子流形是柯西曲面，当且仅当根据 g 它是类空的。见[Budic *et al.*（别克等），1978，Theorem 1]。

② 场 q 被看成是定义在一个抽象瞬间 S 上，它微分同胚于柯西曲面 $\Sigma \subset V$。为了建构 Q，我们选择相对于 (V, g) 的瞬间 Σ，和一个微分同胚 d: $S \to \Sigma$，并且用 d 拖回到 S，q 的全部是作为使 Σ 类空解的 Σ 限制而出现。当然 Q 是独立于 Σ 和 d 的选择的。

③ 假设 σ 是一个相对于 V 上度规 g 的片断，那么 g 对 σ 里的瞬间 Σ_t 的限制就会是合理的瞬时场位形，由此相对于 σ，解 g 应该对应于理论初始数据空间里的曲线。但是如果我们看看相对于 σ 的另一个解 g' 又会怎么样呢？一般来说，限制这个新解到 σ 里 Σ_t 的结果，并不是场的一个瞬时态——因此 σ 不会给我们这样的方法在初始数据空间里联系每一个时空解 g 这样一条轨迹。

量$\dot{q}_{ab}(0)$表示引力场的速度，在其代表长的时间变化率意义上，像通常一样，我们能够定义对应的动量为 $\pi^{ab}:=\frac{\partial L}{\partial \dot{q}}$。对于爱因斯坦—希尔伯特拉格朗日量，我们具有 $\pi^{ab}=(\sqrt{q}k^{ab}-kq^{ab})$(因此动量是二阶对称协变张量)。①我们采取初始数据空间，即(q，π)对的空间$\mathcal{I}\subset T^{*}Q$作为场位形以及在超曲面上解所引入的动量而产生，它们致使类空的出现(具有在抽象流形 S 上的 q 和 π 函数)。就像是在规范对称的理论里期望的，$\mathcal{I}$是 $T^{*}Q$的真正子空间，并且在 $T^{*}Q$上正则辛形式的限制用一个预辛形式 ω 装备$\mathcal{I}$。ω 的规范轨道具有下面的结构：初始数据集合(q，π)和(q'，π')属于同样的规范轨道，当且仅当它们产生于同样解 g 的初始数据。②

2. 建构哈密顿量。运用建构哈密顿量的常用规则所给的拉格朗日量导出哈密顿量 $h\equiv 0$。

3. 建构动力学。令人印象深刻的一般动力学方程，按照由解 $\omega(X_h,\cdot)=dh$ 的矢量场 X_h 产生的动力学轨迹导致动力学轨迹是那些由零矢量场产生的曲面这样的结论。因此在$\mathcal{I}$里的曲率是一个动力学轨迹，当且仅当它总是处在同样的规范轨迹里。然而，这在物理上是无用的——因为通常我们希望具有规范对称理论的动力学轨迹能够表示出从规范轨道到规范轨道所传递的物理信息。但是在当前情况下，还没有办法做到这一点。一个非零哈密顿量可能导致从规范轨道到规范轨道的动力学轨迹，但是这可能在物理上是无意义的(并且比无用还要糟糕)。因为这样的动力学，有可能把我们从初始数据带到后来的瞬时态，前者能够看成是解 g 的瞬时态，而后者不能看成是解 g 的瞬时态。在这么做的时候，结果可能包含跟爱因斯坦场方程里完全不同的动力学信息。

---

① 如果我们计算相对于一个(V，g)的非高斯片断的 $\dot{q}$，那么我们一般会获得跟高斯片断产生的很不一样的答案。但是如果我们用这个在我们场动量定义里的场速度新概念，我们发现我们的新观测者跟我们关于在Σ的每一个点上的场动量值的原始高斯观测者是一致的。因此，非常奇怪的是在广义相对论里场动量依赖于选择的瞬间而不是片断。

② 更准确地，(q_1，π_1)和(q_2，π_2)属于同样规范轨道，当且仅当存在一个解 g、瞬间Σ_1，$\Sigma_2\subset(V, g)$以及微分同胚 d_1: $S\to\Sigma_1$ 和 d_2: $S\to\Sigma_2$，使得对于 $i=1$，2(q_i，π_i)拖回到S，通过的是g 在Σ_i 上约化的初始数据的 d_i。注意到如果$\Sigma_1=\Sigma_2$ 而 $d_1\neq d_2$，那么(q_1，π_1)和(q_2，π_2)不同，确实在(V，g)里的单个柯西曲面的几何的规范等价描述。

初始数据的约化空间。我们能够过渡到$\mathcal{I}'$，即约化的初始数据的空间：由在初始数据空间里点的规范等价类组成的空间里的点。就像解的约化空间，初始数据的约化空间是具有温和奇异性的辛空间。[1] 确实，这一点是假设的，即两个约化空间作为辛空间都是标准同构的，其映射把初始数据的规范轨道取为解的相应规范轨道。[2] $\mathcal{I}'$从$\mathcal{I}$继承了平凡哈密顿量 $h=0$；这引入了在$\mathcal{I}'$上的平凡动力学，根据的是动力学轨迹是对所有 t 的值形式为 $x(t)=x_0$ 的常曲率。

绘景之间的关系。解空间和初始数据空间不同构——这是具有规范对称理论的一般特征。另一方面，正如我们注意到的，解的约化空间和初始数据的约化空间都被认为是同构的。在一个固定背景空间上的理论的情况下，一个片断产生一个单参数族，它是两个空间之间的辛同构。一个是解空间，另一个是初始数据空间，后者具有双重目的，既要它们各自空间的时空对称交织在一起，又要允许我们建构一个在解空间上的变化的表示。在当前情况下我们仅仅具有一个两空间之间的单独标准同构。

时间。不管是解空间还是解的约化空间，我们都没有找到进行时间变换的实数的作用。我们也没有找到一个在初始数据的约化空间上进行时间演化的非平凡作用，因为哈密顿轨迹在那里全部是平凡的，在初始数据的空间上，我们却有非平凡哈密顿轨迹。但是一个初始数据空间上的动力学轨迹一般不能够被看作代表时间演化，比如，没有东西阻止其周期性，甚至当解对应于规范轨道时，所处轨迹也不是任何意义上的周期性。

不过，存在一类不能视为代表动力学的初始数据空间上的动力学轨迹——这些对应于初始数据序列的轨迹可能集结起来形成合理的时空几何（一旦这是可能的，则结果总是场方程的解）。通过初始数据空间的每一个点，事实上存在很多这样的轨迹。但是就像通常具有规范对称的理论里那样，不存在这样的有利方法，以便削减这个量得到通过$\mathbb{R}$ 作用表示时间演化的明确子集。

变化。让我们取些可变物理量，像宇宙的瞬间空间体积。我们如何表示这

---

① 参见[Fischer and Moncrief（费什尔和莫尼克里夫），1996]。

② 如果(q, π)是 g 里柯西曲面的几何，那么在$\mathcal{I}'$跟$\mathcal{S}'$之间的正则同构把$[q, \pi]$送到$[g]$。

样一个在各种空间起作用的量？不管是在解空间还是在解的约化空间，我们都面临我们常见的问题：这些空间里的点表示永久历史，因此在这样的空间上没有函数能够以一个直接方式表示一个可变物理量。过去，我们能够采用下面的策略规避这个问题：(i)我们可能找到出现在哈密顿方法里空间f上的函数，那么使用一个在这个空间跟解的(约化)空间之间的同构的片断相关单参数族，以便找到一个表示所给量行为的后一个空间上的函数参数族；(ii)或者我们可能找到一个解的(约化)空间上的函数，表示给定瞬间有用量的值，那么使用一个解的(约化)空间上的动力学时间变换群，产生一个这样的函数的单参数族。这样的策略这次没有一个起作用，因为我们根本没有一个由瞬间标记的同构的单参数族，也没有解的(约化)空间上的时间变换概念。

我们事实上不比初始数据的约化空间更好，原因是存在太多对应于系统整个历史的点，并且单个函数是不适合表示可变化量的。在初始数据上，我们面临一个不起眼的困难：如果我们试图通过非规范不变性函数表示可变量，那么我们会面临非决定论；如果我们采用了规范不变性函数，那么我们会面临我们在初始数据约化空间里碰到的本质上相同的情形。

渐进平坦区域中的广义相对论

考虑广义相对论的第二个领域是富有启发意义的，在那里人们要求在空间无穷远处是渐进平坦的。这一情况在物理意义上有些边缘化，但是有利于我们澄清我们在空间紧致情况下所遇到的问题的解决。

在此区域我们的空间是$\mathbb{R}^4$，并且在运动学上可能的场是洛伦兹记号度规的赋值，在适当意义上要求V成为在空间无穷远上渐进平坦的。[①] 瞬间也要求满足渐进条件。

---

① 存在几个在空间无限远渐进平坦的概念。在这一节里，所述结果源自于使用了三种不同但是紧密相关的方法：(i)[Andersson，1987]的方法；(ii)[Ashtekar *et al.*(阿西提卡等)，1991]的方法；(iii)[Beig and Ó Murchadha(贝格和莫萨达)，1987]的方法。为了叙述简便，我忽略了这些方法在文字上的差异——我不认为会导致误解。因为方法(i)和(iii)之间的关系可以见[Andersson，1987，Definitions 2.3 and 2.4]以及[Andersson，1989，78]。方法(ii)和方法(iii)两者都由其传播者设定。至于[Beig and Schmidt，1982]方法，见[Ashtekar and Romano(阿西提卡和罗马诺)，1992，Chapter 7]和[Beig and Ó Murchadha，1987，§4 and §5]。

在此设定里自然要考虑$\mathcal{D}^{\infty}(V)$，这个保持边界条件不变性的微分同胚群，而不是全部微分同胚的群。我们找到在无穷远趋于一样的$\mathcal{D}^{\infty}(V)$的子群$\mathcal{D}_0^{\infty}(V)$，是理论的拉格朗日量公式的规范对称的最大群，并且$\mathcal{D}^{\infty}(V)$是具有庞加莱群的$\mathcal{D}_0^{\infty}(V)$的半直积。$\mathcal{D}^{\infty}(V)$的任何元都能够被看作是$\mathcal{D}_0^{\infty}(V)$的元跟作用在无穷远上的庞加莱对称上的一个积。① 这个理论的解空间带有一个预辛形式并且进入规范轨道，具有两个解在同样规范轨道里，当且仅当它们通过一个$\mathcal{D}_0^{\infty}(V)$里的微分同胚区分。②在$\mathcal{D}_0^{\infty}(V)$里微分同胚固定了规范轨道，这些是在$\mathcal{D}_0^{\infty}(V)$而不是在$\mathcal{D}_0^{\infty}(V)$里改变它们的序列。其意义在解的约化空间层面上非常清楚，这是一个支撑庞加莱群表象的辛空间——特别是，对于每一个在空间无限远的时间变换概念，这个概念都容许一个产生这个概念的非零哈密顿量。③

人们也能够给出这个广义相对论截面的哈密顿量处理。④人们建构初始数据空间就像在空间紧致情况里那样，尽管那些条件一定要加在瞬时场位形和动量的渐进行为上。最终空间拥有一个预辛形式。初始数据集合(q, π)和(q', π')属于同样规范轨道，当且仅当存在一个解g和瞬间Σ，$\Sigma' \subset V$使得Σ和Σ'是由$\mathcal{D}_0^{\infty}(V)$的元联系在一起的，并且$g$引入$\Sigma$上$(q, \pi)$和$\Sigma'$上的$(q', \pi')$。⑤就像在解空间上那样，我们拥有一套函数能够被视为在无穷远的庞加莱群的无穷小生成元，对应于在无穷远的时间变换生成元是一个产生规范等价类动力学概念初始数据空间上的哈密顿量，它们当中的每一个都支撑规范轨道到规范轨道(相较于具有规范对称的通常理论的初始数据空间上的动力学概念)。因此由这样一个哈密顿量产生的一般动力学轨迹$x(t)$，将表示通过初始数据空间的非平凡轨迹；同样的哈密顿量将通过初始数据空间里每一个点产生很多轨迹；但是每一个这样的轨迹将符合有关规范轨道在t的每一个值，其中规范轨道是系

① 见[Andersson, 1987, Theorem 2.2]和[Ashtekar *et al.*, 1991, §3.3]。

② 见[Ashtekar *et al.*, 1991, §3]。

③ 见[Andersson, 1987]。

④ 见[Beig and Ó Murchadha, 1987]。

⑤ 跟空间上紧致情况一样，初始数据空间里的不同点能够对应于同样的场位形和由$\Sigma \subset V$上g约化的动量，如果我们用不同的微分同胚把这些张量拖回到抽象瞬间S。

统状态在那个时间所占据的。

人们期待初始数据的约化空间将是辛空间，同构于解的约化空间，并且容许庞加莱群的表示。选择无穷远处时间变换的概念，应该选出初始数据约化空间上的哈密顿量，其动力学轨迹包含理论的动力学：固定一个时间变换概念、相应的哈密顿量以及在初始数据约化空间里的一个任意点，我们发现通过这一点的哈密顿轨迹包含瞬间数据的等价序列，以及发现选出这些堆成合理时空几何种类表示的方法包含理论的一个解。

因此在该情况下的情形跟我们在上面看到的在空间紧致情况下的很不相同，我们有解的约化空间和初始数据约化空间上庞加莱群的表象，并且我们有这些包含在解空间和初始数据空间上的结构里的表象。

我们能够以一种十分熟悉的方式表示可变的物理量，即通过初始数据约化空间上的光滑函数。不考虑特殊情况，这些函数随状态沿着那个空间里的动力学轨迹移动时改变它们的值，并且我们能够计算这些量的变化率，等等。如果我们通过解空间上的函数来寻求表示的变化——这就要求一些后面 7.3 节里发展出来的工具，情况就更加复杂。但是至少在直观上，很清楚所需要做的是什么。因为对于在解的约化空间里的每一点，就每一个无穷远处的时间变换概念来说，存在这个空间里的点的单参数族，对应于所给定点的时间变换，它应该有可能找到——就对应于可变量初始数据约化空间上的任何函数而言——存在一个包含不同时刻那个量值的解约化空间上的单参数族。

广义协变特殊吗?

爱因斯坦相信广义相对论的广义协变性是一个具有重要物理结果的十分特殊的特征。受狭义相对论里观测的激励，假定存在在任何惯性参照系里定律有同样形式的事实跟所有惯性观测者都等价的事实之间的紧密关联(因此不存在绝对速度的概念)，爱因斯坦希望，因为他的引力理论定律在任意坐标系成立，理论可能是所有观测者都等价的那一个(因此没有任何绝对运动的概念)。

但是，众所周知，这些方法不适合这个目的：在广义相对论里存在一个完美可信的(并且坐标无关的)划分，在被加速的观测者和没有被加速的观测者之

间以及在旋转的观测者和没有旋转的观测者之间作出区分。①

爱因斯坦使其理论应该在任意坐标系成立的要求，仅仅是变换成我们第一个弱意义上的广义协变性的坐标语言。上一段指出这个要求并没有爱因斯坦相信的有力结果。更糟的是，克雷茨曼（Kretschmann）在 1917 年已经指出，广义协变的这个弱意义并非十分罕见的性质：很多前广义相对论理论都能够给出一个弱广义协变性的形式。②实际上，存在一个技巧，以固定背景时空上的拉格朗日场论为输入，就会得到所给理论的一个强广义协变形式/相对性作为输出。③

例子 43（人为广义协变性）。设 T_0 是无质克莱因—戈登标量场理论，Ψ 在固定背景时空（V_0，g_0）上传播。T_0 的拉格朗日量是 $L_0(\Psi) := \frac{1}{2} g_0^{ab} \nabla_a \Psi \nabla_b \Psi$，并且对应的运动方程是 $\Box_0 \Psi = 0$，其中 $\Box_0$ 是对应于 g_0 的达朗贝尔 d'Alembert 算符。④ 给定 T_0 我们能够建构一个强的广义协变理论 G 如下。⑤设 V 是一个微分同胚于 V_0 的流形，T 的时空是不具有任何几何的裸流形 V。T 牵涉两个场 X 和 Φ：X 在 V_0 取值而 Φ 取值于 $\mathbb{R}$，一个（X，Φ）对看作运动学上可能的，只要 X：$V \to V_0$ 是一个微分同胚。⑥ T 的拉格朗日量 L 建构如下：对于运动学上可能的（X，Φ），n 形式的 V 上 $L(X, \Phi)$ 是由 V_0 上的 n 形式的 $L_0(\Psi)$ 拖回到 V 的，其中 $\Psi := \Phi \circ X^{-1}$，$L$ 承认 $\mathcal{D}(V)$ 为规范对称的一个群，因此 T 是强广义协变的。注意一个运动学上可能的（X，Φ）对是 T 的一个解，当且仅当 $\Psi = \Phi \circ X^{-1}$ 是 T_0

---

① 爱因斯坦的思路建立在下面观测基础之上：在狭义相对论里洛伦兹变换是用来决定观测者运动状态的时空度规的对称性的，在广义相对论里任意微分同胚肯定不是一个给定解的时空几何的对称——但是这个几何再次在决定观测者的运动状态中起作用。见[Friedman，1983，Chapters II and V]。

② 关于克雷茨曼见[Rynasiewicz（瑞纳齐维兹），1999]。

③ 人们如何在目前情况下使理论各具特色尚不明显。相关讨论和建议见[Sorkin（索金），2002，698]和[Earman，2006，§4]。

④ 对应于洛伦兹度规 g 的达朗贝尔算符，正好定义为黎曼度规 g 的拉普拉斯（Laplacian）算符：跟 $\mathrm{div}_g \circ \mathrm{grad}_g$ 一样，其中 div_g 是 g 的散度算符，而 grad_g 是 g 的梯度算符。

⑤ 见[Lee and Wald，1990，734]或者[Torre，1992，§II]。同样的过程对于具有一阶拉格朗日量的任何标量场也起作用，它描述了场到时空对称的非微分耦合。

⑥ 严格地说，这使我们超出我们拉格朗日场论的正统框架，因为 X 在 V 的不同点取的值并不独立于另外一个。

的一个解。这等价于说一个(X, Φ)对是一个解，当且仅当Φ是克莱因—戈登方程$\Box\Phi=0$的一个解，达朗贝尔算符$\Box$对应于V上度规$g := X^* g_0$。

这说明存在相对普通的理论（像克莱因—戈登场）能够给出强广义协变形式。因此甚至强广义协变也无法区分广义相对论和完美的一般性理论。

然而，很难撼动物理理论中广义相对论的特殊特征与其广义协变性有某种联系的感觉。实际上，有一点是显然的，即目前可以说最好的是，使广义相对论特殊的是其最自然最明白易懂的形式，即广义协变性。但是那仅仅是在说我们还是没有把事情搞清楚，我是这样认为的。

在此关联中，自然要问是不是在广义相对论里表示时间和变化时，我们所遇到的困难会出现，因为例子 43 里的人为地强的广义协变理论。

例子 44（人为广义协变性和时间问题）。让我们回到例子 43 的理论T_0和T，并且我们为了方便起见假设T_0的时空(V_0, g_0)并不承认任何等距，假设我们简单改动T，就会存在通过跟T相联系的辛空间上函数来表示可变量的各种方法吗？

设$\mathcal{S}$是T的解空间，并且设$\mathcal{S}'$是对应的约化空间（即$\mathcal{S}$的规范轨道空间）。跟人们可能期望的一样，两个解(X, Φ)和(X', Φ')处在$\mathcal{S}$的相同规范轨道里，当且仅当存在$d \in \mathcal{D}(V)$使得$X' = X \circ d$和$\Phi' = \Phi \circ d$。当然，空间$\mathcal{S}$是预辛的，而空间$\mathcal{S}'$是辛的。但是由于在同样规范轨道里的解跟Φ的值或者V的全部点上的X就不一致，很难把解的微分同胚等价类看作赋予属性给V的点，因此想把$\mathcal{S}'$看作对应于某些局域拉格朗日量的解空间显然是不可能的。使用那些类似用于上述广义相对论讨论里的步骤，我们能够建构T的初始数据空间$\mathcal{I}$以及相应的约化空间$\mathcal{I}'$，后者可能是辛空间。但是注意哈密顿量$\mathcal{I}$和$\mathcal{I}'$为零。因此，虽然我们能够建构辛空间，我们并不具有跟时间变换和时间演化相联系的非平凡流，那是我们要求在这些空间上通过函数建立我们变化表象的东西。迄今为止，目前的情况看上去很像空间上紧致广义相对论的情况。

但是现在注意单独从T的知识我们能够重构T_0，场X作为其目标空间具有流形V_0，我们取T_0为具有拉格朗日量L_0的V_0上标量场Ψ的理论：设Ψ是T_0的运动学上可能场，并且设X: $V \to V_0$为一个任意微分同胚，那么我们定义$L_0(\Psi)$为V_0上n形式，结果是我们用X^{-1}拖回V_0上n形式$L(X, \Psi \circ X)$，这个

结果独立于X选择。得到的运动方程是$\Box_0\Psi=0$。注意$\Box_0$作为在V_0上唯一度规g_0的达朗贝尔算符，以及场因果性地相对于g_0传播，很自然地我们把g_0视为V_0上的几何结构，并且继续考虑相对于g_0的片断，等等。

有了T_0，我们就能够建构解空间$\mathcal{S}_0$，相对于(V_0, g_0)的片断，我们能够通过在通常方式$\mathcal{S}_0$上的函数表示任何可变量，如标量场支持的体积。

最后，注意到$\mathcal{S}'$是正则辛同构于$\mathcal{S}_0$。① 我们能够把我们的变化表象从后一个空间变换到前一个空间，因此存在一种在此情况下避免时间问题的方法。②

不过，对此方法有个明显的担心。设$\tilde{g}_0$是V上不同于g_0的一个度规，那么$\Box_0$不是$\tilde{g}_0$的达朗贝尔算符，但是假设这样的操作通过$\tilde{g}_0$还是可以定义的，因此按照$\tilde{g}_0$，V_0上的L_0的欧拉—拉格朗日方程并不是克莱因—戈登方程，而是一些不太有名的方程。现在，上述策略是想把T作为同构于(V_0, g_0)的时空上的真正克莱因—戈登场的理论，但是它与基于T是克莱因—戈登理论的我们所得到的数据并不相关。因此，是什么东西阻止我们认为，T真正是服从某些在同构于$(V_0, \tilde{g}_0)$的时空上不太有名的场理论的？在此情况下我们能用$(V_0, \tilde{g}_0)$的片断建立我们的变化表征，等等。

这里有两件事人们会用来回应这个担心：(1)我们找到了通过在$\mathcal{S}'$上的函数来表示变化的自然方法；(2)我们一般要求物理上合理的理论，其场沿着时空度规的零锥传播。这仅仅对于共形等价于g_0的度规$\tilde{g}_0$的T_0来说是正确的。③ (V_0, g_0)的每一个片断也是$(V_0, \tilde{g}_0)$的片断，因为每一个$\tilde{g}_0$共形地跟g_0联系在一起(由于共形联系方法跟线是类时的和超曲面是类空的是一致的)。因此相对于这样一个片断，我们能够考虑一个量，它在我们片断里每一个$\Sigma \subset V$的意义上是共形不变的，这个量在$(V_0, \tilde{g}_0, \Psi)$里的Σ上具有同样的值，因为每一

---

① 通过把解T的等价类$[X, \Phi]$到T_0的解$\Psi=\Phi\circ X^{-1}$的映射。正是在这一点上我们要求假设g_0并不承认等距。一般情况下通过g_0的等距群作用$\mathcal{S}'$是同构于$\mathcal{S}_0$的商。

② 注意我们一定要选择(V_0, g_0)的一个片断，这是为了获得在对应于可变物理量的$\mathcal{S}'$上的函数族。这样的函数告诉我们，场的支持体积多大这样的事情是在空间几何假定了一个给定形式的瞬间上。

③ 回顾V上度规g_0和g_1是共形联系的，如果存在一个正标量Ω: $V\to\mathbb{R}$使得$g_1=\Omega g_0$。

个$\tilde{g}_0$共形地跟g_0联系在一起。这样一个量是由T的解的约化空间上的函数的相同单参数族来表示的，跟我们是否把T视为潜在地在同构于(V_0, g_0)时空上的克莱因—戈登场的理论，或者潜在地作为某些同构于$(V_0, \tilde{g}_0)$时空上的其他种类场的理论没有关系。

7.2 时间问题

在前面第5节和第6节考虑的每一个理论里，理论的动力学内容包含在理论的拉格朗日形式和哈密顿形式里的状态辛空间上的一个流(可能是时间相关的或可能仅仅是局域的)里。在广义相对论里无法理解成作为整个宇宙的理论，这一点使该理论与众不同。当然，这个特征意味着表示变化的标准策略对此理论也是失败的。由于人们并没有对应于在初始数据的约化空间上的时间演化的流，那个空间上就没有函数能够表示可变物理量，因此人们没有办法来要求通过在解的约化空间上的函数表示可变量。

这个联结是时间问题，因为时间在广义相对论里不是通过辛空间上的流来表示的，而变化也不是由瞬间或者整体状态空间上的函数来表示的。①

在对此问题的意义进行讨论之前，弄清楚它的性质和起源是十分重要的。

• 如果人们是通过关注从初始数据空间到初始数据约化空间的过渡来进入时间问题的，那么这个问题就显得特别急迫。因为在从初始数据空间到初始数据约化空间的过渡里，人们区分了初始数据集合，对应于在单个解范围内的柯西面的划分。看上去这牵涉对目前宇宙状态的处理和作为相同态的大爆炸之后其状态的处理。无疑地，按照广义相对论，变化是一种错觉。

但是这个说法太仓促，因为初始数据的约化空间肯定是正规同构于解的约化空间的。②并且在后一个空间，某些点代表了其中有变化的世界(如始于大爆

---

① 时间问题的正规表述是[Kuchař, 1992]和[Isham(艾沙姆), 1993]。关于哲学讨论，见[Belot and Earman, 2001]、[Butterfield and Isham, 2000]以及[Earman, 2002]。对于这个文献的批判反思，见[Maudlin, 2002]和[Healey, 2002]。

② 是在从$[q, \pi]$到$[g]$的映射下，如果(q, g)描述在某些(V, g)的柯西曲面上的瞬时态。

炸的世界)，而有些代表无变化的世界(如爱因斯坦静态解模型化的世界)。因此说广义相对论教给我们肯定的说法，这一点很难理解。

因而我想拒绝以这种说话方式对时间问题进行阐释，更加建设性地讲，我想指出，把注意力集中在解的约化空间而不是初始数据的约化空间来处置时间问题更加有用。在第5节表现好的理论里，初始数据空间和解空间是辛同构的，但是我们仍然认为这两个空间具有不同的表示功能——粗略和启发性地讲，一个适合表示可能的瞬间状态，而另一个适合表示可能的世界。这个区分是依据这样的事实，即相对于一个片断，人们发现对每一个 $t \in \mathbb{R}$，映射一个解到约化在瞬间 $\Sigma_t \subset V$ 上初始数据的映射 T_{Σ_t}，定义了一个在解空间跟初始数据空间之间的不同同构。这自然而然想到后一种空间的点代表了能够发生在不同时间的表示状态(普遍性)，并且认为解空间里的点代表这些态组成的可能世界。这个故事的基本要素一模一样地沿用第6节里各种复杂因素的引入。但是在宇宙学广义相对论的情况下，我们仅仅具有在初始数据约化空间跟解的约化空间之间的单独正则同构。在此情境中，很难否认解的约化空间和初始数据的约化空间在表象上等价。并且看上去很明显，我们应该把解的约化空间里的点解释为代表广义相对论世界而不是瞬间状态——所以我们应该说在初始数据约化空间里的点有相同的情况。因此，我们应该拒绝下面想法，即把约化过程看作是告诉了我们宇宙的早期状态和宇宙的一个后来状态是相同的瞬间状态。

• 由于我们集中在拉格朗日绘景而不是哈密顿绘景，但是却直接进入时间问题，因而我们能够得出结论：这个问题不是牵涉哈密顿方法里的 3+1 分解的人为产物。同样，时间问题作为宇宙学理论的广义相对论特性，而不是在空间无穷远处渐进平坦区域里广义相对论的特性，也不是在固定相对论背景上场论的特性，我认为也不是上述例子43和例子44里人为的强广义协变理论的特性。从这里我们能够总结出下面说法并非时间问题的充分条件：缺乏一个选定片断；可能片断的摇摆；在时空微分同胚群下理论的不变性。显然问题出现在我们采用了一个微分同胚不变性理论来模型化我们把几何取为完全动力学的情形(即我们并没有私下引入任何背景结构，无论是在空间无穷远处还是在其他任何地方)。

对于我至今所说的全部事情，时间问题听起来好像不过是有趣的智力测验

题。就算时间并不像在其早期理论里那样在广义相对论里(甚至在牵涉局域流的无穷小意义里)对称。当然，理论的诱人之处在于它以基本方式改变了时间的本质，并由于理论的成功应用牵涉可变物理量的表象(如水星近日点)，似乎有可能必定存在某些推广前面几节的图景去覆盖广义相对论的方法。而在其将会允许这种推广的研究可能反过来富有启发性的同时，它可能并不是一项很紧迫的事情。

这个问题在我们关注量子化时看上去好像更急迫些。好消息是，立足于约化，人们最终以表示广义相对论真正自由度的辛空间告终。没有沿着这些线的东西，量子化就不可能，但是宇宙学广义相对论哈密顿量的消失意味着两个当前的困难阻碍了广义相对论成功量子化之路。

1. 人们下一步怎么办？通常哈密顿量或者拉格朗日量在量子化里起到重要作用，人们通过这些东西定义量子动力学。在空间紧致广义相对论的情况下，初始数据的约化空间继承了初始数据原始空间的哈密顿量——其为零，使得相应的动力学是平凡的。而引力场的真正自由度的局域拉格朗日场论的说法并没有显得有任何意义，这种方法的前景不太明朗。

2. 人们如何使广义相对论的量子化有意义，这一点并不清楚。在经典理论中，人们只能发现解里面的变化，甚至不能够在动力学层次找到它(通过解空间上的量，等等)，这如何能够在量子水平上实现也不清楚。或许最好的情况是人们有望能够在近似经典解的量子态的子集里讨论近似时间和变化。那似乎是完全可以接受的——那是人们应该瞄准的，甚至在空间和时间的几何自己被量子化的理论里。但是，半经典近似的通常技巧要求哈密顿量。①

7.3 在广义相对论里寻找时间

这最后一节讨论什么可能是有关时间问题的最明显方法。在第5节和第6节讨论的情况里，我们能够表示变化，是通过理论解的(约化)空间上的函数，

---

① 因此，WKB方法的目的是建构量子哈密顿量的近似本征态。基于退相干和融贯态等基础上的分析，目的是要证明由量子哈密顿量驱动的动力学近似于相应经典系统之类的情况，见[Landsman，本书]。

因为我们具有一个片断σ：$S\times\mathbb{R}\rightarrow V$，它把时空分解成空间和时间，因此允许我们分辨对应于在不同瞬间所给量值解的(约化)空间上的函数。但是时空到空间和时间的解无关分解概念在广义相对论里没有意义，因为对于哪一个曲面算作类时以及哪一个超曲面算作类空来说，解是不一样的。有点奇怪的是，不用片断来建构广义相对论的哈密顿量方案是可能的。但是(不必惊讶)没有时空到瞬间的某种分解，问跟着一个给定状态的是哪个状态没有意义(因此也不存在有关哈密顿方法的真正动力学)，也无法建构一个对应于有用量的瞬间值解的约化空间的单参数族。因此寻找一个片断概念的代用品就是自然而然的事，它被应用到解的微分同胚等价类，而不是单个解上——并且希望这将导致时间和变化表象的常见解释。

贯穿整个最后一节，除非特别注明，我限定在只管具有零宇宙常数的四维时空维度里的空间上紧致的真空广义相对论。

让我们从一些定义开始。

定义 45(几何)。在广义相对论解的约化空间里的点称为几何，一种几何是作用在从V到自身的微分同胚群$\mathcal{D}(V)$的解空间上，我们把对应于g的解写成$[g]$，把$[g]$里的解说成是具有几何$[g]$的解。

定义 46(瞬时几何)。从抽象瞬间S到自身的微分同胚群$\mathcal{D}(S)$作用在初始数据空间上，我们称这个作用的轨道为瞬时几何。我们用$<q, \pi>$表示对应于初始数据点的几何，我们把初始数据(q, π)认为是具有几何$<q, \pi>$。[①]

定义 47(对于一个解的时间)。设(V, g)是一个解，对于(V, g)的一个时间是一个通过柯西曲面对(V, g)的一个分割$\{\Sigma\}$，称为时间的瞬间。一个参数时间是一个跟那套瞬间的特定参数化结合在一起的时间，一个仿射参数化时间是一个其瞬间是关于原点选择参数化的时间。[②]

---

① 注意到瞬时几何不是初始数据约化空间里的一个点：由一个不同柯西曲面上的给定解约化的初始数据对应于初始数据约化空间里同样的点，但是(一般)对应于瞬时几何空间里的不同点。

② 我们能够把(V, g)的时间，想成一个(V, g)的柯西曲面空间里的一个非参数化曲线；一个参数化时间是一个这种类型的参数化曲线；一个仿射参数化时间是一个这种类型的仿射参数化曲线。

定义 48(绝对时间)。设 g 是一个解，如果 g 的任何等距映射 $\{\Sigma\}$ 里的瞬间到 $\{\Sigma\}$ 里的瞬间，那么 (V, g) 的时间 $\{\Sigma\}$ 被称为绝对的。如果 g 的每一个等距满足前面的条件，并且保留每对瞬间之间的参数不同，那么一个 g 的仿射参数化时间被称为绝对的。如果 g 的每一个等距映射时间的每一个瞬间到自身，则 g 的一个参数化时间被称为绝对的。

任何整体性双曲线解承认一个参数化时间(因为每一个整体性双曲线解能够通过柯西曲面叶形化，它们都能够给予任意参数化)，但是并非总有可能找到具有庞大等距群解的绝对时间，闵可夫斯基时空就不承认一个绝对时间。①如果一个时空承认时间变换或者嵌入是对称的，那么它就不会承认一个绝对参数化时间。

定义 49(广义相对论时间)。一个(平的、仿射参数化的或者参数化的)广义相对论时间，是一个定义在解空间子集上的映射，它赋予每一个解在其定义域的那个解(平的、仿射或者参数化的)时间，并且以近似光滑的方式做到。

定义 50(广义相对论的几何时间)。一个(平的、仿射参数化的或者参数化的)广义相对论时间被称为几何的，如果它满足下面的条件：(i)其定义域在解空间上$\mathcal{D}(V)$的作用下是闭合的；(ii)如果 g 和 g'是在时间的定义域里，并且对于一些微分同胚 d：$V\to V$ 有 $g' = d^* g$，那么赋予 g'的叶形化是赋予 g 叶形化的 d^{-1}下的像(如果时间是仿射参数化的，那么我们要求这样一个 d 保留在任何两个瞬间之间的时间不同；如果时间是参数化的，那么我们要求这样一个 d 把 t 标记的瞬间映射到 t 标记的瞬间)。我通常会把“广义相对论几何时间”简化为“几何时间”。

注释 51(几何时间是绝对的)。广义相对论赋予解 g 的几何时间，这样的(平的、仿射参数化的、或者参数化的)时间总是绝对的。因为如果 d：$V\to V$ 是一个 g 的等距，那么上述定义中的条件(ii)告诉我们，d 保留的时间赋予了 g，并结合其参数化属性，如果有的话。因此闵可夫斯基时空不是在广义相对

① 在与一个给定参照系相联系的时间变换概念下的一个时间不变性，在相对于那个参照系加强的情况下就不是不变的。同样的论证在德西特(de Sitter)时空里起作用，或者在承认加强对称的其他时空里起作用，见[Moncrief，1992]。

论的任何几何化时间的定义范围内，并且在时间变换或者嵌入之下的解不变性，是在广义相对论的任何参数化几何时间的定义范围之内。

我们能够把相对论的(参数化、仿射参数化或者非参数化)几何化时间，看作是联系解的约化空间里几何[g]跟一个在瞬时几何空间里的(参数化、仿射参数化或者非参数化)曲线 $<q(t), \pi(t)>$ 的方式，我们称这样一个曲线为动力学轨迹。在几何和动力学轨迹之间的对应是以明显的方式建立的：设 g 是一个在给定几何时间的定义范围内的解，并且设($q(t)$, $\pi(t)$)是在初始数据空间里的(参数化、仿射参数化或者非参数化)曲线，它出现于我们考察在赋予 g 时间里的瞬间上的 g 所约化的初始数据时；$<q(t), \pi(t)>$ 是我们寻找的动力学轨迹。①如果 g_1 和 g_2 具有同样几何的解，那么它们是通过某些微分同胚 d：$V \to V$ 相联系的。在这种情况下，d 也与由我们的时间几何赋予它们的叶形相联系，因此 g_1 和 g_2 将对应于在瞬时几何空间里的同样的动力学轨迹。

大量有意义的几何时间的例子为人所知，大多数具有很小的定义范围：(i)在非旋转尘埃解种类里，几何时间是通过唯一的在任何地方正交于尘埃世界线超曲面的族叶形化给定的；(ii)在其等距群是具有类空轨道的三维的这类解范围之内，一个几何时间是通过其等距群的轨道叶形每一个解给定的。②更宽范围的例子很难得到，但是的确存在。

例子 52(CMC 时间)。回顾如果 $\Sigma \subset V$ 是一个(V, g)的柯西曲面，那么我们能够定义具有下面意义的张量 q^{ab} 和 Σ 上 k_{ab} 为：$q^{ab} := g^{ab}|_{\Sigma}$ 是 g 在 Σ 约化的黎曼度规，并且 $2k_{ab}$ 是这个度规按照其世界线正交嵌入 Σ 自由下落观测者的度规变化率。在这些张量之外，我们能够建构平均曲率 k：$\Sigma \to \mathbb{R}$，由 $k := q^{ab}k_{ab}$ 定义，因此 $k(x)$ 仅仅是包含在 x 上的 k_{ab} 的相关信息的矩阵轨迹。如果 k 是一个 Σ

---

① 严格讲，为了建构初始数据空间里的曲线($q(t)$, $\pi(t)$)，我们需要引入(V, g)的片断，其瞬间与给定时间的瞬间一致，使得我们能够把具体瞬间上的态拖到我们抽象瞬间 S 上的态；在我们用 $\mathcal{D}(S)$ 的作用除去初始数据空间达到瞬时态空间时，片断选择任意性消除。

② 方案(i)推广爱因斯坦同时性约定到尘埃宇宙学的环境，见[Sachs and Wu(萨克斯和吴)，1977，§5.3]。注意方案(i)和(ii)不需要在其共同的定义域内一致，见[King and Ellis(金和埃利斯)，1973]。

上的不变函数，一个解 g 的柯西曲面 $\Sigma\subset V$ 被称为恒定平均曲率的曲面，或者简单地叫作 CMC 曲面。回顾一下，除非特别说明，我们限制在具有零宇宙常数的(3+1)空间紧致整体性双曲线真空解范围之内。

适用性。人们普遍认为，一大类爱因斯坦场方程的解，都能够通过 CMC 曲面叶形化。

1)众所周知，这样包含一个 CMC 片断的解，是一个在解空间里的开集。①

2)一度猜想，所有解至少包含一个 CMC 曲面，但是现在知道情况并非如此。②

3)一度猜想，所有承认一个 CMC 片断的解能够由这些片断叶形化。③ 现在相信这种情况仅仅限于某些空间拓扑。④

4)人们相信，在由 CMC 片断叶形化的这类解之内，一个给定空间拓扑的所有解，将会展现不变平均曲率值的相同范围，静态解是唯一的例外(回顾一下，一个解如果它承认类时基灵场是静态的——粗略地讲，一个时间变换群的无穷小产生子)。⑤

---

① 如[Isenberg and Marsden，1982，195]。

② 见[Bartnik and Isenberyg(巴特尼克和伊森伯格)，2004，32]或者[Chruściel *et al.*(克鲁赛尔等)，2005，Corollary 1.3]。相应的空间紧致尘埃解猜测也是错的，见[Bartnik，1988]。

③ 对于推测的原始形式，参见[Isenberg and Marsden，1982，Conjecture 3.2]。这个猜测已经知道对于平直时空是对的[Barbot(巴布特)，2005，§12]和相应的猜测在(2+1)情况下已经知道是对的[Andersson *et al.*，1997]。这个猜测的相应情况对空间上紧致的尘埃解已经知道是错的[Isenberg and Rendall(伊森伯格和伦德尔)，1998]以及在非渐进平坦真空情况下也如此，其中史瓦西(Schwartzchild)解提供了一个反例[Eardley and Smarr(厄德利和斯马)，1979，§Ⅲ]。

④ 对于目前的推测，见[Rendall，1996，Conjectures 1 and 2]。现在相信，对于某些空间拓扑学，类似于史瓦西解的行为能够出现，见[Rendall，1996]和[Andersson，2004，81]。在(3+1)情况下，修正过的推测已经知道对于某些类型的高度对称解是正确的，甚至有些物质形式也是允许的，见[Rendall，1996，Theorems 1 and 2]、[Andersson，2004，81 f. and 95]以及其中的参考文献。

⑤ 对此，见[Rendall，1996，Conjectures 1 and 2]。对于在高度对称情况下的情形和(2+1)维度的情形，见前面两个注释的参考文献。

不变性特征。CMC 叶形化在等距之下表现得特别好。①

设(V, g)是一个解，$\{\Sigma\}$是叶形化 V 的 CMC 曲面的一个集合，并且 d: $V \to V$ 是 g 的一个等距，那么 d 保留叶形$\{\Sigma\}$的不变性。②如果(V, g)是非静态的，那么：(a)g 的任何对称 d 保留$\{\Sigma\}$里的每一个叶；(b)对于任何实数 κ，最多存在一个具有不变平均曲率 κ 的一个柯西曲面。如果(V, g)是静止的，那么：g 是平坦的，并且任何在(V, g)里的 CMC 曲面具有为零的平均曲率。③

CMC 时间。由其 CMC 片断叶形化每一个解，一旦可能的话，在我们所考虑解的种类范围之内我们能够决定一个几何时间。我们能够提出一个仿射参数化几何时间如下：对于非静态解，参数在平均曲率 κ_1 和 κ_2 之间的差异是 $|\kappa_2 - \kappa_1|$；对于静态解，参数在两个片断之间的差异是流逝在这些片断之间的本征时间。如果我们限制在非静态解范围内，并且赋予每一个片断由其平均曲率给定的参数值，那么我们得到参数化的几何时间。

例子 53(宇宙时间)。给定一个解(V, g)，对于 g 的宇宙时间函数是映射 τ: $V \to \mathbf{R} \cup \{\infty\}$，赋予每一个 $x \in V$ 从 x 出发的全部向后因果曲率长度上的上确界。显然，存在很多表现好的空间，其中的 $\tau(x)$ 表现不好，比如，在闵可夫斯基时空里，对于所有事件有 $\tau(x) = \infty$。我们认为解的宇宙时间函数是规则的，如果：(a)对于所有的 x，$\tau(x) < \infty$；(b)沿着每一个向后不可延伸因果曲线有 $\tau \to 0$。如果 τ 是一个在(V, g)上的正规宇宙时间函数，那么：(i)(V, g)是整体性双曲线的；(ii)τ 是一个通常意义上解的时间函数(即它是连续的并且沿着未来方向因果曲线严格增加的)，并且(iii)τ 的水平曲率是(V, g)里的未来柯西曲率(即这些曲率具有空的未来柯西界面)。④在空间上紧致真空$(2+1)$维广义相对论里，已经知道：(a)宇宙时间对于几乎全部时空拓扑学都是规则

① 见[Isenberg and Marsden, 1982, §3]。

② 对于承认加速对称的时空可能不成立，例如闵可夫斯基时空和德西特时空(由于我们要求零宇宙常数，德西特时空就目前的目的来说并不算是空间上紧致的真空解)。注意在非渐进平坦情况下，CMC 片断不变性的问题牵涉更多，见[Bartnik *et al.*, 1990, §5]。

③ 当然，一般一个类时基灵矢量并不保证平坦，但是它在目前考虑的经典解内保证平坦。

④ 见[Andersson *et al.*, 1998, Propositions 2.2 and 2.5 and Corollary 2.6]。

的；(b)在重要的一类解中，宇宙时间跟 CMC 时间是一致的。[①]在其水平曲面是柯西曲面的规则宇宙时间函数的这类时空上，我们通过由恒定宇宙时间曲面对每一个解叶形化来建构广义相对论几何时间；通过宇宙时间函数在每一个叶上取值产生一个广义相对论参数化几何时间来参数化这些叶形，只要我们排除具有时间反射对称的解就行。

广义相对论几何时间，实际上是一种相对于其他被视为演化方法而言的分离理论无穷多变量的方法，其办法是允许我们从解的约化空间里的一点过渡到空间瞬时几何里的(可能非参数化的)曲线。诸如宇宙体积、恒星数目及星系大小这样的瞬时物理量，能够通过瞬时几何空间上实值函数来表示。由此几何时间的选择允许我们以熟悉的方式谈论变化：我们能够核实是否表示相关量的瞬时几何空间上的函数，在对应于出现在给定时空几何里的不同瞬时几何点上取不同的值。如果我们广义相对论的几何时间是参数化的，我们能够计算相关量的变化率(由于我们具有仿射参数化曲线，通过的是对应于给定空间几何的瞬时几何空间)。如果我们有一个参数化的广义相对论几何时间，我们甚至能够模仿我们在前面几节的建构，具体通过理论解的约化空间上的单参数族来表示可变化量：给定在表示我们相关量的瞬时几何空间上的函数 f 以及实数 t，我们通过设 $f_t[g]$ 等于 f 在对应于 $[g]$ 里 t 的瞬时几何上所取的值，来定义解的约化空间上部分定义了的函数 f_t。

跟上面定义的一样，广义相对论里的时间问题有两个主要方面：

1. 在空间上紧致的广义相对论里，时间不通过一个态的辛空间上的流来表示，就像它在早期理论那样；

2. 变化不通过对应于可能瞬时态和世界的空间的辛空间上函数来表示，就像在早期理论里那样。

我们现在看到，如果我们进一步引入一个参数化的几何时间，我们就能够按照通常的方式，通过在解的约化空间上函数的单参数族来解决表示一个可变

① 见[Benedetti and Guadagnini(伯纳黛特和瓜达尼尼)，2001]。当然，一般来说，恒定的宇宙时间的曲面没有 CMC 曲面光滑，因此两个时间概念并不一致，见[Benedetti and Guadagnini，2001，331]或者[Barbot and Zeghib(巴布特和泽吉卜)，2004，§5.4.1]。

量的第二个担忧。

引入几何时间就足以解决第一个担忧吗？任何几何时间都是选择瞬时几何空间的一个子空间，由那些作为柯西曲面瞬时几何出现的 $<q, \pi>$ 构成，它们的出现是通过柯西曲面的瞬时几何挑出来的，靠的是几何时间——比如只有瞬时几何，把空间描述为具有恒定平均曲率的几何时间——能够按照 CMC 片断方案出现。如果我们引入一个仿射参数化广义相对论几何时间，那么我们的确得到一个瞬时几何空间上的流，其出现依照的是这个几何时间（因为这个空间是通过仿射参数化动力学轨迹来划分的，这样的轨迹对应于在给定几何时间的定义范围内的几何）。但是人们并不期望这个空间成为辛的，也不期望会同构于解的约化空间。[①]因此跟我们几何时间相联系的瞬时几何空间上的流，并不是辛空间上的流。并且由于每一个瞬时几何空间上的动力学轨迹对应于解的约化空间里的单个点，我们没有办法把我们在前一种空间上的流传到后一种上的非平凡流。

一个通过在辛流形上的流建立时间表象的自然策略，是想把广义相对论几何时间的选择变成一个非平凡的（但是可能是时间相关的）哈密顿系统的重建。在一个重要的情况下，人们已经知道它可以被实现。

例子 54（CMC 动力学）。我们考虑在上述例子 52 里引入的 CMC 时间。[②]我们对我们的抽象瞬间 S 的拓扑学加以限制。[③]设$\mathcal{M}$是在 S 上具有恒定标量曲率 -1 的黎曼度规空间。[④]余切丛 $T^*\mathcal{M}$是一个辛空间；$T^*\mathcal{M}$的一个元是(γ, p)形式

---

① 直观上说，由给定几何时间产生的瞬时几何空间能够被看成是解的约化空间，具有实线的积（因为每一个几何对应于相对于几何时间的瞬时几何的单参数族）。因此给定几何时间的瞬时几何空间，并不同构于解的空间——它也不能是辛的，因为它是辛空间跟一个奇数维空间的积。

② 对于（3+1）情况的概述，见菲舍尔和莫尼克里夫（Fischer and Moncrief）相关的出版物。详情见[Fischer and Moncrief, 1996]、[Fischer and Moncrief, 1997]以及其中的参考文献。对于（2+1）情况见[Moncrief, 1989]和[Andersson *et al.*, 1997]。下面描述的建构是去参数化的例子，对于这个概念和有限维的应用，见[Beig, 1994, §2]。

③ 我们加两个条件：（i）S 一定是 Yamabe 类型 -1，即唯一 S 承认的恒定的标量曲率黎曼度规具有负标量曲率，这是所引论文采用的结构的重要依据；（ii）S 肯定不承认任何具有正维等距群的黎曼度规，这使得我们免于对奇异商空间的担心。

④ 因为 S 是 Yamabe 类型 1，S 上的任何黎曼度规共形等价于$\mathcal{M}$里的度规。

的，其中 $\gamma\in\mathcal{M}$，而 p 是 S 上二阶对称逆变张量密度，按照 γ 那是无散度并且无迹的。如果存在一个微分同胚 d：$S\to S$ 使得 $(\gamma',\ p')=(d^*\gamma,\ d^*p)$，我们认为 $(\gamma,\ p)$，$(\gamma',\ p')\in T^*\mathcal{M}$ 是等价的。我们用这个等价关系相除时得到空间 $\mathcal{I}^*\coloneqq T^*\mathcal{M}/\mathcal{D}(S)$，该空间从 $T^*\mathcal{M}$ 继承了辛结构。我们称 $\mathcal{I}^*$ 里的点为共形初始数据，对于每一个 $t<0$，存在一个在 $\mathcal{I}^*$ 跟具有恒定平均曲率 t 之间的几何上自然辛同构，①并且存在一个在后者集合跟广义相对论解的约化空间（在其下面，一个瞬时态被送到产生于其中的唯一几何）之间的自然辛同构。因此，对于 $t<0$，我们具有一个在共形初始数据跟解的约化空间之间的辛同构。

相反，给定一个几何 $[g]$ 和时间 $t<0$，我们能够寻找 $\mathcal{I}^*$ 里的点，它按照由 t 标记的同构对应于 $[g]$。对每一个 $t<0$，这么做给我们一条对应于 $[g]$ 的 $\mathcal{I}^*$ 里曲线，在 $\mathcal{I}^*$ 里的一般点会处在很多这样的轨迹上。一般来说，如果 $x\in\mathcal{I}^*$，并且 $t_1\neq t_2$，那么，对应于 $(x,\ t_1)$ 的恒定平均曲率 t_1 的瞬时几何和对应于 $(x,\ t_2)$ 的恒定平均曲率 t_2，就会属于不同的时空几何。如果我们看看对应于在解的约化空间里的全部几何的 $\mathcal{I}^*$ 里的轨迹的完备族，那么我们就会发现，这些都是由 $\mathcal{I}^*$ 的辛结构跟时间相关的哈密顿量 $h(t)$ 结合在一起产生的，后者是 t 的和空间体积的简单函数。②

拿这个例子作为模型，我们引入了跟理论的给定参数化几何时间结合在一起的广义相对论哈密顿量化的概念。假定我们给出了这样一个参数化几何时间，进一步假定我们能够建构一个其点是在抽象瞬间 S 上的张量的辛空间 $\mathcal{I}^*$，并且对于 t 的每一个值，我们能够建构在 $\mathcal{I}^*$ 跟瞬时几何之间的一个几何上的自然同构，后者是按照我们参数化几何时间对应于 t 的。由从瞬时几何空间到解的约化空间的正则映射组成的同构，给了我们一个 $\mathcal{I}^*$ 跟 S' 之间的辛同构的单参

---

① 让我们暂且忽略 $\mathcal{D}(S)$ 对称，给定一个 $(\gamma,\ p)$ 对，并且时间 $t<0$，存在 S 上唯一正标量 ϕ，解 $(\gamma,\ p,\ t)$ 的利克尔诺维奇（Lichernowicz）方程，$\Delta_\gamma\phi-\frac{1}{8}\phi+\frac{1}{12}t^2\phi^5-\frac{1}{8}(p\cdot p)\mu^{-2}\phi^{-7}=0$（这里 Δ_γ 是 γ 的拉普拉斯算符，并且 μ 是 γ 的体积形式）。我们希望 $(q,\ \pi)$ 是由 $q\coloneqq\phi^4\gamma$ 和 $\pi\coloneqq\phi^{-4}p+\frac{2}{3}t\phi^2\gamma^{-1}$ 给定的。

② 空间体积自己是一个 $\mathcal{I}^*$ 上 t 相关函数，因为同样的共形数据在由 t 的不同值增补时，会导致具有不同体积的瞬时几何。

数族。[1]这允许我们把每一个几何[g]跟$\mathcal{I}^*$里曲线$x(t)$结合在一起：对于每一个t，$x(t)$是在$\mathcal{I}^*$的点，它通过由t标记的同构获得到[g]的映射，我们称$x(t)$为跟[g]相联系的动力学轨迹。我们现在考虑以这种方式出现的那类$\mathcal{I}^*$上的动力学轨迹，并且问是否存在一个(可能是时间相关的)$\mathcal{I}^*$上哈密顿量，用$\mathcal{I}^*$的辛结构具体产生它们。如果存在，那么最终的(可能是时间相关的)哈密顿系统，是一个建立在给定参数化几何时间基础上的广义相对论哈密顿量化。

正如我们已经看到的，给定一个参数化的广义相对论几何时间，我们能够通过在解的约化空间上的函数单参数族以熟悉的方式表示可变量。并且，如果我们进一步引入理论的相关哈密顿量化，那么我们能够以一个在辛空间$\mathcal{I}^*$上的(可能是时间相关的)哈密顿流这种熟悉的方式表示时间，我们可以将其点看作是在不同的时候产生的初始数据。因此，这些概念允许我们在固定背景时空里考虑理论时，通过跟片断在第 5 节和第 6 节里取到的同样作用绕开时间问题。

几何时间的引入或者一个相关哈密顿量化的引入会违背广义协变性吗？某个意义上不违背——因为这些概念被置于解的约化空间水平，从而不能区别对待微分同胚解。

但是引入几何时间违背广义相对论精神这一点是真的，就像今天对该理论的一般理解那样——多数人可能喜欢把狭义相对论想成消除了时间和空间之间的任何细微区别，并且认为广义相对论是对狭义相对论的推广，且以一种根本没有恢复这种区分的方式。[2]

不过，需要注意这真的是对一个几何时间比其他的要特殊的异议。通过其每一个哈密顿量化把理论内容想成可以被阐释的，以及通过全部哈密顿量化的集合把理论内容视为已被穷尽了(那即是说，如果我们忽视了具有时间变换或者反射对称的时空)，这似乎完全处在广义相对论的精神之中。

---

① 严格讲，如果由时间参数取值的范围从几何到几何变化，那么同构纯粹是局域性的(正如在 6.1 节那样)。

② 另一方面，很多早期相对论宇宙学家，乐于通过正交于尘埃世界线的曲面取非旋转尘埃宇宙学的自然叶形为区分空间和时间的记号，它们在爱因斯坦电磁理论的解释里为零，在天文学层面重新恢复。见[Belot，2005，§3.2]的讨论和参考文献。

还有，自然要问什么样的考虑可能导致我们认为几何时间或者相关哈密顿量化是正确的那个。①

经典考虑。在上述例子54里概述的CMC哈密顿量化里，广义相对论被重建为时间相关系统。这有点令人不安：我们习惯于认为时间相关哈密顿量只有在一个开放系统遭受外力时才会出现。因此在一个基础范围内遇到一个时间相关的哈密顿系统就令人惊讶。或许这是一个我们不得不学会的东西，即我们这里正在找出广义相对论无穷多的变量之一，并且把它作为时间对待——并且我们期望的是在广义相对论变量之间的全部复杂非线性相互作用。不过，一些有趣的特殊情况被理解成几何时间，它们导致广义相对论的时间与哈密顿量化无关。②

因此我们不能排除这样的可能性，即可能存在广泛的几何时间，允许我们把广义相对论改写为具有非平凡动力学的时间无关哈密顿量理论。③显然这样一个几何时间的建构可能是第一个收获：有可能我们碰巧处在正确时间上，之前被我们所用理论的模糊形式所掩盖——很像如果经典力学最先所给的时间重新参数化不变性公式可能具有的那样，而后来也发现某个参数化族允许方程写成很简单的形式。

量子考虑。是否特殊对待几何时间或者平等对待它们，这个问题有望对量子化有影响(这个计划提供了寻求起先具有非平凡动力学的广义相对论哈密顿形式的主要动机)。因为人们肯定不期望对应于几何时间的不同选择，广义相对论的不同哈密顿形式会具有等价的量子化——至少在等价量子化要求是幺正

① 注意时间哲学里的某些方法和量子力学解释的某些方法，显然要求像通过时间瞬间选定的时空叶形之类的东西。

② CMC时间发生在具有双环拓扑学的空间，见[Moncrief, 1989, 2913]。广义相对论也能够耦合到完美的流——在这种情况下，导出动力学的守恒量是总重子数，见[Moncrief, 1977]和[Moncrief and Demaret, 1980]。

③ 注意给定一个$\mathcal{I}^*$上的非平凡的时间无关哈密顿量，我们能够使用我们的$\mathcal{I}^*$跟$\mathcal{S}'$之间的t相关同构来建构一个$\mathcal{S}'$上的(可能时间相关)哈密顿量。除非后者函数是常数，它可能产生一个解约化空间上的非平凡流。当然，这不应该解释成时间变换，虽然是由产生时间演化的时间无关哈密顿量的对应物产生的。

等价时就没有。①

因此什么是我们能够希望的？长远来说就像下面说的：（1）或许只有一个几何时间会导致一个经验上适当的引力量子理论；（2）或许会有一个几何时间的自然类（如导致时间无关哈密顿量的那种），能够被视为对等价量子理论的保证（或许是在"等价"的自由化意义上）。

似乎更可信的是，量子化广义相对论中困难的解决可能完全在不同的方向，但是弄清楚时间困难的本质在任何意义上都是值得的。

致谢

我要感谢的是：感谢本书编者的邀请及其明智的劝告，以及他们极大的耐性；感谢在蒙特利尔、纽约以及匹兹堡倾听这个冗长材料的听众；感谢文森特·莫尼克里夫和查理斯·托尔有益的信件；感谢理查德·托马斯(Richard Thomas)没有作用的信件；还要感谢弗兰克·阿恩特策留斯(Frank Arntzenius)、慈安·多尔(Cian Dorr)、亚历山大·瓜依(Alexandre Guay)、劳拉·鲁切(Laura Ruetsche)有益的交谈。

参考文献

[Abraham and Marsden, 1978] R. Abraham and J. Marsden. *Foundations of Mechanics.* Perseus Books, Cambridge, MA, 2nd edition, 1978.

[Abraham *et al.*, 1988] R. Abraham, J. Marsden, and T. Ratiu. *Manifolds, Tensor Analysis, and Applications.* Springer-Verlag, Berlin, 2nd edition, 1988.

[Adams *et al.*, 1985] M. Adams, T. Ratiu, and R. Schmid. The Lie group structure of

① 斯通—冯·诺伊曼(von Neumann)定理保证了，除非他们中的一个有古怪，否则两个人开始量子化一个具有线性、有限维相位空间的哈密顿量理论，最终会具有幺正等价的量子化。但是，只要有人考虑无限维或者非线性相位空间，情况就马上发生改变。比如，在经典层面看上去等价的东西会导致不同的量子理论。见鲁切相关的论述和[Gotay, 2000]的讨论、例子及其参考文献。[Gotay and Demaret, 1983]关于最小超空间宇宙学模型，承认导致物理上不同量子化的不同去参数化。

diffeomorphism groups and invertible Fourier operators, with applications. In V. Kac (ed.), *Infinite-Dimensional Groups with Applications*, pages 1-69. Springer-Verlag, Berlin, 1985.

[Ali and Engliš, 2005] T. Ali and M. Engliš. Quantization methods: A guide for physicists and analysts. *Reviews in Mathematical Physics*, 17: 391-490, 2005.

[Anderson and Duchamp, 1980] I. Anderson and T. Duchamp. On the existence of global variational principles. *American Journal of Mathematics*, 102: 781-868, 1980.

[Anderson and Thompson, 1992] I. Anderson and G. Thompson. *The Inverse Problem of the Calculus of Variations for Ordinary Differential Equations*. American Mathematical Society, Providence, RI, 1992.

[Andersson *et al.*, 1997] L. Andersson, V. Moncrief, and A. Tromba. On the global evolution problem in 2 + 1 gravity. *Journal of Geometry and Physics*, 23: 191-205, 1997.

[Andersson *et al.*, 1998] L. Andersson, G. Galloway, and R. Howard. The cosmological time function. *Classical and Quantum Gravity*, 15: 309-322, 1998.

[Andersson, 1987] L. Andersson. Momenta and reduction in general relativity. *Journal of Geometry and Physics*, 4: 289-314, 1987.

[Andersson, 1989] L. Andersson. Momenta and reduction in general relativity. II. The level sets. In R. Bartnik (ed.), *Conference of Mathematical Relativity*, pages 73-88. Australian National University, Canberra, 1989.

[Andersson, 2004] L. Andersson. The global existence problem in general relativity. In P. Chruściel and H. Friedrich (eds.), *The Einstein Equations and the Large Scale Behavior of Gravitational Fields*, pages 71-120. Birkhaüser, Basel, 2004.

[Arms, 1981] J. Arms. The structure of the solution set for the Yang-Mills equations. *Mathematical Proceedings of the Cambridge Philosophical Society*, 90: 361-372, 1981.

[Arnold *et al.*, 1997] V. I. Arnold, V. V. Kozlov, and A. I. Neistadt. *Mathematical Aspects of Classical and Celestial Mechanics*. Springer-Verlag, Berlin, 2nd edition, 1997.

[Arnold, 1989] V. I. Arnold. *Mathematical Methods of Classical Mechanics*. Springer-Verlag, Berlin, 2nd edition, 1989.

[Ashtekar and Romano, 1992] A. Ashtekar and J. Romano. Spatial infinity as a boundary of spacetime. *Classical and Quantum Gravity*, 9: 1069-1100, 1992.

[Ashtekar *et al.*, 1991] A. Ashtekar, L. Bombelli, and O. Reula. The covariant phase space of asymptotically flat gravitational fields. In M. Francaviglia (ed.), *Mechanics, Analysis, and Geometry: 200 Years After Lagrange*, pages 417-450. Elsevier, Amsterdam, 1991.

[Barbot and Zeghib, 2004] T. Barbot and A. Zeghib. Group actions on Lorentz spaces, mathematical aspects: A survey. In P. Chruściel and H. Friedrich (eds.), *The Einstein Equation and the Large Scale Behavior of Gravitational Fields*, pages 401-439. Birkhaüser, Basel, 2004.

[Barbot, 2005] T. Barbot. Flat globally hyperbolic spacetimes. *Journal of Geometry and Physics*, 53: 123-165, 2005.

[Bartnik and Isenberg, 2004] R. Bartnik and J. Isenberg. The constraint equations. In P. Chruściel and H. Friedrich (eds.), *The Einstein Equations and the Large Scale Behavior of Gravitational Fields*, pages 1-38. Birkhaüser, Basel, 2004.

[Bartnik *et al.*, 1990] R. Bartnik, P. Chruściel, and N. Ó Murchadha. On maximal surfaces in asymptotically flat spacetimes. *Communications in Mathematical Physics*, 130: 95-109, 1990.

[Bartnik, 1988] R. Bartnik. Remarks on cosmological spacetimes and constant mean curvature surfaces. *Communications in Mathematical Physics*, 117: 615-624, 1988.

[Beig and Ó Murchadha, 1987] R. Beig and N. Ó Murchadha. The Poincaré group as the symmetry group of canonical general relativity. *Annals of Physics*, 174: 463-498, 1987.

[Beig and Schmidt, 1982] R. Beig and B. Schmidt. Einstein's equations near spatial infinity. *Communications in Mathematical Physics*, 87: 65-80, 1982.

[Beig, 1994] R. Beig. The classical theory of canonical general relativity. In J. Ehlers and H. Friedrich (eds.), *Canonical Gravity: From Classical to Quantum*, 59-80. Springer-Verlag, Berlin, 1994.

[Belot and Earman, 2001] G. Belot and J. Earman. Pre-Socratic quantum gravity. In C. Callender and N. Huggett (eds.), *Physics Meets Philosophy and the Planck Scale*, pages 213-255. Cambridge University Press, Cambridge, 2001.

[Belot, 2003] G. Belot. Symmetry and gauge freedom. *Studies in History and Philosophy of Modern Physics*, 34: 189-225, 2003.

[Belot, 2005] G. Belot. Dust, time, and symmetry. *British Journal for the Philosophy of*

Science, 55: 255-291, 2005.

[Benedetti and Guadagnini, 2001] R. Benedetti and E. Guadagnini. Cosmological time in (2+1)-gravity. *Nuclear Physics B*, 613: 330-352, 2001.

[Bluman, 2005] G. Bluman. Connections between symmetries and conservation laws. *Symmetry, Integrability and Geometry: Methods and Applications*, 1: 011, 2005.

[Brading and Castellani, this volume] K. Brading and E. Castellani. Symmetries and invariances in classical physics. Chapter 13, this volume.

[Budic*et al.*, 1978] R. Budic, J. Isenberg, L. Lindblom, and P. Yasskin. On the determination of Cauchy surfaces from intrinsic properties. *Communications in Mathematical Physics*, 61: 87-95, 1978.

[Butterfield and Isham, 2000] J. Butterfield and C. Isham. On the emergence of time in quantum gravity. In J. Butterfield (ed.), *The Arguments of Time*, pages 111-168. Oxford University Press, Oxford, 2000.

[Butterfield, this volume] J. Butterfield. On symplectic reduction in mechanics. Chapter 1, this volume.

[Butterfield, Unpublished] J. Butterfield. Against*Pointillisme* about mechanics. Unpublished.

[Cannas da Silva, 2006] A. Cannas da Silva. Symplectic geometry. In F. Dillen and L. Verstraelen, eds. *Handbook of Differential Geometry, Volume II*, pp. 79-188, North-Holland, 2006.

[Cantor, 1979] M. Cantor. Some problems of global analysis on asymptotically simple manifolds. *Compositio Mathematica*, 38: 3-35, 1979.

[Choquet-Bruhat and DeWitt-Morrette, 1989] Y. Choquet-Bruhat and C. DeWitt-Morrette. *Analysis, Manifolds, and Physics. Part II*: 92 *Applications.* North-Holland, Amsterdam, 1989.

[Choquet-Bruhat*et al.*, 1977] Y. Choquet-Bruhat, C. DeWitt-Morrette, and M. Dillard-Bleick. *Analysis, Manifolds, and Physics.* North-Holland, Amsterdam, 1977.

[Chruściel *et al.*, 2005] J. Chruściel, J. Isenberg and D. Pollack. Initial data engineering. *Communications in Mathematial Physics*, 257: 29-42, 2005.

[Crampin and Prince, 1988] M. Crampin and G. E. Prince. Alternative Lagrangians for spherically symmetric potentials. *Journal of Mathematical Physics*, 29: 1551-1554, 1988.

[Crnković and Witten, 1987] Č. Crnković and E. Witten. Covariant description of canonical formalism in geometrical theories. In S. Hawking and W. Israel (eds.), *Three Hundred Years of Gravitation*, pages 676-684. Cambridge University Press, Cambridge, 1987.

[Cushman and Bates, 1997] R. Cushman and L. Bates. *Global Aspects of Classical Integrable Systems*. Birkhaüser, Basel, 1997.

[Deligne and Freed, 1999] P. Deligne and D. Freed. Classical field theory. In P. Deligne, P. Etingof, D. Freed, L. Jeffrey, D. Kazhdan, J. Morgan, D. Morrison, and E. Witten (eds.), *Quantum Fields and Strings: A Course for Mathematicians*, volume 1, pages 137-225. American Mathematical Society, Providence, RI, 1999.

[Diacu and Holmes, 1996] F. Diacu and P. Holmes. *Celestial Encounters*. Princeton University Press, Princeton, 1996.

[Diacu, 1992] F. Diacu. *Singularities of the N-Body Problem*. Les Publications CRM, Montréal, 1992.

[Diacu, 2002] F. Diacu. Singularities of the *N-body problem*. In H. Cabral and F. Diacu (eds.), *Classical and Celestial Mechanics: The Recife Lectures*, pages 35-62. Princeton University Press, Princeton, 2002.

[Dirac, 2001] P. A. M. Dirac. *Lectures on Quantum Mechanics*. Dover, New York, 2001.

[Dubrovin *et al.*, 1992] B. A. Dubrovin, A. T. Fomenko, and S. P. Novikov. *Modern Geometry — Methods and Applications. Part I. The Geometry of Surfaces, Transformation Groups, and Fields*. Springer-Verlag, Berlin, 2nd edition, 1992.

[Duval *et al.*, 1990] C. Duval, J. Elhadad, M. Gotay, and G. Tuynman. Nonunimodularity and the quantization of the pseudo-rigid body. In J. Harnad and J. Marsden (eds.), *Hamiltonian Systems, Transformation Groups, and Spectral Transform Methods*, pages 149-160. Les Publications CRM, Montréal, 1990.

[Eardley and Smarr, 1979] D. Eardley and L. Smarr. Time functions in numerical relativity: Marginally bound dust collapse. *Physical Review D*, 19: 2239-2259, 1979.

[Earman, 1986] J. Earman. *A Primer on Determinism*. Reidel, Dordrecht, 1986.

[Earman, 2002] J. Earman. Thoroughly modern McTaggart, or what McTaggart would have said if he had read the general theory of relativity. *Philosophers' Imprint*, 2(3): 1-28, 2002.

[Earman, 2003] J. Earman. Tracking down gauge. In K. Brading and E. Castellani (eds.), *Symmetries in Physics*, pages 140-162. Cambridge University Press, Cambridge, 2003.

[Earman, this volume] J. Earman. Aspects of determinism in modern physics. Chapter 15, this volume.

[Earman, 2006] J. Earman. Two challenges to the requirement of substantive general covariance. *Synthese*, 142: 443-468, 2006.

[Eichhorn, 1993] J. Eichhorn. The manifold structure of maps between open manifolds. *Annals of Global Analysis and Geometry*, 11: 253-300, 1993.

[Fischer and Moncrief, 1996] A. Fischer and V. Moncrief. A method of reduction of Einstein's equations of evolution and a natural symplectic structure on the space of gravitational degrees of freedom. *General Relativity and Gravitation*, 28: 207-219, 1996.

[Fischer and Moncrief, 1997] A. Fischer and V. Moncrief. Hamiltonian reduction of Einstein's equations of general relativity. *Nuclear Physics B (Proceedings Supplement)*, 57: 142-161, 1997.

[Fischer and Moncrief, Unpublished] A. Fischer and V. Moncrief. Nonautonomous dynamicalsystems and the phase portrait of the reduced einstein equations. Unpublished.

[Frauendiener and Sparling, 1992] J. Frauendiener and G. Sparling. On the symplectic formalism for general relativity. *Proceedings of the Royal Society. London. Series A*, 436: 141-153, 1992.

[Friedman, 1983] M. Friedman. *Foundations of Spacetime Theories.* Princeton University Press, Princeton, 1983.

[Gerver, 2003] J. Gerver. Noncollision singularities: Do four bodies suffice? *Experimental Mathematics*, 12: 187-198, 2003.

[Goldstein, 1953] H. Goldstein. *Classical Mechanics.* Addison-Wesley, Cambridge, MA, 1953.

[Gotay and Demaret, 1983] M. Gotay and J. Demaret. Quantum cosmological singularities. *Physical Review D*, 28: 2404-2413, 1983.

[Gotay and Isenberg, 1992] M. Gotay and J. Isenberg. La symplectification de la science. *La Gazette des Mathématiciens*, 54: 59-79, 1992.

[Gotay and Nester, 1980] M. Gotay and J. Nester. Generalized constraint algorithm and special presymplectic manifolds. In G. Kaiser and J. Marsden (eds.), *Geometric Methods in Mathematical Physics*, pages 79-104. Springer-Verlag, Berlin, 1980.

[Gotay *et al.*, 1978] M. Gotay, J. Nester, and G. Hinds. Presymplectic manifolds and the Dirac-Bergmann theory of constraints. *Journal of Mathematical Physics*, 19: 2388-2399, 1978.

[Gotay, 1983] M. Gotay. On the validity of Dirac's conjecture regarding first-class secondary constraints. *Journal of Physics A*, 16: L141-L145, 1983.

[Gotay, 2000] M. Gotay. Obstructions to quantization. In *Mechanics: From Theory to Computation*, pages 171-216. Springer-Verlag, Berlin, 2000.

[Guillemin and Sternberg, 1990] V. Guillemin and S. Sternberg. *Variations on a Theme by Kepler*. American Mathematical Society, Providence, RI, 1990.

[Healey, 2002] R. Healey. Can physics coherently deny the reality of time? In C. Callender (ed.), *Time, Reality, & Experience*, pages 293-316. Cambridge University Press, Cambridge, 2002.

[Healey, Unpublished] R. Healey. Gauging what's real. Unpublished.

[Henneaux and Shepley, 1982] M. Henneaux and L. Shepley. Lagrangians for spherically symmetric potentials. *Journal of Mathematical Physics*, 11: 2101-2107, 1982.

[Henneaux and Teitelboim, 1992] M. Henneaux and C. Teitelboim. *Quantization of Gauge Systems*. Princeton University Press, Princeton, 1992.

[Hojman and Shepley, 1991] S. Hojman and L. Shepley. No Lagrangian? No quantization! *Journal of Mathematical Physics*, 32: 142-146, 1991.

[Iglesias, 1998] P. Iglesias. Les origines du calcul symplectique chez Lagrange. *L'Enseignement Mathématique*, 44: 257-277, 1998.

[Isenberg and Marsden, 1982] J. Isenberg and J. Marsden. A slice theorem for the space of solutions of Einstein's equations. *Physics Reports*, 89: 179-222, 1982.

[Isenberg and Rendall, 1998] J. Isenberg and A. Rendall. Cosmological spacetimes not covered by a constant mean curvature slicing. *Classical and Quantum Gravity*, 15: 3679-

3688, 1998.

[Isham, 1993] C. Isham. Canonical quantum gravity and the problem on time. In L. A. Ibort and M. A. Rodriguez (eds.), *Integrable Systems, Quantum Groups, and Quantum Field Theories*, pages 157-288. Kluwer, Dordrecht, 1993.

[Kay, 1980] B. Kay. Linear spin-zero quantum fields in external gravitational and scalar fields. II. Generally covariant perturbation theory. *Communications in Mathematical Physics*, 71: 29-46, 1980.

[King and Ellis, 1973] A. R. King and G. F. R. Ellis. Tilted homogeneous cosmological models. *Communications in Mathematical Physics*, 31: 209-242, 1973.

[Kondracki and Rogulski, 1986] W. Kondracki and J. Rogulski. On the stratification of the orbit space for the action of automorphisms on connections. *Dissertationes Mathematicae*, CCL: 1-62, 1986.

[Kuchař, 1986] K. Kuchař. Hamiltonian dynamics of gauge theories. *Physical Review D*, 34: 3031-3043, 1986.

[Kuchař, 1992] K. Kuchař. Time and interpretations of quantum gravity. In G. Kunstatter, D. Vincent, and J. Williams (eds.), *Proceedings of the 4th Canadian Conference on General Relativity and Astrophysics*, pages 211-314. World Scientific, Singapore, 1992.

[Kuchař, 1993] K. Kuchař. Canonical quantum gravity. In R. Gleiser, C. Kozameh, and O. Moreschi (eds.), *General Relativity and Gravitation* 1992, pages 119-150. Institute of Physics Publishing, Philadelphia, 1993.

[Lanczos, 1986] C. Lanczos. *Variational Principles of Mechanics.* Dover, New York, 4th edition, 1986.

[Landsman, 1998] N. P. Landsman. Quantization of singular systems and incomplete motions. In M. Rainer and H. -J. Schmidt (eds.), *Current Topics in Mathematical Cosmology*, pages 256-263. World Scientific, Singapore, 1998.

[Landsman, this volume] N. P. Landsman. Between classical and quantum. Chapter 5, this volume.

[Lang, 1999] S. Lang. *Fundamentals of Differential Geometry.* Springer-Verlag, Berlin, 1999.

[Lee and Wald, 1990] J. Lee and R. Wald. Local symmetries and constraints. *Journal of Mathematical Physics*, 31: 725-743, 1990.

[Lee, 2000] J. Lee. *Introduction to Topological Manifolds*. Springer-Verlag, Berlin, 2000.

[Lee, 2003] J. Lee. *Introduction to Smooth Manifolds*. Springer-Verlag, Berlin, 2003.

[Lewis, 1986] D. Lewis. *Philosophical Papers*, volume II. Oxford University Press, Oxford, 1986.

[Lewis, 1999] D. Lewis. Humean Supervenience debugged. In*Papers in Metaphysics and Epistemology*, pages 224-247. Cambridge University Press, Cambridge, 1999.

[Loll, 1992] R. Loll. Canonical and BRST-quantization of constrained systems. In M. Gotay, J. Marsden, and V. Moncrief (eds.), *Mathematical Aspects of Classical Field Theory*, pages 503-530. American Mathematical Society, Providence, RI, 1992.

[Malament, this volume] D. Malament. Classical relativity theory. Chapter 3, this volume.

[Marco, 1990a] J.-P. Marco. Chiurgie et régularisation des variétés de mouvements. *Comptes Rendus des Séances de l'Académie des Sciences*, 310: 125-127, 1990.

[Marco, 1990b] J.-P. Marco. Variétés de mouvements et régularisation des systèmes dynamiques. *Comptes Rendus des Séances de l'Académie des Sciences*, 310: 65-67, 1990.

[Marsden and Ratiu, 1994] J. Marsden and T. Ratiu. *Introduction to Mechanics and Symmetry*. Springer-Verlag, Berlin, 1994.

[Marsden and Weinstein, 1982] J. Marsden and A. Weinstein. The Hamiltonian structure of the Maxwell-Vlasov equations. *Physica D*, 4: 394-406, 1982.

[Marsden *et al.*, 1972] J. Marsden, D. Ebin and A. Fischer. Diffeomorphism groups, hydrodynamics and relativity. In R. Vanstone (ed.), *Proceedings of the Thirteenth Biennial Seminar of the Canadian Mathematical Congress on Differential Geometry and Applications*, volume 1, pages 135-279. Canadian Mathematical Congress, Montreal, 1972.

[Marsden, 1981] J. Marsden. *Lectures on Geometric Methods in Mathematical Physics*. SIAM, Philadelphia, 1981.

[Maudlin, 2002] T. Maudlin. Thoroughly muddled McTaggart, or how to abuse gauge freedom to generate metaphysical monstrosities. *Philosophers' Imprint*, 2(4): 1-19, 2002.

[Maudlin, Unpublished] T. Maudlin. Suggestions from physics for deep metaphysics. Unpublished.

[McGehee, 1975] R. McGehee. Triple collisions in gravitational systems. In J. Moser (ed.), *Dynamical Systems, Theory and Application*, pages 550-572. Springer-Verlag, Berlin, 1975.

[Milnor, 1984] J. Milnor. Remarks on infinite-dimensional Lie groups. In B. DeWitt and R. Stora (eds.), *Relativity, Groups, and Topology II*, 1007-1057. North-Holland, Amsterdam, 1984.

[Moncrief and Demaret, 1980] V. Moncrief and J. Demaret. Hamiltonian formalism for perfect fluids in general relativity. *Physical Review D*, 21: 2785-2793, 1980.

[Moncrief, 1977] V. Moncrief. Hamiltonian formalism for relativistic perfect fluids. *Physical Review D*, 16: 1702-1705, 1977.

[Moncrief, 1989] V. Moncrief. Reduction of the Einstein equations in 2 + 1 dimensions to a Hamiltonian system over Teichmüller space. *Journal of Mathematical Physics*, 30: 2907-2914, 1989.

[Moncrief, 1992] V. Moncrief. Boost symmetries in spatially compact spacetimes with a cosmological constant. *Classical and Quantum Gravity*, 9: 2515-2520, 1992.

[Norton, 1995] J. Norton. Did Einstein stumble: The debate over general covariance. *Erkenntnis*, 42: 223-245, 1995.

[Nutku and Pavlov, 2002] Y. Nutku and M. Pavlov. Multi-Lagrangians for integrable systems. *Journal of Mathematical Physics*, 43: 1441-1459, 2002.

[Olver, 1993] P. Olver. *Applications of Lie Groups to Differential Equations*. Springer-Verlag, Berlin, 2nd edition, 1993.

[Ortega and Ratiu, 2004] J.-P. Ortega and T. Ratiu. *Momentum Maps and Hamiltonian Reduction*. Birkhaüser, Basel, 2004.

[Pflaum, 2001] M. Pflaum. *Analytic and Geometric Study of Stratified Spaces*. Springer-Verlag, Berlin, 2001.

[Prince, 2000] G. E. Prince. The inverse problem in the calculus of variations and its ramifications. In P. Vassiliou and I. Lisle (eds.), *Geometric Approaches to Differential Equations*, pages 171-200. Cambridge University Press, Cambridge, 2000.

[Redhead, 2003] M. Redhead. The interpretation of gauge symmetry. In K. Brading and E. Castellani (eds.), *Symmetries in Physics*, pages 124-139. Cambridge University Press, Cambridge, 2003.

[Reichenbach, 1991] H. Reichenbach. *The Direction of Time*. University of California Press, Berkeley, 1991.

[Rendall, 1996] A. Rendall. Constant mean curvature foliations in cosmological space-times. *Helvetica Physica Acta*, 69: 490-500, 1996.

[Rovelli, 1991] C. Rovelli. Time in quantum gravity: A hypothesis. *Physical Review D*, 43: 442-456, 1991.

[Rovelli, this volume] C. Rovelli. Quantum gravity. Chapter 12, this volume.

[Ruetsche, Unpublished] L. Ruetsche. Quantizing. Unpublished.

[Rynasiewicz, 1999] R. Rynasiewicz. Kretschmann's analysis of general covariance. In H. Goenner, J. Renn, J. Ritter, and T. Sauer (eds.), *The Expanding Worlds of General Relativity*, 431-462. Birkhaüser, Basel, 1999.

[Saari and Xia, 1995] D. Saari and Z. Xia. Off to infinity in finite time. *Notices of the American Mathematical Society*, 42: 538-546, 1995.

[Sachs and Wu, 1977] R. Sachs and H. -H. Wu. *General Relativity for Mathematicians*. Springer-Verlag, Berlin, 1977.

[Schmid, 1987] R. Schmid. *Infinite Dimensional Hamiltonian Systems*. Bibliopolis, Naples, 1987.

[Singer, 2001] S. Singer. *Symmetry in Mechanics: A Gentle, Modern Introduction*. Birkhaüser, Basel, 2001.

[Sorkin, 2002] R. Sorkin. An example relevant to the Kretschmann-Einstein debate. *Modern Physics Letters A*, 17: 695-700, 2002.

[Souriau, 1982] J. -M. Souriau. Géométrie globale du problème à deux corps. *Atti della Accedemia delle Scienze di Torino. Classe di Scienze Fisiche, Matematiche e Naturali*, 117 (Supplement 1): 369-418, 1982.

[Souriau, 1986] J. -M. Souriau. La structure symplectique décrite par lagrange en 1811. *Mathématiques et Sciences Humaines*, 94: 45-54, 1986.

[Souriau, 1997] J. -M. Souriau. *Structure of Dynamical Systems*. Birkhaüser, Basel, 1997.

[Sudbery, 1986] A. Sudbery. A vector Lagrangian for the electromagnetic field. *Journal of Physics A*, 19: L33-L36, 1986.

[Tarkovsky, 1991] A. Tarkovsky. *Time Within Time: The Diaries, 1970-1986*. Seagull, Calcutta, 1991.

[Torre and Anderson, 1996] C. Torre and I. Anderson. Classification of local generalized symmetries for the vacuum Einstein equations. *Communications in Mathematical Physics*, 176: 479-539, 1996.

[Torre, 1992] C. Torre. Covariant phase space formulation of parameterized field theories. *Journal of Mathematical Physics*, 33: 3802-3811, 1992.

[Torre, Unpublished] C. Torre. Local cohomology in field theory (with applications to the Einstein equations). Unpublished.

[Tuynman, 1992] G. Tuynman. What are the rules of the game called BRST? In M. Gotay, J. Marsden, and V. Moncrief (eds.), *Mathematical Aspects of Classical Field Theory*, pages 625-633. American Mathematical Society, Providence, RI, 1992.

[van Brunt, 2004] B. van Brunt. *The Calculus of Variations*. Springer-Verlag, Berlin, 2004.

[Wald, 1984] R. Wald. *General Relativity*. University of Chicago Press, Chicago, 1984.

[Wald, 1994] R. Wald. *Quantum Field Theory on Curved Spacetime*. University of Chicago Press, Chicago, 1994.

[Wallace, 2003] D. Wallace. Time-dependent symmetries: The link between gauge symmetries and indeterminism. In K. Brading and E. Castellani (eds.), *Symmetries in Physics*, pages 163-173. Cambridge University Press, Cambridge, 2003.

[Weinstein, 1981] A. Weinstein. Symplectic geometry. *Bulletin of the American Mathematical Society*, 5: 1-13, 1981.

[Woodhouse, 1991] N. M. J. Woodhouse. *Geometric Quantization*. Oxford University Press, Oxford, 2nd edition, 1991.

[Zuckerman, 1987] G. Zuckerman. Action principles and global geometry. In S.-T. Yau (ed.), *Mathematical Aspects of String Theory*, pages 259-284. World Scientific, 1987.

第三章　经典相对论

大卫·B. 马拉蒙特

1. 引言

这篇论文主要分为两个部分。第一部分，我主要简单介绍经典相对论①的结构。第二部分，我主要针对三个主题进行讨论。

在第一部分(第2节)中，我的论述会限制在以下这些方面。我不会讨论经典相对论的历史发展及其实验依据。就"狭义相对论"自身而言，我不会将其看作一个被"广义相对论"替代的理论。此外我也不会描述已知的关于爱因斯坦(Einstein)方程的精确解(这方面内容在其他章节中有大篇幅的论述②)。

与此不同的是，我将会讨论一些基础的思想，并尽我所能地将其表述清楚。对此我事先假设读者对基础的微分几何有很

① 这里所说的"经典"相对论是指不涉及量子力学的内容。确切地说，就是不讨论弯曲空间中的量子场论以及量子引力理论。针对这些方面的讨论，请参见罗韦利(Rovelli)的论文(本书第十二章)。

② 我这里没有涉及的两个重要论题在本书中的其他文章中被重点关注，分别是相对论的初值公式化[Earman(厄尔曼)，Chapter 14]，以及相对论的哈密顿(Hamiltonian)公式化[Belot(贝洛特)，Chapter 2]。

好的认识，并且至少应该对相对论本身有所了解。[①]

在第三节中，我首先考虑了闵可夫斯基(Minkowski)时空环境下相对同时性关系的地位问题。我在本文中所关注的焦点是从爱因斯坦对同时性做出的“定义”中挑选出来的这种标准关系到底是自然存在的，还是在某种意义上说是强加于我们头脑中的。接着我将从几何学的视角描述牛顿引力理论，一般称之为牛顿—嘉当(Newton-Carton)理论。之所以涉及这方面的内容是想要说明经典相对论在哪些方面与之相符或有所区别。最后，我将考虑在什么程度上时空的整体几何结构能够从其“因果结构”中得到。[②]

2. 相对论的结构

2.1 相对时空

相对论为我们的宇宙(或其子区域，如我们的太阳系)的时空结构确定了一类几何模型。其中每个模型都代表了一个与理论的约束条件相容的可能世界(或世界区域)。我们将分几个部分描述这些模型。我们首先描述“相对时空”的一个总类，并讨论其解释，然后介绍涉及整体时空结构和爱因斯坦方程的约束条件。

我们将用一对符号(M，g_{ab})来表示相对时空。这里 M 是一个平滑的，连

① 对于所需的微分几何基础(我称之为“抽象指数符号”)，可以在一些综述性的著作中找到，具体的例子就是瓦尔德(Wald)于1984年所写的论著及马拉蒙特(Malament)的论著。在巴特菲尔德(Butterfield)论文(本书第一章)的3.1和3.2节中也有部分此方面的内容。在准备第一部分时，我参考了大量的文献。其中最重要的是格洛奇(Geroch)的相关论著及[Hawking and Ellis(霍金和埃利斯)，1972]、[O'Neill(奥尼尔)，1983]、[Sachs and Wu (萨克斯和吴)，1977a；1977b]、[Wald，1984]的论著。

② 从一个稍微不同的视角对经典相对论基础进行的更多讨论，请参见罗韦利的论文(本书第十二章)。

续的四维流形，g_{ab}是 M 中的一个平滑的、关于洛伦兹(Lorentz)标记(1，3)① 的半黎曼(Semi-Riemannian)度规。

我们将 M 解释为世界中点“事件”的流形。② g_{ab}的解释通过一个互相关联的物理原理网络给出。在本节中我们只列出特征上相对简单的三个原理，因为它们只涉及点粒子和光线(仅仅这些对象就足以确定度规，最多还取决于一个常数)。在下一节中我们将列出第四个原理，它涉及(理想)时钟的作用。关于一般物质场的其他原理也将在后面介绍。

我们从回顾一些定义开始。在下面的内容中，$(M,\ g_{ab})$是一个固定的相对时空；令 ∇_a 为一个 M 中由 g_{ab}确定的微分算符，即 M 中一个满足相容性条件 $\nabla_a g_{bc}=0$的唯一(无挠)微分算符。

给定 M 中的一个点 p，以及 p 点切线空间 M_p 中一个矢量 η^a。如果 $\eta^a\eta_a>0$，则 η^a 为类时的(Timelike)；如果 $\eta^a\eta_a=0$，则 η^a 是类光的(Lightlike)；如果 $\eta^a\eta_a\geqslant 0$，η^a 是因果性的；如果 $\eta^a\eta_a<0$，则 η^a 是类空的(Spacelike)。

按照这种办法，g_{ab}给出了 M 中每一个点的切线空间中的“零锥结构”。类时矢量在该结构的内部，类空矢量在锥体的外部。因果关系矢量要么类时要么类光。这个分类能够很自然地扩展到曲线。我们认为这些是如下形式的平滑映

---

① 所述的标记条件是等价于这样一个要求，即在 M 内的每一个点 p 上，切空间 M_p 有一个基 $\overset{1}{\xi}{}^a$，…，$\overset{4}{\xi}{}^a$，使得对于{1，2，3，4}中的所有 i 和 j，如果 $i\neq j$，$g_{ab}\overset{i}{\xi}{}^a\overset{j}{\xi}{}^b=0$。

$$g_{ab}\overset{i}{\xi}{}^a\overset{i}{\xi}{}^b=\begin{cases}+1, & \text{如果 } i=1\\ -1, & \text{如果 } i=2,\ 3,\ 4\end{cases}$$

(这里我们使用的是抽象指标符号。“a”是一个抽象的指标，而“i”和“j”是正常的计数指标。)由此得出结论，在 p 点给定任意向量 $\eta^a=\sum_{i=1}^{4}\overset{i}{k}\overset{i}{\xi}{}^a$，$\rho^a=\sum_{j=1}^{4}\overset{j}{l}\overset{j}{\xi}{}^a$，有：

$$g_{ab}\eta^a\rho^b=\overset{1}{k}\overset{1}{l}-\overset{2}{k}\overset{2}{l}-\overset{3}{k}\overset{3}{l}-\overset{4}{k}\overset{4}{l}$$

$$g_{ab}\eta^a\eta^b=\overset{1}{k}\overset{1}{k}-\overset{2}{k}\overset{2}{k}-\overset{3}{k}\overset{3}{k}-\overset{4}{k}\overset{4}{k}$$

在下文中，我们往往会使用标准约定，为了用度规 g_{ab}下降(抽象)指数，用逆度规 g^{ab}提升指数。因此，举例来说，我们会用 $\eta_a\rho^a$ 或 $\eta^a\rho_a$ 代替 $g_{ab}\eta^a\rho^b$。

② 这里我们将“事件”看作一个中性词，并没有给出特别的意义。有些人可能会愿意谈到“重合点事件的等价类”，或“点事件位置”，或类似的术语。

射γ：$I\to M$，其中$I\subseteq\mathbb{R}$ 是一个（可能是无限的，但不必是开的）间隔。[①] 如果它的切线场$\vec{\gamma}$在每一个点都具有这样的特征，那么γ就是类时的（或类光的、因果性的、类空的）。

如果存在一个从I_2到I_1的平滑映射τ：$I_2\to I_1$，且其导数处处为正，我们就把曲线γ_2：$I_2\to M$称为曲线γ_1：$I_1\to M$的（保持定向）再参数化。类时、类光等性质在再参数化下得到保留。[②]因此在一个明确的意义下，我们的分类也能扩展到曲线的图像。[③]

如果一个曲线γ：$I\to M$的切线场ξ^a满足条件$\xi^n\nabla_n\xi^a=0$，则该曲线被称为（关于g_{ab}的）测地线。测地线的性质在再参数化后不会再保留，所以它不会转移到曲线图像上。但是，适合再参数化的测地线的相关属性当然会保留。（后者对曲线成立，如果它可以被再参数化为一条测地线。）

现在我们叙述前三个解释性原理。对于所有曲线γ：$I\to M$：

C1　γ是类时的，当且仅当其图像$\gamma[I]$是一个有质量的点粒子（即具有正质量的点粒子）的世界线；[④]

C2　γ能够被再参数化为零测地线，当且仅当其图像$\gamma[I]$是一束光的轨迹；[⑤]

P1　γ能够被再参数化为类时测地线，当且仅当其图像$\gamma[I]$是一个有质量

---

① 如果I不是开集，我们可以将平滑理解为这意味着存在一个开放的区间$\bar{I}\subseteq\mathbb{R}$，其中$I\subset\bar{I}$，以及一个平滑映射$\bar{\gamma}$：$\bar{I}\to M$，使得$\bar{\gamma}(s)=\gamma(s)$对所有$s\in I$成立。

② 这是从如下事实得出的，即在上面所描述的那种情况下，$\vec{\gamma}_2=\frac{d\tau}{ds}\vec{\gamma}_1$，其中$\frac{d\tau}{ds}>0$。

③ 曲线和曲线的图像之间的差异（即映射γ：$I\to M$和集$\gamma[I]$之间的差异）很重要。我们认为世界线是后者的例子，也就是将它们解释为点集，而不是参数化的点集。

④ 我们将在以后讨论相对论中质量的概念。就目前而言，我们只把它当作一个基本的粒子属性。

⑤ 对于特定的目的，即使是在经典相对论中，认为光是由“光子”流组成，并认为右侧条件“$\gamma[I]$可能是一个光子的世界线”是有用的。后面的公式使C2看起来更像C1和P1，并使我们注意这样一个事实，即在相对论时空结构中，大质量粒子和零质量粒子（如光子）之间的区别具有直接的意义。

的自由①点粒子的世界线。

在每种情况下，关于几何结构的描述(等式左侧)是与关于粒子或者光束行为的描述(等式右侧)相关联的。

这里还需要给出一些注释和限定条件。

第一，我们在传统意义下的相对论框架内进行研究，并且忽略超过光速的粒子("超光速粒子")存在的可能性(它们的世界线会以类空曲线出现)。

第二，我们要求曲线是平滑的。因此，依赖于模型化点粒子碰撞的方式，我们可能会将注意力集中于那些没有发生过碰撞的粒子上。

第三，该理论需要限制性条件，因为在相对论中"点粒子"的地位是微妙的。问题是，能否将一个粒子本身的质量—能量看作一个环绕度规场 g_{ab} 的源(除了那些恰好存在的其他源以外)。(这里我们为关于爱因斯坦方程的讨论做准备。)如果我们这样认为，那么当接近粒子的世界线时，与 g_{ab} 相关的曲率会突然变大。在这种情况下，我们将不能把世界线作为 M 中曲线的图像。至少要放弃 g_{ab} 必须是 M 中的平滑场的要求。出于这个原因，该原理的更谨慎的公式化将仅限于讨论"测试粒子"，即那些自身质量—能量是可以忽略不计的，或可以为此目的而忽略不计的粒子。

第四，该理论的模态特征(即可能性的对照)是必不可少的。以 C1 为例，事实上，所有类时曲线的图像都是有质量粒子的世界线，这种认识显然是不正确的。其论据是，至少就涉及的相对论的定律来说，它们应该是。当然，在某种情况下做出什么样的判断取决于背景中给定的是什么样的条件。关于一个特定的曲线图像是一个有质量粒子的世界线的陈述必须被如下理解：例如，只要在该路径上没有障碍，该曲线就是可能的测地线。同样，在 C2 中，存在一个暗示性的约束条件，也就是我们考虑在没有中介媒质存在时，即当我们考虑真空中的光线时，这种光线会有什么样的轨迹。

虽然这四点很重要，并且它们还提出了一个有意义的问题，即在物理理论

---

① 这里的"自由粒子"必须被理解为不会受任何力(除了"引力")的粒子。相对论的一个基本原理是引力是时空弯曲的一种表现，而不是使粒子偏移其自然、平直的轨迹(测地线)的外力。在 2.4 节中，我们将进一步讨论这一问题。

的公式化中，理想化与形式化起什么作用的问题，但是它们仍与相对论没什么关系。类似的困难也出现在当人们试图在牛顿引力理论的框架内公式化相应的原理时。

从解释性原理可知度规 g_{ab}是由点粒子和光线的行为决定的(取决于一个常数)。在关于“保形结构”和“投影结构”的一对命题中，我们能精确地表述这个观点。

令 $\bar{g}_{ab}$为 M 中洛伦兹特征的第二平滑度规。如果 M 中存在一个平滑映射 Ω：$M \to \mathbb{R}$ (Ω 被称为保形因子，它当然不需要为常数)，满足 $\bar{g}_{ab}=\Omega^2 g_{ab}$，那么我们说 $\bar{g}_{ab}$保形等价于 g_{ab}。如果 $\bar{g}_{ab}$保形等价于 g_{ab}，则它们在类时、类光等曲线或矢量的分类上是一致的；反之亦然。① 一般来说，保形等价指标并不能决定 M 中的哪些曲线适合作为测地线或者能够再参数化为测地线。但是，事实证明，保形等价指标必然能确定哪些类光曲线是再参数化下的测地线；②反之亦然。③

把这些放在一起，我们有下面的命题。条件(1)和(2)分别对应于 C1 和 C2。

命题 1。令 $\bar{g}_{ab}$为 M 中洛伦兹特征的第二平滑度规，则如下条件是等价的：

(1)关于 M 中的哪条曲线是类时的，$\bar{g}_{ab}$和 g_{ab}是一致的；

(2)关于 M 中的哪条曲线能被再参数化为零测地线，$\bar{g}_{ab}$和 g_{ab}是一致的；

(3)$\bar{g}_{ab}$和 g_{ab}是保形等价的。

---

① 矢量和曲线分别属于三个指标中的任何一个，如果其中两个指标一致，那么它们必然全部一致，并且在此情况下，它们必然是保形等价的，参见[Hawking and Ellis，1972，61]。

② 这是因为从 M 中存不存在(局域)类时曲线和空连接点方面，可以得到表现为一个零测地线的属性。相关的技术性引理阐述如下：

曲线 γ：$I \to M$ 可以在参数化为一个零测地线，当且仅当 γ 是空的，且对所有 $S \in I$，存在一个包含 $\gamma(S)$ 的开集 $O \subseteq M$，使得对所有 S_1，$S_2 \in I$，如果 $S_1 \leqslant S \leqslant S_2$ 且 $\gamma([S_1, S_2]) \subseteq O$，那么在 O 内就没有从 $\gamma(S_1)$ 到 $\gamma(S_2)$ 的类时曲线。这里 $\gamma([S_1, S_2])$ 是限制在区间 $[S_1, S_2]$ 内的 γ 的图像，对此的证明请参见[Hawking and Ellis，1972，103]。

③ 因为如果度规在曲线是零测地线直到再参数化之前的情况下是相同的，那么它们在任意点的矢量是零的情况下也必然相同，我们知道，这意味着这些指标是保形等价的。

在这个意义上说，时空度规 g_{ab} 按照一个有质量的点粒子的可能世界线构成的集合和一个光线的可能轨迹构成的集，由一个保形因子独立地确定。

接下来我们来说投影结构。令 $\overline{\nabla}_a$ 为 M 中的二阶导数算子，我们说 $\overline{\nabla}_a$ 和 ∇_a 是投影等价的，如果它们在哪条曲线被再参数化后是测地线这个方面是一致的。(也就是说，如果对于所有曲线 γ，γ 能被再曲线化为关于 $\overline{\nabla}_a$ 的测地线，当且仅当关于 ∇_a 它能被再参数化。) 并且，如果 $\bar{g}_{ab}$ 是 M 中洛伦兹标记的第二指标，我们就说它与 g_{ab} 是投影等价的，如果与其相关的导数算子 $\overline{\nabla}_a$ 和 ∇_a 是投影等价的话。

根据外尔(Weyl)[1921]的说法，如果 $\bar{g}_{ab}$ 和 g_{ab} 是保形等价且投影等价的话，那么与它们相关联的保形因子必然为常数，这是一个基本的结果。对于我们的目的来说，我们可以很方便地根据我们选取的解释性原理 P1，稍微将其形式修改为仅涉及类时测地线(而非任意测地线)。

命题 2。令 $\bar{g}_{ab}$ 为 M 中的第二平滑度规，且有 $\bar{g}_{ab}=\Omega^2 g_{ab}$。如果关于哪些类时曲线能被再参数化为测地线 $\bar{g}_{ab}$ 和 g_{ab} 也是一致的话，那么 Ω 为常数。

我们看到，时空度规 g_{ab} 是按照一个质点粒子的可能世界线构成的集和一个光线的可能轨迹构成的集，由一个保形因子独立地确定。现有的命题使得我们能够明确在什么意义下它是由上述集与自由重粒子的可能世界线构成的集一起来完全确定的(取决于一个常数)。①

我们对相对时空的描述是极其宽泛的，还可以再加上许多进一步的条件。目前，我们只考虑其中一个。

如果在 M 上存在连续的类时矢量场 τ^a，则 (M, g_{ab}) 被称为是具有时间方向的。假设该条件被满足，那么我们说场 τ^a 和 $\hat{\tau}^a$ 是共向的，如果处处有 $\tau^a\hat{\tau}_a>0$，即在 M 中的每一个点上，τ^a 和 $\hat{\tau}^a$ 都落在零锥的相同的半周期上。共向是(M 中的连续类时矢量场构成的集上的)两个等价类之间的等价关系。(M,

---

① 正如外尔[Weyl, 1950, 103]指出的，……可以表明世界的度量结构完全取决于它的惯性结构和因果结构，因此测定法并不依赖于时钟和刚体，但是仅在惯性影响下的光信号和质点运动就足够了。关于外尔确定时空度量的"因果惯性"方法的详细信息，请参阅[Coleman and Korte (科尔曼和科特), 2001, Section 4.9]。

g_{ab})的时间方向取决于这两个等价类中的哪一个被选择看作“未来”的方向。因此，对M中一个点的非零因果矢量 ξ^a 被说成相对于时间取向T是未来方向的或过去方向的，这取决于在这点上 $\tau^a\xi_a>0$ 还是 $\tau^a\xi_a<0$，其中 τ^a 是T方向的任意连续类时矢量场。受其影响，因果关系曲线 γ：$I\rightarrow M$ 被认为相对于T是未来方向的(或者过去方向的)，如果其正切矢量是那个方向的话。

在下文中，我们假设我们的背景时空(M，g_{ab})在时间上是定向的，并且已被指定了一个特定的时间方向。此外，对于 M 中的给定事件 p 和 q，我们可以将它们记作 $p\ll q$(或者 $p<q$)，如果存在一个未来方向的类时(或者因果)曲线开始于 p，结束于 q 的话。①

2.2 本征时间

到目前为止，我们讨论相对时空结构时既没有涉及“时间”也没有提到“空间”。在本节及下一节中我们将针对它们进行讨论。

令 γ：$[s_1, s_2]\rightarrow M$ 为 M 中的一个未来方向的类时曲线，它具有切线场 ξ^a。我们将其与一个(相对于 g_{ab}的)经历的本征时间通过下式联系了起来：

$$|\gamma| = \int_{s_1}^{s_2}(g_{ab}\xi^a\xi^b)^{\frac{1}{2}}ds$$

这个经历的本征时间在再参数化 γ 下是不变的，而这正是我们将其描述为 γ(图像的)的长度的量。下面是相对论的另一个基本原理。

P2 时钟记录其世界线上经历的本征时间的长度。

我们仍然需要一些限制条件和注释。我们对 C1，C2 和 P1 进行的公式化是粗糙的，对现在的公式化更是这样。我们假定我们知道“时钟”是什么。我们假设它们具有世界线(worldlines)而非世界管(worldtubes)。我们忽略了一个事实，即普通时钟(例如，放在床头柜上的闹钟)在受到极端加速力、潮汐力等外力时不能正常工作(将闹钟扔到墙上试试，就明白了)。此外，这样的一些考虑是重要的，并且它们提出了关于理想化在物理理论公式化过程中的作用这样一些有

① 由此可以直接得出结论，如果 $p\ll q$，则 $p<q$ 成立。一般来说反过来并不成立。但使第二个条件为真实但第一个条件不为真的唯一方法是，从 p 到 q 的未来导向型因果关系曲线是零测地线(或零测地线的再参数化)。参见[Hawking and Ellis，1972，112]。

意义的问题。(有人可能会将“理想时钟”解释为能完美记录世界线上经历的本征时间的点状测量客体，然后认为P2给定了真正的时钟，在适当的条件下，为了适应其准确度，需要一定的理想化。)不过我们所关注的是点粒子的状态，对它们的探讨与相对论没有多少关系。类似的情况也会出现在当人们试图在牛顿的理论框架内将与时钟行为相关的原理进行公式化描述时。

现在假设我们可以确定时空的保形结构，比如说利用光线。那么人们就可以用时钟，而不是自由粒子，来确定保形因子。与命题2相比，人们可以得到下面的简单结果。①

命题3。令 $\bar{g}_{ab}$是 M 中的第二平滑度规，且有 $\bar{g}_{ab}=\Omega^2 g_{ab}$。进一步假设，对于所有的类时曲线 γ：$I\rightarrow M$，两个指标为所有类时曲线赋予相同的长度值，即 $|\gamma|_{\bar{g}_{ab}}=|\gamma|_{g_{ab}}$，那么处处有 $\Omega=1$。(这里 $|\gamma|_{g_{ab}}$是 γ 相对于 g_{ab}的长度。)

P2(在上面那些考虑的模式下)完全描述了相对论时钟的行为。特别是，它说明了时钟的读取与路径有关。如果两个从同一个事件 p 出发的时钟，沿着不同的路径到达事件 q，则一般来说，它们记录下的历经时间将会不同。例如一个记录下的经历时间为3806秒，另一个记录下的为649秒。这是真实的，无论这些时钟是多么相似(我们可能会要求它们来自同一条生产线)。正如P2所断言的，这种情况是因为，每个时钟记录的经历时间是它从 p 到 q 经过的类时曲线的长度，在一般情况下，那些长度是不同的。

假设我们考虑所有从 p 到 q 的未来方向型类时曲线。我们自然会问，是否能使所记录的两个事件之间的历经时间最小化或最大化。对第一个问题的答案是“不能”。事实上，我们有下面的命题。

命题4。设 p 和 q 是在 M 中的事件，且 $p\ll q$。则对于所有的 $\epsilon>0$，存在一条从 p 到 q 的未来方向的曲线 γ，使得 $|\gamma|<\epsilon$。(但这里并不存在长度为0的曲线，这是因为所有类时曲线的长度都不为0。)

虽然要给命题一个可靠的证明还需要一些工作，参见[O'Neill，1983，

---

① 这里，我们不仅为常数确定度规，而且还确定常数。在这里有所区别的是，实际上，我们已经为时空距离选择好了单位。如果我们研究的不是经历的适当时间间隔，而是这种时间间隔的比例，那么我们可以得到一个等价于命题2的更精确的命题。

pp. 294—295]，但它在直观上看起来是合理的。如果存在一个类时曲线连接p和q，那么应该还存在一个锯齿状的类光曲线连接它们，它的长度为0。但是，我们可以将相互连接的类光曲线任意近似为前后摆动的平滑类时曲线。因此，(由长度函数的连续性)我们可以想到，对所有的$\epsilon>0$，存在一个长度小于ϵ的近似类时曲线。见图1。

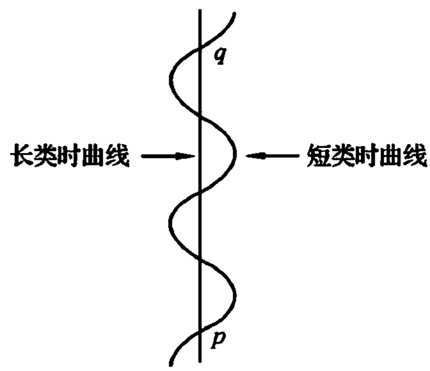


图1　从p到q的一条长类时曲线，一条前后摆动用来近似折线状零曲线的非常短的类时曲线

如果我们将注意力放在时空的局部区域，那么第二个问题(能使所记录的两个事件p和q之间的经历时间最大化吗?)的答案是"肯定的"。在度规是正定的情况下，即度规为$(n, 0)$的形式的情况下，我们知道测地线是最短的线。洛伦兹度规的相应结果是类时测地线是局部最长的曲线。

命题5。令γ: $I \to M$为一个未来方向的类时曲线。那么曲线γ能够被再参数化为测地线，当且仅当对所有$s \in I$，存在一个含有$\gamma(s)$的开集O，使得对于可用于所有满足$s_1 \leqslant s \leqslant s_2$的$s_1$，$s_2 \in I$，如果$\bar{\gamma}=\gamma_{[s_1,s_2]}$的图像包含于$O$中，则$\bar{\gamma}$(及其再参数化的结果)比所有$O$中从$\gamma(s_1)$到$\gamma(s_2)$的其他类时曲线都要长。(这里$\gamma_{[s_1,s_2]}$表示将$\gamma$约束在区间$[s_1, s_2]$中。)

该命题的证明非常类似于正定的情况，可参见霍金和埃利斯[1972, 105]。因此，在所有从p到q的局域时钟中，有一个会记录下从p"自由下落"到q的最大历经时间。为了得到一个时钟记录比极大值稍小的历经时间，人们将不得不加速该时钟。已知加速需要燃料，而燃料不是免费的，所以命题5的推论是(在局域时空中)"节省时间需要耗费金钱"。命题4认为，(在局域时空中)"有

足够的钱，我们想节省多少时间就能够节省多少时间”。

局域时空这一限制是必要的。时钟的行为和加速度之间的联系在一个整体区域上不成立。在一些相对时空中，人们可能发现连接两个事件的未来方向的类时测地线具有不同的长度，并且沿着曲线的时钟记录下的两个事件之间历经时间不同。即使它们都处在自由落体状态。此外——仅根据由连续性因素得到的上述断言——情况可以是这样的：记录两个事件之间历经时间的两个不同时钟，其中在过程中加速的那个时钟可能比处于自由落体状态的那个时钟记录下更长的历经时间。

我们一直考虑的时钟行为和加速之间的关系曾一度被认为是自相矛盾的（我会想到“时钟或双生子佯谬”）。假设两个时钟，A 和 B，在一个适当的小时空区域中，从一个事件进入到另一个事件。进一步假设 A 时钟处在自由落体状态，但 B 在前进过程中一些点上处于加速状态。那么我们知道，A 将比 B 记录到更长的历经时间。这曾经被认为是自相矛盾的，因为我们认为，“相对论否认自由落体运动和加速运动之间‘绝对’可区别性”（如果我们同样也有权力认为是时钟 B 处于自由落体状态，而 A 经历了加速，那么通过等价的推理，应该是 B 记录更长的历经时间）。解决这个矛盾的方法，如果这能叫矛盾的话，是指出相对论没有给出这样的否认。A 和 B 的情况不是对称的。加速运动和自由落体的每一处微小区别在相对论中与在牛顿物理学中都是一样的。

在下文中，除非做出相反的假设，“类时曲线”都应该理解为由经历的本证时间，即弧长度再参数化得到的未来方向的类时曲线。在这种情况下，曲线的切线场 ξ^a 具有单位长度（$\xi^a\xi_a=1$）。并且，如果一个粒子碰巧以曲线的图像作为其世界线，那么，在任意点，ξ^a 被称为粒子在该点的四维速度。

2.3 某一点的空间/时间分解及粒子动力学

令 γ 是一个表示具有四维速度场 ξ^a 的粒子 O 的类时曲线。令 p 是 γ 的图像上的一点，并令 λ^a 为过 p 点的一个矢量。这里存在一个将 λ^a 分解成平行和垂直于 ξ^a 的分量的分解方式：

$$\lambda^a=\underbrace{(\lambda^b\xi_b)\xi^a}_{\text{平行于}\xi^a}+\underbrace{(\lambda^a-(\lambda^b\xi_b)\xi^a)}_{\text{垂直于}\xi^a} \tag{1}$$

它们分别可以被标准地解释为 λ^a 的"时间"分量和"空间"分量。特别是，由垂直于ξ^a 的矢量组成的 M_p 的三维子空间可以被解释为 O 的"无穷小"截面。[①] 如果我们引入正切和正交投影算符：

$$k_{ab}=\xi_a\xi_b \tag{2}$$

$$h_{ab}=g_{ab}-\xi_a\xi_b \tag{3}$$

那么这种分解可以表达为：

$$\lambda^a=k^a{}_b\lambda^b+h^a{}_b\lambda^b \tag{4}$$

我们可以将 k_{ab}和 h_{ab}看作是由 ξ^a 决定的相对时间和相对空间度规。它们是对称的并且满足：

$$k^a{}_b k^b{}_c=k^a{}_c \tag{5}$$

$$h^a{}_b h^b{}_c=h^a{}_c \tag{6}$$

许多标准教科书指出粒子的运动学和动力学可以由这些分解公式重新得到。例如，假设第二粒子 $\overline{O}$ 的世界线经过事件 p 且其在 p 点的四维速度是$\overline{\xi}^a$。由于$\overline{\xi}^a$ 和 ξ^a 都是未来方向的，它们是共向的，即$(\xi^a\overline{\xi}_a)>0$。我们计算由 O 确定的 $\overline{O}$ 的速度。要做到这一点，我们用相对于 O 的 $\overline{\xi}^a$ 的空间的量值除以相对于 O 的时间值得到：

$$v=\overline{O}\text{ 相对于 }O\text{ 的速度}=\frac{\| h^a{}_b\overline{\xi}^b \|}{\| k^a{}_b\overline{\xi}^b \|} \tag{7}$$

给定任意矢量μ^a，如果μ^a 是因果性的，我们将 $\|\mu^a\|$ 理解为$(\mu^a\mu_a)^{\frac{1}{2}}$，且如果 u^a 是类空的，我们将 $\|\mu^a\|$ 理解为$(-\mu^a\mu_a)^{\frac{1}{2}}$。从式(2)、式(3)、式(5)和式(6)，我们有：

$$\| k^a{}_b\overline{\xi}^b \| =(k^a{}_b\overline{\xi}^b k_{ac}\overline{\xi}^c)^{\frac{1}{2}}=(k_{bc}\overline{\xi}^b\overline{\xi}^c)^{\frac{1}{2}}=(\overline{\xi}^b\xi_b) \tag{8}$$

$$\| h^a{}_b\overline{\xi}^b \| =(-h^a{}_b\overline{\xi}^b h_{ac}\overline{\xi}^c)^{\frac{1}{2}}=(-h_{bc}\overline{\xi}^b\overline{\xi}^c)^{\frac{1}{2}}=((\overline{\xi}^b\xi_b)^2-1)^{\frac{1}{2}} \tag{9}$$

所以：

$$v=\frac{((\overline{\xi}^b\xi_b)^2-1)^{\frac{1}{2}}}{(\overline{\xi}^b\xi_b)}<1 \tag{10}$$

---

① 这里，我们简单地认为具有正交性的"相对同时性"的标准验证是理所当然的。在 3.1 节中，我们将返回来考虑其理由。

因此，正如由 O 所测量的，重粒子不可能达到最大速度 1(类似的计算表明，正如由 O 所确定的那样，光传播速度始终为 1)。为便于后面的讨论，我们在这里指出式(10)意味着：

$$(\bar{\xi}^b \xi_b) = \frac{1}{\sqrt{1 - v^2}} \tag{11}$$

相对论中的一个基本事实是，对于世界线上的每个事件，都存在与每个点粒子相关的一个四维动量(或能量—动量)矢量 P^a。在一个具有四维速度 $\bar{\xi}^a$ 的重粒子的情况下，P^a 正比于 $\bar{\xi}^a$，并且(正)比例系数就是我们所说的粒子的质量(或净质量)m，因此，我们有 $P^a = m\bar{\xi}^a$。在光子(或其他质量为 0 的粒子)的情况下，这样的特性是不存在的，因为其世界线是类光(而不是类时)曲线的图像。但我们仍然可以将我们关心的事件的四维动量矢量理解为一个与其世界线相切的未来方向的零矢量。如果我们将四维动量矢量 P^a 看作是基础的，那么我们可以在这两种情况下，重新得到粒子的质量为 P^a 的长度：$m = (P^a P_a)^{\frac{1}{2}}$(在第一种情况下它严格为正，在第二种情况下为 0)。

现在假设一个重粒子 O 在某个事件点上具有四维速度 ξ^a，以及另一个重粒子或光子，在那里具有一个四维动量 P^a。如果我们以 ξ^a 的方式分解 P^a，可以重新得到第二个粒子相对于 O 的能量和三维动量的一般表达式。由式(4)和式(2)，我们有：

$$P^a = (P^b \xi_b)\xi^a + h^a{}_b P^b \tag{12}$$

相对于 O 的能量是第一项 $E = P^b \xi_b$ 的系数。在粒子为重粒子且 $P^a = m\bar{\xi}^a$ 成立的情况下，由式(11)可以得到：

$$E = m(\bar{\xi}^b \xi_b) = \frac{m}{\sqrt{1 - v^2}} \tag{13}$$

注：如果我们没有选择在其中 $c = 1$ 的单位，那么最后一个表达式中的分子为 mc^2，分母为 $\sqrt{1 - \frac{v^2}{c^2}}$。相对于 O 的三维动量是分解后的第二项，即垂直于 ξ^a：$h^a{}_b P^b$ 的 P^a 的分量。在重粒子的情况下，由式(9)和式(11)可得：

$$p = \| h^a{}_b m \bar{\xi}^b \| = m((\bar{\xi}^b \xi_b)^2 - 1)^{\frac{1}{2}} = \frac{mv}{\sqrt{1 - v^2}} \tag{14}$$

解释性原理 P1 断言，自由粒子会经过类时测地线的图像。它可以被认为是牛顿第一运动定律的相对论版本。现在，我们考虑加速度和第二定律的相对论版本。令 γ：$I \to M$ 为一个类时曲线，其图像是一个重粒子 O 的世界线，并令 ξ^a 为 O 的四维速度场。则 O 的四维加速度(或加速度)场为 $\xi^n \nabla_n \xi^a$，即 ξ^a 在方向 ξ^a 上的方向导数。四维加速度矢量正交于 ξ^a。这是很明显的，因为 $\xi^a(\xi^n \nabla_n \xi_a) = \frac{1}{2}\xi^n \nabla_n(\xi^a \xi_a) = \frac{1}{2}\xi^n \nabla_n(1) = 0$。四维加速度矢量在某个点上的大小 $\| \xi^n \nabla_n \xi^a \|$ 就是被我们描述为高斯曲率 γ 的值。它是对 γ 的曲线偏离直线路径的程度的度量(γ 正好就是测地线，如果其曲率处处为零的话)。

我们需要注意时空加速度的概念。考虑这样一个例子：假设你决定要结束一切，跳下帝国大厦，在你最后的时刻加速度是怎样的？在这种情况下，习惯上认为它是相对于地球的加速度。所以有人会说，从你开始跳下到灾难到来之前的整个过程中你经历了加速。但目前看来，这只能说明你周围的事物在倒退。在这段时间内，你并没有加速，你正处于自由落体状态，且沿时空测地线移动，而起跳前和落地后，你正在加速。地面的观测台的地板和地面的人行道，将你推离测地线路径。这里最重要的观点是，我们将“引力场”纳入到了时空的几何结构中，粒子会经过测地线，当且仅当它们“除了引力”外不受任何力的作用时。

任何重粒子的加速度，即其偏离测地线的轨迹，是由(“引力”以外的)力的作用决定的。如果粒子具有质量 $m>0$，且作用在 $\gamma[I]$上的矢量场 F^a 表示作用在粒子上的各种力(除引力)的矢量和，则粒子的四维加速度 $\xi^n \nabla_n \xi^a$ 满足条件：

$$F^a = m\xi^n \nabla_n \xi^a \qquad (15)$$

这就是牛顿第二定律。

考虑这样一个例子：用平滑的反对称场 F_{ab}来表示电磁场(这里反对称是指 $F_{ba} = -F_{ab}$)，如果某点有一个质量为 $m>0$、电荷为 q 和四维速度场为 ξ^a 的粒子，那么在该点由场施加给粒子的力为 $qF^a{}_b\xi^b$。如果我们用该表达式替换式(15)的左侧，那么我们就得到带电粒子在电磁场中运动的洛伦兹定律：

$$qF^a{}_b\xi^b = m\xi^b \nabla_b \xi^a \text{①} \qquad (16)$$

2.4　物质场

在经典相对论中，我们一般理所当然地认为所有发生的事件都可以由各种物质场来描述，例如材料流体和电磁场。②每个场可以用时空流形 M 中的一个或多个平滑张量(或旋量)来表示，并且假设它们都满足包含场和时空度规 g_{ab} 的场方程。

为方便目前的讨论，关于物质场最重要的基本假设如下：

> 与每个质量场$\mathcal{F}$相关的是对称的平滑张量场 T_{ab}，其特征可描述为：对所有 M 中的点 p 和所有过 p 点的未来方向的单位类时矢量 ξ^a，$T^a{}_b\xi^b$ 是相对于 ξ^a 确定的$\mathcal{F}$在 p 点的四维动量密度。

T_{ab}被称为与$\mathcal{F}$相关的能量—动量场。p 点的四维动量 $T^a{}_b\xi^b$ 能够进一步分解为相对于 ξ^a 的时间分量和空间分量，就像前一节中的重粒子的四维动量那样。在第一个分量中 ξ^a 的系数 $T_{ab}\xi^a\xi^b$ 是在 p 点相对于 ξ^a 确定的$\mathcal{F}$的能量密度；第二个分量是在 p 点相对于 ξ^a 确定的$\mathcal{F}$的三维动量密度 $T_{nb}(g^{an}-\xi^a\xi^n)\xi^b$。

其他关于物质场的假设可以通过关于能量—动量张量场的假设得到。举例如下(假设 T_{ab}与物质场$\mathcal{F}$相关)。

弱能量条件：在 M 中的任意一点上，所给出的未来方向型单位类时矢量 ξ^a，满足 $T_{ab}\xi^a\xi^b\geqslant 0$。

主能量条件：在 M 中的任意一点上，所给出的未来方向型单位类时矢量 ξ^a，满足 $T_{ab}\xi^a\xi^b\geqslant 0$，且 $T^a{}_b\xi^b$ 是类时或类光的。

守恒条件：在 M 中的所有点上 $\nabla_a T^{ab}=0$。

第一个条件断言在任意点上由任意观测者所确定的$\mathcal{F}$的能量密度是非负的。

---

①　原文有标号，无注解。

②　在这种情况下，提出的问题是我们如何(或能否)从物质场方面来谈“点粒子”。在本节中，关于该问题我们只会简单陈述。

第二个条件要求在任意点上由任意观测者所确定的$\mathcal{F}$的四维动量密度是一个未来方向的因果性(即类时的或类光的)矢量。此外该条件可以被理解为这样一个观点，即(由任何观测者确定的)能量—动量的传播速度有一个上限。它指出了2.1节中C1原理包含的一些内容，但避免了涉及“点粒子”。①

最后守恒条件断言，$\mathcal{F}$具有的能量—动量是局域守恒的。如果有两个或更多的物质场存在于时空中同一区域，那么每一个单独的场并不是必须满足该守恒条件。它们之间可能会发生相互作用。但是，一个基本的假设是，所有由这些单独的物质场组成的复合能量—动量场满足该守恒条件。能量—动量可以从一个物质场转移到另一个物质场，但它不能被创造或毁灭。

主能量条件和守恒条件有大量的推论支持上述解释。我们这里介绍两个，第一个需要做出一些初步的定义。

令(M, g_{ab})为一个确定的相对时空，令S是M的一个无时子集(即一个其中不存在$p \ll q$的点p和q的子集)。S的相关域$D(S)$是M中所有满足以下特征的点p的集：给定任意没有(过去或未来)端点的平滑因果曲线②，如果其(图像)经过点p，那么它一定穿过S。对于标准物质场，我们至少可以证明一个定理是有效的，即“S上发生了什么完全由整个$D(S)$上发生什么决定”。参见厄尔曼在本本书中的论述。这里，我们只考虑一种特殊情况。

命题6。令S是M的一个无时子集，进一步令T_{ab}为M中的一个平滑对称场，且满足主流能量条件和守恒条件。最后，假定在S上$T_{ab}=0$，那么在所有的$D(S)$上$T_{ab}=0$。

该命题的预期解释是很清楚的。如果能量—动量(在局域范围内)不能传播到类光锥以外，并且如果它是守恒的，且在S上为零，那么它必然在整个$D(S)$上为零。毕竟，它怎么可能“到达”$D(S)$上任意点呢？请注意，我们提出

---

① 这是占主导地位的能量条件的标准公式。如果我们稍微加强一点条件以使得它特别适用于大规模物质场，与C1的符合将更加接近：在M中的任意点p上，如果$T^a{}_b \neq 0$，那么对于p点的所有未来方向型单位类时向量ξ^a来说，$T^a{}_b\xi^b$是类时的。

② 令$\gamma: I \to M$为一条平滑曲线。我们说M中的一个点p，是γ的一个未来端点，如果对所有包含p的开集O，在I中存在一个s_0，使得对所有$s \in I$，如果$s \geqslant s_0$，那么$\gamma(s) \in O$，即γ的图像最终进入并保持在O中(过去的端点也可以被同样定义)。

的命题不需要预先对任意对称场T_{ab}的物理解释做任何假设。所需要的是，它应该满足上述的两个条件。对此的证明，请参见[Hawking and Ellis，1972，94]。

下一个命题[Geroch and Jang(格洛奇和张)，1975]表明，在一定意义上，如果我们假设主能量条件和守恒条件的话，那么我们就可以证明，质点粒子会穿过类时测地线的图像(回顾第2.3节中的原理P1)。关键是找到一种方法，使我们可扩展到用物质场的语言来讨论“点粒子”。

命题7。令γ：$I \to M$为一条平滑曲线。假设对于包含$\gamma[I]$的M中的任意给定的开子集O，存在一个M中的平滑的对称场T_{ab}，使得：

(1)T_{ab}满足主能量条件；

(2)T_{ab}满足守恒条件；

(3)在O外$T_{ab}=0$；

(4)在O内的一些点上$T_{ab} \neq 0$，

则γ是类时的，并且能被再参数化为一条测地线。

该命题可以用下面这种方式来转述。如果时空中的一条平滑曲线使得任意小的自由物体在其世界管中包含该曲线的图像，那么该曲线(被再参数化后)必然是一条类时测地线。实际上，我们将“点粒子”换成了更小的延展粒子的内嵌收敛序列(物体在这里被理解为是“自由的”，当其内部的能量—动量守恒时；如果物体受一个场的作用，那么就只是物体和场的整体能量—动量是守恒的)。

请注意，我们对命题进行公式化时理所当然地认为，我们可以保持背景时空结构(M，g_{ab})不变，当改变M上的场T_{ab}时。只有在T_{ab}被理解为代表测试物体，并且它对背景时空结构的影响是可以忽略的情况下，[①]这才是合理的。还需要注意的是我们并没有在一开始就假定γ是类时曲线，这是基于其他假设的。

我们这里用物质场的语言给出了这个精确的命题，至少在一定程度上得到了(关于自由质点粒子行为的)原理P1。同样，我们也能够得到(关于光的行为的)原理C2，以及在一个波长很小的极限(“几何极限”)下，关于麦克斯韦

---

① 更强的定理已经得到证明，参见[Ehlers and Geroch(埃勒斯和格洛奇)，2004]，在其中并不要求在趋向于极限的过程中的每个阶段上扩展物体的微扰影响完全消失，但要求在一定意义上，它会在极限处消失。

(Maxwell)方程组的解的命题。它有效地断言，当我们度过此极限后，电磁波的波包将被限制在沿着类光测地线(的图像)运动，参见瓦尔德[1984，71]。

现在我们考虑这样一个例子：理想流体由三个物理量来表征，分别是四维速度场 η^a、能量密度场 ρ 和各向同性的压力场 p(后两者是由相对于流体静止的“共动”观测者确定的)。在压力 p 为零的特殊情况下，人们称之为尘埃(dust，零压物质)场。理想流体的具体实例由规定 p 为 ρ 的函数的“状态方程”表征。(这里特别要排除在外的是类似各向异性压力、剪切力和黏度等复杂因素。)虽然一般假设 ρ 是非负的(见下文)，但是一些理想流体(如水可以作为一个很好的近似的例子)会产生负压力。与理想流体相关的能量—动量张量场为：

$$T_{ab}=\rho\eta_a\eta_b-p(g_{ab}-\eta_a\eta_b) \tag{17}$$

请注意，由共动的观测者所确定的(即相对于 η^a 确定的)流体在任意点的能量—动量密度矢量为 $T^a{}_b\eta^b=\rho\eta^a$。因此，我们可以等价地将 ρ 理解为相对于 η^a 的液体的能量密度，即 $T_{ab}\eta^a\eta^b$，或理解为(静止)液体的质量密度，即 $\rho\eta^a$ 的长度。(当然，这里的情况对应于2.3节中考虑的质量为 m，四维速度为 η^a 的点粒子的情况。)

在理想流体的情况下，弱能量条件(WEC)、主能量条件(DEC)和守恒条件(CC)分别为：

$$\mathrm{WEC}\Leftrightarrow\rho\geqslant 0\quad 且\quad p\geqslant-\rho;$$

$$\mathrm{DEC}\Leftrightarrow\rho\geqslant 0\quad 且\quad \rho\geqslant p\geqslant-\rho;$$

$$\mathrm{CC}\Leftrightarrow\begin{cases}(\rho+p)\eta^b\nabla_b\eta^a-(g^{ab}-\eta^a\eta^b)\nabla_b p=0\\ \eta^b\nabla_b\rho+(\rho+p)(\nabla_b\eta^b)=0\end{cases}$$

考虑联合起来等价于守恒条件的这两个方程。第一个是理想流体的运动方程。我们可以认为它是欧拉方程的相对论形式。第二个是连续性(或守恒)方程，某种意义上与经典流体力学相类似。很容易想到的是dust场($p=0$)的特殊情况。在那种情况下，运动方程退化为测地线方程 $\eta^b\nabla_b\eta^a=0$。这是有道理的。在无压力的情况下，液体中的粒子是自由粒子。并且守恒条件退化为 $\eta^b\nabla_b\rho+\rho(\nabla_b\eta^b)=0$。第一项给出了由共动的观测者确定的流体能量密度的瞬时变化率。$\nabla_b\eta^b$ 项给出了由共动的观测者确定的每单位体积内体积的瞬时变化率。

用更熟悉的符号，方程可以写成$\frac{dp}{ds}+\frac{\rho}{V}\frac{dV}{ds}=0$，或者某等价形式$\frac{d(\rho V)}{ds}=0$，这里我们用$s$表示经历的本征时间。它指出(在无压力的情况下，由共动的观测者确定的)包含在一个(无穷小)液滴中的能量仍为常量，即使它的体积发生变化。

在一般情况下，情况比较复杂，因为流体压力对(相对于特殊观测者确定的)能量有贡献，因而对所谓的"有效质量密度"也有贡献。(如果你压缩液滴，它会变重。)在这种情况下，WEC 最终除了要求$\rho\geqslant0$还要求$(\rho+p)\geqslant0$。如果我们取$h^{ab}=(g^{ab}-\eta^a\eta^b)$，那么运动方程可以表示为：

$$(\rho+p)\eta^b\nabla_b\eta^a=h^{ab}\nabla_b p$$

这是"运动学第二定律"应用于(无穷小)液滴的一个例子。等式左边为"有效质量密度×加速度"，右边为作用在液滴上的力。我们可以认为它是(由共动的观测者确定的)压力的负①梯度。再一次说明，这是有道理的。如果液滴左侧的压力大于右侧的压力，那么它将会向右移动。守恒方程中的非零项($p\nabla_b\eta^b$)是必须存在的，因为液滴的能量不是常量，它的体积会由于压力而发生变化。该公式用于计算压力为其能量所给出的贡献。

2.5 爱因斯坦方程

此外，令(M, g_{ab})为一个给定的有时间方向的背景相对时空。

相对论的一个基本思想是，时空结构并不是物理过程发生于其中的一个固定的背景，而是会参与到物理过程中。它指出任意区域的时空度规和物质场会发生动力学相互作用。这种相互作用受爱因斯坦场方程支配：

$$R_{ab}-\frac{1}{2}Rg_{ab}-\lambda g_{ab}=8\pi T_{ab} \tag{18}$$

或者受其等价方程支配：

$$R_{ab}=8\pi\left(T_{ab}-\frac{1}{2}Tg_{ab}\right)-\lambda g_{ab} \tag{19}$$

这里λ是宇宙学常数，$R_{ab}(=R^n{}_{abn})$是里奇(Ricci)张量，$R(=R^a{}_a)$是黎曼曲率

---

① 使用负号是基于我们的符号习惯。

标量，T 是缩并场 T^a_a。[①]我们先给出对式(18)的四点评论，然后再考虑一个为其提供解释的替代公式。

(1)有时"时空度规由物质分布唯一地确定"被认为是"马赫原理"的一种版本。有时候该原理可以由下面的要求得到，即"如果我们首先规定时空流形 M 中的能量—动量分布 T_{ab}，那么严格地存在一个(或最多一个)M 中的洛伦兹度规 g_{ab}，和 T_{ab}一起满足式(18)"。但这里存在一系列问题。总的来说，人们不可能在时空度规中规定能量—动量分布。(例如，我们没有能量密度的概念，除非先有一个体积的概念。)事实上，在典型情况下，度规明显出现在关于 T_{ab} 的表达式中。回顾针对理想流体的式(17)。因此，在找式(18)的解时，我们一般必须同时得到度规和物质场分布的解。

(2)给定 M 中任何平滑度规 g_{ab}，则 M 中肯定存在一个平滑对称场 T_{ab}，和 g_{ab}一起作为式(18)的一个解。通过等式的左边足以定义 T_{ab}。但这样引入的场 T_{ab}，一般来说，并不是与任何已知的物质场相关的能量—动量场，它甚至不能满足2.4节中讨论的弱能量条件。在后者给出的关于 T_{ab}的限制下，爱因斯坦方程是一个针对时空结构的完全非平凡的限制。

根据对 T_{ab}的约束条件的严格性，在经典相对论中可以分三个层次对时空结构进行讨论。在第一个层次上，我们只考虑"精确解"，即在这个解中 T_{ab}是与一个或多个已知的物质场相关的总的能量—动量场。所以，例如，可以为所有理想流体找到展示其特殊对称性的解。在第二个层次上，我们考虑更多的是所谓的"通解"，即在这些解中 T_{ab}满足一个或多个通用约束条件(弱能量条件和主能量条件就是其中的例子)。正是在这一层次上，例如彭罗斯(Penrosé)和霍金奇点定理[Hawking and Ellis, 1972]可以得到证明。最后，在第三个层次上，我们完全抛弃 T_{ab}的约束条件，并且爱因斯坦方程将不再起作用。在这个层次上可以证明许多关于整体结构的结果，例如断言在 M 为紧致的相对时空(M, g_{ab})中存在闭合的类时曲线。

(3)爱因斯坦方程中的宇宙学常数的作用一直存在争议。1917年，考虑到宇宙静态模型(那时候，这被认为是合理地表示真实宇宙所需的必然模型)，爱

---

① 我们使用在其中引力常数 G 为1，同时光速为 c 的"几何单位"。

因斯坦首次引入了($-\lambda g_{ab}$)这个项。[①]但是这样做是有问题的，特别是我们不能认为泊松(Poisson)方程(牛顿引力理论的场方程)是爱因斯坦方程的极限形式，除非 $\lambda=0$。参见下面的第(4)点。在哈勃红移(Hubble's redshift)现象给出了关于宇宙实际上在膨胀的令人信服的证据后，爱因斯坦很快就恢复了方程的初始形式。理论在观测验证之前提出了宇宙膨胀的可能性，这一点毫无疑问是理论的一个重大成就。自那以后宇宙学常数被重新确定以帮助解决理论预测和实际观测之间的差异，在(明显的)差异修正后常数又被弃了。但争论仍在继续。最近对宇宙膨胀加速率的观测令许多宇宙学家们认为，我们宇宙的 λ 值应该为正。参见[Earman，2001]的概述。

关于宇宙学常数值的断言有时被认为是关于"真空的能量—动量张量"的断言。这涉及将($-\lambda g_{ab}$)这项从式(18)的左边移到右边，并且重新将它解释为一个能量—动量场，即取爱因斯坦方程的形式为：

$$R_{ab}-\frac{1}{2}Rg_{ab}=8\pi(T_{ab}+T_{ab}^{VAC})\tag{20}$$

这里 $T_{ab}^{VAC}=\frac{\lambda}{8\pi}g_{ab}$，$T_{ab}$可以被理解为所有物质场的总能量—动量。但是 T_{ab}^{VAC} 现在被理解为与空的空间有关的残余的能量—动量。给定过某点的任意一个单位类时矢量 ξ^a，有($T_{ab}^{VAC}\xi^a\xi^b$)为$\frac{\lambda}{8\pi}$。所以，在这个重新解释下，λ(取决于参数 8π)为时空中任意一个点上的任意观测者所确定的真空能量密度。

应当指出，将 λ 作为宇宙学常数有一定的歧义(相应地作为爱因斯坦方程解的量也有一定的歧义)。我们可以引入(M，g_{ab}，T_{ab})来证明 λ 的某些(或其他)值满足该方程。或者更严格地说，我们可以一劳永逸地证明是否 λ 的某些确定值可以满足该方程，即对所有的模型(M，g_{ab}，T_{ab})都相同。实际上，我们这里有两个版本的"相对论"。参见[Earman，2003]，其中讨论了在两者之间做出选择有何利弊。

---

① 他这样做还有其他原因，参见[Earman，2001]，不过在这里我不对此进行讨论。

(4)考虑爱因斯坦方程与泊松方程(牛顿引力理论中的场方程)之间的关系是有指导意义的。

$$\nabla^2\phi = 4\pi\rho \tag{21}$$

这里，ϕ 为牛顿引力势，ρ 为牛顿质量密度函数。在 3.2 节我们关心的理论的"几何化"公式中，人们用势 ϕ 来替代微分算符，则泊松方程变为：

$$R_{ab} = 4\pi\rho t_{ab} \tag{22}$$

这里 R_{ab}是与新曲线微分算符相关的里奇张量算符，t_{ab}是时间度规。

在广义相对论(19 世纪 20 年代)之后，牛顿万有引力的几何化公式被发现。但现在，在该事实之下，我们可以作为假设调查者来考虑可以作为相对论场方程的那些候选方程，并了解牛顿理论的几何化。是否存在比试图采用或适应式(22)更自然的方法呢？在空间为空($\rho=0$)的情况下，这一策略表明方程 $R_{ab}=0$，对于 $T_{ab}=0$ 和 $\lambda=0$ 来说，当然是爱因斯坦方程式(19)。在我看来，迄今为止，这对后一个方程来说是最好的方法。从牛顿真空方程($R_{ab}=0$)开始，然后简单地保持其完整性！

在一般情况($\rho\neq0$ 的情况)下，可能没有这种简单的推断。事实上，我知道几乎没有关于完整的(有或没有宇宙学常数的)爱因斯坦方程的启发式讨论能令人信服。但是我们可以做如下的一种尝试。与式(22)最接近的形式似乎是这样一种形式：$R_{ab}=4\pi K_{ab}$，这里 K_{ab}是 t_{ab}和 g_{ab}的一个对称张量函数。K_{ab}可能包括 T_{ab}，$g_{ab}T$，$T_a{}^m T_{mb}$，$g_{ab}(T^{mn}T_{mn})$，……以及这些项的线性组合。但除前两项以外，所涉及的其他项都是 T_{ab}中的二阶或更高阶项。所以，例如，在具有能量密度 ρ 和四维速度 η^a 的零压物质场的特殊情况下，它们将包含 ρ^n 出现的可能(其中 $n\geqslant2$)。但是，很可能只有 ρ 的第一阶项出现，如果方程有合理牛顿极限的话。这表明，我们希望得到一个这种形式的场方程：

$$R_{ab} = 4\pi[kT_{ab} + lg_{ab}T] \tag{23}$$

或者等价的①，对于某些实数 k 和 l 来说：

$$R_{ab} - \frac{l}{(k+4l)}Rg_{ab} = 4\pi kT_{ab} \tag{24}$$

---

① 在式(23)中省略"a"和"b"，就得到 $R=4\pi(k+4l)T$。解出 T，并在式(23)替换 T，就得到式(24)。

令 $G_{ab}(k,\ l)$ 为等式左边的场。从守恒条件可知右侧的场是自由发散的，即 $\nabla_a(4\pi k T^{ab})=0$。所以，守恒条件和式(24)都成立，仅当：

$$\nabla_a G^{ab}(k,\ l)=0$$

但是由“比安基(Bianchi)等式”，参见瓦尔德[1984, pp. 39—40]可知：

$$\nabla_a\left(R^{ab}-\frac{1}{2}Rg^{ab}\right)=0 \tag{25}$$

后两个条件表明：

$$\left[\frac{l}{(k+4l)}-\frac{1}{2}\right]\nabla_a(Rg^{ab})=0$$

现在 $\nabla_a(Rg^{ab})=0$ 是一个不合理的约束条件。①因此初始的标量项必须为0。因此，我们得到的结论是仅当 $k+2l=0$ 时，守恒条件和式(24)都成立。在这种情况下，式(23)可以简化为：

$$R_{ab}=4\pi k\left[T_{ab}-\frac{1}{2}g_{ab}T\right] \tag{26}$$

这里仍需论证的是如果式(26)有一个适当的牛顿极限，那么 k 必须为2。为此，我们再一次考虑一个具有能量密度 ρ 和四维速度 η^a 的dust场的特殊情况。那么 $T_{ab}=\rho\eta_a\eta_b$，且 $T=\rho$。如果我们将这些值都带入式(26)并与 $\eta^a\eta^b$ 建立关系，可得：

$$R_{ab}\eta^a\eta^b=2\pi k\rho \tag{27}$$

现在，在牛顿理论中对应于四维速度场的是一个具有单位时间长度的矢量场，即 $t_{ab}\eta^a\eta^b=1$ 的矢量场 η^a。如果我们使泊松方程式(22)的几何形式与 $\eta^a\eta^b$ 建立关系，我们得到 $R_{ab}\eta^a\eta^b=4\pi\rho$。比较 $R_{ab}\eta^a\eta^b$ 的表达式和式(27)中的表达式，我们得出的结论是：在式(26)正好就是 $\lambda=0$ 时的爱因斯坦方程式(19)的情况下，$k=2$。

---

① 这意味着 R 是一个常数，并且因此，如果式(23)成立，则 T 是常数，因为式(23)意味着 $R=4\pi(k+4l)T$。然而，这又是一个关于能量—动量分布 T_{ab} 的不合理的约束。例如，在 $T_{ab}=\rho\eta^a\eta^b$ 的dust场的情况中，$T=\rho$，且因此这种约束意味着，ρ 是常数。这是不合理的，因为它排除了宇宙膨胀的可能性。请回顾第2.4节结尾处的讨论。

现在可总结如下：我们表明，如果从泊松方程的几何化版本式(22)开始，寻找相对论中的对应方程，我们很可能会得到 $\lambda=0$ 的爱因斯坦方程。值得注意的是，如果我们从包含“牛顿宇宙学常数”的式(22)的变形式

$$R_{ab}+\lambda t_{ab}=4\pi\rho t_{ab} \tag{28}$$

开始，我们将得到不限制 λ 值的爱因斯坦方程式(19)。我们可以认为式(28)是

$$\nabla^2\phi+\lambda=4\pi\rho \tag{29}$$

的几何化版本。

现在让我们把如何得到爱因斯坦方程的问题放到一边。无论如何我们已经得到了这个方程，该方程——我们现在取其为式(18)的形式——可以被理解为确定了时空曲率的特定张量测量(左边)和能量—动量张量场(右边)之间存在一种动力学联系。结果是我们可以重新表述这种联系，通过仅将标量看作是相对于任意观测者确定的方式。这种重新表述为理解方程的几何意义提供了一个特定的视角。①

令 S 为 M 中的任意一个平滑类空超曲面。②背景度规 g_{ab} 导出 S 上的一个(三维)度规 ${}^3g_{ab}$。反过来，这个度规决定了 S 上的一个导数算子，一个相关黎曼曲率张量场 ${}^3R^a{}_{bcd}$，以及一个标量曲率场 ${}^3R=({}^3R^a{}_{bca})({}^3g^{bc})$。我们关于爱因斯坦方程的重新表述将关注于某个点上的 ${}^3R$ 的值，这个点是一类特定类空超曲面族穿过的点。③

令 p 是 M 中的任意一点，令 ξ^a 为 p 点的任意一个未来方向的单位类时矢量。考虑所有经过 p 点且垂直于 ξ^a 的测地线的集合，这些曲线(的图像)扫过

① 另一种得到其几何意义的方法是通过测地偏差方程来实现的。参见萨克斯和吴[1977b，114]。

② 我们可以认为这意味着 S 是 M 的一个平滑的、嵌入的、三维子流形，其具有的性质为：任何曲线 γ：$I\rightarrow M$ 在 S 中的图像是类空的。

③ 在三维欧几里得(Euclid)空间中的一个面的案例中，相关黎曼标量曲率 ${}^2R$(取决于一个常数)只是普通高斯表面曲率。在当前背景下我们可以将 ${}^3R$ 看作一个给出了高斯曲面曲率平均值的一个更高的三维等价量。这可以表达得更精确一些，参见[Laugwitz(朗格维兹)，1965，127]。

一个平滑类空超曲面S，至少当曲线被限制在一个足够小的包含p的开集中时。[①] 见图2。我们将其称为测地超曲面(hypersurface)。(我们不能说通过p正交于ξ^a的测地超曲面，因为我们并没有限制生成的测地线能延伸多远。但是，给定任何两条，它们将被限制在一致的包含p的适当的小开集中。)

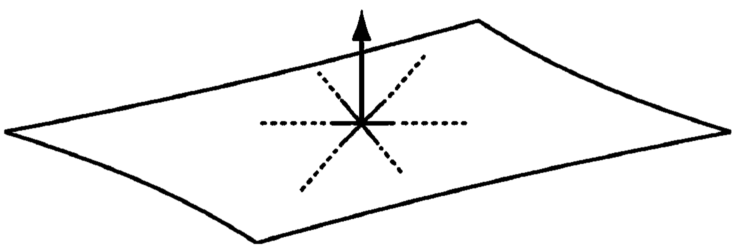


图2 穿过某个点的一个“测地超曲面”是由正交于该点处给定类时矢量的所有方向上投影测地线来建构的

抛开目前的情况不谈，从其自身的角度看，测地超曲面是很有趣的，因为(相对于过一个点的类时矢量)它们是“局域同时性截面”概念的天然的候选对象。现在的问题是根据第一高斯—科达齐(Gauss-Codazzi)方程[Wald, 1984, 258]，在p点我们有：[②]

$$^{3}R = R - 2R_{ab}\xi^a\xi^b \tag{30}$$

在这里，我们用p点处M中的(四维)黎曼标量曲率和里奇张量来表示出p点处S上的(三维)黎曼标量曲率。这样，如果爱因斯坦方程式(18)成立，那么在p点我们有：

① 更准确地说，令S_p为M_p中正交于ξ^a的类空超平面。那么对M_p中任意包含p的足够小开集O，在指数映射 exp: $O \to M$下的映像$(S_p \cap O)$是一个平滑类空超曲面。我们可以把它当作S。参见[Hawking and Ellis, 1972, 33]。

② 令ξ^a——我们使用相同的符号——为p点的初始矢量在S上的一个扩展，其扩展到一个平滑的未来导向型单位类时矢量场，该矢量场处处正交于S。那么第一高斯—科达齐方程断言，在S内的所有点上有：

$$^{3}R = R - 2R_{ab}\xi^a\xi^b + \pi_{ab}h^{ab} + \pi_{ab}\pi^{ab}$$

这里h_{ab}是S中的空间映射场$(g_{ab} - \xi_a\xi_b)$，而π_{ab}是S中的外曲率场$\frac{1}{2}\pounds_\xi h_{ab}$。但是，我们的建构保证了在$p$点$\pi_{ab}$会消失。

$$^{3}R=-16\pi(T_{ab}\xi^{a}\xi^{b})-2\lambda \tag{31}$$

我们也可以简单地反向研究从而恢复爱因斯坦方程，通过假设式(31)对于所有 p 点的单元类时向量 ξ^a(以及所有通过 p 点正交于 ξ^a 的测地超曲面)成立。因此，我们有下面的等价性：

命题8。 令 T_{ab}为 M 上的平滑对称场，并令 p 为 M 上的一个点。爱因斯坦的方程 $R_{ab}-\frac{1}{2}Rg_{ab}-\lambda g_{ab}=8\pi T_{ab}$在 p 点成立，当且仅当所有 p 点的未来方向单元类时向量 ξ^a，所有通过 p 点的测地超曲面正交于 ξ^a，以及 S 的标量曲率^3R，在 p 点满足$^3R=[-16\pi(T_{ab}\xi^a\xi^b)-2\lambda]$。

这个结果在 $\lambda=0$ 的情况下特别具有启发性。因此式(31)直接将能量密度(相对于 ξ^a 来确定)与空间曲率的标量测量(相对于 ξ^a 来确定)等同了起来。

2.6 类时曲线的汇与“公共空间”

在本节中，我们将考虑类时曲线的汇。(我们将其理解为这样一些类时曲线的集合，在集合中的一条曲线或图像经过在该区域内每个点的意义上，它们“填补”了一个时空区域。)我们将它们看作可以表征一群密集群粒子或一个流体的元素。

每一个这样的汇由未来方向的单位类时矢量场(它表示我们粒子群或流体的四维速度场)生成。在下面的讨论中我们直接研究这些生成场。

再一次，令(M, g_{ab})为我们的背景相对时空(赋于其一个时间方向)。令 ξ^a 为 M 上的一个平滑的、未来方向的单位类时向量场(或其中的一些开放子集)。最后，令 h_{ab}为由 ξ^a 确定的空间映射领域。

与 ξ^a 有关的旋转和膨胀张量场可以定义如下：

$$\omega_{ab}=h_{[a}{}^{m}h_{b]}{}^{n}\nabla_m\xi_n \tag{32}$$

$$\theta_{ab}=h_{(a}{}^{m}h_{b)}{}^{n}\nabla_m\xi_n \tag{33}$$

它们是平滑场，在两种指标上都正交于 ξ^a，并满足：

$$\nabla_a\xi_b=\omega_{ab}+\theta_{ab}+\xi_a(\xi^n\nabla_n\xi_b) \tag{34}$$

我们可以给这两个场 ω_{ab}和 θ_{ab}一个几何解释。令 η^a 为粒子 O 的世界线上的一个矢量场，它是“沿着 ξ^a 的流向传导的”，即$\pounds_{\xi}\eta^a=0$，并且在点 p 处正交

于ξ^a(这里$\pounds^{\xi}\eta^a$是η^a关于ξ^a的李导数①)。我们可以将p点的η^a看作一个空间的"连接矢量"，它将O和与其"无限接近"的相邻粒子N之间的间隔相连接起来。在p点N相对于O的瞬时速度由$\xi^n\nabla_n\eta^a$给出，但$\xi^n\nabla_n\eta^a=\eta^n\nabla_n\xi^a$(因为$\pounds_{\xi}\eta^a=0$)。因此，由式(34)以及在$p$点$\xi^a$与$\eta^a$的正交性，在该点我们有：

$$\xi^n\nabla_n\eta^a=(\omega_n{}^a+\theta_n{}^a)\eta^n \tag{35}$$

这里，我们已经简单地将相对速度矢量分解成了两个部分。第一部分，($\omega_n{}^a\eta^n$)正交于η^a(因为ω_{ab}是反对称的)。它给出N在p点相对于O的瞬时旋转速度。

在这种解释的支持下，考虑p点η^a的平方长度($-\eta^a\eta_a$)的瞬时变化率，从式(35)得到：

$$\xi^n\nabla_n(-\eta^a\eta_a)=-2\theta_{na}\eta^n\eta^a \tag{36}$$

因此，变化的计算率完全取决于θ_{ab}。假设$\theta_{ab}=0$，那么，在p点N相对于O的瞬时速度的径向分量会消失；如果$\omega_{ab}\neq0$，N仍然表现出一个相对于O的非零速度，但它只能是一个旋转速度。两个条件($\theta_{ab}=0$和$\omega_{ab}\neq0$)共同描述了"刚性旋转"。

条件$\omega_{ab}=0$本身就描述了无旋流动。通过考虑第二部分的等价公式，我们得到一种关于该条件的相当重要的见解。我们可以说场ξ^a是超曲面正交的，如果存在一个平滑的、真值映射f和g(与ξ^a的定义具有相同的域)，使得对所有的点，有$\xi_a=f\nabla_a g$。请注意，如果该条件满足，那么值为常数g的超曲面是处处正交于ξ^a的。②我们可以进一步说，ξ^a是局域超曲面正交的，如果ξ^a的限制对于每一个足够小开集来说，是超曲面正交的话。

命题9。令ξ^a是定义在M(或M的一些开放子集)上的平滑的、未来方向的单位类时向量场。则下列条件是等价的：

(1) 在任何地方$\omega_{ab}=0$;

(2) ξ^a是局域超曲面正交的。

---

① 这里我们去掉ξ的系数，以避免让人认为$\pounds_{\xi}g_{ab}$是一个三系数张量场。李导数总是相对于(逆变)向量场选取的，所以去掉该系数也不会引起歧义。

② 因为如果η^a为其中一个超曲面的矢量切线，那么$\eta^n\nabla_n g=0$。因此$\eta^n\xi_n=\eta^n(f\nabla_n g)=0$。

我们可以很直接地从条件(2)得到条件(1);①但是反过来却不简单。这是弗罗宾尼斯(Frobenius)定理的一个特例，可参见瓦尔德[1984，436]。条件(2)中"局域"的限定可能会被丢弃，比如如果 ξ^a 的域是简单地相连的。

有一个很好的图像可以说明这个命题。想象一条普通的绳子，在其被自然地拧在一起的状态下，绳子不可能以每个切片都正交于所有纤维的方式被切为无数片。但是，如果绳子先被散开，那么这种切片方式就是可能的。因此正交可切片就等价于纤维散开。这个命题将这种直观的等价性延伸至相对论中的四维"时空绳索"(即世界线的全同性)。可以断言，一致性是无旋的(即没有拧在一起)，当且仅当它至少是局域地超曲面正交的。

假设我们的矢量场 ξ^a 是无旋的，为了简单起见，假设它的定义域是简单地连接的。那么，相对于 ξ^a 或等价地相对于积分曲线的相关集来说，正交于 ξ^a 的超曲面是在给定的"时间"构建"空间"的自然候选对象。这是一个相对于专用空间的公共空间概念，它是相对于个体类时矢量或类时曲线来确定的。②顺便说一下，我们认为，对于前者来说也许最好的选择就是"测地超曲面"，参见2.5节。(在那里给定一个点 p 和一个类时矢量 ξ^a，我们取一个"经过 p 且正交于 ξ^a 的测地超曲面"为由经过 p 且正交于 ξ^a 的测地线所产生的类空超曲面。)

公共空间和专用空间之间的区别，如图3所示。在其中我们认为在闵可夫斯基时空中未来方向的类时半测地线交汇于某些特定的起始点 p，其中一条全等线 L 沿着点 q 被选出来。在 q 点处相交于 L 的专用空间是平直的类空曲面 $S_{专用}$，即生成 $S_{专用}$的度规有一个在任何地方都为零的黎曼曲率张量场 ${}^3R^a{}_{bcd}$。与此相反，过交汇点的q 处的公共空间是一个具有负常数曲率的类空超曲面$S_{公共}$。

---

① 假设 $\xi_a = f\nabla_a g$，那么：

$$\omega_{ab} = h_{[a}{}^m h_{b]}{}^n \nabla_m \xi_n = h_{[a}{}^m h_{b]}{}^n \nabla_m (f\nabla_n g)$$
$$= f h_{[a}{}^m h_{b]}{}^n \nabla_m \nabla_n g + h_{[a}{}^m h_{b]}{}^n (\nabla_m f)(\nabla_n g)$$
$$= f h_a{}^m h_b{}^n \nabla_{[m} \nabla_{n]} g + h_a{}^m h_b{}^n (\nabla_{[m} f)(\nabla_{n]} g)$$

但$\nabla_{[m}\nabla_{n]} g = 0$，因为 ∇_a 是无挠的，且最后一行的第二项也会消失，因为 $h_b{}^n \nabla_n g = f^{-1} h_b{}^n \xi_n = 0$，所以 $\omega_{ab} = 0$。

② 关于"公共空间"和"专用空间"区别的讨论请参见[Rindler(瑞德勒)，1981]和[Page(佩奇)，1983]，这个术语来自于E. A. 米尔恩(Milne)。

如果 ξ^a 是一个在任何地方都处处与汇相切的未来方向单位矢量场，并且 $h_{ab} = (g_{ab} - \xi_a\xi_b)$ 是其相关空间投影场，那么与 h_{ab} 相关的 $S_{公共}$上的曲率张量场具有如下形式 ${}^3R_{abcd} = -\frac{1}{K^2}(h_{ac}h_{bd} - h_{ad}h_{bc})$，这里 K 是沿 L 从 p 到 q 的距离（这是常数曲率 $-\frac{1}{K^2}$ 的三维流形的典型形式）。

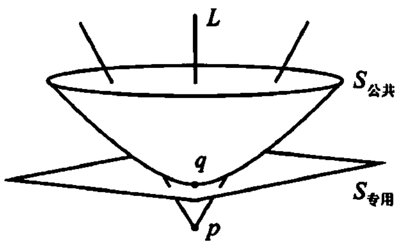


图3　p 点相对于 L 的“专用空间” $S_{专用}$，和 p 点相对于 L 所在的类时曲线的交汇点的“公共空间” $S_{公共}$

我们一直将“公共空间”看作是相对于一个类时曲线的无旋交汇点来确定的。我们可能还会在另一种意义下使用这个术语。例如，在闵可夫斯基时空下考虑“刚性旋转盘表面的几何学”。一个很好的证据是，爱因斯坦意识到，非欧几何学在相对论的发展中起到了重要作用[Stachel（施塔赫尔），1980]。我们需要问的是在何种意义上，旋转盘表面具有一个几何结构。

当然，在闵可夫斯基时空中我们可以确定地将刚性旋转盘模拟为类时曲线的汇（由于盘是二维的，汇将被局限在 M 中的一个三维类时子流形 M' 上），但正是因为盘的旋转，根据背景时空度规 g_{ab}，我们找不到与曲线处处正交的超曲面，也不能将盘的几何结构理解为是由它们导致的——或者严格地说，是由假定的超曲面与 M' 确定的交点决定的二维流形导致的——几何结构。

另一种方法是认为“空间”是由“轨迹的流形”构成的，即认为相交汇的个体类时曲线扮演汇空间点的角色，并认为在此流形下的度规是由背景时空的度规引起的。这种结构对于一个任意的类时曲线的交汇将会无效。重要的是，我们所处理的是“稳定”系统。轨迹流形上产生的度规（当结构有效时）是确定且不变的。至少对于这些系统来说，确实有效。更确切地说，预先考虑下一节的术语，如果所探讨的汇的四维速度场与基灵（Killing）场成正比的话，它是有效

的。该结构的详细介绍参见格洛奇［1971，Appendix］。

因此，关于“公共空间”，我们有两个概念。如果所探讨的汇的四维速度场是无旋的，可用其中的一个概念；如果它与一个基灵场成正比，那么就可以用另一个概念。此外，如果四维速度场是无旋的且正比于基灵场，就像我们处理的“静态”系统的案例，那么这两个关于公共空间的概念本质上是相同的。

2.7 基灵场与守恒量

令 κ^a 为 M 中的一个平滑向量场。我们说，这是一个基灵场，如果$\pounds_\kappa g_{ab} = 0$，即如果度规关于 κ^a 的李微商为零。① 这等价于要求$\pounds_\kappa g_{ab}$产生的“流映射”全是等距的，参见[Wald，1984，441]。出于这个原因，基灵场有时也被称为“平滑单参数等距族的无穷小生成元”或“无穷小对称性”。这个定义条件也可以表示为：②

$$\nabla_{(a}\kappa_{b)} = 0 \tag{37}$$

这是基灵方程。

给定 M 上的任意两个平滑向量场 ξ^a 和 μ^a，由$[\xi, \mu]^a = \pounds_\xi \mu^a$ 定义的括号或对易子场$[\xi, \mu]^a$ 也是平滑的。M 上的平滑向量场的集合形成相对于这样一个操作的李代数，即括号操作在每个时隙处是线性的；它是反对称的（$[\xi, \mu]^a = -[\mu, \xi]^a$）；且对于 M 上所有平滑向量场 ξ^a，μ^a 和 v^a，它满足雅可比（Jacobi）恒等式：

$$[[\xi, \mu], v]^a + [[v, \xi], \mu]^a + [[\mu, v], \xi]^a = 0 \tag{38}$$

事实证明，两个基灵场的括号场也是基灵场。因此，也能得到一个结论，基灵场的集具有天然的李代数结构。

关于时空中平滑对称性及其相关守恒量的讨论，自然是应用基灵场的术语合适。例如，我们可以使用这套术语来精确地得到下面的直观概念。

(1)(M, g_{ab})是稳定的，如果它有一个处处类时的基灵场；

---

① 在这里我们再一次去掉了系数 κ，从而避免让人认为$\pounds_\kappa g_{ab}$是一个三系数张量场。类似的注释适用于我们后面的括号表示法。

② 这是因为$\pounds_\kappa g_{ab} = \kappa^n \nabla_n g_{ab} + g_{nb} \nabla_a \kappa^n + g_{an} \nabla_b \kappa^n$，且$\nabla_a$ 与 g_{ab}相容，即$\nabla_n g_{ab} = 0$。

(2)(M, g_{ab})是静态的，如果它有一个处处类时且局域超曲面正交的基灵场；

(3)(M, g_{ab})是均匀的，如果其基灵场在 M 内的每一个点处跨越切空间；

(4)(M, g_{ab})在空间上是均匀的，如果它有一个平滑单元类时场 ξ^a，使得在 M 内每一个点，其基灵场跨越正交于 ξ^a 的矢量所在的三维空间；

(5)(M, g_{ab})是轴对称的，如果它有一个基灵场，满足：(i)它是处处类空的；(ii)它有一个封闭的积分曲线图像(在"轴"是点集的情况下，它可能是空的，而基灵场则消失)；

(6)(M, g_{ab})是球对称的，如果它有三个基灵场 $\overset{1}{\sigma}{}^a$，$\overset{2}{\sigma}{}^a$，$\overset{3}{\sigma}{}^a$，满足：(i)它是处处类空的；(ii)它们在每一个点上都是线性相关的，即 $\overset{1}{\sigma}{}^{[a}\overset{2}{\sigma}{}^{b}\overset{3}{\sigma}{}^{c]}=0$；(iii)它们表现出相同的对易关系，就像三维旋转群生成元那样：

$$[\overset{1}{\sigma}, \overset{2}{\sigma}]^a=\overset{3}{\sigma}{}^a, \quad [\overset{2}{\sigma}, \overset{3}{\sigma}]^a=\overset{1}{\sigma}{}^a, \quad [\overset{3}{\sigma}, \overset{1}{\sigma}]^a=\overset{2}{\sigma}{}^a \tag{39}$$

通过上一节的讨论，我们应该弄清楚了稳定时空和静态时空之间的区别(回想一下命题9)。粗略地说，在一个稳定时空中，存在一种"类时流动性"，它保持所有的时空距离不变，但是流动性可以表现为旋转(想想一个理想的漩涡)。当我们转而考虑静态时空的定义时，可以排除后者的可能性。

现在我们简单地考虑守恒量的两种类型，其中一个是质点粒子的属性，另一个是扩展客体的属性。令 κ^a 为一个任意基灵场，并令 $\gamma: I\to M$ 为一个具有单位切向量场 ξ^a 的类时曲线，其图像是一个质量 $m>0$ 的点粒子的世界线。考虑量$J=(P^a\kappa_a)$，其中 $P^a=m\xi^a$ 是粒子的四维动量。当然它不需要是 $\gamma[I]$上的常量。但是如果 γ 为测地线，它将是常量。因为在这种情况下，$\xi^n\nabla_n\xi^a=0$，因此，由式(37)得到：

$$\xi^n\nabla_n J=m(\kappa_a\xi^n\nabla_n\xi^a+\xi^n\xi^a\nabla_n\kappa_a)=m\xi^n\xi^a\nabla_{(n}\kappa_{a)}=0 \tag{40}$$

对于自由粒子来说，J 值(解释为质点粒子的属性)是恒定的。

我们将 J 看作关于 κ^a 的守恒量。如果 κ^a 是类时的，且如果由 κ^a 确定的流

映射具有平移的特征①，则 J 被称为（与 κ^a 相关的）粒子的能量。②如果它是类空的，且如果流映射具有平移的特征，则 J 被称为（与 κ^a 相关的）粒子的线性动量的组成部分。最后，如果 κ^a 是类空的，并且如果流映射具有旋转的特征，那么它被称为（与 κ^a 相关的）粒子的角动量的组成部分。

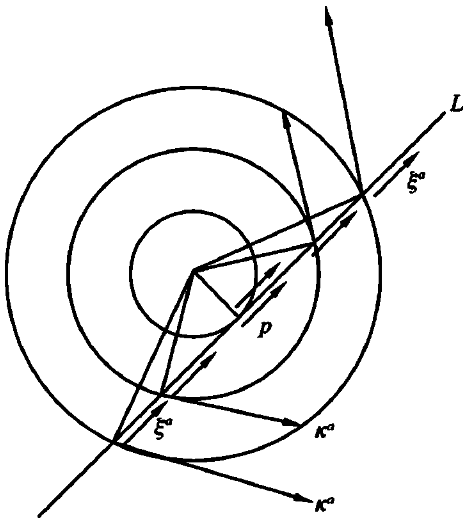


图4　κ^a 是一个旋转基灵场（它处处垂直于一个圆的半径，且与其长度成正比）。ξ^a 是 κ 线上长度固定的切向量场，它们之间的内积是不变的（等价地，κ^a 投影到直线上的长度是常量）

在头脑中形成一个特定的图景是有用的，这个图景可以帮助我们“看到”为什么自由粒子的角动量是守恒的（以这个为例）。它涉及一个类似于欧几里得平面几何中角动量的量。图4展示出欧几里得平面中的一个旋转基灵场 κ^a，测地

① 在闵可夫斯基时空中，存在一种明确的分类方式，将基灵场分为平移、空间旋转及增益（以及它们线性组合的）生成元。一般来说这种分类是不可用的。基灵场仅仅是基灵场。但有时弯曲时空中的基灵场在某些方面类似于闵可夫斯基时空中的基灵场，因此这个术语自然可以继续使用。例如，在渐近平坦时空中，我们可以将基灵场按渐近行为分类。

② 如果 κ^a 中各处都为单位长度，那么这种用法与第2.3节中的用法相符。因为在那里点粒子的能量是相对于单元类时矢量得到的，且在任何点上的能量值被认为是该单元类时矢量与粒子的四维动量矢量的内积。如果 κ^a 至少为固定长度，那么我们总是可以随时对其进行调整，从而实现用法的一致性。但是，一般来说，基灵场、类时的或其他的，并不是固定长度的，所以目前的用法必须被看作是对早前用法的推广。

线(即直线L)的图像，以及与测地线相切的切线场ξ^a。考虑到量$J=\xi^a\kappa_a$，即ξ^a与κ^a沿L的内积。确切地说，与之前方程(40)中相同的证明表明，J是沿L恒定的。①但是，这里我们可以更好地形象化这个断言。

让我们暂时去掉标记，像一个普通欧氏向量运算那样将其记作$\kappa \cdot \xi$(而不是$\xi^a\kappa_a$)。设p是L上最接近于中点的一个点，那里κ不存在。在这一点上，κ平行于ξ。随着我们沿L以任意方向离开p，κ的长度$\|\kappa\|$会增大，但矢量之间的夹角$\angle(\kappa, \xi)$也会随之增大。至少从图像上看(也很容易直接从包含相似三角形的论证来确定)κ投影到直线上的长度是常量，这个结论似乎是可能的。等价地，内积$\kappa \cdot \xi = \cos(\angle(\kappa, \xi))\|\kappa\| \|\xi\|$是常量。

这就是在相对论中我们是如何理解自由粒子的角动量守恒这一原理的。在后一种背景下，我们如何处理洛伦兹度规以及允许曲率的存在都是无关紧要的。这个陈述仍然是向量场的一个特定内积，沿测地线保持不变，而且我们仍然能认为这种恒定性是由两个因素补偿平衡所产生的。

现在让我们转向讨论第二类守恒量，这是扩展客体的一个属性。令κ^a为一个任意基灵场，并令T_{ab}为与物质场相关的能量—动量场。假设它满足守恒条件，那么$(T^{ab}\kappa_b)$就是自由发散的：

$$\nabla_a(T^{ab}\kappa_b) = \kappa_b \nabla_a T^{ab} + T^{ab}\nabla_a\kappa_b = T^{ab}\nabla_{(a}\kappa_{b)} = 0 \tag{41}$$

第二个等号是根据T_{ab}的守恒条件(2.4节中)以及T_{ab}的对称性得到的，第三个等号是基于κ^a是一个基灵场这个事实。因此，将斯托克斯(Stokes)定理应用到矢量场$(T^{ab}\kappa_b)$是很自然的。

考虑一个总能量—动量场为T_{ab}的有界系统，它处于除它之外都为空的宇宙中，那么就存在一个(可能是巨大的)类时世界管，使得T_{ab}在管外不存在(在其边界上也不存在)。令S_1和S_2为(非相交)类空超曲面，它们如图5中那样切割世界管，并令N为处于它们之间(包括边界)的管段。由斯托克斯定理，可得：

$$\int_{S_2}(T^{ab}\kappa_b)\,dS_a - \int_{S_1}(T^{ab}\kappa_b)\,dS_a$$

① 质量m并不起特别作用。

$$= \int_{S_2 \cap \partial N} (T^{ab} \kappa_b) dS_a - \int_{S_1 \cap \partial N} (T^{ab} \kappa_b) dS_a$$

$$= \int_{\partial N} (T^{ab} \kappa_b) dS_a = \int_N \nabla_a (T^{ab} \kappa_b) dV = 0$$

因此，积分 $\int_S (T^{ab} \kappa_b) dS_a$ 与如何选择和世界管相交的类空超曲面 S 无关，并且，在这个意义上说，它是一个守恒量(解释为局限于该管内的系统的一个属性)，“前”相交与“后”相交产生的值相同。同样，背景基灵场 κ^a 的特性决定了我们如何描述所关注的守恒量。如果 κ^a 是类时的，那么我们取 $\int_S (T^{ab} \kappa_b) dS_a$ 为(与 κ^a 相关的)系统的总能量，等等。

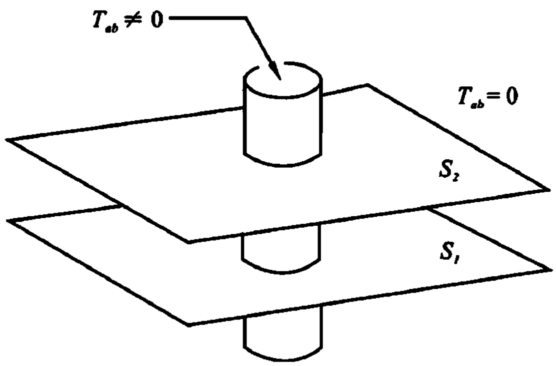


图5 世界管与类空超曲面的交点以外的(相对于背景类时基灵场的)总能量与曲面的选择无关

关于广义相对论中对称性和守恒原理的进一步讨论，请参阅布拉丁(Brading)和卡斯特拉尼(Castellani)的论述。

3. 专题讨论

3.1 闵可夫斯基时空中的相对同时性

在2.3节中我们注意到，当我们相对于一个四维速度矢量，将一个点上的矢量分解为“时间”和“空间”分量进行讨论时，我们理所当然地相信包含正交性的相对同时性的标准认同。这里我们返回来考虑这种认同的理由。

我们不是继续讨论在一个特定的点对切线空间进行分解的问题，而是很方

便在闵可夫斯基时空结构即所谓的"狭义相对论"体系下进行分析。这样做会使其更接近于一种框架，正是在这种框架下，我们对相对同时性关系的状态进行传统的讨论。

闵可夫斯基时空是由以下三个条件描述的相对时空(M, g_{ab})：(i)M是流形$\mathbb{R}^4$；(ii)(M, g_{ab})是平坦的，即g_{ab}使得黎曼曲率处处为零；(iii)(M, g_{ab})在测地线方面是完整的，即(相对于g_{ab}的)每条测地线在两个方向上都可以扩展为任意大的参数值。

根据这些条件，闵可夫斯基时空可以被等同于任意一点上的切线空间，因此在以下意义上它继承了"度规仿射空间"的结构。选择M中的任意一点o，并令V为o点的切线空间M_o，那么就存在一个从$M \times M$到V的映射$(p, q) \mapsto \overrightarrow{pq}$，它具以下两个属性：

(1)对于M中的所有p，q，r，$\overrightarrow{pq} + \overrightarrow{qr} = \overrightarrow{pr}$；

(2)对于M中的所有p，从M到V的诱导映射$q \mapsto \overrightarrow{pq}$是一个双射。①

点集M的三重组分，向量空间V，以及映射$(p, q) \mapsto \overrightarrow{pq}$形成了一个仿射空间。如果我们将$V$上由$g_{ab}$定义的内积添加到这三重组分中，那么它就变成了一个(洛伦兹)度规仿射空间。(为方便起见，我们将暂时去掉标记符号，并对于V中的v和w，用$<v, w>$替换$g_{ab}v^a w^b$。)下面，我们将理所当然地采用这种结构，也就是说我们在闵可夫斯基时空中进行讨论，并在上述意义下将其解释为度规仿射空间，这将大大简化我们的表征方式。

对于正交性，我们还可以使用一个明显的符号。给定M中的四个点p，q，r，s，我们将$<\overrightarrow{pq}, \overrightarrow{rs}> = 0$记作$\overrightarrow{pq} \perp \overrightarrow{rs}$。给定$M$上的一条直线$L$②，如果$\overrightarrow{pq} \perp \overrightarrow{rs}$对$L$上的所有点$r$和$s$成立，我们将其记作$\overrightarrow{pq} \perp L$。

---

① 如果exp是从M_o到M的指数映射，那么我们能取$\overrightarrow{pq}$为M_o中的矢量：

$$\exp^{-1}(q) - \exp^{-1}(p)$$

仿射空间的所有其他标准属性遵循这两个条件。例如$\overrightarrow{pq} = 0 \Leftrightarrow p = q$，对于$M$中所有的$p$和$q$。这里的0是$V$中的零向量。

② 在当前背景下，我们不是只有一种方式来表征一条直线。我们可以把它当作最大扩展测地线的图像，是非平凡的，即不是一个点。等价地，我们可以把它看作是一组点的形式{对于某些$\mathbb{R}$中的ϵ，则r：$\overrightarrow{pr} = \epsilon\, \overrightarrow{pq}$}，其中$p$和$q$为$M$上的任意两个(不同的)点。

现在考虑 M 中的一条类时直线 L。M 中的哪一对点 (p, q) 应该被称为"相对于 L 是同时的"？这正是我们考虑的问题。标准答案是，正是满足 $\overrightarrow{pq} \perp L$ 的那一对。

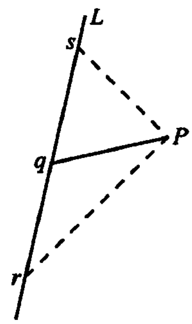


图 6 相对同时性的 $\epsilon = \frac{1}{2}$ 表征：p 和 q 相对于 L 是同时的，当且仅当 q 位于 r 和 s 的中点

在关于相对同时性的传统讨论中，标准答案往往是从"ϵ"值方面的研究得到的。这种联系很容易看到。令 p 为我们的类时直线 L 外的任意一点，那么 L 上就存在固定的（彼此不同的）两个点 r 和 s 使得 $\overrightarrow{rp}$ 和 $\overrightarrow{ps}$ 是未来方向的零向量（见图 6）。现在令 q 为 L 上的任意一点（我们认为相对于 L 它可以被判定为与 p 具有同时性）。那么 $\overrightarrow{rq} = \epsilon\ \overrightarrow{rs}$ 对于一些 $\epsilon \in \mathbb{R}$ 成立，通过简单的计算①可以表明：

$$\epsilon = \frac{1}{2} \Leftrightarrow \overrightarrow{pq} \perp \overrightarrow{rs} \tag{42}$$

所以狭义相对论中相对同时性的标准（正交）关系可以很好地被等价描述为相对同时性的"$\epsilon = \frac{1}{2}$"关系。

另一种等价的公式化涉及"单向光速"。假设光线相对于 L 以速度 c_+ 从 r 传播到 p，然后又相对于 L 以速度 c_- 从 p 传播到 s。我们在 2.3 节中看到，如果我们采取相对同时性的标准原则，那么就会得到 $c_+ = c_-$；事实上，在这种情况下，c_+ 和 c_- 都为 1；反过来也成立。因为如果 $c_+ = c_-$，那么相对于 L 来确定的话，光从 r 传播到 p 与从 p 传播到 s 的时间相同。在这种情况下，L 上的一点 q 可以被确切地认为相对于 L 与 p 是同时发生的，如果它是 r 和 s 的中点的

---

① 首先应该注意，因为 $\overrightarrow{ps}$ 和 $\overrightarrow{pr}$ 为零，有：

$$0 = \langle \overrightarrow{ps}, \overrightarrow{ps} \rangle = \langle \overrightarrow{pr} + \overrightarrow{rs}, \overrightarrow{pr} + \overrightarrow{rs} \rangle = 2\langle \overrightarrow{pr}, \overrightarrow{rs} \rangle + \langle \overrightarrow{rs}, \overrightarrow{rs} \rangle$$

由此可以证明：

$$\langle \overrightarrow{pq}, \overrightarrow{rs} \rangle = \langle \overrightarrow{pr} + \overrightarrow{rq}, \overrightarrow{rs} \rangle = \langle \overrightarrow{pr} + \epsilon\overrightarrow{rs}, \overrightarrow{rs} \rangle = \langle \overrightarrow{pr}, \overrightarrow{rs} \rangle + \epsilon\langle \overrightarrow{rs}, \overrightarrow{rs} \rangle = \left(\epsilon - \frac{1}{2}\right)\langle \overrightarrow{rs}, \overrightarrow{rs} \rangle$$，这就说明了式(42)。

话。因此，我们会再次得到相对同时性的“$\epsilon=\frac{1}{2}$”关系。

那么，采用标准关系是一种惯例还是在某种重要意义上我们被迫采用的呢?

当然，已有大量文献关注于这个问题。[①]但这不是我这篇综述的目的，我想提醒大家注意霍华德·斯坦因(Howard Stein)[1991, pp. 153—154] 的某些观点，在我看来这些观点特别深刻。他抓住了要点，即惯例的确定需要一种背景。

> 关于爱因斯坦相对同时性概念的“惯例”这个问题确实存在两个不同的方面。我们可以假设爱因斯坦本人在其研究的一开始就确定了这种惯例的地位——面临的一个问题是，如何找到一种理论，能圆满地解释它的这种地位，或者我们可以假设(例如)闵可夫斯基时空的地位——那就是，我们已经提出了一个理论，现在试图找到一个最充分的且有启发性的公式。

爱因斯坦面临的问题(部分)是，试图解释为什么我们无法检测到地球相对于“以太(aether)”的任何运动。其解决方案的一个重要部分是建议我们以一种特定的方式(即“$\epsilon=\frac{1}{2}$标准”)来思考同时性，并坚决承认这样做导致的结果。斯坦因强调，当我们并不是从爱因斯坦的初始思路出发考虑它，而是从一个成功理论的高度来理解它时，即当相对论被视为是针对时空结构不变性的论述时，这一建议看起来是多么不同。

> (对于)爱因斯坦，是否有证据确实表明光的速度在各个方向上以及对

---

① 传统位置的经典陈述可能出现在莱辛巴赫(Reichenbach)[1958]和格伦鲍姆(Grunbaum)[1973]中。格伦鲍姆最近回应了对其观点的批判。对许多参考文献中辩论的回顾，请参见贾尼斯(Janis)[2002]。原文献为论文集，未见发表，有一篇同名论文发表于2010年，Foundations of Physics, pp. 1285—1297。

> 于所有的观测者来说必须被视为相同的这个问题(自莱辛巴赫以来经常被讨论),并不是完全适当的。一个人在设计理论时,并没有责任从一开始就表示,其发展的理论是唯一能够给出证据的理论。(但)爱因斯坦的理论一旦得以发展,并已被证明在处理所有相关现象时是成功的,这种情况就完全转变了。因为我们知道,在这个理论中,只有一个“合理”的同时性概念(就这一概念来说,光速确实如爱因斯坦假设的那样),因此要找到一个替代的方法,只有当有人能成功地构建一套在本质上不同的(而且是可行的)关于运动系统的电动力学理论,而不仅仅是简单地提出一种针对同时性的不同的经验判据。事实上还没有这样一种严格的替代性理论(是在原文中就强调的)。

在本节的其余部分中,我的目标是阐述三个基本的唯一性结果,它们彼此密切相关,由此它们可以得到在什么意义上,在闵可夫斯基时空的框架内,“只有一个关于(相对)同时性的‘合理’概念”。

我们首先考虑一个有助于理解的类比。在欧几里得平面几何的一些公式中,角度之间的一致性关系与线段之间的一致性关系(以及其他适用于关于仿射结构的公式化公理的关系)一样也是基础的。但是,假设我们有一个公式,在其中角度之间的一致性关系不是基础的,而我们通过其他基础来定义角度的一致性概念。标准的角度一致性关系肯定能以这种方式定义,而且在一个明确的意义上,它是唯一合理的方式,考虑欧几里得平面上的任意两个角度。(我们首先承认,一个“角”包括两条射线,即具有一个共同起始点的半直线。)无论如何,可以推测这是唯一合理的,即将它们看作是全等的,也就是在“大小”上相等的合理的方式,如果存在欧几里得平面上的一个等距映射能将一个角映射到另一个角上。[①]因此,虽然我们这里的角度一致性概念是通过“定义”引入的,但这不是说在某个意义上,它按照惯例具有这种性质。

如果我们从闵可夫斯基时空几何的角度理解“单向光速”的话,情况就会非

① 在这种背景下,欧几里得平面到其自身的一对一映射是“等距的”,如果保证了线段之间存在一致性关系的话。

常相似。事实上，在各个方向上，以及对于所有惯性观测者来说，光在真空中的速度都一样，自然就可以表达为关于闵可夫斯基时空中角度(一种特殊类型的角)一致性的一种陈述。

我们将“光速角”表示为一个三维组(p, T, N)的形式，其中p是M上的一点，T是起始点为p的指向未来的一条类时射线，N是起始点为p的指向未来的一条类光射线（见图7）。

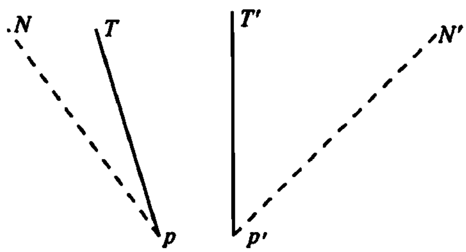


图7　闵可夫斯基时空中的全等“光速角”

那么我们就可以将单向光速的系统属性以映射$(p, T, N)\mapsto v(p, T, N)$的形式表示。我们将$v(p, T, N)$理解为在$p$点具有(半)世界线$T$的观测者赋予具有(半)世界线$N$的光信号的速度。因此，比如说，光速在各个方向上以及对于所有的惯性观测者都相等，就可以表示为下面这个条件：$v(p, T, N) = v(p', T', N')$对所有光速角$(p, T, N)$和$(p', T', N')$成立。

现在，我们很自然地可以将$v(p, T, N)$看作对角度(p, T, N)的“大小”所做的测量。如果我们这样做的话，那么，如同在欧几里得的情况中那样，我们可以期待由背景度规来决定何时两个角的大小相同。也就是说，我们可以把它们看作相等的，当且仅当闵可夫斯基时空的等距映射存在，它可以将一个角映射到另一个上。但在此标准下，所有光速角都是全等的(命题10)。因此，我们回到了(单向)光速在各个方向上，以及对所有的惯性观测者都相等这个原理，即回到了标准的相对同时性关系。

命题10。令(p, T, N)和(p', T', N')为闵可夫斯基时空中的任意两个光速角。那么就存在一个闵可夫斯基时空的等距映射ϕ，使得$\varphi(p) = p'$，

$\varphi[T]=T'$，$\varphi[N]=N'$。①

我们再一次令 L 为 M 上的一个类时直线，并令 Sim_L 为相对于 L 具有同时性的标准关系，因此 $(p, q)\in Sim_L$，当且仅当 $\overrightarrow{pq}\perp L$，对于 M 中所有的 p 和 q 成立。此外，令 S 为 M 上的任意两地关系，我们将其看作一种可能的“相对于 L 的同时性关系”。我们的第二个唯一性结果可以断言，如果 S 满足包括不变性条件在内的下面三个条件，那么 $S=Sim_L$。②

前两个条件是显而易见的：

（S1） S 是一个等价关系（即 S 是自反的，对称的，可传递的）；

（S2） 对于所有 $p\in M$ 的点，存在一个唯一的点 $q\in L$，使得 $(p, q)\in S$。

如果 S 满足条件（S1），那么它就是具有一个等价类的相关族。我们可以将它们看作（相对于 L 的）“同时性截面”。然后，条件（S2）断言，每一个同时性截面严格地与 L 相交于一个点。请注意，如果 $S=Sim_L$，那么条件（S1）和（S2）就成立。因为在这种情况下，与 S 相关的等价类是正交于 L 的超平面，它们严格地与 L 相交于一个点。

第三个条件，即不变性条件的目的是给出这样一个要求，即 S 是由背景闵可夫斯基时空几何结构及 L 自身确定的或定义的。这里不确定的一点是，时间方向是否被看作是背景几何结构的组成部分。让我们假设在那一时刻，时间方向不被考虑在内。

令 φ：$M\to M$ 为 (M, g_{ab}) 的等距映射。我们说它是一个 L 等距映射，如果它保持 L 不变，即对于 M 上所有的点 p，$p\in L\Leftrightarrow\varphi(p)\in L$。$L$ 等距映射构成的集合由下列三种类型的映射生成：（a）L 的方向（“向上”和“向下”）的变换；（b）保证 L 中每个点位置不变的空间旋转；（c）相对于与 L 正交的类空超平面的时间映像。我们说，我们的两地关系是具有 L 不变性的，如果它在上述条件下

① 所需的等距性可以表示为 $\varphi=\varphi_3\circ\varphi_2\circ\varphi_1$ 的形式，其中（i）φ_1 为将 p 变为 p' 的变换；（ii）φ_2 为（基于 p'）的放大，它将 $\varphi_1[T]$ 映射为 T'；（iii）φ_3 为关于 T' 的旋转，它将 $(\varphi_2\circ\varphi_1)[N]$ 映射到 N'。

② 在文献中，我们可以发现许多这种形式的命题，参见［Budden（巴登），1998］。我们的命题只是其中的一个例子。这里存在很多种可能性，它们严格取决于我们如何确定 S 必须满足的那些条件。证明都是完全相同的。

保持不变，即它对于等距映射φ：$M \to M$，以及所有的点p，$q \in M$，满足：

$$(p, q) \in S \Leftrightarrow (\varphi(p), \varphi(q)) \in S \tag{43}$$

现在，我们可以形式化第二个唯一性结果。①

命题 11。令L是一条类时直线，并令S是M上的一个两地关系，它满足条件(S1)和(S2)，以及L不变性。那么$S = Sim_L$。

事实证明，这里并不需要L不变性的全部要求。只要求S对所有(c)型L等距映射保持不变就足够了。②

假设现在我们需要将时间方向看作背景结构的一部分，认为其在确定S时可能是有作用的。那么我们需要重构不变性条件。我们将L等距映射φ：$M \to M$表示为(L，↑)等距映射，如果它(也)具有时间方向，即如果对所有的类时向量$\overrightarrow{pq}$，$\overrightarrow{\varphi(p)\varphi(q)}$具有与$\overrightarrow{pq}$相同的方向。且我们说$S$是($L$，↑)不变的，如果它在所有($L$，↑)等距映射下保持不变。因此，为了实现($L$，↑)不变性，$S$必须在所有(a)类和(b)类$L$等距映射下保持不变，但无须在(c)类下保持不变。

(L，↑)不变性是一个比L不变性更弱的条件，而且事实上它太弱了以至于无法提供我们想要的唯一性结果。结果是除了Sim_L之外还存在许多M上的两地关系S满足条件(S1)，(S2)，以及(L，↑)不变性。与它们相关的“同时性截面”为“平锥面”，它们在(a)类和(b)类L等距映射下保持不变，但在(c)类不是(见图8)。

但如果我们稍微改变一点要求，仍然可以得到一个唯一性结果，并认为同时性不是相对于个体类时线确定的，而是相对于平行类时线的族确定的。让我们正式取M中的一个框架为平行的类时线$\mathcal{L}$的一个集，在M上的每个点都会落在这个集中的一条类时线(且只会落在其中一条)上的这个意义下，这个集是最大的。几乎不需要做任何工作，我们就可以用这个框架而不是直线来重建我们

① 这是最接近于[Hogarth(霍加斯)，1993]中结果的一个变体。

② 证明中的关键步骤如下。设p是M中的点，由条件(S2)可得，L上存在唯一的一个点q，使得$(p, q) \in S$，令φ：$M \to M$为相对于与L正交且包含p点的超平面的一个映射。那么，$\varphi(p) = p$，$\varphi(q) \in L$，且S在φ下不变。因此$(p, \varphi(q)) = (\varphi(p), \varphi(q)) \in S$，由于$\varphi(q) \in L$，由条件(S2)中的唯一性条件，可知$\varphi(q) = q$。但由$\varphi$确定的点就是与$L$正交且包含$p$点的超平面上的点。因此，$p$和$q$都在该超平面上，且$\overrightarrow{pq}$正交于$L$，即$(p, q) \in Sim_L$。

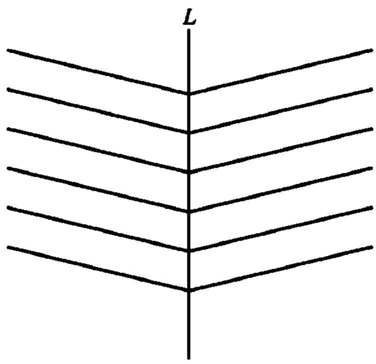


图 8　所展示的"平锥面"是与满足条件(S1)和(S2)及(L，↑)不变性(但不满足 L 不变性)的两地关系 S 相关联的同时性截面

以往的概念。

接下来，令$\mathcal{L}$为一些特定的框架。给定$\mathcal{L}$上的任意两条直线 L 和 L'，$Sim_L = Sim_{L'}$(因为 L 和 L'是平行的，正交于其中一条直线的任何一个向量必然正交于另一条)。因此，我们可以毫无歧义地使用 $Sim_{\mathcal{L}}$(相对于$\mathcal{L}$的同时性的标准关系)。令 φ：$M \rightarrow M$ 为(M，g_{ab})的等距映射。我们说，这是一个$\mathcal{L}$等距映射，如果对于所有$\mathcal{L}$上的 L，直线 $\varphi[L]$也在$\mathcal{L}$上。我们说它是一个($\mathcal{L}$，↑)等距映射，如果它额外还保持时间方向不变。

如果 L 是$\mathcal{L}$上的直线，那么$\mathcal{L}$等距映射集一定包括所有满足上述(a)，(b)，(c)类的 L 等距映射，但它还包括(d)用$\mathcal{L}$上的其他直线替代 L 的变换，以及(e)$\mathcal{L}$上 L 以外的直线上的点的等距映射。如果我们仅限于($\mathcal{L}$，↑)等距映射，那么我们会失去(c)类映射，但我们仍有(a)、(b)、(d)和(e)。这一更大的类别下的不变性足以得到一个唯一性结果。

当然，我们说 S 具有$\mathcal{L}$不变性，如果它在所有$\mathcal{L}$等距映射下保持不变；我们说它具有($\mathcal{L}$，↑)不变性，如果它在所有($\mathcal{L}$，↑)等距映射下保持不变。我们的第三个唯一性结果由下述命题得到。①

命题 12。令$\mathcal{L}$为一个框架，并令 S 为 M 上的一个两地关系。假设 S 满足条

① 它与斯皮尔特斯(Spirtes)[1981]、斯坦因[1991]和巴登[1998]提出的那些命题密切相关。

件(S1)，且$\mathcal{L}$上的一些L满足条件(S2)。此外，假设S是($\mathcal{L}$，↑)不变的，那么$S=Sim_{\mathcal{L}}$。

从命题11到命题12的过渡涉及一个协定。我们去掉S在(c)类映射下不变这个要求，增加S在(d)类和(e)类映射下不变这个要求。这是一个很好的练习，它可以验证在命题12中我们并不需要($\mathcal{L}$，↑)不变性的全部能力。我们只需要S在(a)类和(e)类映射下是不变的，或者可替换为S在(b)类和(d)类映射下不变。

同样，这些结果的诸多变体可以在很多文献中找到。例如，如果承认"基于时间(或时空)的因果理论"，那么我们将会考虑什么样的同时性关系是由(直线L以外的)闵可夫斯基时空的因果结构确定的。令φ: $M\to M$为一个双射。我们说它是因果同构的，如果它保证因果关联的对称性关系，即如果对于M上所有的点p和q，$\overrightarrow{pq}$为因果矢量，当且仅当$\overrightarrow{\varphi(p)\varphi(q)}$为1。我们说它是$L$因果同构的，($L$，↑)因果同构的，以及$\mathcal{L}$因果同构的，或($\mathcal{L}$，↑)因果同构的，如果在每一种情况下，它保持为一个特定的额外结构。由于闵可夫斯基时空的等距映射自动满足因果同构性，因此我们可以将命题11中的不变性条件替换为S在所有L因果同构下不变这个要求；我们可以将命题12中的不变性条件替换为S在所有($\mathcal{L}$，↑)因果同构下不变这个要求。

3.2 牛顿引力理论的几何化

"几何化"牛顿引力理论的方法最早由嘉当(Cartan)[1923；1924]和弗里德里克斯(Friedrichs)[1927]提出，后来经由特劳特曼(Trautman)[1965]、库佐(Künzle)[1972；1976]和埃勒斯[1981]等人逐步发展了起来。

基于以下几点原因，它是有意义的。

(1)它表明相对论的几个特征一旦被认为是其特有的特征，那么我们将不能区分它和(适当的重新公式化的)牛顿引力理论。后者也可以被列入"广义协变性"理论中，在这类理论中：(a)引力是时空曲率的一种表现；(b)时空结构是"动力学的"，即它会参与到物理事件的进展中，而不是为物理事件发展提供一个确定的背景。

(2)它帮助我们了解爱因斯坦的方程是如何"得到"的，至少在空的空间的情况下(回顾2.5节中的讨论)。它还允许我们以自由坐标下的几何语言，精确

地做出一个标准陈述，即牛顿引力理论(或者至少是其广义版本)是广义相对论的“经典极限”请参阅库佐[1976]以及埃勒斯[1981]。

(3)它阐明了牛顿引力势的规范地位。在牛顿理论的几何化公式中，我们研究一个单曲线导数算子$\overset{g}{\nabla}_a$。(在某种意义上)它可以分解为两部分——平面导数算子∇_a和一个引力ϕ——以恢复理论的标准公式。①但在特殊边界条件存在时，这种分解不是唯一的。在物理上，没有一种特别的方式，能够将粒子受到的力分为“惯性力”和“引力”。而且它们也没有任何直接的物理意义。只有它们的“总和”有意义。几何化公式的一个引人注意的特征是它将这两个规范量换成了它们的和。

(4)原因(3)中所给出的说明也解决了或者说消解了有关牛顿万有引力定理的一个旧的概念性问题，即(在宇宙学中)当该定理应用于一个假设是无限的、质量均匀分布的系统时，定理显然不成立，见马拉蒙特[1995]以及诺顿[1995；1999]。

在下文中，我们简要回顾一下牛顿引力理论的几何化公式，并重点探讨一下第(1)点和第(3)点。我们首先给出一类关于我们宇宙的时空结构(或其子区域)的新几何模型，它是足够宽泛的，既包括了标准牛顿理论的模型，也包括了牛顿理论的几何化版本。我们取经典时空的结构为$(M, t_{ab}, h^{ab}, \nabla_a)$，其中(i)$M$是一个平滑的、连续的四维微分流形，(ii)$t_{ab}$为$M$中的一个平滑的、对称的协变张量场，可以将其记作(1，0，0，0)；②(iii)h^{ab}是M中的一个平滑的、

---

① 这里理解的“标准”公式不是在教科书中的形式，而是四维时空结构的“广义协变”理论，在其中引力没有被几何化。

② t_{ab}的标记条件等价于这样的要求，即在M中的每一点上，切空间存在一个基$\overset{1}{\xi}{}^a$，…，$\overset{4}{\xi}{}^a$，使得对于{1，2，3，4}中的所有i和j，有$t_{ab}\overset{i}{\xi}{}^a\overset{j}{\xi}{}^b=0$(如果$i\neq j$)，且：

$$t_{ab}\overset{i}{\xi}{}^a\overset{i}{\xi}{}^b=\begin{cases}1, & \text{如果 } i=1\\ 0, & \text{如果 } i=2, 3, 4\end{cases}$$

同样地，(iii)中h^{ab}的标记条件要求在每一点上，切空间存在一个基$\overset{1}{\sigma}_a$，…，$\overset{4}{\sigma}_a$，使得对于{1，2，3，4}中的所有i和j，有$h^{ab}\overset{i}{\sigma}_a\overset{j}{\sigma}_b=0$(如果$i\neq j$)，且：

$$h^{ab}\overset{i}{\sigma}_a\overset{i}{\sigma}_b=\begin{cases}0, & \text{如果 } i=1\\ 1, & \text{如果 } i=2, 3, 4\end{cases}$$

对称的逆变张量场，可以将其记作(0, 1, 1, 1)；(iv) ∇_a 为 M 中的一个平滑导数算子；(v)存在以下两个条件：

$$h^{ab}t_{bc}=0 \tag{44}$$

$$\nabla_a t_{bc}=0=\nabla_a h^{bc} \tag{45}$$

我们分别将它们称为"正交性"条件和"交换性"条件。

M(像以前一样)被解释为点事件流形，t_{ab}和 h^{ab}分别被理解为 M 上的时间度规和空间度规。对象 t_{ab}，h^{ab}和 ∇_a 合起来表示由经典伽利略相对性动力学预先假设的时空结构。我们简单地回顾一下它们的作用。

在下面，令(M，t_{ab}，h^{ab}，∇_a)为一个确定的经典时空。

首先考虑 t_{ab}。给定过一个点的矢量 ξ^a，它指定一个"时间长度"$(t_{ab}\xi^a\xi^b)^{\frac{1}{2}}\geqslant 0$。矢量 ξ^a 应该被归类为类时的还是类空的，取决于其时间长度为正值还是为零。由 t_{ab}的标记可以知道过任意一点的类空向量构成的子空间是三维的。由此还能得到过每一点都存在一个唯一地取决于符号的协变矢量 t_a，使得 $t_{ab}=t_a t_b$。我们说(M，t_{ab}，h^{ab}，∇_a)在时间上是定向的，如果存在一个(定义在全部时空上的)连续矢量场 t_a，使得上面的分解式在每个点都成立。每一个这样的场定义一个时间方向：每一个类似矢量 ξ^a 相对于 t_a可以被描述为是指向未来方向的，如果 $t_a\xi^a>0$；否则它就是指向过去方向的。我们接下来假设(M，t_{ab}，h^{ab}，∇_a)在时间上是有方向的，且其选择的时间方向为 t_a 的方向。

从交换性条件可以得到 t_a 是闭合的，即 $\nabla_{[a}t_{b]}=0$。因此，至少在局域中，它必须是精确的，即对于一些平滑函数 t 具有 $t_a=\nabla_a t$ 形式。我们称所有这类函数为时间函数。如果 M 存在一个适当的全局结构，例如如果它仅仅是连贯的，那么一个在全局上定义的时间函数 t：$M\to\mathbb{R}$ 必然存在。在这种情况下，时空可以被分解为一个由全局"时间截点"(t 为常数)构成的单参数族。我们可以在一个给定的"时间"点上谈论"空间"。时间函数的不同选择，将导致时间截点的零点不同，但它们都将会导出相同的时间截点，且它们之间的时间间隔相同。

我们说一条平滑曲线是类时的(相对类空的来说)，如果其切向场在每一点上是类时的(相对类空的来说)。在下文中，除非给出相反的说明，我们都应该进一步将"类时曲线"理解为是指向未来方向的，且是由其t_{ab}长度参数化的。在

这种情况下，其切向场 ξ^a 满足归一化条件 $t_a\xi^a=1$。另外，在这种情况下，如果一个粒子恰好存在一个曲线图像，可作为其世界线。那么，在那里的任意一点上，ξ^a 被称为粒子的四维速度，$\xi^n\nabla_n\xi^a$ 被称为其四维加速度。①如果粒子的质量为 m，那么其四维加速度场满足运动方程：

$$F^a=m\xi^n\nabla_n\xi^a \tag{46}$$

这里 F^a（在其世界线的图像上）是一个类空矢量场，它表示作用于粒子上的合力。在这里，我们再次得到牛顿第二定律的新版本，回顾式(15)。需要注意的是该方程具有几何意义，因为四维加速度矢量必然是类空的。②

现在考虑 h^{ab}。它被认为是一个空间度规，但说明它是如何起作用的却有点棘手。在伽利略相对性力学中，对于类时向量，我们没有像四维速度矢量那样的空间长度的概念，因为存在这样一个概念就等同于存在绝对静止的概念（我们可以认为一个粒子处于静止状态，如果其四维速度的空间长度在各处都为0）。但是，对于类空向量，我们确实有一个空间长度的概念，比如四维加速度向量（例如，我们可以用测量杆来确定同时发生的事件之间的距离）。h^{ab} 为我们提供了其中的一种概念，而没有提供另一种。

我们不能认为一个矢量 σ^a 的空间长度为 $(h_{ab}\sigma^a\sigma^b)^{\frac{1}{2}}$，因为后者并没有被很好地定义。（由于 h^{ab} 具有一个简并特征，它是不可逆的，即不存在一个场 h_{ab} 满足 $h^{ab}h_{bc}=\delta^a{}_c$。）但如果 σ^a 是类空的，我们可以用 h^{ab} 间接地给它分配一个空间长度。结果是：(i)矢量 σ^a 是类空的，当且仅当它可以表示为 $\sigma^a=h^{ab}\lambda_b$ 的形式；(ii)如果它可以这样表示的话，那么量 $(h^{ab}\lambda_a\lambda_b)$ 是独立于 λ_a 的选择的。此外，h^{ab} 的特征保证了 $(h^{ab}\lambda_a\lambda_b)\geqslant 0$。因此，如果 σ^a 是类空的，对于相应的 λ_a 的任意选择来说，我们可以将其空间长度看作 $(h^{ab}\lambda_a\lambda_b)^{\frac{1}{2}}$。

---

① 在这里，我们认为对应于 C1：(i) 曲线是类时的，当且仅当其图像可以看作是一个点粒子的世界线时相对应的解释原理是理所当然的，在此阶段我们能够公式化的其他原理对应于 P1 和 P2：(ii)一条类时曲线可以被参数化从而成为一条（相对于 ∇_a 的）测地线，当且仅当其图像可能是一个自由粒子的世界线；(iii)时钟记录沿其世界线经历的 t_{ab} 长度（在这里，与相对性设定相反，我们只考虑大量粒子；直到我们几何化牛顿引力，我们不认为粒子是"自由的"，如果它们受到"引力"的话）。

② 根据相容性条件，$t_a\xi^n\nabla_n\xi^a=\xi^n\nabla_n(t_a\xi^a)=\xi^n\nabla_n(1)=0$。

关于经典时空我们有必要做出最后一个预备性评论。相容性条件式(45)不能确定一个唯一的导数算子，这对于我们的目的而言是至关重要的，我将会说明这一点。这是一个基本结果，即相容性条件 $\nabla_a g_{bc}=0$ 确定一个唯一的导数算子，如果 g_{ab} 是半黎曼度规，即平滑的对称场是可逆的(即非简并的)，但 t_{ab} 和 h^{ab} 都是不可逆的。

由于 h^{ab} 是不可逆的，则我们不能用其既提升指标又降低指标，但我们至少可以用它提升指标，而且这样做有时是很方便的。因此，举例来说，如果 $R^a{}_{bcd}$ 是一个与 ∇_a 相关的黎曼曲率张量场，我们可以将 $R^{ab}{}_{cd}$ 理解为 $h^{bn}R^a{}_{ncd}$ 的简化。

现在，让我们最终来考虑牛顿引力理论。在标准(非几何化的)版本中，我们引入一个平面微分算符 ∇_a 及其引力势 φ，后者可以被理解为一个 M 上的平滑实值函数。一个质量为 m 的点粒子的引力由 $-mh^{ab}\nabla_b\phi$ 给出。(请注意，这是一个由正交性条件确定的类空向量。)依照惯例提升指标，我们也可以将该矢量表示为 $-m\nabla^a\phi$。因此，如果除重力以外粒子不受任何力的作用，并且，如果它具有四维速度 ξ^a，那么它满足运动方程：

$$-\nabla^a\phi=\xi^n\nabla_n\xi^a \tag{47}$$

这里，我们只对式(46)的左边应用 $-m\nabla^a\phi$。再假设 ϕ 满足泊松方程：

$$\nabla^a\nabla_a\phi=4\pi\rho \tag{48}$$

其中，ρ 是牛顿质量密度函数(M 上的另一个平滑实值函数，表达式左侧是 $h^{ab}\nabla_b\nabla_a\phi$ 的简写)。

在该理论的几何化公式中，引力不再被设想为世界中的基本“力”，而是时空弯曲的一种表现(就像在相对论中那样)。我们不认为点粒子偏离了其自然的直线(即测地线)轨迹，而是认为它在弯曲时空中走的就是测地线。因此，我们会面临一个几何问题。从结构(M，t_{ab}，h^{ab}，∇_a)开始，我们是否能找到一个新的微分算符 $\overset{g}{\nabla}_a$，它也是与度规 t_{ab} 和 h^{ab} 相容的，使得关于原来的微分算符 ∇_a 的一条类时曲线满足运动方程式(47)，当且仅当它是关于 $\overset{g}{\nabla}_a$ 的一条测地线？下面的命题主要源于特劳特曼[1965]，该命题断言，严格地存在这样的一个 $\overset{g}{\nabla}_a$，它也记录了与 $\overset{g}{\nabla}_a$ 相关联的黎曼曲率张量场 $\overset{g}{R}{}^a{}_{bcd}$ 的一些事实。

在公式化这个命题时，我们利用了关于微分算符的如下基本事实。给定 M 中任意两个这样的算子 $\overset{1}{\nabla}_a$ 和 $\overset{2}{\nabla}_a$，存在一个唯一的平滑张量场 $C^a{}_{bc}$，它在其协变指标下是对称的，使得对于 M 中的所有平滑场 $\alpha^{a\ldots b}{}_{c\ldots d}$，有：

$$\overset{2}{\nabla}_n\alpha^{a\ldots b}{}_{c\ldots d}=\overset{1}{\nabla}_n\alpha^{a\ldots b}{}_{c\ldots d}+C^r{}_{nc}\alpha^{a\ldots b}{}_{r\ldots d}+\ldots+C^r{}_{nd}\alpha^{a\ldots b}{}_{c\ldots r}-$$
$$C^a{}_{nr}\alpha^{r\ldots b}{}_{c\ldots d}-\cdots-C^b{}_{nr}\alpha^{a\ldots r}{}_{c\ldots d} \tag{49}$$

在这种情况下，我们可以说"对应于 $\overset{1}{\nabla}_a$ 的作用，$\overset{2}{\nabla}_a$ 的作用由 $C^a{}_{bc}$ 给出"。①反之，给定 M 中的任意一个微分算符 $\overset{1}{\nabla}_a$，以及 M 中任意一个平滑对称场 $C^a{}_{bc}$，式(49)可以定义 M 中的一个新的微分算符 $\overset{2}{\nabla}_a$，可参见瓦尔德[1984, 33]。

命题13(几何化定理)。令(M, t_{ab}, h^{ab}, ∇_a)为包含平面 ∇_a($R^a{}_{bcd}=0$)的经典时空。更进一步，令 ϕ 和 ρ 为 M 上满足泊松方程 $\nabla^a\nabla_a\phi=4\pi\rho$ 的平滑真值函数。最后，令 $\overset{g}{\nabla}_a$ 为 M 上的微分算符，它与 ∇_a 的相关作用由 $C^a{}_{bc}=-t_{bc}\nabla^a\phi$ 给出。那么下面的式子成立：

(G1) (M, t_{ab}, h^{ab}, $\overset{g}{\nabla}_a$)是一个经典时空；

(G2) $\overset{g}{\nabla}_a$ 是 M 上唯一的微分算符，使得 M 上具有四维速度场 ξ^a 的所有类时曲线满足：

$$\xi^n\overset{g}{\nabla}_n\xi^a=0\Leftrightarrow-\nabla^a\phi=\xi^n\nabla_n\xi^a \tag{50}$$

(G3) 与 $\overset{g}{\nabla}_a$ 有关的曲率场 $\overset{g}{R}{}^a{}_{bcd}$ 满足：

$$\overset{g}{R}_{bc}=4\pi\rho t_{bc} \tag{51}$$

$$\overset{g}{R}{}^{ab}{}_{cd}=0 \tag{52}$$

---

① 显然，如果对应于 $\overset{1}{\nabla}_a$ 的作用，$\overset{2}{\nabla}_a$ 的作用由 $C^a{}_{bc}$ 给出，那么，反过来说，对应于 $\overset{2}{\nabla}_a$ 的作用，$\overset{1}{\nabla}_a$ 的作用由 $-C^a{}_{bc}$ 给出。在式(49)右侧的总和中，对于 $\alpha^{a\ldots b}{}_{c\ldots d}$ 中的每个指标，存在一个包含 $C^a{}_{bc}$ 的项。在每一种情况下，我们关心的指标与 $C^a{}_{bc}$ 相关，并且系数为 +1 或 -1 的项取决于我们关心的指标处于协变(在下)的位置还是逆变(在上)的位置。在一个特定坐标系统中，$C^a{}_{bc}$ 的组成部分通过(在该坐标系统中)从与 $\overset{2}{\nabla}_a$ 相关联的克里斯托菲符号中减去与 $\overset{1}{\nabla}_a$ 相关联的克里斯托菲符号来获得。

$$\overset{g}{R}{}^{[a\ \ c]}{}_{(b\ \ d)}=0 \tag{53}$$

式(51)是泊松方程的几何化版本。给出其证明过程或多或少地需要根据式(49)直接向前计算。①

我们也可以在相反的方向上进行探讨。在几何化的牛顿引力理论中，我们一开始先给出一个满足式(51)、式(52)和式(53)的曲线微分算符$\overset{g}{\nabla}_a$，以及点粒子(除“引力”外)不受力，它们穿越与$\overset{g}{\nabla}_a$有关的测地线这个原理。式(52)和式(53)作为可积性条件，确保了我们向后研究能回到关于引力势和平面微分算

---

① 这里简单介绍一下。由式(49)，有：

$$\overset{g}{\nabla}_a t_{bc}=\nabla_a t_{bc}+C^r{}_{ab}t_{rc}+C^r{}_{ac}t_{br}=\nabla_a t_{bc}+(-t_{ab}\nabla^r\phi)t_{rc}+(-t_{ac}\nabla^r\phi)t_{br}$$

由相容性条件式(45)，上式最右边第一项为零；由正交条件式(44)，第二项和第三项也为零，因为$(\nabla^r\phi)t_{br}=(h^{rm}t_{br})\nabla_m\phi$，因此$\overset{g}{\nabla}_a$是与$t_{bc}$相容的。许多相同的论证也表明，它是与$h^{ab}$相容的。这就给出了(G1)。

对于(G2)，(暂时)令$\overset{g}{\nabla}_a$为M上的一个任意微分算符，其相对于∇_a的作用由一些场$C^a{}_{bc}$来给出。设p是M中的任意一点，并令ξ^a为通过p点的任意一个类时曲线的四维速度场。然后，由式(49)，有：

$$\xi^n\overset{g}{\nabla}_a\xi^a=\xi^n\nabla_n\xi^a-C^a{}_{rn}\xi^r\xi^n$$

由此可得$\overset{g}{\nabla}_n$满足(G2)当且仅当$C^a{}_{rn}\xi^r\xi^n=-\nabla^a\phi$，或等价地有：

$$[C^a{}_{rn}+(\nabla^a\phi)t_{rn}]\xi^r\xi^n=0$$

对于所有点p处的全部未来方向的单位类时向量ξ^a成立。但是，在任意p点的未来方向的单位类时向量所在空间在那里跨越切线空间M_p，并且括号内的场在其协变指标内是对称的。因此$\overset{g}{\nabla}_n$满足(G2)当且仅当在每一处$C^a{}_{rn}=-(\nabla^a\phi)t_{rn}$。最后，对于(G3)我们应用这样一个事实，即$\overset{g}{R}{}^a{}_{bcd}$能表达为包含$R^a{}_{bcd}$和$C^a{}_{bc}$的项的总和，可参见瓦尔德[1984，184]，因此$C^a{}_{bc}$可由下式替代：

$$\begin{aligned}\overset{g}{R}{}^a{}_{bcd}&=R^a{}_{bcd}+2\nabla_{[c}C^a{}_{d]b}+2C^n{}_{b[c}C^a{}_{d]n}\\&=R^a{}_{bcd}-2t_{b[d}\nabla_{c]}\nabla^a\phi=-2t_{b[d}\nabla_{c]}\nabla^a\phi\end{aligned}$$

这里由正交条件$C^n{}_{b[c}C^a{}_{d]n}$变为0，由相容性条件$\nabla_{[c}C^a{}_{d]b}$变为$-t_{b[d}\nabla_{c]}\nabla^a\phi$。最后的等号是因为我们应用了假设$R^a{}_{bcd}=0$。式(52)和式(53)现在根据$\nabla^{[c}\nabla^{a]}\phi=0$成立，且对式(53)是基于$\nabla^{[c}\nabla^{a]}\phi=0$对任意平滑函数$\phi$成立这个事实。省略“$a$”和“$d$”得到：

$$\overset{g}{R}_{bc}=t_{bc}(\nabla_a\nabla^a\phi)$$

因此式(51)由$\nabla^a\nabla_a\phi=4\pi\rho$这个假设(以及$\nabla_a\nabla^a\phi=\nabla^a\nabla_a\phi$这个事实)得到。

符的标准方程的可能性。[1]于是我们就有了如下的恢复定理或几何化定理，该定理由主要是源于[Trautman，1965]。

命题14(恢复定理)。令$(M, t_{ab}, h^{ab}, \overset{g}{\nabla}_a)$为一个经典时空，在$M$中存在一个平滑的实值函数$\rho$满足式(51)、式(52)和式(53)。那么，至少在局域空间(也可以在全部空间中，例如，如果M仅仅是连贯的)，存在M中的一个平滑的实值函数φ和平面微分算符∇_a，使得下面所有条件成立：

(R1)$(M, t_{ab}, h^{ab}, \nabla_a)$是一个经典时空；

(R2)对于M中具有四维速度场ξ^a的所有类时曲线，几何化条件式(50)成立；

(R3)∇_a满足泊松方程：$\nabla^a \nabla_a \phi = 4\pi p$。

该定理断言了下面这种形式是存在的：给定满足特定条件的$\overset{g}{\nabla}_a$，(至少在局域空间)存在一个M中的平滑函数φ以及一个平面微分算符∇_a，使得$\overset{g}{\nabla}_a$可被看作符号对(∇_a, ϕ)的“几何化”。但是，正如在本节开始时我们声称的那样，我们并不要求其具有唯一性，除非对φ施加一些特殊的边界条件。

假设∇_a是平面的，(∇_a, ϕ)符号对满足(R1)，(R2)和(R3)。令ψ为任意平滑函数(具有与ϕ相同的域)，使得$\nabla^a \nabla^b \psi$在各处都为零，但$\nabla^b \psi$不为零。[2]如果我们设$\overline{\phi} = \phi + \psi$，并令$\overline{\nabla_a}$为一个微分算符，与其相关的$\overset{g}{\nabla}_a$的作用由$\overline{C}^a{}_{bc} = -t_{bc} \nabla^a \overline{\phi}$给出，那么$\overline{\nabla_a}$就是平面的，符号对$(\overline{\nabla_a}, \overline{\phi})$也满足(R1)，

[1] 在这里我谨慎地忽略了一些细节。几何化的牛顿引力理论具有几个不同的公式化版本。对它们之间差异的详细综述，参见贝恩(Bain)[2004]。这里介绍的版本在本质上是特劳特曼[1965]的版本。在其他较弱的公式化版本中，如库佐[1972]的版本，条件式(52)被去掉了，而且也不可能完全回到(关于引力势和平面微分算符的)标准方程，除非关于时空结构的特殊的全局性条件得以满足。

[2] 我们可以认为$\nabla^b \psi$是ψ的“空间梯度”。上述条件要求，$\nabla^b \psi$在所有的类空子流形(“时间截面”)上是常数，但不为零。

(R2)和(R3)。①

$$
\begin{tikzcd}[column sep=small]
(\nabla_a, \phi) \arrow[dr] & \\
& \overset{g}{\nabla}_a \\
(\overline{\nabla}_a, \overline{\phi}) \arrow[ur] &
\end{tikzcd}
$$

但是，由于$\nabla^b\psi$(在某些地方)并不为零，因此符号对(∇_a，ϕ)和($\overline{\nabla}_a$，$\overline{\phi}$)是$\overset{g}{\nabla}_a$的不同分解。相对于第一个符号对，一个(质量为m，四维速度为ξ^a的)点粒子加速度为$\xi^n\nabla_n\xi^a$，且受引力$-m\nabla^a\phi$作用。相对于第二个符号对，其加速度为$\xi^n\overline{\nabla}_r\xi^a=\xi^n\nabla_n\xi^a-\nabla^a\psi$，所受引力为$-m\overline{\nabla}^a\overline{\phi}=-m\nabla^a\phi-m\nabla^a\psi$。

正如本节一开始所建议的那样，我们可以认为恢复结果的这种非唯一性，正是用精确的数学语言说明了牛顿万有引力是一个规范量这个标准陈述。根据给定的论证，如果我们可以取质量为m的一个点粒子所受的力为$-m\nabla^a\phi$，那么我们同样可以等价地取其为$-m\nabla^a(\phi+\psi)$，其中ψ为满足$\nabla^a\nabla^b\psi=0$的任

---

① 从定义$\overline{\nabla}_a$的方式可以断定，($\overline{\nabla}_a$，$\overline{\phi}$)符号对满足条件(R1)和(R2)。论证该断言成立几乎和前面的一个注释中证明几何化定理中(G1)和(G2)的方法完全一样。我们必须证明$\overline{\nabla}_a$是平面的，而符号对($\overline{\nabla}_a$，$\overline{\phi}$)满足泊松方程$\overline{\nabla}^a\overline{\nabla}_a\overline{\phi}=4\pi\rho$。要做出这样的证明，我们需要先证明：(i)$\overline{R}^a{}_{bcd}=R^a{}_{bcd}$；(ii)$\nabla^a\nabla_a\psi=0$；(iii)对于$M$中的所有平滑标量场$\alpha$，$\overline{\nabla}^a\overline{\nabla}_a\alpha=\nabla^a\nabla_a\alpha$。从(ii)和(iii)能够直接得出$\overline{\nabla}^a\overline{\nabla}_a\overline{\phi}=\overline{\nabla}^a\overline{\nabla}_a\phi+\overline{\nabla}^a\overline{\nabla}_a\psi=\nabla^a\nabla_a\phi+\nabla^a\nabla_a\psi=4\pi\rho$。

由几何化定理中的(G2)中的唯一性规定，我们知道涉及∇_a的$\overset{g}{\nabla}_a$的作用由场$C^a{}_{bc}=-t_{bc}\nabla^a\phi$给出。由此可以断定$\overline{\nabla}_a$相对于$\nabla_a$的作用由$\hat{C}^a{}_{bc}=-\overline{C}^a{}_{bc}+C^a{}_{bc}=-t_{bc}\nabla^a(-\overline{\phi}+\phi)=t_{bc}\nabla^a\psi$给出。因此，论证过程几乎和证明几何化定理中(G3)的方法完全一样，我们有：

$$\overline{R}^a{}_{bcd}=R^a{}_{bcd}+2t_{b[d}\nabla_{c]}\nabla^a\psi \tag{54}$$

现在从$\nabla^a\nabla^b\psi=0$可得：

$$\nabla_c\nabla^a\psi=t_c(\xi^n\nabla_n\nabla^a\psi) \tag{55}$$

这里$t_{ab}=t_at_b$，且ξ^n为一个M中的任意未来方向型单位类似矢量场。因此，$t_{b[d}\nabla_{c]}\nabla^a\psi=t_bt_{[d}t_{c]}(\xi^n\nabla_n\nabla^a\psi)=0$。这与式(54)一起给出了(i)，而(ii)能直接从式(55)得到。最后，对于(iii)，我们注意到：

$$
\begin{aligned}
\cdots\cdots\cdots\cdots\overline{\nabla}^a\overline{\nabla}_a\alpha &= h^{ar}\overline{\nabla}_r\overline{\nabla}_a\alpha=h^{ar}\overline{\nabla}_r\overline{\nabla}_a\alpha=h^{ar}(\nabla_r\nabla_a\alpha+\hat{C}^n{}_{ra}\nabla_n\alpha)\\
&=\nabla^a\nabla_a\alpha+h^{ar}t_{ra}(\nabla^n\psi)(\nabla_n\alpha)=\nabla^a\nabla_a\alpha
\end{aligned}
$$

最后的等号是由正交性条件得到的。

意场。

3.3 由“因果结构”回到整体性几何结构

关于相对时空整体结构，可能有许多有趣的和重要的问题值得讨论——奇点的性质和意义、宇宙监督假说、“时间旅行”的可能性等。①但是，我们将仅限于对一个比较特殊的主题简单地做出一些讨论。

在我们关于相对时空结构的讨论中，我们首先用几何模型(M, g_{ab})说明几何结构的一些层次，然后用这些层次定义 M 中的(两地)关系$\ll$和$<$。②后者很自然地被解释为“因果连续性(或可达性)”关系。由此产生的问题是反向推导是否是可能的，即从(M, $\ll$)或(M, $<$)开始，然后将 M 解释为一个基本点集，再回到我们一开始给出的几何结构。这里的问题是由长期关注于时间或时空中的“因果理论”的那部分哲学家提出来的。所讨论的问题还集中在由拉斐尔·索金(Rafael Sorkin)及其同事发展出来的关于量子引力的一个特定方法上。参见例如[Sorkin，1995；2005]。

这里有一种方法可以使问题更加精确化。为方便起见，我们探讨$\ll$的关系。

令(M, g_{ab})和($\overline{M}$, $\overline{g}_{ab}$)为(有时间方向)的相对时空。我们说，它们的基础点集之间的一个双射 φ: $M \to \overline{M}$ 是因果同构的，如果对于 M 中所有的 p 和 q：

$$p \ll q \Leftrightarrow \varphi(p) \ll \varphi(q) \tag{56}$$

现在我们要问：是否因果同构存在一个同胚、一个微分同胚或者保形等距？③

如果不进一步对(M, g_{ab})和($\overline{M}$, $\overline{g}_{ab}$)做出限制，那么关于上述三个问题

---

① 厄尔曼[1995]对其中的许多主题进行了全面的回顾。关于奇点问题，也可以参见[Curiel(库列尔)，1999]。

② 回顾一下，如果存在一个从 p 到 q 的未来方向型类时曲线，那么 $p \ll q$ 成立；如果存在一个从 p 到 q 的未来方向型因果曲线，那么 $p < q$ 成立。

③ 我们预先知道，因果同构不需要是(完全)等距的，因为流形 M 上的保形等价指标 g_{ab} 和 $\Omega^2 g_{ab}$ 确定同样的关系$\ll$。我所能要求的最好情况是，它是保形等距的，即它保证指标为一个保形因子的微分同胚。

的答案肯定都是“否定的”。除非时空的“因果结构”(即由$\ll$确定的结构)是合理的，否则它不会提供任何有用的信息。例如，我们说时空是因果简并的，如果对于所有点p和q，$p\ll q$。两个因果简并时空量之间的双射是因果同构的。但我们肯定能找到底层流形具有不同拓扑结构的因果简并时空，例如哥德尔(Gödel)时空和卷曲的闵可夫斯基时空。

存在一个与此有关的“因果关系”的层次结构，参见[Hawking and Ellis, 1972, Section 6.4]。他们分别以不同的严格程度，要求不可能存在闭合的或“几乎闭合的”类时曲线。这里给出三种提法。

时间次序：不存在闭合的类时曲线(等价地，对所有p，不存在$p\ll p$的情况)。

未来(或过去)可区分性：对于所有点p，以及包含p的所有足够小的开集O，不存在未来方向型(或过去方向型)类时曲线，起始于点p，离开O，还会再返回到O中。

强因果性：对于所有点p，以及包含p的所有足够小的开集O，不存在类时曲线，起始于点p，离开O，还会再返回到O中。

很显然，强因果性包含了未来可区分性和过去可区分性，而且这两个区分性条件(分别都)涵盖了时间次序。标准的例子，如霍金和埃利斯[1972]中的例子表明，上述包含关系反过来不成立，而且两个可区分性条件之间也不相互包含。

“未来可区分性”和“过去可区分性”这两个名称很容易解释。对于所有的p，令$I^+(p)$为集合$\{q: p\ll q\}$，令$I^-(p)$为集合$\{q: q\ll p\}$。那么未来可区分性就等价于下面这个要求，即对于所有的p和q，有：

$$I^+(p)=I^+(q)\Rightarrow p=q$$

用I^-替代I^+就得到等价于过去可区分性的要求。

我们在这里提及这些，是因为事实证明，如果我们仅限于关注未来可区分性和过去可区分性时空，那么对于上述三个问题我们就都能得到肯定的回答。

命题15。令(M, g_{ab})和$(\overline{M}, \overline{g}_{ab})$为(有时间方向的)具有未来可区分性和

过去可区分性的相对时空，并令φ：$M \to \bar{M}$是因果同构的。那么φ就是一个微分同胚，并且它保证g_{ab}是一个保形因子，即$\varphi_{*} g_{ab}$保形等价于$\bar{g}_{ab}$。

马拉蒙特[1977]对此给出了证明。他还给出了一个反例，以证明如果对因果结构的初始约束减弱到仅具有过去可区分性或未来可区分性，那么这个命题就不为真。

致谢

非常感谢杰里米·巴特菲尔德、埃里克·库列尔、约翰·厄尔曼以及克里斯·斯密尼克(Chris Smeenk)对前几稿所做出的评论。

参考文献

[Bain, 2004] J. Bain. Theories of Newtonian gravity and empirical distinguishability. *Studies in the History and Philosophy of Modern Physics*, 35: 345-376, 2004.

[Budden, 1998] T. Budden. Geometric simultaneity and the continuity of special relativity. *Foundations of Physics Letters*, 11: 343-357, 1998.

[Cartan, 1923] E. Cartan. Sur les variétés a connexion affine et la théorie de la relativité généralisée. *Annales Scientifiques de l'Ecole Normale Supérieure*, 40: 325-412, 1923.

[Cartan, 1924] E. Cartan. Sur les variétés a connexion affine et la théorie de la relativité généralisée. *Annales Scientifiques de l'Ecole Normale Supérieure*, 41: 1-25, 1924.

[Coleman and Korté, 2001] R. Coleman and H. Korté. Hermann Weyl: Mathematician, physicist, philosopher. In E. Scholz, editor, *Hermann Weyl's Raum-Zeit-Materie and a General Introduction to his Scientific Work. Birkhäuser Verlag*, 2001.

[Curiel, 1999] E. Curiel. The analysis of singular spacetimes. *Philosophy of Science*, 66: S119-S145, 1999.

[Earman, 1995] J. Earman. *Bangs, Crunches, Whimpers, and Shrieks.* Oxford University Press, 1995.

[Earman, 2001] J. Earman. Lambda: The constant that refuses to die. *Archives for History of Exact Sciences*, 55: 189-220, 2001.

[Earman, 2003] J. Earman. The cosmological constant, the fate of the universe, unimodular gravity, and all that. *Studies in the History and Philosophy of Modern Physics*, 34: 559-577, 2003.

[Ehlers and Geroch, 2004] J. Ehlers and R. Geroch. Equation of motion of small bodies in relativity. *Annals of Physics*, 309: 232-236, 2004.

[Ehlers, 1981] J. Ehlers. Über den Newtonschen Grenzwert der Einsteinschen Gravitationstheorie. In J. Nitsch, J. Pfarr, and E. W. Stachow, editors, *Grundlagen Probleme der Modernen Physik*. Wissenschaftsverlag, 1981.

[Friedrichs, 1927] K. Friedrichs. Eine Invariante Formulierung des Newtonschen Gravitationsgesetzes und der Grenzüberganges vom Einsteinschen zum Newtonschen Gesetz. *Mathematische Annalen*, 98: 566-575, 1927.

[Geroch and Jang, 1975] R. Geroch and P. S. Jang. Motion of a body in general relativity. *Journal of Mathematical Physics*, 16: 65-67, 1975.

[Geroch, 1971] R. Geroch. A method for generating solutions of Einstein's equation. *Journal of Mathematical Physics*, 12: 918-924, 1971.

[Geroch, unpublished] R. Geroch. General Relativity. Unpublished Lecture Notes from 1972, University of Chicago, unpublished.

[Grünbaum, 1973] A. Grünbaum. *Philosophical Problems of Space and Time*. Reidel, 2nd enlarged edition, 1973.

[Grünbaum, forthcoming] A. Grünbaum. David Malament and the conventionality of simultaneity: A reply. In *Collected Papers*, volume 1. Oxford University Press, forthcoming.

[Hawking and Ellis, 1972] S. W. Hawking and G. F. R. Ellis. *The Large Scale Structure of Space-Time*. Cambridge University Press, 1972.

[Hogarth, 1993] M. Hogarth. Standard simultaneity is unique in metrical Minkowski spacetime: A generalization of Malament's theorem. unpublished, 1993.

[Janis, 2002] A. Janis. Conventionality of simultaneity. In E. Zalta, editor, *Stanford Encyclopedia of Philosophy*, 2002. URL = http://plato.stanford.edu/archives/fall2002/entries/spacetime-convensimul/.

[Künzle, 1972] H. Künzle. Galilei and Lorentz structures on space-time: Comparison of the corresponding geometry and physics. *Annales Institute Henri Poincaré*, 17: 337-362, 1972.

[Künzle, 1976] H. Künzle. Covariant Newtonian limits of Lorentz space-times. *General*

Relativity and Gravitation, 7: 445-457, 1976.

[Laugwitz, 1965] D. Laugwitz. *Differential and Riemannian Geometry.* Academic Press, 1965.

[Malament, 1977] D. Malament. The class of continuous timelike curves determines the topology of spacetime. *Journal of Mathematical Physics*, 18: 1399-1404, 1977.

[Malament, 1995] D. Malament. Is Newtonian cosmology really inconsistent? *Philosophy of Science*, 62: 489-510, 1995.

[Malament, unpublished] D. Malament. Lecture Notes on Differential Geometry. available at http://www.lps.uci.edu/home/fac-staff/faculty/malament/FndsofGR.html, unpublished.

[Norton, 1995] J. Norton. The force of Newtonian cosmology: Acceleration is relative. *Philosophy of Science*, 62: 511-522, 1995.

[Norton, 1999] J. Norton. The cosmological woes of Newtonian gravitation theory. In H. Goenner, J. Renn, J. Ritter, and T. Sauer, editors, *The Expanding Worlds of General Relativity.* Birkhäuser, 1999.

[O'Neill, 1983] B. O'Neill. *Semi-Riemannian Geometry.* Academic Press, 1983.

[Page, 1983] D. Page. How big is the universe today? *General Relativity and Gravitation*, 15: 181-185, 1983.

[Reichenbach, 1958] H. Reichenbach. *The Philosophy of Space and Time.* Dover, 1958.

[Rindler, 1981] W. Rindler. Public and private space curvature in Robertson-Walker universes. *General Relativity and Gravitation*, 13: 457-461, 1981.

[Sachs and Wu, 1977a] R. K. Sachs and H. Wu. General relativity and cosmology. *Bulletin of the American Mathematical Society*, 83: 1101-1164, 1977.

[Sachs and Wu, 1977b] R. K. Sachs and H. Wu. *General Relativity for Mathematicians.* Springer Verlag, 1977.

[Sorkin, 1995] R. D. Sorkin. A specimen of theory construction from quantum gravity. In J. Leplin, editor, *The Creation of Ideas in Physics.* Kluwer, 1995.

[Sorkin, 2005] R. D. Sorkin. Causal sets: Discrete gravity. In A. Gomberoff and D. Marolf, editors, *Lectures on Quantum Gravity (Proceedings of the Valdivia Summer School)*. Plenum, 2005.

[Spirtes, 1981] P. Spirtes. Conventionalism in the philosophy of Henri Poincaré. unpublished Ph. D. thesis, University of Pittsburgh, 1981.

[Stachel, 1980] J. Stachel. Einstein and the rigidly rotating disk. In A. Held, editor, *General Relativity and Gravitation. One Hundred Years After the Birth of Albert Einstein*, volume I. Plenum, 1980.

[Stein, 1991] H. Stein. On relativity theory and the openness of the future. *Philosophy of Science*, 58: 147-167, 1991.

[Trautman, 1965] A. Trautman. Foundations and current problems of general relativity. In S. Deser and K. W. Ford, editors, *Lectures on General Relativity*. Prentice Hall, 1965.

[Wald, 1984] R. Wald. *General Relativity*. University of Chicago Press, 1984.

[Weyl, 1921] H. Weyl. Zur Infinitesimalgeometrie: Einordnung der projecktiven und konformen Auffassung. *Nachrichten der Königlichen Gesellschaft der Wissenschaften zu Göttingen; Mathematisch-Physikalische Klasse*, pages 99-112, 1921.

[Weyl, 1950] H. Weyl. *Philosophy of Mathematics and Natural Science*. Princeton University Press, 2nd edition, 1950.

第四章　非相对论量子力学

迈克・迪克森

本章主要介绍源自于或关于非相对论量子理论的一些最重要的哲学问题和基础问题，包含六个主要部分。第 1 节介绍量子理论，包括一些构建，并给出许多哲学问题和基本问题所需要的重要数学结果。这一节最长，同时也最为重要，因为它将为认真的读者提供理解和评价与非相对论量子理论相关的浩瀚文献所需的背景。文献确实非常多——因而绝无可能把它们总结成像本章一样长的章节。因而，在接下来的五节中，我将考虑一些更为重要的问题：量子理论形式体系的基本特征和经验内容、量子不确定性、测量问题和非定域性。当然，除此之外还有许多其他问题可以讨论，而且一些最近的动向也值得考虑。遗憾的是在这里我们没有足够的篇幅完成这些。不过，认真领会本文中的材料是我们了解这些其他问题的起点。

在上述五个问题中，对前两个问题的讨论相对少一些，尤其是在英美哲学文献中。因而，这两节相对于后三节而言，篇幅会长于某些读者的预期。这并不意味着问题的相对重要性，只是我在试图弥补某些特定领域内相对匮乏的讨论。

这里出现的大多数内容——尤其是第 4、第 5 和第 6 节——主要是我对可以在许多地方找到的标准材料的综述。对

此我不打算提供广泛的书目信息。事实上，我也一直试图将参考书目控制到最少。本文并不打算成为对该领域内工作的概况总结，更不是宽泛的附有注解的文献目录。我们鼓励读者寻找另外的资源以补充这里给出的简要说明，而且这样的资源有许多。

最后一节是简短的数学附录，给出了基本的定义和结论，其中大部分是关于希尔伯特(Hilbert)空间和群的理论。它或许满足以下两个目的之一(这将取决于读者)：对在别处学到的概念进行简要的温习，或者促使读者去别处学习这些概念。完全不熟悉这些概念的读者想要仅仅从这里所讲的内容就理解它们是不大可能的。至于与附录相关的小节，会在文章中合适的位置给出参考书目。

虽然在一些要点上，我尽力做到是严谨的，然而大多数时候这里的讨论只是部分严谨的，我有时会指出要做到完全严谨还需要什么。这里也鼓励读者就数学细节翻阅参考文献，并且要时刻牢记这里的大多数讨论并非要完全地在数学上严谨，虽然我也不希望它们有误导性。

1. 理论

本节是对量子理论形式体系的介绍。在对这种方法做简短评述(第1.1节)之后，我将介绍该形式体系的主要要素(第1.2节)，随后给出一个简单但重要的例子(第1.3节)。然后，我将介绍常用的“狄拉克”(Dirac)符号(第1.4节)，通过考虑变换(群)在理论中的作用(第1.5节)——包括动力学变换(运动方程)和最后(第1.6节)对所产生的哲学问题的简短概述——来得出结论。这一节比后面的几节更强烈地依赖于数学附录(第7节)中的内容，因而在适当的时候我会给出参考书目。

1.1 缘何从形式体系开始

为什么对一个物理理论的说明要开始于它的形式体系而不开始于它的基本物理见解，或基本物理原理？后一种方法的问题是，在量子理论情形下，有关它的基本物理见解或基本物理原理是什么还没有取得有意义的一致性。有人主张塌缩假设(随后会讨论到)是理论的核心，还有人主张塌缩假设必须被清除出理论。有人认为理论根本上是非决定论的，而其他人认为仅在一种更深层的决

定论下理论才有意义。有人认为具有确定位置的"粒子"这一熟悉的概念是理论的弊端，而另外的人则认为只有把这样一个概念看作是基本的理论才有意义。

现在，这些不同观点的拥护者们也确实同样试图去拥护理论的不同表述，但是他们并未表明在他们选择之外的表述是错误的，他们也许只是强调其中的错误之处。（事实上，这里将会给出的标准体系在经验上很成功这一点是没有争议的；不同观点的拥护者们最终必须用他们自己的术语来解释这种成功。）因此，虽然在我们讨论的开始，选择单一的表述可能会在某种程度上使我们的观点有倾向性；但是，不像对基本物理见解或基本物理原理的选择，它并不会对核心问题进行判断。

1.2 标准形式体系

现在我们开始对标准形式体系的常规理解进行简短概述，随后我还将对其进行补充与推广。（我并不期望读者仅仅通过阅读这一节就能对该体系的各个方面都有深刻的理解。）

1.2.1 希尔伯特空间

量子力学形式体系通常是用希尔伯特空间理论（7.1 节）中的术语来理解的。一个希尔伯特空间是一个矢量空间（7.1.1 节），它拥有内积（7.1.3 节），也包括由此内积定义的模（7.1.4 节）。一个标准的例子是复数的（模）平方可积序列空间$\mathcal{L}^2$。在该空间中，两个矢量$(x_1, x_2, \cdots)$和$(y_1, y_2, \cdots)$的内积是$\sum_{n=1}^{\infty} x_n^* y_n$。另一个标准例子是在$\mathbb{R}^N$上的（模）平方可积的、勒贝格（Lebesque）可测量的复值函数空间$L^2(\mathbb{R}^N)$，其中我们把两个函数看作是相同的（它们表示同一矢量），它们在当（且仅当）一组勒贝格测度（7.5.4 节）为零时不同。这里，两个矢量$f(x)$和$g(x)$的内积是$\int f^*(x)g(x)dx$，其中$f(x)$和$g(x)$是它们各自等价类中的任意一个。①

① 对于熟悉量子理论的读者而言：空间$\mathcal{L}^2$是海森堡（Heisenberg）"矩阵力学"中用到的空间，而空间$\mathcal{L}^2(\mathbb{R}^N)$是薛定谔（Schrödinger）"波动力学"中用到的空间。作为希尔伯特空间ℓ^2和$L^2(\mathbb{R}^N)$是同构的，意味着两种理论本质上是相同的。

1.2.2 可观测量

理论的“可观测量”——物理量或特征，它们的值或存在性原则上至少可以测量或“观测到”——通常用希尔伯特空间上的自伴算符来表示(7.2.1 节和 7.2.3 节)。(表示的类型——即哪些算符表示哪些可观测量——取决于所描述的物理情形。)通过谱定理(接下来会讨论)，可以把每个可观测量与投影算符的一组谱系等同起来，可观测量本质上是由从可观测量可能值的波莱尔(Borel)集(7.5.5 节)向谱系中元素的映射给出的。本节将简要回顾这些观点。

1.2.2.1 正定算符值测量 采用可观测量的一个更宽泛的概念，即“正定算符值测量”(POVM)通常很有用。在此方法下，我们开始于可观测量的一组“可能值”，它在最一般的情况下被表示为一个局域紧致的拓扑空间 S(7.5.1 节)。对于我们所感兴趣的大多数情形，S 是实数的一个子集，或者可以通过再处理使其成为这样一个集合。

从 S 的波莱尔子集到某个希尔伯特空间$\mathcal{H}$(7.2.2 节)上有界算符的映射 E：$\mathcal{B}(S)\to\mathcal{B}(\mathcal{H})$是一个 POVM，仅当对于这些子集中任何不相交序列 $\Delta_n\subseteq S$，有：

$$E(\Delta_n)\text{对于所有的 } n \text{ 都是正算符;} \tag{1}$$

$$E(S)=\mathbb{I}\text{，是}\mathcal{H}\text{上的单位元;} \tag{2}$$

$$E(\cup_n\Delta_n)=\sum_n E(\Delta_n) \tag{3}$$

在(1)中，若对于所有的 $v\in\mathcal{H}$，有 $<v, Ev>\geqslant 0$，则算符 E 是正定的。$\mathcal{H}$上的正定算符记作$\mathcal{B}(\mathcal{H})^+$。(3)中预期的收敛在$\mathcal{H}$上的弱算子拓扑(weak operator topology)中(7.5.3 节)。另外，当 $n\neq m$ 时，若 $E(\Delta_n\cup\Delta_m)=E(\Delta_n)E(\Delta_m)$，则在 E 映像中的一切都是投影算符；E 被称为一个“投影值测量”(PVM)；且族$\{E(\Delta_n)\}$是一个“谱族”。在这种情况下，$E(\Delta_n)$是相互正交的意味着当 $n\neq m$ 时有 $E(\Delta_m)E(\Delta_n)=0$(零算符)，我们写作 $E(\Delta_m)\perp E(\Delta_n)$。

我们可以从任意的 PVM 复原自伴算符 E。若 $S\subset\mathbb{R}$ 的基数(cardinality)是有限的($S=\{s_1, \cdots, s_N\}$)，则复原就很直接：①

---

① 在(4)中 $E(s_n)$应该严格地写为 $E(\{s_n\})$，因为 $E(\,)$作用于波莱尔集，但这样写太繁复，同时 $E(s_n)$的意义应该足够清晰。在其他地方我也遵循同样的约定。

$$F=\sum_{n=1}^{n}s_nE(s_n) \tag{4}$$

即算符 F 是(相互正交的)投影 $E(s_n)$ 的加权和，权重是可观测量即 S 元素的“可能值”。若 S 是可数无限的(countably infinite)，情形仍然相同，虽然人们肯定会担忧其收敛性。若 S 是不可数无限的，则加权和变为积分，问题就变得相当复杂。但在这些情形中得到的算符 F 都是自伴的。

1.2.2.2 谱定理 谱定理表述了与式(4)相反的过程。同样，有限的情形是最简单的，在那里每个自伴算符 F 都可以写作：

$$F=\sum_n s_nP_n \tag{5}$$

其中 s_n 是实数，P_n 是相互正交的投影。因而一个谱族确定了一个自伴算符，且一个自伴算符确定了一个谱族。因此，PVMs 的形式体系很快与读者或许更加熟悉的、当然也在物理学中广泛应用的一种形式体系(用自伴算符表示的)联系起来，因而 PVMs 形式体系可以看作是用 POVMs 表示的更一般形式体系的特殊情形。无限维空间的情形在概念上类似，但在数学上更加复杂。

注意到对应于投影 P_n 子空间($\mathrm{ran}P_n$，即 P_n 的“范围”)中的每个矢量是 F 的一个本征矢量(第 7.2.1 节)，$\mathrm{ran}P_n$ 因而常常称为 F 的“本征空间”。当 F 的本征空间都是一维的时，F 被称为“最大的”，“最大的”意义会在下文中明确阐述。

最后，注意 F 的谱投影部分地为空间定义了一个正交的，确切地说是正交归一(orthonormal，7.1.4 节)的基矢。在每个 P_n 中，选择一组相互正交且归一化的矢量，$\{e_{n,m}\}_{m=1}^{\dim(P_n)}$。对 F 的核(kernel，F 映射到 0 的子空间，记作 $\ker F$)进行同样处理得到的(即所有这些集合的合集)是一组正交基(若 F 是最大的，则该正交基事实上为基矢中的元素提供了常数乘子)。即使当该基矢不是由 F 唯一地确定(因为它是非最大的)时，我们将把这样的基矢当作“由 F 决定的基矢”。

1.2.3 态

1.2.3.1 概率 用 POVMs 表述的形式体系(还有 PVMs 的特殊情形)描述了一种概率理论，这是由于它为可观测量(波莱尔集合)的值提供了概率，或提供了(等价的且有时更为方便的)可观测量的期望值。我将把概率看作基本的；期望值因而可以以常规方式由对 F 可能值 f_n 的概率测度给出：

$$\mathrm{Exp}(F) = f_1 \Pr(f_1) + f_2 \Pr(f_2) + \cdots \tag{6}$$

正如我们上面注意到的，与其直接考虑可观测量的可能值，我们也可以考虑相应的(谱)投影，认为它在给定的物理情形中表示了那些值。

在投影算符上定义的概率测度 p，当 $P_1 \perp P_2$ 时至少应该满足 $p(P_1 + P_2) = p(P_1) + p(P_2)$。稍后我会阐释该条件。其中的基本观点是，它对应于柯尔莫哥洛夫(Kolmogorov)概率理论中常见的“可加性公理”——见 7.5.6 节。更具体地，现在仅考虑 PVMs 情形，我们要求对希尔伯特空间上投影的概率测度是一个从投影到区间[0，1]的映射 p，其中 p 在相互正交的投影集上是可数可加性的(countably additive)。

处于一个 POVM(而非 PVM)映像中的，算符的可数可加性究竟意味着什么，这是一个有些微妙的问题。特别地，一般而言 POVM 映像中的算符——通常它们被称为“效应”(effect)——并不对应于子空间，且正交性概念并不适用。不过，存在一种对该概念的自然推广。注意到在 PVM 映像中，对于投影 $\{P_i\}$而言，$\mathbb{I} - \sum_i P_i$ 是一个投影(或可能是零算符)的条件等价于$\{P_i\}$是相互正交的条件。① 在正定算符情形中类似的条件是，在 POVM 映像中，对于效应 $\{E_i\}$而言，若$\mathbb{I} - \sum_i E_i$ 是正的(或是 0)，则有 $\Pr(\sum_i E_i) = \sum_i \Pr(E_i)$。

1.2.3.2　态矢量和波函数

1.2.3.2.a　态矢量　归一化的矢量决定着对投影的概率测度，通过：

$$P \text{ 的概率由 } v := \Pr_v(P) := \langle v,\ Pv \rangle \text{ 给出} \tag{7}$$

常见到的表达是 $|\langle \psi,\ v \rangle|^2$，其中 ψ 是来自 $\mathrm{ran}P$ 的归一化矢量。这两种表达是等价的。注意，由矢量 v 和 $e^{i\phi}v$(其中 ϕ 是一个实数)产生的概率是相同的。人们常说“总(overall)相位并不影响概率”。由态 v 给出的自伴算符 F 的期望值是：

$$P \text{ 的期望值由 } v := \mathrm{Exp}_v(F) := \langle v,\ Fv \rangle \text{ 给出} \tag{8}$$

注意一个投影的期望值同样是它的概率。

---

① 论证概要：写出$(\mathbb{I} - \sum_i P_i)$；展开；论证当$\{P_i\}$相互正交时，则结果是$\mathbb{I} - \sum_i P_i$；论证(利用投影是正定的这一事实——这一部分有意义)如果结果是$\mathbb{I} - \sum_i P_i$ 的话$\{P_i\}$是相互正交的；最后，论证$\mathbb{I} - \sum_i P_i$ 是自伴的。

注意到如果 $v \in \operatorname{ran} P$，则$\Pr_v(P)=1$。更一般地，如果 P_n 是 F 相应于本征值f_n 的本征空间，且 $v \in \operatorname{ran} P_n$，则$\Pr_v(P_n)=1$，即 F 拥有值 f_n 的概率（在态 v 中）为 1。这样一个态 v 被称为是 F 的"本征态"——它是 F 的本征空间 ran P_n 内的归一化矢量。在这种情形下，回顾式(4)，用 F 的谱分解来表示它，会使得概率和期望值的计算变得很容易。事实上，即使在处理一般的态时，用 F 的谱分解来表示它，即态用由 F 确定的基矢来表示，常常也很简便。

1.2.3.2.b　叠加　一个系统希尔伯特空间上的每个矢量都是该系统的可能态，这是量子理论的标准假设。该假设常常被表达为"叠加原理"，它声称态矢量的（归一化的）线性叠加仍然是态矢量。

给出一个可观测量 F，叠加原理产生了许多不是 F 本征态的（可能）态。出于简单性，假定 F 是极大的，其本征空间和本征值为$\{P_n\}$和$\{f_n\}$，考虑一个由 F 决定的正交基$\{v_n\}$（由于 F 是最大的，该基矢仅相当于从每个 ran P_n 中选择一个归一化的矢量），则形成了态矢量：

$$v=\sum_n k_n \mid v_n\rangle \tag{9}$$

其中$\sum_n |k_n|^2=1$，且至少有两个非零的系数 k_n。在此种情形下，我们说 v 是 v_n 的叠加。（人们有时会见到"叠加"一词用在如下表达中，即指出一些矢量"在叠加中"，而其他矢量不在叠加中。这种区分相对于给定的基矢才有意义，否则没有意义。每一个矢量都能够通过对某种基矢的选择而处于叠加中。）注意 v 不是 F 的本征态，且赋予 F 多个可能值以非平凡的概率。当然，叠加原理意味着 v 仍然是系统的一个可能态。

1.2.3.2.c　波函数　波函数只是表示态矢量的一种具体方式。把一个量子系统的希尔伯特空间看作是 $L^2(\mathbb{R}^3)$ 的元素常常很简便，在此情形中态矢量是在$\mathbb{R}^3$ 上的复值函数（的等价类）。它们一般满足的运动方程是波动方程的一种（例如，薛定谔方程——见 1.5.2.3.a 小节），且出于此原因——还有该方程历史上是在波动现象理念下得到的——这些函数被称为"波函数"。波的线性叠加可以用波的"重叠(superposing)"来想象，因而有术语"叠加"。

1.2.3.3　盖里森(Gleason)定理　我们可以用非负的迹 1 算符（"密度算符"）来生成概率测度。泛函 Tr[·]是"迹函数"，是从希尔伯特空间上有界的算符向$\mathbb{R}$ 的映射，定义如下：

$$\mathrm{Tr}[F]=\sum_k\langle e_k,\ (F^*F)^{1/2}e_k\rangle \tag{10}$$

其中$\{e_k\}$是$\mathcal{H}$上的一个正交基。（注意 F^*F 是自伴的和正定的。每个正算符 A 都有一个正定的自伴平方根 B，定义为 $B^2=A$，这在事实上为真。）且若 F 本身是正定的，则有 $F=\sqrt{F^2}$和

$$\mathrm{Tr}[F]=\sum_k<e_k,\ Fe_k> \tag{11}$$

迹函数被证明是独立于正交基$\{e_i\}$的选择。此外，迹函数有一个非常有用的特征，即它在其自变数的循环置换中保持不变，例如，对于任意的 A，B 和 C 有：

$$\mathrm{Tr}[ABC]=\mathrm{Tr}[BCA]=\mathrm{Tr}[CAB] \tag{12}$$

令 W 是在希尔伯特空间$\mathcal{H}$上的任意正定算符，且有 $\mathrm{Tr}[W]=1$。令 $E(\,)$是从一些可能值的“谱”S 到正定算符的 POVM，则 $\mathrm{Tr}[WE(\cdot)]$是在由 POVM E 表示的可观测量的可能值（波莱尔集的 σ 代数）上的可数可加性概率测度，如下：

$$\mathrm{Pr}(\Delta)=\mathrm{Tr}[W\,E(\Delta)] \tag{13}$$

可数可加性来源于(3)和迹函数的线性。归一化来自于(2)和 W 具有单位迹这一事实。

当 $E(\cdot)$是一个 PVM 时，式(13)定义了一个在$\mathcal{H}$上投影的可数可加性归一化测度。因而，任何密度算符都产生这样一个测量。上述命题的逆（显然）也为真：希尔伯特空间上投影的每个概率（即可数可加性，归一化的）测度，正如在式(13)中一样是由某密度算符产生的。该定理由盖里森[1957]提出，更精确地表述为：

（盖里森）定理： 令$\mathcal{H}$是维度大于 2 的希尔伯特空间，则每个$\mathcal{H}$（等价地，闭合的子空间）上投影的可数可加性的归一化测度 $\mathrm{Pr}(\cdot)$是由在$\mathcal{H}$上某个迹 -1 正定算符 W 产生的，对于一个投影 P，有：

$$\mathrm{Pr}(P)=\mathrm{Tr}[WP] \tag{14}$$

该论证是非平凡的。盖里森定理可以推广到普遍的 POVMs 情形。也就是说不同结果的可数可加性概率测度同样是由密度算符产生的。事实上，对于 POVMs 而言，对 $\dim(\mathcal{H})>2$ 情形没有限制。同样，该论证是非平凡的。见

[Busch(布施), 2003]。

在这种对量子力学的常见理解中，一个量子系统的运动学的核心是由希尔伯特空间上的 POVMs 和一个态，也就是一个密度算符共同给出的。在许多感兴趣的情形中，人们处理的是 PVMs 而非 POVMs，因而是自伴算符。

最后，注意到对于任意的态矢量 v，我们总能够用密度算符体系来表示 v，通过将其选择为向由 v 张开的子空间上的投影态 P_v 来实现。在这种情形下，对于任意的投影 Q，有 $\mathrm{Tr}[P_vQ]=\langle v, Qv\rangle$。(证明方法：在包含 v 的正交基中求迹。)

1.2.3.4 态的矩阵表示 一个矢量——尤其是一个态矢量 ψ——当然可以用任意的正交基 $\{e_n\}$ 来表示。在此情形中，展开式 $\psi=\sum_n c_n e_n$ 中的系数 c_n 可以看作是"ψ 在 e_n 基矢中的坐标"。事实上，有时候把态表示为这些坐标的列向量很方便(例如，见 1.3.3.2 小节)。

类似的构造对于密度算符也适用。在(正交)基 $\{e_n\}$ 中，将希尔伯特空间 $\mathcal{H}$ 上任意的算符 F 看作是一个元素为 $\langle e_n, Fe_m\rangle$ 的矩阵。从 $\mathcal{H}$ 上的算符到 $N\times N$ (其中 N 可以是无穷的)矩阵的映射事实上与从 $\mathcal{H}$ 上的算符(代数)到 $N\times N$ 矩阵(代数)同构。

特别地，令 W 是 $\mathcal{H}$ 上的一个密度算符，且令 $W_{nm}=\langle e_n, Fe_m\rangle$。现在令 F 是一个可观测量，其本征矢量是 e_n。注意，在此情形中，有 $\langle e_n, Fe_m\rangle=\delta_{nm}$。我们说 F 在基矢 $\{e_n\}$ 下是"对角化的"(因为所有非对角化的值都为0)。若 W 同样在 $\{e_n\}$ 下是对角化的，则 W 赋予 F 的概率表现完全是经典的，特别是经典的"求和定则"成立：

$$\mathrm{Pr}_W(f_n\text{或}f_m)=\mathrm{Pr}_W(f_n)+\mathrm{Pr}_W(f_m) \tag{15}$$

其中 Pr_W 是 W 通过式(14)赋予的概率，且 f_n 是 F 相应于本征矢量 e_n 的本征值。然而，若 W 在 $\{e_n\}$ 中不是对角化的，则一般地式(15)不成立。在这种情形下，我们说在 e_n 之间发生了"干涉"(在态 W 中)。

1.2.3.5 期望值 可以直接得到态(用密度矩阵表示) W 下的可观测量(用自伴算符表示的) F 的期望值是 $\mathrm{Tr}[WF]$。为了理解其原委，用 F 的谱分解来表示它。当 F 仅具有分立谱时，这一点非常容易看出，正如式(5)所示。之后通过迹的线性可得：

$$\mathrm{Tr}[WF]=\sum_n \mathrm{Tr}[WP_n]s_n \tag{16}$$

当 F 拥有连续谱时，我们必须用积分来处理，并且必须谨慎处理积分的定义。注意，表达式 $\mathrm{Tr}[WP_n]$ 是(在态 W 下) F 取值 s_n 时的概率。因而式(16)是 F 可能(谱)值 s_n 的加权和，权重由在态 W 下与那些可能值相关的概率 $\mathrm{Tr}[WP_n]$ 给出。注意式(16)中的求迹一般在由 F 确定的基矢下是最容易计算的。

1.2.3.6 量子概率理论 经典概率理论一般涉及在事件 σ 代数上的测度(7.5.5 节和 7.5.6 节)。这些 σ 代数是用通常的对补集和并集的集合理论操作来定义的。在量子理论中，我们处理的是不同的结构。然而它非常类似于在经典环境下所考虑的结构，至少在数学上是这样的，因而我们常常能够简单地从考虑经典概率理论开始。我们的"样本空间"是所有一维投影的集合。集合理论的补(E')变为"正交补"($E^{\perp}$)；集合理论的并集($E\cup F$)变为"张成(span)"(子空间 E 与 F 的张成，写作 $E\vee F$)；集合理论的交集($E\cap F$)仍为交集(现在写作 $E\wedge F$)；集合理论的"包含"($E\subseteq F$)变为子空间包含(常常写作 $E\leqslant F$)。稍后，我将详细考虑这一结构，它是希尔伯特空间的子空间的"格" $\mathcal{L}$(第 7.4)。现在，我只注意到，它符合恰当的特征：(i) $\mathcal{H}\in\mathcal{L}$；(ii) $E\in\mathcal{L}$ 意味着 $E^{\perp}\in\mathcal{L}$；(iii)对于任何可数的序列，有 $\{E_k\}\in\mathcal{L}$，$\bigvee_k E_k\in\mathcal{L}$。与经典概率理论相类似，量子概率理论则是在这样一个结构上的归一化测度理论。(当然，如果我们考虑的是 POVMs，而非 PVMs，那么这一理论不成立，至少不能以当前这种形式成立；相反，我们考虑的是效应的代数和其上的概率测度。但是在这里我并不关注细节。)

1.2.3.7 路德(Lüder)规则 那条件概率怎么样呢？虽然它的解释可能极富争议，且应用有些复杂，但量子理论中存在关于条件概率的标准表达，称为"路德规则"。事实上，我们能够从基本的考虑中得到它。

回顾基本概率理论中，给定事件 B，事件 A 的条件概率 $\mathrm{Pr}(A\mid B)$ 如下来定义：

$$\mathrm{Pr}(A\mid B):=\frac{\mathrm{Pr}(A\cap B)}{\mathrm{Pr}(B)} \tag{17}$$

这一定义的思想是，给定 B，事件 A(与 B)的概率是在 B 已发生的假定下"重整化的"的 A 与 B 同时发生的概率，即它是"好像" B 的概率为 1 时 A 的概率。事

实上，式(17)是唯一满足条件：若$A\subseteq B$，则$\Pr(A|B)=\Pr(A)/\Pr(B)$的概率测度。换句话说，若A包含在B中，则$\Pr(A|B)$仅是对原初概率测度的重整化，以使得赋予B以概率1。

已经证明，这一条件足以充分确定在希尔伯特空间的闭子空间(或投影)(的格)上条件概率测度的形式，参见巴布(Bub)[1977]。换句话说，令$\Pr_W$是与$\mathcal{H}$上密度算符W相关的概率测度，令P是满足$\Pr_W(P)\neq 0$的子空间(其中当然有$\Pr_W(P)=\mathrm{Tr}[WP]$)，则在$\mathcal{H}$的闭子空间上存在唯一的概率测度$\Pr_{W|P}$(态W中在P'条件下的"概率")对于任意的$Q\leqslant P$满足：

$$\Pr_{W|P}(Q):=\Pr_W(Q|P)=\frac{\Pr_W(Q)}{\Pr_W(P)} \tag{18}$$

该测度由下式给出：

$$\Pr_W(Q|P)=\frac{\mathrm{Tr}[PWPQ]}{\mathrm{Tr}[WP]} \tag{19}$$

式(19)就是"路德规则"。注意，对态矢量$|v\rangle$，通过从$|v\rangle$向P投影，归一化其结果，用新态$(P|v\rangle/\|Pv\|)$来计算Q的概率会得到相同的结果。因而，利用式(7)有：

$$\Pr_{|v\rangle}(Q|P)=\langle Pv|QPv\rangle/\|Pv\|^2 \tag{20}$$

1.2.3.8 混合态和纯态 对应于一维投影(也就是态矢量)的密度算符是"纯"态。这些态赋予那个一维投影以概率1。混合态(即不是纯态的态)并不赋予任意的一维投影以概率1。此外，混合态之所以称为"混合的"是因为它们总能够写成是纯态的线性叠加。事实上，通过谱定理，任意的混合态W可以写为：$W=\sum_n w_nP_n$(因为W是密度算符，有$0\leqslant w_n\leqslant 1$，且$\sum_n w_n=1$)。若一个或更多的$P_n$不是一维的，我们总能够把它写为相互正交的一维投影的和，所以我们可以不失普遍性地假定所有的P_n都是一维的。

这里，系数或"权重"w_n之和必须为1(因为$\mathrm{Tr}[W]=1$)，且事实上w_n是W赋予P_n的概率。所以，我们显然可以把W看作在字面上表示处于纯态P_n的系统的"混合"，比例为w_n，故w_n是从混合态中随机选择的系统处于(纯)态P_n的概率。下面我们将(定性地)探讨混合态的这种解释。

逆命题同样成立，即纯态的任意凸组合仍然也是态，一般是混合态。考虑

一个算符：

$$W=\sum_n w_n P_n \tag{21}$$

其中 P_n 在这里是一维的，但并不需要相互正交（仍然有 $\sum_n w_n=1$）。[①] 这个 W 有单位迹（unit trace，因为求迹函数是线性的），因而它是一个密度算符。但是，注意式(21)一般并不是它的谱分解。

1.2.3.9　本征态—本征值联系　根据量子态的标准解释，当且仅当 W 赋予 F 的一个可能值以概率 1 时（且赋予其他值以概率 0——换句话说，“平凡概率”），处于态 W 的系统拥有可观测量 F 的值。[②] 特别地，注意这种对态的解释不同于经典概率态的一般解释。在经典情形中，概率态是对可能纯态的测度，人们通常假定系统真的处于那些纯态中的一个。

依照法因(Fine)[1973]的说法，赋予确定值的这一规则实际被称为“本征态—本征值联系”。随后（第 5 节）我们将详细考察这一规则的明显影响。

1.2.4　不相容性

这一形式体系的直接结果是如下事实，即存在“不相容的”物理量，至少在最小的意义上若一个态赋予某个物理量以概率 1（例如某个投影），则它必然地赋予其他的量（通过本征态—本征值联系，这些其他的可观测量在该态下并不拥有值——回顾第 1.2.3.9 小节）以非平凡概率（即非 0 非 1）。这一事实直接得自于盖里森定理。（但是需要注意的是，我们可以用其他更为简单的方式表明，在希尔伯特空间的投影中不存在二值的概率测度。）

不相容性和非对易性紧密相联，而事实上这二者有时可以交换使用。考虑两个投影算符，Q 与 Q'。为了使问题简化，我们将通篇假定 Q 与 Q' 是一维的。则若 Q 与 Q' 不对易，即 $[Q, Q']\neq 0$，就不存在这样的态，即赋予 Q 以概率 1，同时赋予 Q' 以 0 或 1。为了证明这一表述，我们将首先表明（下一段）赋予一维

---

① 更为一般的是，如果有些 P_n 不是一维的，则我们要求 $\sum_n w_n \dim P_n=1$，因为一般而言对于投影 P 有 $\mathrm{Tr}[P]=\dim P$。

② 对于无界可观测量如位置和动量，我们自然会主动寻找其他的解释。一种可能的解释是否定它们曾经拥有确定的值，而是考虑粗粒值，例如断言如果态 W 赋予某个区域 Δ 以概率 1，则系统肯定限制在 Δ 中，其中这后一个断言并不意味着在 Δ 中存在某个是系统位置的点。然而，还有其他的方法。例如，见[Halvorson（霍尔沃森），2001]。

投影Q的唯一态是态Q本身。(注意在前一句中，第一次提到Q是表示某个物理量，而第二次是表示一个态。)然后我们将表明(在下一段)Q赋予任意的非对易的Q'以非平凡概率。

令W是赋予(一维的)Q以概率1的态。用W的谱分解来表示它，在由W确定的一个基中求迹，我们直接得到：

$$\mathrm{Tr}[WQ]=\sum_n w_n\langle e_n,\ Qe_n\rangle=1 \tag{22}$$

其中权重w_n(从W的谱分解中得到)的和为1。因而对于某个n，有$Qe_n=e_n$，即W事实上是纯态，且等于Q。所以，赋予一维投影Q以概率1的唯一态是Q本身。①

现在假定(一维的)$Q'\neq Q$且$Q'\not\perp Q$，即Q与Q'不对易(讨论见下文)。则，按照与上述相同的推理，在式(22)中用Q'替代Q，如果$\mathrm{Tr}[WQ']=1$，则W必须是纯态，且处于与Q'相关联的子空间中，即$W=Q'$。但由于我们假定$Q'\neq Q$，故不能够这样。另一方面，若我们想要使$\mathrm{Tr}[WQ']=0$，则$\ker W\subseteq \mathrm{ran}Q'$。(这一推理本质上与上述推理相同。)但这仍然不可能，因为这样的话会有$Q'\perp Q$，而之前我们已经给出W是纯态，且处于与Q相关的子空间中这一结论，并且我们已经假定了$Q'\not\perp Q$。

这一事实在更一般的形式中同样为真。给出两个自伴算符F和G，若F和G不共享本征矢量，则任何赋予F某值以概率1的态必然赋予G多于一个的可能值以非平凡概率(非0非1)。论证(本质上采用与上面相同的推理)留给读者。

上面我表明了如果$Q'\neq Q$且$Q'\not\perp Q$，则Q与Q'不对易。事实上，下列表述为真。对于任意的子空间A和B，相应的投影为P_A和P_B，当且仅当：

$$A=(A\wedge B)\vee(A\wedge B^{\perp})\text{且}B=(B\wedge A)\vee(B\wedge A^{\perp}) \tag{23}$$

时有$[P_A,\ P_B]=0$。这里我们并不限制于一维子空间，但是注意式(23)暗含在析取"$A=B$或$A\perp B$"中，且对于一维子空间而言它们是等价的。这里给出论证的大致理念。注意$A\wedge B$和$A\wedge B^{\perp}$是正交的，所以，若式(23)成立，对于某个$Z\perp A'$我们可以写作$P_A=P_Z+P_{A'}$。(当然，确实有$Z=A\wedge B$和$A'=A\wedge B^{\perp}$。)类

① 这一陈述在更普遍的形式中也是真的。令态W赋予(任意维度的)投影Q以概率1，则$(\mathrm{ran}Q)^{\perp}\subseteq\ker(W)$，如果$Q$是$W$赋予概率1的最小子空间时二者是相等的。

似地，在 $Z\perp B'$ 时有 $P_B=P_Z+P_{B'}$。此外有 $A'\perp B'$。换句话说，条件式(23)意味着 A 与 B“是正交的，除了某些共同的部分(Z)”。则有 $[P_A, P_B]=[P_Z+P_{A'}, P_Z+P_{B'}]=[P_Z, P_Z]+[P_Z, P_{B'}]+[P_{A'}, P_Z]+[P_{A'}, P_{B'}]=0$。

换一种表述方式，我们将只简述大致的想法。若 P_A 与 P_B 对易，则对于任意矢量 v，有 $P_AP_Bv=P_BP_Av$。首先选择 $v\in A$，则 $P_AP_Bv=P_Bv$。一般而言，如果 $P_Aw=w$(这里 $w=P_Bv$)，则要么 $w\in A$，要么 $w=0$。所以，要么是(i)$P_Bv=0$，要么是(ii)$P_Bv\in A$。若(i)对于所有 $v\in A$ 都是真的，则 $B\perp A$ 且式(23)显然成立。若(ii)对于所有的 $v\in A$ 成立，则 $B\leqslant A$ 且式(23)显然成立。利用所涉算符的线性，我们能够表明若(ii)仅对于某些 $v\in A$ 成立，则 P_Bv 必定形成了 A 的一个子空间，且显然这一子空间是 A 和 B 共有的，确切地说它是 $A\wedge B$。类似地，我们能够表明从与 $A\wedge B$ 正交的子空间中选择 v 会产生(i)，故确实有 $A=(A\wedge B)\vee(A\wedge B^{\perp})$。对 $v\in B$ 重复这一论证，我们发现式(23)成立。

不相容性的事实标志着其与经典物理学的重大不同，在经典物理学中态空间的结构和可观测量允许态赋予所有可观测量值以概率1(即在系统所有的“特征”空间中，存在二值概率测度)。因而，量子理论的概率似乎与经典理论的概率有着本质上不同的特征，这一不同总是因为系统的态并不是最大程度确定的而产生。①

1.2.5　正则对易关系

不相容性的一个重要且典型的例子是关于位置和动量可观测量的。事实上，它们满足“正则对易关系”(CCRs)：

$$[P_i, Q_j]=-i\delta_{ij} \tag{24}$$

其中 i 与 j 可以是 x，y 或 z。在这之后，我们将注意力局限于一维情形，写作 $[P, Q]=-i$。向三维情形的推广很简单明了。注意式子右边的常数是与单位算符的隐式(implicitly)相乘。

满足这些对易关系的任意两个可观测量典型地被称为“正则共轭”。这些关系是量子理论的核心，我们将在第4节中细致地讨论它们。现在，我们仅把它们看作是不相容性的重要案例。

---

① 这里我们仅考虑经典物理学实现真正概率测度的情形，且我们忽略那些经典物理学中纯粹是非决定论的情形。见[厄尔曼(Earman)，本书]。

1.2.6 复合系统

1.2.6.1 纠缠态 复合系统用希尔伯特空间的张量积表示(第7.1.9节),例如,由两个粒子组成的一个系统拥有这样一个态,它是在两个粒子各自希尔伯特空间的张量积上的密度算符。在$\mathcal{H}=\mathcal{H}_1\otimes\mathcal{H}_2$上的两种矢量本质上与物理上存在关键性的不同。如果对于$x\in\mathcal{H}_1$和$y\in\mathcal{H}_2$,$\mathcal{H}$上的矢量v能写作$x\otimes y$,则它被称为"可因子化的"。否则,v被称为"不可因子化的"或"纠缠的"。类似的定义适用于$\mathcal{H}$上的算符(也就是密度算符态)。

纠缠态(无论表示为密度算符还是矢量)的存在被证明拥有许多有意义的影响。它与"量子非定域性"相关,也与不能够在经典系统中完成的某些计算和信息理论的(例如,密码的)技术相关联。① 这些态的存在源自于复合系统的纯(矢量)态在线性组合下是闭合的这一要求。换句话说,它来源于态叠加原理在复合系统和简单系统中的应用。

1.2.6.2 双正交分解 关于张量积空间中矢量的一个重要结果是"双正交分解定理",可参见薛定谔[1935b],该定理指出,给定希尔伯特空间$\mathcal{H}$中的矢量v和$\mathcal{H}$的因子分解$\mathcal{H}=\mathcal{H}_1\otimes\mathcal{H}_2$,则存在$\mathcal{H}_1$的正交归一基$\{e_n\}$和$\mathcal{H}_2$的正交归一基$\{f_m\}$满足:

$$v=\sum_n c_n(e_n\otimes f_n) \tag{25}$$

若对于所有的$n\neq n'$有$|c_n|\neq|c_{n'}|$,则这些基是唯一的(在基的每个元素上至多有相位$e^{i\theta}$)。注意,一般而言对于$\mathcal{H}_1$和$\mathcal{H}_2$上的任意基$\{x_n\}$与$\{y_m\}$,v通常是用双重求和表达的:

$$v=\sum_{n,m} c_{nm}(x_n\otimes y_m) \tag{26}$$

此表达与式(25)相区别。

1.2.6.3 约化态

1.2.6.3.a 分迹和约化密度算符 假定给我们一个复合系统的态,并希望从中得到其一个组分的态。若复合态是可因子化的,则此过程直接明了(态$W=W_1\otimes W_2$确定了组分的态分别是W_1与W_2)。但当复合态是纠缠态时又会如何呢?这里我们面临着一个问题。若复合态是纠缠的,则没有显著的意义,因

---

① 见[Bub,本书]。

为在其中它能够被“分解”为相应于一个系统的一“部分”和相应于另一个系统的另一“部分”。

对该问题的通常求解是把复合系统的态看作是由分迹给出的。对于任意希尔伯特空间的张量积$\mathcal{H}=\mathcal{H}_1\otimes\mathcal{H}_2$，“$\mathcal{H}_1$的分迹”是从$\mathcal{H}$上的算符向$\mathcal{H}_2$上算符的一个映射$\mathrm{tr}^{(1)}[\cdot]$。若满足该条件的这一映射是唯一的，则对于在$\mathcal{H}$上的算符 W 和在$\mathcal{H}_2$上的任意可观测量 F_2 而言，算符$\mathrm{tr}^{(1)}[W]$产生 F_2 的期望值与 W 在$\mathbb{I}_1\otimes F_2$ 上产生的相同，可见约赫(Jauch)[1968，§11—8]。这里的思想是$\mathrm{tr}^{(1)}[\cdot]$对系统 1“求迹”，只提取那些适用于系统 2 的复合系统部分。除非 W 是一个“积态”(即 $W=W_1\otimes W_2$)，否则从 W 中得到的约化态必然是混合态。

1.2.6.3.b　正常(Proper)与非正常(Improper)混合　在第 1.2.3.8 小节中我介绍了这样的观点，即一个混合态可以理解为是由每个处于纯态的系统在字面上的混合。当然当我们描述从一个系综中随机选择的某个系统的态，且该系综是由字面上混合处于不同纯态的系统而产生的时，以这种方式来解释混合态是非常正常的。然而，现在我们看到混合态可以另一种方式产生，即作为处于不可因子化复合态的复合系统的一个组分的态。在这些情形下，很显然(组分的)态不应该按照上述解释来理解。事实上，甚至不存在此组分作为其一部分的系综。因而，产生自对复合系统态求分迹的混合态通常被称为“非正常混合”，而那些产生自处于纯态的单个系统的混合通常被称为“正常混合”，这是由德斯帕那特(d’Espagnat)[1971]引入的术语。由非正常混合产生的概率能否合理地理解为是“对真正纯态的无知”(正如对于正常混合可以的那样)则是研究要解释的问题。

1.2.6.4　关联　处于不可因子化态的复合系统将显现出两个(或更多)组分被测可观测量值之间的关联。例如，考虑态矢量 $v=c_1f_1\otimes g_1+c_2f_2\otimes g_2$(其中 c_1 和 c_2 是非零的系数)，假定 f_n 和 g_n 分别是可观测量 F 和 G 的本征矢量。在此态下，在系统 1 的 F 值与系统 2 的 G 值之间存在关联。令 P_{f_n} 和 P_{g_n} 分别是向由 f_n 和 g_n 张成的子空间上的投影，且令 P_v 是向由 v 张成的子空间上的投影。则应用路德规则(19)，我们发现：

$$\mathrm{Pr}_{P_v}(\mathbb{I}_1\otimes P_{g_{n'}}\mid P_{f_n}\otimes\mathbb{I}_2)=\frac{\mathrm{Tr}[(P_{f_n}\otimes\mathbb{I}_2)P_v(P_{f_n}\otimes\mathbb{I}_2)(\mathbb{I}_1\otimes P_{g_{n'}})]}{\mathrm{Tr}[P_v(\mathbb{I}_1\otimes P_{f_n})]}\tag{27}$$

其中$\mathbb{I}_k$是$\mathcal{H}_k$上的单位算符。在包括$f_n\otimes g_{n'}$的基上求迹将表明，这一条件概率当

$n \neq n'$时为0，而当$n = n'$为1。换句话说，F的值(系统1)和G的值(系统2)是全关联的。[①] 考虑其他可观测量将揭示更多的关联(但并不总是全关联)。稍后我们将看到一个例子。

1.2.7 态空间的结构

上面我们注意到(第1.2.3.8节)纯态的每个凸组合仍然是一个态。当然，混合态的一个凸组合同样(通过谱定理)也是纯态的一个凸组合，所以态的集合构成了一个凸集(第7.1.10节)，这一点我随后将详细讨论(第2.2.1节)。在这里我们注意到的要点是，量子理论中态的凸集并不是一个单形(simplex)。

这一点(第7.1.10节)标示了与经典物理学的不同，在经典物理学中每个混合态都可用纯态来唯一分解。因而人们很自然地把混合态当作是对出现在分解中的纯态的无知测量。量子理论中的混合态没有可适用的相应的直接解释，这部分是因为混合态可多重地分解为纯态的一个凸组合。

1.3 简单案例：自旋$\frac{1}{2}$粒子

一些真实计算的实践会增进对形式体系及其所产生问题的理解，虽然这些计算很简单。在这一理念下，我们考虑一个自旋$\frac{1}{2}$粒子的例子。该例子被引用过很多次，但它当之无愧。虽然有一些重要的关于量子理论基础性问题和哲学问题不能在自旋$\frac{1}{2}$粒子情境下显现并得到研究，但仍有许多这样的问题能够在此情况下来研究。

1.3.1 量子理论中自旋的引入

自旋(Spin)是在1924年试图对特定金属电磁辐射光谱作出理解的过程中引入的。在该解释的过程中，电子被假定拥有某种“二值的量子自由度”。[②] 这

---

① 如果两个可观测量具有相同的谱，且一个可观测量的值总是另一个值的负数时，人们会说这两个可观测量是“全反关联的”。他们偶尔也会把术语“全关联”用于如下简单情形，在那里一个可观测量的值总是等于另一个的值。我们对术语“全关联”的使用——仅当给定一个可观测量的值得到另一个的值的条件概率为0或1时，它们在态中是全关联的——包括了这两种情形。

② 讨论见[马希米(Massimi)，2004，Chapters. 2，4]。

一自由度很快与电子的自转关联起来。但由于电子是带电物体，它的自转会产生磁场——电子表现为一个磁子，其北极和南极位于自转轴上。磁的性质正是解释这一现象所需要的。

目前来看，这一解释听起来很不错。然而，几乎很快就发现自转不能是字面意义上的。因此，“自旋”理论在新量子理论背景中发展了起来，其名称照旧，且我们仍旧把电子(正如当前理论告诉我们的，还有其他粒子)的这一磁性质称为“自旋”。

1.3.2　自旋的量子化

自旋是“量子化的”，这一事实在泡利(Pauli)把此性质刻画为“二值自由度”时已经预测到了。这一事实是经典层面上所不曾料想到的。为明白何以如此，需要考虑粒子自旋测量的一个标准方法。(事实上，这一方法对电子并不适用，但它足以说明问题，且对具有自旋的电中性粒子适用。)相关的装置是一对由特定形状以某种特定方式放置的用于产生非均匀磁场的磁体所构成的“斯特恩—盖拉赫(Stern-Gerlach)”装置，该装置中磁场在一个方向上(如北方)强于另一个方向(见图1)。

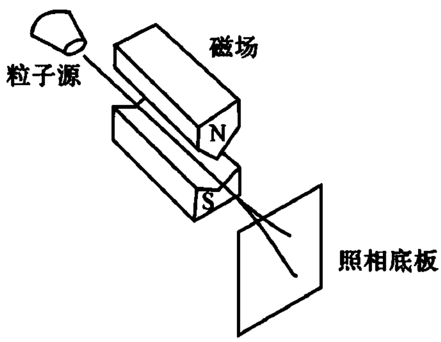


图1　斯特恩—盖拉赫磁场实验

想象一个简单的条形磁体穿过斯特恩—盖拉赫磁场。若北极点垂直向上以使得它接近于上方磁铁，则上方磁铁对(该条形磁铁)北极的下推力要比下方磁铁对南极的上推力大，总的结果将是条形磁铁向下偏转了。若条形磁体进入斯特恩—盖拉赫磁场时是南极向上的，则结果相反，即总体上向上偏转。若反过来，条形磁体水平地进入斯特恩—盖拉赫磁场，则它将径直通过，在其轨迹上

没有总的偏转。最后，若条形磁体既不是垂直穿过也不是水平穿过，则结果将是向上或向下偏转，即是两种极端情形之中的某种。(磁体在两种极端情形之间的轨迹见图1。)

人们的物理直觉可能是，在中间情形中电子的旋转轴应该与磁场方向调整在一条直线上，因而中间情形迅速地演变为一种极端情形。但是，事情并不是这样，而是经典理论中认为自旋的电子会表现如同陀螺，保留其相对于磁场的最初偏转。

现在，想象让自旋粒子系综穿过磁场。与经典期望的结果相反(即从"最大程度的向下"到"最大程度的向上"之间不同偏转程度的分布)，人们发现仅有两种结果："最大程度的向下"和"最大程度的向上"——这些结果见图1。

无论斯特恩—盖拉赫磁场如何朝向，这一结果始终成立。即，注意到我们可以重新设置磁场方向，使得其穿过它的中轴指向空间中的任意方向。让一个电子穿过该装置，我们将再次发现它要么"朝上"，要么"朝下"(相对于空间中的这一新方向)。因此我们可以测量任意方向上粒子的自旋，且我们把测量的可观测量称为"自旋 u"，其中 u 指的是空间中某个确定的方向。此外，关于自旋的这些事实解释了经典的"自旋"至多是对具有"自旋"的粒子的某种特征的隐喻。(不管怎样，在非相对论的量子理论中，电子通常被处理为点粒子，从而它的自旋不能够与任意的空间旋转相一致。正如通常所讲的，自旋没有"经典对应"。)

1.3.3 自旋的量子形式体系

现在，我们来看如何在上文我所概述的形式体系中表示具有自旋的粒子的可观测量和态。我将只考虑与自旋相关的自由度，忽略空间自由度等其他自由度。

1.3.3.1 希尔伯特空间和可观测量 单个自旋$\frac{1}{2}$粒子的希尔伯特空间为$\mathbb{C}^2$，它是一个有两个分量的复列向量空间(第7.3节)。系统的"可观测量"对应于在不同方向(空间的每一个方向)上的"自旋"，每一个都只有两个可能的

值，我们称其为“上”(用数值 +1/2 表示)和“下”(-1/2)。① 在 x，y 和 z 方向上的自旋可观测量用泡利矩阵术语定义为 $S_x=(1/2)\sigma_x$，S_y 和 S_z 相类似(见第7.3.1)。

1.3.3.2　态　纯态可以用模1矢量来表示，或用向这些矢量张成的空间上的投影来表示。例如，考虑态矢量：

$$\psi=\begin{pmatrix}1\\0\end{pmatrix},\chi=\begin{pmatrix}0\\1\end{pmatrix} \tag{28}$$

矢量 ψ 对应于(纯)密度算符(一维投影算符)：

$$W=\begin{pmatrix}1&0\\0&0\end{pmatrix} \tag{29}$$

矢量 ψ 和 χ 是

$$\sigma_z=\begin{pmatrix}1&0\\0&-1\end{pmatrix} \tag{30}$$

的本征态，本征值分别为 +1 和 -1。

注意 S_z 在态 W 中的期望值为：

$$\begin{aligned}\mathrm{Tr}\left[\begin{pmatrix}1&0\\0&0\end{pmatrix}\begin{pmatrix}\frac{1}{2}&0\\0&-\frac{1}{2}\end{pmatrix}\right]&=\mathrm{Tr}\left[\begin{pmatrix}\frac{1}{2}&0\\0&0\end{pmatrix}\right]\\&=(1\quad 0)\begin{pmatrix}\frac{1}{2}&0\\0&0\end{pmatrix}\begin{pmatrix}1\\0\end{pmatrix}+(0\quad 1)\begin{pmatrix}\frac{1}{2}&0\\0&0\end{pmatrix}\begin{pmatrix}0\\1\end{pmatrix}\\&=\frac{1}{2}+0=\frac{1}{2}\end{aligned} \tag{31}$$

---

① 我所讨论的粒子——那些在给定自旋方向上仅有两个自由度(“上”或“下”)——称为“自旋$\frac{1}{2}$粒子”，部分是因为它们关于任意给定轴的角动量或是 $+\hbar/2$(“上”)或是 $-\hbar/2$(“下”)，其中 $\hbar$ 是角动量单位，在常用的单位制下等于 $1.054\times10^{-34}\,\mathrm{kg\ m^2/s}$。(正如通常那样，我采用的单位中 $\hbar=1$。)此外，称这些粒子为“自旋$\frac{1}{2}$”粒子有更深层的群理论原因，但在这里我们不深入探讨。(也存在更大自旋的粒子，这意味着在操作上它们在每个自旋方向上有大于2的自由度。)

回顾我们之前关于在适当选择的基矢下求迹的评论。当然，一般而言一个系统拥有等于某个值 r 的期望值，并不足以说明系统具有值 r。（事实上，r 甚至都可能不在可能值的谱中。）然而，我们可能也注意到了，在此情形下与适当投影算符相关的概率为 1。因此首先注意 S_z 的谱分解为：

$$S_z=\begin{pmatrix}\frac{1}{2} & 0\\ 0 & -\frac{1}{2}\end{pmatrix}=\left(+\frac{1}{2}\right)\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}+\left(-\frac{1}{2}\right)\begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}$$

$$:=\left(+\frac{1}{2}\right)P_{z_+}+\left(-\frac{1}{2}\right)P_{z_-} \tag{32}$$

与 S_z 的值 $+\frac{1}{2}$ 相关的投影为 P_{z_+}，且在态 W 下值是 $+\frac{1}{2}$（对于 S_z）的概率为：

$$\mathrm{Tr}[WP_{z_+}]=\mathrm{Tr}\left[\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}\right]=1 \tag{33}$$

我们把计算的细节留给读者。注意到对矩阵的求迹等于仅把对角线上的数加起来，读者可能想要去论证这一事实。正如我之前在更一般的背景中所指出的，这一表达相当于是 P_{z_+} 在态 W 下的期望值。因此，特别地，如果人们赞同“在态 W 下可观测量 F 取值 r 的概率为 1”意味着“在态 W 下的系统拥有 F 的 r 值”，则我们可以根据式(33)推出系统在态 W 下 S_z 有值 $+\frac{1}{2}$。（我们稍后将对这一解释性原则作更具体的讨论。）

1.3.4 不相容性

最后，注意在此态 W 下，在 x 和 y 方向上的自旋期望值为 0。例如：

$$\mathrm{Tr}[WS_x]=\mathrm{Tr}\left[\begin{pmatrix}1 & 0\\ 0 & 0\end{pmatrix}\begin{pmatrix}0 & \frac{1}{2}\\ \frac{1}{2} & 0\end{pmatrix}\right]=\mathrm{Tr}\left[\begin{pmatrix}0 & \frac{1}{2}\\ 0 & 0\end{pmatrix}\right]=0 \tag{34}$$

这一事实表明（在二维情形下意味着），在态 W 下 $S_x=+\frac{1}{2}$ 和 $S_x=-\frac{1}{2}$ 的概率为 $\frac{1}{2}$，正如我们也可以通过直接计算所确证的那样。首先，注意 S_x 的谱分解为：

$$S_x=\begin{pmatrix}0 & \frac{1}{2}\\ \frac{1}{2} & 0\end{pmatrix}=\left(\frac{1}{2}\right)\begin{pmatrix}\frac{1}{2} & \frac{1}{2}\\ \frac{1}{2} & \frac{1}{2}\end{pmatrix}+\left(-\frac{1}{2}\right)\begin{pmatrix}\frac{1}{2} & -\frac{1}{2}\\ -\frac{1}{2} & \frac{1}{2}\end{pmatrix}$$

$$:=\left(+\frac{1}{2}\right)P_{x_+}+\left(-\frac{1}{2}\right)P_{x_-} \tag{35}$$

同理读者可以证明 $\mathrm{Tr}[WP_{x_+}]=\mathrm{Tr}[WP_{x_-}]=\frac{1}{2}$。

在此特殊情形下，我们确证了之前在理论上做出的一个陈述，即对一个可观测量而言的无色散(dispersion-free，对所有可能的值只产生概率 0 或 1)态，对于某些其他可观测量而言不需要是无色散的。事实上，我早先说过，没有共同本征矢量的非对易可观测量在下述意义上总是不相容的，即在其中一个可观测量上无色散的任意态没有必要在另外一个上也无色散。现在注意到 S_x，S_y 和 S_z 是相互非对易的，且确实没有共同的本征矢量。(在此两维情形中，非对易极大可观测量不能够共享任意的本征矢量。)因此，对于一个可观测量是无色散的态将必然地对其他可观测量产生非平凡的概率。

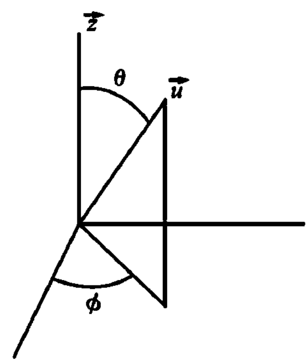


图 2　极角

考虑任意的矢量 u，它在空间上用相对于 z 轴的极角 θ 和 ϕ 来确定，也就是在笛卡尔坐标中，$u=(x, y, z)=(\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$，见图 2，则相关的自旋可观测量用下述矩阵来表示：

$$S_u=\frac{1}{2}\begin{pmatrix}\cos\theta & e^{-i\phi}\sin\theta\\ e^{i\phi}\sin\theta & -\cos\theta\end{pmatrix} \tag{36}$$

对此表达式的一个合理与灵活的解释是注意到 $S_u=S_x\sin\theta\cos\phi+S_y\sin\theta\sin\phi+S_z\cos\theta$。唯一满足对易的一对这种算符是反平行的，即它们对应于反平行方向上的自旋(且这种算符仅仅是其中一个算符乘因子 -1)。

但是，人们应该谨记，盖里森定理对于我们的二维空间并不成立。因此，在此情形下矩阵算符并不定义所有的态。事实上，贝尔(Bell)[1964]表明了如何用一个附加的“隐”参量来定义一个对$\mathbb{C}^2$上投影的无色散测量。此外，量子

力学的态可以通过对隐参量可能的值求平均而得到，隐参量的可能值在其上有适当的概率分布。

1.3.5 布洛赫球(Bloch sphere)

希尔伯特空间$\mathbb{C}^2$用以表示任意的二阶量子系统，这样的系统是量子理论最为感兴趣的，且近些年来随着越来越多的对量子信息与量子计算的兴趣把注意力更多地集中到这样的系统上(因为它们是经典“比特”的量子类比——见巴布在本书第六章内容)，情况更加如此。采用布洛赫球对$\mathbb{C}^2$上纯态的表示通常有助于对这些态的详细研究。注意，$\mathbb{C}^2$上的任意纯态可以用形如$v=\cos(\theta/2)\psi+e^{i\phi}\sin(\theta/2)\chi$的矢量来表示。① 因此，再次参考图2，我们能够把每个不同的纯态表示为在($\mathbb{R}^3$中)单位球表面上的唯一一点，该球通常称为“布洛赫球”。球的“北极”对应于态ψ，“南极”对应于态χ。

但是，事实上“布洛赫球”是一个球体。其内部的点对应于混合态，具体如下。在$\mathbb{C}^2$上的每一个密度矩阵W都可以写作：

$$W=\frac{\mathbb{I}+\vec{r}\cdot\vec{\sigma}}{2} \tag{37}$$

其中$\vec{\sigma}$是泡利矩阵“矢量”(7.3.1节)，$\vec{r}$是$\mathbb{R}^3$上的矢量，有$\|\vec{r}\|\leqslant1$。$\vec{r}$的分量确定了在布洛赫球内部表示相应密度算符的一个点。特别地，注意$\vec{r}=(0,0,1)$对应于由$\theta=0$给出的纯态，如其所应当的那样。

1.4 狄拉克记号

我稍后将回到自旋$\frac{1}{2}$粒子的例子以说明量子理论中的一系列问题。在用到该例子的时候——事实上在文章剩余部分通篇都在用这一例子——掌握“狄拉克左矢(bra)—右矢(ket)”这一方便的记号是非常有用的，物理学家们和哲学家们常常这样用。

---

① 该论述不是说每个矢量可以写作此形式，而是每个纯态都可用此形式来表示。回顾一个总的相因子并不会影响由矢量产生的概率，因而我们可以在不失普遍性的情况下假定ψ的系数是实数。

1.4.1 左矢和右矢

在左矢—右矢记号中，矢量用右矢 $|v\rangle$ 表示（有时也这样称呼）。例如，在上面的讨论中，式(28)中的列向量 ψ 可以表示为 $|z_+\rangle$。对偶空间中的元素（上面讨论中的“行向量”——见第 7.1.8 节）用左矢 $\langle v|$ 表示。在上面的例子中，存在从右矢（列向量）到左矢（行向量）的正常一一映射：

$$\begin{pmatrix} a \\ b \end{pmatrix} \to \begin{pmatrix} a^* & b^* \end{pmatrix} \tag{38}$$

因而左矢以一种显然的方式定义了（连续的）线性函数。令：

$$|v\rangle = \begin{pmatrix} a \\ b \end{pmatrix} \quad 和 \quad |w\rangle = \begin{pmatrix} c \\ d \end{pmatrix} \tag{39}$$

线性函数（左矢）$\langle v|$ 作用于矢量（右矢）$|w\rangle$ 为：

$$\begin{pmatrix} a^* & b^* \end{pmatrix} \begin{pmatrix} c \\ d \end{pmatrix} = a^* c + b^* d \tag{40}$$

在狄拉克（“左矢—右矢”）记号中写作 $\langle v|w\rangle$。（读者可能想要验证这样定义的函数确实是线性的。）当然，正如它所必须是的那样，$\langle v|w\rangle$ 也是 $|v\rangle$ 与 $|w\rangle$ 的内积，由式(38)给出。（在此种记号中，我们继续用 $\|v\|$ 表示矢量的模，而不是用 $\||v\rangle\|$。）

在 $\mathcal{H}$ 是任意（复）希尔伯特空间（至多是可数维的）的一般情形中，我们把 $\mathcal{H}$ 中的元素作为右矢，对偶空间 $\mathcal{H}^*$ 中的元素看作左矢。内积现在写作 $\langle v|w\rangle$，同时它也表示线性函数 $|v\rangle$ 作用于矢量 $|w\rangle$，也是 $|v\rangle$ 与 $|w\rangle$ 的内积。

1.4.2 算符

算符 F 作用于矢量 $|v\rangle$ 写作 $F|v\rangle$。可观测量 F 在态 $|v\rangle$ 中的期望值写作 $\langle v|F|v\rangle$，它在记号上（还有数值上）等价于 $\langle v|Fv\rangle$，后者可以理解为是 $|v\rangle$ 与矢量 $F|v\rangle$ 的内积。类似地也可以定义表达式 $\langle w|F|v\rangle$。

相应地有时称为“矢量直积”的下式：

$$\begin{pmatrix} a \\ b \end{pmatrix} \begin{pmatrix} c & d \end{pmatrix} = \begin{pmatrix} ac & ad \\ bc & bd \end{pmatrix} \tag{41}$$

我们可以通过下式把 $|v\rangle\langle w|$ 定义为 $\mathcal{H}$ 上的算符：

$$(|v\rangle\langle w|)|x\rangle = \langle w|x\rangle|v\rangle \tag{42}$$

注意简单的符号操作将产生相同的结果。

1.4.3 狄拉克记号的应用

正如我刚刚提示的，如果狄拉克记号的真正意义被理解的话，它将是非常有用的，否则的话反而会带来危害。它的功用和危害在于这样的事实中，即它允许人们或多或少忽略各种不同，例如在一个矢量和一个线性函数(一个对偶空间中的元素)之间的不同。它对于“位置无关”的计算也非常有帮助。例如，我们可以在不采用泡利矩阵和类似表达式的情况下讨论自旋$\frac{1}{2}$粒子的理论。考虑$\mathbb{C}^2$上的基$\{|z_+\rangle, |z_-\rangle\}$，其中$|z_+\rangle$是赋予$S_z$值$+\frac{1}{2}$的概率为1的态，以此类推——注意我们不需要担心如何把该态表示为复数的列。进行计算所需要注意到的仅仅是对于空间中的一个矢量u，它相对于z轴用极角θ与ϕ来确定，有：

$$|u_+\rangle = \cos\left(\frac{\theta}{2}\right)e^{-i\frac{\phi}{2}}|z_+\rangle + \sin\left(\frac{\theta}{2}\right)e^{i\frac{\phi}{2}}|z_-\rangle \tag{43}$$

$$|u_-\rangle = -\sin\left(\frac{\theta}{2}\right)e^{-i\frac{\phi}{2}}|z_+\rangle + \cos\left(\frac{\theta}{2}\right)e^{i\frac{\phi}{2}}|z_-\rangle \tag{44}$$

则自旋可观测量表示为：

$$S_u = \frac{1}{2}|u_+\rangle\langle u_+| - \frac{1}{2}|u_-\rangle\langle u_-| \tag{45}$$

注意$\langle z_+|u_+\rangle = \cos\left(\frac{\theta}{2}\right)e^{-i\frac{\phi}{2}}$和$\langle z_-|u_+\rangle = \sin\left(\frac{\theta}{2}\right)e^{i\frac{\phi}{2}}$可以直接从式(43)中得到。因此，在态$W = |z_+\rangle\langle z_+|$中系统的可观测量$S_u$有值$+\frac{1}{2}$的概率可以很快计算为：

$$\mathrm{Tr}[\,|z_+\rangle\langle z_+|(\,|u_+\rangle\langle u_+|\,)] \tag{46}$$

$$= \langle z_+|(\,|u_+\rangle\langle u_+|\,)|z_+\rangle \tag{47}$$

$$= \langle z_+|u_+\rangle\langle u_+|z_+\rangle \tag{48}$$

$$= |\langle z_+|u_+\rangle|^2 \tag{49}$$

$$= \cos\left(\frac{\theta}{2}\right)^2 \tag{50}$$

从第一行得到第二行要用基$\{|z_+\rangle, |z_-\rangle\}$求迹。狄拉克记号的巧妙之处就在于人们能够像这里展示的那样简单地“做符号上合乎常理的事情”然后得到正确的答案。例如，第三行是通过“去括号再合并竖线(bar)”从第二行中得到的。从概念上来说，我们允许算符$|u_+\rangle\langle u_+|$作用于$|z_+\rangle$上，得到矢量$\langle u_+|z_+\rangle|u_+\rangle$，然后取此矢量与$|z_+\rangle$的内积(或将线性函数$|z_+\rangle$作用于$\langle u_+|z_+\rangle|u_+\rangle$)。但是，这一记号的简便性也会使人们忘记概念上的重要区别。

此外要记住，不需要担心矢量和可观测量的明确(即矩阵)表示的简便性也会导致人们写出一些非常荒唐或至少是在物理学上难以理解的态。比如，人们常常看到会写出诸如|死猫⟩或|萨拉看到指针⟩这样的“态”。狄拉克记号很容易诱使人们写下这样的描述，但我们不清楚这样的“态”是否对应于某个纯的矢量态，如果是的话它们的特征是什么，因而这样的表达最好还是当作卡通图片看一看就好了。

1.5　变换

现在我们看到了该如何表示可观测量还有如何计算期望值(和概率)。虽然这些问题确实是理论的核心，但仍然存在对于哲学的和基础的讨论而言很重要的形式体系的其他方面。特别地，本小节讨论变换，包括物理系统态和与这些系统相关的可观测量的变换。因此，我将有机会提到一些对于量子力学基础而言很重要的定理。

1.5.1　*群及其表示*

1.5.1.1　引子　伽利略观察到运动定律并不依赖于它们所使用于其中的“实验室”(参考系)的恒定速度(constant velocity)。(例如，在匀速运动——更精确地说是惯性地运动着的船舱中，伽利略写道“双脚跳起，在每个方向上都经过相等的空间”，这和回到岸上的情形一样。)运动定律既不依赖于位置，也不依赖于它们所使用的时间，也不依赖于朝向的方向。换言之，运动定律在某些变换如递升(boost)(速度的改变)、空间平移、时间平移和旋转下是不变的。这些种类的变换在数学上是用群表示的，在我刚刚提到的“伽利略变换”中的群

通常称为“伽利略群”。[1] 因此，群理论(7.6 节)是研究量子理论“不变性”的自然环境之一。

这么做是因为群的特征恰好是通常认为适用于“不变性变换”的特征。特别地，如果 α 和 β 分别是保持定律不变的变换，则 α 之后 β 的这一组合也是这样一个变换。类似地，如果 α 是这样一个变换，则存在一个“撤销”α 所作变换的变换，即 α 的逆。比如，注意到两个伽利略变换的组合是另一个变换，且每一个这样的变换都有一个逆。[2]

群也出现在其他情况中。假定我们对一封闭物理系统的动力学感兴趣(正如我们将很快表现出的那样)。考虑系统态时间演化的一种方式是将其看作态集合的变换，则所有这些时间演化的集合应该形成一个群。单位算符表示“没有改变”(或不随时间演化的简并情形)。积表示一个阶段演化之后的另一段演化。逆表示“反向”演化，或在时间上后向的演化。(若一个给定理论不是时间可逆的，则我们处理的是一个半群，而不是一个群。)

现在，人们常常通过规定群的积和逆来抽象地明确一个群，而不是把它表示为在某个集合(如系统物理态的集合)上的变换群。最重要的普通案例是群 Z_2，它包含两个元素 x 和 y。乘法规则是：$xy=x$，$yx=x$，$xx=y$ 和 $yy=y$。单位算符(显然)是 y，虽然 x 和 y 都是它们自己的逆。注意到我们明确该群时没有借助于任何具体的数学对象——符号“x”和“y”只是该群的两个元素的名称，它们本身没有更多的数学内容。例如我们也可以把群 Z_2 表示为从任意二元素集合到其自身的映射群，其中 y 为单位映射，x 为交换元素的映射(把其中一个映射到另一个上)。Z_2 的另一个表示是把 x 看作是复共轭*，把 y 看作是**。

1.5.1.2 维格纳(Wigner)定理 在把群看作“对称变换集合”的过程中，这些变换是“对称性”的特有理念，表明它们不应该改变态之间的关系。特别是态

---

① 更具体地，伽利略群是 $(\mathcal{R} \ltimes \mathcal{V}) \ltimes (\mathcal{A} \times \mathcal{T})$，其中 $\times$ 是直积，$\ltimes$ 是半直积，$\mathcal{T}$、$\mathcal{A}$、$\mathcal{V}$ 和 $\mathcal{R}$ 分别是时间变换、空间变换、递升和旋转的(子)群(7.6.2 节)。首先，如果伽利略群被定义为从事件(时空)的欧几里得(Euclid)四维流形(7.5.2 节)E 向保留事件的同时性和同时事件间距离的其自身的仿射(平行线保留)映射集，那么得到的上述提到的子群不全是正规的，正如在人们期待使用直积的地方使用半直积时所表明的那样。

② 更细致的讨论见本书第十三章，布拉丁(Brading)和卡斯特拉尼(Castellani)的文章。

空间上的对称变换应该是这样的，即在态 $|\psi\rangle$ 中的系统对变换前后的可观测量产生相同的概率(至少对于那些假定在该对称性下不变或"随着" $|\psi\rangle$ 变换的可观测量而言，这样的话它们的本征矢量也随着变换了)。该如何表示这样的变换呢?

幺正算符(7.2.6 节)正好符合要求。实际上，我们把一个幺正算符部分地定义为是保持内积的那个算符。维格纳[1931，251]发现存在一个对该事实重要的近反演(near-converse)。

定理(维格纳)： 令 $\mathcal{H}$ 为在 $\mathbb{C}$ 上的希尔伯特空间，令 $T: \mathcal{H}\to\mathcal{H}$ 为对任意 $|w\rangle$，$|v\rangle\in\mathcal{H}$ 满足 $\langle Tw|Tv\rangle=\langle w|v\rangle$ 的一一(但没必要是线性的)映射。则：

$$T|v\rangle=\varphi(v)U|v\rangle \tag{51}$$

其中 U 可以是幺正的，也可以是反幺正的，且 $\varphi(\ )$ 是一个"相位函数"，它是 $\mathcal{H}$ 上的复值函数且它的值拥有模数 1。

任意的反幺正算符 T 可以写作 $T=UK$，其中 K 是"复共轭"算符。因此在取复共轭之后反幺正变换刚好是幺正的变换。例如，时间反演常常与复共轭相关联。

人们常常根据与这些变换"非物理的"性质相关的种种理由，排除掉反幺正的情形，特别是反幺正算符不与单位算符连续关联这一点。为了更精确地表述这一概念，人们需要引入群上的拓扑。在感兴趣的典型情形中，群是通过某组实指标被连续地参数化(7.6.4 节)的，从而使得群事实上形成了一个流形(7.5.2 节)，即它是一个李群(7.6.5 节)。在这些情形中已经给出了一个拓扑。与单位算符连续关联的意义仅在于，在这一情形下人们获得了群变换是从那些"无穷小"变换中建立起来的图景，无穷小变换是指"尽可能完全不对系统进行操作"(单位变换)。当然，假如我们谈论的只是对称性，则没有理由认为要与单位算符连续关联是一个必要条件——只提到最明显的例子，如时间反演或空间反射。另一方面，如果问题中的对称性被假定最终对应于真实的物理过程(如一个封闭系统的动力学演化)，则与单位算符的连续关联开始变得更为必要。

因此，一般用这些映射 T、幺正或反幺正的 U，还有通常对 U 是幺正的假

定(或希望)来表示量子理论中的对称性。

1.5.1.3 投影表示 在表达式(51)中人们不只是(通常)忽略 U 是反幺正的情形，也同样(通常)在寻找能够使 $\varphi(v)$ 等于 1 的映射 T。在此情形中，只是用一个幺正算符群给出关于对称群的表示。这样的表示特别有效，因为关于幺正算符我们知道得很多(一个重要的例子见第 1.5.1.4 节)。但人们并不总是能够如此幸运地找到这种常常被称为“幺正”或“正常”的表示(第 7.6.8 节)。有时候人们必须接受相位函数是非平凡的，在那里表示被称为是“投影的”。

原因如下：令$\mathcal{H}$是一个希尔伯特空间，考虑$\mathcal{H}$中矢量的等价类集合 $P\mathcal{H}$，在其中两个矢量是等价的，当且仅当它们处在相同的一维子空间中。$P\mathcal{H}$是一个投影希尔伯特空间，它的结构由$\mathcal{H}$上不同射线之间的夹角给出(射线归一化表示的内积的模数)。当式(51)中的相位函数为非平凡的时，得到的变换仍然产生 $P\mathcal{H}$ 的一个自同构。(此外，出于计算概率的目的，我们已经观察到，量子理论中的纯态能够像用态矢量表示一样表示为一维投影。因此，群的投影表示仍然可以保留所有的概率是不足为奇的。)所以，虽然普通的表示似乎很容易掌握，但用投影表示很方便，也不会带来问题，而且人们有时候不得不用到它们。

1.5.1.4 斯通(Stone)定理 幺正表示非常好用，因为它们能够通过自伴算符“产生”。首先，给定任意的自伴算符 F，算符 e^{iF}是幺正的。此外，其中 α 为实参数的算符 $e^{i\alpha F}$族形成了幺正算符的一个连续参数化的群，其中 $e^{i\alpha F}e^{i\alpha' F}=e^{i(\alpha+\alpha')F}$。(注意 $\lim_{\alpha\to 0}e^{i\alpha F}=I$，即该群与单位算符连续地关联。)现在假定我们想要把一个连续参数化的群$\mathcal{G}$表示为希尔伯特空间上的一族幺正算符。因为 $e^{i\alpha F}$的便于操作性，人们很容易找到一个产生$\mathcal{G}$表示的 F。我们很幸运有：

定理(斯通[1932])： 令 U_α 为希尔伯特空间$\mathcal{H}$上对$\mathcal{G}$的一个(弱)连续幺正表示。① 则$\mathcal{H}$上存在一个自伴算符 F 满足 $U_\alpha=e^{-iF_\alpha}$。

① 对每个 $|v\rangle$，$|w\rangle\in\mathcal{H}$而言，当且仅当$\langle w|U_\alpha|v\rangle$是 α 的一个连续函数时映射 $\alpha\to U_\alpha$ 是弱连续的。事实上，斯通定理在$\mathcal{H}$有一个可数基(即它是“可分离的”)时的弱条件下成立，在此情形下函数$\langle w|U_\alpha|v\rangle$仅需要是勒贝格可测的。见里斯和纳吉(Riesz and Sz.-Nagy)[1955，137]。

因为在非相对论的(事实上还有相对论的)量子力学中有如此多有趣的群拥有所需的特征，所以斯通定理对于这一理论具有根本的重要性。我们稍后会看到该定理应用方面的一些例子。

1.5.2 动力学

1.5.2.1 关于动力学演化的一些初始假设 系统态的动力学演化只是态空间上的一种变换。我从对一些假设进行简化来入手，这些假设将被证明对于确定动力学演化的形式是充分的。

记住密度算符是混合态，因此是纯态的线性组合。我们假定一个密度算符的演化是由组成它的纯态的演化引起的。这里想说的是，一个密度算符能够表示简单的物理混合，且至少在那种情形中它应该如所描述的那样进行演化。例如，假定我们有由两个不同类型的系统以比例 r 和 $1-r$(有 $0<r<1$)形成的一个混合，第一个系统处于纯态 P_1，第二个系统处于纯态 P_2。相应的密度算符为 $rP_1+(1-r)P_2$。如果系统的演化中二者之间没有相互作用(例如它们可能在物理上相互隔绝)，则人们可以期望的是如果处于 P_n 的系统演化到态 P_n'，则混合态演化到 $rP_1'+(1-r)P_2'$，或者我假定混合态的演化是这样。那样的话，我们就能够把注意力集中到纯态上，从而集中到希尔伯特空间中的(归一化)矢量上。

注意，此论证当然不适用于由分迹产生的密度算符(即“约化密度算符”)。事实上，一般这样的算符并不按照刚才描述的方式来演化。与直接确定它们的演化不同，人们可以从它们所组成的复合系统态的演化中得到约化密度矩阵——且如果复合系统处于非正常混合，则重复这一过程即可。

因此，我们主要关注孤立物理系统的动力学(虽然可以肯定的是系统一般会经历内部的相互作用)，它用一个纯态表示。从而问题变为：系统希尔伯特空间上的哪一个变换是系统态可能的动力学演化呢?

对称性再次派上用场。令 T: $\mathcal{H}\to\mathcal{H}$为从希尔伯特空间$\mathcal{H}$向表示封闭系统时间演化(经过某段给定的时间延续)的希尔伯特空间自身的映射。因为系统是封闭的，在一种我们已经注意到的意义上——对所有的 $|v\rangle$，$|w\rangle\in\mathcal{H}$有 $|\langle v|w\rangle|^2=|\langle Tv|Tw\rangle|^2$——假定 T 有一个对称性似乎是合理的(当然，该假定最终被证明在经验上是成功的)，因此式(51)适用。 我们进一步假定相位函数

$\varphi(\nu)$等同于1，即时间演化是由某个演化算符群的普通表示给出的。最后，回想反幺正算符并不与单位算符连续地关联，这意味着在此情形中如果用它们来表示时间演化，就不能够表示无穷小时间上的演化。在时间是连续的这一假定下，我们被引导到假定时间演化是由某个幺正算符群给出的。

现在注意到，这一结果囊括了我们对下述结论的原初论证，即本质上由于线性假设混合态的演化应遵照纯态的演化。作为在$\mathcal{H}$上的变换，幺正算符U同样产生了对$\mathcal{H}$上算符的变换，且在假定了动力学演化应当不改变纯态与由它们组成的混合态之间的关系（即可以单纯地用希尔伯特空间结构来定义的关系）之后，我们必须用U通过$W \mapsto U^{-1}WU$来产生对混合态的变换。事实上，这一表达式是量子力学中封闭系统标准动力学的最普遍形式。

为找到它的来源，考虑密度算符$W(t)=\sum_n w_n(t)P_n(t)$，并令$\{|\psi_{n,i}(t)\rangle\}$是由$W(t)$决定的正交基，其中下标$n$涉及的是谱投影$P_n(t)$，下标$i$涉及的是$P_n(t)$的维度，我们可以写出：

$$W=\sum_{n,i} w_{n,i}(t)\,|\psi_{n,i}(t)\rangle\langle\psi_{n,i}(t)| \tag{52}$$

根据我们早先的假设，W的演化由$|\psi_{n,i}(t)\rangle$的演化给出，这特别意味着系数$w_{n,i}(t)$是时间无关的。这里要记住我们假定W为封闭（孤立）系统的态。没有该假设的话，这里说到的所有事情都将不是那么看似可信。因此，如果U_t为系统的演化算符，我们可以简单地将其应用于直和项（summands）：

$$W(t)=\sum_{n,i} w_{n,i}(U_t|\psi_{n,i}(0)\rangle)(\langle\psi_{n,i}(0)|U_t^*) \tag{53}$$

$$=U_t\Big(\sum_{n,i} w_{n,i}|\psi_{n,i}(0)\rangle\langle\psi_{n,i}(0)|\Big)U_t^* \tag{54}$$

$$=U_tW(0)U_t^* \tag{55}$$

其中我们用到了U_t的线性。最后，回想对于任意的幺正算符U，有$U^*=U^{-1}$。

最后值得注意的是，对于$\mathcal{H}$上任意给定的幺正映射U，将它看作是“对称性”，对所有的算符F而言，映射$F\mapsto U^{-1}FU$是算符“恰当的”对应对称性，至少在下述意义上如此，即对于任意的$|v\rangle\in\mathcal{H}$和$\mathcal{H}$上任意的算符F，有$U(F|v\rangle)=(UFU^{-1})U|v\rangle$。也就是人们既可以“将算符$F$应用于矢量$|v\rangle$，再根据$U'$作变换”，也可以“根据$U$作变换，之后再将变换过的算符应用于变换过的矢量上”，这两种情形下的结果是相同的。

1.5.2.2 哈密顿量(Hamiltonian) 我们可以说$\mathcal{H}$上的演化是由幺正算符产生的。但是是哪一个幺正算符呢？例如，哪一个算符表示的是自由粒子的演化呢？哪个算符表示的是在某给定势能影响下粒子的演化呢？借助于斯通定理这些问题取得了一些进展。

上文中我论述了动力学演化所拥有的群的所有特征。特别地，令 $U_{1,2}$表示系统从时间 t_1 到时间 t_2 的演化，类似地定义 $U_{2,3}$，则从 t_1 到 t_2 之后再到 t_3 的演化似乎与从 t_1 到 t_3 的演化是相同的，换言之，$U_{1,3}=U_{2,3}U_{1,2}$。记住我们把这些 $U_{m,n}$看作在某个态空间上的操作——进而是对排序(ordering)的操作。

一个略强的但颇有说服力的假设是"时间均匀性(homogeneity)"。想象一个系统在某个时间无关的约束(例如不含时势能)影响下演化。那么，如果 $t_3-t_2=t_2-t_1$，则演化算符 $U_{1,2}$事实上应该和 $U_{2,3}$相同。(记住，这些算符是对整个空间的变换，我们不是在假定给定的单个系统从 t_2到 t_3 与从 t_1 到 t_2"做相同的事情"，而是假定在两个不同时刻处于相同态的两个不同孤立系统在接下来相等的时间延续中做相同的事情。)在此，演化算符仅要求标示相关时间间隔长度的单个参数，从而我们有关系式：$U_tU_{t'}=U_{t+t'}$。我们也假定 U_0为单位算符，意味着"瞬间什么也没有发生"。

注意加法群的类似性。事实上，满足此规则的一组算符 U_t形成了一个半群。但在此情形下也可能存在对应于逆演化(随时间后向的演化)的逆算符。把 U_t 的时间反演写作 U_{-t}，则我们要求 $U_tU_{-t}=U_{t-t}=\mathbb{I}$。那样的话 U_t 就形成了一个单参数群。

最后，我们将增加时间上的连续性假设。特别地，我们将假定群 U_t 的"弱连续性"。之后我们可以应用斯通定理获知，对于表示一个量子系统时间演化的任意群，存在着某个自伴算符 H 满足 $U_t=e^{-iHt}$。

对于给定的系统，我们如何知道要选择哪个算符 H？遗憾的是，此问题通常给出的回答是通过对类似经典系统哈密顿量的"量子化"来选择。对于这样一个系统，哈密顿量通常是经典的总能量。我说"遗憾"，是因为虽然在实践的层面上量子化一般是直观的，但关于在给定量子系统与其经典"类似物"之间的

“类似”这一性质，目前仍没有完全满意的基本说明。[①] 不过最常见的例子是直观的。例如，在一维空间中运动粒子的经典动能为 $p^2/2m$（其中 p 是经典的动量，m 是粒子的质量），量子理论（“量子化的”）哈密顿量为 $P^2/2m$（其中 P 是动量算符）。因而，量子理论中一个自由粒子的时间演化由 $|\psi(t)\rangle = e^{-iP^2t/2m}|\psi(0)\rangle$ 给出。

1.5.2.3 运动方程 在关于量子理论形式体系的最后这部分，我将简要介绍非相对论量子力学中的标准运动方程。许多量子力学的现实实践主要是或精确（在极少情况下可获得解析解）或近似（大多数情况下，要么通过微扰理论的标准技法，要么通过数值近似）地求解这些方程。

1.5.2.3.a 薛定谔方程 考虑时间演化群 $U_t = e^{-iHt}$。这些 U_t 唯一地求解了微分方程：

$$\frac{\partial U_t}{\partial t} = -iHU_t \tag{56}$$

但 $\frac{\partial U_t}{\partial t}|\psi(0)\rangle$ 仅是在时间 t 时 $|\psi(t)\rangle$ 的时间微商，所以允许式(56)两边的算符作用于 $|\psi(0)\rangle$ 上：

$$\frac{\partial}{\partial t}|\psi(t)\rangle = -iH|\psi(t)\rangle \tag{57}$$

方程(57)是不含时哈密顿量的薛定谔方程。类似的方程对于“左矢”也成立：

$$\frac{\partial}{\partial t}\langle\psi(t)| = \langle\psi(t)|(iH) \tag{58}$$

回顾我们之前关于狄拉克记号兼具实用性和简洁性的评价，这里请读者考虑此方程中的项究竟表示什么。当然，混合态的演化仍由式(53)给出。

若 H 依赖于时间，我们仍然能够考虑由幺正算符 $e^{-iH(t)}$ 给出的每一个无穷小演化（从 t 到 $t+dt$）。一般地，从这些无穷小演化来构建有限时间的演化算符是有意义的。但如果哈密顿量不含时的话，则我们当然可以定义 $U_t = e^{-iHt}$。此外，如果系统在时间 $t=0$ 时处于“定态”，即它处于具有固定能量 E 的能量本征态 $|\psi_E(0)\rangle$，有 $H|\psi_E(0)\rangle = E|\psi_E(0)\rangle$ 则演化具有简单的形式 $|\psi_E(t)\rangle =$

---

① 见本书中兰兹曼（Landsman）所撰写的章节。

$e^{-iHt}|\psi_E(0)\rangle = e^{-iEt}|\psi_E(0)\rangle$。也就是说系统仍处于同一个一维子空间中，仅是相位 e^{-iEt}在随时间改变而已。

当然，式(57)表明了期望值是如何改变的。将它直接应用于任意的算符 F，则式(57)表明：

$$\frac{d}{dt}\langle F\rangle = -i\langle [F, H]\rangle + \langle\frac{\partial F}{\partial t}\rangle \tag{59}$$

其中$\langle\cdot\rangle$是处于某个(这里未确定)态中的算符的期望值，H 是哈密顿量。

这就是对"薛定谔演化"的标准说明。然而，人们应该注意到在实践中系统常常不是孤立的，所以有必要在哈密顿量上添加势能以表示某些外部系统 σ_{ext} 对目标系统 σ 的影响。这么做的问题是，试图模拟整个复合系统(σ_{ext}和 σ)通常不切实际，而将 σ_{ext}对 σ 的影响模型化为应用于 σ 的外部势至少有一线成功的机会。一个典型的例子是纳米电子学，其中人们可能对研究施加于分子两端的势差感兴趣。原则上，人们可以把分子的任意一端看作是电极。而实际中，仅是模拟分子就已经非常困难了，把系统看作电极则完全是不可行的。取而代之的是，人们只把它们对分子的影响当作哈密顿量中的势。一般而言这样做会得到非幺正演化(因为系统不是封闭的)。

1.5.2.3.b 薛定谔绘景和海森堡绘景 我们一直在用态的演化来描述动力学。当然人们也可以等价地将态看作是随时间的不变量，而让可观测量来演化。假定系统的态为$|\psi(t)\rangle$，在幺正演化 U_t 下演化。从而可观测量 F 在 t 时刻的期望值为$(\langle\psi(0)|U^*)F(U|\psi(0)\rangle)$。所以我们可以使态随时间不变，即$|\psi(t)\rangle = |\psi(0)\rangle$，而假定可观测量按照 $F(t) = U_t^* F(0) U_t$ 变化。显然，期望值的表达式在两种情形下是相同的。因此，这两种绘景是经验等价的。

第一种绘景(其中态随时间演化，可观测量不变)通常称为"薛定谔绘景"，第二种绘景(在这里态不变，可观测量随时间演化)通常称为"海森堡绘景"。①

1.5.2.3.c 海森堡方程 在海森堡绘景中，可观测量如何随时间变化呢?

---

① 存在第三种绘景，即结合前两种绘景的"相互作用绘景"(有时称为"狄拉克绘景")。在此绘景中，态和可观测量都随时间演化。由其哈密顿量的自由部分引起的系统演化计入态的演化中，且由哈密顿量的其余部分(哈密顿量的"相互作用"部分)引起的系统演化计入可观测量的演化中。

我们将“薛定谔”可观测量(暂时地)标示为“S”，相应的“海森堡”可观测量标示为“H”。我们也假定薛定谔可观测量并不明确含时(正如在基本应用中常见的情形)，则在任意有限时刻我们有 $F^{H}(t)=U_t^* F^S U_t$(见第1.5.2.3.b节)，从而：

$$\frac{dF^{H}}{dt}=\frac{\partial U_t^*}{\partial t}F^S U_t+U_t^* F^S\frac{\partial U_t}{\partial t} \tag{60}$$

$$=iU_t^* H^S U_t U_t^* F^S U_t-iU_t^* F^S U_t U_t^* H^S U_t \tag{61}$$

$$=i[U_t^* H^S U_t,\ F^{H}] \tag{62}$$

$$=i[H,\ F^{H}] \tag{63}$$

其中我们首次用到了式(56)且在每一项中插入了 $U_t U_t^*$，并且之后我们用到了 H^S 与 U_t 对易的事实，所以有 $U_t^* H^S U_t=H^S=H^{H}$。因此我们去掉式(63)中哈密顿量的上标。该方程通常称为“海森堡方程”。①

在海森堡运动方程和经典可观测量(相空间上的函数)$f(x,\ p)$的运动方程之间存在一个重要的相似性，用泊松括号表示为：

$$\{f,\ g\}=\frac{\partial f\partial g}{\partial x\partial p}-\frac{\partial f\partial g}{\partial p\partial x} \tag{64}$$

对易子和泊松括号有着类似的代数特征，此外经典运动方程由 $\frac{d}{dt}f=\{H,\ f\}$ 给出，其中 H 是经典的哈密顿量。因此一种关于经典与量子理论间关系的考察是通过对易子与泊松括号之间的代数相似性进行的。②

1.6 哲学问题概述

非相对论量子理论的哲学在传统上主要关心四个问题(当然这并不是说不

---

① 在薛定谔绘景中也有密度矩阵演化的类似方程：$\frac{d}{dt}W=-i[H,\ W]$。其推导类似于上面给出的那个推导过程，注意正负号的变化。这里并不存在冲突，此方程描述了薛定谔绘景中密度矩阵(态)的时间演化，而方程(63)描述了海森堡绘景下算符(可观测量)的时间演化。

② 狄拉克打算用此类比来定义量子化。结果证明此类比不能够严格完成，至少不能以其最初的形式完成。关于量子化和此代数类比的进一步讨论，见兰兹曼在本书第五章的论述，特别是第4节。对经典力学中泊松流形(进而对类比的经典方面)的进一步讨论，见巴特菲尔德在本书第一章的论述，特别是第5节。

讨论其他问题)。

第一个问题(在第2节中讨论)即量子理论形式体系的起源或依据，它在某种程度上是最为基本的，但也是在某些圈子里最少被讨论的。我们是如何实现对希尔伯特空间和其上的自伴算符而非对相空间和其上函数的应用的?考虑到量子理论哲学家们面临的其他问题在形式体系的某些特征中有所反映，所以对这些问题的回答(或至少是提出它们)将会证明是非常有用的。对于为什么量子理论采用它所采用的形式体系这一问题，也有一些相当有说服力的尝试性说明。下一节将讨论这样的一些尝试。

第二个问题(在第3节中讨论)是关于经验内容的。例如，对一些量子理论创始者们的仔细审视，尤其是尼尔斯·玻尔(Niels Bohr)，揭示出了他们对量子形式体系如何获得其经验意义这一问题的深切关注。受玻尔启发，我将提出一种用于理解形式体系和观测之间联系的方案。

经验内容的议题和不相容性问题，更精确地说是和不确定关系(在第4节中讨论)有着紧密的关联。为什么不可能确定所有可观测量的精确值?它是根据下述更深刻的论述得出的吗?即拥有确定值的可观测量会阻止另一个不相容的可观测量也有精确的值?一些量子理论的创始者们(特别是玻尔)非常关注对下述问题的解释，即对形式体系某些部分的经验确定所需的条件会如何以某种方式，阻止对形式体系其他部分的经验确定所需条件的同时满足。而其他人(特别是爱因斯坦)反而关心如何“避免”不确定原理，他们最初的做法是论证人们能够确定不相容可观测量的精确值，后来是通过在著名的爱因斯坦—波多尔斯基—罗森(Einstein-Podolsky-Rosen)论文中对下述较弱陈述的论证：即它们必定(同时)拥有精确值，即使我们不能够确定(如测量)这些值是什么。

第四个问题(在第5节中讨论)——仍然与前一个问题相关——产生于盖里森定理：若量子理论中不存在无色散(有时在此背景下称为“二值”)的态，则显然一些可观测量有时会处于“不确定”的态。甚至更坏的情况是，在对态与可观测量值之间关系的标准解释下，很容易描述出物理上看似合理的情景，即人们通常确信其确定的可观测量不能够通过量子理论给出的态而被赋予确定的值。薛定谔描述了这样一个著名的情形，其中猫是某种“非活非死”的状态。通过从认识论上对量子态产生的概率进行解释(即就像经典概率通常那样，解释

为对真实值无知的表达)来回避此问题的显著方式，似乎陷入了严重的哲学困境，确切地说是逻辑困境。因此，人们提出了其他的解决方案，它们有时是对形式体系非常奇特的解释。虽然可以看到许多这些解答避免了基本的问题，但每一个都有其严重的隐患以及随之而来的危害。当前对这一问题尚没有普遍一致的对策。

最后一个主要问题(在第6节中讨论)是理论的非定域性，正如用张量积空间表示复合系统时所表明的那样。正如我们所注意到的，这样的空间允许所谓的“纠缠”(不可因子化)态，即表明了在远距离系统之间的强(甚至是全)关联。同样，用在时空中纯粹定域传播共因历史或过程来解释这些关联的显著方式并不奏效。因而人们提出了一系列的定理，其中第一个且最著名的定理是由贝尔提出的[1964]。这一结果与20世纪或更久以前时空理论学家们所使用的某种形式的“定域性”原理倾向，有着明显的冲突。虽然为了解决这一明显冲突，或论证它并不是一个问题，人们做出了许多不同的尝试，但仍没有非常令人满意的解决对策。

2. 运动学形式体系从哪里来?

本节中我们的目标是回顾一些对为什么我们在量子理论中采用我们所使用的这一形式体系进行尝试性的说明。有许多种想法试图从物理上“直观的”公理中“得到”希尔伯特空间形式体系。这里我们应当采取的态度是，认为许多尝试是有价值的——它们能够帮助我们理解量子理论——即使它们都没有成功地从物理意义(更不用说真值)总是明确的公理中推导出形式体系。因此多审视几种通往该形式体系的方法，而不是仅仅依靠其中的一种方法是有意义的。这里我们会考虑几种重要类型的典型。

第一种类型(第2.1节)开始于物理命题的概念，且论证说这些命题是通过希尔伯特子空间得到恰当表示的。第二种类型的两种思路(第2.2节)开始于物理态的概念，论证说这些态被恰当地表示为希尔伯特空间子空间上的概率测度。最后一种思路(第2.3节)与前两种思路相类似，他们都试图用某种抽象的形式(用 C^* 代数)来刻画态空间并将它们“表示”为希尔伯特空间上的态，但不

同的是最后一种思路直接构建了布洛赫球(第1.3.5节)并从中建立了(态空间的)高维希尔伯特空间。(然而，话说回来我们事实上并不会充分审查这些论证，从而注意到它们在细节上的不同。)

我们的意图不是宣称这些方法一劳永逸地阐明了在量子理论中应用希尔伯特空间的真正原因。相反，每一种情形都阐明了量子理论中应用希尔伯特空间的某些有趣或重要的方面，它们是通过从相对无争议的和物理上清晰的原理到希尔伯特空间形式体系的论证来阐述的。这里我们把着重点放在每种方法中那些对理解量子理论可能有用的方面，而忽略仅是技术条件的那些。在每种情形下，我们的讨论将尽可能简略(相对于全文而言)，证明和技术性细节将被省略。

最后，这里还要讨论对于量子理论基础而言很重要的许多概念(如命题格和态的凸空间)。那些不打算了解通往希尔伯特空间形式体系道路的读者通过阅读本节中某些部分内容至少也可以学到其他一些重要的东西。

2.1 从命题到希尔伯特空间

本节中，我们追溯一条开始于物理命题逻辑的通往希尔伯特空间的道路。这里的目的是把量子理论(即希尔伯特空间形式体系)看作是某种意义上满足某些逻辑约束条件的唯一理论。①

2.1.1 命题格

2.1.1.1 命题的物理意义 一开始我们把关于某时刻物理系统的命题集看作是基本的。其要义是这样的命题将采取“系统在时刻 t 有性质 P”的形式。(因此，人们也可以等价地把性质看作是基本的。对于每一个性质，都有表述系统在给定时刻拥有此性质的相应命题。这里我们将只用这些命题来论述。)

2.1.1.2 定义 在逻辑的代数方法中，一般会假定所有命题(符合语法的

① 这一方案有许多个版本，该方案开始于伯克霍夫和冯·诺伊曼(Birkhoff and von Neumann)[1936]。这里的讨论主要是在皮朗(Piron)[1976]理念下进行的。基本上涵盖了此领域的近期著作包括道拉斯·基娅拉(Dallas Chiara)、朱利尼(Giuntini)和格林奇(Greechie)[2004]。更有操作性的方法——某种程度上不同于其他方法——是由路德维希(Ludwig)[1983]提出的。

句子)的集合形成了一个格(第7.4节)。格中的偏序对应于蕴涵$P\leqslant Q$，即P蕴涵Q。格的上限(supremum)(并集)对应于析取(disjunction)，且下限(infimum)(交集)对应于合取(conjunction)。这些认定远不是任意的。以析取为例：在格L中两个元素P和Q的并集，是P和Q二者所蕴含的逻辑上最弱的命题，这可以说就是“析取”的意思。对于合取也可做类似的考虑。最后，假定(这也是很普遍的)格是正交格，格的正交补(orthocomplement)对应于否定。

2.1.1.3 格操作的动机和解释 除了直接诉诸代数逻辑之外，这样的格有时也受到对测量操作性描述的启发。例如，人们用对是—否实验(yes-no experiment)的谈论来检验某些命题的真假。此检验集合上的偏序来自于下述理念，即一个检验Q会在每次某个其他检验P通过时通过，在此情形下我们写作$P\leqslant Q$。然而这里我不打算追求该方法的细节，而是代之以将存在某种用物理术语来理解此逻辑操作和关系的方式视作理所当然。(我并不是在表明某个特殊的方法——尤其是许多操作性方法——是完全令人满意的。)事实上人们可以证明对格理论连接词物理意义的深刻理解是通过命题通往希尔伯特空间的道路中的致命弱点。①

2.1.1.4 命题格上的附加约束 除了已经建立起来的命题格结构，我们还需要假定一系列附加的性质。首先我们假定格有一底一顶(第7.4.5节)0和1，分别对应于逻辑上为假的命题和逻辑上为真的命题。(然后我们需要正交补，即对于所有的P有$P\vee P^{\perp}=1$，因此有$0^{\perp}=1$。)

其次，我们假定命题格是完备的和原子的(第7.4.5节)。该假设等同于这样的理念，即存在一些基本命题——极大确定的命题——从而人们能把较弱命题的真解释为某些基本命题真的后承(这并不是说人们必须要以这种方式来解释它们)。另一种为该假定辩护的方法涉及态集合的凸结构。我们将在下文看到(第2.2.1.3.c节)，有很好的理由可以推断此集合有对应于纯态的极值点，即极大信息的态。该特征表明，每个纯态应当赋予某个极大(逻辑上最强的)命

① 当然我并不是想说量子逻辑理论的倡议者们未曾提出过该问题。例如，约赫[1968]提出了一种用重复测量极限表示的对非对易投影合取(显然它不能够简单地用命题“两个都被测量”来理解)的解释。

题以概率 1，该极大命题应当是命题格中的一个原子。

接下来我们必须要假定格是不可约的(第 7.4.4 节)。不可约性假定虽然并不是没有价值，但它确实拥有接近于物理解释的某些东西，这是根据格理论中的下述定理得出的：

一个正交格不可约，当且仅当其中心(7.4.3 节)是平凡的，即{0，1}。

现在命题处于中心在逻辑上意味着它与所有其他命题相容(第 1.2.4 节)。因此，理解格不可约性的一种方式是将其作为格中心是平凡的这一假定的后承，即每个命题(除了 0 和 1)都至少与一些其他命题不相容。注意此假定在某种意义上是对经典定律 $p=(p\wedge q)\vee(p\wedge q^{\perp})$ 的"极大"违背，因为它断言对于每个 p 定律都不成立。这并不是说不存在一些使定律成立的 p 与 q，而是说对于每个 p，都存在某个使得定律不成立的 q——同样回顾围绕条件(23)的讨论。

我们最后的假定是，L 满足覆盖律特征(7.4.5 节)。遗憾的是虽然给出了某些论证，但其中的动机不是那么清楚。①

2.1.2 皮朗理论

皮朗理论精确刻画了那些是(同构于)希尔伯特空间子空间格的格。他只得到了下述定理：

定理(皮朗[1964])： 若 L 是一个完备的、原子的、不可约的正交模格，满足覆盖律且有至少 4 个正交原子，则它是(同构于)内积空间 V 的子空间格。

事实上，皮朗也可以讨论一下 V 在其上定义的场，但他并没有这样做。最终的结果是，皮朗定理是启发性的，但远不是所期望的终点，即把希尔伯特空间的子空间格刻画为满足某些逻辑约束的唯一结构。

皮朗理论经对如下所述的索勒(Solér)定理的论证而取得了巨大的进展：

① 例如：皮朗[1964]、科恩与斯韦特利奇内(Cohen and Svetlichny)[1987]。

定理(索勒[1995])：若皮朗定理表述中的格L包含一个无穷的规范正交序，则问题中的矢量空间是一个包含实数、复数或四元数的希尔伯特空间。

这一结果是对皮朗理论的重要贡献，虽然它的应用性受到了明显的限制：它并不涵盖有限维希尔伯特空间的情形(如第1.3节中那些描述粒子自旋的情形)。

2.2 从态到希尔伯特空间

2.2.1 态的凸空间方法

现在我们考虑两种从态空间结构开始通往希尔伯特空间的方法。第一种方法始于对态构成一个凸集(第2.2.1.1节)的观察。我们的程序是把一任意的凸集嵌入到矢量空间V中(第2.2.1.2节)，主张(在随后的部分中)逐渐给V增加更多的结构，直到它有足够的结构支持达到下述效果的一个定理，即态的凸集事实上能够表示为希尔伯特空间上的密度算符。

2.2.1.1 态的凸空间 早先我们注意到给定任意两个概率测度p和p'，它们可以组合形成第三个概率测度$q = rp + (1 - r)p'$，其中$0 < r < 1$。很容易验证这样定义的q满足柯尔莫哥洛夫公理(第7.5.6节)，若p与p'都满足的话。事实上，概率测度的任意凸组合(第7.1.10节)都产生另一个概率测度，它称为出现在凸组合中的测度“混合”。

一般的观点是把态的凸空间看作是基本的。关于这一点，除了它是凸的之外我们并不对态空间作出承诺。之后我们将对此空间施加额外的条件，逐渐迫使我们的态空间事实上成为希尔伯特空间上的态空间。换言之，我们将在希尔伯特空间上描绘态空间(密度算符)。①

① 本节中的讨论很大程度上是对哈格(Haag)[1992，§VII.2]所描绘理论的扩展和解释。该理论自阿夫森和舒尔茨(Alfsen and Shultz)[2003]那时起就采取了一种新的形式，下文会简略提到这一点。

2.2.1.2　矢量空间的嵌入(embedding)　开始时，考虑任意的凸集 S。把 S 嵌入一个实的矢量空间 V 在数学上是自然的和简便的，这部分是因为凸组合仅是实线性组合的一种特殊类型，后者是在(实)矢量空间的背景中自然地定义的。

若 S 是由其极值点产生的，则我们当前的任务很容易。(一个凸集的极值点是这样的点，即它本身不是集合中其他点的凸组合。仅当 S 的每个元素都能够表示为极值点的凸组合时，S 才可以由其极值点产生。)那样的话，我们可以通过令 V 为由 S 的极值点产生的自由矢量空间来定义嵌入。(从而直观上，V 是 S 极值点的所有形式的实线性组合。)但是，显然这一方法要求我们假定 S 具有极值点，且正如通常所理解的那样在量子理论中是这样假定的，但值得注意的是我们没有必要如此假定，因为能够从其他考虑中得到这一事实。(在 S 是态的集合这一物理理论背景下，极值点恰好是纯态——回顾第 2.1.1.4 小节。)令 V 是包含 S 为其子集的实矢量空间(即 S 嵌入在 V 中——给定 S 的实线性结构，应该明确的是总可能找到一个矢量空间，它包含一个与作为凸集的 S 同构的子集，这里我们的问题不是表明存在这样一个 V，而是像上面一样从 S 来构建它)。

2.2.1.3　极值点存在的充分条件

2.2.1.3.a　可观测量　我们将把理论中任意的可观测量 f 看作从态到期望值的函数。即 $f(v)$ 是 f 在态 v 中的期望值。事实上，在把态嵌入到矢量空间 V 之后，我们将把可观测量看作是 V 上的实值线性函数(可能是其子集)。显然，被理解为从态到期望值映射的可观测量应该是线性的函数，因为对于任意的态 $v=\sum_n w_n v_n$(这里表示为态 v_n 的凸组合)，我们必有 $f(v)=f(\sum_n w_n v_n)=\sum_n w_n f(v_n)$。否则的话，$f(v)$ 一般而言在数值上将不同于 f 在从混合态 $\sum_n w_n v_n$ 中随机选择的系统上的期望值。

令 O 为可观测量集。我们(尚)不需要对 O 的内容做出承诺，除了要求它分离 V，即对于任意的非零 $v\in V$，存在一个 $f\in O$ 满足 $f(v)\neq 0$。等价地，若 v_1

$\neq v_2$，则存在$f\in O$满足$f(v_1)\neq f(v_2)$。[1] 在S中，这一条件等于要求O足够丰富以对S中的元素做出概率区分。为了推广到所有的V，现在假定O分离了S，并且注意，如果O并没有分离所有的V，则事实上V对于嵌入S“过于大”了。因为考虑所有v的子空间，对于所有的$f\in O$都有$f(v)=0$。之后考虑商空间V/W。存在从V到V/W的一个同态，它是从S到凸集V/W中它的映像的同构，所以我们也可以只用V/W来处理，且这样做事实上是从V中去除了那些对于嵌入S不必要的结构。

2.2.1.3.b 嵌入空间上的拓扑 我们假定O分离了V。现在为V引入一个拓扑(7.5节)。这里的指导思想是，态的“无穷小”变化应该引起期望值的“无穷小”变化。[2] 我们因此在V上引入最粗粒化的拓扑，它使得所有的O都连续，称它为O拓扑。[3] 在此拓扑中，要求S是紧致的在某种程度上是合理的(7.5.1节)。[4] 例如，这一假定保证了那个可观测量(的期望值)是有界的，因为紧致集的连续映像是紧致的。

2.2.1.3.c 极值点的存在 给定这些关于S及其在矢量空间V中的嵌入，还有相关可观测量O的假定，我们可以根据函数分析应用下述定理：

> **定理(克莱茵—米尔曼(Krein-Millman)和鲁丁(Rudin)[1973, 70])**：令S为V的凸子集，O是V上的线性函数的分离集合，且令S在O拓扑中是紧致的，则S具有极值点，且它是包含所有这些点的最小的闭凸集。

---

① 根据哈恩—巴拿赫(Hahn-Banach)定理，此条件总是可以满足的，即人们总是可以找到实现此目的的某些线性函数，对任意赋范的矢量空间而言。我们的空间V事实上是赋范的。

② 虽然表面上合理，但此假设并不具有竞争力。在哈迪(Hardy)方法的背景下对此观点的讨论见第2.3.2节。

③ 此拓扑的基是由B的所有集合(“开球”)给出的，构造如下。选择$v\in V$, f_1, f_2, …, $f_n\in O$和$\epsilon>0$。B则是所有$w\in V$的集合，满足对$i=1$, …, n有$|f_i(w)-f_i(v)|<\epsilon$。

④ 在我们即将在其中操作的具体情境中，紧致性相当于单位元——每个地方都具有值为1的线性函数——是一个可观测量的假设。

人们说 S 是“其极值点凸包的闭包”。换句话说，S 具有极值点，且在 S 是极值点所有凸组合的闭包这一意义上是由这些极值点“产生的”。因此，S 中的每一个元素都能够写为极值点(纯态)的一个凸组合，或是这些态序列的极限。①

2.2.1.4　嵌入的更多特征　回顾将内嵌空间 V 构造为涵盖 S 极值点的自由矢量空间的方法(2.2.1.2 小节)。现在在证实了 S 确实拥有极值点之后，我们几乎能够遵照这种方法。唯一要补充的是我们希望 V 在 O 拓扑下是闭的。因此 V 被构建为涵盖 S 极值点的自由矢量空间的闭包。(这里的“闭包”我们指的是：保证 V 中的每一个开集的闭包同样在 V 中。)在此情形下，S 中的纯态形成了 V 的一个基(因为事实上涵盖纯态的自由矢量空间本质上是纯态的“所有形式的线性组合”)。

我们现在坚称 O 包含一个线性函数 I，它赋予 S 中的每个纯态(极值点)以值 1。这个 I 是唯一的(因为纯态形成了一个基，且 I 按照假定是连续的，从而从纯态唯一地延拓到了所有的 V)。现在，令 V^+ 是由纯态产生的正的凸锥(第 7.1.10 节)，V 中的态则被看作是将 I 取值为 1 的 V^+ 的元素。

事实上，I 以一种自然的方式产生了 V 上的模。对于任意的 $v \in V^+$，定义 v 的模为 $\| v \| = I(v)$。现在对于某些 v_1，$v_2 \in V^+$，任意的 $v \in V$ 可以分解为 $v_1 - v_2$(或者是可以如此分解的一个序列的极限)。② 因此我们可以把 $\| v \|$(现在对于任意的 $v \in V$)定义为 $\| v_1 \| + \| v_2 \|$ 的(涵盖所有这些分解的)下限，且要求模为连续的。V 因而是一个实巴拿赫空间(7.1.7 节)，且(通过构建)V 上的模赋予每个态以模 1(我们现在实现了我们早先的承诺，即 V 原来是一个赋范空间)。

2.2.1.5　面(face)和命题　这些态将赋予概率什么呢？虽然要结束此讨论(在此层次上)我们不需要回答这一问题，但是这样做是有益的：态赋予其概率

---

①　小心！那些熟悉量子理论的人们可能会假定，V 是(或同构于)包含系统态矢量的希尔伯特空间，事实上并不是。正如我们将看到的(一旦我们在 V 上定义了模)，它是一个矢量空间，在其中密度算符形成了单位球面。

②　论证概要：纯态生成 V，从而将 v 写作纯态的线性组合，并将它分解为具有正系数的一部分和具有负系数的一部分。前者显然在 V^+ 中。后者，乘 -1，也在 V^+ 中。

的"命题"是在态的凸集中的态的面(第7.1.10节)。

在经典力学中，人们可以在凸组合下，通过取纯态集的闭包来形成一个凸集的面。在量子理论中，纯化过程一般会在集合中增加新的纯态，所以它对于面的构建是必不可少的。(回顾经典的态也形成了一个单形，而量子态并不会——见1.2.7节。)在两种情形下的面的定义背后的物理理念类似于下述理念，即"可以(从某些初始集合)通过混合而产生的所有态的集合，加上得到的混合态的所有态的集合，在原则上是一个混合态"。

在态的面和关于一个物理系统的命题或物理系统的特性之间有着自然的关联。如上，这里我们可以交换使用术语"特性"和"命题"。同样如上，我们开始于如下最小假设：命题集是一个偏序集(poset)，偏序对应于蕴涵。特别地，我们说对系统的测量断言揭示出系统态的命题(关于某一时刻的一个系统)是态的某个面。

注意到态空间中的一个极值点是一个单元素(singleton)面，反之也成立。因此，对于任意的纯态 v，一种命题形式为"系统在态 v 中"。这种关联是有意义的，因为纯态在直观上被假定是极大信息的态，且如果命题对应于面，则最具体的命题是单元素面(极值点)。

在命题和面之间普遍关联背后的理念叙述如下：假定给你一个其中系统都处于相同态的系综，并且要你决定将系综制备于什么态。给定在极大确定命题和纯态之间的关联，如果你能够确定某个极大确定命题对于系综的每个成员都为真，则你已经做到了——态就是相应的纯态。但假定不存在这样的命题，即对于系综中每个成员没有极大确定命题是真的。这样的话交给你的是一个混合态，但是是什么的混合呢？一般地，你仅能够肯定地确定的是，该混合态将通过来自最小面的态进行混合而产生，其中最小面包含了真实的混合态(因此，真实态赋予最小命题以概率1)。

在标准量子理论术语中，这里的要点是，对于任意的混合态(密度矩阵) W，通过 W 确定得到的逻辑上最强的命题只是 $\mathrm{ran}W$，这里对概念"范围"推广以表示"在希尔伯特空间中的所有矢量下的 W 映像"。(不足为奇的是，$\mathrm{ran}W = V_n P_n$，其中 P_n 是 W 的谱投影。)换言之，$\mathrm{ran}W$ 事实上是形成包含 W 的最小面的纯(矢量)态子空间。

2.2.1.6 锥面的希尔伯特空间表示

2.2.1.6.a 齐次锥 考虑态集合上的自同构(即从态集合向其自身的保留凸结构的映射)。这样的映射似乎对应于可能的态跃迁。此外，它们很自然地被延拓到与 V^+ 自同构的 V 上的线性映射(即这样的映射将 V^+ 映射到其自身且保留了 V^+ 的凸结构，因此它们保持了 V^+ 是由纯态产生的正凸锥这一事实)。向 V^+ 的延拓仅受如下条件的影响：对于任意这样的自同构 f、任意的实数 r 和任意的态 v，有 $f(rv)=rf(v)$。回顾 V 的每一个元素都可以写作 V^+ 中元素的线性组合，所以人们可以看到 f 因而被自然地延拓到了 V 上的线性变换。

当对于锥中的任意两个非极值点 v，v'，T 中有一个将 v 变换到了 v' 的变换时，可以说该锥相对于变换的集合 T 是"齐次的(homogeneous)"。这里的物理理念是，存在某种使系统从任意的非极值点到其他任意非极值点的演化方式。

2.2.1.6.b 有限多纯态的情形 虽然我们不想假定仅存在有限多个纯态，但是下述定理仍然是非常有启发性的：

> **定理(温贝格(Vinberg)[1965])**：一个自对偶的[①]具有有限多极值点的齐次锥的面，一一对应于某些希尔伯特空间(涵盖了实数、复数或四元数)的子空间。

因此，如果人们确信理论的态必定形成了具有上面所讨论的特性的一个凸集合，则在该定理给定的意义上，人们确实得到了希尔伯特空间形式体系。

2.2.1.6.c 一般情形 阿夫森和舒尔茨[2003，414]将此方案做了推广以去除不切实际的(且事实上是假的)假设，即仅有有限多个纯态这一假设。他们的主要结果涉及了许多关于态的凸集结构的技术性假设，这里我们没有足够的篇幅详尽地表述它们。此外，虽然他们可以通过一个 GNS 构建得到希尔伯特空间，但最终，他们得到的并不是希尔伯特空间(至少不是直接地)的特征，而

---

① 我们省略了此条件的物理动机问题。见第 7.1.10 节。

是 C^* 代数态空间的特征。[①]

2.2.2 纯态和 C^* 代数方法

在此意义上(即最终处理的是 C^* 代数的纯态),阿夫森和舒尔茨的定理与我们将要简短概要考虑的兰兹曼[1998, Theorem 3.9.2 and Gorollary 3.9.3]提出的定理相类似。

兰兹曼论证道,纯态集应该被赋予动力学的和概率的两种不同结构,且这两种结构必定以恰当的方式相联系。他之后还补充了特征性的量子条件是什么(描述见下文),并且得到了一个刻画 C^* 代数态空间的定理。

2.2.2.1 泊松(Poisson)结构

2.2.2.1.a 作为泊松流形的态空间 第一种结构对应于我们之前讨论过的动力学。回想我们对海森堡方程的讨论(第1.5.2.3.c节),确切的是(i 乘)对易子具有泊松括号的代数形式这一事实。简而言之,用泊松括号产生的动力学一般理论运作如下。

我们开始于某个态空间(确切地说是流形——第7.5.2节)M。给定 M,人们采用定义在 M 上无穷可微的实数值函数上的泊松括号,将可观测量的演化 $C^\infty(M)$(可观测量)定义为 $C^\infty(M)$上的双线性操作$\{,\}$。这里可回顾式(64),更多的讨论见本书中巴特菲尔德的第一章第5节。同时通过将标量函数 H: $M \to \mathbb{R}$ 选择为哈密顿量,人们可以用可观测量演化的这一定义来定义 M 上的(态的)演化(与海森堡绘景和薛定谔绘景间的等价性相类似)。对于给定的哈密顿量,由此定义的来自 M 的可能动力学路径是系统的"哈密顿曲线"。

记住这一构造被认为是非常普通的。我们基本上明确了某种动力学系统"意味着什么"。为此,兰兹曼对作为泊松流形的态空间施加了"表现良好(well-behavedness)"的附加限制,这里我们避而不谈。

2.2.2.1.b 辛叶子 有可能泊松流形的某些部分是其他部分"不可达的"。特别地,对于任意的哈密顿量,可能不存在连接它们的(分段光滑的)哈密顿曲线。我们说泊松流形 $\langle M, \{,\} \rangle$ 中的两个点 x, $y \in M$ 是辛等价的(symplecti-

---

① 更多关于 C^* 代数、GNS 构造和相关的问题,参见兰兹曼在本书第五章、霍尔沃森在第八章和埃姆什在第十章的论述。

cally equivalent)，如果对于某个哈密顿量它们处在单条(分段光滑的)哈密顿曲线上。这一关系显然是一个等价关系，因此将 M 分割为"辛叶子"(见本书巴特菲尔德撰写的第一章第 5.3.3 节)。

2.2.2.2 跃迁概率结构

2.2.2.2.a 作为跃迁概率空间的态空间 在经典哈密顿量相空间上的唯一动力学结构是它的泊松结构(见本书第一章，特别是其中 5.2.4 节)。然而，在标准量子理论中，存在两种类型的演化：用泊松括号描述的连续的决定性的演化，和通常与"测量"相关的、不连续的、随机的从一个态向另一个态的"量子跃迁"，见 5.4.3 节。

后一种结构反映出这样的事实，即在兰兹曼的方案中量子态空间必定是一个"跃迁概率空间"，这意味着必定存在从空间中的元素对向[0，1]的映射 p，当且仅当 $v=w$ 时满足 $p(v, w)=1$；当且仅当 $p(w, v)=0$ 时有 $p(v, w)=0$。表达式 $p(v, w)$ 读作"从 v 到 w 的跃迁概率"。此外，我们要求这些概率是对称的，即 $p(v, w)=p(w, v)$。

2.2.2.2.b 分支(sector) 一个跃迁概率空间的分支是与其余部分隔离开来的空间区域。即对一个态分支 Q 而言，对于所有的 $v\in Q$ 和所有的 $w\in Q'$(它是所有态的集合 S 中 Q 的补集)，有 $p(v, w)=0$。注意 p 的对称性表明系统不能作出到或开始于一个分支的跃迁。

2.2.2.2.c 叠加 令 $Q\in S$。我们定义 $Q^{\perp}$ 为：

$$Q^{\perp}:=\{v\in S \mid p(v, w)=0\ \forall w\in Q\} \tag{65}$$

即 $Q^{\perp}$ 是 Q 中的每个态不可达(通过单个的概率跃迁)的所有态的集合。我们可以用此定义来刻画"叠加"的普遍概念：态 v 和 w 所有"叠加"类是 $\{v, w\}^{\perp\perp}$，见[Butterfield，1993]。

2.2.2.3 兰兹曼定理 兰兹曼粗略地指出，假定(还有其他)辛叶子对应于分支，则 C^* 代数的态空间是通过其泊松括号和跃迁概率结构所唯一确定的。之后他把量子理论刻画为这样的理论，即在其中"2 球"特征成立，也就是下述定理中的条件(iv)。另一方面，经典理论是用跃迁概率是平凡的——$p(w, v)=\delta_{vw}$——这一条件来刻画的。注意在此情形中分支是单元素。我们下面将会看到关于量子和经典系统之间区别的类似刻画(见第 2.3.2 节)。

实际上，兰兹曼[1998，pp. 104—106]粗略地证明了下述定理：[①]

> 纯态空间 S 是量子系统的纯态空间，当且仅当：(i)S 是一个泊松流形；(ii)S 是一个跃迁概率空间；(iii)S 的辛叶子对应于 S 的分支；(iv)对于任意的 v，$w \in S$，$\{v, w\}^{\perp\perp}$ 作为跃迁概率空间与 $\mathbb{C}^2$ 中的态矢量空间同构。

我已经讨论了条件(i)和(ii)。条件(iii)是这样的要求，即不能够因连续演化发生的事件也不能够因随机演化发生(反之亦成立)。换句话说，若从态 v 得到态 w 是“动力学上不可能的”，则从 v 到 w 的随机跃迁概率为0(反之亦然)。

当然，条件(iv)在我们得到希尔伯特空间形式体系过程中起到了非常重要的作用，因为它本质上要求一组态的所有“叠加”的集合都形成一个跃迁概率空间，后者看起来像是$\mathbb{C}^2$空间上的量子力学纯态，这在之前(第1.3.3.2小节)的自旋背景中讨论过。这种对量子理论或多或少的依赖性是否或在多大程度上最终令人满意是因人而异的，但至少值得注意的是，关于希尔伯特空间形式体系中什么是“本质上量子的”可以(或多或少地)约化为这一假设。

2.3 哈迪公理

最后，我们再一次在试图说明量子形式体系来源的惯例下考虑哈迪[2001；2002]得到的结果。这一方法也开始于态空间的概念，但它的框架与之前的两种方法有很大的不同，故而我们对它进行单独考察。

2.3.1 框架

2.3.1.1 作为概率预示的态 在前一节中已经提到的许多内容中，哈迪把物理理论中的态看作是与在系统给定制备下进行的测量的每个可能结果相关的概率决定因素(因而态与制备相关)。因而，对于给定的制备，对相关态的了解允许我们预言任意测量结果的概率。

---

① 这里并未详细给出此定理的条件。此外，定理的论证要求一些其他的技术性假设，这些假设直接的物理意义或许并不明确。在这里我不考虑它们。

2.3.1.2　自由度　态的一个数学特征是作为所有这些概率的“清单”(当然，一般在“清单”中至少存在不可数的许多项)。但是，一般而言态空间具有某种结构，它允许用哈迪所谓的“置信(fiduciary)”概率的短清单来对态做出某种刻画。在给定的理论中，我们把自由度定义为足以确定态的最小置信概率的数。

2.3.1.3　维度　此外，在单个测量中存在能够以概率1彼此区分的态的集合。换言之，对于集合中的每一组态 v 和 w，若 v 赋予测量某个结果以非零的概率，则 w 赋予该相同结果以概率0。一般而言，能够以这种方式区分的态有最大的数目 N，哈迪称 N 为空间的维度。

2.3.2　公理

哈迪提出了五个“公理”。第一个公理仅仅保证了我们之前关于态能够与制备相关联的假设，也就是说对于给定类型的制备，测量的结果拥有稳定的相对频率。哈迪将其他的公理改写如下(2001)。

子空间　对任意的整数 N，存在具有维度 N 的系统。此外，所有的 N 维系统与具有更高维、但它们的态将其限定在“N 维子空间”的所有系统拥有相同的特性。

复合系统　A 和 B 的自由度和维度分别为 K_A，K_B 和 N_A，N_B，则由二者组成的复合系统的自由度为 $K=K_AK_B$，维度为 $N=N_AN_B$。

连续性　对任意的维度 N、任意的 N 维系统，和该系统任意的两个纯态 v 和 w，存在一个从 v 到 w 的连续可逆变换(连续地与单位元相关联)。

简单性　对于给定的 N，K 取与其他公理一致的最小值。

这些公理中的部分其动机是相对清晰的，但其他的却并不清晰。我们这里并不讨论所有这些公理。“简单性”起作用是因为其他公理表明对有些整数 m 有 $K=N^m$。因为 $m=1$ 时连续性公理被违背，人们得到了经典的概率理论。$m=2$时人们得到了量子理论。

因此，连续性条件无疑是重要的。哈迪所倡导的对它的一种理解是它表达了想得到的东西，即态的“小的改变”势必造成基于该态的预言的“小的改变”。但是，该原则是否在物理上有说服力却并不清楚。毕竟，在哈迪看来经典物理学的态空间是不连续的(参见第2.2.2.3节的开头)，且一般人们并不认为在态

的改变和基于该态的预言的改变之间的关系存在某种严重的问题。

关于连续性的另一种不同理解将它和叠加联系了起来。其基本要点在几何上最容易想象，这里我们点到即止。量子态空间是“连续的”(在哈迪的意义上)，因为对于任意两个纯态，在它们“之间”存在另一个纯态，且事实上这个“中间”态是两个原始态的叠加。换言之，因为叠加原理成立所以连续性严格成立。连续性在经典理论中不成立是因为叠加原理在那里不成立。由此看来，在哈迪的框架中，连续性在经典理论和量子理论间有区别是不足为奇的——尽管该区别可能是不太重要的。

3. 经验内容

本节我们将讨论量子理论形式体系如何获得经验内容的问题。该讨论将自然地导致对不确定性的讨论(第4节)，因为我们这里理解形式体系获得经验内容的特定方式会自然地引向不确定原理(第4节)，正如它所应该的那样。因此本节内容并不只是说明性的。毫无疑问，我们将基于它自然地导致不确定性这一事实，提出一种有利于以特定方式理解形式体系获得其经验内容的论证。

关于经验内容的其中一个问题是如何在理论中建立测量(或更一般的经验观察)的模型。没有这样一个模型的话，很难看到理论是如何给出对经验观察的预言的。我们的第一个任务(3.1节)是对此问题进行讨论。但更难的、并且可以说在哲学上更有意义的问题是形式体系中的元素如何与经验事实完全联系起来。本节中的余下部分(并且是大部分)将关注这一问题。①

在建立形式体系及其经验内容之间联系的现行方法中存在两个要素：对称性和参考系。这些要素是相关的——“合理的”参考系通过特定的对称变换与另一个参考系相联系——但大部分的时候我仍分开来考虑它们。在本节中，我首先用POVMs来构造经验内容问题(3.2节)，之后讨论对称性的作用(3.3节)和物理量定义中的参考系(3.4节)，最后我简要地对形式体系如何获得其经验内容给出说明(3.5节)。

① 更多的材料是根据[Dickson(迪克森)，2004a；2004b]改编和修订的。

3.1 测量

3.1.1 测量的标准说明

在量子理论中，人们常常见到如下对测量的说明：假定系统的态为 W(一个密度算符)，假定人们测量用 POVM E: $\mathcal{B}(S)\to\mathcal{B}(\mathcal{H})^+$表示可观测量，则测量结果将是某个 $\Delta\subset S$，其概率为 $\mathrm{Tr}[E(\Delta)W]$。此外，测量后的态是初始态向结果的“投影”。在 POVM 是一个 PVM 的情形下，投影由 $E(\Delta)WE(\Delta)$ 给出(忽略归一化)。否则的话，通常的过程是采用“投影”(这里可能用词不当)是(再次忽略归一化) $M(\Delta)WM^*(\Delta)$ 的术语来定义“测量”算符 $M(\Delta)=\sqrt{E(\Delta)}$。注意，对于 PVMs 而言后一种规定与前一种是等价的。

关于测量的这一说明虽然对于做出预测有用，但从一种基础性的视角来看，它完全不能令人满意。正如物理学家和哲学家们都反复指出的，测量本身是一个物理过程，特别它是在两个(或更多)物理系统(一个我们称其为“仪器”，另一个我们称之为“被测系统”)间的一种物理相互作用。因而，如果量子理论是我们关于物理系统间相互作用的最好理论，它应当能够以这种方式来描述相互作用以使得结果——相互作用结束时两个系统的态——如同上面所描述的那样。

3.1.2 脉冲测量

沿着那样的线索确实有一种对测量的说明。[①] 它就是所谓的测量的“脉冲模型”。假定我们将要测量一个用自伴算符 F 表示的可观测量。出于简单性考虑，假定被测系统最初处于纯态 $|\psi\rangle$，仪器处于纯态 $|\chi\rangle$。复合系统因而处于态 $|\Psi\rangle=|\chi\rangle|\psi\rangle$。假定除了二者间的相互作用之外，这两个系统中的每一个都自由地演化，(自由)哈密顿量为 H_S 和 H_M。复合系统的总哈密顿量为 $H_{全部}=H_S+H_M+H_I$，其中 H_I 是相互作用哈密顿量(即它表示了系统间交换的能量)。最后，令 Π 为仪器的动量可观测量。

现在对于我们的测量模型而言，我们取 $H_I=g(t)\Pi\otimes F$(此后$\otimes$将不明示)，

① 对于更多的经典讨论——这里或多或少是按照此讨论进行的——见玻姆(Bohm)[1951, Chapter 22]。

其中$g(t)$是由$g(t)=\gamma f(t)$给出的相互作用函数，γ是“相互作用强度”常量，$f(t)$是一个除了t处于0(测量相互作用开始时)和τ(测量结束时)之间以外都为零的函数，且$f(t)$是(纯粹出于简单性)归一化的，即$\int_0^\tau f(t)dt=1$。因此薛定谔方程(1.5.2.3.a)可以写作：

$$\frac{d}{dt}|\Psi\rangle=-i(H_S+H_M+\gamma f(t)\Pi F)|\chi(t)\rangle|\psi(t)\rangle \tag{66}$$

其中我们明确表示出被测系统和测量装置态的时间依赖性。

对于一个“脉冲”测量而言，τ非常小，且γ非常大(相互作用快而强)。如果我们可以假定仪器和系统有低(或为零)的动量，则在时间间隔$[0,\tau]$中相互作用哈密顿量对复合系统演化的影响完全淹没了自由哈密顿量的影响，使得在此期间有：

$$\frac{d}{dt}|\Psi\rangle\simeq-i\gamma f(t)|\chi(t)\rangle|\psi(t)\rangle \tag{67}$$

我们能够轻而易举地求解此方程①以得到在相互作用结束时(结束的瞬间，在自由哈密顿量再次起作用之前)的复合态：

$$|\Psi(\tau)\rangle\simeq\exp\left[-i\int_0^\tau\gamma f(t)\Pi F dt\right]|\chi(0)\rangle|\psi(0)\rangle \tag{68}$$

将$|\psi(0)\rangle$用F(出于简单性，我们假定它是极大的)的(归一化)本征态$|f_n\rangle$来表示：$|\psi(0)\rangle=\sum_n\langle f_n|\psi(0)\rangle|f_n\rangle$(见7.1.4节最后一段)，则式(68)变为：

$$|\Psi(\tau)\rangle\simeq\sum_n\langle f_n|\psi(0)\rangle\exp(-i\gamma f_n\Pi)|\chi(0)\rangle|\psi(0)\rangle \tag{69}$$

其中f_n是F相应于本征矢量$|f_n\rangle$的本征值。现在定义：

$$|\xi_n\rangle=\exp(-i\gamma f_n\Pi)|\chi(0)\rangle \tag{70}$$

因为γ大，这些态是有效正交归一的：②

$$\langle\xi_n|\xi_m\rangle=\langle\chi(0)|\exp[-i\gamma(f_n-f_m)P|\chi(0)\rangle\simeq\delta_{nm} \tag{71}$$

---

① 方程(68)的解由此产生仅是因为相互作用的哈密顿量是用一个时间的标量函数给出的。一般的含时哈密顿量不能够以这种方式来处理。如见科恩—坦诺奇(Cohen-Tannoudji)[1977, pp. 172—175]。

② 在仪器有一个P的连续谱假设下，式(71)中的第二个等式根据黎曼—勒贝格(Riemann-Lebesque)引理成立。

正如我们将在下面更普遍的情况(78)中会看到的，因为Π是指针的“动量”，态 $|\xi_n\rangle$ 是幅度为 γf_n(因而对于大的 γ 而言它们是宏观上可区分的态)的“空间平移”(即在“指针位置可观测量”值中的平移)。这一讨论得到的结果是复合系统的终态为：

$$|\Psi(\tau)\rangle \simeq \sum_n \langle f_n|\psi(0)\rangle |f_n\rangle |\xi_n\rangle \tag{72}$$

注意该态是纠缠的，且事实上它表征了在被测系统的 F 值与指针位置之间的全关联(1.2.6.4 节)。

为了使这一模型与前一小节中的规定相匹配，我们必须采取当仪器指示结果 $|\xi_n\rangle$ 时系统的态“塌缩”到 $|f_n\rangle$ 上这一“规则”的规定。(或者是，复合系统的态投影向 $\mathbb{I}\otimes|\xi_n\rangle\langle\xi_n|$——由于是全关联的——它会使得复合态转变为积态 $|f_n\rangle|\xi_n\rangle$。)我将在第 5.4.3 节中讨论这一“塌缩”规则。

这一模型是否适用于由真实物理装置做出的所有的或确切地说是任意的真实“测量”？在合理的准确度上而言，大概是的。但这里的关键不是做出这样一个陈述，当然也不是罗列出不同种类的物理测量和它们按量子力学被塑造的模型。创建与评价这样的模型是物理学的事情。这里的关键在于量子理论能够在某种程度上提供一个测量的模型。它不需要依赖于前一小节的非说明性内容。

关于此模型的最后一个评论是：它模拟了这样一类测量，其中系统的态经过与测量装置的相互作用后在某种意义上保持不变，即若测量被重复，第二次(随后)得到相同结果的概率为 1。注意，与物理学家们或哲学家们有时所宣称的相反，至少在此意义上测量没有必然“干扰”被测系统的态。按照泡利[1958]的说法，这样的测量常常被称为“第一类测量”。

当然，测量有时候会干扰被测系统。事实上，有时候它会破坏被测系统。系统的态在测量过程中被干扰(即测量要么是不可重复的，要么其结果在重复时不一定相同)的测量常常被称为(也来自于泡利)“第二类测量”。

3.1.3 弱测量

一旦我们开始把测量塑造为真实的物理过程，那么询问在物理条件不同时会发生什么就很自然了。要考虑的一种自然情形是测量“绝热的”情形，也就是相互作用弱而且经历很长时间(在某个适当的尺度上)的情形。

实现这一理念的方案被称为“保护性测量”[Aharonov *et al.* (阿哈罗诺夫

等)，1993]。该理念在 $|\psi(0)\rangle$ 是谐振子(对于我们而言，其关键特征是在可能态之间存在有限的能量差)基态的情形中得到了很好的阐释。当然在 τ 大且 γ 小时，我们不能够忽略由自由哈密顿量引起的演化。令相互作用哈密顿量为 $H_{\mathrm{I}}=g(t)QF$。在此情形下方程(66)的解为：

$$|\Psi(\tau)\rangle = \exp[-i(H_S\tau + H_M\tau + \int_0^{\tau}\gamma f(t')QFdt')]|\chi(0)\rangle|\psi(0)\rangle \tag{73}$$

$$= \sum_n \exp[-i(H_S\tau + H_M\tau + \gamma QF)]\langle f_n|\psi(0)\rangle|\chi(0)\rangle|\psi(0)\rangle \tag{74}$$

现在，因为在被测系统基态和任意激发态之间存在有限的能量差，人们必须在基态上加上有限数量的能量以改变它。根据量子力学的绝热定理，见希夫(Schiff)[1968，pp. 289—291]，如果加到系统上的能量足够小且分散在足够长的时间段内，则它不是累加性的(即增加的总能量并不会越来越大)，而是绝热可忽略的。事实上在适当的小 γ 与大 τ 下，对于除了基态之外的那些态而言，概率幅可以任意的小。换言之，$\exp[-i\gamma QF]$项对 $|\psi(0)\rangle$ 没有净影响，我们只需要考虑它对 $|x\rangle$ 的影响(注意，若我们假定 $|x\rangle$ 的能量谱是连续的或从效果上看是连续的，则对 $|x\rangle$ 存在一个可能的影响)。因此式(74)变为：

$$|\Psi(\tau)\rangle \simeq \sum_n |\phi_n(\tau)\rangle|\psi(\tau)\rangle := |\chi_n(\tau)\rangle|\psi(\tau)\rangle \tag{75}$$

其中我们定义了 $|\phi_n(\tau)\rangle = \exp[-i(\gamma f_n Q + H_M\tau)]|\chi(0)\rangle$

注意态(75)是一个积态——相互作用仅影响仪器，并不会使被测系统和仪器纠缠起来。回顾式(59)可以明白仪器态的这一改变是如何用于获取系统信息的。根据式(75)在态 $|\Psi(\tau)\rangle$ 下取 $\mathbb{I}\otimes\Pi$(其中 Π 是与 Q 共轭的动量)的期望值，我们发现：

$$\frac{d}{dt}\langle\chi(\tau)|\mathbb{I}\otimes\Pi|\chi(\tau)\rangle = -g(t)\langle\psi(\tau)|F\otimes\mathbb{I}|\psi(\tau)\rangle \tag{76}$$

换言之，仪器动量的期望值是被测系统 F 期望值的“指示器”。例如，若系统处于定态，即 $|\psi(\tau)\rangle = |\psi(0)\rangle$，则我们可以对 F 做许多的“保护性”测量，测量相互作用之后仪器动量的平均值，进而获得关于被测系统 F 期望值的信息。

关于此方案有两个重要的评论。首先，正如许多人指出的，为了使此方案奏效人们必须提前知道被测系统的态。(特别地，在此情形下人们必须知道它是谐振子的基态。)否则的话， 我们将不知道它是“被保护的”(即相互作用并不

改变态)。所以在某种重要的意义上，保护性测量并不产生关于系统的任何新信息。此外，有人论证[Uffink(尤菲克)，1999a]说只有与被测系统哈密顿量对易的可观测量 F 才能够以这种方式测量。

然而，尽管有这些限制，人们还是应该质疑在这样一个相互作用下究竟会发生什么。在脉冲测量(无论如何，属于第一类测量)情形中，如果我们测量系统的 F，而该系统的 F 值刚刚已被测量过，我们将不会得到关于被测系统 F 值的任何新信息。虽然如此，我们可能会这样来解释第二次测量的结果：测量装置与系统的相互作用使得它的指示态与系统的态关联了起来，特别是与系统的 F 值关联了起来。也就是，第二次相互作用仍然是物理上的测量，虽然这种测量确实不会带给我们新的信息。人们可能会去类似地谈论保护性测量。当然，它们并不告诉我们那些我们尚且不知道的信息。但是我们要如何理解在此相互作用过程中会发生什么呢？一种似真的理解是仪器的态在改变，因为它在某种程度上对被测系统 F 的期望值敏感(当然，尽管我们已经知道期望值是什么了)。事实上，还有其他什么能够解释仪器态的改变呢？

第二，保护性测量是一类更普遍的测量，即所谓"弱测量"的明确模型。①一般的方案(某种程度上与标准测量的一般方案并行，其中脉冲测量是标准测量的一种模型)叙述如下。考虑一个量子系统，已知其在0时刻处于态 $|\psi_1\rangle$，在 t 时刻处于态 $|\psi_2\rangle$ 中。通常，这一知识是通过通常称为"前选择与后选择(pre-selection 和 post-selection)"的过程获得的。也就是，态 $|\psi_1\rangle$ 是通过对系统的某个系综作标准(即脉冲)测量并且只选择那些第一次测量结果为 $|\psi_1\rangle$ 的系统而被"前选择的"(在对系统的 F"弱测量"之前)。(特别地，人们可以测量相应于 $|\psi_1\rangle\langle\psi_1|$ 的可观测量，之后仅选择那些结果为1的系统。)在"弱测量"之后，再次对系综作一次标准测量，仅选择那些结果相应于 $|\psi_2\rangle$ 的系统。最后得到的系统的系综被称为是"前选择的与后选择的"。对于用算符 F 表示的任意给定的可观测量，定义"在前选择的与后选择的系综上 F 的弱值"为：

$$\frac{\langle\psi_2|F|\psi_1\rangle}{\langle\psi_2|\psi_1\rangle} \tag{77}$$

① 关于弱测量的早期论文是[Aharonov *et al.*，1987]。

注意在上面讨论的保护性测量中，我们假定被测系统的态随时间没有发生改变，故 F 的弱值仅是它在态 $|\psi(0)\rangle$ 中的期望值。①

当然，关于测量仍有许多内容需要讨论。稍后(第5节)我们将考虑对于测量来说或许是最为重要的哲学问题，即测量问题(它已经在我们对脉冲测量后关于态的塌缩毫不掩饰的怀疑态度中初现端倪)。然而，为了防止后面出现问题，现在我们给出如下评论：量子理论为描述测量提供了丰富的结构体系。

3.2 就 POVMs 而言的经验内容问题

描述测量是且仅是形式体系如何与经验观察相联系的整个过程中的一部分。过程的另一部分是关于形式体系与物理事实之间联系的更一般问题。例如，我们被允许用如 S_u 这样的可观测量"表征"在 u 方向上的自旋，但此"表征"的关系究竟是什么呢？在形式体系与物理事实之间的联系该如何做出，或该如何理解呢？(注意，上述对测量的说明已经预设了对这一更基本问题的一种回答。)

必须要清楚的是——这里的问题并非关于如何建造一个自旋测量装置，而是关于它在 u 方向(例如)"有"向上的自旋意味着什么，以及如何在形式体系中获取该意义。即便假定该建造过程不存在限制，仍然存在关于与"测量 S_u"相对应的实验室过程是什么的问题。下面我们将给出对此问题的部分回答。

回顾作为 POVMs 的可观测量是从(波莱尔集的)"可能值"向正定算符的映射。关于上述问题的另一种表述是通过此映射域内数学要素的经验内容进行的。(一旦问题得到回答，则与映射范围内数学要素相关的概率计算就变成了关于经验内容的计算。)事实上，将可观测量看作 POVMs 的一个优势(除了此方案的更大普遍性之外)是因为，它为在形式体系情况下其中的哪一部分在起"表

① "弱值"这一特有概念——更不用说对这些值的解释了——是备受争议的。其主要支持者(和合作者)在对早期工作的充分参考下从正反两方面进行的讨论是阿哈罗诺夫和波特罗(Aharonov and Botero)[2005]。注意阿哈罗诺夫和其他支持者们常常在本身就有争议的量子理论的"二态矢量"形式体系(它关注前选择和后选择系统，并被假定是时间反演不变的)背景下讨论弱值；然而，弱值的概念并没有不可逆转地与量子理论的这一方法绑定在一起，而是仅与前选择和后选择系综的理念绑定在一起，后者的操作意义至少足够清晰。

征”作用的讨论提供了更大的精确性。对一个 POVM E：$\mathcal{B}(S)\to\mathcal{B}(\mathcal{H})^+$而言，$\mathcal{B}(\mathcal{H})^+$的元素在做出表征，而$\mathcal{B}(S)$的元素（确实，最终是 S）在某种意义上是被表征的对象。

为什么我们要用$\mathcal{B}(\mathcal{H})^+$的元素来表征是我们早在第 2 节中就提出的问题（虽然还没有给出对它的确切回答）。但关于映射 E 的域的物理解释是什么，且我们是如何理解此域内的某些要素是用域内的另一些元素来“表征的”这一陈述呢？换言之，在我们描述的数学形式体系的元素与物理事实之间的关系是什么？最后，为什么我们选择一个映射（POVM）而非另一个来表征某个给定的物理量呢？在接下来关于对称性的小节（3.3 节）和关于参考系的小节（3.4 节）中，我们将介绍一些对于描绘关于量子理论经验内容这一过程（3.5 节）最终起作用的材料，这些材料会对本小节中提出的问题进行求解，或提供一些解决的途径。

3.3 对称性

3.3.1 对称群

关于将群与经验内容相联系起来的方式存在一种传统的解释。[①] 取任意群 G，并考虑它对集合 S 的作用。若 S 的两个元素通过 G 的一个元素相关联，则我们称它们是“等价的”。我们可以轻而易举地证明这样的 G 把 S 分离为等价类，然后在下述意义上我们可以说 G 是 S 上的一个“对称”群，即通过 G 的一个元素相联系的 S 的元素在某种重要的意义上来说是“同一的”。（根据 S 元素的某些理论，如果在 S 同一等价类中的不同元素能够拥有“关键性的”不同特征，则可以认为 G 不是此理论中一个对称群。通常而言，“关键性的”意味着什么是不明显的，但在这里讨论的背景下它可能意味着“经验的”或“可观测的”。）例如，令 S 的元素表征宇宙中所有粒子的位置（即 S 是宇宙的一个位形空间），那么 S 中一个点的空间平移的结果表现在宇宙中可以说是没有什么经验

① 这里我主要将群看作是在态的集合上的变换。人们也可以将它们看作是物理定律的形式变换。对此和相关问题的更多讨论，见布拉丁和卡斯特拉尼所撰写的本书第十三章。

上的不同(因为粒子间所有的距离和其他空间关系都不变)。

3.3.2 群与可观测量

我们将要提出的通往经验内容的方法(概要)部分地是用 S(作为一个 POVM 的定义域 E)上的变换群给出的，且要求 E 在某种意义上保留 S 在那些变换下的行为。① 为使这一观点得以展开，我们(至少)对量子理论做出如下两项要求：(1)在希尔伯特空间上对相关群的准确表示；(2)从$\mathcal{B}(S)$到$\mathcal{B}(\mathcal{H})^+$的映射在某种相关的意义上“保留”这些群的作用。

我们在伽利略群$\mathcal{G}$，更具体地说是在空间平移$\mathcal{A}$和加速$\mathcal{V}$(即忽略旋转与时间平移)的案例背景中来考虑这两点。对于任意的 $a \in \mathcal{A}$和 $b \in \mathcal{V}$，令 U_a 和 V_b 分别是对$\mathcal{A}$和$\mathcal{V}$准确表示的相应元素。在此，在给定希尔伯特空间上是否存在对$\mathcal{A}$和$\mathcal{V}$的忠实表示，即要求(1)能否满足是一个开放问题。在无穷维空间上，事实上存在分别扮演上述角色的算符 P(动量)和 Q(位置)：

$$
\begin{aligned}
U_a &= e^{-iPa} \\
V_b &= e^{-iQmb}
\end{aligned}
\tag{78}
$$

其中 m 是粒子的质量，它出现在这里是因为动量是质量乘速度。②

注意位置在$\mathcal{A}$的作用下被平移，在$\mathcal{V}$作用下保持不变。对于速度而言(当然对动量也是)，逆命题成立。

现在令 E_Q：$\mathcal{B}(\mathbb{R}^3)\to\mathcal{B}(\mathcal{H})^+$为位置的 POVM，考虑$\mathcal{G}$在$\mathbb{R}$(或$\mathbb{R}^3$)上的作用(第 7.6.7 节)，其中$\mathbb{R}$被理解为表示了粒子的位置。③ 那样的话，上述的要求

---

① 这里描述的观点极大地受到了与斯科特·塔诺纳(Scott Tanona)的讨论和塔诺纳[2002; 2006]观点的影响。(塔诺纳的观点是与众不同的，明显不同于这里给出的那种观点。)确实，很大程度上正是由于这些讨论我才开始提出关于此问题的观点。

② 人们常常说位置是“平移的生成元”，动量是“递升的生成元”。其原因最终与下述事实有关，即动量(位置)算符出现在无穷小平移(递升)的表达式中。教科书中对有限平移和递升的表达基本上是它们无穷小对应部分的积分。同样注意式(78)仅是一维空间中的情形。用算符“矢量”$\vec{P}=(P_x, P_y, P_z)$和$\vec{Q}=(Q_x, Q_y, Q_z)$取代 P 和 Q，用矢量($\mathbb{R}^3$ 中的)$\vec{a}$和$\vec{b}$取代参数 a 和 v，我们得到三维的版本。注意，这些变化并不改变第 7.6.4 节中关于连续参数化群的定义。

③ 经典物理学中，$\mathcal{G}$自然地作用于相空间，但我们当然可以考虑它对位置或速度约化空间的作用。

(2)相当于要求位置的量子理论表示对任意 $a\in\mathcal{A}$和任意 $b\in\mathcal{V}$拥有相同的特性：

$$U_aE_Q(\Delta)U_a^{-1}=E_Q(\alpha_a(\Delta))\quad(\text{协变性})\tag{79}$$

$$V_bE_Q(\Delta)V_b^{-1}=E_Q(\Delta)\quad(\text{不变性})$$

其中 α_a 是 a 在$\mathcal{B}(\mathbb{R}^3)$上的作用。相应的条件对 E_P，即动量的 POVM 也必须成立(也就是，它应该在平移下不变，在加速下协变)。注意这里我们没有假定式(79)中的 P 和 Q 是式(78)中的那些算符。事实上，式(79)中的这些条件本身就足以确定映射 E_Q 和 E_P，以及表示式(78)(取决于幺正等价性)。下面我们会回到这一点。但在这里，我们把它看作是用式(79)(和关于动量的相应条件)来定义“位置”和“动量”的动机。

对于此人们可能会持有下述两种反对意见中的一种。第一种是这些要求从何而来？为什么我们要坚持位置必须具有这一特别的不变性与协变性，甚至在经典情形中也是如此，或是根本就是如此？对此问题的回答是，至少在这里的讨论中，我们在用相对于$\mathcal{A}$的协变性和相对于$\mathcal{V}$的不变性定义我们所指的“位置”。例如，如果粒子的位置在参考系 F(下面我们将更详细地考虑参考系在此讨论中的作用)中是$\vec{x}$，且如果参考系 F 通过空间平移$\vec{a}$与 F'相关联，则粒子在参考系 F'中的位置为$\vec{x}-\vec{a}$。人们或许会认为位置的这一特征表面上的平凡性是它是“位置”的部分所指这一事实的结果。也就是说，缺乏这一特征的可观测量(POVM)事实上不是“位置”。类似的评论对于递升也成立，并且对于一方是动量另一方是平移和递升的两方之间的关系也成立。相应的评论对于其他可观测量如角动量和自旋都成立。

第二种(反对意见)是强调位置在量子理论背景——即就 POVM 而言的背景，在相关的意义下也是伽利略群(表示)作用的背景——中“指的是相同的事物”，那样的话我们能否获悉位置“非常不同于”我们原本所认为的位置这一信息？(当然，类似的评述对动量也成立，事实上对于许多其他能够通过这种方式来定义的物理量都成立。)对这一反对意见有两种回应。第一，在量子理论背景下(其中位置和动量 POVMs 的确遵从“正确的”不变性和协变性)，我们事实上的确获悉了位置“非常不同于”我们所认为的那样，且这一事实澄清了我们并没有将我们对位置的定义限制得如此狭隘以至于不可能对其现存概念进行实质性的修改。第二种回应是，我们应当将一方面是获悉一个关于现存物理概念的

新内容——在此情形下该概念必须使得在获悉关于它的新内容之前和之后它是"同一个"概念(且我们指出同一个是指它相对于伽利略群某个部分的关系而言的)——和另一方面是发现新的物理概念区分开来。而在这里关于如何做出这样的发现或如何理解它们，我们并不提出建议。①

在本小节的剩余部分，我将深入讨论上面做出的陈述：强调相关的不变性和协变性对于确定位置和动量可观测量是充分的。而在下一小节中，我将转向讨论参考系在可观测量定义中的作用。

3.3.2.1 非本原系和量子可观测量的"唯一性" 作为 POVMs 的位置和动量每一个都产生一个"非本原系"，它是协变系统中的一种特殊情形。② 一般地，一个协变系统是一个集合($\mathcal{H}$, E, S, $\mathcal{G}$, α, $\{U_g\}$)，其中$\mathcal{H}$是希尔伯特空间，E 是一个 POVM，其定义域为 S 且值域是$\mathcal{H}$上的正定算符，$\mathcal{G}$是某个群，α 是$\mathcal{G}$在 S 上的作用，$\{U_g\}$ 是$\mathcal{H}$上对$\mathcal{G}$的幺正表示。如果 E 是一个 PVM，则该集合是一个非本原系。③ 非本原系拥有重要的特征，部分总结于麦基(Mackey)的非本原性定理中。我这里的讨论将采取与[Mackey, 1996]相同的思路，将重点放在论证的结构和假设上，并不赘述在许多地方可以很容易找到的数学细节上。④

在一个非本原系中，S 通常被看作是普通的——例如，它可能是一个普通的度量空间，它伴随有某个局域紧致的且可分离的等距群$\mathcal{G}$(假定它对 S 有一连续且可递的作用——见第 7.6.7 节)。但是为了尽快切入真实的物理问题，我们会很快将其具体化。有了这样的目标，很自然地便可将 S 看作$\mathbb{R}^3$，$\mathcal{G}$看作平移与旋转的半直积(第 7.6.2 节)($\mathcal{A} \ltimes \mathcal{R}$)。然而，在准备应用非本原性定理的过程中，将$S$ 看作是平移($\mathcal{A} = \mathcal{G}/\mathcal{R}$)的拓扑群更为有用，这一平移显然与作为

---

① [Tanona, 2006]对此有特别的帮助。

② 对非本原性及其在量子理论问题中应用的进一步讨论，见本书兰兹曼撰写的第五章。

③ 协变系统可以通过奈马克(Neumark)延拓定理"延拓"到非本原系。见[Cattaneo(卡塔内奥), 1979]。本质上是考虑非本原系并不会失去普遍性。

④ 关于更多的数学细节和同一结论的替代方法，除[Mackey, 1996]之外，见布施、格拉博斯基和拉蒂(Busch, Grabowski and Lahti)[1995]和其中的参考文献。全部细节在[Varadarjan(瓦拉达拉杰), 1995]中可见。

拓扑空间(在给定$\mathcal{A}$上的适当且明显的度规之后，事实上是作为度规空间)的$\mathbb{R}^3$同构。因而，理论要点是S的元素表征了从某个固定起点开始的“位移”，因而表征了一个位置(且进而表征了一个位置可观测量的可能值)。子群$\mathcal{R}$则描述了围绕此起点的旋转。

我们现在要求位置的PVM，即E^Q相对于$\mathcal{G}=\mathcal{A}\ltimes\mathcal{R}$是协变的。(任意$g\in G$在$S=\mathcal{A}$上的作用用一种显然的方式来定义：对于$g\in\mathcal{A}$，它在$a\in\mathcal{A}$上的作用仅是$ga$；$g\in R$作为单位元作用于$\mathcal{A}$上，这样定义的作用是可递的。)结果，给定$\mathcal{G}$在某希尔伯特空间$\mathcal{H}$上的一个表示$\{U_g\}_{g\in\mathcal{G}}$，我们就有了一个非本原系。

在某种形式下将非本原性定理表述如下：

定理(麦基)：令$\{U_g\}_{g\in\mathcal{G}}$是一个可分离的、局域紧致的拓扑群$\mathcal{G}$在可分离的希尔伯特空间$\mathcal{H}$上的幺正表示，令$\mathcal{K}$为$\mathcal{G}$的任意闭子群。令$E$是一个PVM，其定义域为$\mathcal{G}/\mathcal{K}$使得($\mathcal{H}$，$\mathcal{G}/\mathcal{K}$，$\mathcal{G}$，$\alpha$，$\{U_g\}$，$E$)是一个非本原系(其中$\alpha$是$\mathcal{G}$在$\mathcal{G}/\mathcal{K}$上的可递作用)。则对于$\mathcal{K}$在某希尔伯特空间$\mathcal{H}'$上的任意表示$\{V_k\}_{k\in\mathcal{K}}$而言，由$\{V_k\}$在希尔伯特空间$L^2(\mathcal{G}/\mathcal{K})\otimes\mathcal{H}'$上引起的对$\mathcal{G}$的表示(且这一导出表示存在)是与$\{U_g\}$幺正等价的。此外，$E(\Delta)$(如果导出表示通过一个非平凡幺正变换与$\{U_g\}$相联系的话，也可以是其恰当的幺正变换)必须是$L^2(\mathcal{G}/\mathcal{K})$上的乘法算子$\chi_\Delta$与$\mathcal{H}'$上的单位元的张量积。

所以在我们当前考虑的情形中，我们将令$\mathcal{H}$为$L^2(\mathbb{R}^3)$。① 令$\{D_r\}_{r\in\mathcal{R}}$为$\mathcal{R}$在某希尔伯特空间$\mathcal{H}'$上的表示，并考虑$\mathcal{G}$的相关导出表示。这一导出表示必须幺正地等价于$\mathcal{G}$的任意表示，且此外$E^Q(\Delta)$仅仅是$\chi_\Delta\otimes\mathbb{I}'$。

现在考虑最简单的情形，其中$\{D_r\}$是平凡的恒等表示(即$\mathcal{R}$的每个元素都由一个一维希尔伯特空间上的单位元表示)。则麦基定理直接产生了通常的位

① 更精确地，我们应该用平移的拓扑群替代$\mathbb{R}^3$，但我们已经注意到了对于我们的目的而言它们是相同的空间。

置算符薛定谔表示。① 更明确地，对于 $\Delta \in \mathbb{R}^3$，有 $E(\Delta) = \chi\Delta$，其中后者是具有"用 Δ 的特征函数乘 $L^2(\mathbb{R}^3)$ 的一个元素"作用的算符（该算符是一个投影）。对此 PVM 而言，对所有 S 的积分产生了通常的位置算符：

$$Q = \int_{\mathbb{R}^3} \vec{r}\, dE^Q(\vec{r}) \tag{80}$$

它是算符 Q_x，Q_y，Q_z 的"矢量"，该矢量对 $\vec{r} = (x, y, z)$ 而言，对于任意的 $\psi(\vec{r}) \in L^2(\mathbb{R}^3, d\vec{x})$ 有作用 $Q_x\psi(\vec{r}) = x\psi(\vec{r})$，$Q_y$ 和 Q_z 也类似。注意这里所发生的情况：我们一开始时要求表示位置的 PVM 相对于平移和旋转有"正确的"协变性，即它是相关非本原系的一部分，结束时我们通过麦基定理表明由于幺正变换，我们必须选择通常的（"薛定谔"）位置表示。另一种在某种程度上对这一点更具操作性的表述是：假定涉及位置可观测量的观测（或预言）遵从位置可能值空间的对称性（平移和旋转），则所有的位置算符（PVM）表示幺正地等价于薛定谔表示。就确立理论的经验意义而言，我们或许可以说"成为系统的位置"意味着成为在伽利略群相关部分作用下以正确方式进行的变换。麦基定理确立了位置的唯一性（就幺正变换而言），从而定义了它。

事实上，麦基定理还确立了更多内容。比如它还确立了在（适当选择的 POVM）位置（和动量）表示与伽利略群表示之间的关系。为了明白其原委（概述），令$\mathcal{G}$是通过 $a \in \mathbb{R}$（7.6.4 节）连续参数化的任意群。正如我们已经讨论过的，现在如果假定$\mathcal{G}$是一个对称群，则一般而言它应该用"物理上没有任何不同的"希尔伯特空间的变换来量子力学地表示，且这样的变换通常用幺正算符给出（回顾 1.5.1.2 小节中的讨论）。此外（回顾 1.5.1.4 小节），当群通过 a 连续地参数化时（正如我们所感兴趣的情形中那样），这些幺正算符是用$\mathcal{H}$上的自伴算符 F 给出的，满足 $U_a = e^{-iF_a}$。因此，空间平移和递升的幺正表示必须采取这种形式。从而，麦基定理事实上表明了平移和递升能通过幺正等价性用式（78）给出。

重要的是要记住"通过幺正等价性"并不意味着人们可以将不同的幺正变换

---

① 结果是对一个无自旋粒子的描述——见［Mackey, 1996］。$\mathcal{R}$的非平凡表示导致关于粒子的描述是自旋为$\frac{\dim(\mathcal{H}')-1}{2}$。

应用于来自于式(78)的U_a和V_b，同时仍然满足置于位置和动量POVM上的所有(不变性与协变性)条件。这里的要点是位置和动量之间的关系是由那些条件通过麦基的非本原性定理所建立的。(另一方面，人们总能够应用一个全域的幺正变换，但这样的变换类似于经典力学中“将宇宙向右平移五步”的变换。)

事实上，人们可以从这些结果中确立这样的结论，即位置和动量必须满足对易关系的外尔(Weyl)形式。① 特别地，有：

$$U_a V_b = e^{iab} V_b U_a (a,\ b \in \mathbb{R}) \tag{81}$$

正如我们稍后将讨论的(4.1.2节)，这一表达式是P和Q之间正则对易关系式(24)的另一个版本。换言之，位置和动量与伽利略群间的恰当关系这一假定直接导致了它们的不相容性。

因而，位置和动量算符(POVMs)在伽利略群中扮演的角色、伽利略群对它们的作用和它们之间的对易关系(外尔形式)，在我们强调下述两点时就迅速被确定了下来：(1)位置和动量满足上述给出的不变性和协变性条件；(2)位置和动量，还有伽利略群本身都可以在希尔伯特空间上表示。在非相对论理论中否定(1)的可能性很小，同时(2)可以理解为对我们理论来说是真正的量子力学理论的要求。② 最后，人们应当牢记麦基定理的普遍性。我主要在位置和动量背景下讨论它，但类似的解释对于构成相对于某个对称群而言的非本原系(或协变性)的任意可观测量(POVMs)都成立。在布施等(1995年)的书中可以找到一些这样的例子(其中有角动量和自旋的例子)。我们留给读者自己去研究，现在转而考虑量子力学中参考系在物理量定义中的作用。

---

① 细节请参阅瓦拉达拉杰[1985，§V]或麦基[1949；1978]。

② 这里我们必须谨慎一些。感谢乔斯·尤菲克提出此问题。众所周知，经典力学可以表示为关于复希尔伯特空间的理论。见[Bracken(布莱肯)，2003]和其中的参考文献，显然此理念的提出者是格伦沃德(Groenewold)[1946]。然而，希尔伯特空间上的经典可观测量(不足为奇地)构成了对易代数，由线性算符“非标准”积定义，即[Bracken，2003]中的“圈点(odot)”积，此对易代数并不出现在量子理论中。(例如，动力学是通过此“圈点”积定义的李括号来定义的，而不是用由它们在空间上的作用的合成所给出的算符的普通积来定义的。)因此，课本中对此主张更谨慎的表达为，我们要求位置和动量是一般所理解的希尔伯特空间上的算符代数中的算符，即在算符(合成)的通常积下的算符。

3.4 参考系

3.4.1 参考系在量子理论中的身份与作用

在量子理论的早期就有人声称(不是不合理的)，在该理论中没有参考系概念的余地。显然困难在于：依照定义，一个参考系要有一个清晰可辨的位置和运动状态，因为位置和运动状态是相对于它来确定的。但是我们很难看到作为量子参考系的诸如此类的事物如何存在，因为正如我们已经提到过(第1.2.5节)并且在下面会略作详细讨论(第4节)的，标准量子理论不能够描述拥有一个清晰可辨的位置和运动状态(动量)的事物。玻尔(在一种对他的理解看来①)总结说，是我们在约定某物体(通常是测量仪器)作为参考系，且这一约定要求我们经典地处理该物体，因为这一约定要求该物体有明确的位置和动量。(当然，在约定了某个其他的物体来扮演参考系的角色之后，我们也能够返回去并用量子力学来描述该物体。)

但是，约定也分好与坏——正如玻尔本人强调的，并不是每个物体都能够合理地用于定义给定目的的参考系。其中一个重要的原因就是在惯性系与参考系之间存在显而易见的区别。惯性系是在其中运动定律(无论是经典的或是量子的)有效的框架，② 寻找惯性系是经验研究的问题。另一方面，参考系是用以确定如位置、动量、角动量、自旋等物理量的框架。对这些物理量的瞬时定义而言，对参考系的任意约定都是可以的。

然而，对系统随时间的描述而言，参考系的选择远不是任意的。例如，将旋转的坐标系统用作参考系将会引起虚拟的科里奥利(Coriolis)力。当然，人们指的“旋转”是“相对于一惯性系的旋转”，且这里的关键点在于：不是惯性系

---

① 见[Bohr，1935]。对此论文的一些解释性(虽然是有瑕疵的)评论，见[Dickson，2002a；2002b]。关于玻尔坚持经典概念必要性的新近详细解释可以在[Tanona，2002]和[Howard(霍华德)，2003]中找到。当然，许多其他人在此论题上也有表述。那些研究成果中的参考文献会带领感兴趣的读者入门。

② 对“是惯性的意味着什么”的这一理解有很长的一段历史，见[DiSalle(迪沙利)，1990；2002]和[Barbour(巴伯)，1998]。它在牛顿惯性定律应该被理解为存在一个参考系，即一个“惯性系”，在其中牛顿的其他两个定律为真这一陈述的理念中达到了顶点。此理念可以拓展到量子理论中。

的参考系总会引起虚拟力(也就是明显违背运动定律)。因此我们得到的结论是，虽然相对于最方便的那个参考系来描述我们的物理系统肯定是允许的，但同样必要的是我们要知道如何用一个惯性参考系来描述系统，这里我们指的惯性系是在其中运动定律——依情况而定是经典的或是量子的——为真的框架。

在此意义上量子理论中没有什么会排除惯性系的可能性。事实上正如经典物理学那样，量子物理学包含了下述假设(通常是隐含的)：存在某个在其中定律有效的框架(一些坐标系)。(当然这并不是说这样的坐标系可以用作经典意义上的参考系，即同时定义精确的位置和动量。)

如在经典力学中一样，量子力学中人们通过寻找在其中动力学定律为真的坐标系来发现惯性系。在经典力学中，这种寻找常常扩展到用天体确定的框架，见[Ma *et al.* (马等)，1998])。在量子理论中，人们并没有走这么远。典型的测量装置或容纳这些装置的实验室，都足以服务于定义参考系的目标。

3.4.2　相对坐标和绝对坐标

如果量子可观测量是相对于一个参考系(无论是惯性系与否)典型地定义的，且如果这是正确的，则通常在量子理论中观测上有意义的变量在性质上将是相关的。

这一点并不仅适用于明确涉及位置或 $L^2(\mathbb{R}^{3N})$ 空间的测量情形。我们来考虑自旋情形。例如，说系统处于态 $|\psi\rangle = |z_+\rangle$ 在观测上意味着什么？若没有告诉我们空间的哪一个方向算作 z 的话，则系统处于该态这一陈述在观察上是没有意义的。

但是，我们从上面看到，当我们希望将量子运动定律应用于系统时，我们必须在坐标系中来进行，这一坐标系是用在其中运动定律有效的某个框架给出的。我们前面也提到了在量子理论中此框架常常是通过仪器的某个宏观部分来确定的。为了服务于定义在其中运动定律有效的坐标这一目的，仪器的宏观部分必须要满足什么条件呢？我们怎么能够确认这些条件能得到满足呢？

结果是，对最后一个问题的回答与在经典力学中是相同的——我们从未得到过任何“绝对的”参考系，它是惯性的且根据它我们可以检验其他框架的惯性属性(并进而检验它们是否适用于定义在其中定律有效的坐标系)。我们最多只能在经验上尽我们所能地确定运动定律在某个特别的参考系 F 下是有效的，并

进而通过参照F来为其他参考系的使用作证明，特别是通过指明这些其他有效的参考系是通过一适当的对称变换与F相关联这一事实来证明其合理性。

换言之，至少在我们拥有一个真正的相对论理论之前，相对坐标是用绝对坐标来最终定义的，且正如我们现在要描述的那样，它们是用某个对称变换，来如此定义的。①

一开始我们回顾从一个惯性系到另一个的变换是(在当前非相对论的背景下)由伽利略变换给出的。(此外，我们看到对于这些变换在量子理论中的数学形式，我们别无选择。)我们能够用这个事实来得到对下述变换的表述，即从某个给定惯性系的绝对坐标到相对于某个约定的参考系测量得到的坐标的变换。

按照下述方式来思考该情形是有益的。想象实验室中的观测者A，假定A相对于实验室来测量物理量。现在想象一个“外部的”观测者B，他被告知某框架F是惯性的这一信息(或假定该信息)。B同样把F作为参考系：只考虑B的话，实验室和其中包含的东西都是用F给出的坐标描述的。但假定B想要把A的测量描述为是相关的且处于由实验室给出的框架中。B该如何从由F给出的坐标变换到由实验室(它可能也在相对于F运动)给出的(相对)坐标呢?

这一问题的答案几乎直接来自于伽利略变换规则。(当然，在这个例子中我们并不是简单地应用伽利略变换，我们同时也转换到了关系坐标。)正如阿哈罗诺夫和康费尔(Aharonov and Kaufherr)[1984]指出的，正确的变换为：

$$U_{\mathrm{AK}}=e^{-i\Sigma_{n>0}P_n^BQ_0^B} \tag{82}$$

其中P_n^B表示B用以描述系统n的动量可观测量；类似地，Q_n^B表示B的位置可观测量，系统0是实验室本身。注意它们的结果暗中假定了参考客体——“实验室”——在F中惯性地运动。也要注意在此情形中，实验室不在观测范围内。A没有描述实验室(系统0)的坐标，因为A的坐标都是相对于实验室定义的。

快速地作一下审查，注意到有：

$$U_{\mathrm{AK}}U_n^AU_{\mathrm{AK}}^{-1}=Q_n^B-Q_0^B(n>0) \tag{83}$$

从而正如人们所期望的，A仅描述Q_n^A，B“知道的是”$Q_n^B-Q_0^B$。也就是说B可以用纯量子理论的术语来描述A的位置测量是相对于实验室做出的这一事实。

① 这里的讨论主要受到了阿哈罗诺夫和康费尔[1984]工作的启发。

对我们而言这一讨论的主要受益是：一般而言我们处于观测者 A 的角度，而不是 B 的角度。我们并不拥有一个绝对参考系，相反，我们的量是被相对测量(从而被定义，至少在操作上是这样)的。因此，将我们从一个参考系转换到另一个参考系的对称性最终参与了那些量的定义，因为为了使我们的物理理论能够有效地应用于它们，我们必须将它们看作是用“绝对”惯性系坐标来“真正”定义的，且这样的定义涉及了将我们从一个(惯性)系转换到另一个惯性系的对称变换。

这里还有一个有趣的附带问题，叙述如下。观测者 B(原则上，他的观点在这里是真正有效的)有一种特别有意思的观测位置与动量不相容性的方式。假定 A 要测量粒子的位置，且按照脉冲测量(第 3.1.2 节)模型的思路写下一个相互作用哈密顿量。当然，A 只是写下一些类似于 $H_I = g(t)\Pi Q$ 的东西，其中 Q 是 A 拥有的被测系统的位置算符。当 B 将此哈密顿量变换为正确的关系坐标，且求解方程时，B 发现测量的一个结果是实验室本身经历了一个动量的变动，这使得实验室对于定义动量而言变得不再合适(对于 A 而言，因为 A 没有办法测量实验室的动量变动——对于 A 而言实验室是上述测量得以进行的相对参考系)。①

3.5 经验内容的群理论刻画

3.5.1 问题的重构

第 3.3 和第 3.4 节的讨论如何能够帮助我们建立起对量子形式体系经验内容的说明呢？在第 3.2 节提出的所有问题中，我们集中关注如下两个：(1)什么样的经验内容是与 POVM 定义域中的元素相关的？(2)给定对(1)的回答，该如何恰当地选择 POVM 本身？

上面的讨论提出了如下的一般思路：(a)可观测量是参考系依赖的量，这些量是用它们相对于某个对称群的行为来定义的；(b)在指出有效的参考系应

① 细节见[Dickson，2004b]。那里没有充分强调这些观测充其量是理解位置与动量不相容性的第一步。此外，注意在 B 的计算中不包含不确定关系，后者被理解为给位置一动量可以同时测量(或获知或定义)的精度设置一个下限。

该以恰当的方式关联于惯性系——更一般地，它们应该适用于定义我们确实希望定义的那些量——之后，我们应该试图赋予这些事物以某些经验的和观测的意义；(c)类似地，在指出对称变换在物理量定义中的作用之后，我们应该试图赋予这些事物以经验的和观测的意义。(a)的更详细版本将构成对上述问题(2)的回答。(b)和(c)的更详细版本将构成对上述问题(1)的回答。

事实上关于(a)我们已经说过一些了，且相当多的内容出现在关于量子力学的可观测量所遵循的(并且是对可观测量做出定义的)对称性的文献中。不管怎样，这里基本的要点是，(参考系依赖的)可观测量可以通过它们相对于某个适当的对称群所遵循的不变性和协变性来唯一地刻画(取决于平凡变换，如长度标度和全域幺正变换)。在此意义上，"位置"和"动量"等的真正意义就部分地由这些不变性和协变性给出。本小节结束时我将给出关于(b)和(c)的一些初步思考，之后是对一种反对意见的考察。

3.5.2 参考系与变换的经验内容

关于(b)这里给出的基本提法是，参考系把世界描述为某个观测者所见证的那样，且最终是人类所见证的那样。这一提法反映出这样一种观点，依据该观点理论最终是人类的建构。这一观点并不(必然地)意味着物理理论中存在(或有许多)任意性——世界可能仍然支配着像我们这样的观测生物建构理论的方式，若我们能够建构理论的话。它要求理论的"可观测量"与人类的观测能力紧密关联，进而与这些能力的特性紧密关联。

然而，至少从两个相关原因来看，认同这一观点的那些人必须很谨慎。第一，参考系是在不同的方式下被典型地理想化的，这些方式可能不适用于真实的人为观测。例如，它们通常被看作是(空间上)完全不动的。第二，如果给定的参考系在任意时间间隔上使用，则为了有效地(与成功地)应用运动定律(无论是牛顿第二定律、薛定谔方程或是其他什么)，要么它必须是惯性的，要么人们必须知道如何将它与一个惯性系关联起来。但正如先前注意到的，在真实的实践中确定一个给定参考系是否是惯性系非常困难。因此，这里要考虑的建议是参考系最终应当被理解为"有效的(和理想化的)人类视角"，其中有效性概念是用与惯性系的已知关联来阐明的。

关于对称变换本身的经验内容(c)，我再次给出一种人类中心主义的观点。

来自于19世纪多名几何哲学家，如亥姆霍兹(Helmholtz)和以不同方式的庞加莱(Poincaré)，他们的提议是，这样的变换与下述经验相关联，即通过经历从一个参考系角度的观测改变到另一个参考系角度观测时的生理—肌肉感受(physiological-kinesthetic experience)而得到的经验。(例如，考虑与旋转或空间平移相关的生理—肌肉感受。)但是，特殊的变换群与不同种类经验内容间的联系远远超过了本章的范围。我们只是注意到这里勾勒了理论的一个必要因素是为处于量子理论核心的群确立(或认识到)这些关联。①

3.5.3 “绝对”量

那些不是参考系依赖的量又会怎么样呢？当然相对论给出的经验是，虽然许多被测量的量是依赖参考系的，但有一些物理量——或许甚至是最重要的那些物理量——是“绝对的”，即与参考系无关的。例如，考虑定义为 $\tau^2 := t^2 - (x^2 + y^2 + z^2)$ 的时空间隔。在无论人们用什么样的坐标去计算 τ 结果总是相同的这一意义上，它是参考系无关的。事实上，一种对非绝对量的理解认为它们只不过是“从一个特殊视角”来看的绝对量，从而绝对量在某种方式上是基础量，参考系相关的量是导出量。

但是人们要如何测量或观测一个绝对量的值呢？这里做出的断言是，我们不得不在某个参考系下来做。虽然结果并不依赖于参考系，但测量仍然是在参考系中进行的(例如，考虑人们如何测量)。若此断言是正确的，则对于某个量是绝对的这一观点存在两种回应。

较为温和的回应是，允许绝对量在某种意义上甚至是更基础的量。但是，我们这里关心的是量子理论的观测内容——数学形式体系如何与实验观测相联系？如上所述，若观测总是发生在一个参考系内，则一种只考虑参考系依赖量的经验内容的解释是比较公允的。

较为极端的回应是，主张一种与上面提到的态度相反的观点：基本的量是参考系依赖量，“绝对”量是从它们中推导(计算)出的。这一观点事实上与上

① 人们可以采用一种与这里所提议的完全相反的方法，即在对相关变换群的经验说明之前将理论看作某种已具有经验意义的，且之后用那些变换在理论中的影响来确定这些变换的经验内容。这里我不全面阐述该理念。

面表达的态度——(我们的)物理理论基本上是关于我们所观测到的世界的——紧密相关。在此观点下，绝对量的作用是保证不同参考系中的居住者就他们参考系依赖量的值有交流的可能性，且使得可能达成某种一致，而这就是全部。

4. 不确定性

本节致力于审视量子理论中的不确定性。我们将开始于(第4.1节)不确定关系的形式起源，即正则对易关系。之后我们将考虑不确定关系(第4.2节)，它们最低限度地表达了下述事实：两个非对易可观测量(我们将会定义这一概念)的弥散不能够同时任意小。我们接着(第4.3节)将考虑两种根本上不同的理解或解释不确定关系的方式，最后(第4.4节)我们将详细考虑爱因斯坦、波多尔斯基和罗森的著名论证[1935]，他们试图使人们对量子不确定性的基本性产生某种形式上的怀疑，并最终怀疑作为一种对物理实体描述的量子理论的“完备性”。

4.1 正则对易关系

4.1.1 对易关系的表示

先前(1.2.5节)我们注意到量子力学的位置(Q)和动量(P)算符满足正则对易关系(24)。或许更准确地说是，在量子理论中选择Q和P以使得(24)成立。在希尔伯特空间$\mathcal{H}$上选择算符Q和P从而使(24)满足，是选择正则对易关系的一个表示。可以认为正是某个可观测量对——主要是位置和动量——对这些对易关系的满足使得理论是真正“量子的”。

其结果是对于任意的表示，这些算符不能够都有界，且因而$\mathcal{H}$必须是无穷维的(第7.2.2节)。海森堡用无穷维方矩阵在空间l^2(第1.2.1节)上构建了一个表示：

$$Q=\frac{1}{\sqrt{2}}\begin{pmatrix}0 & 1 & 0 & 0 & \cdots\\ 1 & 0 & \sqrt{2} & 0 & \cdots\\ 0 & \sqrt{2} & 0 & \sqrt{3} & \cdots\\ 0 & 0 & \sqrt{3} & 0 & \cdots\\ \vdots & \vdots & \vdots & \vdots & \ddots\end{pmatrix},\ P=\frac{-i}{\sqrt{2}}\begin{pmatrix}0 & 1 & 0 & 0 & \cdots\\ -1 & 0 & \sqrt{2} & 0 & \cdots\\ 0 & -\sqrt{2} & 0 & \sqrt{3} & \cdots\\ 0 & 0 & -\sqrt{3} & 0 & \cdots\\ \vdots & \vdots & \vdots & \vdots & \ddots\end{pmatrix} \quad (84)$$

薛定谔用空间 $L^2(\mathbb{R})$ 上的算符也构建了一个表示，其中 Q 是乘法算符，即对任意的 $f\in L^2(\mathbb{R})$ 有 $Qf(x)=xf(x)$，且 $P=-i\frac{d}{dx}$。

这两个表示事实上是幺正等价的。即存在从 l^2 到 $L^2(\mathbb{R})$ 的同构，在此同构下海森堡算符转到了薛定谔算符。这些表示同构的任意表示被称为正规的(regular)，也存在非正规表示。①

注意因为正则对易关系表示中的算符(至少其中之一)必须是无界的，我们必须谨慎注意它们的定义域(第 7.2.2 节)。因而正则对易关系仅定义在空间的某个(致密)子集上。

4.1.2 外尔关系

P 和 Q 必须是无界的这一事实偶尔会引起麻烦。例如，我们只是注意到这一事实要求人们注意它们的定义域。外尔给出的一种替代方法避免了这一问题。我们考虑算符 U_a 和 V_b(其中 a，$b\in\mathbb{R}$)(第 7.6.4 节)的一对强连续单参数的幺正群。如果它们满足第 3.3.2.1 节中的关系(81)，则我们称它们为一个外尔对(Weyl pair)。根据斯通定理(第 1.5.1.4 节)，U_a 和 V_b 可以写作：

$$U_a=e^{-iaQ},\ V_b=e^{-ibP} \quad (85)$$

其中 Q 与 P 是定义在一个公用(致密)定义域上的无界自伴算符，回顾式(78)，用级数展开的幂从形式上表示这些指数有：

$$e^{-iaQ}=\sum_{n=0}^{\infty}\frac{(-iaQ)^n}{n!} \quad (86)$$

① 这里给出一个简单的例子。考虑在开区间(0，1)上平方可积函数的空间。再次令 Q 和 P 为算符的乘和微商(正如上文给出的)。在此情形中，Q 事实上是有界的(但 P 不是)，因此表示不能够与薛定谔表示幺正地等价，因为在那里两个算符都是无界的。

(e^{-ibP}相类似)，然后代入式(85)，从而我们重新得到正则对易关系。如果式(85)的两边都定义在一个公用的致密子空间上，则这一过程是严格有效的；否则的话，它是提示性的符号操作。最后，注意到算符 e^{iaQ}和 e^{ibP}是有界的，因而不会产生定义域的问题。

4.1.3 冯·诺伊曼的唯一性定理

外尔关系的另一个优良特征是：它们的所有表示都是正规的。换言之，每个外尔对都与由薛定谔位置和动量算符产生的外尔对幺正地等价。冯·诺伊曼[1931]得到的这一结果表明了任意外尔对的生成元必定具有完整的实线谱。

一个深入的问题是关于什么时候正则对易关系的表示是正规的——即除了直接检验幺正等价性之外(这会很难)，人们为何可以说一个给定的表示是正规的？此问题的答案(超出了只是对谱的观测)已经得到了，见[Rellich(瑞利希)，1946]和[Dixmier(狄克斯梅尔)，1958]，但超出了本章的范围。讨论和更多的参考文献见[Summers(萨默斯)，2001]。

4.2 不确定关系

正则对易关系如此重要的一个原因是它们直接产生了不确定关系。回想如果两个有界算符不对易，则存在着一个算符的本征矢量并不是另一个的本征矢量。(类似的评述对无界算符也成立，但在那里我们必须考虑它们可能不会有本征矢量这一事实，因而我们代之以讨论它们谱投影的非对易性。)由此得出结论：存在赋予一个可观测量的可能值以平凡概率(即一个本征值的概率为1，其他为0)且赋予另一可观测量至少两个可能值以非平凡概率(非0或非1)的态。因此，非对易性已经表明了一类“不确定关系”：关于一个可观测量值的确定性意味着关于另一个可观测量值的不确定性。下面，我们将更精确地表述这一理念并考虑对它的解释。

4.2.1 光学推导

海森堡在[1927]中和[1930]中的改良版本里做出了下述论述，试图弄明白或许甚至是想导出位置和动量的不确定关系。假定我们希望通过一个光学显微镜测量一个微小粒子(如一个电子)的位置。孔径角为 θ 的显微镜的分辨能力近似为 $\lambda/\sin\theta$，其中 λ 是光的波长。这一分辨能力确定了我们关于粒子位置在测

量后的不确定度。另一方面，为了使我们能检测到粒子至少需要一个光子撞击它。这一个光子的动量①为 h/λ，且作用角度在角度 θ 范围内是不确定的，因此，传递到粒子上的动量的不确定范围约为 $(h\sin\theta)/\lambda$，且在测量后被测粒子位置和动量的不确定度的积约为 h。

换言之，位置和动量不确定度的积存在下限。注意这一下限只适用于测量之后的情形。事实上，我们可以在以任意精度测量粒子位置之前以任意精度测量其动量，然后我们可以在(光子对粒子的)动量产生影响之前以任意精度(虽然我们仍然不确定它在测量后的动量)确定其位置和动量。

也有对不确定关系的其他推导，它们更明确地依赖于量子理论的形式体系。现在我们考虑其中两种。

4.2.2　波函数推导

位置和动量是通过傅里叶变换相联系的。确实，在薛定谔表象中处理动量时应用傅里叶变换(它是希尔伯特空间上的幺正变换)通常更容易，因而动量算符变为乘法算符(位置算符则变为微分)。但想想波函数发生了什么，一个处于峰值的波函数对应于一个位置明确的态。也就是说，大部分的概率集中在相对小的空间区域。但这样一个函数的傅里叶变换是非常单调的，从而概率均匀地分布在整个实线上，这些实线在变换之后对应于粒子可能的动量。

这一笼统的理念可以在数学上表示得更为精确。考虑一个高斯“波包”(wavepacket)，它是 $L^2(\mathbb{R})$ 中的波函数，作为 x 的函数它仅在某个大小为 $2a$ 的区域有可预见的幅度：

$$\psi(x)=e^{-x^2/2a^2} \tag{87}$$

这一波函数的傅里叶变换(即，变换到“动量表象”)为：

$$\bar{\psi}(k)=e^{-a^2k^2/2} \tag{88}$$

其中 k 是“波数”；动量由 $p=\hbar k$ 给出。这一高斯波包的宽度为 $2/a$。因而，在位置上拥有窄峰(即 a 很小)的波函数在动量上是宽分布的(即 $1/a$ 较大)。具体而言，设定 $\Delta x\approx a$ 和 $\Delta p=\hbar\Delta k\approx 1/a$，我们有 $\Delta x\Delta p\approx\hbar$。(这一表达并不是非常

---

① 应用于光子的关系式 $p=h/\lambda$ 是作为爱因斯坦[1905]对光电效应解释的一部分而引入的，并由德布罗意(de Broglie)[1924]推广到物质粒子。

标准的不确定关系，但这里的推导并不需要很精确。）

4.2.3 代数推导

事实上从量子形式体系到不确定关系有许多种途径。这里我们只考虑一种常见的推导，一部分是因为它会给出“不确定度”的附加意义，一部分也因为这一推导较之前的两种推导更为严密且得到了不确定关系的精确形式。

给定一个可观测量 F，定义 $\Delta F := F - \langle F\rangle$（式子右边是 F 的期望值，这里我们不明确态是什么）。$(\Delta F)^2$ 的期望值是 F 的弥散。事实上，

$$\langle(\Delta F)^2\rangle = \langle F^2 - 2F\langle F\rangle + \langle F\rangle^2\rangle = \langle F^2\rangle - \langle F\rangle^2 \tag{89}$$

它是“弥散”的标准统计概念，物理学家们常常称为“均方差（或均方偏差）”，统计学家们称为“方差”，其平方根是标准偏差。现在，令 F 和 G 是可观测量（自伴算符），则施瓦茨（Schwarz）不等式（第 7.1.3 节）意味着：

$$\langle(\Delta F)^2\rangle\langle(\Delta G)^2\rangle \geqslant |\langle \Delta F\Delta G\rangle|^2 \tag{90}$$

参见樱井（Sakurai）[1985，36]，直接的代数操作将式（90）变换为任意可观测量 F 和 G 的标准不确定关系：

$$\langle \Delta F\rangle\langle \Delta G\rangle \geqslant \frac{1}{2}|\langle[F,\ G]\rangle| \tag{91}$$

例如：

$$\langle \Delta P\rangle\langle \Delta Q\rangle \geqslant \frac{1}{2} \tag{92}$$

或者，如果我们不设定$\hbar = 1$，则右手边为$\hbar/2$。

上面我们说过这一推导会给出不确定度的某些意义。特别是我们现在可以看到，“不确定度”严格指的是可观测量在给定态下的弥散（标准偏差），通常把它理解为在都处于某给定态的系统系综中，对它的值“散布”的测度。

4.2.4 局限性和普遍化

式（91）的推导明确表示任意两个非对易算符会产生某个不确定关系。因此式（91）是非常普遍的。然而，在不用算符表示的量之间也存在“不确定关系”。最著名的是能量—时间不确定关系，众所周知，它的解释是有问题的，这是因为时间不是量子理论中的可观测量（没有表征时间的自伴算符）。另一个例子是相位和光子数（也没有“相位”算符）。对如何理解这些其他的不确定关系，人

们给出了许多不同的提议，但这里我们只注意到一点，即显然它们必须在不同于式(91)的意义上来理解。

此外，式(91)还面临着其他的问题。第一，回想式(91)是态依赖的。事实上，如果我们选择一个态，它是 F(或 G)本征态，则即使 F 和 G 不对易，式(91)两边也都为零，这肯定违背了“不确定度”的精神。(或许“解决之道”是注意到如果 F 和 G 不对易，则一般对处于 F 本征态的系统而言 G 的弥散将是非零的。)第二，要记住的是“弥散”本身会令人误解。即使当大部分的概率集中于 F 可能值的很小区域时，位于远离 F 平均值处的很小概率也会导致其弥散变得很大。对于这些缺陷的处理已经有了一些提议，见[Uffink，1994]。

4.2.5 “波粒二象性”

通常与不确定原理相关的一个量子理论实验是双缝实验(在某种形式上，它早在量子理论提出之前就做过了)。实验设置如下：粒子源(或单色光，即光子)置于一具有两个平行缝的不透明挡板前。挡板后是一个屏幕(如照相底板)。粒子被射向挡板上的缝，屏幕在粒子撞击到时记录每个粒子的位置(见图3)。

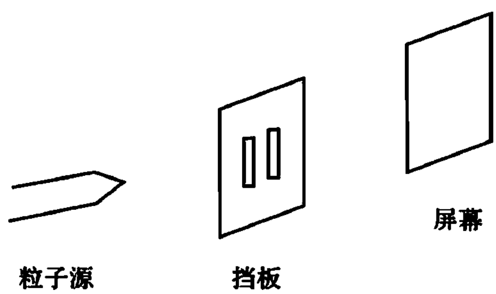


图3 双缝实验

实验要点如下。我们从粒子源处发射一束粒子到挡板，若两个缝都开放时不能确定粒子穿过哪一个缝，则在屏幕上形成干涉图样，这是当某种波穿过双缝时人们所期待的情景，见图4(a)。另一方面，若我们确定粒子穿过了哪个缝，则屏幕上不会出现干涉图样——取而代之显现的是粒子特征的图样(在每个缝之后有一个“斑条”)，见图4(b)。再者，人们可以在实验中一次发出一个粒子，在此情形下人们看到屏幕上的“点”，且如果不确定粒子穿过哪个缝的话

最终这些点将显现出干涉图样，见图4(c)。

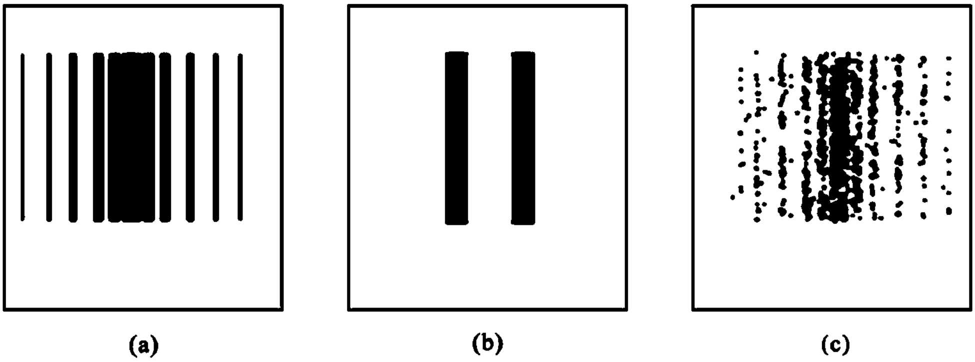


图4 双缝实验的结果

(a)粒子自由地穿过双缝；(b)当确定每个粒子穿过哪个缝时；(c)每次发射的一个粒子自由地穿过双缝时(在屏幕上检测到800个粒子之后的积累)

这一实验阐明了“波粒二象性”：当我们测量粒子(干涉)的类波特性时得到类波行为(干涉图样)，而当我们测量粒子的类粒子特性(粒子穿过哪一个缝)时得到类粒子行为(没有干涉图样)。

它也间接地阐明了不确定关系。考虑到实验中如何以合理的准确度决定粒子穿过哪一个缝——在那种情形下，我们必须以高于 $d/2$ 的准确度来测量粒子的位置，其中 d 是两个缝之间的距离。若干涉图样要在测量时保留下来，则粒子的动量不能被干扰得太多，以免以可观的概率从相长干涉区域[从波动理论的观点看，在这里波通过每个缝都是相干的，也就是图4(c)中许多点出现的区域]被折射到邻近(或任意)的相消干涉区域，也就是图4(c)中少量或没有点出现的区域。粗略的三角法分析(trigonometric analysis)表明，事实上我们位置测量中的不确定度与所要求的动量最小不确定度之积必定违背了位置与动量之间的不确定关系。换言之，如果不确定关系似乎要求在双缝处对粒子位置的测量要足以用于确定(以好的准确度)粒子通过哪个缝的话，那将破坏干涉图样(破坏越严重，位置测量越准确)。

4.3 不确定关系的解释

4.3.1 观察的—认识论的解释

将不确定关系的解释分为两类是有利的：一种单纯用观测上可获得的关于可观测量值的事实来理解不确定度，另一种把观测的不确定性归类为更基本的“本体论的不确定度”或“不准确度(indeterminacy)”。我将依次考察这两种解释。

4.3.1.1 作为不确定度的不确定性 术语“不确定度”和将不确定性理解为标准偏差(弥散)充分表明其对不确定关系是与认知有关的，乃至是操作主义的理解。特别是人们非常倾向于用可观测量真实值观测后的不确定度来理解不确定关系。

这一不确定度是关于单个系统的还是关于系综的仍然是一个问题。在后一情形中——即就像对标准偏差通常的统计理解那样，我们把不确定度理解为反映了系综中(某个可观测量)值的弥散——不确定关系的“解释”并未超出对标准偏差直接的统计理解。当然，此情形中在某种意义上不确定度仍然适用于单个系统，即当这些单个系统是从系综中随机抽选出时。但此外，人们可能会认为不确定度的概念适用于单个系统，与对任意系综的考虑无关。事实上，人们或许会论证说仅在这种情形下我们才真正理解了为什么不确定度在统计层面成立。无论如何，我们来考虑那种可能性。

其理念是在(单个!)被测系统中，测量(或一般说观测)总的来说减小了我们关于某可观测量 F 值的不确定度。不确定原理则被理解为一种论断，对 F 值不确定度的减小意味着对与 F 不相容的那个可观测量值的不确定度的增加。

这一断言的意义足够清晰，但理解为什么不相容的可观测量拥有这样的特征另当别论。假定 F 和 G 是不相容的，且假定我们测量 F，然后再测量 G。为什么第二次测量会破坏我们之前关于 F 值的认识呢？事实上，假定第二次测量与第一次测量隔开且 F 的值在此期间并不被干扰，第二次测量如何破坏我们之前关于 F 值的认识呢？

4.3.1.2 爱因斯坦早期的思想实验 对量子理论的早期批判——尤其是爱因斯坦——质疑了此类问题且确实提出了(思想)实验，这些实验显然想要表

明事实上有可能“击破”不确定原理。爱因斯坦提出的著名实验涉及了一个安装在弹簧上的标准双缝装置(见图3)。其基本理念是，在不干扰粒子本身的情况下通过测量粒子与挡板间的动量交换并用弹簧来确定粒子穿过哪条缝。若粒子源恰好位于双缝之间的平面上，则如果粒子穿过右手边的缝(且缝足够窄)，挡板一般将经历向右的反冲，等等。

4.3.1.3 对爱因斯坦的回复 这里，爱因斯坦对不确定原理的挑战是经验性的，与量子理论所允许的相反，他宣称有可能制备一个系统，其处于精确位置和动量的态(或至少比不确定关系所允许的精确)。因而，对他的回应也必须是经验性的。如果确实有可能制备一个处于位置和动量精确态的系统，则我们应当能够用我们关于位置或动量的知识对系统做出一个可证实的预言。特别地，若刚好在粒子穿过双缝时对它的位置测量在事实上未干扰粒子的动量，则我们应该仍然能在屏幕上看到干涉图样，这与我们在标准双缝实验中看到的一样。另一方面，若干涉图样消除(见第4.2.5节)，且随着位置测量变得越来越精确而形成了“两个斑条”，则不确定原理的认知版本成立。

在当前实验能够分辨的程度上，似乎对粒子穿过哪个缝的测量确实消除了干涉图样。事实上，一个出色的实验似乎要表明的不仅是这些，相反，似乎重要的是是否保留了对测量结果的记录，这里我们指的“记录”是在宇宙可测量的物理态中对结果的编码。该实验[①]本质上是双缝实验，在其中的一个缝后面放置探测器。此外，探测器在下述意义上是“可擦除的”。一旦粒子经过探测器，它就在探测器的态上留下一个痕迹。我们可以选择放大这一“痕迹”以将它转变为指示粒子出现的可辨识信号，或者我们可以完全擦除它，从而探测器的态不再包含关于粒子是否曾经进入探测器的任何可获取的信息。

有了这种可擦除的探测器之后，现在想象一下做如下实验。向挡板发射粒子，一次一个。在粒子穿过挡板之后，要么擦除探测器，要么不这么做。在我们擦除探测器的那些操作中，粒子建立了干涉图样，如图4(c)。在我们不擦除探测器的那些操作中粒子并不建立干涉图样，但代之以“像粒子般的表现”，如

① 例如见[Scully and Walther(斯库利和沃尔特)，1989]和[Walborn *et al.*(沃尔伯恩等)，2002]。

图4(b)。这一实验是相对近期才做的且需要进一步的仔细研究，但它充分表明了正如其他量子力学实验那样，不确定原理的认知版本是一个事实。

4.3.2 本体论的解释

在不确定原理的认知版本显然完好无损的情形下，我们进而追问它为什么是真的。不同的解释会对此问题给出不同的回答：有些人主张“干扰理论”，根据此理论，对一个量的测量会以不可控和不可预言的方式物理地干扰其他量（对应于不相容的可观测量）；其他人认为实在与我们的知识相匹配，即我们不能够知道 G 的值（当知道不相容的 F 的值时），因为事实上 G 没有值。

此观点有许多种版本，这里我们考虑两种。第一种版本常常被认为是量子理论“哥本哈根”解释的一部分，它建立在某种物理量意义理论的确证主义或操作主义版本上。根据这一理论，一个物理量当且仅当它被测量（即进行适当的物理操作并得到适当的结果）时才有值。从而人们必须论证说在同一系统上同时对不相容的可观测量进行测量在物理上是不可能的。

这一论证提出了不确定原理的第二个本体论版本，根据该版本一个物理量仅在它严格定义所要求的条件满足的情形下才是明确的。当然，这样的表述会使该观点听起来几乎是同义反复的。它的真正内容来自于下述论证：确实存在为了使特定物理量是明确的而必须具备的非平凡物理条件，且给定量严格定义所要求的条件不能够同时适用于任意不相容量严格定义所需要的条件。

这样做的话，由于参考系在物理量定义中的作用（第3.4节），此类论证会有破绽。因此，举个例子，“动量”必定指的是“相对于 X 的动量”，其中“X”是定义参考系的某个物理系统。但如果“X”（更精确地是它所定义的参考系）是一个非惯性系统，则它不适用于定义动量（至少对任意的时间段不适合），除非我们知道它与某个惯性系 F 的关系，但这样一来我们就实际是在定义相对于 F 的动量了（第3.4.1节）。类似地，一个非惯性的物理系统也不适合用于定义位置（与前面出于相同的理由）。这些简短的要点回顾了上述更全面的讨论，且我们最好到此为止。

4.4 爱因斯坦—波多尔斯基—罗森论证

目前为止关于不确定关系所说到的每件事情都与下述陈述相一致，即可观

测量F值的确定性在不相容量G的测量中丧失是由G测量引起的一种对F值未知且不可控的物理干扰产生的。事实上，海森堡对不确定关系的“光学”推导恰好应用了这一理念。但由爱因斯坦、波多尔斯基和罗森(EPR)①给出的论证意图表明不确定关系事实上不能够以这种方式来理解。

该论证承认不确定原理的认知版本是真的，即不相容可观测量的值不能够同时得到确证。这里的论点是质疑不确定原理的本体论版本。

4.4.1 不完备性

特别地，其关键点是对量子理论的完备性提出质疑。回顾(第1.2.3.9节)量子理论的标准解释中，一个可观测量F在态为W的系统下，仅当W赋予F的某个可能值以概率1(赋予其他值的概率为0)时才拥有值。这一解释直接意味着不存在能同时赋予所有可观测量(或确切地说是任意两个没有共同本征矢量的可观测量)以确定值的量子态。因此，任何成功表明两个这样的可观测量必定有确定值的论证都意味着量子理论是不完备的——更精确地说，它表明在量子理论的标准解释下量子态并不描述(且不能描述)系统完备的物理态。EPR论证恰恰是要证明这一论述。

4.4.2 普遍实验

一般EPR考虑的实验包含了制备处于下述态的一对粒子(称其为α和β)：

$$|\Psi_{\mathrm{EPR}}\rangle=\frac{1}{\sqrt{N}}\sum_{n=1}^{N}|a_n\rangle|b_n\rangle \tag{93}$$

其中，$|a_n\rangle$和$|b_n\rangle$形成了正交集。② 因而，系统α的可观测量A和系统β的可观测量B(它们对应于本征值a_n和b_n的本征矢量分别为$|a_n\rangle$和$|b_n\rangle$)在此态中全关联(第1.2.6.4节)。

所以假定α和β处于态$|\Psi_{\mathrm{EPR}}\rangle$，虽然它们在空间上是分离的。则在$A$和$B$之间的全关联允许人们如EPR所说的在“不以任何方式破坏”α的情况下仅通过在β上测量B而发现A的值。

---

① 有充分的理由相信EPR[1935]提出的论证不完全是爱因斯坦本人想要的(论文不是由他写的)。例如见[Fine，1986，Chapters. 3—5]。

② 事实上存在式(93)的连续版本，该版本适用于当我们考虑如位置和动量这样具有连续谱的可观测量的情形。见第4.4.3节。

现在到了关键论点上。将从$|a_n\rangle$和$|b_n\rangle$中得到的正交归一化基$\{|a'_n\rangle\}$和$\{|b'_n\rangle\}$表述如下：

$$|a'_n\rangle=\frac{1}{\sqrt{2}}(|a_n\rangle-i|a_{n+1}\rangle)\quad |b'_n\rangle=\frac{1}{\sqrt{2}}(|b_n\rangle+i|b_{n+1}\rangle)\tag{94}$$

其中下标为“$n+1$”的求和是模数N(即，$N+1=1$)。在此基矢下态$|\Psi_{EPR}\rangle$恰好具有相同的形式：

$$|\Psi_{EPR}\rangle=\frac{1}{\sqrt{N}}\sum_{n=1}^{N}|a'_n\rangle|b'_n\rangle\tag{95}$$

为了证明这一点，将式(94)代入式(95)并简化即可。消去形如$-i|a_n\rangle|b_{n+1}\rangle$和$i|a_n\rangle|b_{n+1}\rangle$的“交叉”项。因而存在另外的可观测量$A'$和$B'$(它们的本征矢量分别为$|a'_n\rangle$和$|b'_n\rangle$)，它们在态$|\Psi_{EPR}\rangle$下也是全关联的。此外，$A'$并不与$A$相对易(即不相容)——确实，它们没有共同的本征矢量——且类似地，B和B'也不对易。我们仍然能够通过在β上测量B'从而得到α上A'的值。①

最后，注意这里所有的描述都是量子力学的。人们有时也见到或多或少用经典术语描述的EPR论证。例如，令A和B为位置和动量(见第4.4.3节)，人们可能会倾向于将EPR态的制备描述如下：从共同的源以(大小)相同但方向相反的力发射两个等质量的粒子。它们的位置(到粒子源的距离)和动量(相对于粒子源)则会全关联。这一图像确实诱人，但它是完全错误的。事实上，标准量子力学意味着刚刚描述的态不能够以这样一种方式制备，从而允许人们根据β的位置(或动量)来推知α的位置(或动量)，因为这样的推测会要求我们精确地知道粒子源的位置和动量，而这样的知识已经违背了(认知上的)不确定原理。

4.4.3　位置和动量

上面我们假定了全关联的可观测量拥有离散谱。事实上，在他们考虑一个具体例子(而非上面考虑的抽象情形)时，EPR指的是位置和动量，还有处于位置和动量全关联的态(在它是两个粒子动量的和还有它们位置差的共同本征态这一意义上)。确切地说，对某个确定的x_0，有：

$$\Psi_{EPR}(x_1,x_2)=\int_{-\infty}^{+\infty}e^{(2\pi i/h)(x_1-x_2+x_0)p}dp\tag{96}$$

① 这里所描述的情形在数学上并不严格与EPR考虑的相同，但结果是相同的。

然而，值得指出的是，事实上这个态完全不是系统所允许的态——它不是 $L^2(\mathbb{R}^2)$ 中的矢量。此外，即使这个态能够被制备出来也必然会在时间演化中弥散掉（与有限的势能相关），从而迅速成为不完全关联的态。①

为了克服 EPR 案例中的这些局限，例如，人们可以考虑一个接近全关联态的窄高斯波包，但这样做会使得论证变得越发棘手。且我们应当谨记用位置和动量表述的讨论事实上仅是 EPR 给出的用以表明主要论点的一个例证。

最后，事实上考虑一个涉及有界可观测量的例子会更容易一些。最简单的情形是所谓自旋粒子的"单重"态。一般地，这一情形对应于（对于适当选择的 $|a_n\rangle$ 和 $|b_n\rangle$）上文中的态 $|\Psi_{\mathrm{EPR}}\rangle$，其中 $N=2$。我们将应用上文在全关联讨论中获得的认识来继续讨论可观测量 A 和 A' 及 B 和 B'。

4.4.4 论证

这一实验情形如何能够用于论证量子理论的不完备性呢？EPR 希望建立的结论是 A 和 A' 同时都拥有精确的值。因为它们不具有共同的本征矢量，所以这一结论与量子态的标准解释不一致（见第 1.2.3.9 节）。

得出这个结论的一个诱人的方法是假定一旦在 β 上测量 B，就在 α 上确立了 A 的值，从而我们就能够在 α 上测量 A' 并进而直接确定它的值。然而，此提议的问题也很明显：应用不确定性的干扰理论可以立马推断出结论，即在 α 上对 A' 的测量干扰了之前建立起来的 A 值。这使我们想起这一讨论的要点事实上是在不"以任何方式"干扰 α 的情形下确立 A 和 A' 的确定性。

EPR 的策略涉及两个假设。第一个是他们的"物理实在标准"，它声称每当一个可观测量的值能够被确定地预言时可观测量实际上就拥有该值。至少某些版本的标准解释能够很容易认同这一假设。注意这一标准与上述提到的（第 4.3.2 节）某种确证主义或操作主义观点不一致，在后者看来仅当一个值作为测量结果被得到时系统才拥有该值。在我们介绍第二条假设之前，我们先来看看只在这一假设下能得到什么。

注意如果我们在 β 上测量 B，我们就能够确定性地预言 α 的 A 值。对于 B'

---

① 关于这里的态"弥散"，我们大体上指的是，它接近于 $\mathbb{R}^2$ 上的均匀分布。对这些要点的讨论和更多的文献见[Dickson，2002b]。

和A'也类似。当然，我们不能既测量 B 又测量 B'。因而，代之以考虑对 B 和B' 的实际测量，我们考虑对 B 和 B'非现实但可能的(即反事实的)测量。我们有：

前提 1：可能的是，在β上测量 B，在此情形下 α 拥有 A 的确定值；

前提 2：可能的是，在β上测量 B'，在此情形下 α 拥有 A'的确定值。

根据这两个前提(它们来自于物理实在标准)，EPR 希望推论出：

结论：也许 α 同时拥有 A 和 A'的确定值。

然而，此结论并非来自于前提 1 和 2。事实上逻辑问题在于某种程度上不能够保证可能的条件("B 被测量"和"B'被测量")同时可能(co-possible)的。实际上，正如我们所知道的，它们不是同时可能的。①

因此 EPR 需要另一个前提。他们引入"非干扰"概念用以帮助修补论证：虽然我们据以推断α上 A 确定性的条件(B 测量)与我们据以推断 A'确定性的条件(B'测量)不相容，但它们之间的差别对于 α 并没有影响，因为它们只涉及β环境的改变，而β在空间上是与 α 分离的。

然而，不是任意这样的原则都会起作用。例如，考虑如下原则：

> **弱非干扰性**：若在β上测量 B 且(因而，根据物理实在标准)对于α而言 A 是确定的，则要是我们没有在β上测量 B，α 仍然拥有 A 的确定值(用带撇的可观测量替代不带撇的可观测量也一样)。

这一原则可以用于否定是对 B 的测量使得 A 实现确定值，但它对于得到 EPR 结论是不充分的。作为替代它们的需要：

> **强非干扰性**：若在β上测量 B 且(因而，根据物理实在标准)对于α而言 A 是确定的，则要是我们代之以在β上测量 B'，α 仍然有 A 的确定值(用带撇的可观测量替代不带撇的可观测量也一样)。

① 考虑下述类似论证：(i)有可能纸被点着了，且在此情形中它将化为灰烬；(ii)有可能纸没有被点着，且在此情形中它仍然是完整的；(iii)有可能纸既是完整的，又化为灰烬，显然，该论证是无效的。

弱原则对于得到 EPR 结论不充分是因为引入对 B' 的测量(这与仅仅不测量 B 相反)会破坏此情形中的本质特征，特别是破坏允许从 β 的(B 的值)测量结果推论 α 特性(A 的值)的特征。[①]

当然，强非干扰原则被认为是代表了某个“定域性”概念。特别是其中的理念是：在两个粒子是类空分离的假设下，对 β 做出的行为不能够对 α 的特性造成影响。爱因斯坦的狭义相对论想必许可这一假设。[②] 无论如何，在强非干扰原则下 EPR 对于上述结论有了一个在逻辑上有效的论证。事实上，他们可以(也似乎要求)建立上述结论的更强版本，即用“实际上”取代“可能是”，表述如下：即进一步论证 B(或 B')测量并不会使得 α 有 A(或 A')的值，从而即使当 $B(B')$ 并不被测量时，α 也必须同时拥有 A 和 A' 的值。

4.4.5 对 *EPR* 的回应

事实上，人们可以通过否定定域性来否定强非干扰性。我们将在下文中考虑定域性在量子理论中的地位(第6节)。这里，我们考虑其他两种对 EPR 的回应。

我们已经或多或少地遇到过这些回应之一，注意到 EPR 论证对那些确证主义者或操作主义者的观点没有影响。在这些观点看来，不仅在我们要知道一个物理量的值时需要对它明确测量，它要有一个值也需要明确测量。当然，这样的观点必须否定物理实在标准(牢记，这仅是充分条件而非必要条件)，而许多人发现这一标准非常有说服力。(除了确实拥有问题中的值之外，还有什么其他合理的理由能够使我们确定地预言它的值呢?)

但玻尔提供了一种回应，该回应似乎既不依赖于这种确证主义者或操作主义者的策略，也不依赖于——如他所述——对非定域性的明确认可。特别地，回想 EPR 意图回避对不确定性的“干扰”说明，根据这一说明测量 A' 会物理地干扰 A 的值。这里因为我们仅对 β 做了测量所以远不清楚测量 B' 是如何干扰 α

---

① 反事实的可能世界语义学使这点变得显而易见。虽然最接近“B 被测量且 A 对 α 是确定的”世界的“B 未被测量”世界可能都是“A 对 α 是确定的”世界，但最接近的世界可能是不包含任意“B 被测量”的世界，从而最接近“B 被测量且 A 对 α 是确定的”世界的“B 未被测量但 B' 被测量”世界不需要是“A 对 α 是确定的”世界。

② 见本书中马拉蒙特所撰写的章节。

的A值的——毕竟，这样一个干扰只能是非定域的。玻尔的回应否定了强非干扰性，但没有（因此这一陈述仍然可行）认可对α（例如A的值）物理上直接的（玻尔用到了术语"力学的"）干扰是对β（例如B）任意测量的后果。相反，回顾上述的论点（第4.3.2节），即特定物理量的严格定义性依赖于特定物理条件的满足。

现在像玻尔（和EPR）那样用位置和动量来考虑EPR实验。我们将假定（在不失普遍性的情况下）位置和动量是相对于粒子源定义的。在玻尔看来，只要系统（粒子加上粒子源）依然封闭，也就是说只要总动量守恒，则所有粒子的动量和（即系统的总动量）的严格定义性保持不变。但对β位置的测量引入了对β动量的干扰。系统不再是封闭的，β要么失去动量给外在于系统的客体（测量装置），要么从中得到了动量。但总的动量（相对于粒子源而言，它对于β上的这个外在影响"并不知情"）不再守恒，于是它严格定义（相对于粒子源）的条件不再满足，因而我们建立在总动量严格定义基础上做出的关于α动量的严格定义的推论就不再有效了。换言之，在玻尔看来α动量严格定义所需要的条件（在这一环境下）在我们测量β的位置时已经不再满足了。注意干扰的"旧"类型仍然在起作用——我们实际上假定β的动量受到了对其位置测量的物理干扰——但此外还有另一种干扰类型在起作用，即对β位置的测量"干扰"了α动量严格定义所需要的条件（在这一环境下）。因此玻尔相信他能够通过（本质上）否定强非干扰性来避免EPR的结论，但这种方式并不需要一个非定域的干扰，即在β与α能量或动量交换意义上的干扰。

4.4.6 我们的立场

最后，不论人们对这些关于EPR的回应满意与否（至少在我们考虑第5节和第6节的内容之前，我们不应该再对此做任何判断），应当强调的是没有哪种回应表明EPR论证是没有根据的，更不用说是无效了。确实在某种意义上这些回应意图清晰地阐述一种为量子理论观点辩护的方法，此观点通过以一种一致的和可能似真的方式否定EPR论证中的一个前提来避免EPR的结论。因而EPR的结论不会因为这样的回应而遇到问题。事实上，许多量子理论的哲学家们或多或少地赞同EPR关于标准量子理论不完备的论述，如果不是因为EPR论证的话，则是因为所谓的"测量问题"。接下来我们将考虑这一问题。

5. “测量问题”

测量问题或许是量子理论基础中讨论得最多的问题，也激发了许多不同的、形形色色的量子理论解释。在本节中我将首先回顾该问题(第5.1节)，强调其普遍性。之后我将考虑一些对该问题自然的但最终并不能令人满意的回应(第5.4节)。在最后的小节(第5.5节)中我将提及量子理论解释的一些案例，其中的每种解释都主要致力于测量问题的求解。

5.1 基本问题

5.1.1 “薛定谔的猫”

回想(第1.2.3.2.b节)叠加原理意味着，对任意可观测量 F，若系统能够将 F 的每个(或事实上只是两个)本征值作为它关于 F 的值，则同样可能的是系统不再拥有(假定本征态—本征值联系，第1.2.3.9节)关于 F 的值(因为它处于 F 本征态的叠加)。

若 F 是陌生量子世界中的某个不熟悉的可观测量，或许我们还能够接受这一结果。但它要是我们日常经验中如面包和黄油这类中等大小固态物体的可观测量会怎么样呢？要是“房子的(近似)位置”，或“马的(近似)动量”，等等此类呢？1935年，薛定谔用如下案例阐述了我们面临的此类问题。

> 人们甚至可以提出相当荒谬的情形。把一只猫封闭在一个铁盒子里，并且装有如下装置(必须保证此装置不受猫的直接干扰)：在一台盖革计数器内置入极少量放射性物质，由于数量极少在一小时内也许只有一个原子衰变，但等概率下也可能没有一个原子衰变。假若衰变事件发生，则计数管会放电并通过继电器启动一个锤头打破装有氢氰酸的小烧瓶。如果放置这整个系统一个小时，期间没有原子衰变则人们说猫仍然活着。整个系统的波函数会将此情形表达为活猫与死猫(原谅未给出表达式)各半的混合态或混乱状态，见薛定谔[1935a]。

当然，论点是在某段时间之后原子处于“衰变”与“未衰变”的叠加，因此，锤头、毒药和终端的猫都处在相应的叠加中——在该情形中猫处于“活”与“死”的叠加。但当然我们从未目睹过处于此种状态的猫，很显然量子理论存在严重的问题。

5.2 测量

注意猫本质上充当了衰变—指示器装置——这一残忍的测量仪器。事实上，这正是薛定谔的论点，即人们可以将微观层次(原子)的叠加放大到宏观层次(猫)的叠加，在宏观层次上人们可能更倾向于直截了当地否认叠加(至少是某种叠加，如“活”与“死”的叠加)在物理上有意义，或至少否认它们的存在具有某种像量子理论明显预言的概率那样的东西。意图测量量子力学可观测量的测量装置很明显属于此种类型：它们将某微观量子系统的态“放大”到可以直接观察的仪器(指示器或“指针”)态。①

当然，这种放大是一个物理过程，因而原则上它被描述为薛定谔方程(或某种其他量子理论的运动方程)的解，那些方程是线性的。(事实上，这种线性恰好是叠加原理的部分的根据是：解的任意线性组合仍然是解。)回顾系统的动力学演化可以用一组幺正算符来描述，这些算符当然是线性的。

把给定的(在某仪器与某被测系统之间的)测量相互作用用算符 U 来描述。例子见第 3.1.2 节。令仪器的“测量预备”态为 $|\Psi_0\rangle$，令它的“指针”态为 $|\Psi_n\rangle$。对于被测可观测量 F 的每个本征态 $|f_n\rangle$ 而言，我们假定测量相互作用的结果是用下述结果准确表示的：

$$U(|f_n\rangle|\Psi_0\rangle) = |f_n\rangle|\Psi_n\rangle \tag{97}$$

从而态 $|\Psi_n\rangle$ 表明被测系统拥有 F 的值为 f_n，但之后由于线性可以得到：

$$U(\sum_n k_n|f_n\rangle|\Psi_0\rangle) = \sum_n k_n|f_n\rangle|\Psi_n\rangle \tag{98}$$

现在我们遇到麻烦了。 本征态—本征值联系意味着在这一终态中仪器的

① 然而，回顾我早先(第 1.4.3 节)对如 | 死猫〉这类“纯”态的怀疑态度。这一怀疑给出一种测量问题的出路了吗？并没有。我们这里最好还是用混合态来表述并指出问题，它仅在于下述事实，即在测量结束时仪器态可能会赋予“指针可观测量”所有可能的值以非平凡概率。

指针—可观测量实际上没有值，指针并没有在指示什么。但在典型的成功测量结束时，仪器确实是在指示一个结果，即使被测系统开始时处于被测可观测量本征态的叠加态。标准量子理论似乎与经验的显著事实相冲突，因而有“测量问题”。

5.3 问题的普遍性

上文中描述的“测量问题”至少在三种意义上没有能以充分的普遍性涵盖真正的问题所在。第一，它依赖于非常保守的本征态—本征值联系和“测量”的限定概念；第二，它不能表明叠加的无处不在；第三，它依赖于仪器的宏观指针态是纯态这一有问题的假设，而事实上它们几乎肯定是混合态。在本小节中我将简要考虑这些论点。最后我将简要提出一个相关问题——“经典极限”问题。

5.3.1 “不可行”(No-go)定理

上述论证中导致测量问题的两个重要前提是本征态—本征值联系和对成功“测量”的说明。这两个假设都能够在相当程度上弱化。① 我将依次考察它们。

5.3.1.1 确定性的弱条件 首先，我们会引入一个关于何时可观测量拥有确定值的弱条件，大体如下。仅当复合系统的终态处于态的混合——根据本征态—本征值联系，其中每个态都拥有指针可观测量的一个确定值——时，仪器的终态才赋予仪器的指针可观测量以确定值。换言之，我们现在允许采纳对复合系统混合态的无知解释，承认当复合系统拥有刚刚描述的某种混合态时实际上它正处于混合态中的一个纯态，因而根据本征态—本征值联系仪器拥有指针可观测量的一个确定值。注意该条件适用于复合系统，而不是仪器本身。其中的差别很微小，但很关键。例如，在理想测量中仪器本身总是处于指针可观测量本征态的混合中。然而，这并不是说可以对此混合态进行无知解释。我将在下文稍有不同的情况中讨论这一点(第5.4.5.2.a节)。

5.3.1.2 测量的弱解释 我们也可以按照如下线索弱化对测量的解释。

---

① 对测量问题在标准量子理论背景下“不可解”的证明由来已久，且日趋普遍。对早期历史的回顾，和对该定理可论证的最简单证明，见[Brown(布朗)，1986]。对该问题较新的考察和更广泛的讨论，见[Mittestaedt(米特斯泰特)，1998，Chapter 4]。

令U表示复合系统的态在用"指针可观测量"POVM，即Q对POVM的E测量过程中的时间演化。我们要求，对于复合系统的任意两个态W和W'，若W和W'至少对于E映像中的一个效应在概率上不同，则UWU^{-1}和$UW'U^{-1}$至少对于Q映像中一个效应在概率上不同。直观地，U使得指针可观测量"以某种方式对被测可观测量敏感(即使仅在概率上)"。

然而要注意到虽然这种测量的较弱解释与确定性的弱标准(第5.3.1.1节)一起足以得到测量问题(即在测量结束时指针可观测量的非确定性)，但同样值得留意的是，按照双正交分解定理(第1.2.6.2节)仍存在某些可观测量(一个用于被测系统，一个用于仪器)，对这些可观测量而言，复合系统的态具有与在理想测量中相同的形式。(即态用它的双正交形式表示，被测系统和仪器的基底将是由它们各自的某个可观测量所决定的基底。)另一方面，这一仪器可观测量是否是我们先前确信其确定的那个可观测量，这是另一个问题，其答案可能取决于相互作用的细节。

5.3.2 叠加的普遍存在性

叠加原理宣称可能态的叠加仍然是可能的态，但或许它们极其罕见。那样的话，可能我们就不需要非常在意测量问题了。

然而，事实上微观层次的叠加是普遍存在的。为明白其缘由，我们只需要考虑自旋情形。回想$S_{\vec{u}}$与$S_{\vec{u}'}$是不相容的，除非$\vec{u}=\vec{u}'$或$\vec{u}=-\vec{u}'$。因此，每个自旋1/2粒子几乎在每个自旋方向上都"在叠加中"。类似的评论也适用于光子的极化。且在实验和理论上都有充分的理由相信，许多基本粒子都拥有高度退局域化(de-localized)的波函数。

此外，为了相信这些微观叠加可能会"被放大"，我们并不需要想象如薛定谔所描述的极端情形。虽然这样的放大是测量的重要部分，但没有理由假定它不会经常自然地发生。毕竟，人类的眼睛只对小部分或很少的光子敏感。似乎可行的是，假定许多发生在自然中的其他相互作用都有将某些宏观客体的态与某些微观客体的态关联起来的效果，且这样的关联(即使是非全关联)，对于产生测量问题的那种放大足矣。

因而，有充分的理由相信测量问题在下述几种意义上是极其普遍的：(a)即使在关于"仪器"何时拥有某个宏观可观测量确定值的明显的弱解释(第

5.3.1.1 节)下，量子理论也显然表明它们并不拥有确定值；(b)用于实现情形(a)所需要的在宏观与微观客体之间的那种关联是非常弱的，并且按照双正交分解定理全关联是普遍存在的(第 5.3.1.2 节)；(c)产生情形(b)所要求的那种叠加和相互作用似乎是非常普遍的(本节)。

5.3.3 经典极限

最后，我将快速提及一个在本书其他地方详细讨论过的问题。在某种意义上，测量问题是指在某种程度上量子世界的“古怪性(weirdness)”并不限于微观世界的问题。换言之，“古怪”的量子世界并不(总是)能够很好地与我们更为熟悉的经典世界切合(mesh)，也就是说微观层次(我们可能能够与它们共存)的叠加并不限于那个层次，而是可以出现在宏观层次(随着论证的进行，我们不能够与它们共存)。这一明显的冲突导致了一个更为普遍的问题：量子理论如何过渡到经典理论(毕竟，其在众多条件下对很大范围内的客体而言非常有效)？此问题被以多种方式反复讨论，但在这里我并不考虑它们。我推荐读者参考本书兰兹曼所撰写的第五章(特别是第 5 节和第 6 节)。

5.4 无解

关于测量问题已经提出了许多的求解，我们将在关于解释的单独章节中考虑其中一些。这里，我们考虑的一些求解事实上并不起作用，或至少说要使它们起作用会遇到非同寻常的障碍。

5.4.1 朴素实在论

5.4.1.1 用于拯救的明显解？ 该问题的明显解是放弃本征态—本征值联系。确实，为什么不将量子理论带来的概率解释为是完全认识论的呢？也就是说，为什么不假定每个可观测量在所有时间都有一个明确的值，且不相容性只是表示了对它们之中一个可观测量值的观测以一种不可控的方式干扰了其他量的值？某种类似于玻尔关于不确定性解释的认识论版本——若它起作用的话——可以用于此观点，这样既解释了不确定关系又回避了测量问题。

5.4.1.2 科亨—斯佩克(Kochen-Specker)定理 虽然这一观点乍看很吸引人，但在著名的科亨—斯佩克定理[Kochen and Specker，1967]中失败了。注意朴素实在论试图为每一个可观测量确定一个单一的值。在进一步的要求下，

人们可以表明这样的赋值是不可能的。

5.4.1.2.a　非语境性　这一进一步的要求在各种方式下是能具体指明的。这里我们用术语“非语境性”表示它。注意，对于在整个空间张成的相互正交子空间的每个集合，赋予可观测量以一个值相当于赋予一个子空间以“1”，其他子空间以“0”。(这些子空间是被赋予值的可观测量的本征空间；赋予一给定本征子空间以1等价于赋予该可观测量以相应的本征值。[①])沿此方式，我们将进一步假定所有恰好具有同一本征空间的可观测量拥有“同一”值，在它们拥有与给定共同本征空间相关的本征值这一意义上。这一假定相当于如下要求，即对任意可观测量 G，若拥有值 g，则对于任意的函数 $f(\ )$，可观测量 $f(G)$ 拥有值 $f(g)$。[②]

非语境性包含了此假设，但此外还进一步要求，对给定子空间 P 一个“0”或“1”的赋值是独立于将哪一个相互正交且联合张成子空间 P 的集合考虑为一个要素。这一要求相当于下述条件，即每当两个可观测量 F 和 G 共享一个本征空间 P(但并不必然地共享它们所有的本征空间)时，当且仅当 G 拥有相应于 P 的本征值时 F 才拥有相应于 P 的本征值(当然，本征值一般是不同的)。注意在此情形下 F 和 G 一般并不对易。

5.4.1.2.b　态独立定理　给定非语境性的朴素实在论相当于下述要求：我们能够找到从希尔伯特空间$\mathcal{H}$子空间的格 $L(\mathcal{H})$ 到布尔格(Boolean)$\{0,1\}$的同态。也就是说我们要求映射 $h: L(\mathcal{H})\rightarrow\{0,1\}$ 满足对任意的 P，$Q\in L(\mathcal{H})$，当且仅当 $P\leqslant Q$ 时才有 $h(P)\leqslant h(Q)$。继而有(根据交集、并集和正交补的定义，它们都是用偏序给出的，见第7.4.2节)$h(P)\wedge h(Q)=h(P\wedge Q)$、$h(P)\vee h(Q)=h(P\vee Q)$ 和 $h(P)^{\perp}=h(P^{\perp})$。但注意在$\{0,1\}$中，算符$\wedge$，$\vee$和$\perp$的运算正像经典逻辑中的运算一样(即，$0\wedge0=0$，$0\wedge1=0$，$1\wedge1=1$，$0\vee0=0$，$0\vee1=1$，$1\vee1=1$ 和 $0^{\perp}=1$)。换言之，朴素实在论要求将量子理论看作是来自于(逻辑意义上的)基本的“经典”理论是可能的。

① 这里我忽略了如位置和动量这类谱是连续的(因而它们可能的值并不对应于子空间)可观测量。毕竟，赋予它们值只会使得朴素实在论的生存更加艰难。

② 注意，一般来说，如果 $G=\sum_n g_n P_n$ 是 G 的谱分解，则有 $f(G)=\sum_n f(g_n)P_n$。

这样一幅图像在经典力学中是可行的。事实上，令 Γ 为经典系统的相空间。物理命题(见第 2.1.1 节)是用 Γ 的波莱尔子集格来表示的(其中偏序是由子集包含给出的)，且有可能定义从这一代数到 $\{0, 1\}$ 的一个同态。事实上，选择一个点 $x \in \Gamma$，并将映射 δ_x 定义为：对任意的波莱尔子集 $S \subseteq \Gamma$ 而言，在 $x \in S$ 时 $\delta_x(S) = 1$，否则为零。此 δ_x 就是一个同态。

科亨—斯佩克定理的内容表述为：对于维度大于 2 的希尔伯特空间的子空间格，不存在这样的同态。定理的这一版本有时称为“态独立”，因为它并不依赖于关于系统量子态的任何预设，而仅仅依赖于作为整体的态空间结构。

5.4.1.2.c 态依赖定理 还存在以假设系统处于某给定量子态为开始的科亨—斯佩克定理的其他版本。在此情形下，我们可以引入赋值的另一个条件：它们必须遵循由量子态产生的概率。例如，若态赋予 F 一给定值 f 以概率 0[1]，则该赋值必须赋予相应的本征空间以 0[1]。存在一些特别简单的态依赖科亨—斯佩克定理，最为著名的是 GHZ，即格林伯格—霍恩—泽林格(Greenberger-Horne-Zeilinger)[Greenberger *et al.*, 1989]定理，它在默明(Mermin)[1990]精确描述的形式下，考虑了一个自旋 1/2 粒子的三粒子系统和下述可观测量：

$$S_x^1,\ S_y^1,\ S_x^2,\ S_y^2,\ S_x^3,\ S_y^3 \tag{99}$$

$$S_x^1 \otimes S_y^2 \otimes S_y^3,\ S_y^1 \otimes S_x^2 \otimes S_y^3,\ S_y^1 \otimes S_y^2 \otimes S_x^3 \text{ 和 } S_x^1 \otimes S_x^2 \otimes S_x^3 \tag{100}$$

其中 S_x^1 是 $S_x^1 \otimes \mathbb{I}^2 \otimes \mathbb{I}^3$ 的缩写，且上标表示与每个可观测量相关的粒子。出于简单性，假定每个 $S_u^{(n)}$ 都被“归一化”以拥有本征值 ± 1(而非通常的 $\pm 1/2$)。因此式(99)与式(100)中的所有可观测量的本征值都是 ± 1。

注意式(100)中的可观测量是对易的。所以我们可以考虑一个“GHZ 态”，它是所有这些可观测量共同的本征态。我们将分别考虑本征值为 $+1$，$+1$，$+1$，-1 的本征态。对处于该量子态的一个系统，把确定值赋予可观测量必须要赋予式(100)中的可观测量以上述值。令 $v(\ )$ 是一个从式(99)和式(100)中的可观测量到它们值的映射，且之后考虑下述值的阵列：

$$\begin{matrix} v(S_x^1) & v(S_y^2) & v(S_y^3) \\ v(S_y^1) & v(S_x^2) & v(S_y^3) \\ v(S_y^1) & v(S_y^2) & v(S_x^3) \\ v(S_x^1) & v(S_x^2) & v(S_x^3) \end{matrix} \tag{101}$$

可能的值总是 ±1。前三行中每一行三个值的积必须是 +1，最后一行的积则为 -1，所以所有 12 个数的积必须是 -1。但另一方面，对每个 $n=1$，2，3 和 $u=x$，y 而言，$\nu(S_u^n)$ 在阵列中恰好出现两次，所以所有 12 个数的积必须是 +1，这与我们上面的结论相冲突。所以对于式(99)与式(100)中的所有 10 个可观测量，不存在与 GHZ 态相一致的赋值。

因此朴素实在论面临着严重的困难。通过科亨—斯佩克定理及其类似的 GHZ 定理，对朴素实在论基本约定的最直接解读会引起逻辑上的冲突。虽然提出了许多建议以避免朴素实在论观点的这一问题，但所有这些建议(当然)都违背了科亨—斯佩克定理的一个或更多条件，且那样的话它可能不再是“朴素”实在论了。

5.4.2 系综解释

系综解释，例如[Ballentine(玻仑廷)，1970]，试图通过约定量子态不是关于单个的量子系统来彻底回避测量问题。量子态本质上是统计的，因而仅能够用于描述系统的系综。所以在测量结束时的态不应该看作是对一个装置(和被测系统)在单个测量结束时的描述，而是在所有这些测量结束时对这些装置(和被测系统)的描述。

文献中有两种这类系综解释的类型，但并不总是作出区分。第一种我们称之为“极简主义(minimalist)”，它是紧缩论(deflationary)的观点，即仅当量子态(例如在一个测量结束时)是对类似制备系统一个系综的描述时它才是有意义的。

系综解释的第二种类型作出了更强的论述，即有下述效果的某种论述：除对量子态的紧缩论说明之外不需要任何其他说明，即科学上并没有要求理论描述单个系统。此观点认为，这样一种要求是决定论世界观的人工产物，且一旦我们接受统计理论则量子理论通常的“神秘”会直接消失。但是，量子理论哲学和基础研究领域的绝大多数学者并不信服此观点。

此外，甚至极简主义论述也面临着一个问题，该问题的存在同样表明了为什么更强的论述至少还需要进一步发展。科亨—斯佩克定理已经表明，量子态产生的概率不能够被直接理解为是经典概率。因此系综解释似乎需要转变为其他观点，如稍后将会讨论的量子逻辑观点。

5.4.3 塌缩假设

测量问题的“标准”求解是所谓的量子理论的“塌缩假设”(也称为“投影假设”)。在第3.1.2节中已经提到过塌缩假设，是冯·诺伊曼首次对它进行了详细讨论，参见[1932，351]和该书英文版的第417—418页。该假设可以有许多种表述方式，比如：

> **塌缩假设** 在对处于态W的系统可观测量F的测量中，测量结果将会是对应于F的某本征空间P的一个本征值，且之后系统的态会是$PWP/\mathrm{Tr}[PWP]$。

若该态可以写作为矢量$|\psi\rangle$，则“塌缩”等于将此矢量投影到ran P，即投影到$P|\psi\rangle$，并重新归一化该结果。

在许多方面塌缩假设不能令人满意，但首先值得说明的是，大多数时候对于人们想要从量子理论中作出的大多数预言而言该假设非常有效。

关于塌缩假设常常提到的一个问题是它相当于系统的非连续(和不可逆)演化。这样得到的量子系统演化图景确实是奇怪的：服从薛定谔方程(或某种其他量子运动方程)的连续的、决定论的、可逆的、幺正演化(在希尔伯特空间中)被态的非连续、不可逆、概率的、随机变化所打断。

人们至少想要知道为什么会发生这些打断。显然，在该假设的陈述中它们与测量相联系。但测量有什么物理上的特殊呢？遗憾的是尚没有令人信服的答案。事实上，塌缩假设最明显的问题是它依赖于一个未经分析的“测量”概念。大多数时候理性的人们能够对测量什么时候发生达成一致，但这一事实完全于事无补。因为量子理论基础和哲学领域中的大多数研究者似乎都一致地认为，若量子理论被假定是一个基础的理论，则它应该告诉我们测量什么时候发生(或更一般地说测量是什么)，而不是反过来要我们告诉它。世界不应该依赖于我们“告诉它”什么时候要使一个态塌缩。

5.4.4 宏观—微观区分

我们对薛定谔猫(第5.1.1节)的讨论给出了对“测量是什么”问题的一个回答，且有时候该回答是作为对测量问题的回应而给出的。该提议是，测量发生

于一个微观系统的态以该微观态变得与某个宏观系统的态相关联的方式“被放大”时。对于典型的测量拥有此特征人们并无异议。遗憾的是，此特征依赖于另一个定义不明的概念，即微观与宏观之间的不同。

该区分有时也用于区分“易于可逆的”和“本质上不可逆的”。因为微观系统具有的自由度少，它们的行为常常很容易倒转，而宏观系统的行为在巨大的自由度下倒转——如果在实践中可能的话——是很难的。

在另外的情形中——不论这一要点是用大小还是用可逆性来表示——该理念被假定为，在一恰当的大(或不可逆的)尺度下物理系统表现为经典的。事实上，人们有时见到以这种方式刻画的量子理论“哥本哈根”①解释，特别是它的玻尔版本，人们必须假定测量装置是经典客体，且此假设给出了确定塌缩(测量)何时发生的范围。

然而，虽然这些区分仍然为测量和“经典的”提供了常用的实践刻画，但在大多数研究者看来它们要作为可能的基本理论的基础仍然不够清晰。

5.4.5 退相干

一种最近得到了充分发展的相关观点是，与“环境”的充分相互作用本质上有助于系统态的“塌缩”(“退相干”)。这里我们必须非常谨慎，因为还有一种充分研究的现象称为“退相干”，它涉及了系统与其环境间的相互作用。对于这一现象的物理重要性并没有争议。另一方面，许多人求助于这一现象以作为测量问题的某种求解。此论述尚需详细的审查。

5.4.5.1 退相干现象

5.4.5.1.a 定性描述 将一个物理系统从周围世界(其“环境”)中隔离开来是非常困难的。来自粒子“动物园”几乎每个角度的粒子都在四处乱撞，且它们中的许多甚至能够穿过非常强的壁垒(比如铅墙)。即使对于非常小的系统(如灰尘粒子)，也几乎不可能阻止它与环境的有效相互作用。

物理学家们对此相互作用提出了简单和非常复杂的两种模型，涵盖了从假

① 此“解释”——然而似乎并不是单个的统一观点——如此被称呼是因为它往往与尼尔斯·玻尔及其同事们相关，他们在哥本哈根工作。见[Cushing(库欣)，1994，Chapters. 6，7]和[Beller(贝莱尔)，1999]及其中的文献。

定系统与一个另外未指明热库的相互作用，到关于给定系统在给定类型环境中可能会经历的相互作用的速率、性质和强度的详细模型。有了这些模型，人们能够评估(在极少数、通常高度理想化的情形中，明确确定的)这些相互作用对系统态的影响。

注意这些相互作用将系统与环境纠缠。因而我们从一开始就不再谈论系统的态，而是谈论复合的“系统+环境”的态。当然，一般我们很少或不能够获得相关的环境自由度。(例如，想象一个光子被一个尘埃粒子反弹，之后被困在大气层中或更坏地被扔到太空中。在前一种情形中，在实践上重新获得该光子是不可能的；在后一情形中，则在原则上就是不可能的。)虽然这些自由度在某种意义上“编码”了系统的特定信息(如其位置，因为一般的相互作用是位置依赖的)，但我们总是不能在实践中获得这些信息，且有时在原则上也不能获得这些信息。

因此，虽然系统将开始与环境相纠缠，但我们一般仅能得到其约化态，得不到总的复合态。我们通过对环境自由度的求迹来得到此约化态。在关于此相互作用的许多模型中，得到的结果是系统的约化态近似是位置对角化的，也就是说其(混合)态看起来像是在空间中有确定位置的一个系统的态，指在混合态的每个分支(谱投影)在空间中有确定位置的意义上。在此意义上，退相干可以被看作是在“定域化”系统(参见第5.4.5.2.a节，在那里我们会明确为此混合是非正常的这一事实而犯难)。

此定域化最终来自于如下事实，即在这些模型中系统与环境间的相互作用是位置依赖的(环境与粒子的相互作用仅在粒子附近)。更一般地，若描述系统与其环境间能量交换的相互作用哈密顿量与系统的某可观测量$Q\otimes\mathbb{I}$相对易，则系统的约化态在由Q的本征矢量(或本征空间)挑选的基矢中变得近似对角化。环境被认为是“抑制了”非对角的“干涉”项。此外，因为相互作用哈密顿量与Q对易，已经处于Q本征态的系统常常不变，正如假定相互作用哈密顿量淹没(swamp)了系统自由哈密顿量的影响在物理上是合理的一样。

5.4.5.1.b　案例　退相干一个经典并广为研究的案例，可参见[Joos and Zeh(尤斯和泽)，1985]，涉及空气中的一个尘埃粒子。下面是他们论证的一个简化概要。

令 $|\psi\rangle$ 表示尘埃粒子的初态；令态 $|\psi_q\rangle$ 是粒子确定位置态的一个基矢（每一个粒子都处于用 q 标示的位置）；令 $|E_0\rangle$ 是环境的初态，且考虑粒子与处于一个 $|\psi_q\rangle$ 态的粒子环境中的单个空气分子间的相互作用：作为此单个相互作用的结果，我们将假定粒子 - 环境演化到态 $|\psi_q\rangle \otimes |E_q\rangle$。（特别地，这里我们假定尘埃粒子远重于空气分子，从而该相互作用本质上并不改变尘埃分子的态。）概略地说，尤斯和泽表明，若 $|\psi\rangle$（表示为一个波函数）初始时是高斯（并不必然是位置确定的）态，在 $|\psi_q\rangle$ 基矢中写作 $|\psi\rangle = \Sigma_q c_q |\psi_q\rangle$，则在一个这样的相互作用之后在 $|\psi_q\rangle$ 基矢（回顾第 1.2.3.4 节）中粒子的约化密度算符的矩阵表示将是：

$$W_{qq'} = c_q c_q^* \langle E_q | E_{q'} \rangle \tag{102}$$

其中，每当 q 与 q' 间的距离远大于尘埃粒子的波长时，$|\langle E_q | E_{q'} \rangle| \approx 0$。① 换言之，$W_{qq'}$ 中的非对角元（$q \neq q'$）项通过因子 $|\langle E_q | E_{q'} \rangle|$ 被约掉了。这里直观的是，因为态 $|E_q\rangle$ 和 $|E_{q'}\rangle$ 对应于空气分子从处于两个不同位置（用 q 和 q' 表示）的尘埃粒子处散射开的环境，若这些位置显著不同则环境的相应态也会“显著不同”（即接近于正交。当然，如果 $q = q'$ 则此内积为 1）。

之后尤斯和泽表明，对于环境的一系列不同模型而言，在许多这样的相互作用之后约化密度矩阵中的非对角元以指数衰减，其速率依赖于 $q - q'$（且当 $q = q'$ 时为 0）。速率非常快，根据尤斯[1986]的说法，甚至在高质量真空中半径为 10^{-5} 厘米的尘埃粒子的约化密度矩阵在大概 1 微秒内非常接近于位置②对角化。人们说客体系统（尘埃粒子）的态作为其与环境相互作用的结果而“退相干”了。

5.4.5.2　退相干与测量问题

5.4.5.2.a　退相干并不解决测量问题　退相干似乎将与环境相互作用的系统“定域化”了，因而我们最终或许不再需要担心遭遇到“模棱两可”于两个不同宏观态间的猫了（回顾第 5.1.1 节）。遗憾的是此结论中有一个问题，问题在于系统的

---

① 这里我们依据于在动量与波长之间的德布罗意关系。

② 在此情形中，“非常接近于位置对角化”意味着约化密度算符可以写作表示尘埃粒子局域化在约 10^{-13} 厘米内的态的和。

约化态是非正常混合(回顾第1.2.6.3.b节)的。我们来深入研究这一点。

一方面，我们必须承认在实践中不可能做一个实验以确定退相干的系统的约化态不是一个正常的混合态。要知晓原委，应该像在式(93)中那样考虑两个全关联粒子(类似于客体系统和其环境，它们由于相互作用变得关联起来)的简单情形。α 的约化态是 $W_\alpha = \sum_n (1/N) \mid a_n\rangle\langle a_n \mid$(类似于客体系统的约化态，它"接近位置对角化")。此态对 α 上的每个可观测量做出的预言与式(93)的态做出的相同。为了区分正常的和不正常的 W_α，我们因而需要测量复合系统(α 和 β)的某可观测量，且特别地我们需要测量 α 与 β 之间的关联(因为 α 和 β 并不处于一个纠缠态——因而 α 的态是一个正常混合态——当且仅当在它们之间没有非平凡关联时)。现在，在式(93)的情形中测量这些关联是相对直接的(假定我们有处于相同态的许多粒子对拷贝)，因为事实上它们在许多可观测量中是全关联的。但在这里感兴趣的情形中——一个客体系统及其环境——事情肯定更为复杂，因为关联仅存在于环境的少数粒子和客体系统之间。正如我上面提到的，一般在实践上将这些粒子从环境中重新得到是不可能的，而且即使我们能够这样做也并不是对它们的任意测量都能够完成此任务。事实上，若我们没有对总系统的许多拷贝(且一般而言我们确定没有!)，则所涉及的测量甚至变得更为复杂，因为我们必须测量一个甚至不是积(即具有形式 $F \otimes G$)的可观测量(类似于向方程式(93)中态的投影)。

另一方面，对环境进行这样一个测量在实践上的不可能性本身也不允许假设退相干系统的混合态是(或可以将它当作是)正常的。如前所述，问题还在本征态—本征值联系上。复合系统真正的量子力学态是一个纠缠态，按照本征态—本征值联系，(一般而言)在此态中客体系统并没有一个明确的位置。系统的态"也可能"处于一个正常混合态这一假设事实上与上述陈述不一致。换言之，虽然两个态在观测上很难以区分，但它们(给定本征态—本征值联系)在解释上是不一致的。

当然，人们能够放弃本征态—本征值联系，而且许多解释已经这样做了。我们将简短地讨论其中的一些。这里的要点是，在本征态—本征值联系存在时退相干并不解决测量问题。

5.4.5.2.b 退相干有助于测量问题 然而，退相干有助于测量问题的最普遍形式。回顾(第5.3.2节)测量问题常常并不局限于被看作是测量的相互作

用。给定本征态—本征值联系，在量子与宏观系统间的许多相互作用将宏观系统置于这样的态，该态将违背日常经验中通常所设定的判断方式。换言之，在其最普遍的形式下，“测量”问题仅仅是在日常观察的基矢下，量子理论显然不能够赋予似乎拥有确定值的可观测量以确定值的问题。在一个典型测量的特殊情形中我们或许已经处于求解的半途中了，因为在此情形中装置至少已经处于“期望的”混合态，虽然是不正常的混合态。

但上面提到的更一般的相互作用又如何呢，它不是明确的测量但仍然产生宏观客体的“糟糕”态？退相干有希望得到如下结果：对于“相对大”的系统(例如至少大如一个尘埃粒子)，与环境的相互作用意味着系统的态将变为一个“期望的”混合态(或某种非常接近于此混合态的态——如此接近以至于我们乐于忽略其不同)，虽然是不正常的混合态。但是，即便如此，这一策略仅仅是将普遍问题简化为一个仍然没有满意求解的问题，虽然它给出了许多关注明确地针对测量讨论的理由(因为退相干显然将普遍情形减化为一种至少在形式上类似于测量情形。)

5.5　解释

5.5.1　贯彻此计划的方式

我已经讨论了产生于量子理论的一系列哲学与基础问题，且量子理论的“解释”应当以某种方式或其他方式处理所有这些问题——包括提供对理论经验内容的一种说明和提供对不相容性(进而是“不确定关系”)的一种理解，以及提供对理论中定域性明显失效的某种理解(第6节)。然而，绝大多数解释主要致力于解决测量问题。正如我们在上面看到的，该问题可以被描述为(在其他方式中)下述内容间的冲突：(i)对熟悉的物理客体物理性质的一般理解，包括关于它们的有效推论；(ii)可观测量何时在经验上确定的最简说明在(iii)量子理论背景中。以这一思路来构想会发现导致冲突的是三种因素，因而回避此冲突也有三种一般性的策略：

1. 否认对物理世界和关于物理世界做出的推论的普遍理解总是真实的或是有效的。例如，与表象相反，人们可以否认测量结束时的

“指针”有一个确定(或接近确定)的位置。下文我们将遇到其他种“否认常识”的方式。我们称这些理论为“非常识解释”。

2. 补充或取代对可观测量何时在经验上是确定的最简说明。在某些情形中,理论中增加了附加的“隐”变量。在所有这些情形中,系统被认为拥有标准量子理论所赋予之外的特征,因而所有这样的解释常常被称为“隐变量解释”,他们承认对概念进行了扩展。(因为正如在理论中增加新的物理变量——通常还有它们的动力学——等于提出了一个新的理论,这些解释常常被称为“隐变量理论”。我习惯坚持用术语“解释”,虽然我这样做并不意味着我要做出特别的实质性论述,并且我偶尔也遵从习惯采用术语“理论”。)

3. 迄今为止用某些附加的物理学来补充(或说改变)现在的标准量子理论。这样的解释增加了一条规则(通常被设想为表达了某种动力学定律),该规则使得系统拥有一种我们朴实地认为它们所拥有的特征,或接近于该特征。因为所有这类规则的结果是态从值的叠加“塌缩”到单个的值,所以这些解释常常被称为“塌缩解释”。(同样,它们有时也被称为“塌缩理论”,我偶尔也会用到该术语。)

当然,这些策略也可以进行组合,但一般而言解释用这些策略中的一个来描述它们自身,虽然另外的一个或两个作为附加的结果也会发生。我将相应地在三个标题下面讨论这些解释。

这里没有空间考虑每种解释中的所有或是大部分内容。因而我不得不局限于对三种解释中每一种的两个重要代表给出简短说明。读者应该牢记我在此处所讨论的解释(甚至这里没有讨论到的解释)都存在不同的版本。在大多数情形中,我选择描述那个问题最少因而最打动我的版本,但大多数时候这一表述也是有争议的。

我对这些解释的讨论必然是简短的。我将给出比之前稍多的参考文献,以使感兴趣的读者在文献上有所涉猎。

5.5.2 “非常识”解释

5.5.2.1 量子逻辑解释 我已经提到了(第2.1.1节)希尔伯特空间的子空间

格——关于物理系统的"命题"格——初看起来能够从逻辑上进行解释。量子逻辑解释看重于这一理念，认为量子理论要求从经典到量子逻辑的革命性改变。①

刻画经典与量子逻辑间差别的一种方式是采用分配律(第7.4.3节)的失效——否则的话经典命题格和量子命题格在结构上(逻辑上)是相同的，因而量子逻辑解释的基本理念(特别是经典分配律)失效了。这种失效性被视为允许人们做出如朴素实在论(第5.4.1节)一直都希望做出的那类断言。例如，令 F 和 G 是没有共同本征矢量的两个不相容极大可观测量。它们的本征空间记为 $\{F_n\}$ 和 $\{G_m\}$。作为命题(逻辑上解释为希尔伯特空间中子空间格的要素)，它们断言一个系统拥有相应的本征值作为可观测量的值。则：

$$\{\bigvee_n F_n\} \wedge \{\bigvee_m G_m\} = \mathbb{I} \wedge \mathbb{I} = \mathbb{I} \tag{103}$$

其中$\mathbb{I}$是逻辑真命题。若我们将此合取的前一半解读为断言了可观测量 F 拥有某个值(正如人们常常那样用析取一词来理解存在量词)，且合取的后一半也类似，则式(103)断言了 F 拥有一个值且 G 拥有一个值。确实，此断言的逻辑为真。注意，如果我们能够将分配律应用于式(103)，则我们能够很快将其变为一个逻辑矛盾，因为对于任意的 n 和 m 有 $F_n \wedge G_m = 0$，它是逻辑假命题。

此外，对于任意数目的可观测量，这种析取的类似合取($\wedge$)，相应的论述也是真的。因此(用合取来理解全称量词)量子逻辑解释宣称恢复了所有可观测量(对于给定系统)总是有一个值(对于此系统)的这种理念。果真如此的话，则测量问题将不再是一个问题。当然，此方法同样也引起了一些问题。在质疑逻辑是否是完全可修正的之外，有人论证说量子逻辑解释只是将迷雾从一个地方(量子理论)转移到了另一个地方(逻辑)。无论如何，很明显量子逻辑解释需要做出一些关于为什么经典逻辑事实上在某个范围内有效的论证(这里退相干可能有帮助)。它同样需要解释数学中经典推理的成功。这两个任务等于是

① 伯克霍夫和冯·诺伊曼[1936]属于某种类似量子逻辑解释的最早拥护者。此领域随后的工作建立在他们所做工作的基础上。量子逻辑解释的拥护者们有：芬克尔斯坦(Finklestein)[1962；1969]、普特南(Putnam)[1969]、弗里德曼和普特南(Friedman and Putnam)[1978]以及巴布[1974]。同样见论文集[Hooker(胡克)，1975；1979]。达米特(Dummett)[1976]和吉本斯(Gibbons)[1987]给出了一些著名的批评。迪克森[2001]提到了许多标准的反对意见，这些异议偶尔也出现在上面列出的拥护者们的工作中。

量子逻辑解释中的未解问题[Dickson, 2001]。

5.5.2.2 多事物解释 我们的第二个“非常识”解释案例是“多事物”解释。这些解释通常包含在标题“多世界”解释中，但在这里有效的“世界”概念(就它本身而言是非常有问题的)与对该术语通常的理解相去甚远，因而使用一个抽象一些的词似乎是恰当的。①

无论如何，与量子逻辑解释否认(或至多是再解释)关于物理客体特性显然有效推论的逻辑有效性相反，多事物解释否认(或至多是再解释)关于物理客体特性显然为真的陈述的真实性。

再次考虑式(98)的右边。关于此态之前提出的问题是，在标准解释(采用本征态—本征值联系)下，处于此态的(复合)系统对于可观测量 $F\otimes Q$ 没有具体的值(其中 Q 是指针可观测量)。多事物解释走向相反的极端，宣称式子右边的每一项都对应于某个“真实的”现实。仪器拥有指针可观测量的值既非 0，也非 1，而是所有这些值。

很快这些解释遇到了一些问题。当然，显然的问题是系统如何能够“拥有”多重态，它似乎卷入了直接的冲突。一般地，对此的回应是引入解决该冲突的一个索引特征——例如，指针可观测量的不同值是相对于索引特征的不同值而实现的，索引特征的不同值对应于不同的事物(“世界”，或在某些版本中是“心灵”)。

然而，埃弗雷特(Everett)原始观点的特征似乎在本质上是相关联的。如式(26)中那样考虑一个一般的双系统态。若第一个系统在某种意义上被说成是“真的拥有”态 $|x_j\rangle$，则相对于 $|x_j\rangle$ 第二个系统可以说成是处于态 $|\psi_{相对于x_j}\rangle = K_j\sum_m c_{jm}|y_m\rangle$，其中 K_j 是归一化常数。作为对第二个系统测量的结果，由 $|\psi_{相对于x_j}\rangle$ 产生的概率恰好是那些由原始复合态产生的概率(因为测量局限于第二个系统)。埃弗雷特的原始理念似乎是，系统拥有的态仅是“相对于”其他系统的态而言的。其他人似乎采用某种类似于下述观点的理念，即第一个系统在不同的

① 事实上，我们这里谈论的是一类相当宽泛的解释。远不清楚此类观点的创始者埃弗雷特[1957]的意图是是否引入多世界的观念，多世界似乎是由德维特(Dewitt)引入的，见他和埃弗雷特在[Dewitt and Graham, 1973]中的文章。“多心灵”变种是由阿尔伯特与洛尔(Albert and Loewer)[1988]和唐纳德(Donald)[1990]等人提出的。更精细的分类和更多的文献见巴特菲尔德[1995]的评论和巴雷特(Barrett)[1999]的著作。

“世界”中拥有不同的$|x_j\rangle$，且在那些世界中第二个系统拥有相应的相关态。

这些观点中的每一个都面临着难解的问题。在埃弗雷特的相关联的观点中，需解决的问题更多。考虑时空理论中的类似情形，那时(20世纪以前)相对论者主张唯一的(时空)实体是相对的，但唯一可用的理论(牛顿经典力学)却明显不是相对论的。① 埃弗雷特的相关联的观点似乎处于某种类似的情形。(例如，复合系统本身的量子态如何被赋值，在什么样“角度”上被赋值？且它如何从相对论上来理解？)此外，多事物观点在使得“某事物”(世界、心灵等)的多元性具有某种意义这一点上面临明显的形而上学障碍。

这里我将只指出一个附加的问题，这也是常常被提起的。这些解释的普遍规定是允许叠加中的每一项在刚才描述的意义上对应于一个“实体”。但回想一下给定的态可以用许多方式来分解，且一般依赖于分解而处于叠加中的项是不同的。哪一种分解是“正确的”？

对此问题的一种回答将假设或力主一个优先基，用它来进行分解。对优先基的论证常常建立在退相干(见第5.4.5节)选出它的基础之上，但这里有一个严重的问题，即退相干通常产生的位置的近对角化是否是充分的。② 无论如何，该回答是极其有问题的。因为假定我们有好的理由相信存在一个优先基，它的元素表征了所有客体真正的物理特性。那样的话，我们将几乎不需要奢侈的多事物解释来求解测量问题。在优先基基础上我们只能断言系统总是用此基矢来描述的，从而在此基矢挑选出的特征上来定义概率则不会有障碍(例如，不会有科亨—斯佩克冲突。事实上，下文中第5.5.3.1节我们将考虑一种能够精确地用这些术语来刻画的解释——德布罗意—玻姆理论。)

另一个回答允许所有的分解都对应于实体。这里，退相干“可以”是有用的，因为它指出像我们这样的生物，也就是以一种我们事实上所用的特定方式与它们的环境相互作用的生物实际上也服从退相干，从而我们的感官会与我们

① 例如历史细节见[Brown and Pooley，2002]，对关系主义的进一步讨论见本书中巴特菲尔德撰写的第一章第2.3节。

② 同样值得注意的是，处于接近位置对角化的态并不意味着在其中它是对角化的基矢是接近于位置的某种事物。见[Bacciagaluppi(巴奇伽鲁皮)，2000]。在这些情形中，我们说退相干“挑选出”哪个基矢呢？

所感知到的客体的“正确”态相关联，且事实上也是这样，作为知觉者只要我们能够(在高的概率下)说明世界中的特性正是我们通常相信客体所拥有的那种。当然，仍然需要说明，事实上我们中的每一个都存在许多“拷贝”(索引的每个值都对应着一个拷贝)，但我们把这个问题留给这些解释的拥护者们，他们会以自己的方式赋予此种多元性以意义。

5.5.3 “隐变量”解释

在所有隐变量解释背后的基本理念是量子理论(正如 EPR 论证的！见第 4.4 节)不完备。这些解释提出要给量子理论赋予系统的态补充一个附加的、“隐藏的”态。(该变量不必在任何意义上都无法观测；它们仅在量子理论的眼中是“隐藏的”。)当然，不是所有的方案都能够做到这一点，特别是这样的理论必须重现量子理论的经验成功。这样做相当于重现作为对隐藏态平均的量子概率，这种方式与人们希望从经典统计力学通过对系统的微观态取平均重建经典热动力学的方式相同，见本书第九章。

因此，例如用 λ 标示隐态(其中出于简单起见我们假定 $\lambda \in \mathbb{R}$)。令 $\rho w(\ )$ 是隐态在(量子)态 W 中的分布。最后，令 $\Pr_{\lambda,W}(F=f_k)$ 是隐态 λ 赋予对处于态 W 的系统的 F 测量得到(本征值)f_k 的概率(在决定论的情形中，所有这样的概率当然是 0 或 1)。则经验充分性要求：

$$\int_{\mathbb{R}} \rho W(\lambda) \Pr_{\lambda,W}(F=f_k)\, d\lambda = \mathrm{Tr}[WP_{f_k}] \tag{104}$$

其中 P_{f_k} 是对应于本征值 f_k 的 F 的本征空间。

在此小节中，我将考虑一种决定论的隐变量理论——德布罗意—玻姆理论(第 5.5.3.1 节)和模态解释——一般是非决定论的。

5.5.3.1 德布罗意—玻姆理论 德布罗意—玻姆理论是一种粒子轨迹的决定论理论，某种程度上属于经典的牛顿(更准确说是哈密顿)动力学的模式。[①] 在其最初的形式中，隐变量理论是通过将薛定谔方程(哈密顿量中包含

---

① 其基本理念最初由德布罗意提出，见他最为著名的[de Broglie，1927]。随后(在德布罗意被说服认为该理论并不起作用后)由玻姆加以发展[Bohm，1952]，并在过去的几十年中得到了进一步的发展。关于该理论的不同种解释和观点，见[Cushing，Fine，and Goldstein(库欣、法因与戈德斯坦)，1996]。

势V)分离为实部和虚部而精确构造的，因而方程的解$\psi(x,\ t)$写作$\psi(x,\ t)=R(x,\ t)e^{iS(x,t)}$。虚部具有形式：

$$\frac{\partial S}{\partial t}+\frac{(\nabla S)^2}{2m}+V-\frac{1}{2m}\frac{\nabla^2 R}{R}=0 \tag{105}$$

此方程具有哈密顿方程的形式，势由$V+U$给出，其中U是“量子”势$-\frac{1}{2m}\frac{\nabla^2 R}{R}$。实部具有连续方程的形式，且隐含了概率守恒：

$$\frac{\partial p}{\partial t}+\nabla\cdot p\dot{q}=0 \tag{106}$$

其中$p(x,\ t)=R^2(x,\ t)=|\psi(x,\ t)|^2$被理解为(被约定为)在点$x$(时间$t$)处找到粒子的概率，且：

$$\dot{q}(x,\ t)=\nabla S(x,\ t)/m \tag{107}$$

被解释为粒子的速度(当它在x点处和t时刻时)。

因此，人们能够将单粒子薛定谔方程理解为描述了一个粒子系综(它的每一个“量子态”都相同)，(在空间中)分布满足$p(x,\ t)$。根据式(106)，此分布随时间守恒，因而可以恰当地理解为是概率分布。单个粒子的运动就好像它们是经典粒子那样，即本质上是由具有经典势加上附加的“量子势”的牛顿定律控制的。

从几个理由来看该理论很重要。这里我将提到其中三个理由。

首先，它表明本征态—本征值联系远非解释所必需的原理。事实上，本征态—本征值联系在该理论中被强烈地否定。位置可观测量甚至没有本征态，但是在该理论中每个粒子总是有一个确定的位置。

其次，它为经常提及的量子理论特性提供了一个很好的反例，即量子力学意味着在基本层次上缺乏决定性、基本决定论的不可能性、粒子确定轨迹的不可能性、描绘量子现象的不可能性等影响的反例。而该理论显然描述了具有确定轨迹的粒子，它们在势$V+U$下决定论地运动。

第三，此理论大概是发展得最为充分的解释，且有具体的应用。例如，它对粒子在自旋测量中的如何表现有令人信服的解释，参见[Dewdney, Holland, and Kyprianidis(杜特尼、霍兰德和凯普莱尼迪斯)，1986]。此外，该理论有时也用于处理标准观点难以对付的实际应用和问题。一个成功的例子是对粒子在隧穿一个势垒时在势垒中用多长时间的预言。在标准量子理论中确定这一时间

量(即便作最乐观的估计)在概念上很棘手，因为“粒子”没有确定的轨迹且量子理论中没有时间可观测量。但在德布罗意—玻姆理论中该问题在概念上很明确，参见[Leavens(莱文斯)，1990]，因为人们本质上能够展示在势垒(和量子势)影响下粒子的可能轨迹，进而简单地对每个轨迹下在势垒中所用的时间取平均。

然而，我们不应该忽略该理论面临的一些困难。将此基本理念推广到相对论量子场论中的(非常有意义的)困难先搁置在一旁(因为超出了此文章的范围)，它还面临着其他的困难。这里我谈及其中一个。

尽管初看起来该理论的本体论能够扩展一些想象的空间。但事实上我们面临下述困境。一方面，如果我们将“导向场”(量子势)看作是某种实的势场，它是高度非定域的。(在文献中，此要点有时是用它并不“存在”于$\mathbb{R}^3$而是存在于多粒子系统的位形空间中这一事实表述的。当然，对于单粒子系统而言它们是相同的。)此要点反映在多粒子“引导条件”式(107)的形式中：$\dot{q}_i(x_1, x_2, \cdots, x_n, t) = \nabla_i S(x_1, x_2, \cdots, x_n, t)/m_i$。即，粒子$i$的速度不只依赖于它的位置，同样也依赖于其他粒子的位置。最后，注意到该粒子并不对其范围内量子势的强度敏感，而是对其形式敏感。增加量子势的强度，即在式(105)中用某个大于1的常数乘R，完全不改变它对粒子的影响。对有些人而言，所有这些(或更多的)观测会累积起来得到如下结论：从乐观的角度看量子势需要得到严格的形而上学的阐释，从悲观的角度看量子势太异乎寻常以致很难得到支持。

另一方面，人们可能会否认其实在性。按照引导条件运动的一个粒子在$V=0$(“经典”势为0)下一般仍会偏离经典期望的轨迹，但我们是否要因此假定此“偏离”是由某个附加的非经典势的存在引起的？相反，或许我们会将引导条件看作是对牛顿定律的基本修正。① 在此情形中，我们可以避免量子势作为物理实体的怪异性，但理论变成完全约化论的——粒子唯一实在的特性是它们的时空轨迹。这样一个理论被迫回避提供人们所希望得到的解释(如对粒子间胶

① 确实，这里的基本理念是重建一条“惯性”轨迹的特有概念。见[Pitowsky(皮托斯基)，1991]。

合现象的解释)，而一般在标准量子理论中我们拥有这样的解释。[①]

5.5.3.2 "模态"解释 模态解释[②]在许多方面类似于德布罗意—玻姆解释：它们也假设隐变量(否认本征态—本征值联系)，并且它们也(至少，它们能够)为这些隐变量构造动力学。二者之间有两种主要的区别：(1)一般地，模态解释允许物理量对系统而言是"确定的"，这一点是态依赖的，因而它们能够随时间变化；(2)一般地(典型地作为它们选择哪个物理量是确定的结果)，模态解释是随机的。已经提出了许多版本的模态解释。这里我们集中于这些解释中发展得较好的一类("谱模态解释")，并简要提到一些更新颖的观念。

回顾每个密度算符都可以被唯一地分解为其谱投影的权重和。谱模态解释的核心论述是，在密度算符的谱投影必须拥有值(在此情形下是0或1)的限制下，具有一个确定值的可观测量的集合是能够被一致赋值的最大集合。换言之，将密度算符作为可观测量处理，并赋予它一个值，之后在不遇到科亨—斯佩克类型冲突的情况下尽可能地赋予其他量以值。在一些合理的假设下，已经证明了此观念对于任意给定的态(密度算符)而言产生了将被赋予一个确定值的唯一可观测量集合。[③] 因为态(当然)随时间变化，并且牢记我们在谈论约化密度算符，因此它并不需要幺正地变化，从而确定值的可观测量一般也随时间变化。

许多(虽然不是全部)模态理论学家们致力于为系统的确定特性定义某种动力学。该问题很复杂，因为"同时"发生了两种类型的动力学：系统确定值特性集合的(决定论的)变化，和从一个确定拥有的特性到另一个特性的(随机)转变。巴奇伽鲁皮和迪克森[1999]表明此问题有一种求解，他们还展示了某些案例。然而，正如在德布罗意—玻姆理论中，这样一种动力学能否是洛伦兹(Lorentz)不变的这一点也存在严重的问题。从非相对论到相对论情形的直接推

① 见[Bedard(贝达德)，1999]和[Dickson，2000]。

② 关于模态解释的一般主题有许多不同版本。该术语本身是由范·弗拉森(van Fraasen)[1972]用到的。在20世纪80年代，各种版本接踵而至，例如在科恩[1985]、狄克斯(Dieks)[1988]和希利(Healey)[1989]的关键性工作中，后来巴布[1997]提出了一种相关但有些不同的观点。有用的专著如弗马斯(Vermaas)[2000]，有用的论文集如狄克斯和弗马斯[1998]。

③ 该类型定理的案例见克里夫顿(Clifton)[1995]。巴布和克里夫顿[1996]证明了一个类似的定理，但是是针对巴布[1997]的解释而做的。

广已知是不可用的，但该方法“朴素的”相对论版本能否奏效仍然是一个开放的问题，此问题的求解与更普遍的问题相联系。如果是这样，模态解释的基本理念(“给定特定的限制，使尽可能多的可观测量成为确定的”)能否推广到量子场论。如果能推广的话，似乎最可能的领域将会是代数量子场论，因为模态解释的非相对论量子力学版本本身就是很好的代数表达。①

模态理论解决测量问题了吗？如果是的话，它们之所以能做到这一点是因为规定如何选择确定值的可观测量从而能够挑选出可观测量(比如指针可观测量)，正是这些可观测量的确定性受到了测量问题的威胁。德布罗意—玻姆理论，还有巴布[1997]的模态解释多多少少是用命令的方式来实现这一目的的。只要我们所相信的指针(和猫，等等)所拥有的特性最终能够被理解为随附于组成粒子的轨迹，则德布罗意—玻姆理论就能够给测量问题以令人信服的求解。通常模态解释也必须给出他们对此的论证方式。例如，约化密度算符谱投影的确定性是否足以保证我们认为的宏观客体所拥有特性的确定性？事实证明这一问题非常难以回答。在理想脉冲测量情形中，装置的约化密度算符确实拥有作为其谱投影的确定指针态。但在真实的(非理想)测量中会有潜在的麻烦。人们一度认为退相干挽救了局面，但目前退相干是否完成了这里的任务尚不明晰。②

5.5.4 塌缩解释

5.5.4.1 非动力学的塌缩解释 量子态必须在某种意义上“塌缩”的理念一旦用概率来解释则是非常自然的。且人们不时地提出关于此塌缩何时发生的各种提议。例如，[Dirac，1930]中说：

> 考虑一个对处于态 ψ 系统的观测，包含对可观测量 α 的测量。此观测后系统的态必须是 α 的一个本征态，因为在此态下对 α 测量的结

---

① 例如，见[Halvorson and Clifton，1999]。最初对量子场论模态解释的探索，见[Clifton，2000]。对此探索的讨论尤其是它与洛伦兹不变性的关联，见[Earman and Ruetsche(厄尔曼和鲁切)，2006]，其中也包括了关于模态解释中洛伦兹不变性早期工作的参考文献。也参见本书中霍尔沃森所撰写的第八章第5节。

② 前一种态度见[Bacciagaluppi and Hemmo(巴奇伽鲁皮和海莫)，1996]，后一种见[Bacciagaluppi，2000]。

果必须是确定的。

也就是说塌缩发生在观测(即测量)时。这里狄拉克给出的论证如下。若我们重复对系统中 α 的测量，我们将以概率 1 得到与第一次(与 ψ 无关，且当然假定测量是第一类测量——第 3.1.2 节)相同的结果。假定结果是 a。仅存在一个会赋予 a 以概率 1 的态，即对应于 a 的本征态。

我已经指出(见第 5.4.3 节)这种思路下的说明，往好说需要有补救性措施。[①] 否则的话，它只不过是一个关于塌缩假设的陈述，并没有明确说明是什么使得“测量”不同于其他相互作用。

从原则上做出此区分的困难在于，如果算作测量的相互作用太少的话，我们可能仍然会遭遇到测量问题(因为塌缩可能并不发生在需要用它来得到我们所感受其确定特性的确定性情形中)。然而，如果视为测量的相互作用太多的话，我们最后可能得出一个在经验上为假的理论。例如，若一个穿过双缝装置(图 3)的光子的态总是在其穿过栅栏之后塌缩，则我们将永远不会看到干涉图案，这与实验结果相反。

维格纳[1961]给出了做出此区分(在塌缩发生与没有发生之间)的一种方法，该区分按理说可以避免这两个缺陷并且像是理论的一种解释。维格纳受到了下述论证——常常称为“维格纳的朋友”论证——的启发。

维格纳要求我们想象下述情景。测量装置 μ 在客体系统 σ 上测量某可观测量 F(通过指针可观测量 M)。同时，维格纳(α)和他的朋友(β)准备好观测 μ(通过“测量可观测量”A 和 B)。让我们用 $|m_n\rangle$ 表示 M 的本征态，其他量也类似。[②] 现在，假定 μ 对 σ 的测量已经发生，而 α 和 β 都尚未观测 μ。若 σ 最初处于态 $\sum_n c_n\ |f_n\rangle$，则我们可以将总系统的态写作：

$$\sum_n c_n\,|f_n\rangle\,|m_n\rangle\,|a_0\rangle\,|b_0\rangle \tag{108}$$

---

① 这样的补救尝试很少见到，因此在这里我没有提供给读者的文献列表。或许使非动力学塌缩有意义的最著名尝试恰好是这里讨论的由维格纳给出的这种。

② 这里有一个严重的问题，即一个如“观测到指针处于态 M_n”的物理态是否是纯态，进而是是否存在像 A 和 B 那样的可观测量。见第 1.4.3 节。

根据本征态—本征值联系，在此态中μ和σ并不处于$M \otimes F$的确定态。到这一步标准量子理论会调用塌缩假设以使得对于某个k该态变为$|f_k\rangle|m_k\rangle|a_0\rangle|b_0\rangle$。也就是说在缺乏关于结果的具体知识时，该态变为(无知—可解释的)混合态:

$$\sum_n |c_n|^2 P_{f_n} \otimes P_{m_n} \otimes P_{a_0} \otimes P_{b_0} \tag{109}$$

然而，正如上面注意到的(第5.4.5.2.a节)，式(108)和式(109)在经验上确实不同，即使一般而言在实践中不可能发现该不同。维格纳并未在σ，μ的性质或二者间的相互作用中找到任何保证此态的物理改变发生这一假设的内容。

此外，注意到观测者α和β还未以一种实质性的方式进入视野。但我们(在此论证中)有责任去保证的正是他们的确定经验。维格纳将他自己看作α，且注意到了在此过程的这一阶段，即一旦态由方程式(108)给出，没有什么会威胁到他经验的确定性——事实上，他仍然处于“还没有以概率1观测到”态($|a_0\rangle$)的状态。因而考虑在α观测到仪器之后的态:

$$\sum^n c_n |f_n\rangle|m_n\rangle|a_n\rangle|b_0\rangle \tag{110}$$

现在维格纳感觉到不得不赞同塌缩，因为不是这样的话(根据本征态—本征值联系)他自己的态将会是不确定的。

但维格纳的朋友(β)又如何呢？假定在α之前β观测到仪器。维格纳论证道，给定β对先前心智状态报告的宽容原则，他有义务在β观测的时候将态塌缩，即使他(α)还未观测到μ。因为假定之后(在β后)观测到μ，再然后问β:在你观测到μ之后，你是否感觉到你经历到了确定的结果？你是否观测到它处于一个确定的态？当然β将回答“是”，且假定(正如维格纳那样，此假设是宽容原则)我们要相信β的报告，则我们必须假定该态在β的观测中塌缩了。

因此，任意被应用了宽容原则的观测者将在观测中引起态的塌缩。维格纳相信该原则可以推广到一个“意识”(这里该术语受到了很大的限制)或是拥有“意识”的事物。此外，给定在物理形体和意识之间的明显区分(即给定某种形式的二元论)，人们可以将意识的特异性看作是意识的观测引起态塌缩的原因，而非意识的(例如典型的测量仪器)观测并不引起态的塌缩。

这里暂且不考虑维格纳论证中假设的合理性问题，注意到我们是迂回得到

这一结论的。换言之，维格纳并未给予我们一个来自于意识本性并用此本性来论证意识的观测将塌缩物理态的论证。事实上，维格纳的观点面临着所有二元论面临的同样困难：什么是意识与物质之间的联系？意识是如何使一个系统的物理态塌缩的？对此目前已经提出了许多特设性观点，但没有哪个是特别令人信服的。评论见[Atmanspacher(安特曼斯潘彻)，2004]。

5.5.4.2 动力学约化理论 对塌缩严格定义的另一种策略是将其描述为一个物理过程。当然，在某种意义上它已经被描绘为一个物理过程，因为它是系统物理态的变化。但在这里我所指的内容更多，即要构造一个单独的运动方程，它以某种方式同时统一地包含了连续的“薛定谔”演化和塌缩。沿此方向已经提出了(仍然会提出)许多方案。[①] 这里我们再次只考虑一个案例。

5.5.4.2.a 连续自发定域化的直观说明 这里我考虑的案例是最初由吉拉迪、瑞米尼和韦伯(Ghirardi，Rimini and Weber)[1986]所提出理论的后续版本。该后续版本称为“连续自发定域化”(CSL)，是由佩尔(Pearle)[1989]提出的。用波函数来叙述最为简单。其基本理念是，宇宙中的每个物理系统都同时经历两种类型的演化：“规范的”决定论的“薛定谔演化”和朝向定域化的随机趋势。对于后者，我们指的是系统波函数(在位形空间中)经历的随机(但无穷小)涨落，其净结果是按平均值和以压倒性的概率使波函数更定域化——更多的概率集中于某定域化区域。

此随机演化的强度——它决定了定域化发生有多快——是由系统中的粒子数决定的。[②] 对于拥有微观数目粒子的系统，演化中的随机部分被薛定谔演化所淹没，从而对于这样的系统该演化几乎等同于由标准量子理论给出的演化。但对于拥有宏观数目(约 10^{23}个)粒子的系统，若系统最初处于两个(或更多的)定域态的叠加，这些定域态的定域中心互相远离($>10^{-5}$cm 或类似)，则演化的随机要素作用得非常快，从而(连续地)抑制了除一个项之外的所有叠加项。此随机过程的参数之所以如此是为了使约化到一个态或其他态的概率等于赋予

① 少量的案例，见[Diosi(迪奥西)，1992]、[Ghirardi，Rimini and Weber，1986]、[Gisin(吉辛)，1984]以及[Primas(普利马斯)，1990]。

② 存在一些其他方案，在其中随机演化的强度依赖于其他事物，例如质量。

该态的量子概率(通过最初的叠加给出)。

5.5.4.2.b 连续自发定域化的数学说明 这里给出一些数学细节。在CSL中，波函数(这里写作一维空间中的)的演化是由(非幺正的)[①]算符给出的：

$$U(t) = \exp[-iHt]\exp\left[-\frac{\gamma}{2}t\int N^2(x)\,dx\right]\exp\left[\int N(x)B_x(t)\,dx\right] \quad (111)$$

第一项只是通常的量子力学演化算符(对不含时的哈密顿量 H 而言)。$N(x)$ 是一族"数字密度"算符，它直观地表示了在以 x 为中心的某个区域内的粒子数。或在三维空间中，围绕 x 的某个确定半径的球——这一确定半径是 $N(x)$ 定义的一部分，这里我们不考虑它。$B_x(t)$ 是随机过程场(即随机过程的连续族，每个过程都在点 x 处演化)。因此式(111)中的演化算符是"随机的"——它依赖于 $B_x(t)$ 的取值。注意，虽然对于 $B_x(t)$ 的某个给定赋值人们能够将赋值代入式(111)中，因而它又将是决定论的。

式(111)中的第二项是一个"衰减"项，衰减的速率依赖于 γ(和粒子密度)。式(111)中的第三项是"增强"项，其中增强依赖于随机过程 $B_x(t)$(和粒子密度)的演化。这两项在相互"对抗"，除了在某个定域化区域它增强以外，其最终的结果(在相当高的概率下)是波函数的总衰减。

$B_x(t)$ 的定义使得在给定区域内的增强与在该区域内波函数的振幅直接成正比，因此增强项最终引起波函数在给定区域内增强的概率与在该区域内塌缩发生的量子力学概率相同。注意到随着波函数塌缩到某个区域，它继续塌缩到该区域的概率在增加，因为波函数的振幅在那里会更大，因而在该区域中的随机过程也很可能会增强。

最核心的"技巧"是选择 γ 以使得前一节的陈述为真，即对于拥有小数目粒子的系统，式(111)中的第一项支配着其他两项，因为 $N(x)$ 在各处都会很小，而对于拥有大数目粒子的系统而言，第二项和第三项会占主要地位。关于 γ 的选择会有经验限制，但事实上有可能找到与已知实验事实相一致的 γ 值。

5.5.4.2.c 评价 CSL明显的优点在于它是严格定义的且数学上精确的，

---

① 一般地，波函数的模在CSL中的演化是不守恒的。然而，在任何时候重整化波函数都很容易。

理论能够以一种清晰的方式将“塌缩”描述为一个物理过程。然而，对它也受到了一些质疑，这里我提及两个。

第一个问题相关于该理论面临的或许是最大的理论困难，即相对论版本的构造。虽然在许多不同方向上取得了一些进步，如吉拉迪、格拉西和佩尔(Ghirardi，Grassi and Pearle)[1990]，但仍然存在着严重的问题。此外，这些问题与CSL的标志性特征直接相关，在相对论背景下引入随机过程$B_x(t)$族会引起在标准理论中并不存在的无穷大(且不能够以相同的方式被“重整化”掉)。

第二个问题与CSL中的塌缩从未完成这一事实有关。即该态永远不会成为其支集完全包含于定域化区域内的态。相反，它总是有“尾巴”——波函数在系统被假定的定域化范围外的点处总有非零的幅度。

这里的问题在于人们是否准许称这样一个系统为“定域化的”。若我们保留本征态—本征值联系，则答案可能仍是“否”。毕竟，一个有尾巴的波函数不是位置可观测量的本征态，也不是它的任意粗粒化。事实上，至少某些CSL的拥护者的回应是给出一种对波函数的新理解，正如贝尔[1990]提出的理解为是某种对“密度之类”的直接表示。虽然这一观点避免了直接的问题，但在形而上学上有一些烦人的结论，比如显然在此观点下会在许多场合中存在所有客体的“模糊”(低密度)“拷贝”。

6. 非定域性

回顾第1.2.6.4节，处于非可分解化态的复合系统可能展示出它们所显示(至少是在测量中)特征间的关联。这些关联可能是“非定域的”。首先，非定域性涉及粒子(和在其中展示出关联特征的测量事件)的空间分离。这样的话，会有空间上分离的系统所显示的关联能否用一个“共因”来解释这一问题，其中共因是能够解释两个系统间关联的它们历史中的单个事件。

在6.1节中我将澄清这一问题(见第6.1.1节)，之后考虑一些定理，这些定理在合理的假设下表明了事实上没有这样的共因解释可以采用(见第6.1.2—6.1.3节)。因此量子理论(事实上还有物理世界本身!)在某种意义上表现为“非定域的”。在6.2节中我将考虑一些对这些定理的反响，还有找到量子理论

的一种相对论不变且概念上令人满意的解释前景的意义。

6.1 不可行定理

6.1.1 非定域关联

6.1.1.1 统计关联 术语“关联”来自于统计学，这里也是在该意义上使用的。给定两个随机变量 A 和 B，它们的关联定义为：

$$r_{AB}=\frac{E[(A-\bar{A})(B-\bar{B})]}{\sigma(A)\sigma(B)} \tag{112}$$

其中，$E[\cdot]$是期望值，$\sigma(\cdot)$是标准方差。如果对正的 k(和任意确定的 m)有 $A=kB+m$，则关联为 1，k 为负时关联为 -1。式(112)中的分子称为“协方差”。这里的分母本质上是为了满足归一化。

应当明了的是关联是对一个变量的值对另一变量值“依赖性”的测度。非零的关联一般地是由如下事实引起的：至少对于随机变量某些可能的值 a 和 b 有 $\Pr(A=a\mid B=b)\neq\Pr(A=a)$，即 A 与 B 在统计上不是独立的。

6.1.1.2 单重(Singlet)态关联 一对自旋 1/2 粒子的态：

$$|\psi_{\text{singlet}}\rangle=\frac{1}{\sqrt{2}}(|z_+\rangle|z_-\rangle-|z_-\rangle|z_+\rangle) \tag{113}$$

称为“单重”态。注意它是方程(93)中的态的双粒子类比。考虑两个粒子中每一个的可观测量 S_z。可以把这些可观测量看作是随机变量，因为式(113)中的态产生关于这些可观测量可能值的概率，或更精确地产生可观测量 $S_z^{(1)}\otimes\mathbb{I}^{(2)}$ 和 $\mathbb{I}^{(1)}\otimes S_z^{(2)}$ 可能值的概率，其中上标是粒子的编号。在上述态 $|\psi\rangle$中这些可观测量的协方差(为了避开计算标准方差将这些可观测量归一化以得到本征值 ±1)是：

$$\begin{aligned}r_{S_z^{(1)}S_z^{(2)}}&=\langle\psi|(S_z^{(1)}\otimes S_z^{(2)})|\psi\rangle\cdots\cdots\\&=\frac{1}{2}(\langle z_+|\langle z_-|-\langle z_-|\langle z_+|)((-1)|z_+\rangle|z_-\rangle-(-1)|z_-\rangle|z_+\rangle)\\&=-1\end{aligned} \tag{114}$$

它们的值是完全反关联的。确实，对 $S_u^{(1)}\otimes S_u^{(2)}$(对任意的方向 u)的类似计算表明 $r_{S_u^{(1)}S_u^{(2)}}=-1$，完全反关联在每个方向上都成立。对不同的方向 u 和 u'——为

不失普遍性给定态的球对称性，设定式(36)中的 $\phi=\phi'=0$ 且再次“归一化”可观测量——我们计算得：

$$r_{S_{u}^{(1)}S_{u'}^{(2)}}=-\cos\theta\cos\theta'-\sin\theta\sin\theta'=-\cos(\theta-\theta') \tag{115}$$

这一关联是量子理论所预言的，也被一系列实验检验所确证。

6.1.1.3 关联的共因解释 当然在某种意义上非定域的关联无处不在。考虑：在地球上不同方位的潮汐之间存在关联；在澳大利亚的悉尼与英国的伦敦日间的平均气温之间存在(负)关联；在每天早晨我家前门口出现的报纸与我邻居家前门口出现的报纸上的字数之间存在(正的，近乎完全)关联。这些关联都不足为奇，即使它们是在有空间上分离性质的对象之间的关联。原因很明显：这些“非定域的”关联用共因可以给出一个定域的解释。

按照[Reichenbach，1956，pp. 158—159]，可以用条件概率将共因概念形式化。假定 A 与 B 是关联的。在此情形下，$\Pr(A\wedge B)\neq\Pr(A)\Pr(B)$。事件(比如说，同时发生的)$A$ 与 B 的共因是事件 C，满足：(i) $\Pr(A\mid C)>\Pr(A\mid\neg C)$；(ii) $\Pr(B\mid C)>\Pr(B\mid\neg C)$；(iii) $\Pr(A\wedge B\mid C)=\Pr(A\mid C)\Pr(B\mid C)$；(iv) $\Pr(A\wedge B\mid\neg C)=\Pr(A\mid\neg C)\Pr(B\mid\neg C)$。条件(i)与(ii)说的是 C 在概率上分别与 A 和 B 的发生相关，而条件(iii)与(iv)表示 C“将 A 与 B 隔开”(反之亦然)——即 C 完全解释了 A 与 B 之间的关联。①

因此，例如我们可能注意到我家前门的报纸(A)与我邻居家前门的报纸(B)是由同一版相(C)印刷的；朝向太阳倾斜(C)的同一地球部分地决定了悉尼的气温(A)，同样也部分地决定了伦敦的气温(B)；等等。

正是由于爱因斯坦的广义相对论理论，甚至月球对海洋的引力也是定域的共因：引力“作用力”就像波一样在空间中定域地传播。事实上物理学至少在20世纪都一直在朝着定域理论前进，且事实上理论在某种恰当的意义上是洛伦兹不变的。(“洛伦兹不变”的精确意义是多变的，但大体意思是理论不允许物质和能量的传递快于光速或不允许发出比光快的信号，或最终表述为不允许物质从亚光速到超光速的加速。参考本书第三章第2节。)

① 此构造不能完全令人满意，但对当前目的而言展示了足够好的共因概念。见[Uffink，1999b]和其中的文献。

量子理论允许非定域的关联。我们面临的问题是，能否像这里提到的其他关联一样给这些关联以一个(定域的)共因解释。特别是，式(115)所蕴含的关联能否用共因来解释?

6.1.2 贝尔定理

本质上贝尔定理对该问题的回答是“否”。此外，贝尔对定域理论的预言给出了某些限制。实验表明这些限制并不成立。

6.1.2.1 实验 贝尔考虑的实验本质上与EPR(第4.4节)考虑的是同一个，不同的是他采用了玻姆[1951，Chapter 22]描述的版本。有一个自旋1/2粒子对的源，每个粒子对都处于态$|\psi_{\text{singlet}}\rangle$，粒子被送到斯特恩—盖拉赫装置中，每一个都朝着某个方向(用u和u'来标示)。粒子以下述方式到达这些装置，即测量是在类空分离下做出的。事实上即使对u和u'方向的选择也是在类空分离下做出的。也就是说测量的选择和测量本身都是在空间上充分远、在时间上充分近的地方做出的，以使得任何种类的信号从一个测量事件到另一个的传输必须快于光速。在多次重复实验之后整理实验结果并确定其关联。

事实上为了展开贝尔定理我们只需在每一端考虑三个可能的方向(u_a，u_b和u_c)即可。因而测量装置是在“最后时刻”被选择的，且测量的正是该装置的三个方向，这些装置记录了测量结果。

6.1.2.2 问题的“隐变量”表述 关于我们上面提到的共因问题的一种思考方式是采用所谓的“隐变量理论”术语(第5.5.3节)，且实际中贝尔也是采用这些术语来构想该问题的。回想隐变量描述了粒子的完备态，该态包含量子态并不提供的信息。在决定论情形下，这些完备态为每个粒子的每一个u值都确定了值S_u。从而粒子源处的情形是，将要发射向两个测量装置的粒子对有某个“隐”态。该态已经确定了每个粒子在每个可能方向u_a，u_b和u_c上的自旋值($+1$或-1)。

事实上，回顾EPR论证(第4.4节)与它们关于量子力学“不完备”的结论。那里的要点是定域性(加上物理实在标准)迫使人们得出，在该实验的情况下每个系统必须已经以某种方式拥有了所有S_u的值——它们必须在粒子源处“建立起”正确的关联(因为在类空分离下由于定域性它们不可能在这之后这样做)。

注意到因为完全反关联，对于给定的方向而言隐态必须决定粒子对拥有相

反的自旋。(因而这些隐态能够分为八类，在三个方向中的每一个上有两个可能的自旋赋值。)除此之外，唯一的要求是如式(104)那样有可能通过对隐态的某些分布取平均而恢复量子理论的关联。

6.1.2.3 概率的隐态 更普遍的方法是允许隐态仅是概率地确定自旋。只要粒子源处粒子的隐态将一端的测量事件与另一端的结果分隔开来(第6.1.1.3节)，理论仍然是“定域的”。令 $\Pr_\lambda$ 为用隐态 λ 表示的测量结果的概率(现在我们假定量子态固定为 $|\psi_{\text{singlet}}\rangle$)，则隔离条件如下。对于任意的 u，u'，$k=\pm1$ 和 $k'=\pm1$：

$$\Pr_\lambda(x=k\,|\,i=u,j=u',y=k')=\Pr_\lambda(x=k\,|\,i=u) \tag{116}$$

其中 x 是粒子1的结果，y 是粒子2的结果，i 是粒子1被测量的自旋方向，j 是粒子2被测量的自旋方向。这里直观的想法是，粒子1的结果仅依赖于被测自旋的方向(和 λ)，因而独立于粒子2的结果和粒子2被测量的自旋方向。这一条件或某种类似的条件常常称为“贝尔定域性”。

所以我们最终面临的问题是，是否会存在隐态和在其上的分布 $\rho(\,)$ 使得概率 $\Pr_\lambda$ 同时满足贝尔定域性式(116)并通过如式(104)中的平均再现量子关联式(115)。注意类似的条件对于单翼(single-wing)概率也成立，即那些仅对两个粒子中的一个进行测量得到的概率，这些概率也当然是通过对隐态取平均得到的。

一般而言正如我早先提到的(第5.5.3节)，分布 $\rho(\,)$ 依赖于量子态(回想我们这里假定的是 $|\psi_{\text{singlet}}\rangle$)，但它必定不能依赖于对粒子源有非定域影响的事物，因为 $\rho(\,)$ 被认为表示了在粒子源处产生的粒子对的隐态分布。例如，我们将假定测量方向的选择和测量d结果并不影响 $\rho(\,)$。

6.1.2.4 贝尔定理 贝尔定理表明事实上不存在上面描述的共因解释。

定理(贝尔[1964])：对于任意选择的方向 u_a，u_b 和 u_c，任何满足(i)贝尔定域性(116)和(ii)$\rho(\,)$ 独立于测量事件的隐变量理论将满足不等式：

$$|r_{ac}-r_{bc}|-r_{ab}\leqslant 1 \tag{117}$$

其中 $r_{ab}:=r_{S^{(1)}_{u_a}S^{(2)}_{u_b}}$，类似等等①

① 之后得到了许多其他的不等式，它们具有不同的性质，特别是在实验可检验性方面，见[Clauser and Shimony(克劳泽和西蒙尼)，1978]。

给定量子力学的预测式(115)，对于 u_a，u_b 和 u_c 的许多选择这一不等式都将违背。例如，用欧拉角 ϕ_k 和 θ_k 定义每个 u_k，对 $k=a$，b，c 选择 $\phi_k=0$(从而这些方向是共面的)，则 $\theta_a=0$，$\theta_b=\pi/4$，$\theta_c=\pi/2$。在此情形下，式(117)左手边的量为$\sqrt{2}$。

6.1.2.5 贝尔不等式的实验违背 实验违背了贝尔不等式。当然正如理论的实验实现那样，这里仍然存在一些问题。例如，真实的粒子源通常并不产生以准确相反方向离开源运动的粒子。人们必须要么过滤掉那些方向上不是准确相反的粒子，要么考虑它们并不向精确相反方向运动的事实。这两种解决方案并不完全可行。已有的贝尔不等式实验验证也有其他的问题，虽然实验者们逐渐能够处理这些问题以支持世界对于不等式的违背，也就是支持对量子关联的确证。

仍然有一个问题是实验处理不了的，至少不能完全处理。这一问题指的是对任意给定的粒子对，仅能够做出一对测量(一个粒子一个)。换言之，对每个给定的粒子对我们不能够真正探索整个隐态，而仅能给出它对于一对测量的含义。因而我们必须假定我们得到的隐态系综是具有代表性的。换种方式来表述，考虑我们测量 u_a，u_b 的所有粒子对，还有之后测量 u_b，u_c 的所有粒子对。为了检验贝尔不等式，我们将计算在这两组粒子对之间的关联，然后将结果代入式(117)。这样做时，我们假定了如果我们在 u_a，u_b 对上测量 u_b，u_c(不是在其上测量 u_a，u_b)，则我们得到的关联与我们在真实的 u_b，u_c 测量中得到的关联相同。大多数人认为这种反事实并无大碍——很大程度是因为对它的违背似乎会将人们牵涉进某种异常奇怪的隐匿(conspiracy)理论中，大多数这些理论怎么说都是非定域的。[①] 然而值得指出的是，这一假设的反事实性质并不是由我们碰巧测量到的偶然事实决定的(正如经典物理学中相应的反事实常常所是的那样)，而是由所涉测量之间原则上的不可容性所决定的——同一个粒子的 S_{u_a} 与 S_{u_b}不能够同时被测量，此外这与不确定原理(第4.2节)相一致，是对一个的

① 例如，假设在一个 u_b，u_c 测量中仅涉及一个"特定种类"的隐态则不是非定域的，因为此理论将违背如下条件，即在粒子源处的分布独立于对测量的选择。然而，依赖于检测者无能的模型能够断言——虽然奇特——检测者倾向于只发现"特定种类"的隐态，这取决于做哪种测量。法因的"棱镜"模型就是这样，见[Fine，1991]和[Szabo and Fine(萨博和法因)，2002]。注意检测者的能力会继续改进，并最终在纯经验基础上排除此类理论。

测量将会破坏我们之前关于另一个的知识。

最后，注意到虽然贝尔定理是用所谓的“隐”态来构造的，但它最终并不是关于隐变量理论的定理，而是关于定域性的定理。毕竟，没有什么能够阻止我们将 λ 看作是量子态本身并贯穿到贝尔的推导中。当然这样做时我们将做出量子理论违背(贝尔定域性)的预设，且我们能够在需要时直接检验此违背情况。两种方式下得到的结论是相同的，在这里讨论的意义上(贝尔定域性)，标准量子理论本身是一个非定域的理论。

6.1.3　其他类贝尔定理

流传中的许多其他定理也与定域性问题有关。我们已经看到过其中几个(第5.4.1.2节与第5.4.1.2.c节)，所以这里将只是简略提到它们。

6.1.3.1　作为定域性定理的科亨—斯佩克定理　在恰当的条件下，科亨—斯佩克定理中的非情境性条件可以看作是定域性条件，且定理中追求的从点阵子空间到布尔点阵{0, 1}的同态(回顾第5.4.1.2.b节)是“隐态”。然而，我们不是要在普遍科亨—斯佩克定理的背景中展开这一论点，而只是在GHZ定理情况中考虑这一论点，其中GHZ定理的作者们事实上最初将其(正确地)刻画为没有不等式的类贝尔定理——人们只需要假设三个粒子是类空分离的。

在此背景下，非情境性条件成为如下条件：对(99)中每个可观测量的赋值必须要独立于人们将它看作(100)中哪个可观测量的一部分，即要独立于“测量情境”，该情境是由所测三个粒子的可观测量集给出的。举例来说，粒子1拥有的 $S_x^{(1)}$ 值不能够依赖于我们是否测量 $S_x^{(1)}\otimes S_y^{(2)}\otimes S_y^{(3)}$ 或 $S_x^{(1)}\otimes S_x^{(2)}\otimes S_x^{(3)}$。这一条件是定域性的结果，因为这样的依赖性意味着对粒子1测量的结果会依赖于对粒子2和3哪一个可观测量测量的选择，即使这些选择与粒子1的测量事件处于类空分离。

6.1.3.2　哈迪论证　GHZ简化了贝尔论证的结论，即去掉了不等式，但代价是引入另一个粒子。哈迪[1992]试图做出一个不基于不等式的论证，且只考虑两个粒子。① 他考虑了图5中描绘的实验安排，其中展示了一个“双干涉仪”(称为“双”是因为在 A 点处“重叠”有两个干涉仪)。电子(e^-)与正电子

① 这里的说明来自于[Dickson，1998，pp. 209—211]。

(e^+)分别沿路径 s^- 与 s^+ 进入干涉仪。它们遇到一个分束器，其结果是通过“u”路径与通过“v”路径拥有相同的概率。如果它们都沿着各自的 u 路径则它们在 A 点相遇，并互相湮灭。否则，它们都到达第二个分束器，之后正电子以相同的概率继续通往探测器 C^+ 或探测器 D^+。电子处于相对应的情形。

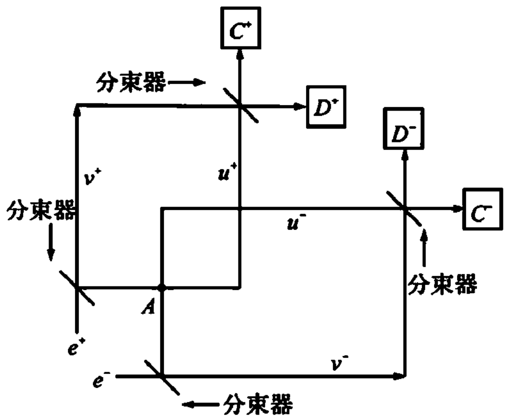


图 5　哈迪实验

哈迪假定隐变量理论将赋予每个粒子以确定的路径(u 或 v)。他进一步宣称粒子是否在给定路径上这个问题应当有一个洛伦兹不变的答案——它的路径不应该依赖于观测者的参考系。否则隐变量理论就不是洛伦兹不变的，且至少在此意义上是非定域的。这里，我们将考虑三种参考系：第一种 F^+ 中正电子的探测(在 C^+ 或 D^+)发生在电子经过其分束器之前；在第二种 F^- 中这两个事件的顺序相反；在第三种中探测是同时的。

电子与正电子的初态分别为 $|s^-\rangle$ 和 $|s^+\rangle$，表示(在它们到达第一个分束器之前)它们分别在路径 s^- 与 s^+ 上。第一组分束器影响了演化：

$$|s^\pm\rangle \to (1/\sqrt{2})(i|u^\pm\rangle + |v^\pm\rangle) \tag{118}$$

类似地，第二组分束器也影响演化：

$$|u^\pm\rangle \to (1/\sqrt{2})(|c^\pm\rangle + i|d^\pm\rangle) \tag{119}$$

$$|v^\pm\rangle \to (1/\sqrt{2})(i|c^\pm\rangle + |d^\pm\rangle) \tag{120}$$

利用式(118)，系统经过 A 点之后的态为(在参考系 F 中)：

$$\frac{1}{2}(-|\gamma\rangle + i|u^+\rangle|v^-\rangle + i|v^+\rangle|u^-\rangle + |v^+\rangle|v^-\rangle) \tag{121}$$

其中$|\gamma\rangle$是在湮灭之后的态。人们可以用式(118)—式(120)中的一个或几个来计算在参考系F、F^+、F^-中不同时刻的态。我将这些计算留给读者。

现在，在F^+中如果正电子在D^+处被探测到，则电子必须在路径u^-(概率为1)上。类似地，在F^-中若电子在D^-处被探测到，则正电子在路径u^+上。然而，在F中，在其中一个粒子经过第二个分束器前粒子对的态与粒子取路径u^+与u^-时的态相正交(由于在此情形下发生湮灭)。

之后考虑实验运行了一段时间，其中探测器D^+与D^-均记录了一次碰撞(这一结果事实上有非零概率)。那样的话，在F^+中的观测者将得到结论：电子经过路径u^-。在F^-中的另一观测者得到正电子经过路径u^+的结论。对F中的观测者而言这两个陈述都必须是真的，但正如我们上面看到的它们不能都为真。因此，哈迪断定，赋予粒子以确定路径的理论不能是洛伦兹不变的。

最后，注意哈迪也隐含地假定了非情境性。隐变量理论仅被要求重现量子理论的实验预言。若理论是情境的，则特别地，粒子的路径可能取决于在该路径上探测粒子(或否)装置的出现(与否)。例如，如果在D^-(在F^-中)中探测到电子的话，这样一个理论并不被要求赋予正电子以路径u^+，除非存在一个恰当的探测器用以确定正电子是否在u^+上。但这样的话探测器会与正电子相互作用，从而所有事情都会改变，包括(最为重要的)隐藏在哈迪论证背后的量子理论计算也会改变。在没有这样的探测器和这样的相互作用时，隐变量理论并不需要关心量子概率。但得到一致认可的是，一个非情境的理论总得遵从那些概率，因为这样一个理论不能够改变它赋予粒子的(隐)态(即路径)，该态依赖于探测器是否在路径u^+上。

6.2 对定理的回应

我们应当如何理解定域性的失效？在本小节中，我将考虑此问题的四种更精确的版本。① 定域性的失效能否用来将一个测量站的信号传输到另一个测量

① 对这些问题和更多问题的广泛讨论在许多地方都有。[Cushing and McMullin(库欣和麦克马林)，1989]、[Butterfield，1992]、[Maudlin，1994]以及[Dickson，1998，Chapters 6—9]可以引导感兴趣的读者入门。

站呢(第6.2.2节)?定域性的失效是否表明测量站之间存在某种因果联系(第6.2.3节)?定域性的失败是否表明(或能否用下述陈述来理解)两个系统在某种意义上并没有真正地区分开(第6.2.4节)?最后,或许从理论物理学的观点来看最为重要的是,定域性的失败对于一个完全洛伦兹不变的(相对论的)量子力学可能性意味着什么(第6.2.5节)?

在提出这些问题之前,我将回顾一种对定域性的重要分析(第6.2.1节),有些人认为该分析有助于解决部分问题。

6.2.1 参数独立性与结果独立性

一旦我们接受了量子理论和经验上充分的隐变量理论的非定域性——令人吃惊的是在多大程度上人们会拒绝这一结论——问题就转变为如何理解贝尔定域性。对此条件最著名的分析是用两个其他条件——常常称为"参数独立性"与"结果独立性"——来进行的。前者表达的要点是一个测量站的参数——对测量仪器的设置(要测量的自旋方向)——不影响另一测量站的结果。后者表达的要点是一个测量站的结果不影响另一测量站的结果。

如同贝尔定域性那样,这些条件是关于概率独立性的陈述:①

参数独立性 对所有的i, j, k, k', λ, 有:

$$\Pr_\lambda(x = k \mid y = k', i = u, j = u') = \Pr_\lambda(x = k \mid y = k', i = u) \quad (122)$$

类似地,也可将粒子1与2的角色互换。

结果独立性 对所有的i, j, k, k', λ, 有:

$$\Pr_\lambda(x = k \mid y = k' \mid i = u, j = u')$$
$$= \Pr_\lambda(x = k \mid i = u, j = u') \times \Pr_\lambda(y = k' \mid i = u, j = u') \quad (123)$$

见第6.1.2.3小节的记号。参数独立性与结果独立性的合取产生了贝尔定域性以及概率理论的某种平凡应用。

该分析的价值在于它更精确地揭示了量子理论是如何违背贝尔定域性的。特别是量子理论违背了结果独立性,但满足参数独立性。关联式(115)的推导

① 关于这些条件存在稍许不同的版本,例如考虑到仪器中的隐变量等。这里我们忽略这些不同。此外,存在人们可能会引入的许多其他定域性条件,见[Dickson, 1998, Chapters 6—9]。[Jarrett(杰瑞特), 1984]给出了关于此区别的第一个清晰陈述。这里陈述的条件,和这里用到的命名法,是由[Shimony, 1986]给出的。

或多或少证明了第一个陈述。[①] 我们将在下一小节中看到证明第二个陈述的一种方式。

6.2.2 信号传递

6.2.2.1 定域性与信号传递

隐藏在将贝尔定域性分析为参数独立性与结果独立性背后的部分的意图最初是，从贝尔定域性中分离出在某种意义上并不违背相对论理论的那部分。特别是人们常常做出的论断是，结果独立性的失效在某种程度上与相对论一致，而参数独立性的失效并不一致。

特别是，人们可以将相对论理解为禁止超光速信号传递。按照这样的论证，对结果独立性的违背并不涉及信号传递的可能性，因为即使它表明一个测量站的结果在概率上依赖于另一个测量站的结果(且测量事件当然是类空分离的)，但结果本身是概率的，也就是说实验者不能够控制结果。但为了用结果间的依赖性来传递信号，有必要对结果进行控制。

另一方面，实验者在控制着参数——事实上参数通常被认为是实验者自由选择的结果。因而，按照这样的论证，对参数独立性的违背意味着实验者能够通过对所在测量站的参数(自旋被测量的方向)操作(在概率上)影响另一测量站的结果。

然而，谨记参数独立性与结果独立性中的概率是由隐态 λ 所产生的那些概率。如果实验者没有对这些隐态进行控制，则参数独立性的失效也并不意味着信号传递的可能性。此外，对隐态的控制意味着对结果独立性的违背事实上也表明了信号传递的可能性，只要由不同隐态产生的结果概率不同。这样的话，实验者事实上能够通过控制隐态在概率的意义上局域地操作结果，使得给定的结果更可能或更不可能，从而借助于结果间的概率依赖性(在概率上)影响另一测量站的结果。

德布罗意—玻姆理论为能够控制隐态的重要意义提供了一个启发性案例。

---

① 严格来说，为了实现式(123)中的概率，人们应当引入两个附加系统来作为仪器，且将仪器设置(对自旋方向的选择)看作是这些附加系统的态。因此就有了一个在四重张量积希尔伯特空间中的态，并考虑由该态产生的概率。然而，以这种方式进行分析并不具有特别的启迪性，并且产生的结果相同，即结果独立性失效了。

很明显该理论满足结果独立性，因为它是完全决定性的理论。也就是说给定参数 i 与 j 和初态(在该理论中，它是粒子的初始位置与量子波函数)，测量的结果就被确定了。因而，特别地由 λ，i 和 j 确定的一个测量结果并不依赖于另一测量的结果。另一方面，德布罗意—玻姆理论确实违背了参数独立性，因为在一个粒子上的测量结果的确一般依赖于对另一粒子哪个自旋方向的测量。(仪器设置中的变化改变了复合系统的波函数，在这种方式下最终改变了量子势，进而改变了两个粒子的轨迹。)然而，想要知道如何通过对参数的改变影响一个粒子(并进而能够控制此影响)就要知道它的精确位置，但粒子的精确位置在此理论中并不知晓。因而对参数独立性的违背不能用于信号传递。

事实上正如库欣［1994］在此背景中指出的，华伦蒂尼(Valentini)［1991；1991b］指出在德布罗意—玻姆理论中，当且仅当粒子的分布不同于由通常量子力学的概率所给出的时信号传递才是可能的。因此，假定玻姆理论确实遵从这一分布，在现象学上我们得到这样的情形：量子理论违背结果独立性并满足参数独立性，而在隐变量(粒子的轨迹)的层次(不可控制的!)上该命题的逆命题才是真的。① 因而人们在把参数独立性的失效当作信号传递可能性的时候应该非常谨慎。

6.2.2.2 无信号传递定理 虽然上文给出的结果独立性与参数独立性之间的区分并不等同于信号传递的不可能性与可能性之间的区分，但量子理论(回想起它遵循结果独立性而非参数独立性)是真的不允许从一个测量站到另一测量站的信号传递(或事实上是一般的非定域信号)。这一断言是量子无信号传递定理的结论。

得到此结论的方案(至少)有两种，一种关注于定域的一般相互作用能够影响空间上分离系统的态，另一种关注于测量能够影响在空间上分离系统的测量结果。我将轮流概述每一种方案。在这两种情形下，我将考虑在(一般是纠缠

---

① 事实上，华伦蒂尼［1991a；1991b］的观点指的是标准量子分布是一种"平衡态"分布，宇宙会自然地弛豫到该分布，该观点得到了经典 H 定理的一种量子类似理论的支持(见尤菲克所撰写的本书第九章)。在此观点中，有可能(确实非常可能)宇宙并不完全处于平衡态分布，这意味着原则上如果人们能够找到并不处于平衡态分布的粒子系综的话，人们能够传递信号。

的)态 W 下的双粒子系统，并假定两个粒子是空间分离的。

6.2.2.2.a 定域相互作用与约化态 现在用两个不同的(由两个不同的哈密顿量产生的)幺正演化 $U^{(1)}\otimes U^{(2)}$ 与 $U^{(1)}\otimes\widetilde{U}^{(2)}$ 来考虑复合系统的演化。无信号传递定理说的是，不论系统的演化是按照 $U^{(1)}\otimes U^{(2)}$ 还是 $U^{(1)}\otimes\widetilde{U}^{(2)}$，在演化结束时，粒子1的约化态是相同的。事实上在第一种演化下粒子1的约化态为(回顾第1.2.6.3.a小节)：

$$
\begin{aligned}
W^{(1)} &= \mathrm{tr}^{(2)}[U^{(1)}\otimes U^{(2)}W(U^{(1)}\otimes U^{(2)})^{-1}] \\
&= \sum_n \langle e_n | U^{(1)}\otimes U^{(2)}W(U^{(1)}\otimes U^{(2)})^{-1} | e_n\rangle \qquad (124)
\end{aligned}
$$

其中$\{|e_n\rangle\}$为粒子2在希尔伯特空间上的某个正交基。但回顾(部分)迹函数并不依赖于此基矢的选择。因此，当在另一种演化下推测约化态 $\mathrm{tr}^{(2)}[U^{(1)}\otimes\widetilde{U}^{(2)}W(U^{(1)}\otimes\widetilde{U}^{(2)})^{-1}]$时，我们仅需要选择一个基矢$\{|\tilde{e}_n\rangle\}$以使得 $\widetilde{U}^{(2)}|\tilde{e}_n\rangle = U^{(2)}|e_n\rangle$。因为 $U^{(2)}$ 与 $\widetilde{U}^{(2)}$是幺正的，这样的一个基矢总是存在的。

也就是说，不存在只与其中一个系统作用就能够影响另一个系统约化态的相互作用，且因为粒子1的边缘概率(仅是粒子1上可观测量测量结果的概率)只依赖于约化态 $W^{(1)}$，所以不存在只涉及粒子2就能够改变粒子1测量结果的统计的相互作用(当然，反之亦成立)。

6.2.2.2.b 单系统测量 但测量又会如何呢？考虑对可观测量$\mathbb{I}\otimes G$ 的测量，即对系统2的 G 的测量，保留系统1不变。令 $G=\sum_n g_nP_n$。当然，我们已经看到如果将复合系统的态投影到该测量结果上，则一般地(例如 W 是单重态)系统1上测量结果的概率将改变。该事实仅是对两个系统间存在关联的一个复述。但要记住的是我们这里谈论的是信号传递的可能性。被限制于系统1附近的观测者不会知道系统2的测量结果，仅能够知道对$\mathbb{I}\otimes G$ 的测量发生或未发生。对于这样一个观测者而言，对粒子2的$\mathbb{I}\otimes G$ 的测量是一个“非选择性”测量，意味着此观测者关于测量后的态能够表述的顶多是复合系统现在处于如下态：

$$
\sum_n (\mathbb{I}\otimes P_n)W(\mathbb{I}\otimes P_n) \qquad (125)
$$

弄明白为什么在一个非选择性测量之后态是式(125)很有意义。假定如果结果事实上是 g_k，那样的话应用塌缩假设态应当为 $P_kWP_k/\mathrm{Tr}[WP_k]$，可将此

表达式与式(19)的路德规则作比较。现在考虑结果 P_k 的概率为 $\mathrm{Tr}[WP_k]$，因而如果我们不知道测量结果(即测量是非选择性的)的话，则态是所有可能结果的加权和，权重由不同结果的概率给出，这也就是说态是式(125)。

这里的论点是，粒子 2 附近的一个实验者(“B”)试图通过选择测量或不测量(粒子 2 的)G 而发送给粒子 1 附近的实验者(“A”)以一个信号。则问题变为 A 能否在粒子 1 的测量结果统计中检测到由 B 对粒子 2 测量引起的任何改变。

答案是“否”。要明白其原委，考虑一个可观测量 $F\otimes\mathbb{I}$，且令 $F=\sum_n f_n Q_n$。在态 W 中，(对 $F\otimes\mathbb{I}$的测量)结果 f_m 的概率为：

$$\mathrm{Tr}[(Q_m\otimes\mathbb{I})W] \tag{126}$$

反过来，如果假定态为式(125)，即做了一个对$\mathbb{I}\otimes G$ 的非选择性测量，则(对 $F\otimes\mathbb{I}$的测量)结果 f_m 的概率为：

$$\mathrm{Tr}[(Q_m\otimes\mathbb{I})\sum_n(\mathbb{I}\otimes P_n)W(\mathbb{I}\otimes P_n)] \tag{127}$$

借助于 $Q_m\otimes\mathbb{I}$的线性，这能够在求和中来进行，且通过迹函数的线性求和的迹则变为求迹的和。此外，由于 $Q_m\otimes\mathbb{I}$与$\mathbb{I}\otimes P_n$(对任意的 n，m)对易，此求和变为：

$$\sum_n \mathrm{Tr}[(\mathbb{I}\otimes P_n)(Q_m\otimes\mathbb{I})W(\mathbb{I}\otimes P_n)] \tag{128}$$

利用式(12)还有对任意投影 P 的事实 $PP=P$，式(128)变为：

$$\sum_n \mathrm{Tr}[(\mathbb{I}\otimes P_n)(Q_m\otimes\mathbb{I})W] \tag{129}$$

回到迹函数中求和，且注意到$\mathbb{I}\otimes P_n$ 构成了对单位算符的一个分解(即 $\sum_n \mathbb{I}\otimes P_n=\mathbb{I}\otimes\mathbb{I}$)，则我们发现式(129)正是式(126)。换言之，A 在粒子 1 上做出的测量不能够确定复合态是最初的 W 还是态式(125)——B 对$\mathbb{I}\otimes G$ 的非选择性测量对 A 在粒子 1 上进行的测量结果概率没有影响。因此，B 不能够通过这样一个测量发送信号给 A。

6.2.3 因果性

贝尔定域性(或其中一个组成条件)的失效是否意味着两个测量站之间的因果联系呢？该问题被广泛地讨论与争论，这里当然我也不能够解决它，但我们可以考虑一些求解方法。

考虑因果性分析的反事实研究法，这被看作是像“如果发生了 C 的话则会发生 E”与“如果没有发生 C 的话则不会发生 E”这类反事实条件句的真值因果联系的充分条件。① 采取对因果性的这一理解的话，似乎违背结果独立性的决定论隐变量理论中的因果联系存在于结果之间，或者违背参数独立性的决定论隐变量理论中的因果联系存在于参数与结果之间。若隐变量理论是概率的，则或许会代之以推论出存在概率因果性的关系，虽然人们需要构造恰当的概率性反事实条件，例如用“E 的概率可以更高(更低)”来取代“将会(不会)发生情形 E”。

另一方面，其他人更偏好于另一种对因果性的说明，其中要求因果联系由能够传输一个“标记”的“因果过程”来承担。② 即在原因上刻印某种“标记”应该留下一条“从原因到结果”的轨迹，且最终体现在结果本身上。关于这一说明的真正意义存在不同的理解，但至少在某些理解——最为清楚的是那些要求标记在空间中连续地移动的——中对参数独立性与结果独立性的违背并不意味着测量站之间存在着因果联系。

最后，还有些人如科列尔(Collier)[1999]曾论证说因果性在信息理论的意义上相当于信息的传输。莫德林(Maudlin)[1994, Chapter 6]论证说在贝尔类实验中存在信息的传输。因此如果人们相信这些论证，则据此因果论说明会得出测量站之间存在因果性。

6.2.4 整体论

围绕贝尔定理的问题导致有些人持有极端观点，将纠缠态看作表征了纠缠粒子对(或 n 元组)的整体论特征。一方面，我们能够认同单重态(向其上的投影)表示的特征不可约化到(且不随附于)单个粒子的特征。(否则的话定域的隐变量理论总归是可能的)。另一方面，有人宣称这些整体论特征说明了我们不能够为量子关联提供一种共因解释，这种量子关联的方式并不违背我们应当关注的所有版本的定域性，特别是它并不意味着对相对论的违背。这里基本的

---

① 因果性常常与这些反事实条件句相联系。刘易斯(Lewis)[1973]因这样的分析而出名，且正如他的研究清楚表明的，人们必须在此基本理念上添加一些附加条件——例如，关于可能世界的相似性条件，其中那些条件将影响所涉反事实条件句的含义，进而影响其真值。

② 例如，见[Salmon(萨尔蒙)，1984]。

论点很清楚：贝尔类实验涉及的“两个”(或更多)粒子并不是真正的“两个”粒子；它们其实是一个对象，因此不管它们离多远都不存在“从一个到另一个的影响”这一问题，因为没有“一个与另一个”，有的只是整个事物。

然而此提议也有一些显而易见的问题。首先，不清楚它是否真的解释了什么；或许它仅是用一种同样神秘的学说从形式上重新表述了问题。此外，考虑到与相对论的明显冲突，不清楚在两个不同客体之间的类空影响与同一客体不同“部分”之间的类空“联系”间的真正的区别是什么(当然，这里我们必须对“部分”的意义非常谨慎)。事实上，贝尔类实验中涉及的事件是时空中精确确定的局域化的事件，也就是宏观的指针读数。这些事件之间存在不能够用定域共因解释的关联。此事实足以产生关于非定域性和与相对论相容性的问题，不论人们是否希望诉诸以某种方式产生这些关联的粒子对的某种“整体性”特征。换言之，我们只是不得不做出整体论本身与相对论不相容的结论。参见[Butterfield，1992]。

其次，正如之前在测量问题背景中讨论过的(第5.3.2节)，纠缠无处不在——即使是我们日常经验中的客体也可能处于纠缠态。我们能否表述清楚这样的观点，即这些明显区别的客体事实上并不是完全独立的客体，而是某个整体对象的某“部分”？或许如果我们要重现日常经验中显而易见的事实的话，至少需要给出一些重要的形而上学的内容。

6.2.5 相对论理论

虽然有人会发现非定域性与直觉知识相反，但它提出的真正问题可以说与相对论理论明显不相容。事实上，假定量子理论的非定域性完全并明确地与相对论理论相容，很难看到人们会对此有反对意见。

并且事实上另一方面人们可能注意到了量子无信号传递定理(第6.2.2.2节)提出了量子理论与相对论理论的某种相容性。因此正如有人所讲的，在两种理论之间可能存在“和谐共处(peaceful existence)”。① 虽然量子关联似乎意味着定域性的失效，且虽然量子态的塌缩瞬时地发生，但这些事实都没有要求发送快于光速信号的能力。更一般地，量子理论的这些特征都没有必要在实验上

① 该术语是由西蒙尼[1978]引入的。

或观测上与相对论冲突。有时会说量子理论遵从相对论规律的字面意义(但违背了此规律的精神)。

此外，即使是超光速信号与相对论不相容的表述也是有争议的。不相容性论证依赖于这样的观点，即这样的信号传递产生不一致的因果循环，因为它允许通往过去的因果传播。有人会指出即使是超光速的因果过程(例如，物质或能量的超光速传播)也并非与相对论不一致。

那么相对论要求什么？有一点似乎是清楚的，且至少大多数物理学家认可的是：如果理论是相对性的就必须得是洛伦兹不变的。这里，我们遇到了真正的问题。标准量子理论的塌缩假设不是洛伦兹不变的，也不清楚如何在保证经验充分性的前提下使它成为洛伦兹不变的。参见[Aharonov and Albert，1981]。

当然，存在相对论的量子理论(见本书第七章和第八章)。这些理论中的方程在必要的方式下是洛伦兹不变的。然而，塌缩假设不是洛伦兹不变的——它明确提出了瞬间的塌缩，而在相对论的量子理论中仍然需要这一假设。

此外，远不清楚最简单且明显解决了测量问题(回顾第5.5节)的量子理论解释能否成为洛伦兹不变的。我们并不知道是否所有隐变量理论必须违背洛伦兹不变性，但关于此陈述的证据非常充分。

另一方面，无信号传递定理似乎意味着量子理论本身在观测上与相对论一致。因此，任何理论——尤其是观测上与量子理论不可区分的任何隐变量理论——在观测上都与相对论相一致。所以，虽然许多隐变量理论在隐变量值演化的层次上明显违背了洛伦兹不变性，但这一违背并非经验可行的。换一种方式表达该论点：这些理论要求一个优先的参考系(正如具有塌缩假设的量子理论本身所要求的)，目前在实验上仍没有一种确定优先参考系的方式。因而这些理论能否令人满意与人们赋予洛伦兹不变性的地位有很大关系。

7. 数学附录

这些附注是为文中的一些定义和标准数学事实提供速览。所有论证过程都省略了。

7.1 希尔伯特空间

7.1.1 矢量空间

一个矢量空间 V 是一个集合，它在加法和域 K 中的"一个纯量乘法"下是闭合的。对于任意的 u，v，$w\in V$ 和 k，$k'\in K$，这些操作必须满足：(可交换性)$v+w=w+v$；(向量结合性)$u+(v+w)=(u+v)+w$；(加法基元)$\exists\vec{0}\in V$ 和 $\forall v\in V$，$v+\vec{0}=v$；(加性逆元)$\forall v\in V$，$\exists -v\in V$，$v+(-v)=\vec{0}$，且一般将 $v+(-w)$ 写作 $v-w$；(纯量结合性)$k(k'v)=(kk')v$；(纯量单位元)对于 $1\in K$(K 中的单位元)有 $1v=v$；(向量分配性)$k(v+w)=kv+kw$；(纯量分配性)$(k+k')v=kv+k'v$。注意：可交换性事实上是根据其他特性得出的；加性逆元是(可证实)唯一的；$\forall k\in K$，$v\in V$，$k\vec{0}=0v=\vec{0}$。在我们考虑的所有情形中，$K=\mathbb{R}$ 或 $K=\mathbb{C}$。集合 V 可交换性地用于指代整个矢量空间还有其中的矢量集。有时出于清晰性，人们用"K 上的向量空间"这一说法。

7.1.2 基和维度

给定一个向量集 $\{v_n\}\subseteq V$，具有形式 $v=\sum_{i=1}^{N}k_nv_n$(有 $k_n\in K$)的任意向量被称为 v_n 的一个线性组合。如果不存在可以写成其他向量线性组合的 v_n，则集合 $\{v_n\}$ 称为线性独立的。(在此情形下，该集合可能有无穷的基数，但注意线性组合总是有限和。)V 中任意极大的线性独立的集合称为是 V 的一个基。可以证明所有这些集合都拥有相同的基数，称为空间的维度，记作 $\dim V$。拥有相同维度的所有向量空间 V 和 V' 是同构的，即存在一个一一映射 m：$V\to V'$，满足对任意的 $k\in K$ 和任意的 v，$w\in V$ 有 $m[k(v+v')]=km(v)+km(v')$。(最后这一条件使得 m 成为一个线性映射，且因为其是一一映射所以成为一个同构。)

7.1.3 内积空间

K 上向量空间 V 的内积是一个从 $V\times V$ 向 K 的映射，记作 $\langle\cdot,\cdot\rangle$，且对所有的 u，v，$w\in V$ 和所有的 $k\in K$ 满足：(非负性)$\langle v,w\rangle\geqslant 0$；(非简并性)当且仅当 $v=\vec{0}$ 时 $\langle v,v\rangle=0$；(半双线性)$\langle u,k(v+w)\rangle=k\langle u,v\rangle+k\langle u,w\rangle$ 和 $\langle v,w\rangle=\langle w,v\rangle^*$，其中 * 表示 K 中的共轭(若 $K=\mathbb{C}$ 时为复共轭，若 $K=\mathbb{R}$ 时为单位映射)。(注意：根据最后两个特征有 $\langle k(v+w),u\rangle=k^*\langle v,u\rangle+$

$k^*\langle w, u\rangle$；因此得名“半双线性”。）拥有内积的矢量空间称为“内积空间”。在一个内积空间中，当且仅当$\langle w, w'\rangle=0$时w正交于w'，写作$w\perp w'$。在内积空间中，当基的要素相互正交时该基称为是“正交的”。

内积的一个重要事实是施瓦茨不等式（Schwarz Inequality），即对任意的v，$w\in V$有$\langle v, v\rangle\langle w, w\rangle\geqslant|\langle v, w\rangle|^2$。

7.1.4 模和正交基

K中向量空间V的模是一个从V到$\mathbb{R}$的函数，记作$\|\cdot\|$，满足对于所有的v，$w\in V$和$k\in K$有：当且仅当$v=0$时$\|v\|=0$；$\|kv\|=|k|\ \|v\|$；$\|v+w\|\leqslant\|v\|+\|w\|$。（注意：因而得到对所有的$v\in V$有$\|v\|\geqslant 0$。）模以一种直观的方式在$V$上定义了一个拓扑（见第7.5节）：对某些$v\in V$和$r\in\mathbb{R}$而言，开球是具有形式$\{x\mid\|x-v\|<r\}$的集合。用另一种方式表达则是：列$\{v_n\}\subseteq V$收敛于“范拓扑中”的矢量$v\in V$以避免$\|v_n-v\|$（在$\mathbb{R}$中）收敛于0。内积通过$\|v\|=\sqrt{\langle v, v\rangle}$来定义一个模。关于通过内积得到模的赋范复向量空间有一个重要事实是“极化恒等式（polarization identity）”：

$$\langle w, v\rangle=\frac{1}{4}(\|w+v\|^2-\|w-v\|^2+i\|w+iv\|^2-i\|w-iv\|^2) \tag{130}$$

换言之，在这样的空间中内积也是通过模得到的。

在一个内积空间中，若一个正交基（第7.1.3节）要素的模都为1，则称它为标准正交的。给定一个内积空间的任意基，通过“格拉姆—施密特（Gram-Schmidt）正交化”来构造一个正交基是可能的，我们将忽略此正交化的细节。注意在一个内积空间中，存在用某个正交基$\{e_n\}$对给定向量v的简明表达，即$v=\sum_n(e_n, v)e_n$。

7.1.5 子空间

向量空间的子集W，如果它在从V中继承来的操作下独立成为一个向量空间的话，则称为一个子空间。类似的定义对于内积空间和希尔伯特空间（下文第7.1.7节中定义的）也成立。在内积空间中，当对于任意的$w\in W$和$w'\in W'$有$w\perp w'$时，子空间W正交于另一子空间W'。

7.1.6 直和

向量空间V和V'的直和（在相同的域内，如$\mathbb{R}$或$\mathbb{C}$）是其要素取自于笛卡

尔乘积$V \times V'$中的向量空间，向量空间操作定义了分量方式。例如，若$V = V' \oplus V''$，则V'和V''是不相交的(或，在内积空间中是正交的)子空间，且V是它们的张成。

7.1.7 巴拿赫空间和希尔伯特空间

一个巴拿赫空间是一个相关于范拓扑完备的赋范向量空间(意味着任意向量序列的极限本身包含于该空间)。一个希尔伯特空间是相关于范拓扑完备的内积空间(这里的范是由内积给出的那个)。三种经典的希尔伯特空间类型是在实数、复数和四元数上的希尔伯特空间。

7.1.8 对偶空间

给定向量空间V，对偶空间(有时记作V^*，但这里的*不是复共轭)是在V上的线性泛函，即是从V到$\mathbb{R}$的线性映射空间。当V拥有拓扑结构（如当V是一个希尔伯特空间)时，我们局限于连续的线性泛函。在V上的一个(连续的)线性泛函ϕ的模是：

$$\|\phi\| = \sup_{v \in V}\{|\phi(v)| \mid \|v\| \leqslant 1\} \tag{131}$$

每个有限维的矢量空间都有与它的模相同的维度。里兹表示定理(Riesz Representation Theorem)规定，对任意的希尔伯特空间$\mathcal{H}$(有限维或无限维)，和在其对偶$\mathcal{H}^*$中的任意连续线性泛函ϕ，存在一个唯一的$v \in \mathcal{H}$满足，对于所有的$w \in \mathcal{H}$，有$\phi(w) = \langle v, w \rangle$。反过来，每个$v \in \mathcal{H}$明显生成一个在$\mathcal{H}$上的连续的线性泛函$\langle v, \cdot \rangle$。换言之，存在一个从希尔伯特空间向其对偶空间的一一映射，Φ：$\mathcal{H} \to \mathcal{H}^*$。此外，$\Phi$是一个等距同构($\|v\| = \|\Phi(v)\|$)和一个“反等距同构”，特别地，对任意的$v \in V$和$k \in \mathbb{C}$有$\Phi(kv) = k^* \Phi(v)$。最后的这个特征源自于内积是半双线性的这一事实。

7.1.9 张量积

两个都在K上的希尔伯特空间$\mathcal{H}_1$和$\mathcal{H}_2$的张量积是第三个希尔伯特空间，即K上的$\mathcal{H} = \mathcal{H}_1 \otimes \mathcal{H}_2$，按照如下方式从$\mathcal{H}_1$和$\mathcal{H}_2$构造而来。首先选择$\mathcal{H}_1$的基$\{e_n\}$和$\mathcal{H}_2$的基$\{f_n\}$，然后做$\{e_n\}$和$\{f_n\}$的笛卡尔积。该集合包含所有形为$(e_n, f_m)$的对，且它被约定为是张量积空间$\mathcal{H}$的一个基。因此，在此阶段，$\mathcal{H}$是由$(e_n, f_m)$所有形式的线性组合(在$K$上)组成的。现在，令$\mathcal{H}_1$和$\mathcal{H}_2$上的内积记

作$\langle\,,\rangle_1$和$\langle\,,\rangle_2$。对所有的v，$w\in\mathcal{H}_1$和x，$y\in\mathcal{H}_2$，通过$\langle v\otimes w,\ x\otimes y\rangle=\langle v,\ x\rangle_1\langle w,\ y\rangle_2$来定义在$\mathcal{H}$上的内积，且通过线性拓展到所有的$\mathcal{H}_1\otimes\mathcal{H}_2$上。最后，在由此内积引起的范拓扑中完备$\mathcal{H}_1\otimes\mathcal{H}_2$。

注意$\dim\mathcal{H}=\dim\mathcal{H}_1\times\dim\mathcal{H}_2$。人们可以(若$\dim\mathcal{H}$不是素数)将一给定的希尔伯特空间“因子化”为一个张量积，这通常有许多种方式。

7.1.10　凸集和锥面

若对于任意的x，$y\in X$，有$rx+(1-r)y\in X$，其中所有的r都在实区间$[0,1]$中，则实向量空间的一个子集X是凸的。换言之，连接x和y的“线段”也在X中。若凸集X中的每一个极值点都能唯一分解为来自X点的凸组合，该凸集就称为是单形的(simplex)。

V中的一个正锥是一个满足对所有的实$r\geqslant 0$和所有的$x\in C$有$rx\in C$的集合$C\subseteq V$(一个负锥反而要求$r\leqslant 0$)。一个凸锥是一个同样凸的锥面。给定一个集合$S\subseteq V$，可以以一种明显的方式构造由S生成的凸集，即在所要求的条件下进行封闭。等价地，可以取包含S的所有凸集的交集。类似的论点对于锥面也成立，当然对于凸锥也成立。

给定V中的一正锥C，和V上的一内积$\langle\cdot,\ \cdot\rangle$，对应于此内积的$C$的对偶是：$C^*=\{y\mid\langle x,\ y\rangle\geqslant 0,\ x\in C\}$。若$C=C^*$，我们说$C$是自对偶的。

凸集X的一个面$F\subseteq X$是一个在“纯化(purification)”下闭合的凸子集，意味着对任意的$v\in F$，若$v=pv_1+(1-p)v_2$，则v_1和v_2也在F中。

7.2　算符

7.2.1　关于算符的基本定义

在K上的向量空间V中的一个线性算符F，是一个从V到其自身的映射，它保留了V的线性结构，即对任意的v，$w\in V$和任意的$k\in K$有：$F(v+w)=F(v)+F(w)$和$F(kv)=kF(v)$。人们通常只写作Fv，且这里我们把F当作一个“算符”(因为我们不讨论非线性的算符)。两个算符F和G，若它们的对易子$[F,\ G]:=FG-GF$为0(即“零”算符)，则它们是对易的。单位算符，记作$\mathbb{I}$，是在V上的算符，满足$\mathbb{I}v=v$，$\forall v\in V$。

给定一算符F，对某个$k\in K$满足$Fv=kv$的任意向量v，被称为F的本征

向量，且 k 是其相对应的本征值。算符无须拥有本征向量。零向量通常不算作算符的本征向量。

向量空间 V 上的算符 F 是可逆的，仅当存在一个算符 G 满足 $FG=\mathbb{I}$，其中 $\mathbb{I}$是在 V 上的单位元。算符 G 记作 F^{-1}。若存在的话它是唯一的。

在张量积空间 $V=V_1\otimes V_2$ 上，考虑两个算符，V_1 上的 F 和 V_2 上的 G。张量积算符 $F\otimes G$ 可以被定义如下：选择 V_1 和 V_2 上的任意基 $\{e_n\}$ 和 $\{f_m\}$，定义 $(F\otimes G)(e_n\otimes f_m)=(Fe_n)\otimes(Gf_m)$，并通过线性拓展到所有的 V。

7.2.2 边界性和连续性

若有一个模 $\|\cdot\|$，当且仅当存在某个 $r\in\mathbb{R}$ 满足对所有的 $v\in V$ 有 $\|Fv\|\leqslant r\|v\|$，则 F 是有界的。当 V 有一个拓扑（就像它是一个希尔伯特空间时那样），人们说算符 F 是连续的，当且仅当 F 就像作为拓扑空间 V 上的函数那样是连续的。一个算符是有界的，当且仅当它是连续的。此外，若 V 的维度是有限的，则 V 上的所有算符都是有界的，即连续的。注意到无界算符并未将整个空间作为它们的范围，因而人们必须了解它们的定义域。

7.2.3 共轭

当$\mathcal{H}$是有限维时，我们如下定义共轭及相关的概念。算符 F 在$\mathcal{H}$上的共轭 F^* 对任意 v，$w\in V$ 满足$\langle Fw, v\rangle=\langle w, F^*v\rangle$。算符 F 是自轭的，若$F=F^*$。关于共轭存在的论证是非平凡的。

当$\mathcal{H}$是无限维时，我们必须更加谨慎，因为如果 F 无界的话，V 可能不在 F 范围内（见第7.2.2节）。在无限维情形中，我们定义 F 的共轭 F^* 如下。F^* 的范围是所有的 $v\in\mathcal{H}$，这使得存在一个 $v'\in\mathcal{H}$满足对所有在 F 范围内的 w 有$\langle Fw, v\rangle=\langle w, v'\rangle$。对于每个这样的 v，定义 $F^*v=v'$。（必须指出 F^* 因而成为一个算符。）最后，为了在无限维情形下做出区分，若对于 F 区域内的所有 v 和 w 有$\langle Fw, v\rangle=\langle w, Fv\rangle$，我们说一个算符 F 是对称的。如果 $F=F^*$，我们认为 F 是自轭的。可以证明，二者的区别在于一个对称的但非自轭的算符的范围将是其共轭算符的范围的恰当子集。

7.2.4 正规算子

正规算子 F 是与其共轭相对易的算符：$FF^*=F^*F$。有了上述评论，可见若$\mathcal{H}$是无限维时必须要注意与算符的范围相关的问题。显然所有的自伴算符都

是正规的，但其逆命题并不成立。如$F=2i\mathbb{I}$。

7.2.5 投影算符

仅在 $PP=P$ 情形，即对任意的 $v\in V$ 有 $P(Pv)=Pv$ 时，算符 P 是幂等的。在 V 上的算符 P 是一个投影算符，仅当它是自轭的和幂等的。每个投影算符 P 对应于一个闭合子空间，即向量子空间 v，有 $Pv=v$。注意当且仅当相应的子空间正交时，$PQ=0$。

7.2.6 幺正算符

向量空间 V 的自同构是一个从 V 到其自身的映射，该映射“保留了 V 的结构”，特别是线性的、内积的和拓扑结构(若 V 拥有后两种结构的话)。令 U 是一个在希尔伯特空间$\mathcal{H}$上的(线性)算符，满足：U 是可逆的(因而 U 是一一映射的)，且 U 保留了内积(即对任意的 v，$w\in\mathcal{H}$，$\langle Uw, Uv\rangle=\langle w, v\rangle$)。这样一个算符称为“幺正的”，且它显然实现了$\mathcal{H}$的一个同构。当然在该情形中，$U$ 也保留了模，即对所有的 v 有 $\|Uv\| = \|v\|$。

足以表明对任意的幺正算符 U，有 $U^*=U^{-1}$。反过来，任意拥有此特征的可逆线性算符都是幺正的。若换作$\langle Uw, Uv\rangle=\langle w, v\rangle^*$和 $U(kv)=k^*Uv$，则 U 是反幺正的。

7.3 希尔伯特空间$\mathbb{C}^2$

拥有 2 个分量的复列向量空间记作$\mathbb{C}^2$。$\mathbb{C}^2$ 的元素以分量方式相加：

$$\begin{pmatrix} a \\ b \end{pmatrix} + \begin{pmatrix} c \\ d \end{pmatrix} = \begin{pmatrix} a+c \\ b+d \end{pmatrix} \tag{132}$$

在此空间上的(线性)算符可以用 2×2 复矩阵来表示。矩阵对向量的操作遵循下述规则：

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} ax+by \\ cx+dy \end{pmatrix} \tag{133}$$

两个矩阵的乘积(它事实上可以通过上述规则而得到)为：

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} e & f \\ g & h \end{pmatrix} = \begin{pmatrix} ae+bg & ce+dg \\ af+bh & cf+dh \end{pmatrix} \tag{134}$$

此空间上的内积由下式给出：

$$\left\langle \begin{pmatrix} a \\ b \end{pmatrix}, \begin{pmatrix} c \\ d \end{pmatrix} \right\rangle = a^* c + b^* d \tag{135}$$

分量为 a 和 b 的向量的模或“长度”则是 $\sqrt{a^* a + b^* b}$。（对比这一表达式与实向量空间 $\mathbb{R}^2$ 中的向量的欧几里得长度。）

7.3.1 泡利矩阵

泡利矩阵（$\mathbb{C}^2$ 上的算符）为：

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{136}$$

这些矩阵拥有许多优良的特征，读者可能希望去证明它们。例如：

$$\begin{aligned} \sigma_x^2 = \sigma_y^2 = \sigma_z^2 &= \mathbb{I} \\ \mathrm{Tr}[\sigma_x] = \mathrm{Tr}[\sigma_y] = \mathrm{Tr}[\sigma_z] &= 0 \\ \sigma_x \sigma_y &= i\sigma_z \\ [\sigma_x, \sigma_y] &= 2i\sigma_z \end{aligned} \tag{137}$$

将最后两个特征进行推广可知，轮换下标它们仍然为真。

人们常常见到表达$\vec{\sigma}$式，它被理解为其分量是三个泡利矩阵的一个“向量”，因而像$\vec{r}\cdot\vec{\sigma}$是对三个泡利矩阵线性组合的一个简略表达，其系数由（实向量）$\vec{r}$的分量给出。

7.4 偏序集和格

7.4.1 偏序集

一个偏序的集合（通常称为偏序集）是一个集合 L 加上其所服从的关系 $\leqslant$，对于所有的 a，b，$c \in L$ 有：（自反性）$a \leqslant a$；（反对称性）若 $a \leqslant b$ 且 $b \leqslant a$，则 $a = b$；（传递性）若 $a \leqslant b$ 且 $b \leqslant c$，则 $a \leqslant c$。关系式 $\leqslant$ 被称为 L 上的偏序。特别需要注意的是一般会存在既非 $a \leqslant b$，也非 $b \leqslant a$ 的 a，$b \in L$。（若在 L 中不存在这样的 a 和 b，则 $\leqslant$ 是 L 上的一个全序。）正如在别处，我用 L 来同时表示偏序集本身和基本的集合。

7.4.2 格

令 L 为一个偏序集。定义两个元素 a，$b \in L$ 的并集为 a 和 b 的最小上界，

即最小的(在$\leqslant$下)满足$a\leqslant c$和$b\leqslant c$的c。a和b的并集(一般地它不需要存在)记作$a\vee b$。定义两个元素a，$b\in L$的交集为a和b的最大下界，即最大的(在$\leqslant$下)c满足$c\leqslant a$和$c\leqslant b$。a和b的交集(一般它并不需要存在)记作$a\wedge b$。其中每一对元素(每个有限集)都有一个交集和一个并集的偏序集称为格。

若每个$a\in L$有一个补集$a^{\perp}\in L$，满足$a\wedge a^{\perp}=0$和$a\vee a^{\perp}=1$，则偏序集(格)L被补充。此情形下的操作$^{\perp}$是一个补集。一个被补充的格L被称为正交补格或一个正交格，若对于所有的a，$b\in L$有：$a\leqslant b$蕴涵着$b^{\perp}\leqslant a^{\perp}$；且$a^{\perp\perp}=a$。此情形下的操作$^{\perp}$是一个正交补。

7.4.3 分配性

格L是满足分配律的，若对所有的a，b，$c\in L$有：$a\wedge(b\vee c)=(a\wedge b)\vee(a\wedge c)$，类似地用$\vee$代替$\wedge$也成立。一般来说，格不是分配性的。格$L$的中心$Z(L)$是所有满足对任意的$p\in L$有$p=(p\wedge z)\vee(p\wedge z^{\perp})$的所有$z\in L$的集合。$Z(L)$是$L$的一个分配子格。

7.4.4 直积和可约性

令L_1和L_2是正交格。则我们能够构造第三个正交格L，它是L_1和L_2的“直积”。作为一个集合，L是集合L_1和L_2的直(笛卡尔)积。则定义对所有的a，$b\in L$，其中$a=(a_1, a_2)$和$b=(b_1, b_2)$，仅当$a_1\leqslant b_1$和$a_2\leqslant b_2$时有$a\leqslant b$。交集、并集和正交补集则以分量方式(从而)类似地定义。一个正交格L是“不可约化的”，若它不与非平凡正交格的直积同构的话。

7.4.5 基元性和覆盖特征

格的“顶”(或“单位元”)(如果存在的话)是元素$\mathbb{I}\in L$，满足对于所有的$a\in L$有$a\leqslant\mathbb{I}$。格的“底”(或“0”)(如果存在的话)是元素$0\in L$，满足对所有的$a\in L$有$0\leqslant a$。偏序集L中的基元是一个非零元素$a\in L$，满足对任意的$b\in L$，若$b\leqslant a$则要么$b=0$要么$b=a$。偏序集是基元性的，若每个非零元素包含(在$\leqslant$下)一个基元。最后，格是完备的，若每个L元素的集合都有一个交集和并集。在一个完备的基元格中，每个元素(除了0)要么是一个基元，要么是基元的并集。

最后，我们说一个格L拥有覆盖特性，若对每个原子$a\in L$和任意的$b\in L$，其中$a\wedge b=0$，$a\vee b$覆盖了b，即没有一个严格处于b和$a\vee b$之间的元素。

7.5 拓扑和测量

7.5.1 拓扑空间

一个拓扑空间是一个集合 S，和一个 S 子集的集合 T，满足：$\emptyset$，$S \in T$（其中 $\emptyset$ 是空集）；T 中任意集合的结合在 T 中；T 中任意集合对的交集在 T 中。T 是 S 的拓扑。T 的元素是开集，它们（在 S 中）的补集是闭集。从一个拓扑空间到另一个拓扑空间的函数 f 是连续的，若每个开集在 f 下的逆映象仍然是一个开集。

拓扑 T 的基 B 是 T 中的一个开集的聚集，满足 T 中的每个开集可以写作为 B 的元素的一个交集。一个常见的例子是具有范的空间（如向量空间），在那里可以将基定义为是"开球"的聚集，即形如 $\{x \mid \|x-y\| < \epsilon\}$ 的集合的聚集，其中 x 和 y 是空间中的点（例如向量），ϵ 是实数。

一个拓扑空间是紧致的，若点的每个序列都有一个收敛于空间中某个点的子序列。空间是局域紧致的，若基本上空间中的每个点有一个紧致的邻域——大体看来，空间的每个小部分都"看起来像"紧致空间的一小部分。

给定两个拓扑空间 S_1 和 S_2，我们可以将它们的笛卡尔积 $S_1 \times S_2$ 构造为集合。则我们将此笛卡尔积上的积拓扑定义如下。令 B_1 和 B_2 分别是 S_1 和 S_2 上的拓扑基，在 $S_1 \times S_2$ 上的积拓扑的基是笛卡尔积 $B_1 \times B_2$。（在多于两个空间的笛卡尔积上对积拓扑的定义更为复杂，但这里我们并不需要它。）

7.5.2 流形

一个流形是一个"局域欧几里得的"拓扑空间，意味着在每个点附近都有一个与 $\mathbb{R}^n$（对某些 n 而言——且 n 是流形的维度）中的开单元球在拓扑上相同的邻域。流形 M 的一个开集 S，与在 S 和 $\mathbb{R}^n$ 的开集间的同胚一起，称为一个坐标图表（chart）。覆盖 M 的图表的一个集合是一个图册（atlas）。现在考虑区域 S，其中两个图表有重叠。因而我们有从 S 到 $\mathbb{R}^n$ 的两个不同映射，它们定义了从 $\mathbb{R}^n$ 的一个子集（第一个图表应用于 S 的范围）到 $\mathbb{R}^n$ 的某个其他子集（第二个图表应用于 S 的范围）的映射 μ。如果在图册中所有重叠的图表生成的所有 μ 都是无限次可微的，则该流形是一个光滑流形。

7.5.3　弱算子拓扑

希尔伯特空间$\mathcal{H}$上算符的弱算子拓扑，是$\mathcal{H}$上有界算子集合$\mathcal{B}(\mathcal{H})$上的最弱拓扑，使得映射$F \mapsto \langle w, Fv \rangle$对任意矢量$v$，$w \in \mathcal{H}$和任意的$F \in \mathcal{B}(\mathcal{H})$是连续的。在弱算子拓扑中，算子的序列$\{F_n\}$收敛于算子$F$，仅当对每个$v$，$w \in \mathcal{H}$，$|\langle w, F_n v \rangle - \langle w, Fv \rangle|$收敛于0。

7.5.4　勒贝格测度

在实线$\mathbb{R}$上，勒贝格测度是将通常距离(间隔的大小)的测量向更复杂的点的集合的自然推广。例如，给定任意开集S，它是不相交间隔的交集，S的勒贝格测度是间隔大小的和。$\mathbb{R}$中单个点的任意可数交集的勒贝格测度为零。该测量能以明显的方式拓展到$\mathbb{R}^3$中的体积上。

7.5.5　波莱尔集

实数的波莱尔集定义如下：给定某个集合S，在S上的σ代数是在补集、可数并集和可数交集下闭合的S的一族子集；在$\mathbb{R}$上的波莱尔代数是包含$\mathbb{R}$开集的最小σ代数(人们必须表明确实存在一个最小的)。实数的一个波莱尔集是$\mathbb{R}$上波莱尔代数的一个元素。注意不是实数的每个子集都是一个波莱尔集，虽然不是波莱尔集的那些子集多少有些怪异。所有的开集和闭集都是波莱尔集。波莱尔代数(进而波莱尔集合)的重要性在于仅是特定的测量理论结果才适用于它们。另一方面，在许多情形下人们能够将重要结果和定义推广到更宽泛的集合种类，例如，推广到在连续函数下是一个波莱尔集合映象的所有集合。但是，我们不打算继续关注这些。

7.5.6　概率测度

令X是基本事件的一个集合(一个“样本空间”)，$\mathcal{A}$是X上的σ代数。概率理论的(柯尔莫哥洛夫)公理则可以表述如下。令p：$\mathcal{A} \to [0, 1]$是一个从$\mathcal{A}$到区间$[0, 1]$的映射，此p是$\mathcal{A}$上的一个概率测度，仅当：(正规化)$p(X) = 1$；(非)对任意的$E \in \mathcal{A}$有$p(E') = 1 - p(E)$；(相加性)对$\mathcal{A}$的元素的任意可数的不相交序列$\{E_k\}$，有$p(\sqcup_k E_k) = \sum_k p(E_k)$。

7.6 群

7.6.1 群和同构

群是一个非空集合$\mathcal{G}$，和一个此集合上的二元运算$*$(称为“积”)，满足：(结合率)对所有的a，b，$c \in \mathcal{G}$有$(a*b)*c=a*(b*c)$；(单位元)存在一个元素$e \in \mathcal{G}$使得对任意的$a \in \mathcal{G}$有$e*a=a*e=a$；(逆)对所有的$a \in \mathcal{G}$，存在$b \in \mathcal{G}$使得$a*b=b*a=e$(此b通常记作a^{-1})；(闭合)对所有的a，$b \in \mathcal{G}$有$a*b \in \mathcal{G}$。除了逆的存在，满足所有这些特征的结构称为一个“半群”。若对所有的a，$b \in \mathcal{G}$，有$m(a*b)=m(a)*m(b)$，则从一个群$\mathcal{G}$到另一个群$\mathcal{G}'$的映射m：$\mathcal{G} \to \mathcal{G}'$是一个同态。(注意左式上的积是在$\mathcal{G}$中取的，而右式中的积是在$\mathcal{G}'$中取的。)由此可得，$m$保留了逆还有从$\mathcal{G}$中的单位元向$\mathcal{G}'$中单位元的映射。若映射$m$是同态的且是一一映射的，则它是一个同构。从$\mathcal{G}$向其自身的同构是$\mathcal{G}$的自同构。通常群积算子$*$是被略去不写的；因此我们之后将把$a*b$写作$ab$，诸如此类。

7.6.2 子群和积

给定一个群$\mathcal{G}$，子群$\mathcal{H}$是正规的，当且仅当对任意的$g \in \mathcal{G}$有$g\mathcal{H}g^{-1} \subset \mathcal{H}$(其中$g\mathcal{H}g^{-1}$是集合$\{ghg^{-1} \mid h \in \mathcal{H}\}$)。给定两个群$\mathcal{H}$和$\mathcal{K}$，群$\mathcal{G}$是它们的直积，当且仅当：(i)$\mathcal{H}$和$\mathcal{K}$是(同构于)$\mathcal{G}$的正规群；(ii)$\mathcal{H} \cap \mathcal{K}=e$，是$\mathcal{G}$中的单位元；(iii)作为一个集合，$\mathcal{G}$(同构于)$\{hk: h \in \mathcal{H}, k \in \mathcal{K}\}$。我们写作$\mathcal{G}=\mathcal{H} \times \mathcal{K}$。构造群直积的常见方式是令$\mathcal{G}$是作为集合的$\mathcal{H}$和$\mathcal{K}$笛卡尔积的一个集合，且对所有的$h$，$h' \in \mathcal{H}$和$k$，$k' \in \mathcal{K}$定义$(h, k)(h', k')=(hh', kk')$。(注意在此情形下$\mathcal{H}$同构于其元素形如$(h, e)$的子群，其中$h \in \mathcal{H}$，$e$是$\mathcal{G}$中的单位元，且$\mathcal{K}$也类似。在此情形中$\mathcal{H}$和$\mathcal{K}$也都是正规子群。)若仅$\mathcal{K}$是$\mathcal{G}$的正规子群，而$\mathcal{H}$是一个非正规子群，则$\mathcal{G}$是$\mathcal{K}$对$\mathcal{H}$的半直积，我们写作$\mathcal{G}=\mathcal{H} \ltimes \mathcal{K}$。

7.6.3 陪集和商

令$\mathcal{H}$是$\mathcal{G}$的一个子群。我们将$\mathcal{G}$中$\mathcal{H}$的(左)陪集定义为集合$\{g\mathcal{H} \mid g \in \mathcal{G}\}$，其中$g\mathcal{H}=\{gh \mid g \in \mathcal{H}\}$。(注意对某些$g$，$g' \in \mathcal{G}$，我们会有$g\mathcal{H}=g'\mathcal{H}$。)$\mathcal{G}$中$\mathcal{H}$的(左)陪集分割了$\mathcal{G}$。它们本身形成了一个群和乘法规则$(g\mathcal{H})(g'\mathcal{H})=(gg')\mathcal{H}$，该群被称为$\mathcal{H}$对$\mathcal{G}$的商，典型地写作$\mathcal{G}/\mathcal{H}$。

7.6.4 连续参数化群

一个群$\mathcal{G}$是通过下述方法连续参数化的，即群作为用 a 标示的集合，且 $g_0 = I$（$\mathcal{G}$上的单位元）；$g_{a+b} = g_a g_b$；极限 $\lim_{n\to\infty} a_n = b$ 蕴涵着$\lim_{n\to\infty} g_{a_n} = g_b$。在第三个条件中，右边的极限要求群是一个拓扑群，即群也是一个拓扑空间。

7.6.5 李群

李群是一个同样为群的光滑流形（第 7.5.2 节），其中乘和逆的群操作是流形上的连续映射。（事实上，李群常常被定义为分析流形，但在这里我们并不考虑这些。）

7.6.6 向量空间表示

群理论的一个重要定理是，每个群$\mathcal{G}$是（同构于）在某个集合上置换群的一个子群。$\mathcal{G}$的另一个常见表示类型是一个向量空间表示，一个从$\mathcal{G}$到 GL（V）的群同态，向量空间 V 的“一般线性群”变换，即在 V 上的可逆线性算子群。

仅当 m 是一一映射时，群$\mathcal{G}$的表示 m：$\mathcal{G}\to$GL（V）是忠实的。非忠实表示忽略了被表示群的结构。表示也能够在下述意义上引入结构。在群下不变的 V 的真子空间（即群是 W 的一个自同构）W 携带了$\mathcal{G}$的一个“子表示”，在 V 上$\mathcal{G}$的表示对 W 的限制本身就是$\mathcal{G}$的一个表示这一意义上。当$\mathcal{G}$的一个表示 m 的（真）子表示存在时，m 被称为“可约的”。否则，它是“不可约的”。如果一个表示是可约的，则它所在的向量空间在某种意义上“大于”用以表示群“所需”的向量空间。

7.6.7 群作用

对任意集合 S，$\mathcal{G}$在 S 上的群作用是一个映射μ：$\mathcal{G}\times S\to S$，满足：(a)对所有的 g，$h\in\mathcal{G}$和 $s\in S$ 有 $\mu(g,\ \mu(h,\ s)) = \mu(gh,\ s)$；(b)对所有的 $s\in S$ 有 $\mu(e,\ s) = s$（其中 e 是$\mathcal{G}$中的单位元）。$\mu(g,\ s)$常常写作 $\mu_g s$。每个 μ_g 事实上是 S 上的一一映射，从而我们也可以将一个群作用定义为从$\mathcal{G}$到 S 上的一一映射群的群同构。有时 μ_g 被称为“g 在 S 上的作用”。$\mathcal{G}$在 S 上的群作用 μ 是可传递的，当且仅当对任意的 s，$t\in S$，存在 $g\in\mathcal{G}$满足 $\mu_g x = y$。若$\mathcal{G}$和 S 都有一个拓扑结构，则若映射 μ 相对于积拓扑$\mathcal{G}\times S$ 是连续的话，$\mathcal{G}$在 S 上的作用是连续的。

7.6.8 酉表示

给定希尔伯特空间$\mathcal{H}$，$\mathcal{H}$上的任意幺正算符 U 实现了 GL（$\mathcal{H}$）的一个自同

构。特别地，映射 $\mathrm{GL}(\mathcal{H})$ 通过 $F \to U^{-1}FU$ 映射到自身，对于每个 $F \in \mathrm{GL}(\mathcal{H})$ 而言。特别地，注意：对任意的 F，$G \in \mathrm{GL}(\mathcal{H})$，可利用对每个可逆算子 A 和 B 有 $(AB)^{-1} = B^{-1}A^{-1}$ 的事实，$U^{-1}FGU = (U^{-1}FU)(U^{-1}GU)$ 和 $U^{-1}F^{-1}U = (U^{-1}FU)^{-1}$（假定 F 是可逆的）。此外，注意此映射是一对一的：对任意的算子 F，存在一个唯一的算子 G 满足 $U^{-1}GU = F$，即 $G = UFU^{-1}$。

同样注意每个幺正算符是 $\mathrm{GL}(\mathcal{H})$ 的一个元素。因此在这里我们有一个关于群的内在同构的普遍构造的范例，对某些 $g \in \mathcal{G}$ 和所有的 $h \in \mathcal{G}$，因为此同构人们通过 $h \mapsto ghg^{-1}$ 将群 $\mathcal{G}$ 映射到其自身。

7.6.9 诱导表示

给定群 $\mathcal{G}$ 和子群 $\mathcal{H}$，$\mathcal{H}$ 在向量空间 W 上的表示在如下意义上“诱导了”$\mathcal{G}$ 的一个表示。人们可以从 W 上的 $\mathcal{H}$ 表示来构造向量空间 V，该向量空间事实上是 W 的拷贝和 $\mathcal{G}$ 在 V 上的表示的直和，W 在 V 内的每个拷贝都携带了 $\mathcal{H}$ 的一个表示。

这里给出了对如何进行构造的粗略描述。令 σ 是 W 上 $\mathcal{H}$ 的一个表示。构建由 σ 诱导的 $\mathcal{G}$ 表示背后的主旨是构造一个向量空间 V，它是 W 拷贝的直和，即 $V = \bigoplus_n W_n$，其中 W_n 是 W 的一个拷贝，且 W 的每个拷贝对应于 $\mathcal{G}/\mathcal{H}$ 的一个元素。由 σ 诱导的 $\mathcal{G}$ 表示 ρ 的定义如下：对每个陪集 n，确定一个 $\mathcal{G}/\mathcal{H}$ 的表示 g_n。注意对任意的 $g \in \mathcal{G}$，对某些 $h \in \mathcal{H}$ 和某些 $m \in \mathcal{G}/\mathcal{H}$，有 $gg_n = g_m h$。令 w_n 是来自于 W_n 的任意向量，对应于（在某些同构下）$w \in W$。（注意，如果我们在每个 W_n 的基上定义一个算子，则通过线性我们对所有的 V 定义了算子。）通过 $\rho(g)w_n = (\sigma(h)w)_m$ 定义 $\rho(g)$（对任意 $g \in \mathcal{G}$），其中 h 和 m 已经在上面给出。右式的表述应当读作“令 $\sigma(h)$ 以与它作用于 $w \in W$ 相同的方式作用于 $w_n \in W_n$，则将结果映射到 W_m 中相应的向量上”。注意这整个的规定预设了在 W 和 W_n 之间的一个同构集合。最后，可以表明上述内容不依赖于对表示 g_n 的选择，因为不同的选择会产生一个同构表示。

致谢

感谢杰里米·巴特菲尔德和约翰·厄尔曼的评论和建议。同样感谢 2004 年 11 月匹兹堡大学讨论班参与者们的有益评论。

参考文献

[Aharonov and Albert, 1981] Yakir Aharonov and David Albert. Can we make sense of the measurement process in relativistic quantum mechanics? *Physical Review Letters*, 24: 359-370, 1981.

[Aharonov and Botero, 2005] Yakir Aharonov and Alonso Botero. Quantum averages of weak values. *Physical Review A*, 72: 052-111, 2005.

[Aharonov and Kaufherr, 1984] Yakir Aharonov and M. Kaufherr. Quantum frames of reference. *Physical Review D*, 30: 368-385, 1984.

[Aharonov *et al.*, 1987] Yakir Aharonov, David Albert, A. Casher, and Lev Vaidman. Surprising quantum effects. *Physics Letters A*, 124: 199-203, 1987.

[Aharonov *et al.*, 1993] Yakir Aharonov, Jeeva Anandan, and Lev Vaidman. Meaning of the wavefunction. *Physical Review A*, 47: 4616-4626, 1993.

[Albert and Loewer, 1988] D. Albert and B. Loewer. Interpreting the many worlds interpretation. *Synthese*, 77: 195-213, 1988.

[Alfsen and Shultz, 2003] Erik M. Alfsen and Frederic W. Shultz. *Geometry of State Spaces of Operator Algebras.* Birkhäuser, Boston, 2003.

[Atmanspacher, 2004] Harald Atmanspacher. Quantum approaches to consciousness. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy.* Stanford University, 2004. http://plato.stanford.edu/archives/win2004/entries/qt-consciousness/.

[Bacciagaluppi and Dickson, 1999] Guido Bacciagaluppi and Michael Dickson. Dynamics for modal interpretations. *Foundations of Physics*, 29: 1165-1201, 1999.

[Bacciagaluppi and Hemmo, 1996] Guido Bacciagaluppi and Meir Hemmo. Modal interpretations, decoherence, and measurements. *Studies in History and Philosophy of Modern Physics*, 27: 239-277, 1996.

[Bacciagaluppi, 2000] Guido Bacciagaluppi. Delocalized properties in the modal interpretation of a continuous model of decoherence. *Foundations of Physics*, 30: 1431-1444, 2000.

[Ballentine, 1970] L. Ballentine. The statistical interpretation of quantum mechanics. *Reviews of Modern Physics*, 42: 358-381, 1970.

[Barbour, 1989] J. Barbour. The Discovery of Dynamics: *A Study from a Machian Point of View of the Discovery and the Structure of Dynamical Theories.* Oxford University Press, Oxford, 1989.

[Barrett, 1999] J. Barrett. *The Quantum Mechanics of Minds and Worlds.* Oxford University Press, Oxford, 1999.

[Bedard, 1999] K. Bedard. Material objects in Bohm's interpretation. *Philosophy of Science*, 66: 221-242, 1999.

[Bell, 1964] John Bell. On the Einstein-Podolsky-Rosen paradox. *Physics*, 1: 195-200, 1964.

[Bell, 1990] John Bell. Against measurement. In A. Miller, editor, *Sixty-Two Years of Uncertainty*, pages 17-31. Plenum, New York, 1990.

[Beller, 1999] Mara Beller. *Quantum Dialogue.* University of Chicago Press, Chicago, 1999.

[Birkhoff and von Neumann, 1936] G. Birkhoff and J. von Neumann. The logic of quantum mechanics. *Annals of Mathematics*, 37: 823-843, 1936.

[Bohm, 1951] David Bohm. *Quantum Theory.* Prentice-Hall, Inc., Englewood Cliffs, NJ, 1951.

[Bohm, 1952] David Bohm. A suggested interpretation of the quantum theory in terms of 'hidden' variables, I and II. *Physical Review*, 85: 166-193, 1952.

[Bohr, 1935] Niels Bohr. Can quantum-mechanical description of physical reality be bonsidered complete? *Physical Review*, 48: 696-702, 1935.

[Bracken, 2003] A. J. Bracken. Quantum mechanics as an approximation to classical mechanics in Hilbert space. *Journal of Physics A*, 36: L329-L335, 2003.

[Brown and Pooley, 2002] Harvey Brown and Oliver Pooley. Relationalism rehabilitated? I: Classical mechanics. *British Journal for the Philosophy of Science*, 53: 183-204, 2002.

[Brown, 1986] Harvey Brown. The insolubility proof of the quantum measurement problem. *Foundations of Physics*, 16: 857-870, 1986.

[Bub and Clifton, 1996] Jeffrey Bub and Rob Clifton. A uniqueness theorem for interpretations of quantum mechanics. *Studies in History and Philosophy of Modern Physics*, 27B: 181-219, 1996.

[Bub, 1974] Jeffrey Bub. *The Interpretation of Quantum Mechanics.* D. Reidel, Dordrecht, 1974.

[Bub, 1977] Jeffrey Bub. Von Neumann's projection postulate as a possibility conditionalization rule in quantum mechanics. *Journal of Philosophical Logic*, 6: 381-390, 1977.

[Bub, 1997] Jeffrey Bub. *Interpreting the Quantum World.* Cambridge University Press, Cambridge, 1997.

[Busch *et al.*, 1995] P. Busch, M. Grabowski, and P. Lahti. *Operational Quantum Physics.* Springer-Verlag, Berlin, 1995.

[Busch, 2003] Paul Busch. Quantum states and generalized observables: A simple proof of Gleason's theorem. *Physical Review Letters*, 91: 1-4, 2003.

[Butterfield, 1992] Jeremy Butterfield. Bell's theorem: What it takes. *British Journal for the Philosophy of Science*, 43: 41-83, 1992.

[Butterfield, 1993] Jeremy Butterfield. Interpretation and identity in quantum theory. *Studies in History and Philosophy of Science*, 14: 443-478, 1993.

[Butterfield, 1995] Jeremy Butterfield. Worlds, minds, and quanta. *Aristotelian Society Supplementary Volume*, 69: 113-158, 1995.

[Cattaneo, 1979] U. Cattaneo. Mackey's imprimitivity theorem. *Commentari Mathematici Helvetici*, 54: 629-641, 1979.

[Clauser and Shimony, 1978] J. Clauser and A. Shimony. Bell's theorem: Experimental tests and implications. *Reports on Progress in Physics*, 41: 1881-1927, 1978.

[Clifton, 1995] Rob Clifton. Independently motivating the Kochen-Dieks modal interpretation of quantum mechanics. *Commentari Mathematici Helvetici*, 46: 33-57, 1995.

[Clifton, 2000] Rob Clifton. The modal interpretation of algebraic quantum field theory. *Physics Letters A*, 271: 161-177, 2000.

[Cohen and Svetlichny, 1987] D. Cohen and G. Svetlichny. Minimal supports in quantum logics. *International Journal of Theoretical Physics*, 27: 435-450, 1987.

[Cohen-Tannoudji *et al.*, 1977] C. Cohen-Tannoudji, D. Bernard, and F. Laloë. *Quantum Mechanics.* John Wiley & Sons, New York, 1977.

[Collier, 1999] J. D. Collier. Causation is the transfer of information. In H. Sankey, editor, *Causation, Natural Laws and Explanation*, pages 279-331. Kluwer, Dordrecht, 1999.

[Cushing and McMullin, 1989] J. T. Cushing and E. McMullin, editors. *Philosophical Consequences of Quantum Theory.* University of Notre Dame Press, Notre Dame, IN, 1989.

[Cushing *et al.*, 1996] J. T. Cushing, A. Fine, and S. Goldstein, editors. *Bohmian Mechanics and Quantum Theory: An Appraisal*, volume 184 of *Boston Studies in the Philosophy of Science.* Kluwer Academic Publishers, Boston, 1996.

[Cushing, 1994] J. T. Cushing. Quantum Mechanics: *Historical Contingency and the Copenhagen Hegemony.* University of Chicago Press, Chicago, 1994.

[dalla Chiara *et al.*, 2004] M. dalla Chiara, R. Giuntini, and R. Ghreechie. *Reasoning in Quantum Theory: Sharp and Unsharp Quantum Logics.* Kluwer Academic Publishers, Dordrecht, 2004.

[de Broglie, 1924] Louis de Broglie. *Recherches sur la Théorie des Quanta.* PhD thesis, Faculté des Sciences de Paris, 1924. Appears in Annales de Physique, series 10, vol III, pp. 22-128, 1925.

[de Broglie, 1927] Louis de Broglie. La nouvelle dynamique des quanta. In H. Lorentz, editor, *Rapports et Discussions du Cinquième Conseil de Physique Solvay*, pages 105-141, Paris, 1927. Gauthier-Villars.

[d'Espagnat, 1971] Bernard d'Espagnat. *The Conceptual Foundations of Quantum Mechanics.* W. A. Benjamin, Inc., Reading, MA, 1971.

[Dewdney *et al.*, 1986] C. Dewdney, P. Holland, and A. Kyprianidis. What happens in a spin measurement? *Physics Letters A*, 119: 259-267, 1986.

[DeWitt and Graham, 1973] B. S. DeWitt and N. Graham, editors. *The Many-Worlds Interpretation of Quantum Mechanics.* Princeton University Press, Princeton, 1973.

[Dickson, 1998] Michael Dickson. *Quantum Chance and Nonlocality.* Cambridge University Press, Cambridge, 1998.

[Dickson, 2000] Michael Dickson. Are there material objects in Bohm's theory? *Philosophy of Science*, 67: 704-710, 2000.

[Dickson, 2001] Michael Dickson. Quantum logic is alive $\wedge$ (it is true $\vee$ it is false). *Philosophy of Science*, 68: S274-S287, 2001.

[Dickson, 2002a] Michael Dickson. Bohr on Bell: A proposed reading of Bohr and its implications for Bell's theorem. In J. Butterfield and T. Placek, editors, *Proceedings of the NATO*

Advanced Research Workshop on Modality, Probability and Bell's Theorem, Amsterdam, 2002. IOS Press.

[Dickson, 2002b] Michael Dickson. The EPR experiment: A prelude to Bohr's reply to EPR. In Michael Heidelberger and Friedrich Stadler, editors, *History of Philosophy of Science-New Trends and Perspectives*, volume 9 of *Institute Vienna Circle Yearbook*, pages 263-276, Dordrecht, 2002. Kluwer.

[Dickson, 2004a] Michael Dickson. Quantum reference frames in the context of EPR. *Philosophy of Science*, 71: 655-668, 2004.

[Dickson, 2004b] Michael Dickson. A view from nowhere: Quantum reference frames and uncertainty. *Studies in History and Philosophy of Modern Physics*, 35: 195-220, 2004.

[Dieks and Vermaas, 1998] Denis Dieks and Pieter Vermaas, editors. *The Modal Interpretation of Quantum Mechanics.* The Western Ontario Series in Philosophy of Science. Kluwer Academic Publishers, Dordrecht, 1998.

[Dieks, 1988] Denis Dieks. The formalism of quantum theory: An objective description fo reality? *Annalen der Physik*, 7: 174-190, 1988.

[Diosi, 1992] L. Diosi. Quantum measurement and gravity for each other. In P. Cvitanovic, I. Pervical, and A. Wirzba, editors, *Quantum Chaos-Quantum Measurement.* Kluwer, Dordrecht, 1992.

[Dirac, 1930] P. A. M. Dirac, editor. *The Principles of Quantum Mechanics.* Clarendon Press, Oxford, 1st edition, 1930.

[DiSalle, 1990] Robert DiSalle. Conventionalism and the origins of the inertial frame concept. In Arthur Fine, Micky Forbes, and Linda Wessels, editors, *PSA 1990*, East Lansing, MI, 1990. The Philosophy of Science Association.

[DiSalle, 2002] Robert DiSalle. Space and time: Inertial frames. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy.* Stanford University, 2002. http://plato.stanford.edu/archives/sum2002/entries/spacetime-iframes.

[Dixmier, 1958] J. Dixmier. Sur la relation i(pq-qp) = 1. *Compositio Mathematica*, 13: 263-270, 1958.

[Donald, 1990] M. Donald. Quantum theory and the brain. *Proceedings of the Royal Society of London*, A427: 43-93, 1990.

[Dummett, 1976] Michael Dummett. Is logic empirical? In H. Lewis, editor, *Contemporary British Philosophy*. Allen and Unwin, London, 1976.

[Earman and Ruetsche, 2006] John Earman and Laura Ruetsche. Relativistic invariance and modal interpretations. *Philosophy of Science*, forthcoming, 2006.

[Einstein *et al.*, 1935] A. Einstein, B. Podolsky, and N. Rosen. Can quantum-mechanical description of physical reality be considered complete? *Physical Review*, 47: 777-780, 1935.

[Einstein, 1905] Albert Einstein. Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt. *Annalen der Physik*, 17: 132-149, 1905.

[Everett, 1957] H. Everett. Relative state formulations of quantum mechanics. *Reviews of Modern Physics*, 29: 454-462, 1957.

[Fine, 1973] Arthur Fine. Probability and the interpretation of quantum mechanics. *British Journal for the Philosophy of Science*, 24: 1-37, 1973.

[Fine, 1986] Arthur Fine. *The Shaky Game: Einstein, Realism and the Quantum Theory*. University of Chicago Press, Chicago, 1986.

[Fine, 1991] Arthur Fine. Inequalities for nonideal correlation experiments. *Foundations of Physics*, 21: 365-378, 1991.

[Finkelstein, 1962] D. Finkelstein. The logic of quantum physics. *Transactions of the New York Academy of Sciences*, 25: 621-637, 1962.

[Finkelstein, 1969] D. Finkelstein. Matter, space, and logic. In M. Wartofsky and R. Cohen, editors, *Boston Studies in the Philosophy of Science*, volume 5, pages 199-215. Humanities Press, New York, 1969.

[Friedman and Putnam, 1978] Michael Friedman and Hilary Putnam. Quantum logic, conditional probability, and interference. *Dialectica*, 32: 305-315, 1978.

[Ghriardi *et al.*, 1986] G.-C. Ghriardi, A. Rimini, and T. Weber. Unified dynamics for microscopic and macroscopic systems. *Physical Review D*, 34: 470-491, 1986.

[Ghriardi *et al.*, 1990] G.-C. Ghriardi, R. Grassi, and P. Pearle. Relativistic dynamic reduction models-General framework and examples. *Foundations of Physics*, 20: 1271-1316, 1990.

[Gibbons, 1987] P. Gibbons. *Particles and Paradoxes: The Limits of Quantum Logic*. Cambridge University Press, Cambridge, 1987.

[Gisin, 1984] N. Gisin. Quantum measurement and stochastic processes. *Physical Review Letters*, 52: 1657-1660, 1984.

[Gleason, 1957] A. Gleason. Measures on the closed subspaces of a Hilbert space. *Journal of Mathematics and Mecha-nics*, 6: 885-893, 1957.

[Greenberger *et al.*, 1989] D. M. Greenberger, M. A. Horne, and A. Zeilinger. Going beyond Bell's theorem. In M. Kafatos, editor, *Bell's Theorem, Quantum Theory, and Conceptions of the Universe*, pages 69-72. Kluwer, Dordrecht, 1989.

[Groenewold, 1946] H. J. Groenewold. On the principles of elementary quantum mechanics. *Physica*, 12: 405-460, 1946.

[Haag, 1992] Rudolf Haag. *Local Quantum Physics.* Springer-Verlag, Berlin, 1992.

[Halvorson and Clifton, 1999] Hans Halvorson and Rob Clifton. Maximal beable subalgebras of quantum mechanical observables. *International Journal of Theoretical Physics*, 38: 2441-2484, 1999.

[Halvorson, 2001] Hans Halvorson. On the nature of continuous physical quantities in classical and quantum mechanics. *Journal of Philosophical Logic*, 30: 27-50, 2001.

[Hardy, 1992] Lucien Hardy. Quantum mechanics, local realistic theories, and Lorentz-invariant realistic theories. *Physical Review Letters*, 68: 2981-2984, 1992.

[Hardy, 2001] Lucien Hardy. Quantum theory from five reasonable axioms. quant-ph/0101012, 2001.

[Hardy, 2002] Lucien Hardy. Why quantum theory? In J. Butterfield and T. Placek, editors, *Proceedings of the NATO Advanced Research Workshop on Modality, Probability and Bell's Theorem*, Amsterdam, 2002. IOS Press.

[Healey, 1989] Richard Healey. *The Philosophy of Quantum Mechanics: An Interactive Interpretation.* Cambridge University Press, Cambridge, 1989.

[Heisenberg, 1927] Werner Heisenberg. Ueber den anschaulichen Inhalt der quantentheoretischen Kinematik and Mechanik. *Zeitschrift für Physik*, 43: 172-198, 1927.

[Heisenberg, 1930] Werner Heisenberg. *Die Physikalischen Prinzipien der Quantenmechanik.* Hirzel, Leipzig, 1930. English translation *The Physical Principles of Quantum Theory* published by University of Chicago Press, Chicago.

[Hooker, 1975] C. A. Hooker, editor. *The Logico-Algebraic Approach to Quantum Me-*

chanics, volume 1. D. Reidel, Dordrecht, 1975.

[Hooker, 1979] C. A. Hooker, editor. *The Logico-Algebraic Approach to Quantum Mechanics*, volume 2. D. Reidel, Dordrecht, 1979.

[Howard, 2003] Don Howard. What makes a concept classical? http://www.nd.edu/dhoward1/Classcon.pdf, 2003.

[Jarrett, 1984] Jon Jarrett. On the physical significance of the locality conditions in the Bell arguments. *Noûs*, 18: 569-589, 1984.

[Jauch, 1968] J. M. Jauch. *Foundations of Quantum Mechanics.* Addison Wesley, Reading, MA, 1968.

[Joos and Zeh, 1985] E. Joos and H. D. Zeh. The emergence of classical properties through interaction with the environment. *Zeitschrift für Physik*, B59: 223-243, 1985.

[Joos, 1986] E. Joos. Quantum theory and the appearance of a classical world. In D. Greenberger, editor, *New Techniques and Ideas in Quantum Measurement Theory*, volume 480 of *Annals of the New York Academy of Science*, pages 6-13, New York, 1986. New York Academy of Science.

[Kochen and Specker, 1967] S. Kochen and E. P. Specker. The problem of hidden variables in quantum mechanics. *Journal of Mathematics and Mechanics*, 17: 59-87, 1967.

[Kochen, 1985] S. Kochen. A new interpretation of quantum mechanics. In P. Mittelstaedt and P. Lahti, editors, *Symposium on the Foundations of Modern Physics*, pages 923-928, Singapore, 1985. World Scientific.

[Landsman, 1998] N. P. Landsman. *Mathematical Topics Between Classical and Quantum Mechanics.* Springer-Verlag, New York, 1998.

[Leavens, 1990] C. Leavens. Transmission, reflection, and dwell times within Bohm's causal interpretation of quantum mechanics. *Solid State Communications*, 74: 923-928, 1990.

[Lewis, 1973] David Lewis. Causation. *Journal of Philosophy*, 70: 556-567, 1973.

[Ludwig, 1983] G. Ludwig. *Mathematical Foundations of Quantum Mechanics.* Springer-Verlag, New York, 1983.

[Ma *et al.*, 1998] C. Ma, E. F. Arias, T. M. Eubanks, A. L. Fey, A-M. Gontier, C. S. Jacobs, O. J. Sovers, B. A. Archinal, and P. Charlot. The international celestial reference frame as realized by very long baseline interferometry. *The Astronomical Journal*, 116: 516-

546, 1998.

[Mackey, 1949] G. W. Mackey. Imprimitivity for representations of locally compact groups. *Proceedings of the National Academy of Sciences*, 35: 156-162, 1949.

[Mackey, 1978] G. W. Mackey. *Unitary Group Representations in Physics Probability and Number Theory*, volume 55 of *Math Lecture Notes Series.* Benjamin Commings Publishing Company, Menlo Park, CA, 1978.

[Mackey, 1996] G. W. Mackey. The relationship between classical and quantum mechanics. In L. A. Coburn and M. A. Rieffel, editors, *Perspectives on Quantization.* American Mathematical Society, Providence, RI, 1996.

[Massimi, 2004] Michela Massimi. *Pauli's Exclusion Principle: The Origin and Validation of a Scientific Principle.* Cambridge University Press, Cambridge, 2004.

[Maudlin, 1994] Tim Maudlin. *Quantum Non-Locality and Relativity.* Blackwell, Oxford, 1994.

[Mermin, 1990] David Mermin. Simple unified form for the major no-hidden variables theorems. *Physical Review Letters*, 65: 3373-3376, 1990.

[Mittelstaedt, 1998] Peter Mittelstaedt. *The Interpretation of Quantum Mechanics and the Measurement Process.* Cambridge University Press, Cambridge, 1998.

[Pauli, 1958] W. Pauli. *Die allgemeinen Prinzipien der Wellenmechanik*, volume 1 of *Handbuch der Physik*, pages 1-168. Springer, Berlin, 1958.

[Pearle, 1989] P. Pearle. Combining stochastic dynamical state-vector reduction with spontaneous localization. *Physical Review A*, 39: 2277-2289, 1989.

[Piron, 1964] C. Piron. Axiomatique quantique. *Helvetica Physica Acta*, 37: 439-468, 1964.

[Piron, 1976] C. Piron. *Foundations of Quantum Physics.* W. A. Benjamin, Reading, MA, 1976.

[Pitowsky, 1991] Itamar Pitowsky. Bohm quantum potentials and quantum-gravity. *Foundations of Physics*, 21: 343-352, 1991.

[Primas, 1990] H. Primas. Induced nonlinear time evolution of open quantum objects. In A. Miller, editor, *Sixty-Two Years of Uncertainty*, pages 259-280. Plenum, New York, 1990.

[Putnam, 1969] Hilary Putnam. Is logic empirical? In *M. Wartofsky and R. Cohen*, volume

5 of *Boston Studies in the Philosophy of Science*, pages 181-206. Humanities Press, New York, 1969.

[Reichenbach, 1956] Hans Reichenbach. *The Direction of Time.* University of Los Angeles Press, Berkeley, 1956.

[Rellich, 1946] F. Rellich. Der Eindeutigkeitssatz fur die Losungen der Quantunmechanischen Vertauschungsrelationen. *Nachrichten Akad. Wiss. Göttingen*, Math. Phys. Klasse II, 11A: 107-115, 1946.

[Riesz and Sz. -Nagy, 1955] Frigyes Riesz and Béla Sz. -Nagy. *Functional Analysis.* Frederick Ungar Publishing Co. , New York, 1955.

[Rudin, 1973] W. Rudin. *Functional Analysis.* McGraw-Hill, New York, 1973.

[Sakurai, 1985] J. J. Sakurai. *Modern Quantum Mechanics.* Benjamin/Cummings, Menlo Park, CA, 1985.

[Salmon, 1984] Wesley Salmon. *Scientific Explanation and the Causal Structure of the World.* Princeton Univeristy Press, Princeton, NJ, 1984.

[Schiff, 1968] Leonard Schiff. *Quantum Mechanics.* McGraw-Hill, New York, 3rd edition, 1968.

[Schrödinger, 1935a] Erwin Schrödinger. Die gegenwartige Situation in der Quantenmechanik. *Naturwissenschaften*, 23: 807-812; 823-828; 844-849, 1935. Translated as 'The Present Situation in Quantum Mechanics' by John D. Trimmer, *Proceedings of the American Philosophical Society*, 124, 323-338.

[Schrödinger, 1935b] Erwin Schrödinger. Discussion of probability relations between separated systems. *Proceedings of the Cambridge Philosophical Society*, 31: 555-563, 1935.

[Scully and Walther, 1989] Marlan O. Scully and Herbert Walther. Quantum optical test of observation and complementarity in quantum mechanics. *Physical Review A*, 39: 5229-5236, 1989.

[Shimony, 1978] Abner Shimony. Metaphysical problems in the foundations of quantum mechanics. *Inernational Philosophical Quarterly*, 18: 3-17, 1978.

[Shimony, 1986] Abner Shimony. Events and processes in the quantum world. In R. Penrose and C. Isham, editors, *Quantum Concepts in Space and Time*, pages 182-203. Clarendon Press, Oxford, 1986.

[Solèr, 1995] M. P. Solèr. Characterization of Hilbert spaces with orthomodular spaces. *Communications in Algebra*, 23: 219-243, 1995.

[Stone, 1932] Marshall Stone. On one-parameter unitary groups in Hilbert space. *Annals of Mathematics*, 33: 643-648, 1932.

[Summers, 2001] Stephen Summers. On the Stone-von Neumann uniqueness theorem and its ramifications. In Miklos Rédei and Michael Stöltzner, editors, *John von Neumann and the Foundations of Quantum Physics*, volume 8 of *Institute Vienna Circle Yearbook*, pages 135-152, Dordrecht, 2001. Kluwer.

[Szabó and Fine, 2002] Lószló Szabó and Arthur Fine. A local hidden variable theory for the GHZ experiment. *Physics Letters A*, 295: 229-240, 2002.

[Tanona, 2002] Scott Tanona. *From Correspondence to Complementarity: The Emergence of Bohr's Copenhagen Interpretation of Quantum Mechanics.* PhD thesis, Indiana University, 2002.

[Tanona, 2006] Scott Tanona. Theory, coordination, and empirical meaning in modern physics. In Mary Domski and Michael Dickson, editors, *Synthesis and the Growth of Knowledge.* Open Court, Peoria, IL, 2006.

[Uffink, 1994] Jos Uffink. Two new kinds of uncertainty relations. In D. Han, editor, *Proceedings of the Third international Workshop on Squeezed States and Uncertainty Relations*, volume 3270 of *NASA Conference Publications*, pages 155-161, Washington, DC, 1994. NASA.

[Uffink, 1999a] Jos Uffink. How to protect the interpretation of the wave function against protective measurements. *Physical Review A*, 60: 3474-3481, 1999.

[Uffink, 1999b] Jos Uffink. The principle of the common cause faces the Bernstein paradox. *Philosophy of Science*, 66: S512-S525, 1999.

[Valentini, 1991a] A. Valentini. Signal-locality, uncertainty and the subquantum J-theorem. I. *Physics Letters A*, 156: 5-11, 1991.

[Valentini, 1991b] A. Valentini. Signal-locality, uncertainty and the subquantum J-theorem. II. *Physics Letters A*, 158: 1-8, 1991.

[van Fraassen, 1972] Bas van Fraassen. A formal approach to the philosophy of science. In R. Colodny, editor, *Paradigms and Paradoxes: The Philosophical Challenge of the Quantum*

Domain, pages 303-366, Pittsburgh, 1972. University of Pittsburgh Press.

[Varadarajan, 1985] V. S. Varadarajan. *Geometry of Quantum Theory*. Springer-Verlag, Berlin, 1985.

[Vermaas, 2000] Pieter Vermaas. *A Philosopher's Understanding of Quantum Mechanics: Possibilities and Impossibilities of a Modal Interpretation*. Cambridge University Press, Cambridge, 2000.

[von Neumann, 1931] J. von Neumann. Die Eindeutigkeit der Schrödingerschen Operatoren. *Mathematische Annalen*, 104: 570-578, 1931.

[von Neumann, 1932] J. von Neumann. *Mathematische Grundlagen der Quantenmechanik*. Springer, Berlin, 1932. English translation *Mathematical Foundations of Quantum Theory* published by Princeton University Press, Princeton, 1955.

[Walborn *et al.*, 2002] S. P. Walborn, M. O. Terra Cunha, S. Pádua, and C. H. Monken. Doubleslit quantum eraser. *Physical Review A*, 65: 033-818, 2002.

[Wigner, 1931] Eugene P. Wigner. *Gruppentheorie und ihre Anwendung auf die Quantenmechanik der Atomspektren*. Vieweg, Braunschweig, 1931.

[Wigner, 1961] Eugene P. Wigner. Remarks on the mind-body question. In I. J. Good, editor, *The Scientist Speculates*, pages 284-302, London, 1961. Heinemann.

第五章　在经典与量子之间

N. P. 兰兹曼

"但最糟糕的事情是，我确实不能明确(矩阵力学)向经典理论的过渡。"见海森堡(Heisenberg)给泡利(Pauli)的信(1925年10月23日)。[①]

"亨德里克·洛伦兹(Hendrik Lorentz)把建立经典与量子理论之间正确关系作为之后研究的最基本的问题。该问题困扰着他，同样也困扰着普朗克(Planck)。"见[Mehra and Rechenberg(梅拉和雷琴堡)，2000，721]。

"因而量子力学在物理学理论中占据着一个非常特殊的位置：它涵盖了作为一个极限情形的经典力学，然而同时它自身的构建又需要该极限情形。"见[Landau and Lifshitz(朗道和利弗席兹)，1977，3]。

1. 引言

多数现代物理学家和哲学家们承认，适当的量子力学解释

① "Aber das Schlimmste ist, daβ ich über den übergang in die klassische Theorie nie Klarheit bekommen kann"(德语，意同上引文，译者注)，见[Pauline(波林)，1979，251]。

应该至少满足两个标准。第一，它必须要阐明其数学体系的物理意义，并由此获得理论的经验内容。所有量子理论的创始者们都清楚地认识到了这一点（我们仅以一种复述的方式指出）。① 第二（这是本论文的主题），它必须至少说明经典世界的表象。② 正如我们上面第二段引文中所表明的，普朗克看到了它指出的困难，且他首次注意到了他的黑体辐射公式在高温处的极限趋同于经典的表述。虽然玻尔（Bohr）相信量子力学应该通过经典物理学来解释，但在理论的创立者中，他似乎是唯一对从量子理论中得到经典物理学这一问题缺乏领会的。然而，通过对应原理（玻尔是为了回答上述第一个问题而非第二个而提出的），玻尔做出了对该问题最为深刻的贡献之一。海森堡最初认识到了该问题，但却错误地相信他已在关于不确定关系的著名论文中解决了它。③ 爱因斯坦（Einstein）出了名地不相信量子理论的基础本性，而薛定谔（Schrödinger）从一开始就意识到了该问题，之后用他极其著名的猫来突出这一问题，且在他职业生涯的不同阶段为该问题的解决做出了重要的技术贡献。埃伦费斯特（Ehrenfest）表述了用他名字命名的著名定理。冯·诺伊曼（von Neumann）也看到了困难所在，并通过他对量子力学中测量过程的著名分析来解决这一困难。

这一问题事实上甚至比量子理论创立者们预见到的更为严重。薛定谔猫在

① 量子理论的历史在许多书中都有表述。最详细的介绍是[Mehra and Rechenberg，1982—2001]，但有一些篇幅较小的作品也与该多卷系列同样优秀，如[Jammer（雅默），1966；van der Waerden（范德瓦尔登），1967；Hendry（亨得利），1984；Darrigol（达里戈尔），1992]和[Beller（贝莱尔），1999]。更多的信息也可以在传记中找到，如[Heisenberg，1969；Pais（派斯），1982；Moore（摩尔），1989；Pais，1991；Cassidy（卡西迪），1992；Heilbron（海伯伦），2000；Enz（恩兹），2002]，等等，也参见[Pauli，1979]。尤根·雷恩（Jurgen Renn）领导的关于矩阵力学历史的一项新计划正在进行。

② 哥本哈根解释表明了这两点是非常不同的，该解释专门针对第一点而完全忽略第二点。虽然如此，在量子力学的其他绝大多数解释中，致力于满足所涉及的两个标准的不同机制间存在实质性的交叠。

③ “人们可以看到从微观力学到宏观力学的过渡现在非常容易理解：经典力学完全是量子力学的一部分。”见海森堡给玻尔的信，1927年3月19日，正好在3月23日海森堡[1927]投稿之前。来自《量子物理学历史档案》（*Archives for the History of Quantum Physics*）中“玻尔科学通信”（Bohr's Scientific Correspondence）。

实验上的实现让大多数物理学家们感到比原先预想的要轻松[Leggett(莱格特),2002；Brezger *et al.*(布瑞格等)，2002；Chiorescu *et al.*(秋瑞库等)，2003；Marshall *et al.*(马歇尔等)，2003；Devoret *et al.*(德沃尔特等)，2004]。此外，棘手的叠加绝不仅限于物理实验室，由于其混沌运动，土星的卫星亥伯龙(尺度约为纽约的大小)被估计会在20年内散布在它的轨道上，如果把它处理为一个孤立的量子力学波包的话[Zurek and Paz(朱瑞克和帕兹)，1995]。而且，退相干理论学家们已经说明了“测量”不仅是由实验物理学家们在他们实验室中开展的过程，而且要是没有人类干预的话无时无刻不在自然中发生。在概念方面，如玻姆(Bohm)、贝尔(Bell)和他们的追随者是一方面，量子宇宙学家们是另一方面的不同派别都论证道，在客体和观测者之间的“海森堡界线(Heisenberg cut)”可能不能够成为物理学基本理论的基础。① 在过去的几十年里，这些以及其他显著的见解已经引起了量子力学解释，特别是引起了用量子力学来说明经典物理学对这些问题的重要性的广泛关注。

下面将对这些观点进行详细讨论，而且事实上我们对经典和量子力学间关系的讨论将部分是历史性的。然而，除了那些(历史性的)之外将是技术化的和数学上严谨的论述。由于手头的问题是如此复杂，以至于在该领域中不严谨的数学几乎肯定会导致不可靠的物理学和概念上的混乱(尽管在理论物理学中到处有欠缺数学的人取得不可否认的成功)。除了冯·诺伊曼，量子力学先驱者们并不持有这一态度，虽然要承认他们的许多理念仍然是当前讨论的核心，但是这些理念本质上却没有解决这一问题。因而我们假定读者熟悉量子力学的希尔伯特(Hilbert)空间形式体系，② 并且在论文的某些部分(特别是第6节和第4节中的部分内容)同样需要熟悉 C^* 代数的基本理论和它在量子理论中的应

---

① 更不用说由量子宇宙学家们提出的从某种量子引力理论得到经典时空的问题了。这当然也是从量子理论中得到经典物理学总纲领中的一部分，但遗憾的是本文不会讨论它。

② 除了成熟的代表作，如[Mackey(麦基)，1963；Jauch(约赫)，1968；Prugovecki(普如古维奇)，1971；Reed and Simon(里德和西蒙)，1972]或[Thirring(西凌)，1981]。读者可以参考更新的书，如[Gustafson and Sigal(古斯塔夫森和西加尔)，2003]或[Williams(威廉姆斯)，2003]，也见[Dickson(迪克森)，2005]。

用。① 除此之外，预先了解一些量子理论的概念问题是有益的。②

哪一种理念解决了用量子理论说明经典世界表象的问题？在我们看来，没有。虽然自量子力学创立之初就提出了许多新的理念，且几乎可以肯定这些理念在问题的最终解决中将发挥作用，而应该总会找到最终解决。这些理念无疑包括：

·小的普朗克常数极限$\hbar \to 0$（微局部分析数学领域时代的到来）；

·具有自由度N的大系统的极限$N \to \infty$（仅在C^*代数方法出现之后以一种严肃的方式被研究）；

·退相干和一致历史。

数学上，第二个极限可以看作是第一个极限的特殊情形，尽管其背后的物理情况应该是非常不同的。不管怎样，在详细分析之后我们得到的结论是，上述三个理念单独任何一个并不能够说明经典世界，但是通过结合这三者还是有些许希望，人们在未来有可能这样做。

由于本论文的论题是未竟的事业，到目前为止人们对于经典与量子理论之间的关系，有可能采用一些内在一致却相互矛盾的哲学立场。有两种极端的立场，不论是否坚持其中之一，记住它们总是有用的，它们是：

1. 量子理论是基本的和普遍有效的，经典世界仅拥有"相对的"或"视角化的"存在性；

2. 量子理论是近似的和导出的理论，可能是不正确的，经典世界绝对地存在。

有如下一种立场，我们对测量问题③的最新理解致使该立场内在不一致。

---

① 对C^*代数的物理倾向性介绍见[Davies（戴维斯），1976；Roberts and Roepstorff（罗伯茨和勒普斯托夫），1969；Primas（普利马斯），1983；Emch（埃姆什），1984；Strocchi（斯多奇），1985；Sewell（休厄尔），1986；Roberts，1990；Haag（哈格），1992；Landsman（兰兹曼），1998；Araki（荒木），1999]和[Sewell，2002]。权威的数学教科书包括[Kadison and Ringrose（卡迪孙与林罗斯），1983；1986]和[Takesakil（竹崎），2003]。

② 值得信赖的著作包括，如[Scheibe（沙伊贝），1973；Jammer，1974；van Fraassen（范·弗拉森），1991；d'Espagnat（德斯帕那特），1995；Peres（佩雷斯），1995；Omnes（翁内斯），1994；1999；Bub（巴布），1997]和[Mittelstaedt（米特斯泰特），2004]。

③ 特别见[Mittelstaedt，2004]

3. 量子理论是基本的和普遍有效的，且经典世界(仍)绝对地存在。

在某种意义上，立场1源自于海森堡[1927]，但在当代却开始于埃弗雷特(Everett)[1957]。[①] 目前，绝大多数的退相干理论学家、一致历史学家和模态解释者似乎都支持这一点。立场2无疑有着很长且强大的渊源，包括爱因斯坦、薛定谔和贝尔等人。最近的支持来自莱格特和“自发塌缩”理论学家如佩尔(Pearle)、吉拉迪(Ghirardi)、瑞米尼(Rimini)、韦伯(Weber)和其他人。正如我们将在第3节中看到的，玻尔的立场不能依照这些条件来分类，我们的三种立场具有一种本体论的特性，他大概认为每一种都不具有吸引力。

当然，我们必须确定所涉及的术语指的是什么。对于量子理论，我们指的是包含本征向量—本征值联系的标准量子力学。[②] 量子力学的模态解释[Dieks(狄克斯)，1989a；1989b；van Fraassen，1991；Bub，1999；Vermaas(弗马斯)，2000；Bene and Dieks(贝内和狄克斯)，2002；Dickson，2005]否认该联系，使得其立场接近或等同于立场1。当我们一般地谈论量子理论时，我们既不认可也不否认投影假设。

说“经典世界”指的是什么有点困难。在当前的讨论中，我们显然不能把经典世界定义为独立于观测而存在的世界——正如玻尔那样，见3.1——但是它也不能被看作由现有经典物理学规律描述的那部分世界。因为，若立场1是正确的，则这些定律仅是近似有效的，如果可能的话。因而，我们将其简单地表述如下：

经典世界是观测揭示给我们的那样，其行为——在适当的准确性下——满足经典物理学规律。

---

① 注意，虽然立场1绝不意味着所谓的多世界解释，在我们看来后者同样“仅是无意义的词汇大杂烩”[Leggett，2002]。

② 令A是希尔伯特空间$\mathcal{H}$上的自伴算符，有相应的投影值测量$P(\Delta)$，$\Delta\subset\mathbb{R}$，故有$A=\int dP(\lambda)\lambda$。本征向量—本征值联系规定，当且仅当$A$确定地取$\Delta$中的某个值时，系统的态$\Psi$处于$P(\Delta)\mathcal{H}$。特别地，若$\Psi$是本征值为$\lambda$的$A$的本征向量时，故有$P(\{\lambda\})\neq 0$和$\Psi\in P(\{\lambda\})\mathcal{H}$，则$A$在态$\Psi$中取值$\lambda$的概率为1。一般地，在态$\Psi$中，可观测量$a$取$\Delta$中的某个值(“在测量时”)的概率$p_{\Psi}(\Delta)$由玻恩—冯·诺伊曼(Born-von Neumann)规则$p_{\Psi}(\Delta)=(\Psi,\ P(\Delta)\Psi)$给出。

对于“在适当的准确性下”意味着什么应该没有多少怀疑的空间：灰颜色的存在并不意味着黑和白不存在！

我们可以将经典世界的绝对存在性按照玻尔的方式定义为其存在性独立于观测者或测量装置。与摩尔[1939]关于外部世界存在的论证相对比：

> 如何来证明？通过举起我的两只手，当我用右手做特定的手势时说“这是一只手”，当我再用左手做特定的手势时补充说“这是另一只”。

那些坚持立场1的人们主张说经典世界仅作为相对于特定具体要求的表象而存在，其中涉及的具体要求可以是一个观测者（海森堡），一组观测者和态（如在退相干理论中），也可以是由特定的一致历史集合定义的某个宇宙粗粒，等等。若观测者这一概念是在充分抽象和普遍的意义上理解的，则同样可以构造立场1，宣称经典世界仅从观测者（或相应的可观测量类别）的视角而言存在。① 例如，薛定谔猫“悖论”在引入适当的视角时就马上消解了，参见6.6节。

另一方面，那些坚持立场2的人们相信，经典世界在绝对意义上存在（正如摩尔那样）。因而，立场2类似于常识实在论，虽然在立场1与2之间的区分很大程度上与科学实在论问题无关。② 因为立场1的拥护者们通常仍然相信一些观测者无关实体的存在（即在量子领域的某处），但否认该实体组成了我们周围的观测世界。这为经典世界这一重要的概念作出了一个相当模糊的解释：除了围绕多世界解释无意义的讨论之外，一个有趣的结论是这样的见解，即若量子力学是基本的，则经典世界的观念内在地是不清晰的和粗略的。因此，在这点上太过于精确就会事与愿违。③

---

① 术语“视角化的(perspectival)”是理查德·希利(Richard Healey)向作者提议的。

② 此语境中关于实在论的更详尽讨论见[Landsman，1995]。词语“客观的”和“主观的”作出的区分似乎是无益的：陈述“我的孩子们是世界上最可爱的人”乍一看是主观的，但可以通过“柯拉斯·兰兹曼发现他的孩子们是世界上最可爱的人”的重构转换为客观的陈述。类似地，命题（或许是由于退相干）“局域的观测者发现世界是经典的”是完全客观的，虽然它描述了一个主观经历。也见[Davidson（戴维森），2001]。

③ 参见[Wallace（华莱士），2002；2003]，同样见[Butterfield（巴特菲尔德），2002]。这一点在玻尔和海森堡那里仍然成立，见[Scheibe，1973]。

虽然立场1被立场2的拥护者们认为是防守性的，如果不是因为胆怯的话，但一个非常有意义的数学事实是，迄今为止它似乎得到了量子力学数学体系的支持。在对他称为“为了一切实用目的(FAPP = For All Practical Purposes)”的测量问题求解(和对经典世界是量子理论的表象进行解释的更普遍尝试)的嘲讽中，贝尔[1987；2001]和在他之后的其他人错误地把一种深刻的认识论立场当作是一种可怜的防御性举措。① 事实上，我们要将立场2推荐给胆小者，因为证实或证伪立场1似乎都是整场争论真正的挑战，我们把本论文中的技术化内容看作是对进展的一个梳理，该进展实际上在朝着对它的证明迈进。确实，总结我们的结论，我们声称有好的证据支持：

1. 经典物理学在极限$\hbar \to 0$或$N \to \infty$时从量子理论中产生，假如系统处于特定的“经典”态，并只由“经典”的可观测量来记录；

2. 退相干和一致历史将很可能解释，为什么系统处于这些态，并以这样的方式被观测到。

然而，即使有一天该方案能够起作用，关于经典世界是量子理论的表象的解释将以对众所周知的问题“从‘与’到‘或’”的外部解决为基础：若量子力学预言了具有特定概率的不同可能结果，为什么我们仅看到一个?②

本论文更详细的大纲，可以参考上述的内容列表。绝大多数的哲学讨论出现在关于哥本哈根解释的第3节中，因为无论其优势何在，它不可否认地设定了整个关于经典与量子关系的讨论平台。③ 论文的其他小节将是几乎纯技术化的。除此之外我们试图要避免争议，但当不可避免时争议仅限在第3至第6节增加的后记中。最终的后记(第8节)表达了我们对此论题的深刻思考。

① 在此文献中对“精确性”的坚持使人想起普朗克对不可逆性的绝对本性长期坚持的信念[Darrigol，1992；Heilbron，2002]。需要指出，虽然普朗克的固执由于历史的偶然使得他跨出了量子理论的第一步，但他最终放弃了它，而与玻耳兹曼(Boltzmann)站在一边。

② 不得不承认，我们应该把对此问题的坚持归功于立场2的拥护者们。

③ 我们不在理论约化与理论间关系的哲学背景下讨论量子力学的经典极限，参见[Scheibe，1999]和[Batterman(巴特曼)，2002]。

2. 早期历史

本节是对量子力学创始者们关于经典与量子关系的主张与贡献的简要回顾。更多的细节可以在引言中引用的著作和将要引用的具体文献中找到。可观的(但不完备的)书目见于[Gutzwiller(古兹维勒),1998]。量子理论的早期历史有其自身的意义,因为它是历史上最为重要的科学革命之一。虽然这段历史不是本论文的主要焦点,但对我们的论题有着特别的重要性。因为对普朗克工作的常见和错误的解释(即他引入了类似于"量子假设"的观念,见下文3.2节)似乎会引起这样的信念,即量子理论和普朗克常数与自然中普遍的非连续性有关。确实,这种非连续性有时甚至被认为标志着经典与量子物理的根本不同。这种信念在玻尔的作品中异常明显,甚至在今天依然能引起共鸣。

2.1 普朗克和爱因斯坦

经典物理学与量子理论间的关系是如此的微妙和混乱,以至于历史学家和物理学家甚至不能就经典被量子取代的确切方式达成共识!正如达里戈尔[2001]所说:"在过去的二十年里,历史学家(和物理学家)关于普朗克于1900年在其黑体理论中引入的量子的意义存在分歧。这种混乱的源头是托马斯·库恩(Thomas Kuhn)关于反传统论题的发表,该论题是指普朗克并不打算用他的能量子表达量子的非连续性。"

众所周知,如果比较[Mehra and Rechenberg,1982a]等文献,普朗克最初在他建立不可逆性的绝对本性的方案的语境中得到了黑体辐射的维恩(Wien)定律(与玻耳兹曼的概率方法相竞争,并最终取得了成功)。当1900年10月新的高精度测量结果否定了维恩定律,普朗克第一次推测出了他关于正确定律的表达式:

$$E_\nu/N_\nu = h\nu/(e^{h\nu/kT}-1) \tag{1}$$

顺便引入了两个新的自然常数 h 和 k①,随后在1900年12月4日,给出了对这一定律的理论推导。在此定律中他引入了下述理念,即构成黑体的振子的能量

---

① 从而,普朗克引入了玻耳兹曼常数 k,并第一个写下了公式 $S=k\log W$。

是以 $\varepsilon_\nu = h\nu$ 为单元量子化的(其中 ν 是给定振子的频率)。该推导通常被看作是量子理论的诞生，诞生的日期就是刚才提到的日期。

然而，现在清楚的是[Kuhn，1978；Darrigol，1992；2001；Carson(卡尔森)，2000；Brush(布鲁斯)，2002]：往好了说，普朗克并不知道他所说的振子的能量；往不好了说，他赋予振子以一连续的能量谱。从技术上说，特别是他的经验定律的推导最终被证明产生了期望的结果(这依赖于玻耳兹曼的熵概念)，[①] 普朗克不得不计算给定的能量数额 E_ν 在频率 ν 时分配给一定数量的振子 N_ν 有多少种方式。当然，方式有无穷多种，因而为了找到一个有限数目的回答，普朗克追随着玻耳兹曼把 E_ν 分解为具有等同大小 ε_ν 的大数目 A_ν 份，从而有 $A_\nu\varepsilon_\nu = E_\nu$。[②] 现在，正如我们所知，尽管玻耳兹曼最后在他对气体的相应计算中令 $\varepsilon_\nu \to 0$，普朗克发现若他假定关系式 $\varepsilon_\nu = h\nu$，则会诞生他经验性的黑体定律。

但是，该假设并不意味着普朗克量子化了振子的能量。事实上，在他对给定分布的定义中，他计算了能量在 $(k-1)\varepsilon_\nu$ 和 $k\varepsilon_\nu$ ($k \in \mathbb{N}$)间的振子的数量，就像玻耳兹曼对于气体采取的类似方式，而不是具有能量 $k\varepsilon_\nu$ 的振子的数目，就像绝大多数的物理学家对此过程的解释那样。更一般地，有大量的书面证据表明，普朗克本人绝不相信或表达他有了量子化的能量，在 1920 年诺贝尔奖演讲中，他把对能量子 ε_ν 的正确解释归于爱因斯坦。确实，对量子理论最早期阶段的最新理解是这样的，即是爱因斯坦而非普朗克，在 1900—1905 年期间，清楚地认识到了普朗克的辐射定律标志着与经典物理学的分裂[Buttner *et al.*(布特纳等)，2003]。这一洞察让爱因斯坦进行了能量量子化。他是以两重方式进行的，都与普朗克的振子相关——被爱因斯坦解释为是现代方式下的谐振子——且借助于他的光子概念以一种紧密关联的举措进行。尽管普朗克理所当然引入了以他名字命名的常数，并且因此成为了用 h 标志的理论的创立之父，但是光子的引入使得爱因斯坦至少成为量子理论之母。爱因斯坦本人也把光子

---

① 虽然事实是，普朗克只是在 1914 年左右才转变到不可逆性的玻耳兹曼方法。

② 问题中的数目则由 $(N+A-1)!/(N-1)!A!$ 给出，在记号中略去了对 ν 的依赖性。

看作是他最为革命性的发现，因为他关于他相关论文的写法并不与他关于相对论所体现的自信相匹配："关于光的辐射和能量的属性，是非常革命性的。"①

最后，在本文看来需要提及的是，爱因斯坦[1905]与普朗克[1906]首次对量子理论的经典极限作出评论，见下文第5节的导言。

2.2 玻尔

玻尔卓越的原子模型强化了他的理念，即量子理论是关于量子的理论。②因为该模型同时突出了经典与量子物理学间的冲突，并且带来了通过玻尔同样卓著的对应原理来解决这一冲突的方法的萌芽，这里关于此模型值得说几句。③玻尔的原子模型指出了卢瑟福(Rutherford)的太阳系型原子的辐射不稳定性。④根据洛伦兹的电动力学，一个加速的电子应该辐射，因为设想的电子绕核的环形或椭圆形运动是加速运动的特殊情形，电子应该连续地损失能量，并螺旋式地撞向原子核。⑤玻尔通过三种同时的运动来对抗这种不稳定性，每一种都非常有原创性。

1. 他引入了量子化条件，该条件仅挑选了一组离散的被允许的电子轨道(这些轨道随后用经典力学来描述，例如在玻尔对里德(Ryolberg)堡常数 R 的计算中)。

2. 他用发生在不可预测时刻的具有不可预测目的性的量子跃迁取

---

① "(本论文)是关于光的辐射和能量属性，且是非常革命性的。"也见派斯[1982]的序言。

② 虽然当时玻尔在实践上追随着所有的物理学家拒绝爱因斯坦的光子，因为他相信在一个量子跃迁期间，原子以球面波形式发射电磁辐射。要是采用辐射的光子图景他的模型将获得连贯性，但事实上玻尔是关于光子最后的主要反对者，拒绝该观念直到1925年。也见于布莱尔·波勒斯(Blair Bolles)[2004]。

③ 更详细的处理见[Darrigol，1992]，也见[Liboff(立波夫)，1984]和[Steiner(斯坦纳)，1998]。

④ 太阳系提供了卢瑟福原子的直观形象，但卢瑟福本人的图像更接近于土星圆环，而非围绕太阳运行的行星。

⑤ 此外，拥有多于一个电子的任何卢瑟福式的原子在力学上是不稳定的，因为电子之间相互排斥，与行星之间的相互吸引相反。

代了洛伦兹所要求的连续辐射，在跃迁中原子发出光，光的能量等于电子跃迁轨道间的能量差。

3. 他通过引入基态概念阻止了量子跃迁时原子的崩溃，电子不会降低到基态之下。

通过这些假设(关于这些假设当时还不存在所谓的基础)，① 玻尔解释了氢原子的光谱，包括对 R 异常准确的计算。此外，他提出了在随后十年中被指定为寻找量子力学的关键的引导性原理，即对应原理，参见[Darrigol，1992]，和[Mehra and Rechenberg，1982a，pp. 249—257]。

一般地，电子在一个具体的量子跃迁中失去的能量与它在跃迁前运转轨道上经典地辐射的能量(即根据洛伦兹的理论)二者之间没有关系。确实，在基态上电子完全不能够通过量子跃迁辐射，而根据经典电动力学它应该不停地辐射。但是，玻尔看到了在非常大轨道(即，那些拥有大的主量子数 n 的轨道)的相反情形中，发出辐射的频率 $\nu=(E_n-E_{n-1})/h$(有 $E_n=-R/n^2$)近似对应于经典理论中最低阶谐波的频率(在将经典理论应用到的电子在初始轨道中运动时)。② 此外，相关谱线被测量到的强度(它在理论上应该与量子跃迁的概率相关，该概率在早期量子论中不可理解)，类似地由经典电动力学给出。这一特性，在简单情形下能够通过精确计算或实验被证明，成为关于无法证明情形的引导原理，且有时甚至拓展到低量子数情形，尤其是在经典理论预言了选择定则之时。

需要强调的是，玻尔的对应原理关注的是辐射的特性，而非它们本身的力

① 到目前为止，玻尔原子模型在数学上被证明的是基态的存在，见[Griesemer *et al.* (格里斯默等)，2001]，其中的参考文献里有迄今为止可得的最一般性的证明和原子在与电磁场耦合后激发态的亚稳定性，见[Bach *et al.* (巴奇等)，1998；1999]与[Gustafson and Sigal，2003]。仅当辐射场退耦后能量谱是离散的，才能对氢原子光谱做常规计算，这最早是由薛定谔和外尔(Weyl)做出的。也见5.4节的结尾。

② 类似地，更高阶的谐波对应于量子跃迁 $n\to n-k$，其中 $k>1$。

学轨道。[1] 这与当代文献中所谓的对应原理不完全相同。[2] 事实上，虽然当代这个对应原理目前有着特定的有效范围(正如我们将在第5节中详细看到的)，但玻尔从未支持过类似的原理，甚至公开表示过反对这一原理:[3]

> “正是在珀塞尔(Purcell)的办公室中，珀塞尔和其他人让玻尔休息了几分钟(在1961年对哈佛大学物理系访问期间)。当时他们正在进行一般性的讨论，玻尔评论道：‘人们说经典力学是量子力学在 h 趋于0时的极限。’然后，珀塞尔回忆说，玻尔摇了摇手指，走到黑板前写下 e^2/hc。随着他在 h 下划了三横，玻尔转过身说，‘你们看到了 h 是在分母上。’”

2.3 海森堡

海森堡的论文[1925]《关于运动学和力学关系的量子论的重新解释》(*Über die quantentheoretische Umdeutung kinematischer und mechanischer Beziehungen*)[4]一般被看作是量子力学发展的转折点。甚至派斯(他不是海森堡的朋友)[5]也承认海森堡的论文标志着“伟大的一跃——或许是最伟大的——在20世纪物理学的发展中”。海森堡究竟完成了什么？从我们论题的视角看这个问题尤其令人感兴趣。

在那时，原子物理学处于危机状态，对此不同阵营以不同方式应对。玻尔的方法最好被描述为危机控制(damage control)：他的量子理论是经典力学的杂交体，通过特设性的量子化规则进行调整，同时不惜一切代价保持电动力学的

---

① 同样，它尚需以严密的方式来确证。

② 当代版本的典型案例是：“非相对论的量子力学建立在玻尔的对应原理之上：‘当普朗克常数 h 可以相对于如质量和距离等其他参数被看作很小时，量子力学接近于经典的牛顿理论。’”见[Robert(罗伯特)，1998，44]。将其归于玻尔在历史上是不正确的！

③ 引用自[Miller(米勒)，1934，313]

④ 关于运动学和力学关系的量子论的重新解释。英文翻译见[van der Waerden，1967]。

⑤ 例如在[Pais，2000]中，宣称描述了“科学的天才”，海森堡显然并不在列。

经典性。[①] 爱因斯坦是第一个认识到需要量子化经典电动力学的物理学家，但鉴于他在广义相对论上的胜利，梦想着拥有奇异解的经典场理论成为量子现象的最终解释。玻恩领导着激进的阵营，其中包括泡利：他看到了需要用一种全新的力学取代经典力学，[②] 这种新力学建立在满足微分方程的离散量基础之上。[③] 这是黑暗中的一跃，尤其是在泡利为对应原理而发愁之时［Hendry，1984；Beller，1999］。

海森堡明智地处于玻尔与玻恩之间是海森堡的才能展示。[④] 他的重新解释(Umdeutung)的意义是保持经典的运动方程，[⑤] 同时将那里用到的数学符号转换为(随后被认识到是)矩阵。因而，他的重新解释 $x \mapsto a(n, m)$ 是现在被称为量子化映射 $f \mapsto Q_\hbar(f)$ 的早期形式，其中 f 是经典的可观测量，即相空间的一个函数，从希尔伯特空间中的一个向量，或更抽象地，从某个 C^* 代数的一个元素的意义上来讲，$Q_\hbar(f)$ 是一个量子力学的可观测量。见下文第 4 节。正如海森堡认识到的，这一举措意味着量子力学可观测量的非对易性，正是此特性，而非像“量子假设”(见下文 3.2 节)之类的某种东西，是量子力学的定义性特征。事实上，大多数之后关于量子物理学和实践中对经典与量子联系的所有考虑都依赖于海森堡的重新解释理念之上。这甚至适用于整个数学体系(见 2.5 节)。

我们这里在一种不严格的意义上使用术语“可观测量”。现在认识到文献［Mehra and Rechenberg，1982b；Beller，1999；Camilleri(卡米勒里)，2005］中海森堡的下述陈述是障眼法，即他的形式体系可以从物理上解释为用可观测量

---

① 我们引用自［Mehra and Rechenberg，1982a，pp. 256—257］：“因而，在 1920 年代早期，尼尔斯·玻尔对如何推进原子理论有了明确的观点。他想要最大化地利用他所谓的‘更二象性的描述’……在其中，原子被看作是拥有分立态并发出分立频率辐射的力学系统，辐射频率(以一种非经典的方式)由定态间的能量差来确定，另一方面，辐射不得不用经典电动理论来描述。”

② 甚至在海森堡的论文之前，玻恩发明了量子力学这一名称。

③ 克莱默(Kramers)早先就提到了这一想法。

④ 同样是字面上的！海森堡大多数的时候往来于哥本哈根和哥廷根。

⑤ 重新解释的这一重要方面立即被狄拉克(Dirac)［1926］领会到：“在一篇最近的论文中，海森堡提出了一种新的理论，该理论提出不是经典力学的方程有某种过错，而是物理结果由之导出的数学操作需要修正……量子与经典理论之间的对应不在于当 $\hbar \to 0$ 时的极限一致，因为事实上对两种理论的数学操作在许多情形下遵循相同的定律。”

取代原子轨道，这是由他与泡利的讨论激发而产生的。事实上，在量子力学中任意力学量都必须被"重新解释"，不论它是否是可观测量。正如海森堡[1969]回忆的，爱因斯坦谴责他造成物理学承认可观测量这一概念的假象，并解释说是理论决定了什么能够被观测到。重新思考可观测性问题则使得海森堡作出了他对量子力学的第二大贡献，即他的不确定关系。

这些关系是海森堡自己对本文开头引述问题的回答。确实，矩阵力学最初是极端抽象和形式化的方案，缺乏的不只是形象化，也缺乏态(见下文)的概念。虽然这些特征最初合乎玻恩、海森堡、泡利和约尔丹(Jordan)的意愿，但薛定谔工作的成功迫使他们放弃他们极端的立场，并寻找支持他们的数学的半经典图景，这是相当大的逆转[Beller，1999；Camilleri，2005]。海森堡[1927]找到了这样的图景，宣称他的不确定关系提供了"量子理论的运动学和力学的直觉内容"(正如他论文的名称)。他的理念是经典世界通过观测产生自量子力学："轨迹仅因为我们观测它才成为存在。"①这一理念开始变得非常有影响，可以看作是在引言中立场 1 的来源。

2.4 薛定谔

量子力学的历史很大程度上已经被这样的洞见所澄清，即正如普遍相信的那样，海森堡和薛定谔并没有发现理论的两种等价构造，是海森堡[1925]确定了可观测量的数学性质，而薛定谔[1926a]找到了对态的描述。② 矩阵力学缺乏态的概念，而同样地波动力学最初不包含可观测量，只是在试图把波动力学与矩阵力学关联起来时薛定谔[1926c]才引入了位置与动量算符③：

$$Q_{\hbar}(q^j) = x^j$$

$$Q_{\hbar}(p_j) = -i\hbar\frac{\partial}{\partial x^j} \tag{2}$$

① "Die Bahn entsteht erst dadurch, daβ wir sie beobachten."(德语，意思同上)

② 也见[Muller(穆勒)，1997]。

③ 这里 $j=1, 2, 3$。按现代术语，右手边的表达是在希尔伯特空间 $\mathcal{H}=L^2(\mathbb{R}^n)$ 上的无界算子。更多细节见第 4 节。表达式 x^i 是一个乘法算子，即 $(x^j\Psi)(x)=x^j\Psi(x)$，而显然 $(\partial/\partial x^j\Psi)(x)=(\partial\Psi/\partial x^j)(x)$。

这为薛定谔方程提供了新的基础：①

$$\left(-\frac{\hbar^2}{2m}\sum_{j=1}^{n}\frac{\partial^2}{\partial x_j^2}+V(x)\right)\Psi = i\hbar\frac{\partial\Psi}{\partial t} \tag{3}$$

通过把左手边解释为 $H\Psi$，其中 $H = Q_\hbar(h)$ 是用经典的哈密顿(Hamiltonian)量 $h(p,q) = \sum_j p_j^2/2m + V(q)$ 表示的。因而薛定谔建立了现在用他名字命名的算符理论，② 并由此给出了关于海森堡就经典可观测量重新解释的理念中现在看来仍最为重要的例子。

随后，通过对狄拉克、泡利和薛定谔特定观点的修正与扩展，冯·诺伊曼[1932]出色地通过希尔伯特空间的概念将这两部分统一了起来。他也给出了玻恩、泡利、狄拉克和约尔丹跃迁概率公式的抽象形式，因而完成了量子力学的数学构建。

但是，薛定谔并不这样看待他的贡献。他想把波动力学打造成一个关于实体的成熟的经典场理论，而不仅仅是对世界概率描述的一半(即现代用法中态空间那一半)，该描述仍然需要引入他如此憎恶的量子跃迁[Mehra and Rechenberg, 1987; Gotsch(高斯克), 1992; Bitbol and Darrigol(比特鲍罗和达里戈尔), 1992; Bitbol, 1996; Beller, 1999]。粒子被假定以波包的形式出现，但很快海森堡、洛伦兹和其他人指出，在实际情形中这样的波包随时间的推移会弥散开来。最初薛定谔忽视了这一点[1926b]，他把他的直觉建立在谐振子的特殊情形上。在积极的方面，在他通过对波包的应用从波动力学推导经典粒子力学不成功的尝试过程中，薛定谔[1926b]给出了现在称为相干态(coherent state)的第一个例子。这里，一个量子波函数 Ψ_z 用一个"经典的"参量 z 来标示，以这样一种方式 $\Psi_z(t)$ 的量子力学时间演化就由 $\Psi_{z(t)}$ 近似给出，其中 $z(t)$ 表示某种关联的经典时间演化，见下文的 4.2 或 5.2 小节。这被证明是理解从量子到经典力学过渡的一个非常重要的理念。

---

① 或者是相应的不含时方程，在右手边有 $E\Psi$。

② 见[Reed and Simon, 1972; 1975; 1987; 1979; Cycon(赛康), 1987; Hislop and Sigal(希斯洛普和西加尔), 1996; Hunziker and Sigal(洪齐克和西加尔), 2000; Simon, 2000; Gustafson and Sigal, 2003]。薛定谔方程的数学起源也参见[Simon, 1976]。

与此同时，在同一篇论文中，薛定谔[1926b]提出了玻尔对应原理的下述波动力学版本：经典的原子态应该来自于大量(比如说至少10,000个)高激发态(即具有非常大量子数的能量本征函数)的叠加。经过很多年对这种理念有限的理论兴趣之后，对原子物理学中波包的兴趣在20世纪80年代晚期再次被唤起，这归功于建立在激光(如抽运—探测与相位调制)之上的现代实验技术的发展。关于近期的技术回顾见[Robinett(罗宾奈特)，2004]或稍早些的通俗说明见[Nauenberg *et al.* (瑙恩伯格等)，1994]。大体来说，显现出的图景是这样的：所说类型的定域波包初始时的时间演化基本具有经典的周期性，正如薛定谔所期望的，但后来在经历了许多的轨道后就弥散了。因此，在第二个阶段，概率分布近似地充满了经典轨道(虽然并不均匀)。甚至更惊奇的是，在更长的时间尺度上存在一种波包复原现象，其中波包恢复了它初始时的定域化。之后整个循环再一次开始，从而我们看到周期性行为，但并不是所期望的经典类型。因而即使是在被天真地想象为完全经典的体系中，波动现象继续起作用，引起相当异乎寻常和意料之外的行为。虽然波包复原的严格数学描述还未形成，但总的图像(同时建立在“理论物理学”式的数学与实验基础上)是足够清晰的。

波包是否是按照薛定谔的意图演化的是有争议的(且无关紧要的)，参见[Littlejohn(雷托约翰)，1986]，确定的是，他关于经典与量子关系的其他核心观点相当有影响力。当然，这是薛定谔[1926b]从经典力学的哈密顿—雅可比(Hamilton-Jacobi)体系对他波动方程的“推导”。这引起了WKB近似与相关的方法，见5.5节。

无论如何，在薛定谔期待为他的波函数做经典解释和海森堡想要避开[Beller，1999]的地方，玻恩与泡利很快认识到了它正确的、概率性的意义。因而他们剥夺了波函数单纯的物理本性，且有效地将其降级为概率幅的纯数学地位。且由此玻恩与泡利再次把量子力学与经典力学间的联系看作是几乎不可理解的！正是这种不可理解性使海森堡提出了他的不确定关系。

2.5 冯·诺伊曼

虽然创造了量子力学的希尔伯特空间体系，冯·诺伊曼的著作[1932]仍然

可以看作是对海森堡重新解释观念的数学贯彻。冯·诺伊曼事实上提出了如下量子理论的重新解释:

相空间 $M \mapsto$ 希尔伯特空间$\mathcal{H}$;

经典可观测量(即 M 上的实值可测量函数)$\mapsto \mathcal{H}$上的自伴算符;

纯态(看作 M 中的点)$\mapsto \mathcal{H}$的单位向量(事实上是射线);

混合态(即 M 上的概率测量)$\mapsto \mathcal{H}$上的密度矩阵;

M 上可测量的子集$\mapsto \mathcal{H}$的闭线性子空间;

补集合$\mapsto$正交补;

子集的并$\mapsto$子空间的闭线性张成;

子集的交$\mapsto$子空间的交;

是—否问题(即 M 上的特征函数)$\mapsto$投影算符。[①]

这里出于简单性的考虑,我们假定希尔伯特空间$\mathcal{H}$上的量子可观测量 R 是有界算符,即 $R \in \mathcal{B}(\mathcal{H})$。冯·诺伊曼事实上是依靠把量子力学态公理化刻画为线性映射 Exp: $\mathcal{B}(\mathcal{H}) \to \mathbb{C}$ 来做到将经典混合态重新解释为密度矩阵的,其中映射满足 $\mathrm{Exp}(R) \geqslant 0$ 当 $R \geqslant 0$ [②], $\mathrm{Exp}(1) = 1$ [③],且在算符交换集上有可数可加性。因为他证明了这样一个映射 Exp 必定是由密度矩阵 ρ 给出的,根据是 $\mathrm{Exp}(R) = \mathrm{Tr}(\rho R)$。[④] 一个单位向量 $\Psi \in \mathcal{H}$通过 $\psi(R) = (\Psi, R\Psi)$ 定义了冯·诺伊曼意义上的一个纯态,其中 $R \in \mathcal{B}(\mathcal{H})$,我们称这个态为 ψ。类似地,$\mathcal{H}$上的一个密度矩阵 ρ 通过 $\rho(R) = \mathrm{Tr}(\rho R)$ 定义了一个态(一般是混合态),并称之为 ρ。在当前表述中,冯·诺伊曼定义的在$\mathcal{B}(\mathcal{H})$上的态会被称为正常态。在量子

---

① 稍后,他自然又补充了用模格对布尔格的重新解释,并接着用量子逻辑重新解释了经典逻辑[Birkhoff and von Neumann(伯克霍夫与冯·诺伊曼),1936]。

② 即当 R 是具有正谱的自伴算符,或等价地当对某个 $S \in \mathcal{B}(\mathcal{H})$ 有 $R = S^* S$ 时。

③ 其中 Exp(1) 中的 1 是$\mathcal{H}$上的单位算符。

④ 这一结果被普遍误解(显然也包括冯·诺伊曼本人)为是在量子力学中排斥隐变量的定理,见[Scheibe, 1991]。然而,贝尔将冯·诺伊曼在态的定义中的线性假设刻画为"愚蠢的(silly)"很不准确,因为它同时在经典力学和量子力学中成立。事实上,冯·诺伊曼的定理确实排除了量子力学的隐变量扩展,该隐变量本质上是经典的,而且物理学家们最初寻找的正是这些扩展。对此问题的最新讨论见[Rédei and Stöltzner(瑞德和斯托尔兹纳),2001]和[Scheibe, 2001]。

物理学的 C^* 代数构造中，这一公理化一直保留到今天。这里为了与可能的超选择定则相适应，用更一般的可观测量代数取代$\mathcal{B}(\mathcal{H})$[Haag，1992]。

在量子力学的数学公理化之外，这一公理化(连同它随后通过 C^* 代数构造的推广一起)是关联经典与量子力学所有重要工作的基础，冯·诺伊曼通过他对测量问题的分析为经典与量子间的关系做出了贡献。① 因为在这里经典与量子物理学之间的明显冲突在测量问题处达到顶点，有必要在这里总结冯·诺伊曼对此问题的分析。对测量问题更一般的讨论同样见[Wheeler and Zurek(惠勒和朱瑞克)，1983；Busch *et al.*(布施等)，1991；Auletta(奥莱塔)，2001]与[Mittelstaedt，2004]。

测量问题的本质是，有些态从未在自然中见过，虽然它们不只是量子力学所允许的(在它普遍有效的假设下)，在典型测量情形的预测中也会出现。考虑系统 S，它的纯态在数学上由希尔伯特空间$\mathcal{H}_S$中的归一化向量(更具体地是射线)描述。人们想要测量可观测量$\mathcal{O}$，它在数学上是用$\mathcal{H}_S$上的自伴算符 O 表示的。冯·诺伊曼假定 O 具有离散谱，这种简化并不掩盖测量问题中的基本问题。因此 O 拥有实本征值为 o_n 的单位本征向量 Ψ_n。为了测量$\mathcal{O}$，将系统与具有希尔伯特空间$\mathcal{H}_A$和"指针"可观测量$\mathcal{P}$的仪器 A 耦合，$\mathcal{P}$由$\mathcal{H}_A$上的自伴算符 P 表征，拥有离散的本征值 p_n 和单位本征向量 Φ_n。总系统 $S+A$ 的纯态则对应于张量积$\mathcal{H}_S\otimes\mathcal{H}_A$上的单位向量。一个好的("第一类")测量是这样的，即在测量之后，Ψ_n 关联于 Φ_n，即对于适当选择的初态 $I\in\mathcal{H}_A$，态 $\Psi_n\otimes I$(在 $t=0$ 时)几乎立即演化到 $\Psi_n\otimes\Phi_n$。这确实可以通过适当的哈密顿量来实现。

由薛定谔猫所突出的测量问题现在产生了，若选择 S 的初态为 $\sum_n c_n\Psi_n$

---

① 冯·诺伊曼[1932]回避了对量子力学的经典极限或(也许)是量子化概念的讨论。在后一方向上，他宣称"若量$\mathfrak{R}$拥有算符 R，则量 $f(\mathfrak{R})$ 拥有算符 $f(R)$"，且"若是$\mathfrak{R}$，$\mathfrak{S}$，…，拥有算符 R，S，…，则量$\mathfrak{R}+\mathfrak{S}+\cdots$拥有算符 $R+S+\cdots$"。然而，虽然他极富清晰与精确的盛名，冯·诺伊曼仍然对跃迁$\mathfrak{R}\mapsto R$ 的意义相当不清楚。将$\mathfrak{R}$构造为一个经典可观测量，其量子力学的对应是 R，这是相当有诱惑力的，从而上述引述可以看作是量子化的公理。但是，这样一种解释既不受相关文献的支持，也不受我们现在对量子化理解的支持(参见第4节)。例如，一个量子化映射$\mathfrak{R}\mapsto\mathcal{Q}_\hbar(\mathfrak{R})$不能够满足$f(\mathfrak{R})\mapsto f(\mathcal{Q}_\hbar(\mathfrak{R}))$，即使对如$f(x)=x^2$ 这样非常合适的函数而言。

(有 $\sum |c_n|^2 = 1$)，则叠加原理导致这样的结论，即耦合系统的终态是 $\sum_n c_n \Psi_n \otimes \Phi_n$。现在，从根本上说冯·诺伊曼所说的就是，若限制到系统 S 的终态，则得到的密度矩阵是混合态 $\sum_n |c_n|^2[\Psi_n]$，其中$[\Psi]$是向一个单位向量 Ψ 上的正交投影，① 从而，从系统自身的视角看，测量似乎引起了从纯态 $\sum_{n,m} c_n \overline{c_m} \Psi_n \Psi_m^*$ 向混合态 $\sum_n |c_n|^2[\Psi_n]$ 的转变，其中干涉项 $\Psi_n \Psi_m^*$ 在 $n \neq m$ 时消失。这里算符 $\Psi_n \Psi_m^*$ 通过 $\Psi_n \Psi_m^* f = (\Psi_m, f)\Psi_n$ 来定义；特别地，$\Psi\Psi^* = [\Psi]$。② 类似地，仪器——考虑到它本身——从纯态 $\sum_{n,m} c_n \overline{c_m} \phi_n \phi_m^*$ 演化到混合态 $\sum_n |c_n|^2[\Phi_n]$。这仅仅是一个数学定理(假定有可能以期望的方式将系统与仪器耦合)，而非对自然中存在两种不同时间演化的提议，即按照薛定谔方程的幺正传播，与此同时还有上述“塌缩”过程。

无论如何，这一举措本身并没有解决测量问题。③ 第一，在给定的情形下不允许采取对混合态的无知解释(即假定系统真的是处于其中一个态 Ψ_n)；参见如[Mittelstaedt, 2004]。第二，即使允许采取无知的解释，仍可以通过再一次考虑到系统的其他组成部分而重建该问题(即最初的叠加 $\sum_n c_n \Psi_n \otimes \Phi_n$)。

冯·诺伊曼至少很好地意识到了第二点，他对此的回应是对链的构造：重新把 $S+A$ 定义为系统，将它与新的仪器 B 等耦合。这样在最后会产生一个后测量态 $\sum_n c_n \Psi_n \otimes \Phi_n \otimes \chi_n$(在可以指望的不言自明的记号中，假定向量 χ_n 形成了一个正交集)，将它限制到 $S+A$ 上就是混合态 $\sum_n |c_n|^2[\Psi_n] \otimes [\Phi_n]$。将后一个态局限于 S 又会是 $\sum_n |c_n|^2[\Psi_n]$。这一过程显然被重复，构造的关键明显在于把给定系统的叠加传递到链之外的任意系统。于是出现了下述结果，对原始系统的终态而言在哪里“切断链条”(即不考虑链条中的哪一部分)

① 即$[\Psi]f = (\Psi, f)\Psi$，在狄拉克记号下有$[\Psi] = |\Psi\rangle\langle\Psi|$。

② 在狄拉克记号下我们有 $\Psi_n \Psi_m^* = |\Psi_n\rangle\langle\Psi_m|$。

③ 即使在量子力学的系综解释下也没有解决，该解释是冯·诺伊曼在写他的这本书时所不幸坚持的。

是无关紧要的，只要它在某处被切断。冯·诺伊曼[1932]用优美的散文和其他人建议把观测行为看作是切断，但更好且更一般的简单说法是，链条的某个终端在描述中被忽略。

则测量问题的任务就是要：

1. 构造一适当的链条，同时在那里有合适的切断，在哪里做出切断无关紧要，只要被切断；

2. 构造一适当的时间演化以完成测量；

3. 为混合态的无知解释辩护。

正如我们将要看到的，这些问题在哥本哈根解释与退相干方案中，以一种概念上不同但数学上类似的方式得到了处理。（概念的主要不同在于后者也致力于解决解释经典世界表象这一更艰巨的问题，而前者似乎认为这是理所当然的。）

结束本节时我们要说，虽然有一些闪光点，但量子力学的创始者们远未解决从他们理论的特定体系得到经典力学这一问题。

3. 哥本哈根：再评述

所谓的量子力学“哥本哈根解释”要追溯到玻尔、海森堡和泡利在1927年前后首次讨论和构建的观点。与最开始就存在的“政党路线”观点相对比，人们常常指出的是：在20世纪20年代晚期，事实上玻尔和海森堡关于量子力学的解释存在鲜明地不同，他们从未真正达成统一理论[Hooker（胡克），1972；Stapp（斯坦普），1972；Hendry，1984；Beller，1999；Howard，2004；Camilleri，2005]。例如，他们从未就互补性概念达成一致（见3.3节）。更一般地，海森堡通常把他的观点建立在量子理论的数学体系上，而玻尔的立场主要是哲学导向的。然而，也存在清晰可辨的核心观点，对此他们都认同，且因为这一核心与经典与量子之间的关系紧密相关，我们将对它进行详细讨论。

这里主要的原始素材是玻尔的科摩（Como）讲座，他对EPR的回应和他写

给爱因斯坦的随笔[Bohr, 1927; 1935; 1949]。[①] 玻尔[1985; 1996]和梅拉与雷琴堡[2001]给出了对这些论文产生和接收的历史性讨论。对这些论文引出的大量文献进行选择，我们仅提到相对较新的工作[Hooker, 1972; Scheibe, 1973; Folse(福尔斯), 1985; Murdoch(莫多克), 1987; Lahti and Mittelstaedt (拉蒂和米特斯泰特), 1987; Honner(霍纳), 1987; Chevalley(谢瓦莱), 1991; 1999; Faye(法耶), 1991; Faye and Folse, 1994; Held(赫尔德), 1994; Howard, 1994; Beller, 1999; Faye, 2002]与[Saunders(桑德斯), 2005]。与玻尔争论的对手见[Heisenberg, 1930; 1943; 1958; 1984a, b; 1985]、相关的二手文献[Heelan(希兰), 1965; Horz(赫尔茨), 1968; Geyer *et al.* (盖伊尔等), 1993; Camilleri, 2005]与[Pauli, 1933; 1949; 1979; 1985; 1994], 还有[Laurikainen(劳里凯宁), 1988]与[Enz, 2002]。

正如维特根斯坦(Wittgenstein)和许多其他思想家一样，在"早期"玻尔和"晚期"玻尔之间作出区分有助于理解玻尔。[②] 尽管他的思想中有许多的连续性(见下文)，分界点是他对 EPR 的回应[Bohr, 1935],[③] 他后来做出的主要转变是，他在 1935 年之后强烈坚持客体和观测者不可分的统一性集中在现象概念上。在 EPR 之前，玻尔同样相信客体和观测者都是量子理论完备描述的必要因素，但之后他认为虽然它们的相互作用从来不能被忽略，至少在逻辑上它们可以分开来考虑。在 1935 年之后，玻尔慢慢开始宣称，客体和观测者再也不具有分立的特性，而是共同形成了一个"现象"。相应地，他的互补性概念也改变了，越来越关注于这样的观点，即实验条件的确定对于量子理论中必要的经典概念的明确使用是至关重要的[Scheibe, 1973; Held, 1994]。同样见下文 3.3 节。这一发展在玻尔最终对量子世界存在性的否定中达到了顶点。

---

① 这些论文事实上分别是与泡利、罗森菲尔德(Rosenfeld)和派斯合作完成的。

② 这里我们与赫尔德[1994]与贝莱尔[1999]一起反对霍华德[1994]与桑德斯[2005]。同样见[Pais, 2000, 22]:"但是，玻尔的科摩演讲并没有赢得喝彩。他本人后来对他在那里用到的表述表示不满，如'观测对现象的干扰'。这样的语言引起了长期围绕该主题的相当大的混乱。"

③ 这一回应是有问题的，正如 EPR 本身一样。因此，关于二者都有相当多的文献，以这样的事实为标志，即同样有能力且博学的两方评论者设法直接互相攻击，而同时双方都宣称要解释和重构玻尔"真正"的意思。

"不存在量子世界。只存在抽象的量子物理学描述。认为物理学的任务是发现自然是什么的想法是错误的。物理学关心的是关于自然我们能够说些什么……我们人类依赖于什么？我们依赖于我们的语言。我们的任务是与他人交流经验与想法。我们飘荡在语言中。"引自[Petersen(皮特森)，1963，8]。①

3.1 经典概念学说(doctrine)

虽然有这种转变，但似乎也可以看到玻尔在他的整个职业生涯中坚持了一个核心思想：

"不论现象超出经典物理学的解释范围有多远，对所有证据的说明必须用经典术语来表达……论证简要如下，即通过实验这个词我们指的是一种情境，在其中我们可以告诉别人我们做了什么、我们学到了什么，因而对实验安排和对观测结果的说明必须在对经典物理学术语的恰当应用下用精确的语言来表达。"见[Bohr，1949，209]。

简而言之，这就是玻尔关于经典概念的学说。虽然他的许多描绘和事迹可能会表明另外的方面，但玻尔这里并不是要表达物理学的目标是对实验的描述这一观点。② 事实上，他只是指出"无歧义的"交流的需求，对此他明显感觉到了量子力学带来的威胁。③ 上述引述中有争议的部分在于，他把无歧义的交流方式等同于经典物理学的语言，包括粒子和波等。我们将简要地分析玻尔这一认识的具体论证，但需要说明的是，就像在实践中他对于量子力学的所有基础性评论一样，玻尔将他的论证看作是自明的、必要的、无须进一步分析的[Scheibe，1973；Beller，1999]。然而，年轻的海森堡恰好在这一点上与玻尔

① 对此有争议的引述的有趣讨论见[Mermin(默明)，2004]。

② 它常常但被错误地与爱因斯坦的信念——即物理学的目标是描述实体——相对比。近期的讨论见[Landsman，2006b]。

③ 这里"无歧义的"意味着"客观的"[Scheibe，1973；Chevalley，1991]。

产生了冲突，因为海森堡觉得是量子理论抽象的数学体系（而非玻尔的语言与图像世界）提供了那些无歧义交流的途径。①

对于经典物理学，玻尔无疑指的是牛顿、麦克斯韦（Maxwell）和洛伦兹的理论，但这并不是重点。② 对玻尔而言，经典物理学的确定性这一特性是指它的客观性，即它能够以观测者无关的方式来研究：

> “目前为止所有对经验的描述都建立在下述假设上，即有可能严格区分客观的行为和观测的方法，它已经内在于语言的普通约定中。这一假设不只是被日常经验所完全证明，甚至还构成了经典物理学的整体基础。”见[Bohr，1958，25]。③

同样见[Hooker，1972；Scheibe，1973]与[Howard，1994]。海森堡[1958，55]也持这种观点：④

> “在经典物理学中科学开始于信念——或我们应该说开始于假

---

① 很难不同意贝莱尔[1999]的结论，即玻尔只是不能够理解 1925 年之后的量子力学体系，他把自己用语言与图像来理解该理论的需要转变为一种深刻的哲学必要性。

② 否则，人们应当惊讶为什么对此目的不应该采用亚里士多德的物理学和教学法，它是沟通我们对世界朴素印象更为有效的方式。相对比，自牛顿以来物理学的本质就是揭开现象背后的实在。事实上，牛顿本人强调过他的物理学致力于那些胜任自然哲学的人，而不是那些相信朴素表象的“俗人（ye vulgar）”。现在知道亚里士多德的物理学是错的这一事实应当不足以取消其用于玻尔目的的资格，因为相同的评论也可以对牛顿等人的物理学做出。

③ 虽然这一引用是典型的强势论调，玻尔常常也把一些其他特性看作是经典物理学的本质，如决定论、时空概念与动力学守恒定律的结合使用和形象化描述的可能性。然而，这些特性在某种意义上是第二位的，正如玻尔把它们看作是在经典物理学中隔离一个客体可能性的结果。例如：“在理想因果性背后的假设是，客体的行为是唯一确定的，完全独立于它是否被观测”[Bohr，1937]；而现在假设是否定的：“（量子力学中）对理想因果性的放弃仅逻辑地建立在下述基础之上，即我们不再处于谈论一个物理客体自主行为的状态中”[Bohr，1937]。见[Scheibe，1973]。

④ 正如卡米勒里[2005，161]所说的：“对于海森堡而言，经典物理学是对理想客观性的最完全表达。”

象？——即我们可以在不提及我们自身的情况下描述世界或至少是部分世界。这事实上在很大程度上是可能的。我们知道不论我们是否看到它，伦敦城都存在。因此可以说经典物理学仅仅是理想化过程，在其中我们能够在不提及我们自身的情况下谈论部分世界。它的成功导致了对世界客观描述的普遍观念。”

类似地，在他“量子假设”(见3.2节)的基础上，玻尔开始相信，量子物理学的确定性特性恰好是相反的，即观测者(或仪器——玻尔并不区分二者，也从未赋予观测者的意识以特殊地位，或采取物理学的主观观点)角色的不可或缺性。通过把无歧义的交流看作是客观的描述，转而宣称它是经典物理学的本质，玻尔总结说，尽管量子物理学本身也需要完全用经典物理学术语来描述。因而他的经典概念学说有一个认识论的源头，产生自对人类知识条件的分析。① 在那种意义上可以说它在精神上是康德主义的[Hooker，1972；Murdoch，1987；Chevalley，1991；1999]。

对此，玻尔本人曾这样说：“他们的做法很巧妙，但有意义的是做得正确。”②见[Rosenfeld，1967，129]。经典概念学说当然是巧妙的，但它正确吗？正如我们看到的，玻尔的论证开始于这样的陈述，即经典物理学因独立于观测者是客观的(或无歧义的)。事实上，当前广泛相信的是，量子力学导致了相反的结论，即“量子实在”(无论它是什么)是客观的，虽然在德斯帕那特[1995]的术语中是隐晦的，而“经典实在”仅相对于特定的说明而存在：这是引言中讨论过的立场1。③ 那些不同意立场1的人在这点上也不能采用立场2(否认量子

① 例如，见玻尔[1958]的题目！

② “他们”指EPR。

③ 事实上，最近出现了一种精确化玻尔量子力学哲学的尝试，调整了经典可观测量的先验地位以适应某种版本的模态解释，见[Dieks，1989b；Bub，1999；Halvorson and Clifton(霍尔沃森和克里夫顿)，1999；2002]和[Dickson，2005]。这让人们相信二者有可能和平共处，两种最敌对的量子力学解释，即哥本哈根解释和玻姆解释[Cushing(库欣)，1994]现在在作者刚刚引述的意义上在模态解释中找到了一个共同的归宿！无论人们是否同意巴布[2004]对模态解释的批评，玻尔对经典概念必要性的坚持没有受到其当前任何版本的拥护。

理论的基本性质)，因为那当然也不是玻尔心中所想。不幸的是，在他对玻尔最为坦率的辩护中，即使是海森堡也无法找到比不具说服力的评论“经典概念的使用是普遍人类思维方式的最终结果”更好的对玻尔学说的论证。[①]

在我们看来，玻尔学说的动因需要依据我们当前对量子理论的理解做出修改，我们将在3.4节中做这项工作。无论如何，不管其动因是什么，该学说本身似乎值得保留，除去它清晰地描述了实验实践这一事实，它还提供了对量子力学概率论性质的有力说明(参见下一小节)。

3.2 客体与仪器：海森堡界线

用经典概念描述量子物理学听起来似乎是不可能的，甚至是自相矛盾的，参见[Heisenberg，1958]。举例来说，它排除了对世界的完全量子力学描述：“不论现象超出经典物理学解释范围有多远，对所有证据的说明必须用经典术语来表达。”但同时，它也排除了对世界的纯经典描述，因为在经典物理学之下我们拥有量子理论。[②] 玻尔量子力学哲学的迷人之处在于，他给出了对这一明显自相矛盾的情形出色的解答。

他和海森堡提出的求解第一步是，把寻求其描述的系统分为两个部分：一部分是客体，用量子力学描述；而另一部分是仪器，被当作经典的来处理。尽管在文献中有数不清的说法与此相反(即与玻尔坚持认为自然的一个分立领域在本质上是经典的这一想法相反)，无须质疑的是玻尔和海森堡都相信量子力学的基本和普遍本性，并把对仪器的经典描述看作是在本体论中没有对应的纯认识论举措，即表达了这样的事实，给定的量子系统是被作为测量装置而使用

① 且类似地：“我们被迫采用经典物理的语言，只是因为我们没有其他用来表达结果的语言。”见[Heisenberg，1971，130]。尽管后期海森堡非常深刻地思考过这一问题，见他的[Heisenberg，1942]和[Camilleri，2005]。莫多克 [1987，pp. 207—210]万般无奈地尝试通过诉诸于斯特劳森[1959]而将经典概念学说提升到深刻的哲学论证。

② 此特殊情形使得非常难以给哥本哈根解释以实在论的说明，因为量子实在被否定，而经典实在既不基本也不真实。

的。① 例如："所有仪器的构造和运转好像光圈和快门，用来确定实验安排的形状与记时，或像是用以记录原子客体位置的照相底板，对这二者的选择依赖于物质的特性，它们本身在本质上是由作用量子决定的。"见[Bohr，1948，315]。也如："仅在对所关心的过程进行量子力学描述与经典描述实际等价的区域中，我们能够自由地做出分界。"引自[Bohr，1935，701]。②

客体与仪器间的分隔在这里通常被称为海森堡界线，它在量子力学的哥本哈根解释中占据了绝对核心的地位。③ 因而有这样的观点，即量子力学的客体是专门通过它对经典地描述的仪器的影响得以研究的。尽管在经典描述中，仪器是一个量子系统，且被假设受到了它与其深层(量子)客体量子力学耦合的影响。

在描述中同时包括客体和仪器的必要性起初被宣称是所说的"量子假设"的结果。这一概念在玻尔的思想中起到了关键的作用，他的科摩演讲[Bohr，1927]甚至以"量子假设与原子理论的近期发展"为题。在那里他表述其内容如下："量子理论的本质是量子假设(每个原子过程都有着本质的离散性)，对经典理论而言是陌生的，用普朗克作用量子来刻画。"④甚至更为明显的，在他对

① 关于玻尔特别见[Scheibe，1973]和[Heisenberg，1958]。哈格(他同时认识玻尔和海森堡)在20世纪90年代他关于量子力学的大多数演讲中也指出问题的要点。在这方面我们不同意霍华德[1994]，他宣称按照玻尔的看法对仪器的经典描述等于是挑选了其"所能是的(beables)"量子力学代数的一特别的(最大的)阿贝尔子代数，该选择由测量环境所支配。但拥有对易代数远远不能满足经典描述，因为在典型案例中用此方式的人们仅得到一半的正则经典自由度。找到对量子力学系统的经典描述是更深刻的问题，我们将在此论文中不断回到这一点上。

② 这最后一点表明，该界线与在自然中微观与宏观领域的划分有某种关系，尽管此划分通常有助于划出界线(当界线被很好地定义时)，但这绝不是一个原则性问题，参见[Howard，1994]。特别地，因为哥本哈根解释中微观与宏观的区分被模糊而认为该解释不清楚，据此所有对于该解释的反对意见都是没有事实根据的。

③ 泡利[1949]在说到海森堡界线为对现代物理学提供了适当的普遍化时，称可知客体和认识主体间的关系是一种旧康德式的对立："在此方式下，现代物理学将一方面是认识主体和另一方面是已知客体间的旧对立推广到了在观测者或观测工具与被观测系统间划界的观念。"他之后将此界线称作是人类知识的必要条件(原文中引号部分内容保留有德文，译者注)。

④ 代之以"离散性"，玻尔也选择性地用词语"不连续性"或"个体性"。他很少省略像"本质的"这类强化。

EPR 的回应中[Bohr, 1935, 697]："事实上，受作用量子存在的限制在客体与测量媒介间有限的相互作用意味着——由于客体对测量装置反作用的不可控性，若这些测量装置是为了服务于他们的目的——最终必须放弃经典的理想因果性并彻底修正我们关于物理实在问题的态度。"同样，海森堡的不确定关系起初也是由上述形式的量子假设所启发的。根据 1927 年前后玻尔与海森堡(的观点)，这种"本质的离散性"通过仪器在它们的相互作用中引起了对客体"不可控的干扰"。虽然"量子假设"并未得到成熟的量子力学数学体系的支持，且基本上是过时的，但玻尔和海森堡关于对量子力学客体的测量引起对客体"不可控的干扰"这一直觉确实是非常正确的。①

事实上，这种干扰的原因并不在"量子假设"中，而是在纠缠现象中，正如 3.4 节中的进一步讨论。也就是说，从冯·诺伊曼测量理论(见 2.5 节)的观点看，海森堡界线仅仅是冯·诺伊曼链的两步案例，有着特别的特征，即在量子力学的相互作用发生之后，第二个环节(即仪器)是经典地描述的。后一特征不仅支持玻尔的哲学议题，更为重要的是，也足以保证对于测量完成后产生的混合态而言无知解释的适用性。② 冯·诺伊曼所有对该界线位置任意性的分析在这里都适用，因为人们总可以通过把最初的客体与任意其他想要的纯量子力学系统进行耦合，进而扩展对量子力学客体的定义，对于经典部分也可以类似地操作。因而海森堡界线的两步本质包含这样的可能性，即第一个环节或客体事实上的本质是一条很长的链条(只要它是量子力学的)，第二个环节也类似（只要它是经典的)。③ 玻尔与海森堡认识到了这一任意性，该任意性满足本小节中第二段[1935]对玻尔引述所表达的限制，且玻尔至少是发现了它重大的哲学意义。

---

① 即使玻尔后来放弃了它，参见[Beller, 1999]。精确分析的话，耦合于经典仪器时受干扰的是系统的量子力学态(而不是像位置和动量这样的经典可观测量的某个精确值，正如在玻尔和海森堡早期作品中所讲的)。

② 在纯量子力学的冯·诺伊曼链中，系统加仪器的终态是纯态，但如果仪器是经典的，则后测量态是混合态。

③ 泡利[1949]也说："虽然这样的(海森堡)界线的存在是人类知识的必要条件，作为实用的结果且因而是部分自由选择的，其定位在某种程度上是任意的。"(原文中保留有德语部分，译者注。)

在客体与仪器间的相互作用引起了测量对前者的“干扰”，但也只是且恰好是对后者的经典描述(通过对终态的无知解释)使得这一干扰“不可控”。① 在哥本哈根解释中，概率完全是因为我们通过经典的眼镜来看量子世界而产生的。

“正是测量媒介作用说明了经典线路的必要性，原则上妨碍了在正常量子现象中测量仪器对于原子客体反作用的准确控制。”见[Bohr，1956，87]。

“人们可以认为这些不确定性是客观的，即它们仅仅是我们用经典物理学描述实验这一事实的结果，它们在细节上并不依赖于观测者。人们也可以认为它们是主观的，即它们反映了我们关于世界的不完备知识。”见[Heisenberg，1958，pp. 53—54]。

因而，产生了这样的图景：虽然海森堡界线的量子力学方面是用薛定谔方程描述的(这是决定论的)，而经典的方面满足牛顿定律(这也是决定论的)②，产生不可预言性是因为作为仪器的量子系统被近似为经典系统。因而得到的概率反映了由决定(或需要)去忽略仪器的量子力学自由度而产生的无知。所以量子理论的概率本性不是内在的，而是外在的，且同样完全是经典概念学说的结果，经典概念学说同样也说明了这一本性。

数学上，关于这一观念最简单的例子如下。取有限维希尔伯特空间$\mathcal{H}=\mathbb{C}^n$，因而产生的可观测量代数$\mathcal{A}=M_n(\mathbb{C})$(即$n\times n$矩阵)。单位向量$\Psi\in\mathbb{C}^n$以常规的方式确定了一个量子力学态。现在这样来描述此量子系统，通过忽略除对角化矩阵外的所有可观测量就好像该系统是经典的。则态瞬间塌缩向了在n个点集合上的概率测量，概率由玻恩规则$p(i)=|(e_i,\Psi)|^2$给出，其中e_i $(i=1,\cdots,n)$是$\mathbb{C}^n$的标准基。类似地，用$|\Psi(x)|^2$表示对波函数$\Psi\in L^2(\mathbb{R}^3)$的概率解释，玻恩—泡利规则由此产生，若人们忽略$L^2(\mathbb{R}^3)$上除位置算符外

① 海森堡[1927]没有在其关于不确定关系的论文中对这些要点进行明确区分，但之后玻尔对此进行了澄清。见[Scheibe，1973]。

② 见[Earman，1986；2005]。

的所有可观测量。①

虽然这整个图景很诱人，但它在实践中适用与否尚不可知！对于人们可能确实把量子系统描述为是“好像它是经典的”，不存在任何先验的保证。玻尔和海森堡显然把他们如此热爱的大山和小溪的经典世界的存在看作是理所当然的，前者大概是在经验基础上，后者在他不确定关系的基础上——两人都几乎对涉及的极端复杂的数学和概率问题视而不见。在我们看来，使哥本哈根解释有意义的主要困难在于对涉及的经典描述进行辩护。这一困难是本论文的主要论题，其中第 6 节与当前的讨论非常相关。

3.3 互补性

除了给定链条中界线的精确位置之外，海森堡界线的概念还服从某种任意性，因为原则上可以以许多不同和不相容的方式来构造该链条。玻尔用他称为“互补性”的术语对这一任意性进行了分析。②

玻尔从未给出互补性的精确定义，③ 但他仅限在一些案例中来分析它。④一个突出的案例是量子系统“因果”地⑤描述与时空描述之间的互补性，其中守恒律在该量子系统中成立，时空描述必然具有统计特征。这里玻尔的观点似乎是，一个原子的定态(即能量的本征态)与在时空轨道中运动的电子不相容——

① 技术上，人们将限制 Ψ——将其看作是在 $B(L^2(\mathbb{R}^3))$ 的 C^* 代数上的一个态——为 $C_0(\mathbb{R}^3)$ 的 C^* 代数，由在 $L^2(\mathbb{R}^3)$ 上的所有乘法算子给出，$L^2(\mathbb{R}^3)$ 由在无穷大处等于零的 $x\in\mathbb{R}^3$ 的连续函数定义。这一限制产生了 一个在 $\mathbb{R}^3$ 上的概率测量，这恰好是最初由泡利提出的那个普通的概率测量。

② 不巧且有代表性地，玻尔再一次将互补性看作是思考的必需品，而非将它看作确实所是的真正迷人的可能描述模式。

③ 也许他偏爱这一方法，因为他感觉到一个定义只能够揭示要定义的事物的一部分：他最喜欢的互补性的案例之一是在定义与观测之间的互补性。

④ 我们在这里避免讨论在真理与清晰度、科学与宗教、思想与感觉、客观性和内省之间的互补性，即使有如下事实，即在此基础上玻尔的传记作者派斯[1997]才将玻尔看作是自康德以来最伟大的哲学家。

⑤ 在如下事实的角度看来玻尔对 "因果的" 一词的用法非常混乱，即在英国经验主义传统中因果性常常被在时空描述的意义上来解释。但玻尔的“因果的”指的是与时空描述的互补性！

对该问题的讨论见5.4节。然而，海森堡[1958]认为互补性的这一案例意味着，不在其上做测量的系统确定地按照薛定谔方程演化，而快速的连续测量产生一个时空轨迹，量子理论仅能够统计地预言该轨迹的精确形式[Camilleri，2005]。换句话说，这一案例恰好重现了这样的图景，根据这一图景海森堡[1927]相信他建立了在经典与量子力学间的联系，参见2.3节。

玻尔的另一个主要案例在于粒子与波之间的互补性。这里他的主要目的是说明杨氏双缝实验的意义。该实验在经典图像化中熟知的困难是，粒子描述看起来是不可能的，因为一个粒子不得不经过一个缝，破坏了在检测屏上逐渐建立起来的干涉图样，而波描述似乎与波打到屏幕上的点状位置不相容。因而，玻尔建议说，当这些经典描述中的每一个都不完备时，把它们联合起来以对实验做完备的描述是必要的。

更深层的认识论论点似乎是，虽然对微观系统量子力学进行描述的完备性似乎受到了经典概念学说的危及，但它事实上通过引入两种“互补的”描述(即“若想要达到目的”，对给定量子系统和必须要经典描述的测量装置的描述)得到了修复。不幸的是，虽然这种笼统的观念很有吸引力，但玻尔案例给出的互补性的精确定义并不明晰。首先，在决定论与时空描述之间的互补性概念在只考虑到经典物理学时是相互协调的，但在量子力学中显然相互冲突。其次，在经典物理学中，对某些实体的波动描述与对同一实体的粒子描述相冲突，而量子力学中这些描述以某种方式共同存在。①

沙伊贝[1973，32]注意到“(玻尔的作品)明显地倾向于现象间的互补性这一首选表达”，其中玻尔现象是指量子系统和用以研究它的经典描述的实验安排不可分的联合(或整体)，见下文第2项。玻尔关于互补性的一些早期案例可以归为这一类，而其他的则不可以[Held，1994]。对玻尔的许多学生(包括笔

① 更严峻的是，玻尔以冲突的方式混淆了这些案例。在对原子中电子束缚态的讨论中，他同时将决定论和粒子作为互补对中的一方，波和时空作为另一方。在他对电子—光子散射的描述中，他反过来处理：这次决定论和波形成一方，粒子和时空为另一方，参见[Beller，1999]。

者)而言，迷雾尚需拨开。① 虽然如此，下述数学的解释可能会在冯·诺伊曼量子力学体系的框架中赋予互补性观念以某些意义。②

1. 海森堡[1958]认为量子力学系统的互补性图景与它等价的数学表征等同。例如，他把 x 和 p 的互补性看作为我们现在所谓的对在 $L^2(\mathbb{R}^n)$ 和它向动量空间的傅里叶变换间的正则对易关系(CCR)的薛定谔表征。此外，他感觉到在量子场论中，由于二次量子化，粒子和波给出了量子理论的两种等价描述模式。因而，对于海森堡而言互补性图景是经典的，因为存在基础的经典变量，无须仪器在场，且这样的图景并不相互冲突，而是(一致)等价的。同样见[Camilleri，2005，88]，在她看来"海森堡从未接受玻尔的互补性论证。"

2. 泡利[1933]仅是说，当两个可观测量相应的算符不对易时它们是互补的。③ 因此，它是根据海森堡的不确定关系而得到的，不确定关系即互补的可观测量不能够同时以任意精度被测量。这说明了(但绝不是证明)它们应该独立被测量，通过采用相互排斥的实验安排。互补性的后一特征在玻尔晚期的作品中得到了强调。④ 这一方法使得互补性概念变得清晰且数学上精确，且或许因此有少数确实在他们的工作中应用了互补性概念的物理学家们倾向于追随泡利

① 即使爱因斯坦[1949，674]也承认在他与玻尔争论期间，他从未理解互补性的概念，"此外，我从未能够理解其精确构造，虽然我花了许多精力在其中。"关于玻尔—爱因斯坦争论的观点见[Landsman，2006b]。

② 此运用非常有悖于玻尔的精神，他曾说："冯·诺伊曼的方法……并不解决问题，因而产生了潜在的困难。"见[Scheibe，1973，11]，引用自费伊阿本德(Feyerabend)的论文。

③ 更精确地，人们应该很可能会要求问题中的两个算符生成可观测量的围绕空间代数，从而泡利意义上的互补性真的是在可观测量一给定代数的两个对易子代数间定义的(仍然假设它们共同生成了后者)。

④ 玻尔早期的作品并不完全符合泡利的方法[Scheibe，1973；Held，1994]。在玻尔对双缝实验的讨论中，粒子和波形成了一互补对，而泡利的互补可观测量是位置和动量，它们是玻尔互补对中的一方面。对于在相互排斥实验安排、非对易可观测量和海森堡不确定关系间关系的精确分析，见[Busch *et al.*，1998]和[De Muynck(德马尔凯)，2002]。

和晚期玻尔(的工作)。①

3. 笔者提出，可观测量和纯态是互补的。因为在基本量子力学的薛定谔表征中，大体来说，前者是由位置和动量算符产生的，而后者由波函数给出。玻尔关于互补性的一些其他案例也与这一解释一致(至少在我们忽略测量时的波函数塌缩时)。这里我们捕捉到这样的观念，即互补对的两个方面对于完全描述都是必要的，虽然在可观测量与态之间所谓的相互冲突是模糊的。同样，这种对互补性的解读依赖于正则对易关系的具体表示。能从这种思想体系中得到什么并不是非常清楚，也许需要用更多的细节来推进。例如，在量子场论中，可观测量再一次充当着时空描述，尤其是在哈格[1992]的代数描述中。

3.4 后记：用纠缠态拯救?

既然玻尔的"量子假设"是模糊的与过时的，考虑霍华德[1994]在纠缠态基

① 采用此观点后，很容易通过下述猜想来理解在位置与动量间的互补性：$\mathcal{B}(L^2(\mathbb{R}^n))$上的任意正常纯态$\omega$(即在$C^*$代数意义上被看作态的任意波函数)是由$\{\omega\mid L^\infty(\mathbb{R}^n),\ \omega\mid FL^\infty(\mathbb{R}^n)F^{-1}\}$对来确定的(换言之，通过它对位置与动量的限制)。这里，$L^\infty(\mathbb{R}^n)$是$L^2(\mathbb{R}^n)$上乘法算子的冯·诺伊曼代数，即由位置算符产生的冯·诺伊曼代数，而$FL^\infty(\mathbb{R}^n)F^{-1}$是其傅里叶变换，即由动量算符产生的冯·诺伊曼代数。这里的理念是它对$\omega\mid L^\infty(\mathbb{R}^n)$和$\omega\mid FL^\infty(\mathbb{R}^n)F^{-1}$的每个限制给出了$\omega$的一个经典图景。这些限制是，一个$\mathbb{R}^n$上的测量被解释为位置空间，而$\mathbb{R}^n$上的另一个测量被解释为动量空间。不幸的是，该猜想是错的。巴克霍尔兹(D. Buchholz)给出了(私人通信)下述反例：将ω看作由波函数$\Psi(x)\sim\exp(-ax^2/2)$定义的态，有$\mathrm{Re}(a)>0$，$\mathrm{Im}(a)\neq 0$，和$|a|^2=1$。则ω取决于$\mathrm{Im}(a)$，而$\omega\mid L^\infty(\mathbb{R}^n)$和$\omega\mid FL^\infty(\mathbb{R}^n)F^{-1}$则不依赖于$\mathrm{Im}(a)$。甚至对$2\times 2$矩阵的$C^*$代数的类似猜想也有一个反例，由霍尔沃森发现：若$A$是由$\sigma_x$产生的对易$C^*$代数，$B$由$\sigma_y$生成，则$\sigma_z$的两个不同本征态在$A$和$B$上重合。一种改进我们猜想的方式可能是希望，在薛定谔图像中，若两个态在两个给定的对易冯·诺伊曼代数上总是重合，则它们必须是等价的。但这仅对某些"真实的"时间演化是真的，因为通常的哈密顿量$H=0$产生了上述反例。我们将此问题留待将来研究。在写作时，霍尔沃森[2004]仅囊括了对位置与动量间互补性的合理数学解释，通过将其关联于CCR的表征理论。他表明在位置算符拥有本征态的任意表示中，不存在动量算符，反之亦然。

础上对玻尔物理学哲学的“重建”会有意义。[①] 他的实例或许可以通过诉诸普利马斯[1993]给出的对量子物理学中需要经典概念的分析而得到加强。[②] 普利马斯提议把一个“量子客体”定义为一个不与其环境有着他所谓的“EPR 关联”的物理系统 S。这里“环境”包括仪器、观测者和宇宙的其他可能必要的部分，等等。在基本的量子力学中，这种意义上的量子客体仅在非常特别的态中存在：若$\mathcal{H}_S$是系统 S 的希尔伯特空间，$\mathcal{H}_E$是环境 E 的希尔伯特空间，形如 $\sum_i c_i\Psi_i\otimes\Phi_i$（多于一项）的任何纯态根据定义将 S 与 E 关联；唯一不关联的纯态是那些对于单位向量 $\Psi\in H_S,\Phi\in H_E$ 具有形式 $\Psi\otimes\Phi$ 的态。$S+E$ 上纯的 EPR 关联态局限到 S 上是混合态，从而(想要成为的)量子客体“并不拥有它自己的纯态”。换句话说，EPR 关联态 ω 在 S 的限制与它在 E 的限制并不共同地确定 ω。更一般地，若$S+E$ 的态是 EPR 关联的，对 S 态的完全刻画需要 E(反之也成立)。但(与玻尔相反!)数学要优于文字：对 EPR 关联最清晰的刻画可以用算符代数给出，正如下面所做的。在这种思想下，论文的余下部分将以一种非常普遍与抽象的方式进行。[③] 关于下面部分特别参见[Werner(维纳)，1989]。

---

① 我们很少发现玻尔本人曾那样想的证据。经许可我们引用泽(Zeh)的话，他接着玻尔作出的类似于上文 3.2 节中关于量子假设的陈述，写道：“对量子理论这些早期解释的随后修正(这是更大系统量子纠缠态的重要作用所要求的)似乎未被许多物理学家注意到。”见[Joos *et al.*（尤斯等)，2003，23]。关于纠缠态有趣的历史视角也见[Howard，1990]，实验情形参见[Raimond *et al.*（莱蒙德等)，2001]。

② 也见[Amann and Primas(阿曼和普利马斯)，1997]和[Primas，1997]。

③ Summers and Werner(萨默斯和维纳)[1987]给出了甚至更普遍的结果，其中下面的张量积$\mathcal{A}\otimes\mathcal{B}$被任意 C^* 代数$\mathcal{C}$所取代，后者将$\mathcal{A}$和$\mathcal{B}$包含为 C^* 子代数。

令$\mathcal{A}$和$\mathcal{B}$是 C^* 代数,[①] 有张量积$\mathcal{A}\hat{\otimes}\mathcal{B}$。[②] 如果描述得稍具体些，可考虑如上两个希尔伯特空间$\mathcal{H}_S$和$\mathcal{H}_E$，有张量积$\mathcal{H}_S\otimes\mathcal{H}_E$，假定当$\mathcal{B}$是$\mathcal{B}(\mathcal{H}_E)$本身，或是其(范数闭与对合的)换位子代数时有$\mathcal{A}=\mathcal{B}(\mathcal{H}_S)$。则张量积$\mathcal{A}\hat{\otimes}\mathcal{B}$是$\mathcal{B}(\mathcal{H}_S\otimes\mathcal{H}_E)$的一个(范数闭与对合的)子代数，$\mathcal{B}(\mathcal{H}_S\otimes\mathcal{H}_E)$是在$\mathcal{H}_S\otimes\mathcal{H}_E$上所有有界算符的代数。

在$\mathcal{A}\hat{\otimes}\mathcal{B}$上的积态是具有形式 $\omega=\rho\otimes\sigma$ 的态，其中$\mathcal{A}$上的态 ρ 和$\mathcal{B}$上的态 σ 可以是纯态也可以是混合态。[③] 我们说在$\mathcal{A}\hat{\otimes}\mathcal{B}$的态 ω 是可分解的，当它是积

---

① 回顾 C^* 代数是一个复代数$\mathcal{A}$，$\mathcal{A}$在满足对所有 A，$B\in\mathcal{A}$有 $\|AB\|\leqslant\|A\|\,\|B\|$ 的模 $\|\cdot\|$ 中是完备的，且拥有演化 $A\to A^*$ 满足 $\|A^*A\|=\|A\|^2$。一个基本例子是$\mathcal{A}=\mathcal{B}(\mathcal{H})$，它是希尔伯特空间$\mathcal{H}$上所有有界算子的代数，具备常规的算子模和伴随。根据盖尔范德—奈马克(Gelfand-Naimark)定理，对某希尔伯特空间$\mathcal{H}$而言任意 C^* 代数与$\mathcal{B}(\mathcal{H})$的闭模自伴子代数相同构。另一重要案例是$\mathcal{A}=C_0(X)$，它是局域紧致的豪斯道夫(Hausdorff)空间 X 上的所有连续复值函数空间，X 在无穷大处为零(在如下意义上，即对每个 $\varepsilon>0$ 存在一紧致子集 $K\subset X$ 满足对所有的 $x\notin K$ 有 $|f(x)|<\varepsilon$)，装备上确界模 $\|f\|_\infty:=\sup_{x\in X}|f(x)|$，且演化由(点态的)复共轭给出。根据盖尔范德—奈马克引理，对某个局域紧致的豪斯道夫空间 X 而言，任意对易的 C^* 代数与 $C_0(X)$ 同构。

② 两个(或更多)C^* 代数的张量积不是唯一的，且我们这里需要所谓的射影张量积$\mathcal{A}\hat{\otimes}\mathcal{B}$，它被定义为在最大 C^* 交叉范数中代数张量积$\mathcal{A}\otimes\mathcal{B}$的完备化。射影张量积的选择保证了在$\mathcal{A}\otimes\mathcal{B}$上的每个态通过连续性拓展到$\mathcal{A}\hat{\otimes}\mathcal{B}$上的一个态；相反，因为$\mathcal{A}\otimes\mathcal{B}$在$\mathcal{A}\hat{\otimes}\mathcal{B}$中是紧致的，后者上的每个态都由其在前者上的值所唯一地确定。见[Wegge-Olsen(韦奇—奥尔森)，1993，Appendix T]或[Takesaki，2003，Vol. I，Chapter Ⅳ]。特别地，积态$\rho\otimes\sigma$和其混合态 $\omega=\sum_i p_i\rho_i\otimes\sigma_i$ 如下面所认为的在$\mathcal{A}\hat{\otimes}\mathcal{B}$上是定义明确的。若$\mathcal{A}\subset\mathcal{B}(\mathcal{H}_S)$和$\mathcal{B}\subset\mathcal{B}(\mathcal{H}_E)$是冯·诺伊曼代数，正如拉希奥(Raggio)[1981；1988]分析的，它处理空间张量积$\mathcal{A}\overline{\otimes}\mathcal{B}$很容易(且充分)，$\mathcal{A}\overline{\otimes}\mathcal{B}$被定义为是在$\mathcal{B}(\mathcal{H}_S\otimes\mathcal{H}_E)$中$\mathcal{A}\otimes\mathcal{B}$的二次对易子(或弱完备化)。对于在$\mathcal{A}\otimes\mathcal{B}$上的任意正常态，通过连续性拓展到$\mathcal{A}\overline{\otimes}\mathcal{B}$上的一正常态。

③ 我们采用在代数框架中很平常的态的概念。因而 C^* 代数$\mathcal{A}$上的一个态是一线性函数ρ：$\mathcal{A}\to\mathbb{C}$，它对所有 $A\in\mathcal{A}$在$\rho(A^*A)\geqslant0$ 中都是正的，且在 $\rho(1)=1$ 中是归一化的，其中 1 是$\mathcal{A}$的单位元。若$\mathcal{A}$是冯·诺伊曼代数，则人们有正常态的概念，它满足加法连续性条件。若$\mathcal{A}=\mathcal{B}(\mathcal{H})$，则冯·诺伊曼态的一个基本定理表明，在$\mathcal{A}$上的每个正常态 ρ 是由$\mathcal{H}$上的密度矩阵$\hat{\rho}$给出的，从而对每个 $A\in\mathcal{A}$有 $\rho(A)=\mathrm{Tr}(\hat{\rho}A)$。特别地，对某个单位向量$\Psi\in\mathcal{H}$而言，在$\mathcal{B}(\mathcal{H})$(被看作是冯·诺伊曼代数)上的正常纯态必然拥有形式 $\psi(A)=(\Psi,A\Psi)$。

态的混合态，即当$\omega=\sum_i p_i \rho_i \otimes \sigma_i$时，其中系数$p_i>0$满足$\sum_i p_i=1$。[①] 可分解的态$\omega$仅当它是纯态的积时是纯态。这有着重要的结果，即它分别在$\mathcal{A}$和$\mathcal{B}$上的限制$\omega_{|\mathcal{A}}$和$\omega_{|\mathcal{B}}$都是纯的。[②] 另一方面，在$\mathcal{A}\hat{\otimes}\mathcal{B}$上的态当它不是可分解的时可以说成是EPR关联的[Primas，1983]。一个EPR关联的纯态拥有这样的特性，即它在$\mathcal{A}$或$\mathcal{B}$上的限制是混合态。

拉希奥[1981]证明了下述两个条件是等价的：

· $\mathcal{A}\hat{\otimes}\mathcal{B}$上的每个态都是可分解的；

· $\mathcal{A}$或$\mathcal{B}$是对易的。

换句话说，EPR关联态仅当$\mathcal{A}$与$\mathcal{B}$都是对易的时候才存在。[③] 正如人们可能期望的，这一结果与贝尔不等式紧密相关。即考虑不等式：

$$\left|\sup\{\omega(A_1(B_1+B_2)+A_2(B_1-B_2))\}\right|\leqslant 2 \tag{1}$$

其中ω是在$\mathcal{A}\hat{\otimes}\mathcal{B}$上的定态，上确界是对所有自伴算符$A_1$，$A_2\in\mathcal{A}$，$B_1$，$B_2\in\mathcal{B}$而取的，每个的模都$\leqslant 1$。式(1)当且仅当$\omega$是可分解的时候成立[Baez(贝兹)，1987；Raggio，1988]。所以，在某个态ω下仅当代数$\mathcal{A}$和$\mathcal{B}$都是不对易时不等式(1)才被违背。另一方面，若式(1)满足，则我们知道存在经典的概率空间和概率测度(因而是一个“隐变量”理论)，它们可以重现给定的关联[Pitowsky(皮托斯基)，1989]。正如巴奇伽鲁皮(Bacciagaluppi)[1993]所强调的，这样一个描述并不需要整个背景是经典的；正如我们所看到的，要使贝尔不等式成立，仅需代数$\mathcal{A}$和$\mathcal{B}$中的一个是对易的。

关于玻尔这告诉我们什么？若我们跟随普利马斯[1983]把一个(量子)客体描述为是不与其环境有EPR关联的系统，则刚刚回顾的数学结果带给我们两种可能性。第一，我们可能口头上赞成玻尔把代数$\mathcal{B}$(如上，在最大可能的意义

---

① 这里允许无限和。更精确地，ω是可分解的，若它处于$\mathcal{A}\hat{\otimes}\mathcal{B}$上积态凸包的$\omega^*$闭中。

② $\mathcal{A}\hat{\otimes}\mathcal{B}$上态$\omega$对$\mathcal{A}$的限制$\omega_{|\mathcal{A}}$由$\omega_{|\mathcal{A}}(A)=\omega(A\otimes 1)$给出，其中1是$\mathcal{B}$的单位元，等等。

③ 拉希奥[1981]证明了关于冯·诺伊曼代数和正常态的这一理论。他的论证由巴奇伽鲁皮[1993]推广到C^*代数上。

上，解释为是环境可观测量的代数）当作一种可以看成是对易的描述。在那种情形下，我们的客体就真的是在其某个态中的“客体”。但是显然这并不是唯一的可能性。因为甚至在基本量子力学中——其中有$\mathcal{A}=\mathcal{B}(\mathcal{H}_S)$与$\mathcal{B}=\mathcal{B}(\mathcal{H}_E)$——系统仍然是在普利马斯意义上的“客体”，只要$S+E$的总态$\omega$是可分解的。一般来说，对于纯态这仅意味着$\omega=\psi\otimes\phi$，即总态是纯态的积。为了做到这一点，我们必须以合适的方式来确定海森堡界线，并随后希望给定的积态在时间演化下保持不变，见[Amann and Primas，1997]与[Atmanspacher *et al.*（安特曼斯潘彻等），1999]和其中的文献。这样在$\mathcal{A}$上选择了某些“鲁棒的(robust)”或“稳定的”态，这种方式与退相干方法中的方式相同。我们将在第7节中继续这一讨论(特别见7.1节中的第6点)。

4. 量子化

海森堡[1925]的重新解释(reinterpretation)观念提出，有可能对一个已知其经典描述的物理系统构建量子力学描述。正如我们看到的，这一可能性由薛定谔所实现[1925c]，他在原子物理学情境中建立了对最简单案例(2)和(3)的重新解释。这一早期的案例惊人地成功，几乎所有的原子和分子物理学都建立在它之上。

量子化理论试图理解这一案例，使其在数学上精确，将其推广到更为复杂的系统中。有必要在一开始就说明，正如整个的经典—量子界线，量子化的本性尚未得到很好的理解。这一事实由存在许多相互竞争的量子化过程所体现，我们下面将回顾含义最为清楚的量子化过程。① 其中对量子化首次进行数学上严密讨论的是[Mackey，1968]和[Souriau（苏里奥），1969]，更新且全面的讨论有如[Woodhouse（伍德豪斯），1992；Landsman，1998]和[Ali and Englis（阿里与英吉利），2004]。

① 量子化的路径积分方法仍然在发展中，且目前对基础争论尚无影响，因此我们这里不讨论它。见[Albeverio and Høegh-Krohn（阿尔贝维里奥和霍埃格—克罗恩），1976]和[Glimm and Jaffe（格利姆和杰夫），1987]。

4.1　正则量子化与非本原系(system of imprimitivity)

建立在(2)上的方法常常称为正则量子化。除了数学严密性问题之外，人们可以仅支持麦基[Mackey，1992，283]的观点这边，他写道："这一模型很简洁与优美，但它乍看起来非常的随意并有特设性。很难理解人们如何猜想到它，且一点也不明确该如何修改它以适用于与$\mathbb{R}^r$不同的空间模型。"

冯·诺伊曼揭开了[1931]量子化的神秘面纱，他(沿着海森堡、薛定谔、狄拉克和泡利先前的启发性建议)认识到(2)不仅仅提供了对正则对易关系

$$[Q_\hbar(p_j), Q_\hbar(q^k)] = -i\hbar\delta_j^k \tag{1}$$

的一个表征，同时(满足正则性条件)①也是仅有的不可约化的(根据幺正等价性)这类表征。特别地，由海森堡和薛定谔做出的看似不同的量子理论构造(分别通过引入态和可观测量得到修正，参见第2节)仅涉及对(1)看似不同但幺正等价的表征：在矩阵与波之间的不同可以说只是希尔伯特空间中坐标系的不同。此外，任何其他可想象到的量子力学构造——现在仅被定义为(1)的(正则)希尔伯特空间表征——都需要与海森堡和薛定谔的构造相等价。②

从而，这把在$\mathbb{R}^n$上运动粒子的量子化问题转化为正则对易关系(1)。虽然这些对易关系存在(如它们所示)数学严密的理论[Jørgensen and Moore(约根森和摩尔)，1984；Schmüdgen(施莫真)，1990]，但它们是有问题的。首先，从技术上来说所涉及的算符是无界的，为了表示物理上的可观测量它们必须是自伴的，但在它们各自自伴性的领域内，左手边的对易子是不明确的。第二，也是更为重要的是，(1)依赖于在$\mathbb{R}^n$全域坐标选择的可能性，这排除了向任意位

---

① 要求将无界算符$Q_\hbar(p_j)$和$Q_\hbar(q^k)$整合到$2n+1$维海森堡群H_n的幺正表示中，H_n即为唯一的连通且单连通李群，具有$2n+1$维李代数，其生成子X_i，Y_i，$Z(i=1, \cdots, n)$满足李括号$[X_i, X_j]=[Y_i, Y_j]=0$，$[X_i, Y_j]=\delta_{ij}Z$，$[X_i, Z]=[Y_i, Z]=0$。因而，冯·诺伊曼关于正则对易关系表征的唯一性定理(正如他本人确实认识到的)真的是对H_n幺正表示的唯一性定理，其中中心元素Z映射到$-i\hbar^{-1}1$，其中$\hbar\neq 0$是一固定常数。例如，见[Corwin and Greenleaf(科温和格林利夫)，1989]或[Landsman，1998]。

② 这并不与海森堡绘景和薛定谔绘景间对立的问题相关，它们是关于可观测量的时间演化与态时间演化的对立。

形空间的简单推广。第三，即使我们通过找到(1)的表征而量子化了 p 和 q，其他可观测量的量子化问题仍然存在——考虑到哈密顿量和薛定谔方程。

大约50年前，麦基给自己设置了为正则量子化寻找意义的任务，结果见[Mackey，1968；1978；1992]和下述的简单讨论。虽然笔者现在只是把麦基用约化表征和非本原系术语对量子化的构造看作是通往建立在形变理论和广群上的当前理解的跳板(参见下文4.3节)，物理学基础中常常用到麦基的方法(十分正确地)，建议人们熟悉它。无论如何，麦基[1992，283](继续前一个引用)用一些证明宣称他的量子化方法“去除了许多的神秘”。

像绝大多数量子化方法一样，麦基在量子力学中赋予动量和位置非常不同的角色，虽然在经典力学中 p 和 q 通过正则变换可以相互交换：①

1. 位置算符 $Q_\hbar(q^j)$ 全部被 $\mathbb{R}^n$ 上的单投影值测量 P 所取代，② 它在 $L^2(\mathbb{R}^n)$ 上是作为乘法算子由 $P(E)=\chi_E$ 给出的。给定这一量 P，任意由(可测量的)函数 f: $\mathbb{R}^n\to\mathbb{R}$ 定义的乘法算子可以表征为 $\int_{\mathbb{R}^n} dP(x)f(x)$，它在适当范围内是确定的和自伴的。③特别地，位置算符 $Q_\hbar(q^j)$ 可以通过选择 $f(x)=x^j$ 从 P 中重建，即：

$$Q_\hbar(q^j) = \int_{\mathbb{R}^n} dP(x)x^j \tag{2}$$

2. 动量算符 $Q_\hbar(p_j)$ 全部用一个单幺正群表征 $U(\mathbb{R}^n)$ 取代，在 $L^2(\mathbb{R}^n)$ 上通过 $U(y)\Psi(x) := \Psi(x-y)$ 定义。每个 $Q_\hbar(p_j)$ 都可以通过

$$Q_\hbar(p_j)\Psi := i\hbar\lim_{t_j\to 0} t_j^{-1}(U(t_j)-1)\Psi \tag{3}$$

---

① 对于负号，确实是那样。这在 $\mathbb{R}^n$ 全域上是真的，在任意辛流形上是局域为真的，其中局域达布(Darboux)坐标并不在位置和动量之间做出区分。

② 空间 Ω 上的投影值测量 P 是从波莱尔(Borel)子集 $E\subset\Omega$ 到 $\mathcal{H}$ 上投影的映射 $E\mapsto P(E)$，该空间具有波莱尔结构(即具备通过拓扑定义的可测量集合的 σ 代数)，值在希尔伯特空间 $\mathcal{H}$ 中，该映射对所有可测量的 E，$F\subset\Omega$ 满足 $P(\emptyset)=0$，$P(\Omega)=1$，对所有相互不相交的 $E_i\subset\Omega$ 的所有可数集合 E，$F\subset\Omega$ 满足 $P(\emptyset)=0$，$P(\Omega)=1$，$P(E)P(F)=P(F)P(E)=P(E\cap F)$。对所有相互不相交的 $E_i\subset\Omega$ 的所有可数集合 $P(\cup_{i=1}^{\infty}E_i)=\sum_{i=1}^{\infty}P(E_i)$ 满足。

③ 这一范围包含所有的 $\Psi\in\mathcal{H}$，它们有 $\int_{\mathbb{R}^n} d(\Psi,P(x)\Psi)\,|f(x)|^2<\infty$。

从 U 中重建，其中 $U(t_j)$是在 $x^j=t_j$ 且 $x^k=0$ 时的 U，当 $k\neq j$ 时。①

因此，它意味着用$(P,\ U)$取代$(Q_\hbar(q^k),Q_\hbar(p_j))$并不会失去普遍性。对易关系(1)在这里被

$$U(x)P(E)U(x)^{-1}=P(xE) \tag{4}$$

所取代，其中 E 是$\mathbb{R}^n$ 的一个(波莱尔)子集，且 $xE=\{x\omega \mid \omega\in E\}$。在该重构的基础上，麦基提出了如下对正则对易关系的广泛推广：②

对于给定群 G 在空间 Q 上的作用，非本原系$(\mathcal{H},\ U,\ P)$包含了一个希尔伯特空间$\mathcal{H}$，在$\mathcal{H}$上关于 G 的一个幺正表征 U，和在 Q 上的一个投影值测量 $E\mapsto P(E)$，其值处于$\mathcal{H}$中，从而式(4)对于所有的 $x\in G$ 与所有的波莱尔集 $E\subset Q$ 都成立。

在物理学中，这样一个系统描述了在位形空间 Q 中运动粒子的量子力学，其中 G 在 Q 上通过对称变换起作用，更详细的讨论见 4.3 节。当一切都光滑时③，G 的李代数$\mathfrak{g}$中的每个元素 X 定义了一个在$\mathcal{H}$上的广义动量算符④：

$$Q_\hbar(X)=i\hbar dU(X) \tag{5}$$

这些算符满足广义正则对易关系⑤：

$$[Q_\hbar(X),Q_\hbar(Y)]=i\hbar Q_\hbar([X,Y]) \tag{6}$$

此外，用算符术语表示为：

$$Q_\hbar(f)=\int_Q dP(x)f(x) \tag{7}$$

其中f是在 Q 上的一个光滑函数，且 $x\in\mathfrak{g}$，此外有：

$$[Q_\hbar(X),\ Q_\hbar(f)]=i\hbar Q_\hbar(\xi_X^Q f) \tag{8}$$

---

① 根据斯通(Stone)定理[Reed and Simon，1972]，此算符是在所有 $\Psi\in H$ 集合上定义的且自伴的，在此集合上极限存在。

② 所有的群和空间都被假定是局域紧致的，且作用和表征被假定是连续的。

③ 即 G 是李群，Q 是流形，且 G 作用是光滑的。

④ 此算符是在向量 $\Psi\in\mathcal{H}$范围上定义且自伴的，在此 $dU(X)\Psi:=\lim_{t\to 0}t^{-1}(U(\exp(tX))-1)\Psi$存在。

⑤ 正如前面在(1)的背景中注意的，对易关系(6)、(8)和(9)并不在所涉算符的自伴性范围内成立，但在更小的共同核心范围内成立。

其中 ξ_X^Q 是 Q 上通过 G 作用①定义的一个正则向量场，且：

$$[Q_\hbar(f_1),\ Q_\hbar(f_2)]=0 \tag{9}$$

在$\mathbb{R}^n$ 上的初等量子力学对应于 $Q=\mathbb{R}^n$ 与 $G=\mathbb{R}^n$ 时的特殊情形，具有常规的加法群结构。为了表明这一点，我们用名称(p_j)来表示$\mathbb{R}^3$ 的标准基(在其外观上是$\mathbb{R}^3$ 的李代数)，且取$f_1(q)=q^j$，$f_2(q)=f(q)=q^k$。对于 $X=p_j$ 且 $Y=p_k$ 来说，等式(6)则写作$[Q_\hbar(p_j),\ Q_\hbar(p_k)]=0$，等式(8)产生正则对易关系式(1)，且式(9)描述了位置算符的对易性，即$[Q_\hbar(q^j),\ Q_\hbar(q^k)]=0$。

为了涵盖自旋，取 $G=E(3)=SO(3)\ltimes\mathbb{R}^3$，即欧几里得运动群(Euclidean motion group)，它以明显(定义)的方式作用于 $Q=\mathbb{R}^3$。$E(3)$的李代数是作为向量空间的$\mathbb{R}^6=\mathbb{R}^3\times\mathbb{R}^3$；我们通过对$\mathbb{R}^3$——它外观为 $SO(3)$的李代数——的第一个拷贝基(J_i)来扩展$\mathbb{R}^3$(即$\mathbb{R}^3$ 的李代数)的第二个拷贝基(p_j)，发现$Q_\hbar(J_i)$仅仅是通常的角动量算符。②

麦基对关于正则对易关系(1)不可约化表征的冯·诺伊曼[1931]唯一性定理的推广是他的非本原性(imprimitivity)定理。该定理适用于对某些(闭的)子群$H\subset G$有 $Q=G/H$ 的特殊情形，且该定理表明(依据幺正等价性)下面二者间存在双射对应：

1. 在 G/H 上左平移 G 的非本原系$(\mathcal{H},\ U,\ P)$；
2. H 的幺正表征 U_χ。

这一对应保留了不可约化性。③

例如，冯·诺伊曼的定理通过选择 $G=\mathbb{R}^3$ 和 $H=\{e\}$(从而，如上 $Q=\mathbb{R}^3$)就恢复为麦基定理的特殊情形：这里正则对易关系的(正则)不可约表征的唯一

---

① 即 $\xi_X^Q f(y)=d/dt\,|_{t=0}[f(\exp(-tX)y)]$。

② 前面段落中的对易关系现在通过熟悉的关系$[Q_\hbar(J_i),\ Q_\hbar(J_j)]=i\hbar\epsilon_{ijk}Q_\hbar(J_k)$，$[Q_\hbar(J_i),\ Q_\hbar(p_j)]=i\hbar\epsilon_{ijk}Q_\hbar(p_k)$和$[Q_\hbar(J_i),\ Q_\hbar(q^j)]=i\hbar\epsilon_{ijk}Q_\hbar(q^k)$进行了推广。

③ 特别地，给定 U_χ，三元数$(\mathcal{H}^\chi,\ U^\chi,\ P^\chi)$是非本原系，其中$\mathcal{H}^\chi=L^2(G/H)\otimes\mathcal{H}_\chi$携带了由 $U_\chi(H)$引起的表征 $U^\chi(G)$，且 P^χ 就像乘法算子那样作用。反过来，若$(\mathcal{H},\ U,\ P)$是非本原系，则存在幺正表征 $U_\chi(H)$使得三元数$(\mathcal{H},\ U,\ P)$幺正等价于刚刚描述的三元数$(\mathcal{H}^\chi,\ U^\chi,\ P^\chi)$。例如，对 $G=E(3)$和 $H=SO(3)$，人们有$\chi=j=0,\ 1,\ 2,\ \cdots$和$\mathcal{H}^j=L^2(\mathbb{R}^3)\otimes\mathcal{H}_j$，其中$\mathcal{H}_j=\mathbb{C}^{2j+1}$携带了给定表征 $U_j(SO(3))$。

性是从平凡群的不可约表征唯一性中得到的。更形象的例子是$G=E(3)$和$H=SO(3)$（从而有$Q=\mathbb{R}^3$），在这里，相关非本原系的不可约表征通过自旋$j=0$，1，…来分类。麦基把这看作是对自旋表现为纯量子力学自由度的解释。① 虽然自旋没有经典类似物的观点广泛地被量子理论的先驱们所接受，② 它现在已经过时了（见下文4.3节）。尽管有此不幸的错误解释，麦基的正则量子化方法的强有力与清晰性很难被超越，并且有许多有趣的应用。③

我们仅提及在物理学哲学中特别有趣的一个，即牛顿—维格纳位置算符，正如怀特曼（Wightman）[1962]所分析的。④ 这里，普遍的问题是，在某些希尔伯特空间$\mathcal{H}$上$G=E(3)$的给定幺正表示U能否拓展到一个关于$H=SO(3)$（且所以有$Q=\mathbb{R}^3$，如上）的非本原系统，在那种情形下，U（而不是相关的量子系统）被认为是在$\mathbb{R}^3$上被定域化。根据维格纳（Wigner）（1939）的建议，即一个相对论粒子是由庞加莱（Poincaré）群P的不可约化表征U所描述的，我们通过将$U(P)$限制到子群$E(3)\subset P$⑤ 得到表征$U(E(3))$。则按照之前的分析，由$U(P)$描述的粒子是定域化的，当且仅当$U(E(3))$是由$SO(3)$的某个表征得到的。当然，这一问题可以解决，其结果是大而重的粒子的任意自旋可以定域化在$\mathbb{R}^3$中（相应的位置算符恰好是牛顿—维格纳算符），但是无质量的粒子当且

① 根据通常的论证（维格纳定理），人们可以用$SU(2)$取代$SU(3)$，从而得到$j=0$，1/2，…。

② 这一观点要追溯到泡利[1925]那里，他谈论了一种"在电子的量子理论特征（没有经典描述）中的矛盾"，该特征随后被古德米特和乌伦贝克认为是自旋。大概第一个关注自旋的经典对应的是苏里奥[1969]。另一个关于自旋的误解是，其最终的解释必须在相对论的量子力学中找到。

③ 这要求审查关于麦基非本原性定理"最好的"可能论证。麦基自己的论证相当有利于测量理论，且并未阐明其结果的来源。大概最短的论证来自于[Orsted（奥斯特），1979]，但简短论证给出的内容也相当有限。正好相反，真正清晰的论证将数学陈述简化为同义反复。然而，这样的论证往往要求一强大的机制以使此简化奏效。此风格的两种通往非本原性定理的不同路径见[Echterhoff *et al.*（埃希特霍夫等），2002]和[Landsman，2006a]。

④ 弗莱明与巴特菲尔德（Fleming and Butterfield）[2000]给出了在相对论量子理论中对粒子定域化的最新介绍。也见[De Bièvre（德比埃夫勒），2003]。

⑤ 严格来讲，这取决于在闵可夫斯基空间中对惯性系的选择，再加上相应调整的坐标，从而使得问题中的位形空间$\mathbb{R}^3$由$x^0=0$给出。

仅当它们的螺旋度小于 1 时才可以被定域化在$\mathbb{R}^3$中。特别地，光子(和引力子)不能够在此意义上被定域化在$\mathbb{R}^3$中。①

为了鉴别我们之后给出的关于相空间量子化与形变量子化的资料，给出麦基方法的 C^* 代数重构是有帮助的。首先，根据谱定理[Reed and Simon，1972；Pedersen(佩德森)，1989]，空间 Q 上在希尔伯特空间$\mathcal{H}$中取值的投影值测量 $E\mapsto P(E)$，通过对应(7)式等价于$\mathcal{H}$上对易 C^* 代数 $C_0(Q)$ 的非退化表征 π。② 其次，若此外$\mathcal{H}$也携带了 G 的幺正表征 U，非本原系的定义条件(4)(给定在 Q 上的 G 作用)等于协变条件：

$$U(x)Q_\hbar(f)U(x)^{-1}=Q_\hbar(L_xf) \tag{10}$$

对于所有的 $x\in G$ 和 $f\in C_0(Q)$，其中 $L_xf(m)=f(x^{-1}m)$。因而对 Q 给定的 G 作用而言，非本原系就与一个协变的非退化表征 $C_0(Q)$“相同”了。第三，从 Q 上的 G 作用我们能够构建一个特定的 C^* 代数 $C^*(G,Q)$，即由作用定义的所谓的变换群 C^* 代数，它具有这样的特征，即它的非退化表征双射地(且“自然地”)对应于 $C_0(Q)$ 的协变非退化表征，且因而对应于给定 G 作用的非本原系[Effros and Hahn(爱弗罗斯和哈恩)，1967；Pedersen，1979；Landsman，1998]。在量子物理学的 C^* 代数方法中，$C^*(G,Q)$是在 Q 上运动的一个粒子可观测量的代数，满足由 G 作用定义的对称性，它的非等价不可约表征对应于系统可能的超选分支(sectors)[Doebner and Tolar(多布纳和托拉尔)，1975；Majid(马吉德)，1988；1990；Landsman，1990a；1990b；1992]。③

---

① 将光子看作是量子化的光波，垂直于其传播方向有两个可能的偏振，此最后的结果在物理上是非常合理的。

② 希尔伯特空间$\mathcal{H}$上的 C^* 代数$\mathcal{A}$的一个表示是一个线性映射 π: $\mathcal{A}\to B(\mathcal{H})$，满足对所有的 A，$B\in\mathcal{A}$有$\pi(AB)=\pi(A)\pi(B)$和 $\pi(A^*)=\pi(A)^*$。这样一个表征被称为非退化的，当对所有的$A\in\mathcal{A}$，$\pi(A)\Psi=0$ 意味着 $\Psi=0$ 时。

③ 麦基方法的另一个重构，或不如说是对它的拓展，是由艾沙姆(Isham)[1984]给出的。在试图将整个理论简化为群表征中的一个问题的过程中，他提出拥有位形空间 G/H 的粒子的可能量子化是由“正则群”$G_C=G\ltimes V$ 的非等价不可约表征给出的，其中 V 是携带了对 G 表征的最低维的向量空间，在其中 G/H 是偶向量空间 V^* 中的轨道。所有相关的非本原系事实上对应于 G_C 的幺正表示，但此群拥有许多其他表征，它们的物理解释是模糊的。

4.2 相空间量子化与相干态

在麦基的量子化方法中，Q 是系统的位形空间；相关的位置坐标对易，参见式(9)。这体现在上述讨论的在 Q 上的投影值测量与对易的 C^* 代数 $C_0(Q)$ 表征之间的对应。通过在图景中增加对称群 G 并强化关系式(4)或等价地，式(8)或式(10)，就引入了量子力学中典型的可观测量的非对易性(和相关的不确定关系)。正如我们指出的，这一过程破坏了在经典力学中相空间变量位置与动量间的对称性。

麦基方法的这种多少不能令人满意的特征可以通过用系统的相空间代替 Q 而避免，该相空间下文称为 M。① 在这一方法中，通过对麦基方案作一背叛性的微小修正而引入非对易性。即，他借以开始的在 M 上的投影值测量 $E \mapsto P(M)$ 现在被 M 上的正定算符值测量(positive-operator-valued measure)或 POVM 取代，仍然在某希尔伯特空间$\mathcal{K}$上取值。这是一个从 M 的(波莱尔)子集 E 向$\mathcal{K}$上的正定有界算符集合体的映射 $E \mapsto A(E)$②，对不相交的波莱尔集 E_i 的任意可数集合体而言满足 $A(\emptyset)=0$，$A(M)=1$ 和 $A(\cup_i E_i) = \sum_i A(E_i)$。③ 一个对所有(波莱尔)$E$，$F \subset M$ 满足 $A(E\cap F)=A(E)A(F)$ 的 POVM 恰好是一个投影值测量，故 POVM 是对后者的推广。④ 关键在于，一个给定的 POVM 如下来定义一个量子化的过程，即规定一个经典可观测量 f(即在相空间 M 上的可测函

---

① 这里读者可以考虑到最简单的情形 $M=\mathbb{R}^6$，粒子的 p 和 q 空间在$\mathbb{R}^3$ 上移动。更一般地，若 Q 是位形空间，相关的相空间是余切丛 $M=T^*Q$。即使更一般的相空间，也就是任意辛流形，也可以包含于该理论中。关于结果的文献见[Busch *et al.*, 1998; Schroeck (施罗克), 1996]和[Landsman, 1998; 1999a]。

② 在$\mathcal{K}$上的有界算符 A 被称为是正定的，当对所有的 $\Psi \in \mathcal{K}$有 $(\Psi, A\Psi) \geq 0$。因此，它是自伴的，谱包含于$\mathbb{R}^+$ 中。

③ 这里无限和是在弱算符拓扑中取的。注意上述条件要求 $0 \leq A(E) \leq 1$，在此意义上对所有的 $\Psi \in \mathcal{K}$有 $0 \leq (\Psi, A(E)\Psi) \leq (\Psi, \Psi)$。

④ 这产生了所谓的量子理论操作方法，其中可观测量并不用自伴算符(或等价的，用它们相关的投影值测量)表征，而是用 POVM 表征。在其上 POVM 定义的空间 M 是测量装置结果的空间；POVM 由 A 和此装置的校准过程同时决定。在态 ρ 下实验结果处于 $E \subset M$ 中的概率可被取作 $\mathrm{Tr}(\rho A(E))$。见[Davies, 1976; Holevo (霍尔夫), 1982; Ludwig, 1985; Schroeck, 1996; Busch *et al.*, 1998]和[De Muynck, 2002]。

数，出于简单性假定其有界）由下述算符来量子化：①

$$Q(f) = \int_M dA(x)f(x) \tag{11}$$

因而，从位形空间上的投影值测量到相空间上的正定算符值测量间的表面上的微小变动带给量子化一个全新的前景，事实上将这一任务简化为找到这样一个 POVM 的问题。②

这一问题的解通过奈马克的延拓（dilation）定理得到了很大的简化。③ 这表明，在希尔伯特空间$\mathcal{K}$中，给定 M 上的 POVM$E \mapsto A(E)$，则存着一个希尔伯特空间$\mathcal{H}$，携带了 M 上的投影值测量 P 和一个等距入射（isometric injection）$\mathcal{K} \hookrightarrow \mathcal{H}$，满足对于所有的 $E \subset M$（其中$[\mathcal{K}]$是从$\mathcal{H}$向$\mathcal{K}$的正交投影）有：

$$A(E) = [\mathcal{K}]P(E)[\mathcal{K}] \tag{12}$$

把这与麦基的非本原性定理结合起来，得到对后者的有力推广[Poulsen（波尔森），1970；Neumann，1972；Scutaru（斯库塔鲁），1977；Cattaneo（卡塔内奥），1979；Castrigiano and Henrichs（卡斯特奇诺和亨里克斯），1980]。首先，对空间 M 的群 G 给定作用定义一个广义非本原系（$\mathcal{K}$，U，A），它作为 M 上的 POVM A 在希尔伯特空间$\mathcal{K}$中取值，和 G 在$\mathcal{K}$上的幺正表征 V 满足对所有的 $x \in G$ 和 $E \subset M$：

$$V(x)A(E)V(x)^{-1} = A(xE) \tag{13}$$

参见(4)。现在假定 $M = G/H$（和 M 上相关的正则左作用），则广义非本原性定理表明，对此作用而言广义非本原系（$\mathcal{K}$，V，A）必然是（幺正等价于）对

---

① 定义式(11)右边的最简单方式是确定 $\Psi \in \mathcal{K}$，并通过 $p_\Psi(E) = (\Psi, A(E)\Psi)$ 定义 M 上的概率测量 p_Ψ。则人们将 $Q(f)$ 通过其期望值 $(\Psi, Q(f)\Psi) = \int_M dp_\Psi(x)f(x)$ 定义为算符。表述(11)推广了(7)，且同样推广了算符 $f(A) = \int_{\mathbb{R}} dP(\lambda)f(\lambda)$ 的谱求解，其中 P 是通过自伴算符 A 定义的投影值测量。

② Q的一个重要特征是它在下述意义上是正定的，即如果对 $x \in M$，$f(x) \geqslant 0$ 成立，则对所有 $\Psi \in \mathcal{K}$有 $(\Psi, Q(f)\Psi) \geqslant 0$。换言之，$Q$作为一个从 C^* 代数 $C_0(M)$ 向 C^* 代数 $\mathcal{B}(\mathcal{H})$ 的投影，是正定的。

③ 例如见[Riesz and Sz.-Nagy（里斯和纳吉），1990]。然而，更好的是将奈马克定理看作是斯坦斯普林（Stinespring）定理的特殊情形，正如在[Landsman，1998]和下文中说明的那样。

相同作用下非本原系($\mathcal{H}$, U, P)的简化。换言之，奈马克定理中的希尔伯特空间$\mathcal{H}$携带了一个幺正表征 $U(G)$，它与投影[$\mathcal{K}$]对易，且表征 $V(G)$只是 U 对$\mathcal{K}$的限制。此外，POVMA 具有形式(12)。麦基的非本原性定理完全描述了($\mathcal{H}$, U, P)的结构，因而人们拥有了对广义非本原系的完全分类。① 人们有：

$$\mathcal{K} = p\,\mathcal{H};\ \mathcal{H} = L^2(M) \otimes \mathcal{H}_\chi \tag{14}$$

其中 L^2 是相对于在 $M = G/H$ 上的适当测量而定义的，② 希尔伯特空间$\mathcal{H}_\chi$携带了 H 的一个幺正表征，且 p 是在由 $U_\chi(G)$引起的表征 $U^\chi(G)$的交换子群(commutant)中的投影。③ 量子化(11)由下式给出：

$$Q(f) = p\,f\,p \tag{15}$$

其中 f 作为乘法算符作用于 $L^2(M) \otimes \mathcal{H}_\chi$，即$(f\Psi)(x) = f(x)\Psi(x)$。特别地，对相空间的区域 $E \subset M$ 有 $P(E) = \chi_E$(作为乘法算符)，从而有 $Q(\chi_E) = A(E)$。因此，在态 ρ(即在$\mathcal{K}$上的密度矩阵)下，系统定域化于 E 的概率由$\mathrm{Tr}(\rho A(E))$给出。

在一种比麦基方法更自然的方式中，协变的 POVM 量子化允许人们一开始就引入时空对称，通过取 G 为伽利略群或庞加莱群，并选择 H 使得 G/H 是一个物理的相空间(G 正则作用于其上)。见[Ali *et al.*, 1995]和[Schroeck, 1996]。

另一种在相空间(其中出现与前一对称性有交叠的对称性)上构造 POVM 的有效方法④建立在相干态上⑤。对相空间 M 而言希尔伯特空间$\mathcal{H}$中相干态的

---

① $V(G)$必须是由 $U_\chi(H)$引起的某表征 $U^\chi(G)$的子表征。

② 在 G/H 是辛的(从而它典型地是 G 的余伴随轨道)物理相关情形中，人们应该取刘维尔(Liouville)测量的倍数。

③ $U^\chi(g)$(其中 $g \in G$)的明确形式依赖于对投影 π: $G \to G/H$(即 $\pi \circ \sigma = \mathrm{id}$)截面的选择 σ: $G/H \to G$。若在 G/H 上定义 $L^2(G/H)$的测量是 G 不变的，则精确的表述是 $U^\chi(g)\Psi(x) = U_\chi(s(x)^{-1}gs(g^{-1}x))\Psi(g^{-1}x)$。

④ 假定存在一个向量 $\Omega \in \mathcal{K}$相对于某个截面 σ: $G/H \to G$ 和一个 G 不变测量 μ 满足 $\int_{G/H} d\mu(x)\,|(\Omega, V(\sigma(x))\Omega)|^2 < \infty$，且对于所有 $h \in H$ 有 $V(h)\Omega = U_\chi(h)\Omega$，其中 U_χ: $H \to \mathbb{C}$ 是一维的。则(取$\hbar = 1$)向量 $V(\sigma(x))\Omega$(适当归一化的)构成了在 G/H 上的一相干态族[Ali *et al.*, 1995; Schroeck, 1996; Ali, Antoine and Gazeau(阿里、安托万和盖泽奥), 2000]。例如，海森堡群的相干态(20)具有此形式。

⑤ 对相干态的一般讨论见[Klauder and Skagerstam(克劳德和史卡格斯丹), 1985; Perelomov(佩列洛莫夫), 1986; Odzijewicz(奥德瓦尔兹), 1992; Paul and Uribe(保罗和乌里韦), 1995; 1996; Ali *et al.*, 1995]和[Ali *et al.*, 2000]。

最小定义是对普朗克常数$\hbar$的某个固定值，目前人们有入射[①] $M \hookrightarrow \mathcal{H}$，$z \mapsto \Psi_z^{\hbar}$，使得对所有的 $z \in M$ 有：

$$\|\Psi_z^{\hbar}\| = 1 \tag{16}$$

且对单位模的每一个 $\Phi \in \mathcal{H}$(这里 μ_L 是在 M 上的刘维尔测量，且 $c_{\hbar} > 0$ 是一适当常数)有：[②]

$$c_{\hbar}\int_M d\mu_L(z) \mid (\Psi_z^{\hbar}, \Phi) \mid^2 = 1 \tag{17}$$

条件(17)保证了我们可以通过下式在$\mathcal{K}$中定义一个在 M 上的 POVM：[③]

$$A(E) = c_{\hbar}\int_E d\mu_L(z)[\Psi_z^{\hbar}] \tag{18}$$

式(11)则简单地解读为(为了稍后的应用添加Q的$\hbar$依赖性和上标 B)：

$$Q_{\hbar}^B(f) = c_{\hbar}\int_M d\mu_L(z) f(z)[\Psi_z^{\hbar}] \tag{19}$$

薛定谔[1926b]提出的由来已久的例子是 $M = \mathbb{R}^{2n}$，$\mathcal{H} = L^2(\mathbb{R}^n)$，且：

$$\Psi_{p,q}^{\hbar}(x) = (\pi\hbar)^{-n/4} e^{-ipq/2\hbar} e^{ipx/\hbar} e^{-(x-q)^2/2\hbar} \tag{20}$$

方程(17)则在 $d\mu_L(p, q) = (2\pi)^{-n} d^n p d^n$ 和 $c_{\hbar} = \hbar^{-n}$时成立。人们可以证明$Q_{\hbar}^B(p_j)$和$Q_{\hbar}^B(q^j)$与薛定谔算符式(2)相一致。此例子表明相干态不需要是相互正交的。事实上，采用表达 $z = p + iq$，对于式(20)中的态有：

$$\mid (\Psi_z^{\hbar}, \Psi_{\omega}^{\hbar}) \mid^2 = e^{-\mid z-\omega \mid^2/2\hbar} \tag{21}$$

此结果的意义稍后会给出。

在一般情形中，很容易对 POVM(18)证明奈马克延拓定理。改变记号从而向量 $\Psi_z^{\hbar}$ 现在位于$\mathcal{K}$中，人们发现：

$$\mathcal{H} = L^2(M, c_{\hbar}\mu_L) \tag{22}$$

嵌入 W：$\mathcal{K} \hookrightarrow \mathcal{H}$由$(W\Phi)(z) = (\Psi_z^{\hbar}, \Phi)$给出。$\mathcal{H}$上的投影值测量 P 只是 $P(E) = \chi_E$(作为一乘法算符)，且向 $W\mathcal{K}$上的投影 p 由下式给出：

---

① 此入射作为从 M 到$\mathbb{P}\mathcal{H}$——$\mathcal{H}$的投影希尔伯特空间——的映射必须是连续的。

② 这里也可能发生其他测量，例如见[Bonechi and De Bièvre(布奈奇和德比埃夫勒)，2000]。

③ 回顾$[\Psi]$是对单位向量 Ψ 的正交投影。

$$p\Psi(z) = c_{\hbar}\int_M d\mu_L(w)(\Psi_z^{\hbar}, \Psi_w^{\hbar})\Psi(w) \tag{23}$$

因此，式(19)幺正地等价于式(15)，其中f作为一乘法算符作用于$L^2(M)$上。①

因而式(15)、式(22)或其可能的扩展式(14)构成了相空间量子化的本质。

我们以与前一节相同的方式结束此节，即通过指出 POVM 的 C^* 代数意义。这是相当容易的。在$\mathcal{H}$中 M 上的投影值测量与在$\mathcal{H}$中 $C_0(M)$的非退化表征相同，在希尔伯特空间$\mathcal{K}$中 M 上的 POVM 只不过是非退化的完全正定映射 φ: $C_0(M)\to\mathcal{B}(\mathcal{K})$。② 因此，奈马克的延拓定理成为了斯坦斯普林［1955］定理的一种特殊情形：若Q: $\mathcal{A}\to\mathcal{B}(\mathcal{K})$是完全正定映射，存在一个希尔伯特空间$\mathcal{H}$携带有对 $C_0(M)$的表征 π 和一个等距入射$\mathcal{K}\hookrightarrow\mathcal{H}$，满足对所有的$f\in C_0(M)$有$Q(f) = [\mathcal{K}]\pi(f)[\mathcal{K}]$。采用$Q(C_0(M))$术语，协变条件(13)变为 $U(x)Q(f)U(x)^{-1} = Q(L_xf)$，如同式(10)。

4.3　形变(deformation)量子化

目前为止，我们在一种启发的方式上使用“量子化”一词，把我们的理解建立在历史的连续性上，而非公理化的基础之上。在本节与下一节中我们将通过引入两种备选的以公理化方式审视量子化的方案来澄清问题。我们从时间上最新的方法开始，但该方法在概念上与刚刚讨论的材料更为接近。这就是形变量子化，它起源于贝雷辛(Berezin)［1974；1975a；1975b］、维伊(Vey)［1975］和拜耶等(Bayen *et al.*)［1977］的工作。这里我们沿着瑞菲尔(Rieffel)［1989a；1994］提出的形变量子化的 C^* 代数方法，因为它不仅在数学上含义清楚且精

---

① 这引起了在相干态与具有一个重构核的希尔伯特空间之间的紧密关联，见［Landsman，1998］或［Ali *et al.*，2000］。

② 在 C^* 代数间的映射 φ: $\mathcal{A}\to\mathcal{B}$被称为是正定的，每当对 $A\geq 0$ 有 $\varphi(A)\geq 0$ 时；这样一个映射被称为是完全正定的，若对于所有的 $n\in\mathbb{N}$ 有映射 φ_n: $\mathcal{A}\otimes M_n(\mathbb{C})\to\mathcal{B}\otimes M_n(\mathbb{C})$是正定的。这里$M_n(\mathbb{C})$是$n\times n$ 复矩阵的 C^* 代数，该映射用基本张量上 $\varphi\otimes\mathrm{id}$ 的线性拓展来定义。当$\mathcal{A}$是对易的时，非退化正定映射$\mathcal{A}\to\mathcal{B}$对任意$\mathcal{B}$自动是完全正定的。

确，而且也相当接近于物理实践。[①] 由于采用了数学语言，这种方法当然很自然地属于量子物理学的普遍 C^* 代数方法。

形变量子化的关键理念是量子化应该通过具有正确经典极限的特征来定义。所以，普朗克“常数”$\hbar$被当作一个变量，从而对它的每一个值我们都应该拥有一种对应的量子理论。关键的要求是当$\hbar\to 0$ 时该量子理论家族收敛于基础的经典理论。[②] 对此理念的数学实现非常漂亮，即可观测量的经典代数“黏合(glued)”到可观测量的量子代数族中，以这样的方式即经典理论真正形成了包含恰当量子理论（对于$\hbar>0$ 的每一个值都有一个理论）的空间边界。技术上，这是通过 C^* 代数连续场的概念完成的。[③] 接下来的内容好像不需要技术化，但近 15 年来的研究表明这样做恰好产生了关于量子化的正确定义。

令 $I\subset\mathbb{R}$ 是$\hbar$在其中取值的集合，人们通常有 $I=[0,1]$，但当相空间是紧致的时，$\hbar$常常在$(0,1]$的一个可数子集中取值。[④] 相同的情形发生在无限系统理论中，见第 6 节。无论如何，I 应该包含 0 作为极限点。从而，I 上的 C^* 代数连续场包括了 C^* 代数$\mathcal{A}$，C^* 代数$\{\mathcal{A}_\hbar\}_{\hbar\in I}$的一个集合，和对每个$\hbar\in I$ 而言的满射同态(surjective morphism)$\varphi_\hbar$：$\mathcal{A}\to\mathcal{A}_\hbar$，从而有：

1. 对所有的 $A\in\mathcal{A}$，在 $C_0(I)$ 中函数$\hbar\mapsto\|\varphi_\hbar(A)\|_\hbar$；[⑤]
2. 任意 $A\in\mathcal{A}$的模为 $\|A\|=\sup_{\hbar\in I}\|\varphi_\hbar(A)\|$；
3. 对任意的$f\in C_0(I)$和$A\in\mathcal{A}$，存在着元素$fA\in\mathcal{A}$，对它而言对所有的$\hbar\in I$有 $\varphi_\hbar(fA)=f(\hbar)\varphi_\hbar(A)$。

---

① 对形变量子化 C^* 代数方法的广泛讨论也见[Landsman，1998]。在形变量子化的其他方法中，如星花积理论中，$\hbar$是形式参数而非实参数。特别是，极限$\hbar\to 0$ 的意义是模糊的。

② 对此极限的更多评论参见第 5 节的前言。

③ 相同概念的三种不同方法见[Dixmier(狄克斯梅尔)，1977；Fell and Doran(菲尔和多兰)，1988]和[Kirchberg and Wassermann(基希贝格与瓦塞尔曼)，1995]。我们的定义遵从后者，用一任意局域紧致的豪斯道夫空间替代 I，就会得到普遍定义。

④ 参见[Landsman，1998]，但相比较非对易环面(torus)的案例见[Rieffel，1989a]，在那里人们量子化每个$\hbar\in(0,1]$的紧致相空间。此类型的更多案例见夏目与内斯特(Natsume and Nest)[1999]，夏目、内斯特和英格特(Natsume，Nest and Ingot)[2003]及霍金斯(Hawkins)[2005]的讨论。

⑤ 这里 $\|\cdot\|_\hbar$是在 C^* 代数$\mathcal{A}_\hbar$中的模。

理论要点是 C^* 代数族$(\mathcal{A}_\hbar)_{\hbar\in I}$是通过确定在从 $\amalg_{\hbar\in[0,1]}\mathcal{A}_\hbar$ 上的一个拓扑而聚集在一起的(不相交并集)。然而，事实上这一拓扑的定义非常迂回，通过对丛的连续部分空间的确定来定义。① 即，场的连续部分根据定义是 $\prod_{\hbar\in I}\mathcal{A}_\hbar$ 的一个元素$\{A_\hbar\}_{\hbar\in I}$(等价地，是映射$\hbar\mapsto A_\hbar$，其中 $A_\hbar\in\mathcal{A}_\hbar$)，对于 $\prod_{\hbar\in I}\mathcal{A}_\hbar$ 而言存在 $A\in\mathcal{A}$，它满足对所有的$\hbar\in I$有 $A_\hbar=\varphi_\hbar(A)$。其结果是，C^* 代数$\mathcal{A}$可能真的与场的连续部分空间相统一，若我们这样做的话，同构 $\varphi_\hbar$ 仅仅是在$\hbar$上的赋值映射。②

物理上，$\mathcal{A}_0$是底层经典系统可观测量的对易代数，且对每个$\hbar>0$，非对易 C^* 代数$\mathcal{A}_\hbar$被假定是在普朗克常数的值为$\hbar$时对应量子系统的可观测量代数。则代数$\mathcal{A}_0$具有形式 $C_0(M)$，其中 M 是定义经典理论的相空间。相空间比任意拓扑空间拥有更多的结构，它是流形，在其上可以定义泊松(Poisson)括号$\{,\}$。例如，在 $M=\mathbb{R}^{2n}$上人们有熟悉的表达：

$$\{f,g\} = \sum_j \frac{\partial f}{\partial p_j}\frac{\partial g}{\partial q^j}-\frac{\partial f}{\partial q^j}\frac{\partial g}{\partial p_j} \tag{24}$$

技术上，M 被当作是一泊松流形。这是一个在 $C^\infty(M)$上装备了李括号$\{,\}$的流形，拥有如下特征，即对任意的$f\in C^\infty(M)$，映射 $g\mapsto\{f,g\}$确定了一个对由点态乘法给出 $C^\infty(M)$对易代数结构的推导。因此此映射由向量空间 ξ_f 给出，称为f的哈密顿量向量场，即有 $\xi_f g=\{f,g\}$。辛流形是泊松流形的特殊例子，用哈密顿量向量场穷尽了切丛特征来刻画。一个泊松流形通过其辛叶子来分叶(foliate)：给定辛叶子 L 用下述特征来刻画，即在每个 $x\in L$ 处切空间 T_xL

---

① 这是有关对易 C^* 代数的盖尔范德—奈马克定理的重申，它通过 C^* 代数 $C_0(X)$在一局域紧致豪斯道夫空间 X 上确定了拓扑。简单说，在(局域平凡的)向量束理论中，塞尔—斯万(Serre-Swan)定理允许人们根据 E 的连续截面的空间 $\Gamma_0(E)$——它被看作是(有限产生的投影)$C_0(X)$模——来重构在向量束 $E\xrightarrow{\pi}X$ 上的拓扑。例如见[Gracia-Bondiá *et al.*(格拉夏—邦迪亚等)，2001]。我们对 C^* 代数连续场的定义中的第三个条件在下述精确意义上使$\mathcal{A}$成为 $C_0(I)$模，即存在从 $C_0(I)$到$\mathcal{A}$的乘子中心的非退化同构。这一特征也可以代替我们的条件 3。

② 作为 C^* 代数，$\mathcal{A}$的结构对应于点态标量乘法、加法、共轭和在截面上的乘法算符的操作。

$\subset T_x M$ 通过在 x 处的哈密顿向量场的集合来张成。因此，任意哈密顿量向量场在 M 上通过给定点的流作为一个整体处于包含该点的辛流形叶子中。泊松流形的最简单案例是 $M=\mathbb{R}^{2n}$，其泊松括号为式(24)；此流形甚至是辛的。①

有了这些预备知识，我们的基本定义如下：②

相空间 M 的形变量子化包括 C^* 代数 $(\mathcal{A}_\hbar)_{\hbar\in[0,1]}$ 连续场，有 $\mathcal{A}_0=C_0(M)$，伴随有一个自伴③线性映射族 $\mathcal{Q}_\hbar$：$C_c^\infty(M)\to\mathcal{A}_\hbar$，$\hbar\in(0,1]$，满足：

1. 对每个 $f\in C_c^\infty(M)$，由 $0\mapsto f$ 和 $\hbar\mapsto\mathcal{Q}_\hbar(f)$ $(\hbar\neq 0)$ 定义的映射是给定连续场的连续截面；④

2. 对所有的 f，$g\in C_c^\infty(M)$，我们有：

$$\lim_{\hbar\to 0}\left\|\frac{i}{\hbar}[\mathcal{Q}_\hbar(f),\mathcal{Q}_\hbar(g)]-\mathcal{Q}_\hbar(\{f,g\})\right\|_\hbar=0 \tag{25}$$

人们可能赋予其明显的连续性特征，如：

$$\lim_{\hbar\to 0}\|\mathcal{Q}_\hbar(f)\mathcal{Q}_\hbar(g)-\mathcal{Q}_\hbar(fg)\|=0 \tag{26}$$

或：

$$\lim_{\hbar\to 0}\|\mathcal{Q}_\hbar(f)\|=\|f\|_\infty \tag{27}$$

---

① 对泊松流形的力学导向的介绍见[Marsden and Ratiu(马斯登和拉蒂乌)，1994]；基本事实也见[Landsman，1998]或[Butterfield，2005]。关于泊松流形的经典数学论文有[Weinstein(温斯坦)，1983]。

② 这里 $C_c^\infty(M)$ 代表了在具有紧致支柱的 M 上的光滑函数空间，这是 $\mathcal{A}_0=C_0(M)$ 的一个稠密子代数。映射 $\mathcal{Q}_\hbar$ 是否可以从 $C_c^\infty(M)$ 拓展到 $C_0(M)$ 的问题必须要在一个个案例基础上回答。在这样一种拓展下，若它存在，条件(25)将失去其意义，因为泊松括号 $\{f,g\}$ 并非是对所有 f，$g\in C_0(M)$ 定义的。

③ 即 $\mathcal{Q}_\hbar(\bar{f})=\mathcal{Q}_\hbar(f)^*$。

④ 等价地，人们可以将族 $(\mathcal{Q}_\hbar)_{\hbar\in[0,1]}$ 通过 $\mathcal{Q}_0=\mathrm{id}$ 拓展到 $\hbar=0$，且表明 $\hbar\mapsto\mathcal{Q}_\hbar(f)$ 是一个连续截面。同样，人们可以用 φ_0 的单个截面 $\mathcal{Q}$：$C_c^\infty(M)\to\mathcal{A}$ 来取代此映射族，并定义 $\mathcal{Q}_\hbar=\varphi_\hbar\circ\mathcal{Q}$：$C_c^\infty(M)\to\mathcal{A}_\hbar$。

就成了此定义的一个必然结果。[①] 然而，条件(25)转换了 C^* 代数环境，且是证明(与其他因素一起)量子动力学收敛于经典动力学的关键因素，[②] 见第5节。映射$Q_\hbar$是在普朗克常数取值$\hbar$处的量子化映射，我们认为它是目前为止已知的对海森堡最初关于经典可观测量重新解释的最精确构造。它与目前用到的启发性符号$Q_\hbar$拥有相同的解释：算符$Q_\hbar(f)$是量子力学的可观测量，其经典的对应为f。

这被证明是量子化的一个卓有成效的定义，首先是因为绝大多数很好理解的量子化案例都能满足它[Rieffel，1994；Landsman，1998]，其次是因为它提出了许多不同的引人注目的新案例[Rieffel，1989a；Natsume and Nest，1999；Natsume *et al.*，2003；Hawkins，2005]。局限于前者，例如我们注意到式(19)和式(20)定义了相空间$\mathbb{R}^{2n}$上的形变量子化(拥有标准泊松括号)，若人们把$\mathcal{A}_\hbar$当作是在希尔伯特空间$L^2(\mathbb{R}^n)$上紧致算符的 C^* 代数。这被称为是$\mathbb{R}^{2n}$(作为一个相空间)的贝雷辛量子化；[③] 明确地，对于 $\Phi\in L^2(\mathbb{R}^n)$人们有：

$$Q_\hbar^B(f)\Phi(x)=\int_{\mathbb{R}^{2n}}\frac{d^npd^nqd^ny}{(2\pi\hbar)^n}f(p,q)\,\overline{\Psi^\hbar_{(p,q)}(y)}\Phi(y)\Psi^\hbar_{(p,q)}(x) \tag{28}$$

此量子化拥有正定性的显著特征，[④] 其更著名的姊妹，即外尔量子化(Weyl quantization)[⑤]并不拥有该特征。后者也是$\mathbb{R}^{2n}$的形变量子化，拥有相同的C^*代数连续场，但在其量子化映射上不同于贝雷辛量子化：

---

① 它们是必然的这一点不应该忽略下述事实，即尤其是式(27)是经典与量子力学之间的精彩联系。

② 这一见解常常归功于狄拉克[1930]，他第一次认识到量子力学中的对易子与经典力学中泊松括号之间的相似性。事实上，M上的泊松结构是通过连续场结构与条件(25)共同唯一确定的。因此对$Q_\hbar$的选择是次要的。

③ 在文献中，$\mathbb{R}^{2n}$上的贝雷辛量子化常常被称为反威克(Wick)量子化或序(ordering)，而在紧致复流形上它有时被称为托普利兹(Toeplitz)或贝雷辛—托普利兹量子化。基于其他相空间的相干态常常也能定义形变量子化，见[Landsman，1998]。

④ 作为结果，式(28)不仅对$f\in C_c^\infty(\mathbb{R}^{2n})$有效，甚至对所有$f\in L^\infty(\mathbb{R}^{2n})$有效，且从$C_c^\infty(\mathbb{R}^{2n})$到$L^\infty(\mathbb{R}^{2n})$的拓展$Q_\hbar^B$是连续的。

⑤ 原始文献是外尔[1931]。最新的物理学导向的但数学上详细的处理见[Dubin *et al.*(杜宾等)，2000]和[Esposito *et al.*(埃斯波西托等)，2004]。从形变量子化视角来讨论也见[Rieffel，1994]与[Landsman，1998]，无限维案例见[Binz *et al.*(宾兹等)，2004]。

$$Q_\hbar^W(f)\Phi(x) = \int_{\mathbb{R}^{2n}} \frac{d^npd^nq}{(2\pi\hbar)^n} e^{ip(x-q)/\hbar} f\left(p, \frac{1}{2}(x+q)\right)\Phi(q) \tag{29}$$

虽然缺乏好的正定性和连续性特征，[①] 但外尔量子化比贝雷辛量子化拥有更好的对称性特征。[②] 虽然由于这些区别，且这些区别表明了具体的量子化过程缺乏唯一性，但外尔和贝雷辛量子化都重塑了薛定谔的位置和动量算符(2)。[③] 此外，若$f \in L^1(\mathbb{R}^{2n})$，则$Q_\hbar^B(f)$和$Q_\hbar^W(f)$是迹类(trace class)，有：

$$\mathrm{Tr}\, Q_\hbar^B(f) = \mathrm{Tr}\, Q_\hbar^W(f) = \int_{\mathbb{R}^{2n}} \frac{d^npd^nq}{(2\pi\hbar)^n} f(p,q) \tag{30}$$

外尔量子化和贝雷辛量子化通过下式相联系：

$$Q_\hbar^B(f) = Q_\hbar^W(e^{\frac{\hbar}{4}\Delta_{2n}} f) \tag{31}$$

其中 $\Delta_{2n} = \sum_{j=1}^{n}(\partial^2/\partial p_j^2 + \partial^2/\partial(q^j)^2)$，根据它可以表明外尔和贝雷辛量子化在下述意义上是渐近相等的(asymptotically equal)，即对任意的$f \in C_c^\infty(\mathbb{R}^{2n})$，有：

$$\lim_{\hbar \to 0} \| Q_\hbar^B(f) - Q_\hbar^W(f) \| = 0 \tag{32}$$

麦基的量子化方法也在形变量子化背景中找到了其自然归属。令李群 G 作用于流形 Q——解释为位形空间——上，如 4.1 节中那样。结果是相对应的经典相空间是流形$\mathfrak{g}^* \times Q$，具备所谓的泊松结构的半直积[Marsden *et al.*，1984；Krishnaprasad and Marsden(克瑞斯纳普贤德和马斯登)，1987]。相对于 G 的李代数$\mathfrak{g}$的基矢(T_a)，其拥有结构常数 C_{ab}^c，即 $[T_a, T_b] = \sum_c C_{ab}^c T_c$，问题中的泊松括号由下式给出：

$$\{f,g\} = \sum_a \left(\xi_a^M f \frac{\partial g}{\partial \theta_a} - \frac{\partial f}{\partial \theta_a}\xi_a^M g\right) - \sum_{a,b,c} C_{ab}^c \theta_c \frac{\partial f}{\partial \theta_a}\frac{\partial g}{\partial \theta_b} \tag{33}$$

其中 $\xi_a^M = \xi_{T_a}^M$。为展示此长等式的意义，我们考虑一些特殊情形。首先，取$f = X \in \mathfrak{g}$ 和 $g = Y \in \mathfrak{g}$（它被看作是在对偶$\mathfrak{g}^*$上的线性函数）。这产生了：

---

① 然而，利用分布理论中的技术(保留量子力学中典型的希尔伯特空间环境)，外尔量子化可以从 $C_c^\infty(\mathbb{R}^{2n})$ 拓展到更大的函数空间。经典的处理见霍尔曼德尔(Hörmander)[1979；1985a]。

② 但外尔量子化在仿射辛群 $\mathrm{Sp}(n, \mathbb{R}) \ltimes \mathbb{R}^{2n}$ 下是协变的，而贝雷辛量子化仅在其子群 $\mathrm{O}(2n) \ltimes \mathbb{R}^{2n}$ 下是协变的。

③ 这要求将映射 $Q_\hbar^W$和$Q_\hbar^B$形式拓展到 M 上的无界函数，如 p_j 和 q^j。

$$\{X, Y\} = -[X, Y] \tag{34}$$

随后，假定g只依赖于位置q。这产生了：

$$\{X, g\} = -\xi_X^M g \tag{35}$$

最后，假定$f=f_1$和$g=f_2$仅依赖于q，这明确给出了：

$$\{f_1, f_2\} = 0 \tag{36}$$

4.1节中已经在量子层次考虑过两个最简单的物理相关案例。首先，取$G=\mathbb{R}^n$（作为一个李群）和$Q=\mathbb{R}^n$（作为一个流形），G通过平移作用于Q。则式(34)—式(36)产生了泊松括号$\{p_j, p_k\}=0$，$\{p_j, q^k\}=\delta_j^k$，和$\{q^j, q^k\}=0$，表明在此情形中$M=\mathfrak{g}^* \times Q=\mathbb{R}^{2n}$是在$\mathbb{R}^n$中运动粒子的标准相空间，参见式(24)。第二，在$G=E(3)$和$Q=\mathbb{R}^3$情形下产生了一相空间$M=\mathbb{R}^3\times\mathbb{R}^6$，其中$\mathbb{R}^6$是刚刚考虑过的无自旋粒子的相空间，且$\mathbb{R}^3$是包含作为经典自由度自旋的附加内部空间。事实上，在刚刚描述过的$\mathbb{R}^6$上的泊松括号之外，式(34)—式(36)产生了更多的泊松括号$\{J_i, J_j\}=\epsilon_{ijk}J_k$，$\{J_i, p_j\}=\epsilon_{ijk}p_k$，和$\{J_i, q^j\}=\epsilon_{ijk}q^k$。①

一方是式(34)、式(35)和式(36)，另一方是式(6)、式(8)和式(9)的两方之间分别的类似性不是偶然的：所涉泊松括号是刚刚提到的对易关系的经典对应。这一观测通过关联麦基的非本原系与形变量子化的基本定理而变得精确[Landsman，1993；1998]，人们设置C^*代数族如下：

$$\mathcal{A}_0 = C_0(\mathfrak{g}^* \times Q) \tag{37}$$

$$\mathcal{A}_\hbar = C^*(G, Q)$$

其中$C^*(G, Q)$是C^*代数的变换群，通过G在Q上的给定作用来定义（参见4.1节结尾），具有连续场结构，且人们能够定义量子化映射$Q_\hbar$：$C_c^\infty(\mathfrak{g}^*\times Q)\to C^*(G, Q)$从而得到相空间$\mathfrak{g}^*\times Q$的形变量子化。其结果是，对于刚刚考虑过的$X$，$Y\in\mathfrak{g}$和$f=f(q)$类型的特殊函数，不等式：

$$\frac{i}{\hbar}[Q_\hbar(f), Q_\hbar(g)] - Q_\hbar(\{f, g\}) = 0 \tag{38}$$

① 这些是已讨论过的角动量间对易关系的经典对应。

恰好成立，且对于$\hbar \to 0$不仅是渐近成立，正如形变量子化的基本公理(25)所要求的那样。

这一结果通过非本原系澄清了麦基量子化的地位。在关系式(4)背后的经典理论不是在 Q 上运动的无结构粒子的通常相空间 T^*Q，而是 $M = \mathfrak{g}^* \times Q$。出于简单性，我们局限于过渡情形 $Q = G/H$(有正则左 G 作用)。则仅当 $H = \{e\}$ 并因此有 $Q = G$ 时 M 与 T^*Q 重合；① 一般地，相空间$\mathfrak{g}^* \times (G/H)$定域地拥有形式 $T^*(G/H) \times \mathfrak{h}^*$(其中$\mathfrak{h}^*$是 H 李群的对偶)。用$\mathfrak{h}^*$描述的内部自由度是对经典自旋的推广，正如我们看到的，它出现在 $G = E(3)$和 $H = SO(3)$情形下。所有这些都仅是用李广群描述的一大类形变量子化中的一种特殊情形，见[Landsman, 1998; 1999b; 2006a]和[Landsman and Ramazan(兰兹曼和拉马赞), 2001]。②

4.4 几何量子化

由于对抽象 C^* 代数的使用，形变量子化非常复杂，且近年来很技术化。历史上，它之前是一种更具体和传统的方法，称为几何量子化。③ 这里的目标是首先通过具体给定的希尔伯特空间$\mathcal{H}(M)$“量子化”相空间 M，其次将经典的可观测量(即 M 上的实值光滑函数)映射到$\mathcal{H}$上的自伴算符(毕竟它在冯·诺伊

---

① 对于李群 G 有 $T^*G \cong \mathfrak{g}^* \times G$。

② 类似的分析可以应用于艾沙姆[1984]的量子化方案中。正则群 G_C 的幺正不可约表征代表着与 C^*代数 $C^*(G_c)$群非退化表征之间的双射对应[Pedersen, 1979]，后者是泊松流形$\mathfrak{g}_c^*$(即 G_c 的李代数对偶)的形变量子化。此泊松流形将 G_C 的余伴随轨道包含作“不可约”经典相空间，其中仅有一个是人们最初以为他在量子化的余切丛 $T^*(G/H)$(半直积余伴随轨道的分类见[Landsman, 1998])。所有其他轨道仅仅是人们应当避免的废弃品，也见[Robson(罗伯森), 1996]。若人们习惯了广群，则不需要正则群方法。

③ 几何量子化是由科斯坦特(Kostant)[1970]和苏里奥[1969]独立引入的。之后在最初形式体系基础上进行的主要处理有[Guillemin and Sternberg(吉耶曼和斯特伯格), 1977; Sniatycki(斯耐蒂斯基), 1980; Kirillov(基列洛夫), 1990; Woodhouse, 1992; Puta(普塔), 1993; Chernoff(切诺夫), 1995; Kirillov, 2004]和[Ali and Englis, 2004]。当代阶段(建立在应用狄拉克算符和 K 理论使用基础上的)是由博特(Bott)在 1990 年代早期通过未发表的评论开启的，见[Vergne(维赫涅), 1994]和[Guillemin *et al.*, 2002]。后现代(即函数的)时期由[Landsman, 2005]发起。

曼的量子力学体系中扮演可观测量的角色)。① 原则上，这一方案较之形变量子化应该将几何量子化与无界自伴算符在量子力学中扮演的基础作用结合得更好，但实际上几何量子化仍然遇到一些问题。② 然而，把形变量子化与几何量子化看作竞争对手是错误的，正如我们在下面章节中看到的，它们自然地联合在一起，形成了一个假想的量子化函子的“互补”部分。

事实上，在我们看来，几何量子化最好与相空间量子化在4.2节中的具体构造进行对比(即在它 C^* 代数抽象化与随后如4.3节被纳入形变量子化之前)。③ 对于几何量子化也同样，从希尔伯特空间 $L^2(M)$ 开始，④ 然后试图从中构造$\mathcal{H}(M)$，虽然这种方式典型地不同于式(14)。

然而，在这样做之前，几何量子化过程首先试图定义从 $C^\infty(M)$ 到 $L^2(M)$ 上(一般是无界的)算符集的线性映射$\mathcal{Q}_\hbar^{pre}$，它形式上满足：

$$\frac{i}{\hbar}[\mathcal{Q}_\hbar^{pre}(f),\ \mathcal{Q}_\hbar^{pre}(g)]-\mathcal{Q}_\hbar^{pre}(\{f,\ g\})=0 \tag{39}$$

即当$\mathcal{Q}=\mathcal{Q}_\hbar^{pre}$时的式(38)，还有非退化特征：

$$\mathcal{Q}_\hbar^{pre}(\chi_M)=1 \tag{40}$$

其中χ_M是 M 上的函数，同样等于1，且右边的1是 $L^2(M)$ 上的单位算符。这样一个映射被称为是 M 的预量子化(prequatization)。⑤ 对于 $M=\mathbb{R}^{2n}$，用它标准的泊松括号(24)来装备，一个预量子化映射是由下式给出的，即在 $\Phi\in$

① 在几何量子化中，相空间通常被看作是辛流形，除了唯一的例外[Vaisman(魏斯曼)，1991]。为什么开始于泊松流形的更一般类型不常见，其原因将在下一节中明确。

② 除了关于由几何量子化定义的算符的定义域和自伴特征这些相当技术化的问题之外，关键点是人们在几何量子化过程中必须做出的各种不同的数学选择不能够都通过物理学论证来证明，虽然理论的物理学特征依赖于这些选择。极化的概念是主要的相关案例。此外，正如我们要看到的，人们不能够在标准几何量子化中充分量子化许多函数。我们在4.5节中引入的几何量子化的函数方法部分是为了消解这些问题。

③ 也见[Tuynman(图尔曼)，1987]。

④ 正如在式(17)等式子中那样，相对于刘维尔测量乘一适当的因子 $c_\hbar$ 来定义，在几何量子化中，此因子并不是非常重要，因为很少有研究极限$\hbar\to 0$。对于 $M=\mathbb{R}^{2n}$，在 M 上的测量是 $d^npd^nq/(2\pi\hbar)^n$，$L^2(M)$相对于此测量来定义。

⑤ 预量子化的理念要早于几何量子化，见[van Hove(范·霍夫)，1951]与[Segal，1960]。

$L^2(M)$上：

$$Q_\hbar^{pre}(f)\Phi = -i\hbar\{f,\Phi\} + \left(f - \sum_j p_j \frac{\partial f}{\partial p_j}\right)\Phi \tag{41}$$

这一表达最初是对$\Phi \in C_c^\infty(M) \subset L^2(M)$定义的，在这一域中当$f$是实值时$Q_\hbar^{pre}(f)$是对称的①，注意这里的算符是无界的，即使$f$是有界的。② 这看起来很复杂，但是，较简单的表达$Q_\hbar(f)\Phi = -i\hbar\{f, \Phi\}$满足式(38)，并不满足式(40)，且式(41)中第二项的目的是为了满足后者，虽然它也满足前者。③ 例如，我们有：

$$Q_\hbar^{pre}(q^k) = q^k + i\hbar\frac{\partial}{\partial p_k}$$

$$Q_\hbar^{pre}(p_j) = -i\hbar\frac{\partial}{\partial q^j} \tag{42}$$

对一般相空间M，当M是可预量子化时，我们可以构建一个满足式(39)与式(40)的映射$Q_\hbar^{pre}$，对这一记号的完整解释需要一些微分几何。④ 假定这样的话，则预量子化在构建物理学所感兴趣的那种酉群表征过程中是非常有效

---

① 在希尔伯特空间$\mathcal{H}$的致密子空间$\mathcal{D} \subset \mathcal{H}$上定义的算符$A$称为是对称的，当对所有的$\Psi$，$\Phi \in \mathcal{D}$有$(A\Psi, \Phi) = (\Psi, A\Phi)$时。

② 正如已经提到的，自伴性是几何量子化中的一个问题，我们这里不讨论此问题。贝雷辛量子化在此方面要优于几何量子化，因为它将有界函数映射到有界算符上。

③ 人们可能会批评，相对于其同样自然的对应$Q(fg) = Q(f)Q(g)$，几何量子化过程强调式(39)，前者未能被$Q_\hbar^{pre}$所满足，且事实上除了简单的$Q(f) = f$，即作为在$L^2(M)$上的乘法算符，任何已知的量子化过程都不满足。

④ 辛流形(M, ω)在某个固定的$\hbar$值处被称为是可预量子化的，当其允许有联络∇的一个复线丛$L \to M$(称为预量子化线丛)满足$F = -i\omega/\hbar$时，其中F是联络的曲率，由$F(X, Y) = [\nabla_X, \nabla_Y] - \nabla_{[X,Y]}$来定义。这是当且仅当$[\omega]/2\pi\hbar \in H^2(M, \mathbb{Z})$时的情形，其中$[\omega]$是辛形式中的德拉姆(de Rham)上同调类。如果是这样的话，预量子化通过公式$Q_\hbar^{pre} = -i\hbar\nabla_{\xi_f} + f$来定义，其中$\xi_f$是$f$的哈密顿向量场(见4.3节)。此表达是在$L$的紧支柱光滑截面的空间$C_c^\infty(M, L) \subset L^2(M)$上定义且对称的，且很容易证明它满足式(39)和式(40)。为了得到作为特殊情形的(41)，注意对于拥有正则辛形式$\omega = \sum_k dp_k \wedge dp^k$的$M = \mathbb{R}^{2n}$，有$[\omega] = 0$，从而$L$是平凡丛$L = \mathbb{R}^{2n} \times \mathbb{C}$。有$A = -\frac{i}{\hbar}\sum_k p_k dq^k$的联络$\nabla = d + A$满足$F = -i\omega/\hbar$，且这逐渐产生了式(41)。

的。即假定一个李群 G 以"正则"方式作用于相空间 M。这意味着存在一个映射 μ: $M \to \mathfrak{g}^*$，称为动量映射，对于每个 $X \in \mathfrak{g}$ 有 $\xi_{\mu X} = \xi_X^M$,① 且 $\{\mu_X, \mu_Y\} = \mu_{[X,Y]}$。见[Abraham and Marsden(亚伯拉罕和马斯登)，1985；Marsden and Ratiu，1994；1998；Belot，2005；Butterfield，2005]，等等。之后借助于在 $L^2(M)$ 上的反对称无界算符，通过下式得到 G 的李代数 $\mathfrak{g}$ 的表征 π:

$$\pi(X) = -i\hbar Q_\hbar^{pre}(\mu_X) \tag{43}$$

其常常取幂于一个 G 的酉表征。②

正如名称所示，预量子化是尚未量子化。例如，$M = \mathbb{R}^{2n}$ 的预量子化并不再现薛定谔的波动力学：算符(42)并不幺正地等同于(2)。事实上，作为正则对易关系(1)的表征(42)的载体，希尔伯特空间 $L^2(\mathbb{R}^{2n})$ 包含携带了表征(2)的 $L^2(\mathbb{R}^n)$ 和无限重数(multiplicity)[Ali and Emch，1986]。这种情形常常用这样的陈述来表达，即"预量子化是可约的"或预量子化希尔伯特空间 $L^2(M)$"太大"。但两种陈述都有误导性：$L^2(M)$ 事实上在 $Q_\hbar^{pre}(C^\infty(M))$ 的作用下是不可约的[Tuynman，1998]，也可说例如 $L^2(\mathbb{R}^n)$"大于"$L^2(\mathbb{R}^n)$ 在这些希尔伯特空间幺正同构的观点看来是非数学的。这里真正所意味着的是，典型的案例 $L^2(M)$ 在某种李代数的作用下一般是可约的，而人们希望它是不可约的。这也适用于(2)，它定义了海森堡群的李代数表征。更一般地，在相空间 M 携带有李群 G 的可迁作用的情形下，我们能够期望此 G 作用的量子化通过希尔伯特空间上的酉算符是不可约的，$L^2(M)$ 在 $\mathfrak{g}$ 的表征(43)中常常是高度不可约的。③

相空间量子化也遇到这个问题。相空间量子化取代复杂表达式(41)，用式

---

① 这里 $\mu_X \in C^\infty(M)$ 是用 $\mu_X(x) = \langle \mu(x), X \rangle$ 定义的，且 ξ_X^M 是 M 上的向量场，用 G 作用来定义。因此这一条件意味着对所有的 $f \in C^\infty(M)$ 和所有 $y \in M$ 有 $\{\mu_X, f\}(y) = d/dt_{|t=0}[f(\exp(-tX)y)]$。

② 在希尔伯特空间 $\mathcal{H}$ 的致密子空间 $\mathcal{D} \subset \mathcal{H}$ 上定义的算符 A 称为是反对称的(Skew-symmetric)，当对所有的 Ψ, $\Phi \in \mathcal{D}$ 有 $(A\Psi, \Phi) = -(\Psi, A\Phi)$。若人们有 $\mathcal{H}$ 上李群 G 的酉表征 U，则李代数 $\mathfrak{g}$ 表征 dU 由反对称算符组成，使得人们对下述内容抱有希望，即用反对称算符给出的 $\mathfrak{g}$ 给定表征可以积分(或取幂)到 G 的幺正表征。见[Barut and Racka(巴鲁特和拉卡)，1977]或[Jφrgensen and Moore，1984]和其中的文献。

③ 这可以在所谓的轨道方法情境中变得精确。

(11)简单地“相空间量子化”在 $L^2(M)$上的$f \in C^\infty(M)$，f作为一个乘法算符。① 在$C^\infty(M)$作用下，希尔伯特空间 $L^2(M)$当然是高度不可约的。② 把 $L^2(M)$一个适当的子空间表示为：

$$\mathcal{H}(M) = pL^2(M) \tag{44}$$

其中 p 是投影。可将其看作是携带 M 或是 $C^\infty(M)$“量子化”的希尔伯特空间，这可以看作是该约化问题的解决方法，因为若该过程成功的话，对投影 p 的选择使得 $pL^2(M)$在 $pC^\infty(M)p$ 下是不可约的。此外，实践中任何在 M 上的函数都可以用这种方式量子化，虽然要牺牲掉式(38)，正如我们看到的，用它的渐近版本式(25)取代它。关于经典与量子理论下可观测量代数表征的可约性与不可约性的讨论见 6.3 节。

我们把对几何量子化的处理限制在这样的情形下，即采用与上述相同策略的情形，假定最终的希尔伯特空间也具有式(44)的形式。③ 但它与相空间量子化的本质不同在于，它的第一步是式(41)(或其向更一般相空间的推广)而非仅在式子右边拥有$f\Phi$。④ 此外，在几何量子化中我们仅仅量子化了经典可观测量集 $C^\infty(M)$的子空间，它由满足下式的那些函数组成：

$$[Q_\hbar^{pre}(f),\ p] = 0 \tag{45}$$

若函数$f \in C^\infty(M)$满足该条件，则我们定义f的“几何量子化”为：

$$Q_\hbar^G(f) = Q_\hbar^{pre}(f) \upharpoonright \mathcal{H}(M) \tag{46}$$

这是严格的定义，因为由于式(45)，算符$Q_\hbar^{pre}(f)$现在将 $pL^2(M)$映射到其自身上。所以因为式(39)，对于$Q_\hbar = Q_\hbar^G$则式(38)成立；在几何量子化中，我们仅

---

① 对无界的f，此算符是在所有 $\Phi \in L^2(M)$的集合上定义的，对此集合有$f\Phi \in L^2(M)$。

② 即每个(可测量的)子集 $E \subset M$ 定义了一个投影χ_E，且$\chi_E L^2(M)$在所有乘法算符f下是稳定的。人们实际上可以不理会该问题的困扰并就此打住，但那样的话就是在希尔伯特空间背景中简单地处理经典力学[Koopman(库普曼)，1931]。此形式体系甚至被证明在各态历经理论中是非常有用的[Reed and Simon，1972]。

③ 几何量子化传统上是基于极化概念的。此方法产生了一个最终的希尔伯特空间 $\mathcal{H}(M)$，它不是 $L^2(M)$的一个子空间，除了在所谓的(反)全纯情形中之外。

④ 它在想法上也不同于相空间量子化，即投影 p 应当单独从 M 的几何中构建，因此得名“几何量子化”。

拒绝去量子化那些使式(38)不再有效的函数。

虽然取得了一些可观的初步胜利,① 但不存在普遍的方法能够保证成功完成几何量子化的目标。因此，几何量子化仍然像一个黑客工具，其适用性很大程度上取决于用户的创造力。

无论如何，我们熟悉的例子 $M=\mathbb{R}^{2n}$很好理解，且我们在其背景下展示了这一方法的一般要义，通过取 $n=1$ 进一步简化。用 $z=p+iq$ 与$\bar{z}=p-iq$ 取代 M 上的正则坐标$(p,\ q)$是很方便的，且几何量子化的数学工具包使得在 $L^2(\mathbb{R}^2)$ 中审视下述方程的解空间变得很自然②：

$$\left(\frac{\partial}{\partial\bar{z}}+\frac{z}{4\hbar}\right)\Phi(z,\ \bar{z})=0 \tag{47}$$

这些方程处于 $L^2(\mathbb{R}^2)=L^2(\mathbb{C})$中的一般解是：

$$\Phi(z,\ \bar{z})=e^{-|z|^2/4\hbar}f(z) \tag{48}$$

其中 f 是一个全纯(holomorphic)函数，满足：

$$\int_{\mathbb{C}}\frac{dzd\bar{z}}{2\pi\hbar i}e^{-|z|^2/2\hbar}\,|f(z)|^2<\infty \tag{49}$$

则投影 p 是向包含这些解的 $L^2(\mathbb{C})$的闭子空间上的投影。③希尔伯特空间 $pL^2(\mathbb{C})$以一种自然的方式(即不选择基矢)幺正等价于 $L^2(\mathbb{R})$。条件(45)归结为 $\partial^2 f(z,\ \bar{z})/\partial\bar{z}_i\partial\bar{z}_j=0$；特别地，坐标函数 q 与 p 是可量子化的。变换到

---

① 如幂零群的轨道方法和紧致群最新理解的波莱尔—威尔(Borel-Weil)方法，参见[Kirillov，2004]和引用的大多数其他书目。

② 我们用规范等价形式 $A=\frac{i}{2\hbar}(\sum_k q^k dp_k-\sum_k p_k dq^k)=-\frac{i}{\hbar}\sum_k p_k dq^k+\frac{i}{2\hbar}d(\sum_k p_k q^k)$ 代替 1 形式 $A=-\frac{i}{\hbar}\sum_k p_k dq^k$ 来定义联络 $\nabla=d+A$，前者拥有相同的曲率。用此新的 A 表达，即在复坐标中读作 $A=\sum_k(z_k d\bar{z}_k-\bar{z}_k dz_k)/4\hbar$，方程(47)就是 $\nabla_{\partial/\partial\bar{z}}\Phi=0$。这是在几何量子化体系中所谓全纯配极变换(holomorphic polarization)的一个案例。

③ $\mathbb{C}$ 上满足(49)的所有全纯函数的集合是相对于内积 $(f,g)=(2\pi\hbar i)^{-1}\int_{\mathbb{C}}dz\,d\bar{z}\exp(-|z|^2/2\hbar)\,\overline{f(z)}g(z)$ 的希尔伯特空间，称为伯格曼—福克(Bargmann-Fock)空间 $\mathcal{H}_{BF}$。此空间可能通过$f(z)\mapsto\exp(-|z|^2/2\hbar)f(z)$嵌入到 $L^2(\mathbb{C})$中，且此嵌入的图像当然是 $pL^2(\mathbb{C})$。

$L^2(\mathbb{R})$上，我们发现算符$Q_\hbar^G(q)$和$Q_\hbar^G(p)$与薛定谔的表达(2)相吻合。特别地，海森堡群H_1与$L^2(\mathbb{C})$上的无限重数一起不可约地作用于$pL^2(\mathbb{C})$。

4.5 尾声：量子化的函子性

量子化的一个非常重要的方面是，它与对称性和约束的相互作用。事实上，描述自然的基本理论，即电动力学、杨—米尔斯(Yang-Mills)理论、广义相对论，可能也有弦理论，先验地被构造为约束系统。约束与约化的经典方面很好理解，① 通过辛约化过程整理了许多重要的例子。这其中一个特别的例子是马斯登—温斯坦约化：若一个李群 G 以正则方式通过动量映射 μ：$M\to\mathfrak{g}^*$(参照4.4 节)作用于相空间 M，则我们能够构建另一个相空间 $M/\!/G=\mu^{-1}(0)/G$。② 物理上，在 G 是一个规范群且 M 是一个无约束的相空间情形下，$\mu^{-1}(0)$是约束的超曲面(即是 M 的子空间，其上由规范对称性所定义的约束成立)，且 $M/\!/G$ 是仅包含物理自由度的系统的真正相空间。

不幸的是，处理约束量子系统的正确方法仍然是思考与争论的源头。③ 实践中，所有量子化(就像在前一节中讨论的那种)的精确结果都是关于无约束系统的。相应地，我们想要通过把问题化归为无约束情形来量子化一个约束系统。在如下情形适用时可以这么做。我们首先量子化无约束的相空间 M(假定是问题中最容易的部分)，随后采用辛约化的量子版本。最后，通过抽象方式证明这样约束的量子理论等价于下述理论，即一开始就约化到经典层面，之后再到量子化约束的经典相空间(通常在实践中是不可能完成的任务)的理论。

悲剧的是，表述了“量子化与约化相对易”的足够有力的定理，在这个意义

---

① 见[Gotay *et al.*(戈塔伊等)，1978；Binz *et al.*，1988；Marsden，1992；Marsden and Ratiu，1994；Landsman，1998；Butterfield，2005]和[Belot，2005]。

② 技术上，M 必须是一个辛流形，且若 G 恰当且自由地作用于$\mu^{-1}(0)$，则$M/\!/G$仍然是一个辛流形。

③ 关于约束系统量子化的不同视角，参见[Dirac，1964；Sundermeyer(桑德梅尔)，1982；Gotay，1986；Duval *et al.*(杜瓦尔等)，1991；Govaerts(古维特斯)，1991；Henneaux and Teitelboim(亨诺和泰特尔鲍姆)，1992]和[Landsman，1998]。

上仍然是难以捉摸的。① 例如，目前为止它阻碍了对四维杨—米尔斯理论的严格量子化，这是克雷数学研究所(Clay Mathematical Institute)的千禧年(Millenium)问题中的一个，悬赏1百万美金。②

以一种更精神化的论调，数学家尼尔森(E. Nelson)公开说“一次量子化是一个谜，但二次量子化是一个函子”。二次量子化的函子性，即涉及福克空间的一个构建，见[Reed and Simon，1975]，几乎是微不足道的事情，深层的数学与概念问题在“一次”量子化的可能函子性中，它仅意味着在我们目前所讨论的意义上的量子化。这最初被认为是意味着，相空间M的正则变换α应该通过$\mathcal{H}(M)$上的酉算符$U(\alpha)$来量子化，在这种方式下$U(\alpha)\mathcal{Q}_\hbar(f)U(\alpha)^{-1}=\mathcal{Q}(L_\alpha f)$，参见(10)。这仅在一些特殊情况下可能，例如当$M=\mathbb{R}^{2n}$和$\alpha$是线性辛映射时，且更一般地当$M=G/H$是齐次的，且有$\alpha\in G$时(见4.2节结尾)。③结果是，量子化的函子性广泛被认为是一条死胡同。④

然而，此结论中建立的所有不可行定理都开始于错误和单纯的范畴，无论是在经典还是在量子层面。⑤ 很有可能的是，我们事实上能通过对范畴的更精细选择来使量子化成为函子，额外令人高兴的是形变量子化与几何量子化统一起来了：前者是量子化函子中的对象部分(object part)，而后者(经过适当的重新解释)是箭头部分(arrow part)。令人惊奇的是，在这种构建下，陈述“量子化与约化相对易”成为量子化函子性的特殊情形[Landsman，2002；2005]。

---

① 所谓吉耶曼—斯特伯格猜测[Guillemin-Sternberg，1982]——现在是一个定理[Meinrenken(迈因克)，1998；Meinrenken and Sjamaar(迈因克和斯玛尔)，1999]——只处理了其中G和M紧致的马斯顿—韦恩斯坦约化情形。这里“量子化与约化相对易”定理的数学深刻性可见一斑，它远不能称为规范理论，因为在规范理论中群和空间不仅是非紧致的，甚至是无限维的。

② 见http：//www.claymath.org/millennium/。

③ 正则变换可以在恰当意义上量子化，即通过所谓的傅里叶积分操作在$\hbar\to0$时变得精确，见[Hörmander，1971；1985b]和[Duistermaat(杜斯特马特)，1996]。

④ 见[Groenewold，1946；van Hove，1951；Gotay *et al.*，1996]和[Gotay，1999]。

⑤ 典型地，我们认为经典范畴包括作为对象的辛流形、作为箭头的辛同构，量子范畴拥有作为对象的C^*代数和作为箭头的自同构。

为了说明其主要观点，我们回到上节解释过的几何量子化 $M=\mathbb{R}^2\cong\mathbb{C}$ 上。把 $pL^2(\mathbb{C})$①看作是问题中正确的希尔伯特空间，可以用完全不同的方式来理解，这为几何量子化方案的强有力重建铺平了道路，而几何量子化将最终定义量子化函子。即$\mathbb{C}$ 支撑了一个特定的线性一阶微分算符 $\not{D}$，它完全是通过其作为相空间的几何来定义的，称为狄拉克算符。② 这一算符由下式给出③：

$$\not{D}=2\begin{pmatrix} 0 & -\dfrac{\partial}{\partial z}+\dfrac{\bar{z}}{4\hbar} \\ \dfrac{\partial}{\partial \bar{z}}+\dfrac{z}{4\hbar} & 0 \end{pmatrix} \tag{50}$$

它作用于 $L^2(\mathbb{C})\otimes\mathbb{C}^2$ 上(作为一适当定义的无界算符)。这一算符具有一般形式：

$$\not{D}=\begin{pmatrix} 0 & \not{D}_- \\ \not{D}_+ & 0 \end{pmatrix}$$

这样一个算符的指标由下式给出：

$$\operatorname{index}(\not{D})=[\ker(\not{D}_+)]-[\ker(\not{D}_-)] \tag{51}$$

其中$[\ker(\not{D}_\pm)]$表示 $\ker(\not{D}_\pm)$的(幺正)同构类，被看作是算符适当代数的表征

① 或者是伯格曼—福克空间$\mathcal{H}_{BF}$。

② 特别地，这是所谓的自旋^{c}狄拉克算符，由$\mathbb{C}$ 的复结构来定义，与预量子化线丛相耦合，见[Guillemin *et al.*, 2002]。

③ 相对于狄拉克矩阵 $\gamma^1=\begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}$和 $\gamma^2=\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$。

空间。① 在当前情形中，我们有 $\ker(\not{D}_+)=pL^2(\mathbb{C})$ 与 $\ker(\not{D}_-)=0$，②这里我们把 $\ker(\not{D}_+)$ 看作是海森堡群 H_1 的表象空间。因此，$\mathrm{index}(\not{D})$ 给出相空间ℂ几何量子化的模幺正等价性，它被看作是 H_1 表象空间的“形式差异”。

这一过程可以推广到任意的相空间 M，其中 $\not{D}$是由 M 的相空间几何与量子化要求自然地定义的特定算符。③ 这被证明是几何量子化最有希望的构造——在某种代价下。④ 因为原先用希尔伯特空间来量子化相空间的目标现在已经被更抽象的过程所取代，在后者中量子化的结果是可观测量量子代数表象空间的

---

① (51)的左手边真正应该写为 $\mathrm{index}(\not{D}_+)$，因为 $\mathrm{coker}(\not{D}_+)=\ker(\not{D}_+^*)$ 和 $\not{D}_+^*=\not{D}_-$，但因为指标自然地相关于作为整体的 $\not{D}$，我们误将 $\mathrm{index}(\not{D})$ 写成为 $\mathrm{index}(\not{D}_+)$。在有限维向量空间之间的线性映射 $L: V\to W$ 的常规指标定义为 $\mathrm{index}(L)=\dim(\ker(L))-\dim(\mathrm{coker}(L))$，其中 $\mathrm{coker}(L)=W/\mathrm{ran}(L)$。基本线性代数产生了 $\mathrm{index}(L)=\dim(V)-\dim(W)$。这是令人称奇的，因为它独立于 L，而 $\dim(\ker(L))$ 和 $\dim(\mathrm{coker}(L))$ 对 L 非常地敏感。例如，取 $V=W$ 和 $L=\varepsilon\cdot 1$。若 $\varepsilon\neq 0$ 则有 $\dim(\ker(\varepsilon\cdot 1))=\dim(\mathrm{coker}(\varepsilon\cdot 1))=0$，而对于 $\varepsilon=0$，我们有 $\dim(\ker(0))=\dim(\mathrm{coker}(0))=\dim(V)$！类似地，通过(47)等对几何量子化的通常定义在深层辛结构的干扰下是不稳定的，而通过(51)改进过的定义则是稳定的。为了从上述关于指标的概念过渡到后者，我们首先写出 $\mathrm{index}(L)=[\ker(L)]-[\mathrm{coker}(L)]$，其中$[X]$是作为ℂ模的线性空间 X 的同构类。这一表达是 $K_0(\mathbb{C})$ 的一个元素，且我们通过下述认识来恢复早先的指标，即类$[X]$由 $\dim(X)$ 和相应的同构 $K_0(\mathbb{C})\cong\mathbb{Z}$ 完全决定。当一个更复杂的有限维 C^* 代数作用于 V 和 W 上时，伴随有 $\ker(L)$ 和 $\mathrm{coker}(L)$ 在$\mathcal{A}$作用下是稳定的特征，我们可以将$[\ker(L)]-[\mathrm{coker}(L)]$和 $\mathrm{index}(L)$ 定义为所谓 C^* 代数的 K 理论群 $K_0(\mathcal{A})$ 一个元素。在特定技术条件下，指标的这一概念可以推广到无限维希尔伯特空间和 C^* 代数中，见[Baum *et al.*（鲍姆等），1994]和[Blackadar（布莱卡达），1998]。K 理论指标在$\mathcal{A}=C^*(G)$是某个局域紧致群 G 的 C^* 代数群时是最好理解的。在$M=\mathbb{R}^2$ 例子中，人们可以将 G 看作是海森堡群 H_1，从而有 $\mathrm{index}(\not{D})\in K_0(C^*(H_1))$。对此 K_0 群的描述见[Elliott *et al.*（埃利奥特等），1993]。

② 因为$\left(-\frac{\partial}{\partial z}+\frac{\bar{z}}{4\hbar}\right)\Phi=0$ 意味着 $\Phi(z,\bar{z})=\exp(|z^2|/4\hbar)f(\bar{z})$，当且仅当$f=0$ 时它处于 $L^2(\mathbb{C})$中。

③ 任意辛流形携带了一个与辛形式相容的复结构，产生了一个自旋^{c}狄拉克算符，仍见[Guillemin *et al*, 2002]。若 $M=G/H$，或更一般地，如果 M 携带了具有紧致商 M/G 的李群 G 的一个正则作用，则 $\mathrm{index}(\not{D})$ 定义了 $K_0(C^*(G))$ 的一个元素。在完全普遍性下，$\mathrm{index}(\not{D})$ 应当是$K_0(\mathcal{A})$ 的一个元素，其中$\mathcal{A}$是量子系统可观测量的 C^* 代数。

④ 在有利的方面，在 $\not{D}$的连续形变下指标的不变性似乎消除了传统量子化过程中关于不同“算符序”的模糊性，经典理论并不规定该算符序。

特定同构类之间的形式差异。为了展现这里涉及的抽象化程度，假定我们忽略可观测量(在刚考虑的案例中如位置与动量)的作用。那样的话希尔伯特空间$\mathcal{H}$的同构类[$\mathcal{H}$]完全是由其维度 dim($\mathcal{H}$)来刻画的，那样的话有 ker($\not{D}_-$) ≠0，因而量子化在 index($\not{D}$)形式下甚至可以是负数！我们是不是疯了？

不完全是这样。几何量子化的上述图景确实与物理学非常不相关，除非用形变量子化来作补充。在此背景下用某个固定的$\hbar$值便于处理，从而形变量子化仅仅是将某个 C^* 代数$\mathcal{A}(P)$与给定相空间 P 联系起来。① 在寻找量子化的一种范畴解释时，假定经典范畴$\mathfrak{C}$的对象是相空间 P 因而是很自然的，② 而量子范畴$\mathfrak{Q}$的对象是 C^* 代数。③ 假想的量子化函子的对象部分要被形变量子化，符号上写作 $P \mapsto Q(P)$。

若几何量子被重新解释为是假想的量子化函子的箭头部分的话，则一切都那么协调。为实现这一点，经典范畴$\mathfrak{C}$中的箭头不应该被看作是在相空间之间的映射，而是辛双模(symplectic bimodules)$P_1 \leftarrow M \rightarrow P_2$。④ 更精确地说，$\mathfrak{C}$中的箭头是这些双模合适的同构类。⑤ 类似地，量子范畴$\mathfrak{Q}$中的箭头并非如单纯以为的那样是 C^* 代数的态射(morphisms)，而是 C^* 代数双模的特定同构类，具备广义的狄拉克算符的附加结构。⑥

在已经将量子化映射Q：$\mathfrak{C} \rightarrow \mathfrak{Q}$的对象部分定义为形变量子化之后，我们现在在对(51)适当推广的意义上假定箭头部分是几何量子化，细节见[Lands-

① 这里 P 没有必要是辛的；它可能是一个泊松流形，且为了使泊松流形和辛流形分开，今后我们将前者记作 P，把记号 M 留给后者。

② 严格来讲，为成为此范畴中的一个对象，泊松流形 P 必须是可积的，见[Landsman，2001]。

③ 出于引入 K 理论的技术原因，这些必须是可分离的。

④ 这里 M 是一个辛流形，P_1 和 P_2 是可积的泊松流形；映射 $M \rightarrow P_2$ 是反泊松的，而映射 $P_1 \leftarrow M$ 是泊松的。这样的双模(常常称为对偶对)由卡拉肖夫(Karasev)[1989]和温斯坦[1983]引入。为了使之作为$\mathfrak{C}$中的箭头出现，辛双模必须满足一系列正则性条件[Landsman，2001]。

⑤ 为了使箭头合成是结合的，这是必要的；这是由辛约化过程的推广来给出的。

⑥ 范畴$\mathfrak{Q}$只不过是卡斯帕洛夫(Kasparov)引入的范畴 KK，该范畴的对象是可分离的 C^* 代数，且它的箭头是所谓的卡斯帕洛夫群 $KK(A, B)$，由卡斯帕洛夫积 $KK(A, B) \times KK(B, C) \rightarrow KK(A, C)$组成。见[Higson(希格森)，1990]和[Blackadar，1998]。

man，2005]。则我们猜想Q是一个函子，在此能够并且已经被检验的情形中，Q的函子性正好是如下陈述，即量子化与约化相对易。①

因而海森堡的重新解释理念在量子化函子中找到了其最终的实现。

5. 极限$\hbar\to0$

早些时候就有人认识到，普朗克常数趋于零的极限$\hbar\to0$应当在对从量子理论到经典世界的说明中发挥作用。严格来说，$\hbar$是一个有量纲的(dimensionful)常数，但实践中人们是通过构造一个$\hbar$与其他参数无量纲的组合来研究给定量子理论的半经典体系，如果它确实是能够被改变的$\hbar$的无量纲版本，则这一组合重新进入理论。最早的案例是普朗克的辐射公式(1)，其中温度T是相关变量。爱因斯坦[1905]与普朗克[1906]的发现，即在极限$\hbar\nu/kT\to0$下这一公式趋向于经典的均分律$E_\nu/N_\nu=kT$，确实可能是对量子理论$\hbar\to0$极限的第一次应用。②

另一个案例是哈密顿量为$H=-\frac{\hbar^2}{2m}\Delta_x+V(x)$的薛定谔方程(3)，其中$m$是相关粒子的质量。这里人们可以通过引入$H$典型的能量度量$\epsilon$如$\epsilon=\sup_x|V(x)|$，也可以引入典型的长度尺度$\lambda$如$\lambda=\epsilon/\sup_x|\nabla V(x)|$（若这些量是有限的），而转到无量纲的参数。采用无量纲的变量$\tilde{x}=x/\lambda$，重新标度的哈密顿量$\tilde{H}=H/\epsilon$则是无量纲的，且等于$\tilde{H}=-\tilde{\hbar}^2\Delta_{\tilde{x}}+\tilde{V}(\tilde{x})$，其中$\tilde{\hbar}=\hbar/\lambda\sqrt{2m\epsilon}$，且$\tilde{V}(\tilde{x})=V(\lambda\tilde{x})/\epsilon$。这里$\tilde{\hbar}$是无量纲的，且人们可以研究它在其中很小的体系[Gustafson and Sigal，2003]。我们的最后一个案例发生在大的量子系统的理论中，将在下一节中处理。接下来的所有讨论中变量$\hbar$都将指普朗克常数这样的无量纲版本。

---

① 在辛流形M上的正则G作用和动量映射μ：$M\to\mathfrak{g}^*$产生了一个对偶对$pt\leftarrow M\to\mathfrak{g}^*$，它在$\mathfrak{C}$中解释为是从拥有一点的空间$pt$到$\mathfrak{g}^*$的箭头。此箭头和从$\mathfrak{g}^*$到$pt$的箭头$\mathfrak{g}^*\hookrightarrow0\to pt$组成了$pt\leftarrow M//G\to pt$。若$G$是连通的，则在这两个对上的量子化函子等价于吉耶曼—斯特伯格猜测，见[Landsman，2005]。

② 这里爱因斯坦[1905]通过在固定的T和$\hbar$处令$\nu\to0$，取$\hbar\nu/kT\to0$，而普朗克[1906]是在固定的ν和$\hbar$处取$T\to\infty$。

正如我们将要论证的，虽然极限$\hbar \to 0$本身不能说明经典的世界，但它确实产生了一些真正令人欣喜的数学结果。进而，这些数学结果引起了几乎不可避免的结论，即正在考虑的极限确实是相关于从量子理论中恢复经典物理学的那个。因而本节打算罗列那些令人欣喜的结果，它们可能是量子理论基础研究者们所直接感兴趣的。

还有另一个对$\hbar \to 0$极限更技术化的应用，即通过用相关的经典客体来近似化态与可观测量的时间演化，从而在量子力学中进行计算。这一尝试就是熟知的半经典分析。数学上，对$\hbar \to 0$极限的这一应用与从量子力学中恢复经典力学的目标紧密关联，但在概念上非常不同。接下来我们将试图指出相关的区别。

5.1 再议相干态

正如薛定谔[1926b]预见的，相干态在极限$\hbar \to 0$中扮演了重要的角色。我们回顾4.2节中，对于普朗克常数的某个固定值$\hbar$，对于相空间M而言希尔伯特空间$\mathcal{H}$中的相干态是由入射$M \hookrightarrow \mathcal{H}$与$z \mapsto \Psi_z^\hbar$确定的，以使得式(16)与式(17)成立。接下来，我们说$\Psi_z^\hbar$聚集在$z \in M$处，这一说法是由关键例子(20)所证明的。

为了与经典极限相关，考虑到其对$\hbar$的依赖性相干态必须满足附加的特征，这一特征同样大大地澄清了它们的性质[Landsman，1998]。我们要求对于每一个$f \in C_c(M)$与每一个$z \in M$，从$\hbar$在其中取值的集合I，即通常$I = [0, 1]$，但无论如何总是包含零作为极限点，到$\mathbb{C}$的下述函数是连续的：

$$\hbar \mapsto c_\hbar \int_M d\mu_L(w) \mid (\Psi_w^\hbar, \Psi_z^\hbar) \mid^2 f(w) \quad (\hbar > 0) \tag{1}$$

$$0 \mapsto f(z) \tag{2}$$

在(19)看来，式(2)的右边与$(\Psi_z^\hbar,\ Q_\hbar^B(f)\Psi_z^\hbar)$相同。特别地，这一连续性条件意味着：

$$\lim_{\hbar \to 0}(\Psi_z^\hbar,\ Q_\hbar^B(f)\Psi_z^\hbar) = f(z) \tag{3}$$

这意味着聚集在$z \in M$处的相干态中，经典可观测量f的相空间量子化(19)的量子力学期望值的经典极限，正好是f在态z下的经典期望值。这一解释依赖于将经典态看作是在相空间M上的概率测度，在其中M的点在狄拉克测度(即

δ 函数)的外观下是纯态。此外，可以表明[Landsman，1998]，所有函数(1)—(2)的连续性意味着下述特征：

$$\lim_{\hbar\to 0}|(\Psi_w^\hbar, \Psi_z^\hbar)|^2=\delta_{wz} \tag{4}$$

其中 δ_{wz}是普通的克罗内克(Kronecker)δ(即对于所有 $z\in M$，每当有 $w\neq z$ 时 $\delta_{wz}=0$，且 $\delta_{zz}=1$)。这也有常见的物理解释：聚集在 w，$z\in M$ 处的两个相干态之间的量子力学跃迁概率的经典极限，与在 w 和 z 之间的经典(且平凡的)跃迁概率相等。换言之，当$\hbar$变得很小时，在 w 和 z 不同值处的相干态变得越来越相互正交。① 这产生了一个有趣的结果，即对于所有的$f\in C_c(M)$有：

$$\lim_{\hbar\to 0}(\Psi_w^\hbar, Q_\hbar^B(f)\Psi_0^\hbar)=0(w\neq z) \tag{5}$$

特别地，下述薛定谔猫类型的现象发生在经典极限下：若 $w\neq z$ 且在$\hbar\in[0, 1]$上有连续函数$\hbar\to c_w^\hbar\in\mathbb{C}$ 与$\hbar\to c_z^\hbar\in\mathbb{C}$，使得：

$$\Psi_{w,z}^\hbar=c_w^\hbar\Psi_w^\hbar+c_z^\hbar\Psi_z^\hbar \tag{6}$$

对于$\hbar\geqslant 0$ 与 $|c_w^0|^2+|c_z^0|^2=1$ 是单位向量，则：

$$\lim_{\hbar\to 0}(\Psi_{w,z}^\hbar, Q_\hbar^B(f)\Psi_{w,z}^\hbar)=|c_w^0|^2f(w)+|c_z^0|^2f(z) \tag{7}$$

因此由向量 $\Psi_{w,z}^\hbar$确定的(典型地)纯态 $\Psi_{w,z}^\hbar$(在 C^* 代数$\mathcal{A}_\hbar$上，映射$Q_\hbar^B$在其中取值)族②在某种意义上趋向于由式(7)右边确定的在 $C_0(M)$上的混合态。本节最后将精确化这一表述。

毋庸置疑，薛定谔的相干态(20)满足我们的公理，人们也可以从(21)直接证明(4)。因此，根据(32)外尔量子化也有相同的特征(3)，只要$f\in\mathcal{S}(\mathbb{R}^{2n})$③，即：

$$\lim_{\hbar\to 0}(\Psi_z^\hbar, Q_\hbar^W(f), \Psi_z^\hbar)=f(z) \tag{8}$$

类似地，对于$Q_\hbar^W$式(5)也成立。

---

① 对于跃迁概率概念的一般意义，见[Mielnik(梅尔尼克)，1968；Cantoni(坎托尼)，1975；Beltrametti and Cassinelli(贝尔特拉梅蒂和卡西内利)，1984；Landsman，1998]与下文6.3 节。

② 例如，对于 $M=\mathbb{R}^{2n}$，每个$\mathcal{A}_\hbar$等于 $L^2(\mathbb{R}^n)$上紧致算符的 C^* 代数，$L^2(\mathbb{R}^n)$上的每个向量态肯定是纯态。

③ 这里$\mathcal{S}(\mathbb{R}^{2n})$是在无穷处迅速衰退的平滑测量函数的普通施瓦茨(Schwartz)空间。

此外，文献中许多被称作是相干态的构造满足(16)、(17)和式(4)，可参见[Landsman，1998]，① 则得到的一般结果为，聚集于 $z\in M$ 处的一个相干态是对作为一个量子力学纯态的 z 的重新解释(正如上面解释过的，它被看作是一个经典的纯态)。②

虽然相干态有广泛的应用(且有些人说它美)，人们不得不超越相干态以获得量子力学极限$\hbar\to 0$ 更完整的图像。适当地推广态的连续场概念，③ 这是相对于相空间 M 的给定形变量子化来定义的，参见4.3节。若对于每一个$\hbar\in[0,1]$，或更一般地，对于包含0为一个极限点的离散子集[0，1]，在$\mathcal{A}_\hbar$上有一个态 $\omega_\hbar$，则每当对于每个$f\in C_c^\infty(M)$函数$\hbar\mapsto\omega_\hbar(Q_\hbar(f))$在[0，1]上是连续的时，可以称随之而来的态的集合为一个连续场。这一概念事实上是由 C^* 代数连续场内在地定义的，因而独立于量子化映射$Q_\hbar$。特别地，有：

$$\lim_{\hbar\to 0}\omega_\hbar(Q_\hbar(f))=\omega_0(f) \tag{9}$$

方程(3)或(8)表明，相干态确实是态连续场的例子，拥有每个 $\omega_\hbar$ 都为纯态的附加特征。至于其中所有态 $\omega_\hbar$ 是混合态的例子，我们提到量子力学的配分函数收敛于它们在统计力学中的经典对应属于此行列，见[Lieb(利布)，1973；Simon，1980；Duffield(达菲尔德)，1990]与[Nourrigat and Royer(努里加和罗耶)，2004]。最后，我们遇到了这样的现象，即纯的量子力学态可能会收敛于

---

① 例如，佩列洛莫夫[1986]引入的相干态类型适用于下述设定[Simon，1980]。令 G 是一个紧致连通李群，$\mathcal{O}_\lambda$是积分余伴随轨道，对应于最大的权重 λ。这里可以认为 $G=SU(2)$和 $\lambda=0$，1/2，1，…。注意$\mathcal{O}_\lambda\cong G/T$，其中 T 是 G 中最大的环面，其中权重是相对于它而确定的。令$\mathcal{H}_\lambda^{\text{hw}}$是具有最大权重 λ 的不可约表征 $U_\lambda(G)$的承载空间，包含了最大的权重向量 Ω_λ。对于 $G=SU(2)$有著名的自旋j的希尔伯特空间$\mathcal{H}_j^{\text{hw}}=\mathbb{C}^{2j+1}$，在其中 Ω_j 是在 z 方向上自旋为j的向量。对于$\hbar=1/k$，$k\in\mathbb{N}$ 定义$\mathcal{H}_\hbar:=\mathcal{H}_{\lambda/\hbar}^{\text{hw}}$。选择投影 $G\to G/T$ 的一个截面 σ：$\mathcal{O}_\lambda\to G$，则我们得到相干态 $x\mapsto U_{\lambda/\hbar}(\sigma(x))\Omega_{\lambda/\hbar}$，该相干态相关于在$\mathcal{O}_\lambda$上的刘维尔测量和 $c_\hbar=\dim(\mathcal{H}_{\lambda/\hbar}^{\text{hw}})$。这些态显然不是对所有在(0，1]中的$\hbar$定义的，而是仅对分立的 1/$\mathbb{N}$ 集合定义的。

② 这一理念也被如下事实所确证，即至少薛定谔相干态是最小不确定度的态。

③ 在量子化不同数学方法中对此概念的应用基本上是不专业的。对于 C^* 代数背景见[Emch，1984；Rieffel，1989b；Werner，1995；Blanchard，1996；Landsman，1998]与[Nagy，2000]。

混合的经典态。第一个这样的例子刚才在(7)中已经展示过了，其他相关例子是能量本征态与 WKB 态(见下文5.4、5.5与5.6节)。

5.2 量子动力学向经典运动的收敛(convergence)

非相对论量子力学建立在薛定谔方程(3)上，后者更一般地写作：

$$H\Psi(t)=i\hbar\frac{\partial\Psi}{\partial t} \tag{10}$$

初始值为 $\Psi(0)=\Psi$ 的形式解为：

$$\Psi(t)=e^{-\frac{it}{\hbar}H}\Psi \tag{11}$$

这里我们假定 H 是系统希尔伯特空间$\mathcal{H}$上一给定的自伴算符，从而根据斯通定理，此解确实存在且幺正地演化，参见[Reed and Simon，1972]与[Simon，1976]。等价地，我们可以通过下式将态的时间演化(薛定谔图像)转变为算符的时间演化(海森堡绘景)：

$$A(t)=e^{-\frac{it}{\hbar}H}Ae^{-\frac{it}{\hbar}H} \tag{12}$$

我们这里局限于$\mathbb{R}^n$中的粒子运动，以使$\mathcal{H}=L^2(\mathbb{R}^n)$。[①] 那样的话，$H$ 往往由类似于(3)的形式表达给出(在某具体的定义域内)。[②] 现在首先印入脑海的是埃伦费斯特定理(Ehrenfest's theorem)[1927]，它表明在定义域$Q_\hbar(q^j)=x^j$与$\partial V(x)/\partial x^j$中，对任意的(单位)向量 $\Psi\in L^2(\mathbb{R}^n)$，有：

$$m\frac{d^2}{dt^2}\langle x^j\rangle(t)=-\left\langle\frac{\partial V(x)}{\partial x^j}\right\rangle(t) \tag{13}$$

记号为：

$$\langle x^j\rangle(t)=(\Psi(t),x^j\Psi(t)) \tag{14}$$

$$\left\langle\frac{\partial V(x)}{\partial x^j}\right\rangle(t)=\left(\Psi(t),\frac{\partial V(x)}{\partial x^j}\Psi(t)\right)$$

这看起来像在态 ψ 中 x 期望值的牛顿第二定律，很微小但关键性的不同在于，牛顿想要看到式(13)右边的$(\partial V/\partial x^j)(\langle x\rangle(t))$。此外，即使不包括这一点，埃

---

① 对 N 体薛定谔算符的新近考察见[Hunziker and Sigal，2000]。

② 则人们不得不证明在一更大范围内的自伴性(或对它的缺乏)，在此范围内算符是闭的。

伦费斯特定理也绝不足以拥有经典的行为，因为它并不保证如$\langle x\rangle(t)$的行为像一个点粒子那样。接下来的内容可以看作是对埃伦费斯特定理的强化，以使得它确实为恰当算符期望值给出合适的经典运动方程。

我们假定量子哈密顿量有更普遍的形式：

$$H=h(Q_\hbar(p_j),\ Q_\hbar(q^j)) \tag{15}$$

其中h为经典的哈密顿量(即在经典相空间$\mathbb{R}^{2n}$上定义的函数)，$Q_\hbar(p_j)$与$Q_\hbar(q^j)$是由(2)给出的算符。每当这一表达是不清晰的时，如$h(p,q)=pq$情形，人们不得不假定一具体的量子化方法，如外尔量子化$Q_\hbar^W$，参见(29)，使得在形式上拥有：

$$H=Q_\hbar^W(h) \tag{16}$$

事实上，在被引用的文献中，甚至有更大类的量子哈密顿量是用这里解释的方法来处理的。量子哈密顿量H具有明确的(且非常突出的)$\hbar$依赖性，且对于$\hbar\to 0$人们则期望式(11)或式(12)以一种或另一种方式关联于经典哈密顿量h流。这一关系已经被薛定谔(1926a)所预见，几乎在量子力学诞生之后很快便由著名的WKB近似形式化，参见[Landau and Lifshitz, 1977]与下文5.5节。对这一方法和类似的近似方法在数学上的严格理解则很晚，即在从其最初的偏微分方程背景[Hörmander, 1965; Kohn and Nirenberg(科亨和尼伦伯格), 1965; Duistermaat, 1974; 1996; Guillemin and Sternberg, 1977; However(豪艾沃), 1980; Hörmander, 1979; 1985a; 1985b; Grigis and Sjöstrand(格里吉斯和舍斯特兰德), 1994]到量子力学$\hbar\to 0$极限的研究中，一种称为微局域分析的技术中被适应时才做到。现在由许多主流人物完成的许多综述解释了这一适应(常常称为半经典分析)及其结果，著名的有[Robert, 1987; 1988; Helffer(海弗), 1988; Paul and Uribe, 1995; Colin de Verdière(科林·德韦德耶尔), 1998; Ivrii(伊维依), 1998; Dimassi and Sjöstrand(迪马西和舍斯特兰德), 1999]与[Martinez(马丁内斯), 2002]，同样见[Robert, 1992]中的论文。更具体的文献将在下文给出。①

正如前面提到的，半经典分析给出的H与h之间的关系是把双刃剑。一方

① 关于半经典渐近的启发性理论，[landau and Lifshitz, 1977]非常有价值。

面，我们得到方程(11)或(12)的近似解，或用经典数据表示的对小值$\hbar$而言的近似能量本征值与能量本征函数，有时称为准模(quasi-modes)。这是结果通常出现的方式，人们用底层的经典理论来计算在特定体系内量子理论的具体特性。但是，另一方面，在某些工作中相同的结果常常被重新解释为，对经典动力学从量子力学中突现的部分解释。我们感兴趣的正是半经典分析的后一方面，这一方面文献中不怎么提及。在本小节与接下来的三小节中，我们将局限于最简单类型的结果，然而它们仍然为通过这些方法可以达到并理解什么提供了好的评判。同样地，我们也像之前一样只处理通常的平坦相空间 $M=\mathbb{R}^{2n}$。

将经典与量子动力学关联起来的最为简单的结果如下：[1]

> *若经典哈密顿量 $h(p, q)$ 对 p 与 q 至多是二次的，且(12)中的哈密顿量由(16)给出，则：*

$$Q_{\hbar}^{W}(f)(t)=Q_{\hbar}^{W}(f_t) \tag{17}$$

这里f_t 是经典运动方程 $df_t/dt=\{h, f_t\}$ 的解；等价地，也可写作：

$$f_t(p, q)=f(p(t), q(t)) \tag{18}$$

其中 $t\mapsto(p(t), q(t))$是具有初始条件为$(p(0), q(0))=(p, q)$的经典哈密顿流 h。这对于所有合适的f都成立，如$f\in\mathcal{S}(\mathbb{R}^{2n})$。

这一结果用经典的术语解释了量子动力学，但通过将(17)与(9)结合可以达到其相反过程。对于态$(\omega_{\hbar})$的任意连续场而言，这产生了：

$$\lim_{\hbar\to 0}\omega_{\hbar}(Q_{\hbar}(f)(t))=\omega_0(f_t) \tag{19}$$

特别地，对于薛定谔的相干态(20)，我们得到：

$$\lim_{\hbar\to 0}(\Psi_{(p,q)}^{\hbar}, Q_{\hbar}(f)(t)\Psi_{(p,q)}^{\hbar})=f_t(p, q) \tag{20}$$

现在，尽管式(17)仅仅反映了外尔量子化好的对称特征，[2] 且对$Q_{\hbar}^{B}$而言是假的，但方程(20)事实上对一大类实在的哈密顿量和任意与$Q_{\hbar}^{W}$渐近相等的形变

---

① 更一般地，叶戈洛夫(Egorov)定理表明对一大类哈密顿量我们有$Q_{\hbar}^{W}(f)(t)=Q_{\hbar}^{W}(f_t)+O(\hbar)$。见[Robert, 1987; Dimassi and Sjöstrand, 1999]和[Martinez, 2002]。

② 方程(17)在仿射辛群下等价于外尔量子化的协变性。

量子化映射$Q_\hbar$都有效。亨普(Hepp)[1974]首次建立了此类型的结果，此方向上进一步的工作包括[Yajima(矢岛)，1979；Hogreve *et al.*(霍洛威等)，1983；Wang(王)，1986；Robinson(罗宾逊)，1988a；1988b；Combescure(卡姆斯克)，1992；Arai(新井)，1995；Combescure and Robert，1997；Robert，1998]和[Landsman，1998]。

这类令人印象深刻的结果在薛定谔绘景中也适用。式(17)的对应部分是，对于任意适当光滑的经典哈密顿量h(即使是含时的)，即至多是相空间$\mathbb{R}^{2n}$上正则坐标p和q的二次函数，我们可以构造广义的相干态$\Psi^{\hbar}_{(p,q,C)}$，用h的形式规定的一组经典参数C来标示，从而有：

$$e^{-\frac{it}{\hbar}Q^{W}_{\hbar}(h)}\Psi^{\hbar}_{(p,q,C)}=e^{iS(t)/\hbar}\Psi^{\hbar}_{(p(t),q(t),C(t))} \tag{21}$$

这里，$S(t)$是与由h决定的经典轨迹$(p(t),q(t))$相关的作用，$C(t)$是同样有经典解释的微分方程特定系统的解[Hagedorn(哈格多恩)，1998]。薛定谔的相干态(20)是标准谐振子哈密顿量的一种特殊情形。对于更一般的哈密顿量，我们则有一渐近的结果[Hagedorn and Joye(哈格多恩和乔伊)，1999；2000]①：

$$\lim_{\hbar\to 0}\left\| e^{-\frac{it}{\hbar}Q^{W}_{\hbar}(h)}\Psi^{\hbar}_{(p,q,C)}-e^{iS(t)/\hbar}\Psi^{\hbar}_{(p(t),q(t),C(t))}\right\|=0 \tag{22}$$

再一次，乍一看这样的结果只对用经典运动理解量子动力学有贡献。正如之前提到过的，通过对$Q^{W}_{\hbar}(f)$类型的适当的$\hbar$依赖可观测量取期望值，它们可以转变为关于从量子力学中突现经典运动的陈述。

对于有限的$\hbar$，(22)中的第二项是对第一项的很好近似——对于某些$\gamma>0$，随着$\hbar\to 0$误差甚至与$\mathcal{O}(\exp(-\gamma/\hbar))$一样小，当$t$小于所谓的埃伦费斯特时间时：

$$T_E=\lambda^{-1}\log(\hbar^{-1}) \tag{23}$$

其中λ是哈密顿量的典型的逆时间度量，例如对于混沌系统而言它是最大的李雅普诺夫(Lyapunov)指数。② 这是典型的时间度量，在其上对于拥有广义哈密顿量的含时薛定谔方程的波包解，半经典的近似往往是正确的[Ehrenfest，

① 类似的陈述也见[Paul and Uribe，1995；1996]，还有在式(20)之后列出的文献。

② 回顾此节全文，我们假定$\hbar$通过一恰当的重新标度被设定为无量纲量。

1927；Berry *et al.*（贝里等），1979；Zaslavsky（扎斯拉维斯基），1981；Combescure and Robert，1997；Bambusi *et al.*（班布西等），1999；Hagedorn and Joye，2000］。① 例如，埃伦费斯特［1927］本人估计道，对1克的质量，仅在自由运动大约10^{13}年后波包宽度才变为其2倍。然而，朱瑞克和帕兹［1995］估计说对于土星的卫星亥伯龙，埃伦费斯特时间是20年量级的！这显然指出了从量子力学中推导经典行为（表象）方案的严重问题，这一点影响了对该理论的所有解释。

最后，我们并没有讨论将极限$t\to\infty$与极限$\hbar\to 0$结合起来的重要问题，这应当以下述方式来完成，即T_E保持不变。这双重极限对于量子混沌有特别的重要性，见罗伯特［1998］，和5.6节中引用的大多数文献。

5.3 维格纳函数

量子力学的极限$\hbar\to 0$常常用维格纳［1932］引入的所谓的维格纳函数来讨论。② 每个单位向量（即波函数）$\Psi\in L^2(\mathbb{R}^{2n})$在经典相空间$M=\mathbb{R}^{2n}$上定义了这样一个函数$W_\Psi^\hbar$，通过要求对每个$f\in\mathcal{S}(\mathbb{R}^{2n})$有：

$$(\Psi, Q_\hbar^W(f)\Psi) = \int_{\mathbb{R}^{2n}} \frac{d^npd^nq}{(2\pi)^n} W_\Psi^\hbar(p,q)f(p,q) \tag{24}$$

① 这里人们应该区分含时薛定谔方程的两种不同近似方法。第一，我们有量子力学波包的半经典传播，也就是它的传播遵循计算得到的参数时变特性，而波包如经典运动方程所示依赖于该参数。刚刚引用过的文献表明，直到$t\sim T_E$它对波包的完全量子力学传播都近似得很好。第二，我们有含时WKB近似（对可积系统）和它向混沌系统（它典型地涉及成千上万的项，而非一个）的推广。这第二个近似在更长时间尺度上有效，典型地$t\sim\hbar^{-1/2}$［O'Connor，Tomsovic and Heller（奥康纳、汤姆苏威克和海勒），1992；Heller and Tomsovic，1993；Tomsovic and Heller，1993；2002；Vanicek and Heller（万尼塞克和海勒），2003］。麻烦的是，玻仑廷（Ballentine）在过去数年里论述说，即使在远大于埃伦费斯特时间的尺度内，典型地$t\sim\hbar^{-1/2}$，波包的半经典传播也近似于其量子力学传播［Bllentine *et al.*，1994；Ballentine，2002；2003］。这一论述建立在下述标准之上，即量子的和经典的（即刘维尔）概率在这样的时间尺度上近似相等，但此标准的有效性有赖于量子力学的"统计"或"系综"解释。按照该解释，一个纯态提供了对类似制备系综特定统计特征的描述，但不需要为单个系统提供完备描述，见［Ballentine，1970；1986］。虽然一度被冯·诺伊曼、爱因斯坦和波普尔所支持，但现在这一解释彻底过时了。

② 最初的背景是量子统计力学，人们也可以写下混合态的（24）。关于此的一个综述见［Hillery *et al.*（希勒里等），1984］。

这样一个函数的存在性可以通过明确将其写为如下形式来得到证明：

$$W_{\Psi}^{\hbar}(p,q) = \int_{\mathbb{R}^{2n}} d^n v e^{ipv} \overline{\Psi(q+\frac{1}{2}\hbar v)}\Psi(q-\frac{1}{2}\hbar v) \tag{25}$$

换言之，量子态 Ψ 下，经典可观测量 f 的外尔量子化的量子力学期望值在形式上等于对应于分布 W_{Ψ} 的 f 经典期望值。然而，后者不能够看作是一个概率分布，因为它不一定是正定的。① 虽然有此缺陷，维格纳函数仍然拥有一些诱人的特征。例如，有：

$$Q_{\hbar}^{W}(W_{\Psi}^{\hbar}) = \hbar^{-n}[\Psi] \tag{26}$$

这一多少有些反常的结果意味着，若由 Ψ 定义的维格纳函数被看作是经典的可观测量（虽然它有明显的 $\hbar$ 依赖性），则它的外尔量子化刚好是向 Ψ② 的（$\hbar^{-n}$ 乘以）投影算符。此外，人们也可以得到跃迁概率的如下公式：③

$$|(\Phi,\Psi)|^2 = \hbar^n \int_{\mathbb{R}^{2n}} \frac{d^n p d^n q}{(2\pi)^n} W_{\Psi}^{\hbar}(p,q) W_{\Phi}^{\hbar}(p,q) \tag{27}$$

这一表达拥有直接的直观吸引力，因为右边的被积函数是由相空间中两个维格纳函数重叠的区域支撑的，这刚好与跃迁概率的理念相协调。

维格纳函数对正定性的潜在缺乏可以得到补救，注意到贝雷辛的形变量子化方案，参见(28)，类似地通过下述方法定义了相空间上的函数 $B_{\Psi}^{\hbar}$：

$$(\Psi, Q_{\hbar}^{B}(f)\Psi) = \int_{\mathbb{R}^{2n}} \frac{d^n p d^n q}{(2\pi)^n} B_{\Psi}^{\hbar}(p,q) f(p,q) \tag{28}$$

形式上，(28)与(26)立刻产生了采用薛定谔的相干态(20)的表示：

$$B_{\Psi}^{\hbar}(p,\ q) = \hbar^{-n} |(\Psi_{(p,q)}^{\hbar},\ \Psi)|^2 \tag{29}$$

这一表达被证明是正定的。$B_{\Psi}^{\hbar}$ 的存在性可以通过回顾贝雷辛量子化映射 $f \mapsto Q_{\hbar}^{B}$ 从 $C_0(\mathbb{R}^{2n})$ 到 $\mathcal{B}(L^2(\mathbb{R}^n))$ 是正定的来详细地论证。这意味着，对每个（单位）向量 $\Psi \in L^2(\mathbb{R}^n)$，映射 $f \mapsto (\Psi,\ Q_{\hbar}^{B}(f)\Psi)$ 从 $C_c(\mathbb{R}^{2n})$ 到 $\mathbb{C}$ 是正定的，根据测量

① 事实上，它甚至可能不在 $L^1(\mathbb{R}^{2n})$ 中，因而其总质量不一定被确定，更不用说等于1了。赫德森（Hudson）[1974]给出了由纯态定义维格纳函数正定性的条件。混合态情形见[Bröcker and Werner（布洛克和维纳），1995]。

② 换言之，W_{Ψ} 是投影算符$[\Psi]$的外尔符号。

③ 此公式是严格定义的，因为 $\Psi \in L^2(\mathbb{R}^{2n})$ 意味着 $W_{\Psi}^{\hbar} \in L^2(\mathbb{R}^{2n})$。

理论的里兹(Riesz)定理，必定存在一个在$\mathbb{R}^{2n}$上的测量μ_Ψ，使得$(\Psi, Q_\hbar^B(f)\Psi) = \int d\mu_\Psi f$。则，这一测量恰好由$d\mu_\Psi(p, q) = (2\pi)^{-n} d^n p d^n q B_\Psi^\hbar(p, q)$给出。若$(\Psi, \Psi) = 1$，则$\mu_\Psi$是一个概率测度。相应地，即使是$\hbar$依赖的，$B_\Psi^\hbar$仍定义了一个在相空间上真实的经典概率分布，采用相空间人们可以在某种程度上将量子力学形象化。

对于$\hbar$的有限值，维格纳与贝雷辛分布函数是不同的，因为量子化映射$Q_\hbar^W$与$Q_\hbar^B$不同。在$B_\Psi^\hbar$与$W_\Psi^\hbar$之间的联系很容易被计算出是：

$$B_\Psi^\hbar = W_\Psi^\hbar * g^\hbar \tag{30}$$

其中$g^\hbar$是高斯函数：

$$g^\hbar(p, q) = (2/\hbar)^n \exp(-(p^2 + q^2)/\hbar) \tag{31}$$

这是物理学家们眼中的贝雷辛函数，① 也就是将其看作是用高斯函数改动的维格纳函数，以使其成为正的。但是因为$g^\hbar$随着$\hbar \to 0$收敛于狄拉克δ函数，即相对于在分布意义上的测量$(2\pi)^{-n} d^n p d^n q$，很显然，根据(30)关于分布我们得到：②

$$\lim_{\hbar \to 0}(B_\Psi^\hbar - W_\Psi^\hbar) = 0 \tag{32}$$

同样见(32)。因此在对极限$\hbar \to 0$的研究中，采用维格纳函数没有优势；正相反，在极限过程中它们对正定性的普遍缺乏使得它们比贝雷辛函数更难处理。③例如，人们可能乐于将相干态的渐近行为(8)写为形式$\lim_{\hbar \to 0} W_{\Psi_z^\hbar}^\hbar = \delta_z$。虽然这在分布的意义上确实是真的，相应的极限：

$$\lim_{\hbar \to 0} B_{\Psi_z^\hbar}^\hbar = \delta_z \tag{33}$$

在(概率)测量的意义上存在，因而该极限是在大量检验函数集合上定义的。④

---

① "贝雷辛"函数$B_\Psi^\hbar$是由胡斯米(Husimi)[1940]从一种不同的视角引入的，因此物理学家们实际上称其为胡斯米函数。

② 方程(32)应该被解释为在由$B_\Psi^\hbar - W_\Psi^\hbar$定义的$\mathcal{D}(\mathbb{R}^{2n})$或$\mathcal{S}(\mathbb{R}^{2n})$上分布的极限。这两个函数对$\hbar > 0$都是连续的，但在极限$\hbar \to 0$下失去了此特征，一般地收敛于分布。

③ 见[Robinett, 1993]与[Arai, 1995]。应该指出式(32)表达了作为在$\hbar$无关的检验函数上的分布的维格纳和贝雷辛函数之间的渐近等价性。即使在极限$\hbar \to 0$中，人们有时也会对研究$O(\hbar)$现象感兴趣，在此情形中人们应该做出一个选择。

④ 即$C_0(\mathbb{R}^{2n})$中的那些，而不是在$\mathcal{D}(\mathbb{R}^{2n})$或$\mathcal{S}(\mathbb{R}^{2n})$中的。

这里和接下来，我们将混用记号：若μ^0是在$\mathbb{R}^{2n}$上的某概率测度，$(\Psi^\hbar)$是在$L^2(\mathbb{R}^n)$中用$\hbar$(或也许是其他标号)标示的单位向量列，则通过定义对于$\hbar\to 0$，$B^\hbar_{\Psi^\hbar}\to\mu^0$意味着对任意的$f\in C_c^\infty(\mathbb{R}^{2n})$，有：①

$$\lim_{\hbar\to 0}(\Psi^\hbar, Q_\hbar^B(f)\Psi^\hbar)=\int_{\mathbb{R}^{2n}} d\mu^0 f \tag{34}$$

5.4 能量本征态的经典极限

在(33)处理了相干态$\Psi_z^\hbar$之后，本小节中我们讨论更困难的问题，即计算量子哈密顿量为H的本征态$\Psi_n^\hbar$的极限测量μ^0。因而我们假定H拥有用$n\in\mathbb{N}$(根据简便性可以规定包括零或不包括零)标示的能量本征态$E_n^\hbar$，且由于H对此参数的明确依赖性而依赖于$\hbar$。相关的本征态$\Psi_n^\hbar$则通过定义满足：

$$H\Psi_{\mathbf{n}}^\hbar=E_n^\hbar\Psi_{\mathbf{n}}^\hbar \tag{35}$$

这里我们包含了这样的可能性，即本征值$E_n^\hbar$是简并的，从而标示从$\mathbf{n}$扩展至n。例如，对于一维谐振子，我们有$E_n^\hbar=\hbar\omega(n+\frac{1}{2})(n=0, 1, 2, \cdots)$，其中没有重数，但对于氢原子玻尔本征值$E_n^\hbar=-m_e e^4/2\hbar^2 n^2$(其中$m_e$是电子质量，$e$是其电量)是简并的，它有著名的本征函数$\Psi^\hbar_{(n,l,m)}$[Landau and Lifshitz, 1977]。因此在此情形下有$\mathbf{n}=(n, l, m)$，其中$n=1, 2, 3, \cdots$，满足$l=0, 1, \cdots, n-1$，以及$m=-l, \cdots, l$。

在任何情形下，令$\mathbf{n}\to\infty$都有意义，这当然意味着$n\to\infty$，且额外地涉及将$\mathbf{n}$中的其他指标扩展至无穷大(如上，$\mathbf{n}\to\infty$满足适当的限制条件)。之后若同时令$\hbar\to 0$并且$E_n^\hbar\to E^0$收敛于某个“经典的”值E^0，则人们期待玻尔式的经典行为。取决于人们如何规定可能的其他指标在此极限下的表现，则这里也可能涉及关于与H对易的算符的本征值的类似的渐近条件——可积情形详见下文。我们将这些本征值(包括$E_n^\hbar$)集合记为$\mathbf{E}_{\mathbf{n}}^\hbar$。所以在能量级$E_n^\hbar$非简并的地方，指

① 因为$Q_\hbar^B$可以从$C_c^\infty(\mathbb{R}^{2n})$拓展到$L^\infty(\mathbb{R}^{2n})$，人们可以忽略下述约定，即$\mu^0$在此定义中是一个概率测度，若人们对所有$f\in L^\infty(\mathbb{R}^{2n})$或仅对在$C^*$代数的$C_0(\mathbb{R}^{2n})$的单元化(unitization)中所有$f$要求收敛性。

标 E 就是E。一般地，我们将本征值 $\mathbf{E}_{\mathbf{n}}^{\hbar}$ 随着$\hbar\to 0$ 与 $\mathbf{n}\to\infty$ 时的集体极限记作 $\mathbf{E}^0$。

例如，对于氢原子，在总的角动量上有另外的算符 J^2，且在 z 方向上有角动量算符 J_3。H 本征值为 $E_n^{\hbar}$ 的本征函数 $\Psi_{(n,l,m)}^{\hbar}$ 也是本征值为 $j_n^2=\hbar^2 l(l+1)$ 的 J^2 的本征函数，也是本征值为 $j_3^{\hbar}=\hbar m$ 的 J_3 本征函数。随着 $n\to\infty$ 与$\hbar\to 0$，则可以令 $l\to\infty$ 和 $m\to\pm\infty$ 以使得 $j_{\hbar}^2$ 与 $j_3^{\hbar}$ 接近具体常数。

则兴趣目标现在是相空间上的测量，它是作为贝雷辛函数(29)的极限而得到的，即：

$$\mu_{\mathrm{E}}^0=\lim_{\hbar\to 0,\mathbf{n}\to\infty}B_{\Psi_{\mathbf{n}}^{\hbar}}^{\hbar} \tag{36}$$

虽然量子力学的先驱们毋庸置疑地对这样的量感兴趣，但直到 20 世纪 70 年代才得到详细的结果。有两种情形得到了很好的理解：在本节中我们讨论可积情形，将混沌与更一般的各态历经运动留到 5.6 节。

物理学文献中论证了，对于可积系统，极限测量μ_{E}^0(以 δ 函数的形式)聚集于与 $\mathbf{E}^0$ 相关的不变环面上[Berry，1977a]。① 独立地，数学家们开始研究一个非常类似于 μ_{E}^0 的量，该量通过在黎曼(Riem)流形 M 上拉普拉斯(Laplacian)量本征函数的极限列来确定。这里基础的经典流也是哈密顿量，相应的轨迹是给定度规的测地线，例如见[Klingenberg(克莱因根贝格)，1982；Abraham and Marsden，1985；Katok and Hasselblatt(卡托克和哈苏尔布拉特)，1995]或[Landsman，1998]。② 随之而来的图像很大程度上确证了物理学家们的说法：在可积情形下极限测量 μ_{E}^0 聚集于不变量环面上。见[Charbonnel(卡波内尔)，1986；Zelditch(泽尔蒂奇)，1990，1996a；Toth(托斯)，1996；1999；Nadirashvili *et al.*(纳季拉什维利等)，2001]与[Toth and Zelditch，2002；2003a；

① 事实上，这一结论是从维格纳函数体系中得到的。关于贝里和他的合作者们对此论题的评论见[Ozorio de Almeida(奥瑞欧·德阿梅达)，1988]。

② 可积各态历经运动的最简单例子是 n 环面，其中测地线是线的投影，还有球，其测地线是大圆[Katok and Hasselblatt，1995]。

2003b]。[①] 最后，作为从微局域分析到半经典分析变换的一部分(对比5.2节)，这些结果被推广到量子力学中[Paul and Uribe，1995；1996]。

现在我们给出可积系统(刘维尔类型)中的一些细节，这里包括作为特殊情形的氢原子。可积系统是通过在 $2p$ 维相空间 M 上有 p 个独立的[②]经典可观测量($f_1=h$, f_2, …, f_p)这样的特征来定义的，其中那些可观测量相互间的泊松括号全都为零[Arnold(阿诺德)，1989]。则人们希望存在适当的量子化方案 $Q_\hbar$，在此方案下相应的量子可观测量($Q_\hbar(f_1)=H$, $Q_\hbar(f_2)$, …, $Q_\hbar(f_p)$)都是自伴的且相互对易在共同的核(core)上。[③] 这确实是氢原子的情形，其中(f_1, f_2, f_3)可以看作是(h, j^2, j_3)，j^2 是总的角动量，j_3 是它的 z 分量，[④]H 由式(16)给出，$J^2=Q_\hbar^W(j^2)$ 与 $J_3=Q_\hbar^W(j_3)$。一般地，能量本征函数 $\Psi_{\mathbf{n}}^\hbar$ 将会是算符($Q_\hbar(f_1)$, …, $Q_\hbar(f_p)$)的共同本征函数，从而有 $\mathbf{E}_{\mathbf{n}}^\hbar=(E_{n_1}^\hbar, \cdots, E_{n_p}^\hbar)$，与 $Q_\hbar(f_k)\Psi_{\mathbf{n}}^\hbar=E_{n_k}^\hbar\Psi_{\mathbf{n}}^\hbar$。我们假定子流形 $\cap_{k=1}^p f_k^{-1}(x_k)$ 是紧致的，且对于每个 $x\in\mathbb{R}^p$ 都是连通的，从而根据刘维尔—阿诺德定理它们是环面[Abraham and Marsden，1985；Arnold，1989]。

令 $\hbar\to0$ 与 $\mathbf{n}\to\infty$，从而对于某些点 $E^0=(E_1^0, \cdots, E_p^0)\in\mathbb{R}^p$ 有 $E_{n_k}^\hbar\to E_k^0$，由此得出在(36)中定义的极限测量 $\mu_{\mathbf{E}}^0$ 聚集于不变环面 $\cap_{k=1}^p f_k^{-1}(E_k^0)$ 上。这一环面一般是 p 维的，但对于奇点 E^0 而言它可能是更低维的。特别地，在不变环面是一维的特殊情况下，$\mu_{\mathbf{E}}^0$ 集中于经典的轨道。当然，对于 $p=1$(其中任意的哈密顿量系统是可积的)，这一异常的情形是普遍的。只要用形成谐振子运动闭

---

① 这些论文在忽略 $\hbar\to0$ 的情况下考虑了极限 $n\to\infty$。事实上，一个物理学家会说他们令 $\hbar=1$，那样的话 $E_n\to\infty$。在此过程中物理学家们认为微观的 $E\sim O(\hbar)$ 和宏观的 $E\sim O(1)$ 体系分别对应于 $E\sim O(1)$ 和 $E\to\infty$。

② 即处处有 $df_1\wedge\cdots\wedge df_p\neq0$。为了避免记号上的混淆，我们用 $2p$ 而不是 $2n$ 来表示相空间的维度。

③ 关于量子可积系统没有普遍的理论。奥什兰斯基和佩列洛莫夫(Olshanetsky and Perelomov)[1981；1983]是很好的入门资料。

④ 事实上，若 μ 是 $\mathbb{R}^3$ 上标准 $SO(3)$ 作用的动量映射，则有 $j^2=\sum_{k=1}^3\mu_k^2$ 和 $j_3=\mu_3$。

轨道的椭圆来考虑$\mathbb{R}^2$的叶状结构(foliation)即可。①

在此视角下玻尔的氢原子图像成了什么样呢?② 事实上，许多内容确证了他非凡的物理直觉。玻尔计算的能级是由薛定谔方程给出的那些，因而在成熟的量子力学中也是正确的。他的轨道仅在“对应原理”极限$\hbar\to 0$，$n\to\infty$下有字面意义，然而在那里情况要比人们在可积系统中期待的甚至更好：因为开普(Kepler)勒问题的高度对称性[Guillemin and Sternberg，1990]，我们可以构造这样的本征函数，即其极限测量μ^0聚集于任意希望的经典轨道[Nauenberg，1989]。③ 为了恢复波包的运行，我们需要从具有很高量子数的大量能量本征态来构造波函数，如在2.4节中说明的那样。对于有限的n和$\hbar$，玻尔的轨道似乎没有什么意义，正如海森堡[1969]在其探索时期已经认识到的!④

5.5 WKB 近似

人们可能期待量子力学极限$\hbar\to 0$截面的中心在WKB近似周围，正如实践中所有教科书都将它们对经典极限的讨论建立在这一概念之上。虽然这一方法的范围事实上相当有限，但它确实值得一提。出于简单性，我们限制在不含时

① 考虑球面上的测地线运动是有启发作用的，这一例子可以看作是没有径向自由度的氢原子(从而所考虑的简并也发生在氢原子中)。若我们在球谐函数Y_l^m(它是球面上拉普拉斯量的本征函数)中以$\lim m/l=\cos\varphi$的方式令$l\to\infty$和$m\to\infty$，则不变环面一般是二维的，且发生于$\cos\varphi\neq\pm 1$时；用$\varphi\neq 0$，π这样的值标示的不变环面包括了所有大圆(通过为测地线的每一点增加单位长度的速度和与测地线相切的方向，它们被看作是相空间的一部分)，这些大圆与z轴间的夹角是φ(更精确地，所考虑的夹角是在通过给定大圆的平面法线与z轴间的那个)。然而，对于$\cos\varphi=\pm 1$(即$m=\pm l$)而言，仅存在一个满足$\varphi=0$的大圆，即赤道($\varphi=\pi$情形对应于以相反方向通过的同一赤道)。因此在此情形中的不变环面是一维的。读者可能会诧异于不变环面明确地依赖于变量的选择，但这一特征是所谓简并系统典型具有的，见[Arnold，1989]。

② 这里我们忽略与电磁场的耦合。

③ 对给定的主量子数n，我们通过用适当的径向波函数乘以球谐函数Y_{n-1}^{n-1}来构造本征函数$\Psi^{\hbar}_{(n,n-1,n-1)}$。随着$n\to\infty$和$\hbar\to 0$，极限测量(36)则聚集于一条轨道(而不是一个不变环面)上。现在，在一般可积系统的可能性之外，我们可以用开普勒问题的$SO(4)$对称性对佩列洛莫夫[1986]群理论相干态的构造来找到想要的本征函数，也见[De Bièvre，1992]和[De Bièvre *et al.*，1993]。

④ 晚期玻尔通过他因果描述互补于时空图景的理念承认了这一点，见3.3节。

情形。① 在其最初的构造中，不含时的 WKB 方法试图通过如下类型的波函数来对不含时薛定谔方程 $H\Psi=E\Psi$ 求近似解：

$$\Psi(x)=a_\hbar(x)e^{\frac{i}{\hbar}S(x)} \tag{37}$$

其中 $a_\hbar$ 允许$\hbar$的展开是幂级数。假定哈密顿量 H 具有形式(15)，将拟设的(37)代入薛定谔方程，并用$\hbar$展开，产生了最低阶的经典(不含时)哈密顿—雅可比方程：

$$h\left(\frac{\partial S}{\partial x},\ x\right)=E \tag{38}$$

用下面所谓的齐次的(homogeneous)输运方程来补充：②

$$\left(\frac{1}{2}\Delta S+\sum_k\frac{\partial S}{\partial x^k}\frac{\partial}{\partial x^k}\right)a_0=0 \tag{39}$$

特别地，E 应当是经典上允许的能量值。即使在它适用时(见下文)，在大多数感兴趣的情形下，拟设的(37)仅局域有效(在 x 处)，这产生了严重的问题。这些问题被证明是使用作为对希尔伯特空间$\mathcal{H}=L^2(\mathbb{R}^n)$选择结果的坐标表征的人为现象，且是可以避免的[Maslov and Fedoriuk(马斯洛夫和罗德里克)，1981]：WKB 方法在用辛几何表示的几何重构中真正获得认可。很完备的介绍见[Arnold，1989；Bates and Weinstein(贝茨和温斯坦)，1995]和[Dimassi and Sjöstrand，1999]，深入的说明见[Guillemin and Sternberg，1977；Hörmander，1985a；1985b]和[Duistermaat，1974；1996]。

导向这一重构的基本发现是，S 很少被全域地定义为是在位形空间$\mathbb{R}^n$ 上的光滑函数，它通过$\mathcal{L}=\{(p=dS(x),\ q=x),\ x\in\mathbb{R}^n\}$在相空间 $M=\mathbb{R}^{2n}$上定义了一个子流形$\mathcal{L}$。这一子流形是拥有两个定义特征的拉格朗日量：第一，$\mathcal{L}$是 n 维的；第二，对辛形式(即 $\sum_k dp_k\wedge dq^k$)到$\mathcal{L}$的限制为零。哈密顿—雅可比方程(38)则保证了，拉格朗日子流形$\mathcal{L}\subset M$ 包含于在 M 中的常数能量 E 的曲面 $\sum_E$

① 对于含时情形，参见[Robert，1998]和其中的文献。

② 这里仅说经典的哈密顿量 $h(p,\ q)=p^2/2m+V(q)$。对 $a_\hbar=\sum_j a_j\hbar^j$ 中的展开系数 $a_j(x)$ 而言，$\hbar$的高阶项产生了更多的非齐次输运方程。这些问题可以用递归法得到解决，从式(39)开始。

$= h^{-1}(E)$ 上。因此，任意开始于$\mathcal{L}$的哈密顿运动方程的解都仍然在$\mathcal{L}$中。

一般地，WKB 近似的出发点是拉格朗日子流形 $\mathcal{L} \subset \sum_E \subset M$，而非局域地定义它的某函数 S。通过对几何量子化过程的某种适应，在适当的条件下，人们可能会将适当希尔伯特空间中的单位向量 $\Psi_{\mathcal{L}}$ 与$\mathcal{L}$联系起来，对于小的$\hbar$它是对本征值为 E 的 H 本征函数好的近似。这一策略在可积情形中是成功的，其中非简并的环面(即那些具有最大维 n 的环面)提供了这样的拉格朗日子流形 M，当$\mathcal{L}$满足(广义的)玻尔—索末菲量子化条件时，相关的单位向量 $\Psi_{\mathcal{L}}$ 则被证明是很好地精确定义的。事实上，在可积情形下(36)中的测量 μ_E^0 就是这样被普遍计算的。

若基础的经典系统不可积，则对于要定义的不变环面它可能仍然非常接近于可积情形。这样的系统被称为准可积的或可积系统的微扰，且是由柯尔莫哥洛夫—阿诺德—莫瑟(Kolmogorov-Arnold-Moser，KAM)理论描述的，见[Gallavotti(加拉沃蒂)，1983；Abraham and Marsden，1985；Ozorio de Almeida，1988；Arnold，1989；Lazutkin(拉祖特金)，1993；Gallavotti *et al.*，2004]和许多其他著作。在这样的系统中，WKB 方法继续为与保留下来的不变环面相关的能量本征态提供近似[Colin de Verdière，1977；Lazutkin，1993；Popov(波波夫)，2000]，但已经失去了它的某些吸引力。

在一般的系统中，尤其是混沌系统中，WKB 方法几乎是没有用的。事实上，维纳[1995]的下述定理表明，由 WKB 函数(37)定义的测量μ_E^0 聚集于由 S 定义的拉格朗日子流形$\mathcal{L}$上：

> 对每个$\hbar>0$ 令 $a_\hbar$ 在 $L^2(\mathbb{R}^n)$ 中，点态极限 $a_0=\lim_{\hbar\to 0}a_\hbar$ 也在 $L^2(\mathbb{R}^n)$中，① 且假定 S 几乎在每个地方都是可微的。则对于每个 $f\in C_c^\infty(\mathbb{R}^{2n})$ 有：

$$\lim_{\hbar\to 0}\left(a_\hbar e^{\frac{i}{\hbar}S}, Q_\hbar^B(f)a_\hbar e^{\frac{i}{\hbar}S}\right)=\int_{\mathbb{R}^n}d^nx\,|a_0(x)|^2 f\left(\frac{\partial S}{\partial x},x\right)\tag{40}$$

① 这一假设并不是在维纳[1995]中做出的，他在式(37)中直接假定 $\Psi=a_0\exp(iS/\hbar)$。

正如我们很快要看到的，这一行为对于各态历经系统而言是不可能的，且总的来说这足以决定混沌系统中 WKB 的命运(除了或许可以作为黑客的工具)。

但是，注意对于给定能级 E，目前为止的讨论都关注于在 $\sum_E$ 上经典轨迹(在那里由于能量守恒它们被迫保留)的特征。现在，相空间的其余部分而非 $\sum_E$ 可能也与 H 的谱特征有关，这一点属于量子力学的本质。例如，对形如 $h(p, q)=p^2/2m+V(q)$ 的简单经典哈密顿量，这涉及了所谓的经典禁区 $\{q \in \mathbb{R}^n \mid V(q)>E\}$ (和 p 的任意值)。这里经典运动可以不具有像可积性或各态历经性的特征，因为它并不存在。然而，并且或许是反直觉地，这里恰好是稍加采用 WKB 近似往往最为有效。对于经典禁区中的 $q=x$，假设式(37)应该被下式取代：

$$\Psi(x)=a_\hbar(x) e^{-\frac{S(x)}{\hbar}} \tag{41}$$

其中这一次 S 遵循“虚时间”的哈密顿—雅可比方程,① 即：

$$h\left(i \frac{\partial S}{\partial x}, x\right)=E \tag{42}$$

且输运方程(39)保持不变。例如，可以推断出在一维中拥有类型(3)的哈密顿量的 WKB 函数(41)假定了在禁区中的形式：

$$\Psi(x) \sim e^{-\frac{\sqrt{2m}}{\hbar} \int^{|x|} dy \sqrt{V(y)-E}} \tag{43}$$

它同时解释了量子力学中的隧穿效应(即波函数向禁区中的传播)和此效应在极限$\hbar \to 0$ 下消失的事实。然而，即使这里对 WKB 方法的采用现在很大程度上正被阿格蒙(Agmon)[1982]提出的技巧取代，例如可参见[Hislop and Sigal, 1996]和[Dimassi and Sjöstrand, 1999]的评论。

5.6 尾声：量子混沌

经典力学中的混沌大概在牛顿时代就已经知道了，且众所周知庞加莱

---

① 这一术语来自于拉格朗日形式体系，在那里通过将 $\tau \in \mathbb{R}$ 代入 $t=-i\tau$ 使经典作用 $S=\int dt\, L(t)$ 被 iS 取代。

(1892—1899)①很关注它，但它与量子理论的关联(和潜在威胁)显然是由爱因斯坦[1917]在一篇“完全被忽略40年”的论文[Gutzwiller，1992]②中首次认识到的。当前，量子混沌的研究是所有物理学中最为兴盛的研究之一，正如数不清的会议手册和关于此论题的专著所表明的那样，这些文献包括从古兹维勒[1990]的经典之作到由克维塔诺维奇(Cvitanovic)等[2005]完成的在线版《伟大的任务》(*opus magnum*)。③ 然而，目前仍未完全理解该论题，且这一论题为在经典与量子力学间的相互作用提供了一个引人入胜的检验平台。

人们应当区分量子混沌领域中的不同目标。关于量子混沌的大多数论文和著作关注于对某些具体给定量子系统的半经典分析，这些量子系统拥有一个作为其经典极限的混沌系统。这意味着人们试图用与基础的经典运动相关的数据来近似(对小的$\hbar$)一个合适的量子力学表述。麦克·贝里甚至将此目标描述为量子混沌的“圣杯(Holy Grail)”。5.2节中描述的方法有助于实现此目标，但该方法在很大程度上依赖于动力学的性质。因而本节中我将集中讨论具体到混沌运动的技术细节和结果。

历史上，第一种专门应用于混沌系统的半经典近似理论的新工具是所谓的古兹维勒迹公式。④ 大体来讲，此公式用基础的经典哈密顿量的周期性(即闭的)轨道近似了量子哈密顿量的本征值。⑤ 古兹维勒迹公式并非开始于波函数

① 历史背景也见[Diacu and Holmes(迪阿库和霍姆斯)，1996]和[Barrow-Green(拜罗—格林)，1997]。

② 正是对相同氦原子的研究使得海森堡相信需要用一种本质上新的“量子”力学来取代玻尔与索末菲的不充分的旧量子论。见[Mehra and Renchenberg，1982b]和[Cassidy，1992]。混沌系统的另一个微观案例是在外磁场中的氢原子。

③ 其他令人满意的书目包括，如[Guhr *et al.*(古尔等)，1998]，[Haake(哈克)，2001]和[Reichl(雷克尔)，2004]。

④ 这一归因建立在古兹维勒[1971]基础之上。类似的结果也由Balian and Bloch(巴里安和布洛赫)[1972；1974]独立得到。建立在迹公式基础上的、在数学上有启发性的对半经典物理学的精彩说明也见[Gutzwiller，1990]和[Brack and Bhaduri(布拉克和比哈杜里)，2003]。数学上详细的讨论与论证可以在[Colin de Verdière，1973；Duistermaat and Guillemin，1975；Guillemin and Uribe，1989；Paul and Uribe，1995]和[Combescure *et al.*，1999]中找到。

⑤ 这样的轨道是致密的，但在混沌经典系统中也是刘维尔测度为零的。庞加莱首次认识到了它们的关键作用。

(如 WKB 近似那样)，而开始于传播子 $K(x, y, t)$。物理学家们将此写作 $K(x, y, t)=\langle x|\exp(-itH/\hbar)|y\rangle$，而数学家们将它看作是下式中的格林函数：

$$e^{-\frac{it}{\hbar}H}\Psi(x)=\int d^n y K(x,y,t)\Psi(y) \tag{44}$$

其中 $\Psi\in L^2(\mathbb{R}^n)$。其(分布的)拉普拉斯变换：

$$G(x,y,E)=\frac{1}{i\hbar}\int_0^\infty dt K(x,y,t)e^{-\frac{itE}{\hbar}} \tag{45}$$

同时包含了关于谱和本征函数的信息，因为若前者是离散的，人们有：

$$G(x,y,E)=\sum_j\frac{\Psi_j(x)\overline{\Psi_j(y)}}{E-E_j} \tag{46}$$

有可能通过下式类型的表述来近似 K 或 G 本身：

$$K(x,y,t)\sim(2\pi i\hbar)^{-n/2}\sum_P\sqrt{|\det V_P|}\,e^{\frac{i}{\hbar}S_P(x,y,t)-\frac{1}{2}i\pi\mu_P} \tag{47}$$

其中求和是对在时间 t 内从 y 到 x 的所有经典路径 P 进行的(即解决了经典运动方程的路径)。这样一个路径拥有一相关的作用量 S_P，马斯洛夫指标 μ_P，和范·弗莱克(Van Vleck)[1928]行列式 $\det V_P$[Arnold，1989]。对混沌系统人们常常不得不在求和中包括数以千计的路径，但人们这样做产生的近似被证明是相当成功的[Heller and Tomsovic(海勒和汤姆苏威克)，1993；Tomsovic and Heller，1993]。对拥有离散谱的量子哈密顿量，且拥有混沌运动的基础的经典哈密顿量，古兹维勒迹公式是对下式的半经典近似：

$$g(E)=\int d^n x G(x,x,E)=\sum_j\frac{1}{E-E_j} \tag{48}$$

它具有形式：

$$g(E)\sim g_0(E)+\frac{1}{i\hbar}\sum_P\sum_{k=1}^\infty\frac{T_P}{2\sinh(k\chi_P/2)}e^{\frac{ik}{\hbar}S_P(E)-\frac{1}{2}i\pi\mu_P} \tag{49}$$

其中 g_0 是给出态平均密度的光滑函数。这一次，求和是对在能量 E 处经典哈密顿量所有(主要的)周期路径 P 进行的，有相关的作用量 $S_P(E)=\oint pdq$ (对于给定 E，其中的动量 p 由 P 确定)、周期 T_P 和稳定性指数 χ_P(这是对周边轨迹偏离 P 有多快的衡量)。自爱因斯坦[1917]表示受挫以来，这第一次表明了半经典近似与混沌系统有某些关系。

关于能级的另一重要进展是形成了下述两个关键性的猜想:①

· 若用经典哈密顿量 h 定义的经典动力学是可积的，则 H 的谱是“不关联的”或“随机的”。见[Berry and Tabor(贝里和塔博尔)，1977]。

· 若用 h 定义的经典动力学是混沌的，则 H 的谱是“关联的”或“规则的”。见[Bohigas, Giannoni and Schmit(波依加斯、詹诺尼和施密特)，1984]。

这里用到的关联性和随机性概念可以用如能级间隔的分布和本征值的对关联函数来精确化，初步的处理见[Zelditch, 1996a]和[De Bievre, 2001]，进一步的细节见本节中引用的大部分文献。②

我们现在不考虑本征值而代之以本征函数，且回到极限测量(36)。在非(准)可积情形中，主要结果是:

对遍历(ergodic)经典运动,③ 极限测量 μ_E^0 与在恒定能量表面 $\sum_E = h^{-1}(E)$ 上引起的(归一化)刘维尔测量相重合。④

---

① 严格来讲，这两个猜想都是错误的，例如，谐振子构成了第一个猜想的反例。更多信息见[Zelditch, 1996a]。然而，这两个猜想在更深层的意义上被认为是对的。

② 量子混沌的这一方面在数论中有应用，且甚至可能会导向对黎曼假设的论证，例如见[Sarnak(萨纳克)，1999; Berry and Keating(贝里与基廷)，1999]和其他许多近期的论文。另一个相关的联系是在能级与随机矩阵间的，这一联系与刚刚提到的那个相关，特别见[Guhr *et al.*, 1998]。对所有这些与实践物理学的清晰关联见[Mirlin(梅林)，2000]。

③ 遍历是混沌动力学系统所拥有的最弱的特征。见[Katok and Hasselblatt, 1995; Emch and Liu, 2002; Gallavotti *et al.*, 2004]和多不胜数的其他书目。

④ 在 $\sum_E$ 上未归一化的刘维尔测量 μ_E^u 是通过 $\mu_E^u(B) = \int_B dS_E(x)(\|dh(x)\|)^{-1}$ 来定义的，其中 dS_E 是在 $\sum_E$ 上的表面元素，$B \subset \sum_E$ 是波莱尔集。若 $\sum_E$ 是紧致的，则 $\sum_E$ 上归一化的刘维尔测量 μ_E 由 $\mu_E(B) = \mu_E^u(B)/\mu_E^u(\sum_E)$ 给出。它是 $\sum_E$ 上的一个概率测度，反映了下述事实，即本征向量 $\Psi_n^\hbar$ 被归一化到单位长度从而用于定义量子力学的态。

此结果第一次是在关于紧致双曲线黎曼流形上的遍历短程线(geodetic)运动的数学文献中提出的[Snirelman(斯蒂尔曼), 1974], 它随后被普遍地证明了[Colin de Verdière, 1985; Zelditch, 1987]。① 泽尔蒂奇[1991]、杰拉德和雷特纳姆(Gerard and Leichtnam)[1993]、泽尔蒂奇和泽沃斯基(Zelditch and Zworski)[1996]还有其他人证明了其他遍历系统的这一特征, 据我们所知完备的一般化论证尚待给出。

关于$\mathbb{R}^n$上薛定谔算符的类似版本是在物理学文献[Berry, 1997b; Voros(弗罗斯), 1979]中独立论述的, 且随后由海弗、马丁内斯和罗伯特(Helffer, Martinez and Robert)[1987], 卡波内尔[1992]和保罗与乌里韦[1995]在对势的特定假设下所证明。因而在适当假设下, 对任意的$f\in C_c^\infty(\mathbb{R}^{2n})$, 我们有:

$$\lim_{\hbar\to0,n\to\infty}(\Psi_n^\hbar, Q_\hbar^B(f)\Psi_n^\hbar) = \int_{\Sigma_E} d\mu_E f \tag{50}$$

其中μ_E仍然是在$\sum_E\subset\mathbb{R}^{2n}$上(归一化的)刘维尔测量(假定此空间是紧致的)。特别地, 在遍历情形中μ_E^0仅依赖于E^0, 且对能量本征函数$(\Psi_n^\hbar)$的(几乎)每个列而言只要$E_n^\hbar\to E^0$它都与E^0相同。② 因而对极限测量的支持一致地扩展到了动力学上可能的相空间的最大部分。

对于遍历经典运动μ_E^0是在所陈述的条件下$\sum_E$上的刘维尔测量这一情形, 其结果为"疤痕"(scars)现象留下了空间, 据此现象在混沌系统中极限测量有时集中于周期性的经典轨道。文献中对这一术语的使用是在两种些许不同的方式下进行的。"强"疤痕在极限$\hbar\to0$下幸存, 且聚集于稳定封闭轨道,③ 它们可能

① 在$\hbar=1$的黎曼情形中, 余球丛(cosphere bundle)S^*Q, 即由单位长度的1形式(one-forms)组成的余切丛T^*Q的子丛, 扮演了$\sum_E$的角色。紧致双曲空间提供了遍历短程线运动的低维案例。在外规范场中运动粒子这一物理上重要的案例参见[Zelditch, 1992a]。评论也见施尼雷尔曼(A. I. Shnirelman)为拉祖特金[1993]写的附录和[Nadirashvili *et al.*, 2001]。

② 该结果不一定对所有给定极限行为的列$\Psi_n^\hbar$有效, 而只对"几乎所有"这样的列有效(技术上, 是对密度为1的一组列有效)。对此的简单说明见[De Bièvre, 2001]。

③ 当对γ的每个邻域U存在γ的邻域$V\subset U$, 满足对所有的$z\in V$和所有的t有$z(t)\in U$时, 轨道$\gamma\subset M$被称为稳定的。

来自于本征函数的“独特”列。这些主要是在数学文献中考虑的，参见[Nadirashvili *et al.*，2001]和其中的文献。

另一方面，在物理学文献中，一个疤痕的概念通常指对$\hbar$有限值而言在不稳定封闭轨道周围函数$B^{\hbar}_{\Psi_n^{\hbar}}$的反常聚集，参见式(29)，总体研究见[Heller and Tomsovic，1993；Tomsovic and Heller，1993；Kaplan and Heller(卡普兰和海勒)，1998a；1998b]和[Kaplan，1999]。已证明这样的伤痕对于相关量子系统能谱的说明是至关重要的。这些疤痕在(36)的(双重)极限下不存留的原因是，此极限是相关于$\hbar$无关的光滑检验函数而定义的。物理学上，这意味着随着$\hbar\to 0$人们对越来越多的德布罗意(de Broglie)波长取平均，逐渐损失关于单个波长尺度的信息[Kaplan，1999]。因此为了以一种数学上可行的方式来得到它们，人们应当将式(36)重新定义为逐点的极限[Duclos and Hogreve(杜库拉斯和霍洛威)，1993；Paul and Uribe，1996；1998]。无论如何，在引述的数学结果与物理学家们的发现之间不存在冲突。

量子混沌的另一个目标是在一给定量子力学模型中辨识出混沌现象。这里有些复杂，即人们不能够简单地用在相空间中的离散轨迹来重复混沌的经典定义，因为(根据时间演化的幺正性)在量子力学中$\|\Psi(t)-\Phi(t)\|$对薛定谔方程的解而言在时间t处是常数。然而，这仅仅表明内在量子混沌存在，它的定义必须不同于经典混沌。① 现在在量子理论的代数构造中已经很大程度上达成了这一点[Benatti(贝纳蒂)，1993；Emch *et al.*，1994；Zelditch，1996b；1996c；Belot and Earman，1997；Alicki and Fannes(埃里克和法纳斯)，2001；Narnhofer(奈霍芬)，2001]。最近在此方面的“启发式”文献中最显著的进展是对如下量的研究：

$$M(t)=|\left(e^{-\frac{it}{\hbar}(H+\Sigma)}\Psi,e^{-\frac{it}{\hbar}H}\Psi\right)|^2 \tag{51}$$

其中Ψ是一相干态(或高斯波包)，且$\sum$是哈密顿量H的某个微扰[Peres，1984]。贾拉伯特和帕斯塔斯基(Jalabert and Pastawski)[2001]发现在特定体系

① 正如贝洛特和厄尔曼[1997]指出的，经典力学的考普曼构造排除了经典混沌，若它是用希尔伯特空间中的轨迹来构造的话。从经典到量子混沌概念的过渡可以通过首先重构混沌的经典定义(一般是在相空间中引入轨迹特征)来衔接。

内 $M(t)$ 独立于 $\sum$ 的细节形式，且按照 $\sim\exp(-\lambda t)$ 来衰减，其中 λ 是基础经典系统（最大）的李雅普诺夫指数。这普遍被看作是在此领域上的突破。详细说明和进一步发展见[Cucchietti（库基耶蒂），2004]。

无论如何，经典混沌在量子力学的 $\hbar\to 0$ 极限下出现的可能性绝非是在上述意义上由内在量子混沌的存在所能预言的。① 因为即使在量子动力学被证明是内在非混沌的不可能情形中，其经典极限对于允许多种经典运动是充分突出的，这些经典运动在量子理论中没有定性的对应。这种可能性不只得到了大多数关于量子混沌的文献的确证（其中少有对内在量子混沌运动概念的利用），更多的也得到了不完备运动的可能性的确证。这是一种动力学类型，在其中哈密顿量向量场的流仅定义在一特定时间 $t_f<\infty$（或开始于一初始时间 $t_i>-\infty$）内，这意味着运动方程对 $t>t_f$（或 $t<t_i$）无解。② 则结论性的观点是，幺正量子动力

① 出于下述原因，由[Ford（福特），1988]和其他人给出的论证必须被抛弃。他们的论证结果的大概含义是，由于量子力学不能够在其经典极限下产生混沌因而量子力学是错的。也可参见[Belot and Earman, 1997]。事实上，采用相同的论证，这些作者们同时也能够“证明”相反的陈述，即作为拥有非简并谱的量子理论的经典极限产生的任意经典动力学必须是遍历的。因为量子遍历流的单纯定义显然是，量子的时间演化清除掉了所有在能量 E 处的态；但对于非简并谱，按照本征函数的定义这是同义反复。

② 最简单的例子是拥有测地流的不完备黎曼流形 Q，在此类别中，拥有平坦度规的情形 $Q=(0,1)$ 很难符合简单性。显然，粒子在有限时间内到达两个边界点中的一个，且不知之后会怎么样（或甚至是否存在）。其他的例子来自于 $Q=\mathbb{R}^n$ 上的势 V，它具有经典动力学是不完备的这一特征，见[Reed and Simon，1975]和[Gallavotti，1983]。稍有不同的是，宇宙本身因为大爆炸和可能的大挤压（Big Crunch）而拥有不完备的动力学。

学，虽然内在地完备，也可能拥有如其经典极限那样不完备的运动。①

6. 极限 $N\to\infty$

本节中我们将表明当系统的尺度变得很大时，在什么程度上经典物理学可以近似地从量子理论中突现。严格的经典行为是一种理想化的、适用于其尺度为无穷大的极限，我们在符号上将此极限记作“$\lim N\to\infty$”。正如我们将要看到的，从数学上来讲此极限是前一节中讨论过的极限 $\hbar\to 0$ 的特殊情形。更多的是，我们将表明形式上极限 $N\to\infty$ 甚至可以归类于 C^* 代数连续场和形变量子化的主题(见 4.3 节)之下。因而在假定一个系统是无限的时涉及的理想化的“哲学”本质与在给定(有限)尺度的量子系统中假定 $\hbar\to 0$ 是相同的，特别地，第 1 节中的导言式的评论在这里也适用。

与此类似的讨论是关于从统计力学对热动力学的推导[Emch and Liu,

① 目前宇宙的量子化是未知的，但在黎曼流形——不论其完备与否——上的短程线运动是用 $H=-\frac{\hbar^2}{2m}\Delta$ 量子化的，或许还有与里奇(Ricci)标量 R 成比例的附加项，见[Landsman，1998]，其中 Δ 是拉普拉斯量，且在 $Q=\mathbb{R}^n$ 上的量子化不论经典动力学完备与否都由薛定谔方程(3)给出。在这两种情形中，以及可能更一般的情形中，经典运动的不完备性常常(但不总是)反映在量子哈密顿量在其自然的初始域 $C_c^\infty(Q)$ 上对基本自伴性的缺乏。例如，若 Q 作为一个黎曼流形是完备的，则 Δ 在 $C_c^\infty(Q)$ 上本质上是自伴的[Chernoft，1973；Strichartz(斯特里沙兹)，1983]；且若 Q 是不完备的，则拉普拉斯量常常不能够在此域内本质地自伴，反例见[Horowitz and Marolf(霍罗威茨和马若夫)，1995]。人们可能将后一特征看作是量子力学的不完备性，虽然一个哈密顿量在 $C_c^\infty(Q)$ 上不能是本质自伴的这一情况常常可以通过对边界条件(可能在无穷处)的选择来拓展(必须以一种非唯一的方式)到一个算符的自伴性。按照斯通定理，用每个自伴拓展定义的量子动力学是幺正的(且因而是对所有时间定义的)。类似地，虽然对于将势中的(不)完备经典运动与相应薛定谔算符的本质自伴性(缺乏)关联起来，尚不能做出普遍的论述，但完备性通常意味着本质的自伴性，反之亦然。见[Reed and Simon，1975]第 X.1 节的附录，在那里读者也可能找到经典上不完备但量子力学上完备的运动，以及经典上完备但量子力学上不完备的运动。当前讨论的中心要点是：正如亨普[1974]第一次注意到的，不同的自伴拓展拥有相同的经典极限，指在(20)的意义上，或类似的标准下，即拥有给定的不完备经典动力学。这证明了完备的量子动力学可以拥有不完备的运动作为其经典极限。然而，此领域中仍有许多东西需要理解，也可见[Earman，2005；2006]。

2002；Batterman，2005]。例如，在理论上相位跃迁仅发生在无限系统中，但实践中人们每天都看到它们。因此似乎通过无限数目的水分子来近似化一锅10^{23}个开水分子是有效的。所得到的基本点是，在微观与宏观体系间的区分是不精确的，除非人们允许无限系统是一种理想化，因而人们能够简单地说微观系统是有限的，而宏观系统是无限的。这一过程最终由它所引起的结果而得以证实。

类似地，在量子理论的语境中，经典行为并不发生在有限系统中(当$\hbar>0$固定时)，而正如我们看到的，它发生在无限系统中。对于给定宏观世界所观测到的经典本性，① 最后人们总结出所涉及的理想化显然是有效的。人们不应该被下述事实所迷惑，即此近似化中涉及的粒子数误差(即$\infty-10^{23}=\infty$)要比在真实系统中的粒子数大得多。若问题中所有这10^{23}个粒子是单个被追踪的，近似化确实是不值当的，但关键是，每当对$N=10^{23}$个粒子取平均都可以通过对任意大数N(则人们可能会令其趋于无限)取平均而得到很好的近似时，极限$N\to\infty$是有效的。下面我们将给出此论证的精确版本。

虽然有我们上面的开场白，但无限系统的量子理论有其自身的特征，这些特征需要一个独立的章节来论述。我们的处理是对例如蒂林[1983]、斯多奇[1985]、布拉特利和罗宾逊(Bratteli and Robinson)[1987]、哈格[1992]、荒木[1999]和休厄尔[1986；2002]等人的补充，关于更多的信息和无限量子系统可以参考他们。6.1和6.5节中的理论是一个重构，是依据C^*代数连续场和由拉希奥与维纳[1989；1991]、达菲尔德与维纳[1992a；1992b；1992c]和达菲尔德、鲁斯(Roos)和维纳[Werner，1992]完成的关于所谓量子平均场系统的一系列卓越论文中更为基础部分的形变量子化。这些模型来自于博戈留波夫(Bogoliubov)[1958]和哈格[1962]对超导体的BCS理论的处理，更多重要的贡献由蒂林和外尔勒(Wehrl)[1967]、蒂林[1968]、亨普[1972]、亨普和利布[1973]、瑞克斯[1984]、默西奥和斯多奇(Morchio and Strocchi)[1987]、杜福勒和瑞尔切尔斯(Duffner and Riechers)[1988]、博纳[1988；1989；2000]、昂

① 有著名的介观例外[Leggett，2002；Brezger *et al.*，2002；Chiorescu *et al.*，2003；Marshall *et al.*，2003；Devoret *et al.*，2004]。

纳斯塔尔(Unnerstall)[1990a; 1990b]、贝格雷洛和默西奥(Bagarello and Morchio)[1992]以及休厄尔[2002]等人完成。

6.1 宏观可观测量

我们将要研究的大量子系统由 N 个拥有可观测量 $\mathcal{A}_1$ 的单个代数的单个量子系统所组成。几乎所有的特征已经在最简单的情形 $\mathcal{A}_1=M_2(\mathbb{C})$(即复 2×2 矩阵)中显现了，所以随着抽象性的增加牢记这一例子仍有效。① 接下来的目的是描述在何等精确的意义上宏观可观测量(即那些通过对无限数目的位置取平均而获得可观测量)是“经典的”。

根据单个 C^* 代数 $\mathcal{A}_1$，我们来构造一个在(1)上的 C^* 代数 $\mathcal{A}^{(c)}$ 的连续场如下：

$$I=0\cup1/\mathbb{N}=\{0,\ \cdots,\ 1/N,\ \cdots,\ \frac{1}{3},\ \frac{1}{2},\ 1\}\subset[0,\ 1] \tag{1}$$

我们取：

$$\mathcal{A}_0^{(C)}=C(\mathcal{S}(\mathcal{A}_1)) \tag{2}$$

$$\mathcal{A}_{1/N}^{(C)}=\mathcal{A}_1^N$$

其中 $\mathcal{S}(\mathcal{A}_1)$ 是 $\mathcal{A}_1$ 的态空间(具有弱 * 拓扑)② 且 $\mathcal{A}_1^N=\hat{\otimes}^N\mathcal{A}_1$ 是 $\mathcal{A}_1$ 的 N 个拷贝的(空间)张量积。③ 这说明了 $\mathcal{A}^{(c)}$ 的尾标 c：它指的是极限代数 $\mathcal{A}_0^{(c)}$ 是经典的或对易的这一事实。

例如，取 $\mathcal{A}_1=M_2(\mathbb{C})$。每个态都由一个密度矩阵给出，该密度矩阵具有形式：

---

① 在更具普遍性的相反方向上，值得注意的是下述背景事实上包含了在 $\mathbb{R}^n$ 中(如在 $\mathbb{Z}^n$ 中那样)一般点阵上定义的量子系统。因为我们可以重新标示事物以使得 $\mathcal{A}_{1/N}$ 位于包含在半径为 N 的球中所有点阵点 Λ 的可观测量代数之下。如此一来，极限 $N\to\infty$ 对应于极限 $\Lambda\to\mathbb{Z}^n$。

② 在此拓扑中，当对每个 $A\in\mathcal{A}_1\omega_\lambda(A)\to\omega(A)$ 时，有 $\omega_\lambda\to\omega$。

③ 当 $\mathcal{A}_1$ 是有限维时张量积是唯一的。一般地，在这里我们需要投影张量积。这里的理论要点相同：在 $\hat{\otimes}^N\mathcal{A}_1$ 上的任意张量积态 $\omega_1\otimes\cdots\otimes\omega_N$——通过 $\omega_1\otimes\cdots\otimes\omega_N(A_1\otimes\cdots\otimes A_N)=\omega_1(A_1)\cdots\omega_N(A_N)$ 在初等张量上定义的——根据连续性拓展到了 $\hat{\otimes}^N\mathcal{A}_1$ 上的态。

$$\rho(x, y, z)=\frac{1}{2}\begin{pmatrix}1+z & x-iy\\ x+iy & 1-z\end{pmatrix} \tag{3}$$

对某个$(x, y, z)\in\mathbb{R}^3$满足$x^2+y^2+z^2\leqslant 1$。因此$\mathcal{S}(M_2(\mathbb{C}))$与$\mathbb{R}^3$中的三球$B^3$同构(作为一紧致凸集)。纯态正好是边界上的点,[①] 即有$x^2+y^2+z^2=1$的密度矩阵(因为这些矩阵独自定义了一维的投影)。[②]

为了定义场的连续截面，我们引入对称化映射：j_{NM}: $\mathcal{A}_1^M\to\mathcal{A}_1^N$，定义为：

$$j_{NM}(A_M)=S_N(A_M\otimes 1\otimes\cdots\otimes 1) \tag{4}$$

其中我们有$N-M$份单位元$1\in\mathcal{A}_1$的拷贝，从而得到$\mathcal{A}_1^N$的一个元素。对称化算符S_N: $\mathcal{A}_1^N\to\mathcal{A}_1^N$由对下式(线性和连续的)拓展给出：

$$S_N(B_1\otimes\cdots\otimes B_N)=\frac{1}{N!}\sum_{\sigma\in\mathfrak{S}_N}B_{\sigma(1)}\otimes\cdots\otimes B_{\sigma(N)} \tag{5}$$

其中$\mathfrak{S}_N$是在N个元素和所有$i=1, \cdots, N$的$B_i\in\mathcal{A}_1$上的置换群(即对称群)。例如，j_{N1}: $\mathcal{A}_1\to\mathcal{A}_1^N$由下式给出：

$$j_{N1}(B)=\overline{B}^{(N)}=\frac{1}{N}\sum_{k=1}^{N}1\otimes\cdots\otimes B_{(k)}\otimes 1\cdots\otimes 1 \tag{6}$$

其中$B_{(k)}$是被看作在$\mathcal{A}_1^N$中$\mathcal{A}_1$的第k个拷贝的一个元素B。正如我们的记号$\overline{B}^{(N)}$所表明的，这仅仅是在$\mathcal{A}_1$的所有拷贝上对B“取平均”。更一般地，在构造$j_{NM}(A_M)$时，涉及M位置的算符$A_M\in\mathcal{A}_1^M$是对$N\geqslant M$位置的平均。当$N\to\infty$时这意味着我们构造了一个对M粒子算符的宏观平均。

---

① 凸集K的极边界$\partial_e K$包括了所有的$\omega\in K$，对它而言，对某个$p\in(0, 1)$和ρ, $\sigma\in K$, $\omega=p\rho+(1-p)\sigma$意味着$\rho=\sigma=\omega$。若$K=\mathcal{S}(\mathcal{A})$是$C^*$代数$\mathcal{A}$的态空间，则极边界由$\mathcal{A}$上的纯态组成，$\mathcal{S}(\mathcal{A})$的剩余部分由混合态组成。若$K$嵌入到一个向量空间中，则极边界$\partial_e K$可能会也可能不会与$K$的几何边界$\partial K$相重合。在$K=B^3\subset\mathbb{R}^3$情形中会，但对于$\mathbb{R}^2$中的等边三角形则不会，因为$\partial_e K$只是由三角形的隅角组成，而几何边界也包括了三角形的边。

② 式(3)拥有形式$\rho(x, y, z)=\frac{1}{2}(x\sigma_x+y\sigma_y+z\sigma_z)$，其中$\sigma_i$是泡利矩阵。这产生了在$\mathbb{R}^3$与$SO(3)$在其在$\mathbb{C}^2$上的自旋$\frac{1}{2}$表征$\mathcal{D}_{1/2}$中李代数之间的同构。这一同构将$\mathbb{R}^3$上$SO(3)$的定义作用与其在$M_2(\mathbb{C})$上的伴随作用缠结在一起。即对于任意的旋转$R$，我们有$\rho(R_{\mathbf{x}})=\mathcal{D}_{1/2}(R)\rho(\mathbf{x})\mathcal{D}_{1/2}(R)^{-1}$，随后我们会用到这一点(见6.5节)。

我们说在 $A_N \in \mathcal{A}_1^N$ 时列 $A=(A_1, A_2, \cdots)$ 是对称的，当对于某固定的 M 和所有 $N \geqslant M$ 有下式时：

$$A_N = j_{NM}(A_M) \tag{7}$$

换言之，对称列的尾部完全是由“被平均的”或“密集的”可观测量组成，它们在极限 $N \to \infty$ 下成为了宏观的可观测量。这样的列拥有重要的特征，即它们在此极限下对易。更精确地，若 A 和 A' 是对称列，则有：

$$\lim_{N\to\infty} \| A_N A'_N - A'_N A_N \| = 0 \tag{8}$$

作为一有启发作用的特殊情形我们取 $A_N = j_{N1}(B)$ 和 $A'_N = j_{N1}(C)$，有 $B, C \in \mathcal{A}_1$。我们很快根据 $k \neq l$ 时的关系 $[B_{(k)}, C_{(l)}]=0$ 得到：

$$[\overline{B}^{(N)}, \overline{C}^{(N)}] = \frac{1}{N}\overline{[B, C]}^{(N)} \tag{9}$$

例如，若 $\mathcal{A}_1 = M_2(\mathbb{C})$ 且对于 B 和 C 而言取自旋 1/2 算符 $S_j = \frac{\hbar}{2}\sigma_j$，有 $j=1, 2, 3$（其中 σ_j 是泡利矩阵），则有：

$$[\overline{S}_j^{(N)}, \overline{S}_k^{(N)}] = i\frac{\hbar}{N}\epsilon_{jkl}\overline{S}_l^{(N)} \tag{10}$$

这表明对一个粒子算符取平均产生了在形式上类似于所涉及的一个粒子算符的对易关系，但伴随普朗克常数 $\hbar$ 被变量 $\hbar/N$ 所取代。对常数 $\hbar=1$ 而言，这产生了区间(1)，在其上我们定义了 C^* 代数连续场，对 $\hbar$ 的任意其他常数值而言场在 $I=0\cup\hbar/\mathbb{N}$ 上被定义，这当然只改变了所涉及的 C^* 代数的标示。

我们回到一般情形，用列 $A=(A_0, A_1, A_2, \cdots)$ 来定义具有纤维(2)的场截面，如前有 $A_0 \in \mathcal{A}_0^{(c)}$ 和 $A_N \in \mathcal{A}_1^N$（即对应的截面是 $0 \mapsto A_0$ 和 $1/N \mapsto A_N$）。之后我们通过宣称列 A 定义了一连续截面来完善我们对连续场的定义，当且仅当：

- $(A_1, A_2, \cdots)$ 是近似对称的，在如下意义上，即对任意的 $\varepsilon>0$ 存在一个 N_ε 和一个对称列 A' 满足对所有的 $N \geqslant N_\varepsilon$ 有 $\| A_N - A'_N \| < \varepsilon$；①

- $A_0(\omega) = \lim_{N\to\infty} \omega^N(A_N)$，其中 $\omega \in \mathcal{S}(\mathcal{A}_1)$ 和 $\omega^N \in \mathcal{S}(\mathcal{A}_1^N)$ 是 ω 的 N 个拷贝的张量积，由对下式的（线性的和连续的）扩展所定义：

$$\omega^N(B_1 \otimes \cdots \otimes B_N) = \omega(B_1)\cdots\omega(B_N) \tag{11}$$

---

① 一个对称列显然是近似对称的。

根据对近似对称列的定义，此极限存在。①

不难证明对连续截面的这种选择确实在有纤维(2)的 $I=0\cup 1/\mathbb{N}$ 上定义了 C^* 代数的连续场。关键点在于：

$$\lim_{N\to\infty}\|A_N\| = \|A_0\| \tag{12}$$

每当(A_0，A_1，A_2，…)满足上述两个条件时。② 很容易表明对称列下的情形，③且据此得到近似对称列的情形。

与式(8)相一致，我们得到结论，即在极限 $N\to\infty$ 下宏观可观测量将它们自身组织到一个与 $C(\mathcal{S}(\mathcal{A}_1))$ 同构的对易 C^* 代数中。

6.2 准定域可观测量

在量子理论的 C^* 代数方法中，无限系统常常用 C^* 代数的归纳极限和相关的准定域可观测量方法来描述[Thirring，1983；Strocchi，1985；Bratteli and Robinson，1981；1987；Haag，1992；Araki，1999；Sewell，1986；2002]。为了得到下述情形中的这些概念，我们做如下处理[Duffield and Werner，1992c]。

列 $A=(A_1, A_2, \cdots)$（如上所述，其中 $A_N\in\mathcal{A}_1^N$）被称为定域的，当对某个固定的M 和所有的 $N\geqslant M$，我们有 $A_N=A_M\otimes 1\otimes\cdots\otimes 1$（其中有 $N-M$ 份单位元 1

---

① 若(A_1，A_2，…)关于(7)对称，对于 $N>M$ 我们有 $\omega^N(A_N)=\omega^M(A_M)$，从而列($\omega^N(A_N)$)的尾部独立于 N。在近似对称情形中我们很容易证明($\omega^N(A_N)$)是一个柯西(Cauchy)列。

② 给定式(12)，这一论述来自于兰兹曼[1998]中的命题II.1.2.3 和$\mathcal{S}(\mathcal{A}_1)$上以所说方式产生的函数集 A_0 在 $C(\mathcal{S}(\mathcal{A}_1))$（具有上确界模）中是致密的这一事实。这产生自斯通—外尔施特拉斯(Stone-Weierstrass)定理，按照该定理我们推论出所涉函数甚至穷尽了$\mathcal{S}(\mathcal{A}_1)$。

③ 假定式(7)，从而对 $N\geqslant M$ 有 $\|A_N\|=\|j_{NN}(A_N)\|$。通过 C^* 公理 $\|A^*A\|=\|A^2\|$，足以证明对于 $A_0^*=A_0$ 式(12)成立，这意味着对所有 $N\geqslant M$ 有 $A_M^*=A_M$，进而有 $A_N^*=A_N$，则我们得到 $\|A_N\|=\sup\{|\rho(A_N)|, \rho\in\mathcal{S}(\mathcal{A}_1^N)\}$。因为 A_N 的特殊形式，我们要用所有置换恒定态的集合$\mathcal{S}^p(\mathcal{A}_1^N)$的上确界来代替$\mathcal{A}_1^N$上所有态集合$\mathcal{S}(\mathcal{A}_1^N)$的上确界，而前者相应地又可以用 $\mathcal{S}^p(\mathcal{A}_1^N)$极边界 $\partial\mathcal{S}^p(\mathcal{A}_1^N)$的上确界来代替。后者是由形如 $\rho=\omega^N$ 的所有态组成的，这是广为人知的[Størmer(斯托默)，1969]，也见 6.2 节，因而有 $\|A_N\|=\sup\{|\omega^N(A_N)|, \omega\in\mathcal{S}(\mathcal{A}_1)\}$。这事实上等于 $\|A_M\|=\sup\{|\omega^M(A_M)|\}$。现在$\mathcal{A}_0^{(c)}$中的模为 $\|A_0\|=\sup\{|A_0(\omega)|, \omega\in\mathcal{S}(\mathcal{A}_1)\}$，且根据 A_0 的定义我们有 $A_0(\omega)=\omega^M(A_M)$。因此得到了(12)。

$\in \mathcal{A}_1$)时，参见(4)。一个列被说成是准定域的，当对任意的 $\varepsilon>0$，存在一个 N_ε 和一个定域的列 A' 使得对所有的 $N \geqslant N_\varepsilon$ 有 $\| A_N - A'_N \| < \varepsilon$。在此基础上，我们定义 C^* 代数($\mathcal{A}_1^N$)族的 C^* 代数归纳极限为：

$$\overline{\cup_{N\in\mathbb{N}} \mathcal{A}_1^N} \tag{13}$$

C^* 代数($\mathcal{A}_1^N$)族是关于由 $A_N \mapsto A_N \otimes 1$ 给出的包含映射 $\mathcal{A}_1^N \hookrightarrow \mathcal{A}_1^{N+1}$ 的。作为一个集合，(13)包含了当 $\lim_{N\to\infty} \| A_N - B_N \| = 0$ 时在等价关系 $A \sim B$ 下的准定域列 A 的所有等价类 $[A] \equiv A_0$。在 $\overline{\cup_{N\in\mathbb{N}} \mathcal{A}_1^N}$ 上的模为：

$$\| A_0 \| = \lim_{N\to\infty} \| A_N \| \tag{14}$$

且剩余的 C^* 代数结构是以一种明显的方式，即 $A_0^* = [A^*]$ 其中 $A^* = (A_1^*, A_2^*, \cdots)$，从准定域列中继承而来的。正如记号所表明的，每个 $\mathcal{A}_1^N$ 都是作为一个 C^* 代数包含在 $\overline{\cup_{N\in\mathbb{N}} \mathcal{A}_1^N}$ 中的，通过将 $A_N \in \mathcal{A}_1^N$ 看作是定域(且因而是准定域的)列 $A = (0, \cdots, 0, A_N \otimes 1, A_N \otimes 1 \otimes 1, \cdots)$，且正如刚刚所说明的那样在 $\overline{\cup_{N\in\mathbb{N}} \mathcal{A}_1^N}$ 中形成了其等价类 A_0。[①] 隐含于(13)是所研究的无限系统的“那个”可观测量代数这一普遍观念中的假定是，通过定域性或某种其他人为限制，系统的无限尾部是不可及的，因而该可观测量必须任意接近(即在模上)于形如 $A_N \otimes 1 \otimes 1, \cdots$ 的算符，对某个有限的 N 而言。

这将我们导向了在 $0 \cup 1/\mathbb{N}$ 上的第二个 C^* 代数连续场 $\mathcal{A}^{(q)}$，其上有纤维：

$$\mathcal{A}_0^{(q)} = \overline{\cup_{N\in\mathbb{N}} \mathcal{A}_1^N}$$

$$\mathcal{A}_{1/N}^{(q)} = \mathcal{A}_1^N \tag{15}$$

因而指标 q 提醒我们有这样的事实，即极限代数 $\mathcal{A}_0^{(q)}$ 是由准定域或准量子力学可观测量组成的。我们用 $0 \cup 1/\mathbb{N}$ 上一个 C^* 代数连续场 $\mathcal{A}^{(q)}$ 的结构来装备 C^* 代数(15)的集合，通过宣称连续截面的形式为 $(A_0, A_1, A_2, \cdots)$，其中 $(A_1, A_2, \cdots)$ 是准定域的，且 A_0 如刚才所说明的那样是通过此准定域列来定义的。[②] 对于 $N<\infty$，此场拥有与前一小节中连续场 $\mathcal{A}$ 相同的纤维：

---

① 当然，被设为零的表值 $A_1, \cdots, A_{N-1}$ 是任意的。

② 根据式(14)和[Landsman, 1998]中的命题 II. 1. 2. 3 得到这定义了一个连续场。

$$\mathcal{A}_{1/N}^{(q)}=\mathcal{A}_{1/N}^{(c)}=\mathcal{A}_1^N \tag{16}$$

但纤维$\mathcal{A}_0^{(q)}$完全不同于$\mathcal{A}_0^{(c)}$。特别地，若$\mathcal{A}_1$是非对易的，则$\mathcal{A}_0^{(q)}$也是，因为它包含了所有的$\mathcal{A}_1^N$。

在C^*代数连续场$\mathcal{A}^{(q)}$与$\mathcal{A}^{(c)}$间的关系可以用两种不同(但有关的)方式来研究。第一，我们可以构造所有C^*代数$\mathcal{A}_1^N(N<\infty)$，还有单个希尔伯特空间上的$\mathcal{A}_0^{(c)}$和$\mathcal{A}_0^{(q)}$的具体表征，这一方法产生了传统意义上的超选规则。下一小节将继续使用这种方法。第二，我们可以关注态(ω_1，$\omega_{1/2}$，…，$\omega_{1/N}$，…)(其中$\omega_{1/N}$是在$\mathcal{A}_1^N$上的态)的那些分别允许在$\mathcal{A}_0^{(c)}$和$\mathcal{A}_0^{(q)}$上的极限态$\omega_0^{(c)}$和$\omega_0^{(q)}$的族，使得继而产生的态($\omega_0^{(c)}$，ω_1，$\omega_{1/2}$，…)和($\omega_0^{(q)}$，ω_1，$\omega_{1/2}$，…)族分别是在$\mathcal{A}^{(c)}$和$\mathcal{A}^{(q)}$上态的连续场(参见5.1节结尾)。

现在$\mathcal{A}_0^{(q)}$上的任意态$\omega_0^{(q)}$通过限制在$\mathcal{A}_1^N$上定义的态$\omega_0^{(q)}{}_{|1/N}$，且继而产生的$\mathcal{A}^{(q)}$上的态场显然是连续的。反过来，$\mathcal{A}^{(q)}$上态的任意连续场($\omega_0^{(q)}$，ω_1，$\omega_{1/2}$，…，$\omega_{1/N}$，…)变得任意接近于上述N很大时的场的类型。① 然而，对$\mathcal{A}_0^{(q)}$上给定态$\omega_0^{(q)}$对$\mathcal{A}_1^N$的限制$\omega_0^{(q)}{}_{|1/N}$在$N\to\infty$时可能不会收敛于$\mathcal{A}_0^{(c)}$上的态$\omega_0^{(c)}$。这里称$\overline{\cup_{N\in\mathbb{N}}\mathcal{A}_1^N}$上不具有此特征的态$\omega_0^{(q)}$是经典的。换言之，$\omega_0^{(q)}{}_{|1/N}$是经典的，当在$\mathcal{S}(\mathcal{A}_1)$上存在一个概率测度$\mu_0$满足：

$$\lim_{N\to\infty}\int_{\mathcal{S}(\mathcal{A}_1)}d\mu_0(\rho)\left(\rho^N(A_N)-\omega_{0|1/N}^{(q)}(A_N)\right)=0 \tag{17}$$

对每个(近似的)对称列(A_1，A_2，…)而言。为了分析此概念我们需要引入关于一般C^*代数及其表征的简短内容。

· C^*代数$\mathcal{B}$的态空间$\mathcal{S}(\mathcal{B})$中的叶形线(folium)是凸的，$\mathcal{S}(\mathcal{B})$的模闭子空间$\mathcal{F}$拥有下述特征，即若$\omega\in\mathcal{F}$且$B\in\mathcal{B}$则$\omega(B^*B)>0$，从而"约化"态ω_B：$A\mapsto\omega(B^*AB)/\omega(B^*B)$必须在$\mathcal{F}$中[Haag *et al.*，

① 对于任意固定的准列(A_1，A_2，…)和$\varepsilon>0$，存在一个N_ε满足对所有的$N>N_\varepsilon$有$|\omega_{1/N}(A_N)-\omega_{0|1/N}^{(q)}(A_N)|<\varepsilon$。

1970]。[1] 例如，若 π 是在希尔伯特空间$\mathcal{H}$上$\mathcal{B}$的表征，则在$\mathcal{H}$上所有密度矩阵的集合(即在$\mathcal{B}$上的 π 正常态)[2]构成了叶形线$\mathcal{F}_\pi$。特别地，$\mathcal{B}$上的每个态 ω 通过其 GNS 表征 π_ω 定义了叶形线$\mathcal{F}_\omega=\mathcal{F}_{\pi_\omega}$。

· 若 π 的亚表征都不(幺正地)等价于 π' 的子表征，则两个表征 π 和 π'被称为是不相交的，写作 $\pi\perp\pi'$，反之亦然。当 π 没有与 π'不相交的子表征时，它们被称为是准等价的，写作 $\pi\sim\pi'$，反之亦然。[3] 准等价是在表征集合上的等价关系 ~。见[Kadison and Ringrose, 1986, Chapter 10]。

· 类似地，当两个态 ρ，σ 相应的 GNS 表征拥有这些特征时，它们被称为准等价的($\rho\sim\sigma$)，或不相交的($\rho\perp\sigma$)。

· 态 ω 称作原初的(primary)，当它相应的冯·诺伊曼代数 $\pi_\omega(\mathcal{B})''$是因子时。[4] 等价地，当且仅当 $\pi_\omega(\mathcal{B})$ 的每个子表征准等价于 $\pi_\omega(\mathcal{B})$ 时 ω 是原初的，这是当且仅当 $\pi_\omega(\mathcal{B})$ 不允许(非平凡)分解为两个不相交子表征的直和时的情形。

现在，在$\mathcal{S}(\mathcal{B})$的叶形线与$\mathcal{B}$的准等价表征类之间存在双射对应，即当且仅当 $\pi\sim\pi'$时有$\mathcal{F}_\pi=\mathcal{F}_{\pi'}$。此外，(正如从 GNS 构造中看到的)任意叶形线$\mathcal{F}\subset\mathcal{S}(\mathcal{B})$对于某个表征 $\pi(\mathcal{B})$ 具有形式$\mathcal{F}=\mathcal{F}_\pi$。注意若 π 是内射的(injective)(即忠实的)，则相应的叶形线在弱*拓扑中按照菲尔(Fell)定理在$\mathcal{B}$中是致密的。所

① 也参见[Haag, 1992]。名称"叶形线"是非常糟糕的选择，因为$\mathcal{S}(\mathcal{B})$绝不是通过其叶形线来分叶(foliated)的，例如，一条叶形线也可能包含子叶形线。

② $\mathcal{B}$上的态 ω 称为 π 正常的，当它对某个密度矩阵 ρ 而言具有形式 $\omega(B)=\mathrm{Tr}\,\rho\pi(B)$。因此，$\pi$ 正常态是在冯·诺伊曼代数 $\pi(\mathcal{B})''$上的正常态。

③ 等价地，两个表征 π 和 π'是不相交的，当且仅当没有 π 正常态是 π'正常的，且反之亦然，当且仅当每个 π 正常态是 π'正常的时这两个表征是准等价的，且反之亦然。

④ 作用于希尔伯特空间上的冯·诺伊曼代数$\mathcal{M}$称为一个因子，当其中心$\mathcal{M}\cap\mathcal{M}'$是平凡的，即由单位元的倍数组成。

以在$\mathcal{B}$是单的情形中,① 任意叶形线在态空间中是弱*致密的。

两个态不需要不相交或准等价。然而，这种二分法在原初态类中适用。因此两个原初态要么不相交要么准等价。若ω是原初的，则在π_ω叶形线上的每个态也都是原初的，且与ω准等价。另一方面，若ρ和σ是原初的和不相交的，则有$\mathcal{F}_\rho \cap \mathcal{F}_\sigma = \emptyset$。当然，纯态是原初的。② 此外，在热力学中，纯相是用原初 KMS 态来描述的[Emch and Knops(埃姆什和诺普斯)，1970；Bratteli and Robinson，1981；Haag，1992；Sewell，2002]。这种在原初态与某种“纯度”间的显著关系得到了我们对宏观可观测量描述的确证:③

· 若$\omega_0^{(q)}$是在$\mathcal{A}_0^{(q)} = \overline{\cup_{N\in\mathbb{N}} \mathcal{A}_1^N}$上的一个经典原初态，则它在$\mathcal{A}_0^{(c)} = C(\mathcal{S}(\mathcal{A}_1))$上对应的极限态$\omega_0^{(c)}$是纯的，且因此由$\mathcal{S}(\mathcal{A}_1)$中的一点给出。

· 若$\rho_0^{(q)}$和$\sigma_0^{(q)}$是在$\mathcal{A}_0^{(q)}$上的经典原初态，则有:

$$\rho_0^{(c)} = \sigma_0^{(c)} \Leftrightarrow \rho_0^{(q)} \sim \sigma_0^{(q)} \tag{18}$$

$$\rho_0^{(c)} \neq \sigma_0^{(c)} \Leftrightarrow \rho_0^{(q)} \perp \sigma_0^{(q)} \tag{19}$$

正如在(17)中，在$C(\mathcal{S}(\mathcal{A}_1))$上有极限态$\omega_0^{(c)}$的一般经典态$\omega_0^{(q)}$通过下式定义了一个在$\mathcal{S}(\mathcal{A}_1)$上的概率测度$\mu_0$:

$$\omega_0^{(c)}(f) = \int_{\mathcal{S}(\mathcal{A}_1)} d\mu_0 f \tag{20}$$

它描述了宏观可观测量在那个态中的概率分布。正如我们看到的，此分布对原初态而言是一个δ函数。无论如何，它对$\omega_0^{(q)}$的微观细节不敏感，在对$\omega_0^{(q)}$的定域修正并不影响极限态$\omega_0^{(c)}$的意义上[Sewell，2002]。也就是说，很

① 在它没有闭双边理想的意义上。例如，对任意的n矩阵代数$M_n(\mathbb{C})$是单的，正如其无限维的相似，也就是在希尔伯特空间上所有紧致算符的C^*代数。无限量子系统准定域可观测量的C^*代数往往也是单的。

② 因为对应的 GNS 表征π_ω是不可约的，故$\pi_\omega(\mathcal{B})'' = \mathcal{B}(\mathcal{H}_\omega)$是一个因子。

③ 这些陈述很容易从休厄尔[2002]的第 2.6.5 小节中得到，而后者建立在亨普[1972]基础上。

容易根据(8)和 GNS 表征是循环的这一事实来强化上面的第二个陈述：

> 在经典态 $\omega_0^{(q)}$ 的叶形线 $\mathcal{F}_{\omega_0^{(q)}}$ 中的每个态都必然是经典的，且在 $\mathcal{A}_0^{(c)}$ 上拥有与 $\omega_0^{(q)}$ 相同的极限态。

为使这一讨论更加具体，现在我们在 $\overline{\cup_{N\in\mathbb{N}}\mathcal{A}_1^N}$ 上确定一类重要的经典态。我们说一个在此 C^* 代数上的态 ω 是置换不变的，当它向 $\mathcal{A}_1^N$ 的每一个限制在 $\mathcal{A}_1^N$ 上的对称群 $\mathfrak{S}_N$ 的自然作用下是不变的，即 $\sigma\in\mathfrak{S}_N$ 将一基本张量 $A_N=B_1\otimes\cdots\otimes B_N\in\mathcal{A}_1^N$ 映射到 $B_{\sigma(1)}\otimes\cdots\otimes B_{\sigma(N)}$，参见式(5)。斯托默[1969]分析了在 $\mathcal{S}(\mathcal{A}_0^{(q)})$ 中所有置换不变态的集合 $\mathcal{S}^{\mathfrak{S}}$ 的结构。就像任意的紧致凸集，(弱*闭)凸剥离了其极边界 $\partial_e\mathcal{S}^{\mathfrak{S}}$。后者由所有无限积态 $\omega=\rho^\infty$ 组成，其中 $\rho\in\mathcal{S}(\mathcal{A}_1)$。也就是说若 $A_0\in\mathcal{A}_0^{(q)}$ 是一等价类 $[A_1, A_2, \cdots]$，则有：

$$\rho^\infty(A_0)=\lim_{N\to\infty}\rho^N(A_N) \tag{21}$$

参见式(11)。等价地，ω 对任意 $\mathcal{A}_1^N\subset\mathcal{A}_0^{(q)}$ 的限制由 $\otimes^N\rho$ 给出。因此 $\partial_e\mathcal{S}^{\mathfrak{S}}$ 明显与 $\mathcal{S}(\mathcal{A}_1)$ 同构(作为一紧致凸集)，且 $\mathcal{S}^{\mathfrak{S}}$ 中的原初态正好是 $\partial_e\mathcal{S}^{\mathfrak{S}}$ 中的元素。

在 $\mathcal{S}^{\mathfrak{S}}$ 中的一般态 $\omega_0^{(q)}$ 拥有唯一的分解：①

$$\omega_0^{(q)}(A_0) = \int_{\mathcal{S}(\mathcal{A}_1)} d\mu(\rho)\rho^\infty(A_0) \tag{22}$$

其中 μ 是在 $\mathcal{S}(\mathcal{A}_1)$ 上的概率测量，且有 $A_0\in\mathcal{A}_0^{(q)}$。② 则如下对该抽象理论的精彩呈现[Unnerstall, 1990a, b]显然来自于式(17)和式(22)：

---

① 这是因为 $\mathcal{S}^{\mathfrak{S}}$ 是所谓的鲍厄(Bauer)单形(simplex)[Alfsen(阿夫森), 1970]。这是一个紧致凸集 K，它的极边界 $\partial_e K$ 是闭的，且对 K 而言每一个 $\omega\in K$ 作为受 $\partial_e K$ 支撑的概率测度都有唯一的分解，在如下意义上，即 K 上的每一个连续仿射函数 a 有 $a(\omega)=\int_{\partial_e K} d\mu(\rho)a(\rho)$。对于一个有单位元的 C^* 代数 $\mathcal{A}$ 而言，态空间 $K=\mathcal{S}(\mathcal{A})$ 上的连续仿射函数恰好是 $\mathcal{A}$ 中的元素 A，通过 $\hat{A}(\omega)=\omega(A)$ 它被重新解释为 $\mathcal{S}(\mathcal{A})$ 上的函数 $\hat{A}$。例如，一个对易的单位的 C^* 代数 $\mathcal{A}$ 的态空间 $\mathcal{S}(\mathcal{A})$ 是一个鲍厄单形，该鲍厄单形由前态空间 $\mathcal{P}(\mathcal{A})$ 上所有(正规波莱尔)的概率测度组成。

② 这是经典概率理论中德菲尼蒂(De Finetti)表征定理的一个量子对应[Heath and Sudderth(希思和苏德斯), 1976; van Fraassen, 1991]，也见[Hudson and Moody(哈德森和穆迪), 1975/76]和[Caves *et al.*(科夫斯等), 2002]。

若 $\omega_0^{(q)}$ 是置换不变的，则它是经典的。在$\mathcal{A}_0^{(c)}$上相关的极限态 $\omega_0^{(c)}$ 是用下述事实来说明的，即式(20)中的测量 μ_0 与式(22)中的测量 μ 一致。①

6.3 超选规则

无限量子系统常常与超选规则或分支(sector)概念联系在一起，这一概念是由威克、怀特曼和维格纳[1952]在希尔伯特空间$\mathcal{H}$上标准量子力学背景下引入的。其基本理念可以在玻色子/费米子或"单价化"(univalence)的超选规则案例中阐明。② 这里我们有在$\mathcal{H}$上旋转群 $SO(3)$的投影幺正表征$\mathcal{D}$，对围绕某轴的任意 2π 旋转 $R_{2\pi}$有$\mathcal{D}(R_{2\pi})=\pm 1$。特别地，在玻色子态 Ψ_B 上有$\mathcal{D}(R_{2\pi})\Psi_B=\Psi_B$，而在费米子态 Ψ_F 上规则为$\mathcal{D}(R_{2\pi})\Psi_F=-\Psi_F$。现在的论点是 2π 旋转并未完成什么，因而它不能够改变系统的物理态。这一要求显然在$\mathcal{H}$中玻色子态的子空间$\mathcal{H}_B\subset\mathcal{H}$上也成立，但它也同时满足于费米子态的子空间$\mathcal{H}_F\subset\mathcal{H}$，因为 Ψ 和在 $|z|=1$ 下的 $z\Psi$ 描述了相同的物理态。然而，若 $\Psi=c_B\Psi_B+c_F\Psi_F$(有 $|c_B|^2+|c_F|^2=1$)，则$\mathcal{D}(R_{2\pi})\Psi=c_B\Psi_B-c_F\Psi_F$，它并不与 Ψ 成比例，且明显地描述了一个真正不同于 Ψ 的物理态。

该问题解决的办法是通过宣称$\mathcal{D}(R_{2\pi})\Psi$ 和 Ψ 并不描述相同的物理态来否定这一结论，并且这是通过假定对于$(\Psi_B, A\Psi_F)\neq 0$ 不存在物理的可观测量 A (在他们通常的数学外形，即$\mathcal{H}$上的算符下)来实现的。因为在那种情形下，对任意可观测量 A，我们有：

$$(c_B\Psi_B\pm c_F\Psi_F, A(c_B\Psi_B\pm c_F\Psi_F))=|c_B|^2(\Psi_B, A\Psi_B)+|c_F|^2(\Psi_F, A\Psi_F) \tag{23}$$

从而对任意的 $\Psi\in\mathcal{H}$有$(\mathcal{D}(R_{2\pi})\Psi, A\,\mathcal{D}(R_{2\pi})\Psi)=(\Psi, A\Psi)$。因为任意量子力学的预言最终依赖于物理可观测量 A 的期望值$(\Psi, A\Psi)$，从而得到的结论

① 事实上，相应于表征$\otimes_{[\omega\in S^2]\pi}$(准等价类)的$\mathcal{S}(\mathcal{A}_0^{(q)})$中叶形线$\mathcal{F}^3$中的每个态都是经典的。

② 当代的数学处理见[Giulini(朱利尼)，2003]。

是 2π 旋转确实不改变系统。这被转述为认为 $c_B\Psi_B+c_F\Psi_F$ 类型的叠加是不相干的；而叠加 $c_1\Psi_1+c_2\Psi_2$ 是相干的，其中 Ψ_1 和 Ψ_2 同时要么在$\mathcal{H}_B$要么在$\mathcal{H}_F$中。$\mathcal{H}$的每个子空间$\mathcal{H}_B$和$\mathcal{H}_F$被认为是超选分支，且对于任意可观测量 A，$\Psi_B\in\mathcal{H}_B$和 $\Psi_F\in\mathcal{H}_F$，有$(\Psi_B,\ A\Psi_F)=0$ 的陈述被称为是超选规则。①

对于此解决办法我们付出的代价是，形如 $c_B\Psi_B+c_F\Psi_F$ 的态是混合态，其中 $c_B\neq0$ 和 $c_F\neq0$，正如人们从(23)中看到的。更一般地，每当 A 是一可观测量，$\Psi\in\mathcal{H}_\lambda$，$\Phi\in\mathcal{H}_{\lambda'}$，且 $\lambda\neq\lambda'$时，若在$(\Psi,\ A\Phi)\neq0$ 下有$\mathcal{H}=\oplus_{\lambda\in\Lambda}\mathcal{H}_\lambda$，且若此外对每个 λ 和每个 Ψ，$\Phi\in\mathcal{H}_\lambda$存在一个满足$(\Psi,\ A\Phi)\neq0$ 的可观测量 A，则子空间$\mathcal{H}_\lambda$被称为$\mathcal{H}$上的超选分支。另外，发生超选分支的一个主要影响是，在 $\Psi\in\mathcal{H}_\lambda$下(且对至少两个 λ 有 $c_\lambda\neq0$)，$\Psi=\sum_\lambda c_\lambda\Psi_\lambda$ 类型的单位向量定义了混合态：

$$\psi(A)=(\Psi,A\Psi)=\sum_\lambda|c_\lambda|^2(\Psi_\lambda,A\Psi_\lambda)=\sum_\lambda|c_\lambda|^2\psi_\lambda(A)$$

这一过程是相当有特设性的。哈格和其合作者们提出了超选理论的一种更深刻方法，关于该方法的介绍见[Roberts and Roepstorff，1969]。这里的出发点是给定量子系统可观测量$\mathcal{A}$的抽象 C^* 代数，同时超选分支被重新解释为是$\mathcal{A}$(满足特定的选择标准，见下文)不可约表征的等价类(在幺正同构下)。在上面讨论的超选分支的具体希尔伯特空间方法与抽象的 C^* 代数方法之间的联系由下述引理给出[Hepp，1972]：②

> 两个 C^* 代数$\mathcal{A}$上的纯态 ρ 和 σ 定义了不同的分支，当且仅当对包含单位向量 Ψ_ρ，Ψ_σ 的希尔伯特空间$\mathcal{H}$上的每个表征 $\pi(\mathcal{A})$ 而言，满足对所有的 $A\in\mathcal{A}$有 $\rho(A)=(\Psi_\rho,\ \pi(A)\Psi_\rho)$和 $\sigma(A)=(\Psi_\sigma,\ \pi(A)\Psi_\sigma)$，则我们得到对所有的 $A\in\mathcal{A}$有$(\Psi_\rho,\ \pi(A)\Psi_\sigma)=0$。

---

① 在 Ψ 和 Φ 之间一个普通的选择规则中，对于哈密顿量 H，我们仅有$(\Psi,\ H\Phi)=0$。

② 亨普证明了此引理的一个更普遍版本，在该版本中"两个 C^* 代数$\mathcal{B}$上的纯态 ρ 和 σ 定义了不同的分支，当且仅当……"由"两个 C^* 代数$\mathcal{B}$上的纯态 ρ，σ 是不相交的，当且仅当……"所取代。

然而，实际上物理学中用到的绝大多数典型的 C^* 代数$\mathcal{A}$的不可约表征在物理上与数学构造无关。这样的表征可以通过某种选择标准排除出考虑范围。此选择标准是什么这一点则依赖于语境。例如，在量子场论中此概念在所谓的 DHR 理论中变得精确，由罗伯茨[1990]、哈格[1992]、荒木[1999]和霍尔沃森[2005]所评述。在前两节讨论过的理论种类中，我们取可观测量$\mathcal{A}$的代数为$\mathcal{A}_0^{(q)}$——本质上是由于人类的局限性——且出于教学法的原因将$\mathcal{A}_0^{(q)}$的不可约表征(等价种类)定义为超选分支，今后仅当它们等价于由在$\mathcal{A}_0^{(q)}$上一置换不变的纯态给出的 GNS 表征时常常只称作分支。特别地，这样的态是经典的。在此选择标准中，前一小节中的结果一般意味着在$\mathcal{A}_1$上的纯态和$\mathcal{A}_0^{(q)}$的分支之间存在着双射对应。对易 C^* 代数$\mathcal{A}_0^{(c)}$的分支仅仅是$\mathcal{S}(\mathcal{A}_1)$上的点，注意在$\mathcal{A}_1$上的一个混合态定义了$\mathcal{A}_0^{(c)}$上的一个纯态！关于$\mathcal{A}_1$的分支在与$\mathcal{A}_0^{(c)}$中那些分支相关联中的角色将在 6.5 节中澄清。

不论什么样的模型和选择标准，完全从在可观测量$\mathcal{A}$代数上纯态的角度来考虑超选分支，而忽略$\mathcal{A}$本身及其表征是有启发作用的(且在某种程度上与实验实践相一致)。要这样做，我们用跃迁概率空间结构来装备$\mathcal{A}$上纯态的空间$\mathcal{P}(\mathcal{A})$[von Neumann，1981；Mielnik，1968]。① 在集合$\mathcal{P}$上的跃迁概率是一个函数：

$$p:\ \mathcal{P}\times\mathcal{P}\to[0,\ 1] \tag{24}$$

它满足：

$$p(\rho,\ \sigma)=1\Leftrightarrow\rho=\sigma \tag{25}$$

和

$$p(\rho,\ \sigma)=0\Leftrightarrow p(\sigma,\ \rho)=0 \tag{26}$$

拥有这样一个跃迁概率的集合被称为跃迁概率空间。现在，若我们定义：②

$$p(\rho,\ \sigma):=\inf\{\rho(A)\mid A\in\mathcal{A},\ 0\leqslant A\leqslant 1,\ \sigma(A)=1\} \tag{27}$$

---

① 简明评论也见[Beltrametti and Cassinelli，1984]或[Landsman，1998]。

② 这一定义适用于$\mathcal{A}$是有单位元的情形，一般情形见[Landsman，1998]。类似的公式在任意紧致凸集的极边界上定义了一个跃迁概率。

则 C^* 代数$\mathcal{A}$的纯态空间$\mathcal{P}(\mathcal{A})$恰好具有此结构。为了给出更精巧的公式，注意因为纯态是原初的，两个纯态ρ，σ 要么是不相交的($\rho\perp\sigma$)，要么是(准，因此幺正地)等价的($\rho\sim\sigma$)。在第一种情形中，式(27)导致：

$$p(\rho,\ \sigma)=0 \quad (\rho\perp\sigma) \tag{28}$$

在第二种情形中根据卡迪孙跃迁定理[Kadison and Ringrose，1986]得到来自由 ρ 定义的 GNS 表征 $\pi_\rho(\mathcal{A})$的希尔伯特空间$\mathcal{H}_\rho$包含了一单位向量 Ω_σ(在相位上是唯一的)，满足：

$$\sigma(A)=(\Omega_\sigma,\ \pi_\rho(A)\Omega_\sigma) \tag{29}$$

如此一来，方程(27)产生了广为人知的表达式：

$$p(\rho,\ \sigma)=|\ \Omega_\rho,\ \Omega_\sigma|^2 \quad (\rho\sim\sigma) \tag{30}$$

特别地，若$\mathcal{A}$是对易的，则：

$$p(\rho,\ \sigma)=\delta_{\rho\sigma} \tag{31}$$

对于$\mathcal{A}=M_2(\mathbb{C})$，我们得到：

$$p(\rho,\ \sigma)=\frac{1}{2}(1+\cos\theta_{\rho\sigma}) \tag{32}$$

其中 $\theta_{\rho\sigma}$是在 ρ 与 σ——被看作是双球面 $S^2=\partial_e B^3$ 上的点，参见(3)等——之间的角距离，它是沿着一个大圆来测量的。

现在可以对任意跃迁概率空间$\mathcal{P}$来定义超选分支了。若每当 ρ 与 σ 不处于相同的子集中时有 $p(\rho,\ \sigma)=0$，则$\mathcal{P}$的子集族被称为是正交的。空间$\mathcal{P}$被称为可约的，如果它是两个(非空)正交子集的并集；如果不是这样的话，它被称为不可约的。$\mathcal{P}$的一个分支是一个子集$\mathcal{C}\subset\mathcal{P}$，满足$\mathcal{C}$与$\mathcal{P}\setminus\mathcal{C}$正交。$\mathcal{P}$的一个不可约分支是被称为一个(超选)分支。因此$\mathcal{P}$是其分支的不相交并集。对于$\mathcal{P}=\mathcal{P}(\mathcal{A})$，则通过在态与由 GNS 构造给出的表征之间的对应重塑了超选分支的代数定义(对选择标准模)。例如，在对易情形$\mathcal{A}\cong C(X)$中，$X\cong\mathcal{P}(\mathcal{A})$中的每个点是其自身的小分支。

6.4　简单案例：无限自旋链

我们将在一个简单的案例中展示超选分支的发生，在该案例中可观测量代数是$\mathcal{A}_0^{(q)}$，有$\mathcal{A}_1=M_2(\mathbb{C})$。令$\mathcal{H}_1=\mathbb{C}^2$，从而$\mathcal{H}_1^N=\otimes^N\mathbb{C}^2$是$\mathbb{C}^2$的 N 个拷贝的

张量积。显然$\mathcal{A}_1^N$以一种自然的方式（即类分支）作用于$\mathcal{H}_1^N$。这定义了$\mathcal{A}_1^N$上的一个不可约表征π_N，它事实上是其唯一的不可约表征（取决于幺正等价性）。特别地，对于$N<\infty$，其可观测量代数为$\mathcal{A}_1^N$的量子系统（如有N个二阶系统的链）没有超选规则。我们将C^*代数$(M_2(\mathbb{C}))^N$在$N\to\infty$时的极限“$(M_2(\mathbb{C}))^\infty$”定义为对于$\mathcal{A}_1=M_2(\mathbb{C})$的归纳极限$\mathcal{A}_0^{(q)}$，正如6.2小节中介绍的那样，见(13)。“$\otimes^\infty\mathbb{C}^2$”的定义涉及的内容有些多，见[von Neumann, 1938]。

对任意希尔伯特空间$\mathcal{H}_1$，令Ψ是一列$(\Psi_1, \Psi_2, \cdots)$，有$\Psi_n\in\mathcal{H}_1$。此列的空间$\mathbf{H}_1$显然是一个向量空间。现在令Ψ和Φ是两个这样的列，且写作$(\Psi_n, \Phi_n)=\exp(i\alpha_n)|(\Psi_n, \Phi_n)|$。若$\sum_n|\alpha_n|=\infty$，我们定义$(\Psi, \Phi)$的（前）内积为零。若$\sum_n|\alpha_n|<\infty$，我们给出$(\Psi, \Phi)=\Pi_n(\Psi_n, \Phi_n)$（当然它可能仍然为零）。$\mathbf{H}_1$的（向量空间）商通过有$(\Psi, \Psi)=0$的列$\Psi$空间，可以在约化内积中完备到一个希尔伯特空间$\mathcal{H}_1^\infty$，称为希尔伯特空间$\mathcal{H}_1$的完备无限张量积（在指数集合$\mathbb{N}$上）。① 我们将此构造应用于$\mathcal{H}_1=\mathbb{C}^2$。若$(e_i)$是$\mathbb{C}^2$的某个基，则$\mathcal{H}_1^\infty$的一个正交归一化基包括了所有不同的无限行（strings）$e_{i_1}\otimes, \cdots, e_{i_n}\otimes\cdots$，其中$e_{i_n}$被看作是$\mathbb{C}^2$中的一个向量$e_i$。② 我们将多指数$(i_1, \cdots, i_n, \cdots)$简单地记作$I$，相应的基矢记作$e_I$。

此希尔伯特空间$\mathcal{H}_1^\infty$携带了对$\mathcal{A}_0^{(q)}$自然忠实的表征π：若$A_0\in\mathcal{A}_0^{(q)}$是一个等价类$[A_1, A_2, \cdots]$，则有$\pi(A_0)e_I=\lim_{N\to\infty}A_N e_I$，其中$A_N$作用于$e_I$的前$N$个分支，同时保留其他分支不变。③ 现在关键点在于，虽然每个$\mathcal{A}_1^N$不可约地作用于$\mathcal{H}_1^N$，但因此构造的$\mathcal{H}_1^\infty$上的表征$\pi(\mathcal{A}_0^{(q)})$是高度不可约的。此中的缘由在于，

---

① 每个固定的$\Psi\in\mathcal{H}_1$定义一个非完备张量积$\mathcal{H}_\Psi^\infty$，将其定义为由所有Φ组成的$\mathcal{H}_1^\infty$的闭子空间，对Φ而言有$\sum_n|(\Psi_n,\Phi_n)-1|<\infty$。若$\mathcal{H}_1$是可分离的，则$\mathcal{H}_\Psi^\infty$也是可分离的（与$\mathcal{H}_1^\infty$相反，它是$\mathcal{H}_\Psi^\infty$的一个不可数直和）。

② 所有这些行的集合的基数等于$\mathbb{R}$的基数，因而是不可分离的，如所述的那样。

③ 事实上，这产生了一种将$\overline{\cup_{N\in\mathbb{N}}\mathcal{A}_1^N}$定义为以所述方式作用于$\mathcal{H}_1^\infty$的所有$\mathcal{A}_1^N$并集的范数闭包的替代性方案。

根据定义$\mathcal{A}_0^{(q)}$的(准)定域元素(几乎)不改变在$\mathcal{H}_1^\infty$中的向量的无限尾部，因而具有不同尾部的向量处于不同的超选分支中。在没有关于$\mathcal{A}_0^{(q)}$上元素的准定域条件时，不会产生超选规则。例如，采用$\mathbb{C}^2$普通基矢术语：

$$\left\{\uparrow=\begin{pmatrix}1\\0\end{pmatrix},\ \downarrow=\begin{pmatrix}0\\1\end{pmatrix}\right\} \tag{33}$$

向量$\Psi_\uparrow=\uparrow\otimes\uparrow\cdots\uparrow\cdots$(即"上"向量的无限积)和$\Psi_\downarrow=\downarrow\otimes\downarrow\cdots\downarrow\cdots$(即"下"向量的无限积)位于不同的分支中。对任意的$A\in\mathcal{A}_0^{(q)}$，内积$(\Psi_\uparrow,\ \pi(A)\Psi_\downarrow)$为零的原因在于，对定域可观测量$A$，我们有$\pi(A)=A_M\otimes 1\otimes\cdots 1\cdots$，对于某个$A_M\in\mathcal{B}(\mathcal{H}_M)$，因而所考虑的内积涉及了无限多的因子$(\uparrow,\ 1\downarrow)=(\uparrow,\ \downarrow)=0$。对于准定域的$A$，算符$\pi(A)$可能有一个小的非平凡尾部，但根据近似论证内积仍然为零。

更一般地，初步的分析表明，每当对于单位向量u，$v\in\mathbb{C}^2$，其中$u\neq v$，当$\Psi_u=\otimes^\infty u$和$\Psi_v=\otimes^\infty v$时，$(\Psi_u,\ \pi(A)\Psi_v)=0$。在$\mathcal{A}_0^{(q)}$上相应的向量态$\psi_u$和$\psi_v$，即$\psi_u(A)=(\Psi_u,\ \pi(A)\Psi_u)$等，显然是置换不变的，且因而是经典的。如在(3)中那样将$\mathcal{S}(M_2(\mathbb{C}))$看作$B^3$，由$\psi_u$定义的$\mathcal{A}_0^{(c)}$上相应的极限态$(\psi_u)_0$由$\partial_e B^3=S^2$(即双球面)的(赋值于)点$\tilde{u}=(x,\ y,\ z)$给出，对此点而言相应的密度矩阵$\rho(\tilde{u})$是到$u$的投影算符。其结果是，$\psi_u$和$\psi_v$不相交，参见式(19)。我们得到结论，每个单位向量$u\in\mathbb{C}^2$确定了一个超选分支π_u，即相应态ψ_u的GNS表征，且每一个这样的分支都实现为$\mathcal{H}_1^\infty$上的子空间$\mathcal{H}_u$。此外，因为在$\mathcal{A}_0^{(q)}$上的一个置换不变态当且仅当其具有形式ψ_u时是纯的，我们找到了系统中所有的超选分支。因此接下来我们将关注于($\mathcal{H}_1^\infty$的)子空间和(π的)子表征：

$$\mathcal{H}_{\mathfrak{s}}=\oplus_{\tilde{u}\in S^2}\mathcal{H}_u \tag{34}$$

$$\pi_{\mathfrak{s}}(\mathcal{A}_0^{(q)})=\oplus_{\tilde{u}\in S^2}\pi_u(\mathcal{A}_0^{(q)})$$

其中π_u只是π对$\mathcal{H}_u\subset\mathcal{H}_1^\infty$的限制。

在出现超选分支时，我们可以构造用于区分不同分支的算符，同时该算符也是每个分支中单元的倍数。在量子场论中该算符通常是全域电荷，在我们的案例中则是宏观可观测量。为明白这一点，我们回到6.1小节。不难表明对任意近似对称列$(A_1,\ A_2,\ \cdots)$，极限为：

$$\overline{A}=\lim_{N\to\infty}\pi_{\Xi}(A_N) \tag{35}$$

在$\mathcal{B}(\mathcal{H}_{\Xi})$上的强算符拓扑中存在[Bona，1988]。此外，若$A_0\in\mathcal{A}_0^{(c)}=C(\mathcal{S}(\mathcal{A}_1))$是由给定列定义的函数，① 则映射$A_0\mapsto\overline{A}$在$\mathcal{H}_{\Xi}$上定义了一个$\mathcal{A}_0^{(c)}$的忠实表征，我们也称其为$\pi_{\Xi}$（记号通用）。简单的计算实际上表明对$\Psi\in\mathcal{H}_u$有$\pi_{\Xi}(A_0)\Psi=A_0(\hat{u})\Psi$，或换言之有：

$$\pi_{\Xi}(A_0)=\oplus_{\hat{u}\in S^2}A_0(\hat{u})1_{\mathcal{H}_u} \tag{36}$$

因此$\pi_{\Xi}(A_0)$确实是这里所谈论的算符。

为了表明这有多精妙，值得注意的是即使对于对称列而言极限$\lim_{N\to\infty}\pi(A_N)$在$\mathcal{H}_1^{\infty}$上并不存在，② 更不用说在强拓扑中了。在积极的一方面，可以表明$\lim_{N\to\infty}\pi(A_N)\Psi$作为冯·诺伊曼代数$\pi(\mathcal{A}_0^{(q)})''$的一个元素而存在，每当由$\Psi$定义的向量态$\psi$位于由所有置换不变态产生的叶形线$\mathcal{F}^{\Xi}$中时[Bona，1988；Unnerstall，1990a]。

得到的这一发现是在冯·诺伊曼代数背景下宏观可观测量一般理论的一部分[Primas，1983；Rieckers（瑞克斯），1984；Amann，1986；1987；Morchio and Strocchi，1987；Bona，1988；1989；Unnerstall，1990a；1990b；Breuer（布洛伊尔），1994；Atmanspacher *et al.*，1999]，它是对上面解释过的雷格和维纳[1989；1991]，达菲尔德和维纳[1992a，b，c]和达菲尔德等[1992]的纯C^*代数方法的补充。③ 在我们看来，后者的优势在于：它在概念上对极限$N\to\infty$（且

---

① 回顾$A_0(\omega)=\lim_{N\to\infty}\omega^N(A_N)$。

② 例如，我们取列$A_N=j_{N1}(\mathrm{diag}(1,-1))$和向量$\Psi=\uparrow\downarrow\downarrow\uparrow\uparrow\uparrow\uparrow\downarrow\downarrow\downarrow\downarrow\downarrow\downarrow\downarrow\downarrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\uparrow\cdots$，其中↑的$2^N$个因子的序列之后是↓的$2^{N+1}$个因子，依此下去。从而$\mathcal{H}_1^{\infty}$中的列$\{\pi(A_N)\Psi\}_{N\in\mathbb{N}}$发散：其中$N$涵盖所有$2^n$且$n$为奇数的子列收敛于$\frac{1}{3}\Psi$，而其中$N$涵盖所有$2^n$且$n$为偶数的子列收敛于$-\frac{1}{3}\Psi$。

③ 瑞克斯和他的合作者们在C^*代数和冯·诺伊曼代数方法情境中研究了真实的模型。参见[Honegger and Rieckers（奥涅格和瑞克斯），1994；Gerisch *et al.*（格瑞琪等），1999；Gerisch *et al.*，2003]和许多其他论文。关于宏观可观测量全部不同的方法，见[van Kampen（范·坎朋），1954；1988；1993；Wan and Fountain（万与方丹），1998；Harrison and Wan（哈里森和万），1997；Wan *et al.*，1998；Fröhlich *et al.*（弗罗利希等），2002]和[Poulin（波林），2004]。

由此将一个大系统理想化为无限的系统)的过渡是非常令人满意的，尤其是当我们用 C^* 代数连续场来重构时。这里无限系统宏观可观测量的对易的 C^* 代数 $\mathcal{A}_0^{(c)}$ 被以一种连续的方式与相应有限系统的非对易代数 $\mathcal{A}_1^N$ 黏合在一起，且继而产生的 C^* 代数 $\mathcal{A}^{(c)}$ 连续场的连续截面恰好描述了有限系统的宏观量子可观测量是如何收敛于经典的可观测量的。另一方面，相关的有限系统的微观量子可观测量收敛于无限量子系统的量子可观测量，且此收敛是通过 C^* 代数 $\mathcal{A}^{(q)}$ 连续场的连续截面来描述的，这完全避免了超选规则的表达方式，相反后者展示了一种在有限与无限系统间令人震惊的不连续性：因为超选规则在有限系统中不存在！①

6.5　泊松结构和动力学

我们现在转而讨论目前为止考虑的这类无限系统的时间演化。我们开始于下述发现，即一个有限维 C^* 代数 $\mathcal{B}$(接下来出于简单性假定它是单元的)的态空间 $\mathcal{S}(\mathcal{B})$ 是自然状态中的泊松流形(参见 4.3 小节)。类似的陈述在无限维情形中也成立，必要的主要论证读者可参见脚注。② 我们写为 $K=\mathcal{S}(\mathcal{B})$。

首先，元素 $A\in\mathcal{B}$ 定义了一个在 $\mathcal{B}^*$ 上的线性函数 $\hat{A}$，并通过 $\hat{A}(\omega)=\omega(A)$ 定义在 K 上(也就是通过限制)。对于这样的函数我们定义泊松括号为：

---

① 我们这里指的是在传统意义上与单 C^* 代数不可约表征不等价的超选规则。由于拓扑的原因，一些有限维系统是用允许不等价不可约表征的非单(non-simple) C^* 代数来描述的[Landsman，1990a，b]。

② 这是那些脚注中的第一个。当 $\mathcal{B}$ 是无限维时，态空间 $\mathcal{S}(\mathcal{B})$ 不再是一个流形，更不用说是一个泊松流形了，但它是一个泊松空间[Landsman，1997；1998]。这是对泊松流形的推广，该推广把泊松流形的关键特性变为了定义。这个特征是泊松流形经过其辛叶子的叶状结构[Weinstein，1983]，且相应的定义如下：一个泊松空间 P 是一个具有形式 $P=\cup_\alpha S_\alpha$ 的豪斯道夫空间(不相交的并)，其中每个 S_α 是一个辛流形(可能是无限维的)且每个入射 ι_α: $S_\alpha\to P$ 是连续的。此外，我们有一个分离点的线性子空间 $F\subset C(P,\mathbb{R})$，它具有这样的性质，即每个 $f\in F$ 对每个 S_α 的限制是光滑的。最后，如果 f, $g\in F$，则 $\{f,g\}\in F$，其中泊松括号是由 $\{f,g\}(\iota_\alpha(\sigma))=\{\iota_\alpha^*f,\iota_\alpha^*g\}_\alpha(\sigma)$ 定义的。显然，一个泊松流形 M 定义了一个泊松空间，如果取 $P=M$，$F=C^\infty(M)$，且 S_α 是由给定的泊松括号所定义的辛叶子。因此，在上面的定义中我们指的是流形是 S_α，还有 P 的辛叶子。

$$\{\hat{A},\ \hat{B}\} = i\,\widehat{[A,\ B]} \tag{37}$$

这里引入因子 i 是为了使两个实值函数的泊松括号保持实值，因为正好在 A 自伴时 $\hat{A}$ 是 K 上的实值，且若 $A^* = A$ 和 $B^* = B$，则 $i[A,\ B]$是自伴的，而$[A,\ B]$是反自伴的(skew-adjoint)。一般地，对$f,\ g \in C^\infty(K)$，我们将：

$$\{f,\ g\}(\omega) = i\omega([df_\omega,\ dg_\omega]) \tag{38}$$

解释如下。① 令$\mathcal{B}_{\mathbb{R}}$是$\mathcal{B}$的自伴部分，将 K 解释为是$\mathcal{B}_{\mathbb{R}}^*$的子空间。因为态 ω 对所有 $A \in \mathcal{B}$满足 $\omega(A^*) = \overline{\omega(A)}$，它通过其在自伴元素上的值来得到确定。之后，我们把在 ω 处的切空间等同于：

$$T_\omega K = \{\rho \in \mathcal{B}_{\mathbb{R}}^* \mid \rho(1) = 0\} \subset \mathcal{B}_{\mathbb{R}}^* \tag{39}$$

在 ω 处拥有(实巴拿赫空间)商的余切空间为：

$$T_\omega^* K = \mathcal{B}_{\mathbb{R}}^{**}/\mathbb{R}\,1 \tag{40}$$

其中单位元 $1 \in \mathcal{B}$被看作是通过正则嵌入$\mathcal{B} \subset \mathcal{B}^{**}$的$\mathcal{B}^{**}$的一个元素。因此，$\omega \in K$ 处的微分形式 df 和 dg 定义了$\mathcal{B}_{\mathbb{R}}^{**}/\mathbb{R}\,1$ 的元素。式(38)中的对易子因而定义如下：我们将 $df_\omega \in \mathcal{B}_{\mathbb{R}}^{**}/\mathbb{R}\,1$ 提升(lift)到$\mathcal{B}^{**}$，并且采用有限维向量空间中典型的自然同构$\mathcal{B}^{**} \cong \mathcal{B}$。② 这个提升中的任意性是 1 的倍数，它退出了对易子。因而 $i[df_\omega,\ dg_\omega]$是$\mathcal{B}_{\mathbb{R}}^{**} \cong \mathcal{B}_{\mathbb{R}}$ 的一个元素，函数 ω 的值在其上被确定。③ 这样就完成了对泊松括号的定义，人们很容易重新得到作为式(38)的特殊情形的式(37)。

达菲尔德和维纳[1992a]确定了 K 上给定泊松结构的辛叶子。④ 即：

> 当且仅当对某个幺正 $U \in \mathcal{B}$有 $\rho(A) = \sigma(UAU^*)$时，两个态 ρ 和 σ 位于$\mathcal{S}(\mathcal{B})$同一个辛叶子上。

---

① 在无限维情形中 $C^\infty(K)$被定义为是 K 上光滑函数的交集，相对于其巴拿赫流形结构和 K 上的弱*连续函数空间 $C(K)$。(38)中的微分形式 df 和 dg 也要求一个恰当的定义，技术细节见[Duffield and Werner，1992a；Bona，2000]和[Odzijewicz and Ratiu，2003]。

② 在无限维情形中我们用在$\mathcal{B}^{**}$与$\mathcal{B}$的包络冯·诺伊曼代数之间的正则同化(identification)来定义对易子。

③ 若$\mathcal{B}$是无限维的，这里我们将$\mathcal{B}^*$看作是冯·诺伊曼代数$\mathcal{B}^{**}$的准对偶(predual)。

④ 无限维特殊情形也见[Bona，2000]，在那里$\mathcal{B}$是紧致算符的 C^*代数。

当ρ和σ是纯态时，这是当且仅当相应的 GNS 表征$\pi_\rho(\mathcal{B})$和$\pi_\sigma(\mathcal{B})$是幺正等价时的情形，① 但一般地该关系仅在一个方向上成立：若ρ和σ处于同一个叶子，则它们拥有幺正等价 GNS 表征。②

正是根据$K=\mathcal{S}(\mathcal{B})$辛叶子的这一特征得到了纯态空间$\partial_e K=\mathcal{P}(\mathcal{B})$从$K$中继承了泊松括号，因而其本身变成了一个泊松流形。③ 这给出了在$\mathcal{B}$的超选分支与$\mathcal{P}(\mathcal{B})$上的泊松结构之间的重要关联[Landsman，1997；1998]：

> C^*代数$\mathcal{B}$的纯态空间$\mathcal{P}(\mathcal{B})$的分支作为一个跃迁概率空间与其作为泊松流形的辛叶子一致。

例如，当$\mathcal{B}\cong C(X)$是对易的时，在X上的所有(正规波莱尔)概率测量的空间$\mathcal{S}(C(X))$得到了一个等同于零的泊松括号，正如其极限边界X那样。根据式(31)可如X中的分支是它的点，从而也是其辛叶子(从它们的定义和为零的泊松括号角度看来)。最简单的非对易情形是$\mathcal{B}=M_2(\mathbb{C})$，对此态空间$K=\mathcal{S}(M_2(\mathbb{C}))\cong B^3$的辛叶子是拥有常数半径的球。④ 半径为 1 的球由对应于

---

① 参见[Kadison and Ringrose，1986]中的定理 10.2.6。

② 论证的关键一步是注意到，$f\in C^\infty(K)$的哈密顿向量场$\xi_f(\omega)\in T_\omega K\subset\mathcal{A}_{\mathbb{R}}^*$是由$\langle\xi_f(\omega), B\rangle=i[df_\omega, B]$给出的，其中$B\in\mathcal{B}_{\mathbb{R}}\subset\mathcal{B}_{\mathbb{R}}^{**}$和$df_\omega\in\mathcal{B}_{\mathbb{R}}^{**}/\mathbb{R}1$。例如，这通过式(37)给出了$\xi_{\hat{A}}\hat{B}=i\,\widehat{[A, B]}=\{\hat{A}, \hat{B}\}$，正如它所应当的那样。若$\varphi_t^h$表示在$t$时刻$h$的哈密顿流，其结果，参见[Duffield *et al.*，1992，Prop. 6.1]或[Duffield and Werner，1992a，Prop. 3.1]，是对某个幺正$U_t^h\in\mathcal{B}$有$\langle\varphi_h^t(\omega), B\rangle=\langle\omega, U_t^h B(U_t^h)^*\rangle$。例如，若$h=\hat{A}$则$U_t^h=\exp(itA)$。

③ 更一般地，是一个泊松空间。$\mathcal{P}(\mathcal{B})$的结构作为一泊松空间是由兰兹曼[1997；1998]在不借助于完全态空间或达菲尔德和维纳[1992a]的工作的情形下引入的。

④ 用所谓富比尼—施图迪(Fubini-Study)辛结构的倍数来装备，关于此概念见[Landsman，1998]或微分几何的任意合适书目。更一般地，$M_n(\mathbb{C})$的纯态空间是投影空间$\mathbb{P}\mathbb{C}^n$，它仍然是用富比尼—施图迪辛结构来装备的。若人们将$M_\infty(\mathbb{C})$定义为是在可分离的希尔伯特空间$\mathcal{H}$上的紧致算符的C^*代数，这一论述甚至在$n=\infty$时也成立，那样的话我们有$\mathcal{P}(M_\infty(\mathbb{C}))\cong\mathbb{P}\mathcal{H}$。参见[Cantoni，1977；Cirelli *et al.*(萨拉里等)，1983；Cirelli *et al.*，1990；Landsman，1998；Ashtekar and Schilling(阿西提卡和希林)，1999；Marmot *et al.*(马尔莫等)，2005]等文献。

$M_2(\mathbb{C})$上纯态的B^3中的点组成，所有K的内部辛叶子都来自于$M_2(\mathbb{C})$上的混合态。

$\mathcal{P}(\mathcal{B})$的分支与辛叶子间的重合是在跃迁概率结构与泊松结构间的相容性条件。它分别是具体选择(27)和(38)的典型结果，因而也是量子理论的典型结果。在经典力学中我们可以自由地用任一泊松结构装备一个流形M，还可以将$C_0(M)$用作是可观测量的对易C^*代数。跃迁概率(31)，它来自于对易情形中的式(27)，显然是经典物理学中恰当的那些，但因为M的辛叶子几乎可以是任意的，刚才谈论到的一致性并不成立。

然而，在跃迁概率结构与泊松结构之间存在一个相容性条件，它是经典与量子理论共有的。这是哈密顿流幺正性的特征，在当前背景下我们如下来构造此特征。① 首先，我们在有代数可观测量$\mathcal{B}$的量子理论中定义时间演化(在阿贝尔群$\mathbb{R}$对$\mathcal{B}$的自同构作用，也就是自同构的单参数群α对$\mathcal{B}$作用的意义上)是哈密顿的，当$A(t)=\alpha_t(A)$对某个自伴元素$H\in\mathcal{B}$满足海森堡方程$i\hbar dA/dt=[A,H]$时。在该情形下$\mathcal{P}(\mathcal{B})$上的相应流——即$\omega_t(A)=\omega(A(t))$——在那种情形中也等于说是哈密顿的。在具有泊松流形M的经典力学中，我们类似地说M上的流是哈密顿的，当它对某个$h\in C^\infty(M)$而言是哈密顿向量场ξ_h的流时。等价地，可观测量$f\in C^\infty(M)$的时间演化由$df/dt=\{h, f\}$给出，参见式(18)等。关键是，在两种情形中流在下述意义上是幺正的，即：

$$p(\rho(t), \sigma(t))=p(\rho, \sigma) \tag{41}$$

对所有的t和所有的ρ，$\sigma\in P$有$P=\mathcal{P}(\mathcal{B})$——具有跃迁概率(27)和泊松括号(38)，或者$P=M$——具有跃迁概率(31)和任意的泊松括号。②

在$P=\mathcal{P}(\mathcal{B})$和$P=M$这两种情形中，哈密顿流拥有下述特征(这直接来自于辛叶子的定义)：对所有(有限的)时间t，点$\omega(t)$处于与$\omega=\omega(0)$相同的P的辛叶子上。特别地，在量子理论中，$\omega(t)$和ω必须处于相同的分支中。在无

① 所有这些可以提升为一个在其中经典与量子理论都适用的公理化结构，见[Landsman, 1997; 1998]。

② 在量子理论中流是对任意t定义的。在经典动力学中，式(41)对所有t成立，$\rho(t)$和$\sigma(t)$是为式(41)而定义的。

限系统的量子理论中，自同构的时间演化很少是哈密顿的，我们是在一个更弱的假设下得到类似结论的。也就是，对于某些 $\omega \in \mathcal{P}(\mathcal{B})$ 而言，若$\mathcal{B}$上给定的自同构的参数群 α 是在 GNS 表征 $\pi_\omega(\mathcal{B})$ 中生效的，① 则 $\omega(t)$ 和 ω 处于相同的分支中，且因此在$\mathcal{P}(\mathcal{B})$的相同辛叶子中。

为了阐明这些概念，我们回到 C^* 代数$\mathcal{A}^{(c)}$的连续场，参见式(2)。在每个纤维代数$\mathcal{A}^{(c)}_{1/N}$的纯态空间上的正则 C^* 代数的跃迁概率(27)，对 $N<\infty$ 而言收敛于在对易极限代数$\mathcal{A}^{(c)}_0$上的经典跃迁概率(31)，这是不足为奇的。类似地，在每个$\mathcal{P}(\mathcal{A}^{(c)}_{1/N})$上的 C^* 代数泊松结构(38)收敛于零。然而，我们根据量子力学的极限$\hbar\to 0$ 知道，在 C^* 代数连续场的极限代数上产生经典行为的过程中，人们应当重新调整对易子，见 4.3 节和第 5 部分。因而我们用下式取代$\mathcal{A}^{(c)}_{1/N}$的泊松括号(38)：

$$\{f,\ g\}(\omega)=iN\omega([df_\omega,\ dg_\omega]) \tag{42}$$

因而，空间$\mathcal{P}(\mathcal{A}^{(c)}_{1/N})$上重新调整过的泊松括号被证明收敛于在$\mathcal{P}(\mathcal{A}^{(c)}_0)=\mathcal{S}(\mathcal{A}_1)$上的正则泊松括号(38)，而不是从极限代数$\mathcal{A}^{(c)}_0$的对易本性中期望得到的零括号。因此，纤维代数$\mathcal{A}^{(c)}_1$的完全态空间$\mathcal{S}(\mathcal{A}_1)$的辛叶子变成了纤维代数$\mathcal{A}^{(c)}_0$的纯态空间$\mathcal{S}(\mathcal{A}_1)$的辛叶子。这毫无疑问是对经典相空间及其泊松结构在量子理论中起源的揭示。

更精确地，我们拥有如下结果[Duffield and Werner, 1992a]：

> 若 $A(A_0,\ A_1,\ A_2,\ \cdots)$和 $A'=(A'_0,\ A'_1,\ A'_2,\ \cdots)$是由对称列定义的$\mathcal{A}^{(c)}$的连续截面，② 则列

① 此假设意味着在$\mathcal{H}_\omega$上存在$\mathbb{R}$ 的幺正表征 $t\mapsto U_t$，使得对所有 $A\in\mathcal{B}$和所有 $t\in\mathbb{R}$ 有 $\pi_\omega(\alpha_t(A))=U_t\pi_\omega(A)U_t^*$。

② 该结果并不对所有连续截面(即对所有近似对称列)成立，因为极限函数 A_0 和 A'_0 可能是不可微的，因而它们的泊松括号并不存在。这一问题发生在所有形变量子化案例中。然而，使该陈述成立的列的类要远大于只是对称的列的类。对于 A 和 B 使式(43)有意义的充分条件是 $A_N=\sum_{M\leq N} j_{NM}(A_M^{(N)})$ (有 $A_M^{(N)}\in\mathcal{A}_1^M$)，使得 $\lim_{N\to\infty}A_M^{(N)}$ (在模中)存在，且有 $\sum_{M=1}^{\infty} M\sup_{N\geq M}\{\|A_M^{(N)}\|\}<\infty$，见[Duffield and Werner, 1992a]。

$$(\{A_0, A_0'\}, i[A_1, A_1'], \cdots, iN[A_N, A_N'], \cdots) \tag{43}$$

定义了$\mathcal{A}^{(c)}$的一个连续截面。

这得自于一个简单的计算。换言之，虽然对易子$[A_N, A_N']$列收敛于零，但重新调整的对易子 $iN[A_N, A_N'] \in \mathcal{A}_N$收敛于宏观可观测量$\{A_0, A'_0\} \in \mathcal{A}_0^{(c)} = C(\mathcal{S}(\mathcal{A}_1))$。虽然在一个大量子系统的经典极限上采用量子化术语来重新解释这一结果显得有悖常理（这与取经典极限相反），但这么做在形式上是可能的（参见4.3小节）。如果我们取：

$$\hbar = \frac{1}{N} \tag{44}$$

通过在必要时利用选择公理，我们设计了一个赋予场的连续截面$A = (A_0, A_1, A_2, \cdots)$以一给定函数$A_0 \in \mathcal{A}_0^{(c)}$的过程。我们将此写作$A_N = \mathcal{Q}_\hbar(A_0)$，且类似地$A_N' = \mathcal{Q}_\hbar(A_0')$。这一选择不需要使得列(43)被赋予于$\{A_0, A_0'\}$，但因为后者是式(43)的唯一极限，它必须是：

$$\lim_{N\to\infty} \| iN[\mathcal{Q}_\hbar(A_0), \mathcal{Q}_\hbar(A_0')] - \mathcal{Q}_\hbar(\{A_0, A_0'\}) \| = 0 \tag{45}$$

同样注意到式(27)就是式(12)。因此，参见式(25)和周围的文本：

> 由式(2)定义的C^*代数$\mathcal{A}^{(c)}$连续场和作为连续截面的近似对称列（和它们的极限）对任意的量子化映射$\mathcal{Q}$，产生了相空间$\mathcal{S}(\mathcal{A}_1)$的一个形变量子化，其用泊松括号(38)来装备。

对于动力学，这意味着：

> 令$H = (H_0, H_1, H_2, \cdots)$是由一个对称列定义的$\mathcal{A}^{(c)}$的连续截面，① 令$A = (A_0, A_1, A_2, \cdots)$是$\mathcal{A}^{(c)}$的一个任意连续截面（即一个近

① 再者，该结果事实上对更多的哈密顿量类型成立，即那些满足确定的条件的哈密顿量[Duffield and Werner, 1992a]。每个哈密顿量H_N处于$\mathcal{A}_1^N$中且因而是有界的这一假设在点阵模型中是自然的，但一般是不合需要的。

似对称列)。则下述列定义了$\mathcal{A}^{(c)}$的一个连续截面:

$$(A_0(t),\ e^{iH_1t}A_1e^{-iH_1t},\ \cdots,\ e^{iNH_Nt}A_Ne^{-iNH_Nt},\ \cdots) \tag{46}$$

其中$A_0(t)$是拥有经典哈密顿量H_0的运动方程的解。①

换言之,对哈密顿量H_N的有界对称列,限制于宏观可观测量的量子动力学收敛于拥有哈密顿量H_0的经典动力学。分别对比式(12)和式(46)中$\hbar$和N的位置,我们会欣喜地发现这是对式(44)的再确证。

相反,一旦考虑了由这样的哈密顿量定义的动力学极限$N\to\infty$,准定域可观测量的表现会不尽如人意。也就是说,如果$(A_0, A_1, \cdots)$是连续场$\mathcal{A}^{(q)}$的一个截面,且$(H_1, H_2, \cdots)$是哈密顿量的任意有界对称列,则列:

$$(e^{iH_1t}A_1e^{-iH_1t},\ \cdots,\ e^{iNH_Nt}A_Ne^{-iNH_Nt},\ \cdots)$$

在$N\to\infty$时没有极限,因为它不能够通过某个$A_0(t)$拓展到$\mathcal{A}^{(q)}$的连续截面。事实上,这正是最初在此语境下引入宏观可观测量的确切原因[Rieckers, 1984; Morchio and Strocchi, 1987; Bona, 1988; Unnerstall, 1990a; Raggio and Werner, 1989; Duffield and Werner, 1992a]。相反地,自然的有限N哈密顿量作为在$\mathcal{A}^{(q)}$上的单参数自同构群,满足恰当的定域性条件,该定域性条件排除了定义对称列的全域平均。对此哈密顿量而言,$\mathcal{A}_1^N$上时间演化的极限$N\to\infty$存在。

6.6 尾声:宏观可观测量与测量问题

在一篇知名论文中,亨普[1972]提出,宏观可观测量与超选规则应当在量子力学测量问题的求解中发挥作用。他假定说,一个宏观的仪器可以被理想化为无限量子系统,该量子系统的可观测量$\mathcal{A}_A$代数拥有不相交的纯态。参考我们在2.5小节对语境与记号的讨论,亨普的基本观点(对此他宣称没有原创性)

① 见式(18)及其上下文。

是，作为测量过程的结果，系统与仪器的初态向量$\Omega_I = \sum_n c_n \Psi_n \otimes I$演化为终态向量$\Omega_F = \sum_n c_n \Psi_n \otimes \Phi_n$，其中每个$\Phi_n$处于仪器希尔伯特空间的一个不同的超选分支中(换言之，在$\mathcal{A}_A$上的对应态φ_n是相互不相交的)。因此，虽然初态ω_I是纯态，但终态ω_F是混合态。此外，由于ω_n的不相交性，终态ω_F拥有关于纯态的唯一分解$\omega_F = \sum_n |c_n|^2 \psi_n \otimes \varphi_n$，从而允许一种真正的无知解释。亨普因此在某种理由上宣称，测量“使得波包塌缩”，正如量子测量理论所要求的那样。

即使除了从最终的混合态中所有项的集合到一个实际测量结果过渡的概念问题外，亨普本人还指出了此方案的一个严重数学问题。即如果初态是纯态，它必须处于一个特定的超选分支(或态的等价类)中，但这样的话，如果时间演化是哈密顿的，或更一般地是自同构的(正如我们在前一小节中看到的)，则终态必须处于相同的分支中。此外，亨普[1972]本人论证了一个更普遍的引理：

> 若C^*代数$\mathcal{B}$上的两个态ρ，σ是不相交的，且α：$\mathcal{B} \to \mathcal{B}$是$\mathcal{B}$的自同构，则$\rho \circ \alpha$与$\sigma \circ \alpha$也是不相交的。

为了得到与上面相反的结论，我们将$\mathcal{B}$看作是系统与仪器共同的可观测量代数，通过选择$\alpha = \alpha^{-1}_{t_F - t_I}$来计算时间回溯，其中$\alpha_t$是描述了系统与仪器共同的时间演化(且$t_I$与$t_F$分别是测量前与测量后的时间)的$\mathcal{B}$上的单参数自同构群。然而，亨普指出，如果人们允许测量要无限长时间来完成这种可能性，则这一结论可以避免。因为极限$A \mapsto \lim_{t\to\infty} \alpha_t(A)$(假如它在适当的，即弱的意义上存在)并不必然产生$\mathcal{B}$的自同构。因此态——在薛定谔图像中用$\omega_t(A) \equiv \omega(\alpha_t(A))$来描述其演化——在无限长时间下可能会留在其分支上，这种可能性亨普事实上已经在一系列模型中展示了，同样参见[Frigerio(弗里赫里奥)，1974；Whitten-Wolfe and Emch(惠顿—沃尔夫和埃姆什)，1976；Araki，1980；Bona，1980；Hannabuss(汉纳布斯)，1984；Bub，1988；Landsman，1991；Frasca(弗拉斯卡)，2003；2004]和其他许多论文。

虽然反对下述结论遭到了一些批评，即量子力学的测量需要一个无限的仪

器，且必须要花无限长的时间[Bell，1975；Robinson，1994；Landsman，1995]，虽然这一做法与冯·诺伊曼[1932]的理念相悖(在冯·诺伊曼被广泛采纳的描述中测量实际上是瞬时的)，但该结论得到了当前观点的响应，即量子理论普遍有效且经典世界没有绝对的存在性这一观点，参见引言。此外，量子力学的测量只不过是一个具体的相互作用，与散射过程属于同类型，且这样一个过程需要无限长时间来完成这一点是没有争议的。确实，散射在某个有限时间之后结束意味着什么？哪个时间？正如我们将在下一节中看到的，退相干理论也要求极限 $t\to\infty$，而且很大程度上是出于相同的数学原因。在退相干和亨普的方法中，极限行为事实上很快就达到了(在相关的时间尺度上)，我们需要令 $t\to\infty$ 是为了使 $\sim\exp(-\gamma t)$(在 $\gamma>0$)项为零，而不只是非常小。关于此极限意义不那么实用的观点也参见[Primas，1997]。

亨普方法一个更为严重的问题在于他关于无限测量仪器可观测量准局域代数(在我们的案例种类中是$\mathcal{A}_0^{(q)}$)的时间演化是自同构的这一假设。然而，情况绝不总是这样，参见6.5节末尾列出的文献。正如我们已经看到的，在有限 N 处对于某个自然哈密顿量(因而它是自同构的)的时间演化而言，动力学在准局域可观测量代数上不拥有极限 $N\to\infty$——更不用说是一个自同构的极限了。

然而，如果我们采用宏观可观测量代数$\mathcal{A}_0^{(c)}$，在该代数上(在适当的假设下，见6.5节)$\mathcal{A}_1^N$上哈密顿量的时间演化确实拥有一个 $N\to\infty$ 极限，则亨普的结论仍然有效。因为正如6.3节中指出的，$\mathcal{A}_0^{(q)}$的每一个超选分支定义了$\mathcal{A}_1$上的一个纯态，反过来也被该纯态所定义，而该纯态依次又定义了$\mathcal{A}_0^{(c)}$的一个分支。现在后一个分支仅是对易 C^* 代数$\mathcal{A}_0^{(c)}$的纯态空间$\mathcal{S}(\mathcal{A}_1)$上的一个点，因而上面引述的亨普引理归结为如下陈述：如果 $\rho\neq\sigma$，则对于任意自同构 α 有 $\rho\circ\alpha\neq\sigma\circ\alpha$。当然，这只是任意哈密顿量的时间演化的一个平凡特征，仍可再次得到在$\mathcal{A}_0^{(c)}$上从一个纯的前测量态到一个混合的后测量态间的过渡在有限时间内是不可能的。为了避免这一结论，我们应该避免极限 $N\to\infty$，它是 $t\to\infty$ 极限的根源，见[Jassens(杰森)，2005]。

所有这些形式体系对薛定谔猫意味着什么呢？在我们看来，它确证了这样一个印象，即悖论的表象建立在含糊的措辞之上。事实上，因为人们在两个相

互矛盾的解释间摇摆不定，问题才会产生。①

一种解释认为，一个波西米亚式的理论学家，当他(她)摘下眼镜然后仔细考虑着猫是死是活的问题时，或茫然或犹疑。这样一个人对猫的研究是专门从它宏观可观测量的角度进行的，所以他(她)必须采用代数$\mathcal{A}_0^{(c)}$上的后测量态$\omega_F^{(c)}$。如果$\omega_F^{(c)}$是纯态，它位于$\mathcal{P}(\mathcal{A}_1)$中(除非前测量态是混合态)。这样一个态对应于$\mathcal{A}_0^{(q)}$的单个超选分支$[\omega_F^{(q)}]$，因而猫是死的或是活的。另一方面，如果$\omega_F^{(c)}$是混合态(如果薛定谔想要的话，则会是这样)，那么一开始就不会有问题：在宏观可观测量层次上人们只拥有关于猫的统计描述。

另一种解释中，一个勤快的、有使不完的劲的实验物理学家，他研究了猫的详细微观构造。对他(她)而言，对于某个大的N，猫总是处于$\mathcal{A}_1^N$上的一个纯态中。这一次生或死的论题并不是懒散的观测和结论的问题，而是完全无止尽的实验和计算问题。从这样一个观测者的角度来看，猫处于两个态的相干叠加中没有什么不妥，微观上这两个态实际上是相当接近对方的——至少暂时是那样。

无论哪种解释，“这样的谜是不存在的”，见维特根斯坦的《逻辑哲学论》(*TLP*)，§6.5。

7. 为什么是经典的态与可观测量?

> “我们在未知的海岸上发现了一个陌生的脚印。我们设计了一个又一个深刻的理论以说明它的来源。最后，我们在重建留下脚印的生物的过程中获得了成功。瞧！它就是我们自己。”见[Eddington(爱丁顿)，1920，pp. 200—201]。

第5与第6节的结论是，量子理论可能产生处于某个态且相关于某个可观测量的经典行为。例如，我们已经看到在极限$\hbar\to 0$下，相干态与$Q_\hbar(f)$形式的算符分别是适当的，而在极限$N\to\infty$下，我们又应该采用如在6.2节中定义的

① 互补性是不是以某种隐秘的方式重新出现的呢?

经典态(名称是一个符号!)和宏观可观测量。相反,如我们采用这经典态的叠加,或有着错误极限行为的可观测量,则不会产生经典物理学。因而,为什么世界一般处于这样的态,为什么我们用涉及的可观测量来研究这个世界,这些问题都没有答案。这些问题在测量问题中找到了它最初的原型(参见2.5节),但这个问题其实代表了更大的难题。

在过去的25年里,[①] 关于这一问题提出了两种深刻与独创性的答案。

7.1 退相干

第一种答案的名称是退相干。前期的论文包括[van Kampen, 1954; Zeh, 1970; Zurek, 1981; 1982][②]与[Joos and Zeh, 1985],一些最近的评论有[Bub, 1999; Auletta, 2001; Joos *et al.*, 2003; Zurek, 2003; Blanchard and Olkiewicz(布兰查德和奥利基威克茨), 2003; Bacciagaluppi, 2004]与[Schlosshauer(施洛斯豪尔), 2004]。[③] 更多的参考文献会在适当的时候给出。有了这些评论(与其精彩之处)就不需要在本文中给出对退相干的详细处理,更何况在本论文撰写时退相干在概念上和数学上都处于过渡阶段(在下文会变得很明显)。因而,我们不同于前面章节的开头,只给出一些个人的评论。

1. 数学上,退相干归结为在 $S+A$(即系统与仪器)之外,在冯·诺伊曼链(见2.5节)上多增加一个环节。但是,在概念上,退相干与原先的方法有着很大的区别就在于采取了这一步:原先(在冯·诺伊曼、伦敦与鲍厄、维格纳等人那里)[④]链指向观测者,而在退相干中它离开了观测者。也就是说,第三个和最后的环节现在被看作是环境(在完全字面的意义上与该词的直觉意义相同)。特别地,在现实的模型中,环境被作为无限系统处理(加上极限 $N\to\infty$),这就有了这样的结果,即(在指针具有离散谱的简单模型中)后测量态 $\sum_n c_n\Psi_n \otimes$

---

① 虽然有人说退相干的基本理念可追溯到海森堡和路德维希(Ludwig)。

② 也见[Zurek, 1991]和随后在《当代物理》(*Physics Today*)[Zurek, 1993]上的争论,它们引起了对退相干的广泛关注。

③ 网页 http://almaak.usc.edu/~tbrun/Data/decoherence_list.html 包含了关于退相干的详尽文献列表。

④ 见[Wheeler and Zurek, 1983]。

$\Phi_n \otimes \chi_n$（其中χ_n是相互正交的）仅在极限$t \to \infty$下达到。然而，正如在6.6节中提到的，无限时间只在数学上需要以使得$\sim \exp(-\gamma t)$类型的项（有$\gamma > 0$）为零，而不仅是非常小：在许多模型中内积(χ_n, χ_m)事实上可以忽略，因为在非常短的时间尺度内$n \neq m$。①

如果仅从需要$N \to \infty$（对于环境）和$t \to \infty$类型极限的观点来看，我们认为，退相干最好与引言中的立场1联系起来，因为它的目标是说明从被看作普遍有效的量子力学中得到经典世界的近似表象。但是，退相干被表述为支持量子力学基础的几乎任何观点，批判性的评论参见[Bacciagaluppi，2004]和[Schlosshauer，2004]，也见下文第3点。

2. 最初，退相干是作为测量问题（以一种2.5节结尾处提到的精确形式）的一种解决方案进入视野的。因为态$\sum_n c_n \Psi_n \otimes \Phi_n \otimes \chi_n$对$S+A$的约束（即对环境的自由度求迹）在极限$t \to \infty$下是混合的，这意味着对不同的$n$值态$\Psi_n \otimes \Phi_n$之间的量子力学干涉"退局域化"到环境中了，如果不观察环境的话（即在描述时忽略掉）从而是无意义的。不幸的是，将无知解释应用到$S+A$测量后的混合态上是不合适的，即使是从引言中的第1种立场来看。无知解释仅当环境在描述范围内且是经典的时（有对易的可观测量C^*代数）才有效。但是，后一假设[Primas，1983]使得测量问题的退相干求解是循环论证。②

事实上，正如巴奇伽鲁皮[2004]所正确指出的那样，退相干实际上加剧了测量问题。而之前这一问题被认为是人为的，且只与那些非常异常的实验室情形相关（正如它们对于物理学基础而言是重要的），但现在很清楚，环境（取代了一个实验物理学家）对一个量子系统的"测量"无处不在且一直在发生，所以现在的情形变得比以前更加不可思议，以前的情形是这样一个测量之后存在一个单一的结果。因而，这样的退相干并没有为测量问题提供一个解决方案

① 参见[Joos *et al.*，2003]中第66—67页上的表3.1和3.2。

② 另一方面，对环境的处理就好像它是经典的，这可能是对哥本哈根思想中将测量仪器当作它是经典的来处理的改进（参见第3节）。

[Leggett, 2002;[1] Alder(阿尔德), 2003; Joos and Zeh, 2003], 而实际上是依附于这样一个解决方案。

3. 对这一见解有许多不同的回应。占支配地位的一种是，把退相干与量子力学的某种解释结合起来，从而退相干找到了归属，而反过来涉及的解释又常常通过退相干而得到加强。在该背景下，最流行的是多世界解释，该解释在经过多年的默默无闻与嘲讽之后，突然开始随着退相干的流行而吹响了蓬勃发展的号角。例如，见[Saunders, 1993; 1995; Joos *et al.*, 2003]与[Zurek, 2003]。在量子宇宙学圈子里，一致性历史方法已经成为退相干受欢迎的搭档，常常与多世界结合在一起，见下文。退相干在模态解释中的重要性由狄克斯[1989b]、贝内和狄克斯(Bene and Dicks)[2002]所强调，而实践中退相干的所有作者都以这样或那样的方式在口头上支持玻尔。对所有这些结合的批判性评论见[Bacciagalluppi, 2004]与[Schlosshauer, 2004]。

在我们看来，没有哪个已经建立的量子力学解释能胜任这一任务，这为真正的新观点留下空间。一种新的观点是回到环境：不要“对它求迹”，正如在退相干理论最初的背景中那样，环境不应该被忽略！测量的本质现在被认识到是测量结果(或“记录”)在环境中的冗余。正是这种关于基本量子客体信息的冗余“客观化了”它，信息对于许多的观测者是可达的，没有必要干扰客体[2][Zurek, 2003; Ollivier *et al.* (奥利弗等), 2004; Blume-Kohout and Zurek(布卢姆—科胡特和朱瑞克), 2004; 2005]。这一见解(称为“量子达尔文主义”)产生了朱瑞克[2003]的量子力学“存在性的”解释。

4. 对退相干不能够(事实上所有其他方法也不能够)解决测量问题(在一种不能够赢得广泛认同的意义上)的另一种反应是某种更为悲观的(或许有人会说是实用的)观点，即所有解释量子世界的尝试都被放弃了，产生了“物理学在最

---

① 事实上，莱格特的论证仅适用于引言中的立场3，且失去了其对立场1的反对效力。因为他的论点是，退相干只是移除了一个给定态(薛定谔猫类型的)成为叠加态的证据，并指责那些宣称这解决了测量问题的人们，那些人犯了承认移除犯罪证据将取消犯罪这样的逻辑错误。但是根据立场1，犯罪仅相对于证据而确定！但是，莱格特坚持在引言结尾处提到的“从‘与’到‘或’问题”是非常正确的。

② 这样的客观化被认为产生了“对存在的操作定义”[Zurek, 2003, 749]。

基本层次上的真正目标变为信息的表征与操作”这一观点[Bub，2004]。这里“测量仪器最终在某个层次上仍是黑匣子”，且有人得到结论说，理解测量(或就事论事地说，EPR关联)的所有努力都是无意义的且不会有结果的。①

5. 这是一个量子物理学家的噩梦吗?② 也不尽然。反过来看：退相干理论的真正意义不是解决测量问题，而是给出了测量问题不存在的条件！也就是，前面提到的对从$S+A+\mathcal{E}$的态$\sum_n c_n\Psi_n\otimes\Phi_n\otimes\chi_n$向$S+A$的一个态$\Psi_n\otimes\Phi_n$过渡的解释，在退相干的核心中是这样的论述，即第二个态中的每一个对于环境的耦合都是鲁棒的。正如目前所假定的，如果哈密顿量使得$\Psi_n\otimes\Phi_n$与环境的某个初态$I_{\mathcal{E}}$的张量结合在一起，事实上演化到$\Psi_n\otimes\Phi_n\otimes\chi_n$。这意味着每个$\Psi_n\otimes\Phi_n$态在与环境耦合后都仍然是纯态，随后限制到原初系统加仪器也是纯态，因而最后环境不会对它们有影响。换言之，退相干的真正意义是，环境超选现象(einselection，即environment-induced superselection)，其中一个态是“环境超选的”是指(给定某个相互作用哈密顿量)它拥有前面提到的稳定性特征。从而论述变为，环境超选的态常常是经典的，或至少经典态(在一种本节开始时提到的意义上)是经典的，正是因为它们与环境的耦合是鲁棒的。假如这一方案确实产生了经典世界(这将在后面详述)，它对此给出了一种动力学的解释。但即使尚未达到这一目标，环境超选概念的重要性也不能够被夸大；在我们看来，它是自纠缠(当然，环境超选试图取消它!)以来量子理论中最为重要且强有力的观点。

6. 测量问题，和相关的在系统与仪器为一方和环境作为另一方之间的区别，现在可以从退相干理论中忽略掉。继续3.4节的讨论，退相干的目标应该只是发现与环境$\mathcal{E}$耦合的客体$\mathcal{O}$的鲁棒的或环境超选的态，也包括由此引起的动力学(给定$\mathcal{O}+\mathcal{E}$的时间演化)。但是，这种寻找必须包括从总的$\mathcal{S}+\mathcal{E}$中对客体$\mathcal{O}$的正确识别，即作为一个真正拥有该鲁棒态的子系统。因而，在客体与仪器间的海森堡界线是可移动的(对比3.2节)这一哥本哈根观念，一般将不会扩展

① 退相干正是在对测量从量子信息(或熵)到经典信息转变的描述中出现并展现出了某种强大的力量。或许因为这个原因这样的想法也说服了朱瑞克[2003]。

② [Kent，2000]。与麦克·科尔马赫(McCormmach)[1982]的题目双关。

为在客体与仪器之间的“普利马斯—朱瑞克”界线。在传统的物理学术语中，该问题是找到一个量子系统的正确“外衣(dressing)”以使得至少它的某些态在与环境的耦合中够鲁棒[Amann and Primas，1997；Brun and Hartle(布伦与哈特)，1999；Omnès，2002]。换言之：什么是一个系统？为了标记这一视角的变换，我们现在把记号从$\mathcal{O}$(即客体)改为$\mathcal{S}$(即系统)。在退相干理论中求解此问题的各种工具现在也有了——并在不断修正中，对信息论中概念的依赖也在增加[Zurek，2003]——但对它而言正确的背景似乎是一致历史体系，见下文。

7. 现在已经发现了多种动力学体系，每一种都产生一种不同的鲁棒态[Joos *et al.*，2003；Zurek，2003；Schlosshauer，2004]。这里H_S是系统的哈密顿量，H_I是系统与环境间的相互作用哈密顿量，$H_{\mathcal{E}}$是环境哈密顿量。正如所述，这里不需要提及测量、客体与仪器。

· 在$H_S \ll H_I$体系下，对于适当的哈密顿量鲁棒态是量子测量理论中传统的指针态。这一体系符合冯·诺伊曼的观点[von Neumann，1932]，即量子测量几乎是瞬时的。此外，若也有$H_{\mathcal{E}} \ll H_I$——无论在或不在测量语境下——退相干机制被证明是普遍的，独立于$\mathcal{E}$与$H_{\mathcal{E}}$的细节[Strunz *et al.*(斯特伦茨等)，2003]。

· 若$H_S \approx H_I$，则(至少在量子布朗运动模型中)鲁棒态是相干态(可能是传统的薛定谔类型，也可能如5.1节中定义的那样具有更一般的性质)，见[Zurek *et al.*，1993]和[Zurek，2003]。当然，这一情形对于我们在第5节中引用物理相关的结果具有极高的重要性，且——若仅由于此——退相干理论会从与量子随机性分析的数学详细结果的相互结合中受益更多。①

· 最后，若$H_S \gg H_I$，则鲁棒态被证明是系统哈密顿量H_S的本征态[Paz and Zurek，1999；Ollivier *et al.*，2004]。从5.5与5.6节中我们对这些态的讨论看来，这表明鲁棒态没有必要是经典的。应该提到的是，在这一背景下退相干理论很大程度上与标准原子物理学一致，其中原子被看作是系统$\mathcal{S}$，辐射场扮演环境$\mathcal{E}$的角色。以数学为主的导论见[Gustafson and Sigal，2003]，更完整

① 参见[Davies，1976；Accardi *et al.*(阿卡迪等)，1990；Parthasarathy(巴萨拉席)，1992；Streater(斯特赖特尔)，2000；Kümmerer(古美雷尔)，2002；Maassen(马森)，2003]等。

的(数学)版本见[Bach *et al.*, 1998; 1999]。

8. 关于上文对能量本征态角色的澄清，退相干也有关于量子混沌的重要论述[Zurek, 2003; Joos *et al.*, 2003]。参见2.4节中我们对波包复原的讨论，我们看到在原子物理学中波包并不在长时间尺度上表现为经典的。或许很令人惊奇，这对于某个混沌的宏观系统也成立，参见引言中提到的亥伯龙情形与5.2节的结尾。退相干现在用经典的概率分布取代了基本的叠加，这反映了极限经典动力学的混沌特征。再者，从系统加环境相关的纯态向单个被观测系统态的过渡，仍然笼罩着迷雾。但通过承认这一过渡，退相干单个阐明了经典混沌并且至少避免了与观测间的最明显的冲突。①

9. 鲁棒性和环境超选形成了态的一方或退相干的薛定谔绘景。当然，也应该有相应的可观测量一方或退相干的海森堡绘景。但这里两种图象间的转换远比在封闭系统的量子力学中微妙得多。在薛定谔绘景中，环境超选的全部意义在于，绝大多数的纯态只是从场景中消失。这可以在具有希尔伯特空间$\mathcal{H}_S=\mathbb{C}^2$的二阶系统上精彩呈现[Zurek, 2003]。参见式(33)，若↑与↓正好是系统在与环境恰当耦合后的鲁棒向量态，且如果我们把密度矩阵分别看作是北极$(0, 0, 1)\in B^3$和南极$(0, 0, -1)\in B^3$，则按照退相干，随着$t\rightarrow\infty$所有其他态都朝向联结北极与南极的轴(即z轴与B^3的交集)运动。在海森堡绘景中，在相同极限下，除了这两个态外的所有纯态的消失对应于系统的所有可观测量代数$M_2(\mathbb{C})$向它的对角化(且因此是对易的)子代数$\mathbb{C}\oplus\mathbb{C}$约化。因为正是后一个代数包含了足够的元素以区分出↑与↓，而不包含检测这些纯态间干涉项的可观测量。

10. 为了以一种更抽象与一般的方式来理解，我们回顾纯态与可观测量之间的数学关系[Landsman, 1998]。从一个给定系统的可观测量的C^*代数$\mathcal{A}$到它的纯态的过渡是熟知的：作为一个集合，纯态空间$\mathcal{P}(\mathcal{A})$是总态空间$\mathcal{S}(\mathcal{A})$的极边界。为了从$\mathcal{P}(\mathcal{A})$重建$\mathcal{A}$，$\mathcal{P}(\mathcal{A})$需要通过(27)用一个跃迁概率空间结构(见6.3节)来装备。每一个$A\in\mathcal{A}$通过$\hat{A}(\omega)=\omega(A)$在$\mathcal{P}(\mathcal{A})$上定义了一个函

① 尽管应当提到任何解释从量子到经典转变的成功机制都应该具有此特征，从而到最后退相干可能被证明在这里只是转移了人们的注意力。

数$\hat{A}$。现在，在$\mathcal{A}$是有限维（因而是矩阵代数的直和）的简单情形中，我们可以表明每个$\hat{A}$函数是具有$\hat{A}=\sum_i p_{\omega_i}$形式的有限线性组合，其中$\omega_i\in\mathcal{P}(\mathcal{A})$且$\mathcal{P}(\mathcal{A})$上的基本函数$p_\rho$由$p_\rho(\sigma)=p(\rho,\sigma)$来定义。反过来，每个这样的线性组合对某个$A\in\mathcal{A}$定义了一个函数$\hat{A}$。因而，$\mathcal{A}$可看作是在线性态空间$\mathcal{P}(\mathcal{A})$上的函数，其的元素只是跃迁概率和它们的线性组合。$\mathcal{A}$的代数结构可能从作为具有跃迁概率的泊松空间$\mathcal{P}(\mathcal{A})$的结构来重建（参见6.5节）。在此意义下$\mathcal{P}(\mathcal{A})$唯一地确定了可观测量的代数，它是纯态空间的可观测量。例如，由具有经典跃迁概率(31)的两个点组成的空间产生了对易代数$\mathcal{A}=\mathbb{C}\otimes\mathbb{C}$，而在$\mathbb{R}^3$中具有跃迁概率(32)的单位双球产生了$\mathcal{A}=M_2(\mathbb{C})$。

这一重建过程可以推广到任意的C^*代数[Landsman，1998]，并定义在相关于退相干的薛定谔绘景与海森堡绘景之间的精确联系。这些绘景是等价的，但在实践中这些重建过程可能很难进行下去。

11. 出于此原因在海森堡绘景中有对退相干的直接描述是很有意义的。布兰查德和奥利基威克茨［2003］提出了这样一个描述，部分基于奥利基威克茨［1999a，b；2000］之前的结果。数学上，他们的方法比薛定谔绘景更强有力，关于退相干的绝大多数文献建立在薛定谔绘景上。令$\mathcal{A}_S=\mathcal{B}(\mathcal{H}_S)$和$\mathcal{A}_\mathcal{E}=\mathcal{B}(\mathcal{H}_\mathcal{E})$，假定总哈密顿量$H$作用于$\mathcal{H}_S\otimes\mathcal{H}_\mathcal{E}$和环境的一个固定态，后者用密度矩阵$\rho_\mathcal{E}$表示（常常被看作是热平衡态）。若$\rho_S$是在$\mathcal{H}_S$上的一个密度矩阵（从而总态是$\rho_S\otimes\rho_\mathcal{E}$），退相干（更一般地是开放系统的量子理论）的薛定谔绘景方法建立在下式的时间演化上：

$$\rho_S(t)=\mathrm{Tr}_{\mathcal{H}_\mathcal{E}}(e^{-\frac{it}{\hbar}H}\rho_S\otimes\rho_\mathcal{E}e^{\frac{it}{\hbar}H})\tag{1}$$

另一方面，海森堡绘景建立在相关算符的时间演化上，对于$A\in\mathcal{B}(\mathcal{H}_S)$由下式给出：

$$A(t)=\mathrm{Tr}_{\mathcal{H}_\mathcal{E}}(\rho_\mathcal{E}e^{\frac{it}{\hbar}H}A\otimes 1e^{-\frac{it}{\hbar}H})\tag{2}$$

由此产生了由下式表达的薛定谔绘景与海森堡绘景间的等价性：

$$\mathrm{Tr}_{\mathcal{H}_S}(\rho_S(t)A)=\mathrm{Tr}_{\mathcal{H}_S}(\rho_SA(t))\tag{3}$$

更一般地，令$\mathcal{A}_S$与$\mathcal{A}_\mathcal{E}$是单位的C^*代数，具有空间张量积$\mathcal{A}_S\otimes\mathcal{A}_\mathcal{E}$，它具有时间

演化 α_t 和在$\mathcal{A}_{\mathcal{E}}$上的固定态 $\omega_{\mathcal{E}}$。通过对 $P_{\mathcal{E}}(A\otimes B)=A\omega_{\mathcal{E}}(B)$的线性与连续拓展这定义了一个条件期望值 $P_{\mathcal{E}}$: $\mathcal{A}_S\otimes\mathcal{A}_{\mathcal{E}}\to\mathcal{A}_S$，从而通过下式定义了$\mathcal{A}_S$上一个约化时间演化 $A\mapsto A(t)$：

$$A(t)=P_{\mathcal{E}}(\alpha_t(A\otimes 1)) \tag{4}$$

见阿里奇和兰迪(Alicki and Lendi)[1987]。在我们的背景中，这一普遍性对于连续经典相空间的潜在突现是至关重要的，见下文。① 现在的关键点是，退相干是由作为矢量空间(而非一个 C^* 代数)的$\mathcal{A}_S=\mathcal{A}_S^{(1)}\oplus\mathcal{A}_S^{(2)}$的分解来描述的，其中$\mathcal{A}_S^{(1)}$是一个 C^* 代数，对于所有 $A\in\mathcal{A}_S^{(2)}$具有 $\lim_{t\to\infty}A(t)=0$ 的(弱)特征，而 $A\mapsto A(t)$对于每个有限的 t 是$\mathcal{A}_S^{(1)}$上的一个自同构。因而，$\mathcal{A}_S^{(1)}$是退相干之后可观测量的有效代数，正是$\mathcal{A}_S^{(1)}$上的纯态在之前讨论的那种意义上是鲁棒的或是环境超选的。

12. 例如，若$\mathcal{A}_S=M_2(\mathbb{C})$，态↑与↓在退相干下是鲁棒的，则$\mathcal{A}_S^{(1)}=\mathbb{C}\oplus\mathbb{C}$和$\mathcal{A}_S^{(2)}$由对角为零的 2×2 矩阵组成。在此例子中，$\mathcal{A}_S^{(1)}$是对易的因而是经典的，但这并不是一般情形。但若这是一般情形的话，$\mathcal{A}_S^{(1)}$上的自同构时间演化在其结构空间上产生一个经典的流，它应该是哈密顿流，利用第 6 节中的技巧来处理。② 无论如何，每当$\mathcal{A}_S^{(1)}$有一个非平凡中心时，就会有退相干系统的某种经典行为。③ 若该中心是离散的，则其上引起的时间演化必然是平凡的，且我们得到了典型的测量情形，其中所涉中心是由向具有离散谱的指针可观测量本征态的投影产生的。对于$\mathcal{A}_S$是类型 I 因子的情形这是共同的。然而，类型Ⅱ与Ⅲ因子可能会产生具有非平凡时间演化的连续经典系统，见[Lugiewicz and Olkiewicz(路奇维卡和奥利基威克茨)，2002；2003]。我们在这里不可能公平地评判完整的技术细节和这里涉及的复杂性。但我们想要强调关于量子场论和热动力学极限理论，退相干的当前语境应该为专业人士在量子理论基础中学习算符代数理论提供重要的启示。

---

① 由于技术原因布兰琪和奥基维施[2003]假定$\mathcal{A}_S$是具有平凡中心的冯·诺伊曼代数。

② 因为在前一个脚注中假定$\mathcal{A}_S^{(1)}$是对易的冯·诺伊曼代数，人们应当以一种非直接的方式定义结构空间；见[Blanchard and Olkiewicz，2003]。

③ 这是可能的，即使当$\mathcal{A}_S$是因子！

7.2 一致性历史

这样做的同时，我们应该更加努力，同时熟悉一致性历史。格里菲斯(Griffiths)是研究量子理论这一理论的先锋[1984]，随后出现翁内斯[1992]等人。盖尔曼和哈特(Gell-Mann and Hartle)[1990；1993]也独立提出了类似的观点。类似于退相干，一致性历史方法也是许多评论[Hartle，1995]和创立者著作[Omnès，1994；1999；Griffiths，2002]中的主题。也可参考基弗(Kiefer)[2003]和哈利维尔(Halliwell)[2004]的评论，多克和肯特(Dowker and Kent)[1996]、肯特[1998]、巴布[1999]、巴锡和吉拉迪(Bassi and Ghirardi)[2000]的批判，还有对一致性历史理论不同的数学重建与再解释[Isham，1994；1997；Isham and Linden，1994；1995；Isham *et al.*，1994；Isham and Butterfield，2000；Rudolph(鲁道夫)，1996a；1996b；2000；Rudolph and Wright(鲁道夫和怀特)，1999]。

一致性历史与退相干之间的关系有些特殊。一方面，退相干是自然的机制，通过它适当的历史集合变得(近似)一致，但另一方面二者又似乎有着非常不同的出发点。退相干开始于下述观念，即(量子)系统与它们环境的自然耦合，因而必须被作为开放系统来处理，而一致性历史的目标是处理封闭量子系统，如宇宙，对测量或观测者没有先验的讨论。但是，这种区分仅是过去式，正如我们在前一小节的第6项中看到的，系统与其环境间的分界线应该被看作是动力学的实体，要根据特定的稳定性标准进行划出，因而即使在退相干理论中，从一开始我们就应该真正地把系统及其环境作为一个整体来研究。① 而这正是一致性历史学家们所做的。

如同前一节，也出于同样的原因，我们对一致性历史的处理也以罗列开放讨论项来进行。

1. 量子理论一致性历史构想的出发点是约定的：我们有一个希尔伯特空间$\mathcal{H}$，一个态ρ，它被看作是所考虑的总系统的初态(在$\mathcal{H}$上实现为一个密度矩阵)和一个哈密顿量H(定义为$\mathcal{H}$上的一个自伴算符)。这一总系统可能是整个宇宙，这一点是非约定的。总系统的每个特性α在数学上用$\mathcal{H}$上的投影P_α表

① 这使得在系统的“开放”与“封闭”间的区分有点“舍本逐末”，正如在退相干理论中系统与其环境的总体被当作封闭系统来处理。

示，例如，若 α 是能量取值 ϵ 的特性，则算符 P_α 是向相关本征空间的投影（假定 ϵ 属于 H 上的离散谱）。在海森堡绘景中，P_α 随时间演化到 $P_\alpha(t)$ 满足(12)，注意 $P_\alpha(t)$ 仍然是一个投影。

一个历史 $\mathbb{H}_A$ 是一个特性（或命题）链 $(\alpha_1(t_1),\cdots,\alpha_n(t_n))$，用 n 个不同的时间 $t_1<\cdots<t_n$ 来标示，这里 A 是多重标示的，同时体现了特性 $(\alpha_1,\cdots,\alpha_n)$ 和时间 $(t_1,\cdots,t_n)$。这样一个历史表明，$i=1,\cdots,n$ 时，在时刻 t_i 每个特性 α_i 成立。该历史可能被看作是对所有时间定义的集合 $\{\alpha(t)\}_{t\in\mathbb{R}}$，但是出于简单性我们常常假定仅对时间 t 的有限集有 $\alpha(t)\neq 1$（其中 1 是总是成立的平凡特征），该时间集合精确而言是 $\{t_1,\cdots,t_n\}$。海森堡提出的一个案例是把 α_i 看作是下述特征，即穿过威尔逊云室的一个粒子在其相空间的一个单元 $\Delta_i\subset\mathbb{R}^6$ 中被找到，则历史 $(\alpha_1(t_1),\cdots,\alpha_n(t_n))$ 表示了这样的态，在其中粒子在时刻 t_1 处于单元 Δ_1 中，随后在时刻 t_2 处于单元 Δ_2 中，等等。在时刻的间隔中没有关于粒子行为的表述。另一个历史例子是由双缝干涉实验提供的，其中 α_1 是粒子在 t_1 时刻在粒子源处的发射（描述中常常忽略），α_2 是粒子在 t_2 时刻穿过（例如）上面的缝，α_3 是 t_3 时刻屏幕的某个位置 L 处检测到粒子。正如我们都知道的，这一历史存在一个潜在的问题，该问题将在下文的当前框架中予以澄清。

一致性历史学家们基本的论述似乎是，量子理论应该就是做出关于历史发生可能性的预言。这些概率空间意味着什么仍然是不清楚的——或许除了当概率接近于 0 或 1 时，或者除了当具体到某个测量环境中时，见[Hartle，2005]——但我们首先来看看什么时候我们能够定义它们以及如何定义。对于历史 $\mathbb{H}_A$ 相关于态 ρ 的概率，唯一可能有意义的数学表达（在量子力学中）是[Groenewold（格伦沃德），1952；Wigner，1963]：

$$p(\mathbb{H}_A)=\mathrm{Tr}(C_A\rho C_A^*) \tag{5}$$

其中：

$$C_A=P_{\alpha_n}(t_n)\cdots P_{\alpha_1}(t_1) \tag{6}$$

注意 C_A 一般不是一个投影（因而不是一个特征）本身（除非所有的 P_{α_i} 相互对易）。特别地，当 $\rho=[\Psi]$ 是一个纯态（通过某个单位向量 $\Psi\in\mathcal{H}$ 来定义）时，我们很容易有：

$$p(\mathbb{H}_A)=\|C_A\Psi\|^2=\|P_{\alpha_n}(t_n)\cdots P_{\alpha_1}(t_1)\Psi\|^2 \tag{7}$$

当 $n=1$ 时这正好产生了玻恩规则。反过来，从玻恩规则对式(5)的推导见艾沙姆(1994)。①

2. 无论人们怎么看待量子力学的形而上学，当概率仅赋予一单个历史时无论如何它是没有意义的(除非它恰好等于0或1)。至少人们应当拥有某种类似于历史样本空间(或事件空间)的东西，它的每个(可测量的)子集被赋予某个概率从而通常的(柯尔莫哥洛夫)规则被满足。即使对于单个时间 t 和单个投影 P_α(即 $n=1$)，这都是(著名的)问题。在此情形中，该问题通过找到一自伴算符 A 而得以解决，对于此算符 P_α 是谱投影，从而样本空间被当作 A 的谱 $\sigma(A)$，有 $\alpha\subset\sigma(A)$。当然，给定 P_α，对 A 的选择绝非唯一的，不同的选择可能会引起不同的和不相容的样本空间。在实践中，人们通常开始于 A，并从其谱投影 $P_\alpha=\int_\alpha dP(\lambda)$ 中得到 P_α，给定 A 的谱求解是 $A=\int_{\mathbb{R}} dP(\lambda)\lambda$。随后，则人们要么粗粒化要么精粒化此样本空间。前者是通过找到一个划分 $\sigma(A)=\coprod_i\alpha_i$(不相交并集)，且仅承认由 α_i 生成的 σ 代数的元素、而不是 $\sigma(A)$ 的所有(可测量的)子集为事件(伴随着相关的谱投影 P_{α_i})。为了进行精粒化，人们用与 A 对易且相互对易的算符来补充 A，从而新的样本空间是继而产生的相互对易算符族的联合谱。

无论何种情形，接下来采用投影 P_α 而非样本空间的子集 α 都很简便，则上述讨论等于是将在$\mathcal{H}$上的给定投影扩展到$\mathcal{H}$上所有投影的格$\mathcal{P}(\mathcal{H})$的某个布尔子格上。② 任意态 ρ 以常规方式在此子格上定义了一个概率测度[Beltrametti and Cassinelli，1984]。

3. 将此推广到多时刻情形并不容易，在某种程度上可以通过下述设计得以简化(Isham，1994)。令$\mathcal{H}^N=\otimes^N\mathcal{H}$，其中 N 是与给定集合中历史相关的所有时刻 t_i 集合的基数，③ 且对给定历史$\mathbb{H}_A$，定义：

---

① 对玻恩规则的全新推导也见[Zurek，2005]，也有之后在[Schlosshauer，2004]中的讨论。

② 该子格被假定是$\mathcal{P}(\mathcal{H})$的单位元，即在$\mathcal{H}$上的单位算符，零投影也是。此评论也适用于下面讨论的$\mathcal{P}(\mathcal{H}^N)$的布尔子格。

③ $N=\infty$ 情形见上文的数学参考书目。

$$\mathbb{C}_A = P_{\alpha_n}(t_n)\otimes\cdots\otimes P_{\alpha_1}(t_1) \tag{8}$$

这里可以说 $P_{\alpha_i}(t_i)$ 作用于用 t_i 标示的张量积 $\mathcal{H}^N$ 中 $\mathcal{H}$ 的拷贝。注意 $\mathbb{C}_A$ 是 $\mathcal{H}^N$ 上的一个投影，而(6)中的 C_A 一般并非 $\mathcal{H}$ 上的投影。此外，如上给定 $\mathcal{H}$ 上的密度矩阵 ρ，通过下式将退相干函数 d 定义为从历史对到 $\mathbb{C}$ 的映射：

$$d(\mathbb{H}_A,\ \mathbb{H}_B) = \mathrm{Tr}(C_A\rho C_B^*) \tag{9}$$

一致性历史方法的核心论点现在可以概括如下。历史集合 $\{\mathbb{H}_A\}_{A\in\mathbb{A}}$ 可以看作是一个样本空间，在其中态 ρ 通过(5)定义了一个概率测度，它当然等于：

$$p(\mathbb{H}_A) = d(\mathbb{H}_A,\ \mathbb{H}_A) \tag{10}$$

前提是：

(a)算符 $\{\mathbb{C}_A\}_{A\in\mathbb{A}}$ 形成了 $\mathcal{H}^N$ 上所有投影点阵 $\mathcal{P}(\mathcal{H}^N)$ 的一个布尔子格；

(b) $d(\mathbb{H}_A,\ \mathbb{H}_B)$ 的实部在 $\mathbb{H}_A$ 与 $\mathbb{H}_B$ 不相交时为零。①

那样的话，集合 $\{\mathbb{H}_A\}_{A\in\mathbb{A}}$ 被称为是一致的。很重要的是认识到历史的一个给定集合的可能一致性不仅仅(平凡地)依赖于此集合，此外也依赖于动力学和初态。

历史的一致集合概括了在单个时刻的对易投影族。通过 $d(\mathbb{H}_A,\ \mathbb{H}_B)$ 本身为零来取代第二个条件并没有很大的损失，在此情形中历史 $\mathbb{H}_A$ 和 $\mathbb{H}_B$ 被称为是“退相干”。② 例如，在双缝实验中历史对 $\{\mathbb{H}_A,\ \mathbb{H}_B\}$ 不是一致的，其中 $\alpha_1=\beta_1$ 是在 t_1 时刻粒子源处粒子的发射，$\alpha_2(\beta_2)$ 是粒子在 t_2 时刻穿过上(下)缝，且 $\alpha_3=\beta_3$ 是在屏幕的某位置 L 处检测到粒子。然而，当粒子对任意一个缝的穿越在记录装置并不包含在历史中(若是的话，什么也不会得到)的情况下被记录时，它变得是一致的。这是量子测量理论中冯·诺伊曼链的联想(reminiscent)，它确实

---

① 这意味着 $\mathbb{C}_A\mathbb{C}_B=0$；等价地，对至少一个时间 t_i 有 $P_{\alpha_i}(t_i)P_{\beta_i}(t_i)=0$。此条件保证了概率(10)在不相交历史上是可加的。

② 一致历史学家们以一种不同于退相干理论学家的方式使用这一术语。根据定义，当且仅当上述条件(a)成立时，仅涉及一单个时间的任意两个历史是一致的(或确实是“退相干”)；条件(b)在那种情形中一般也满足，且仅当考虑多于一个时刻时才有意义。然而，在退相干理论中，在某个给定时刻的约化密度矩阵一般根本不“退相干”，(原先的)退相干理论中的全部意图是提供模型，在这些模型中由于系统与其环境的耦合而发生退相干(如果只是近似地)。阐述了这一点之后，在模型语境中，在多时间历史的一致性(或退相干)和约化密度矩阵的退相干之间存在着紧密的关联：使前者(近似)实现的动力学机制通常是与引起后者动力学机制相同的。

为退相干提供了一个抽象的背景(参见前一节中的第1项)。另一种选择是，该历史集可以通过忽略 α_2 和 β_2 来做到一致。关于采用一致性历史语言对双缝实验更深入的讨论见[Griffiths，2002]。

更一般地，通过简单地省略某些特征进行的粗粒化常常很有希望使给定不一致集变得一致。如果最初的历史就已经一致了，这样做绝不会使它变得不一致。另一方面，精粒化(通过嵌入到一个更大的集合)却是危险的，它可能使一致集变得不一致。

4. 它究竟意味着什么？对一致集合的每个选择定义了一个"论域(在特定论述中使用全部术语的完整集合)"，在其中人们能够应用经典概率理论和经典逻辑[Omnès，1992]。在此意义上一致性历史学家们非常忠实于哥本哈根精神(正如他们大多数人承认的)：为了理解量子世界，必须通过经典的眼镜来看它。在我们看来，目前没有可信的实例支持玻尔立场的绝对必要性(参见3.1节)，但接受它的话，一致性历史方法优于哥本哈根之处在于，它不依赖于作为量子力学解释的一个先验要素的测量。[①] 在将一个系统概念转变为一个动力学变量方面，它也比退相干方法更强有力。不同的一致集描述了不同的系统(且进而描述了不同的环境，环境定义为宇宙的剩余部分)，参见前一节的第6条。[②] 换言之，对一致集的选择归结为是对"相关变量"的选择，对比于描述所忽略的那些"不相关"变量。正如文献所强调的，对作为论域的某个一致集的辨识行为本身不过是对作为整体宇宙的一个粗粒化。

5. 但是这些概念上的成功也要付出代价。首先，一致集被证明并不存在于现实的模型中(至少如果集合中的历史包含多于一个时间变量是如此)。该理论一开始时就认识到了这一点，对此的回应是我们不得不处理近似的一致集，对它们而言 $d(\mathbb{H}_A, \mathbb{H}_B)$(的实部)只是非常小。此外，即使是关于一个历史的定义常常也不能够用投影来给出。例如，在海森堡的云室案例中(见上述第1条)，由于他自己的不确定原理，不可能写出相应的投影 P_{α_i}。很常见的替代者是$P_\alpha = Q_h^B(\chi_\Delta)$，参见式(19)和式(28)，但在式(21)看来这一算符不满足 $P_\alpha^2 =$

① 关于一致性历史与哥本哈根解释还有其他解释间联系的分析见[Hartle，2005]。

② 技术上，作为发生在给定历史中的投影的对易子。

P_α，因而它不是一个投影(虽然它满足投影 $P_\alpha^* = P_\alpha$ 的第二个定义特征)。这只是反映了任意量子化方法的常规特征 $Q(f)^2 \neq Q(f^2)$，和采用近似投影的必要性[Omnès，1997]。确实，这一点要求采用正定算符代替投影来对整个一致性历史方法进行重构[Rudolph，1996a；1996b]。

这些可能不是严重的问题，事实上，对经典性只是在近似的意义上(概念上和数学上)产生自量子理论的认识是深刻的(见引言)，倒不如把它看作是在祝愿一致历史方法目前为止已确证了它。

6. 这里可能更棘手的是，一致性决不意味着经典性具有了赋予经典概率与使用经典逻辑的能力(在给定的一致集下)。而恰恰相反，薛定谔猫态与那些每个时刻看来都是经典的却随时间完全按照非经典轨迹的历史，都是一致性条件自身所允许的[Dowker and Kent，1996]。但除了对那些仍然相信地球处于宇宙中心和/或者人类是有特权的观测者的那些人而言，这是一个真正的问题吗?它只是似乎是那样——至少在一致历史学者们看来——量子理论所展现的本体论图像是远远比我们从玻尔那里继承来的更“不人道的”(或有人会说“晦涩的”)，在大多数的一致集并不与我们观察的世界有明显的关系这一意义上。为了使这更容易被理解，这里不会诉诸“互补性”。例如，玻尔要求量子世界的互补性图像是经典的，这里经典的意义比一般的历史集更强，且除此之外玻尔要求我们只考虑两种这样的图像，这与提供给我们的无数一致集相对立。我们的结论是，尽管退相干并没有解决测量问题而是加剧了它(见前一节的第 2 条)，同样一致性历史事实上也使得解释量子力学的问题变得比以前所认为的更加困难。无论如何，毫无疑问的是一致性历史学家们显著地深化了我们对量子理论的理解——至少提供了一个好的记述方式!

7. 在识别至少一些(近似地)显示(近似的)经典行为的一致集任务中，已经取得了相当的进展，这里“经典”一词是在其充分的意义上而言的[Gell-Mann and Hartle，1993；Omnès，1992；1997；Halliwell，1998；2000；2004；Brun and Hartle，1999；Bosse and Hartle(博斯和哈特)，2005]。事实上，在我们看来这种类型的研究形成了一致性历史方法主要的具体成果。关键是找到具有如下三个决定性特征的一致集 $\{H_A\}_{A\in\mathbb{A}}$：

(a)它的元素(即历史)是具有经典解释的命题串(string)；

(b)该集合中每一个勾画出经典轨迹的历史(即对作为经典运动方程的解)拥有接近于单位1的概率(10)，同时每一个沿着经典上不可能轨迹的历史的概率接近于0;

(c)为了达到一致性描述是要充分粗粒化的，但也是要充分精粒化的，从而把其从(b)中得到的决定论运动方程变成为一个封闭系统。

当这些目标实现时，正是在这种意义上(不多也不少)，一致性历史方法可以在某些理由下宣称它表明了(或甚至是说明了)"量子宇宙是如何变为经典的"[Halliwell，2005]。

具有可认可解释的量子化的经典可观测量，如在第5项上提到的算符$Q_{\hbar}^{B}(\chi\Delta)$，以及6.1节中研究的那种宏观可观测量和流体变量，即对守恒流的空间积分，这些都是拥有经典解释的命题案例。它们表征了经典性的三个不同层次，原则上它们通过相互的精细或粗粒化相联系。① 第一种仅在极限$\hbar\rightarrow0$下被充分粗粒化以达到一致性(参见第5节)，而后两种依据它们本身的性质已经被粗粒化了。即使这样，为了达到三个目标(a)—(c)，初态也不得不在某种意义上是"经典的"。

所有这些都给人留下深刻的印象，但我们想要表达我们的观点，即退相干和一致性历史都不能单独说明经典世界的表象。虽然这些方法是有希望的，但它们至少需要与在第5与第6节中那类极限方法相结合——更不用说需要新的形而上学了！因为即使承认退相干引起了薛定谔猫类型叠加的消失，或一致历史学家们给了我们一致的集合，其中没有哪个元素包含了它们特征间的这样的叠加，但这绝不意味着它们足以说明经典相空间和在其上由经典运动方程决定的流的突现。因为，目前为止在第5与第6节中引用到的方法从未与退相干和/或一致性历史理论相结合，对经典世界从量子理论中产生的完全解释仍处于其早期阶段。这不仅在技术层面上是真的，在概念也是这样，现在做到的仅仅是一个小小的开始。在积极的方面，这为物理学基础领域中有数学思维的研究者们留下充满吸引力的挑战！

---

① 对这些联系的研究与本论文的方案相关，但实际在本质上属于经典物理学，考虑从牛顿方程对纳维叶—斯托克斯(Navier-Stokes)方程的推导。

8. 尾声

作为发人深省的结束语，我们不应该忘记无论在量子力学中识别一个“经典领域”取得了怎么样的成绩，该理论仍会包含另一个领域，即纯量子世界。那是年轻的海森堡第一个通达的地方，如果不是通过他的数学，则或许是通过他最喜欢的作曲家——贝多芬的音乐。关于视野之外的这个世界的描述，从未有比霍夫曼(Hoffmann)[1810]在他关于贝多芬器乐的文章中做得更好的，我们发现从中引用某些片段来结束本文非常合适。①

当我们要把音乐当成一门独立的艺术时，就只能谈到器乐，器乐拒绝来自于其他艺术(诗意)的协助，器乐独有的艺术特质只有从它自身当中才能表现出来。器乐是所有艺术之中最浪漫的，我们甚至可以说它是唯一真正浪漫的，因为器乐的题材是无限的。俄耳浦斯(Orpheus)的竖琴打开了地狱的大门，音乐为人类打开了一个未知的世界，这个世界与人们周围的感官世界完全不同，在这个世界里，人们要将所有他关于确定性的感觉抛诸身后，好把自己投入无可名状的渴望之中……

贝多芬的器乐为我们打开了一个神秘与不可度量的王国。闪烁的光束穿透这个王国的漫漫黑夜，我们得以窥见上下翻腾的巨大阴影。它们将我们包围起来，越来越紧，并且把我们消灭掉……贝多芬的音乐拨动了恐惧、敬畏、惊异和痛苦的杠杆，并唤醒了构成浪漫艺术特质的那种无限渴望。他之所以是一个纯粹的浪漫主义作曲家，原因盖出于此，在他看来声乐容不下不确定的渴求性格，而只是用艺术的语言来描绘确定的情绪，代之以在无限的王国里去感受……

除了极其辉煌而又深沉的c小调，又有哪一部贝多芬的器乐作品在更高的程度上证实了这一点呢？这部非凡的作品是怎样一步一步地发展，把听众不容分说地引入无限的精神王国。只有这一个作曲家真正钻进了和声的奥秘之中，所以他得以借助和声影响人们的心灵；数字的比例如果缺乏天才的语法学家不

① 翻译版权：英格丽·施韦格曼(Ingrid Schwaegermann)[2001]。中文版参见《德语国家经典散文》，叶廷芳，李伯杰主编，上海文艺出版社，2004年。译者注。

过是僵死的算术题，对于这位作曲家却是神奇的药水，他可以用它催生出一个让人心醉神迷的世界来……

无论在什么地方，只要器乐只是作为音乐通过自己而产生作用，而不是服务于某个特定的戏剧目的，就必须避开一切无关宏旨的嬉戏成分及一切调笑的噱头。欢乐的感觉来自一个未知的国度，较之我们这个处处受到限制的世界灿烂美好得多。它在我们胸中点燃了内在的、欢乐的生命。为了这些欢乐的感觉，深邃的心灵在寻找一种更高级的表述，它远远超过贫乏的语言之能事，因为这些语言仅属于有限的尘世乐趣。

致谢

作者受惠于斯特蕃·德比埃夫勒、杰里米·巴特菲尔德、狄尼斯·狄克斯、杰姆·哈特、吉斯·图尔曼、斯蒂芬·泽尔蒂奇和沃伊塞克·朱瑞克对本论文各版草稿的细致评论。最终的版本很大程度上得益于第 7 届派恩(Pine)"经典—量子交界"会议(2005 年 5 月)。作者还希望表达对李·哥里克(Lee Gohlike)和第 7 届派恩会议主委会的邀请以及其他发言者德沃尔特、哈特、海勒、特霍夫特、霍华德、古兹维勒、詹森(M. Janssen)、A. 莱格特、彭罗斯、斯塔普(P. Stamp)和朱瑞克与作者分享他们的观点的感谢。

参考文献

[Abraham and Marsden, 1985] R. Abraham and J. E. Marsden. *Foundations of Mechanics*, 2nd ed. Addison Wesley, Redwood City, 1985.

[Accardi *et al.*, 1990] L. Accardi, A. Frigerio, and Y. Lu. The weak coupling limit as a quantum functional central limit. *Communications in Mathematical Physics*, 131: 537-570, 1990.

[Adler, 2003] S. L. Adler. Why decoherence has not solved the measurement problem: A response to P. W. Anderson. *Studies in History and Philosophy of Modern Physics*, 34B: 135-142, 2003.

[Agmon, 1982] S. Agmon. *Lectures on Exponential Decay of Solutions of Second-Order*

Elliptic Equations. Princeton: Princeton University Press, 1982.

[Albeverio and Høegh-Krohn, 1976] S. A. Albeverio and R. J. Høegh-Krohn. *Mathematical Theory of Feynman Path Integrals.* Berlin: Springer-Verlag, 1976.

[Alfsen, 1970] E. M. Alfsen. *Compact Convex Sets and Boundary Integrals.* Berlin: Springer, 1970.

[Ali *et al.*, 1995] S. T. Ali, J. -P. Antoine, J. -P. Gazeau, and U. A. Mueller. Coherent states and their generalizations: A mathematical overview. *Reviews in Mathematical Physics*, 7: 1013-1104, 1995.

[Ali *et al.*, 2000] S. T. Ali, J. -P. Antoine, J. -P. Gazeau, and U. A. Mueller. *Coherent States, Wavelets and their Generalizations.* New York: Springer-Verlag, 2000.

[Ali and Emch, 1986] S. T. Ali and G. G. Emch. Geometric quantization: modular reduction theory and coherent states. *Journal of Mathematical Physics*, 27: 2936-2943, 1986.

[Ali and Englis, 2004] S. T. Ali and M. Englis. Quantization methods: a guide for physicists and analysts. 2004. arXiv: math-ph/0405065

[Alicki and Fannes, 2001] A. Alicki and M. Fannes. *Quantum Dynamical Systems.* Oxford: Oxford University Press, 2001.

[Alicki and Lendi, 1987] A. Alicki and K. Lendi. *Quantum Dynamical Semigroups and Applications.* Berlin: Springer, 1987.

[Amann, 1986] A. Amann. Observables in W^*-algebraic quantum mechanics. *Fortschritte der Physik*, 34: 167-215, 1986.

[Amann, 1987] A. Amann. Broken symmetry and the generation of classical observables in large systems. *Helvetica Physica Acta*, 60: 384-393, 1987.

[Amann and Primas, 1997] A. Amann and H. Primas. What is the referent of a non-pure quantum state? In S. Cohen, R. S., Horne, M. A., and J. Stachel (eds.), *Experimental Metaphysics: Quantum Mechanical Studies in Honor of Abner Shimony.* Dordrecht: Kluwer Academic Publishers, 1997.

[Arai, 1995] T. Arai. Some extensions of the semiclassical limit $\hbar \to 0$ for Wigner functions on phase space. *Journal of Mathematical Physics*, 36: 622-630, 1995.

[Araki, 1980] H. Araki. A remark on the Machida-Namiki theory of measurement. *Progress in Theoretical Physics*, 64: 719-730, 1980.

[Araki, 1999] H. Araki. *Mathematical Theory of Quantum Fields.* New York: Oxford University Press, 1999.

[Arnold, 1989] V. I. Arnold. *Mathematical Methods of Classical Mechanics.* Second edition. New York: Springer-Verlag, 1989.

[Ashtekar and Schilling, 1999] A. Ashtekar and T. A. Schilling. Geometrical formulation of quantum mechanics. *On Einstein's Path (New York, 1996)*, pages 23-65. New York: Springer, 1999.

[Atmanspacher *et al.*, 1989] H. Atmanspacher, A. Amann, and U. Müller-Herold (eds.). *On Quanta, Mind and Matter: Hans Primas in Context.* Dordrecht: Kluwer Academic Publishers, 1999.

[Auletta, 2001] G. Auletta. *Foundations and Interpretation of Quantum Mechanics.* Singapore: World Scientific, 2001.

[Bacciagaluppi, 1993] G. Bacciagaluppi. Separation theorems and Bell inequalities in algebraic quantum mechanics. In P. Busch, P. J. Lahti, and P. Mittelstaedt (eds.), *Proceedings of the Symposium on the Foundations of Modern Physics (Cologne, 1993)*, pages 29-37. Singapore: World Scientific, 1993.

[Bacciagaluppi, 2004] G. Bacciagaluppi. The role of decoherence in quantum theory. In E. N. Zalta (ed.), *Stanford Encyclopedia of Philosophy*, (Winter 2004 Edition), 2004. Online only at http://plato.stanford.edu/archives/win2004/entries/qm-decoherence/.

[Bach *et al.*, 1998] V. Bach, J. Fröhlich, and I. M. Sigal. Quantum electrodynamics of confined nonrelativistic particles. *Advances in Mathematics*, 137: 299-395, 1998.

[Bach *et al.*, 1999] V. Bach, J. Fröhlich, and I. M. Sigal. Spectral analysis for systems of atoms and molecules coupled to the quantized radiation field. *Communications in Mathematical Physics*, 207: 249-290, 1999.

[Baez, 1987] J. Baez. Bell's inequality for C^*-algebras. *Letters in Mathematical Physics*, 13: 135-136, 1987.

[Bagarello and Morchio, 1992] F. Bagarello and G. Morchio. Dynamics of mean-field spin models from basic results in abstract differential equations. *Journal of Statistical Physics*, 66: 849-866, 1992.

[Ballentine, 1970] L. E. Ballentine. The statistical interpretation of quantum mechanics. *Reviews of Modern Physics*, 42: 358-381, 1970.

[Ballentine, 1986] L. E. Ballentine. Probability theory in quantum mechanics. *American Journal of Physics*, 54: 883-889, 1986.

[Ballentine, 2002] L. E. Ballentine. Dynamics of quantum-classical differences for chaotic systems. *Physical Review*, A65, 062110-1-6, 2002.

[Ballentine, 2003] L. E. Ballentine. The classical limit of quantum mechanics and its implications for the foundations of quantum mechanics. In A. Khrennikov (ed.), *Quantum Theory: Reconsideration of Foundations-2*, pages 71-82. Växjö: Växjö University Press, 2003.

[Ballentine *et al.*, 1994] L. E. Ballentine, Y. Yang, and J. P. Zibin. Inadequacy of Ehrenfest's theorem to characterize the classical regime. *Physical Review*, A50: 2854-2859, 1994.

[Balian and Bloch, 1972] R. Balian and C. Bloch. Distribution of eigenfrequencies for the wave equation in a finite domain. III. Eigenfrequency density oscillations. *Annals of Physics*, 69: 76-160, 1972.

[Balian and Bloch, 1974] R. Balian and C. Bloch. Solution of the Schrödinger equation in terms of classical paths. *Annals of Physics*, 85: 514-545, 1974.

[Bambusi *et al.*, 1999] D. Bambusi, S. Graffi, and T. Paul. Long time semiclassical approximation of quantum flows: a proof of the Ehrenfest time. *Asymptotic Analysis*, 21: 149-160, 1999.

[Barrow-Green, 1997] J. Barrow-Green. *Poincaré and the Three Body Problem*. Providence, RI: American Mathematical Society, 1997.

[Barut and Raçka, 1977] A. O. Barut and R. Raçka. *Theory of Group Representations and Applications*. Warszawa: PWN, 1977.

[Bassi and Ghirardi, 2000] A. Bassi and G. C. Ghirardi. Decoherent histories and realism. *Journal of Statistical Physics*, 98: 457-494. Reply by R. B. Griffiths. *ibid.*, 99: 1409-1425, 2000. Reply to this reply by A. Bassi and G. C. Ghirardi. *ibid.*, 99: 1427, 2000.

[Bates and Weinstein, 1995] S. Bates and A. Weinstein. *Lectures on the Geometry of Quantization. Berkeley Mathematics Lecture Notes*, 8. University of California, Berkeley. Re-issued by the American Mathematical Society, 1995.

[Batterman, 2002] R. W. Batterman. *The Devil in the Details: Asymptotic Reasoning in*

Explanation, *Reduction*, *and Emergence*. Oxford: Oxford University Press, 2002.

[Batterman, 2005] R. W. Batterman. Critical phenomena and breaking drops: Infinite idealizations in physics. *Studies in History and Philosophy of Modern Physics*, 36: 225-244, 2005.

[Baum *et al.*, 1994] P. Baum, A. Connes, and N. Higson. Classifying space for proper actions and K-theory of group C^*-algebras. *Contemporary Mathematics*, 167: 241-291, 1994.

[Bayen *et al.*, 1978] F. Bayen, M. Flato, C. Fronsdal, A. Lichnerowicz, and D. Sternheimer. Deformation theory and quantization I, II. *Annals of Physics*, 110: 61-110, 111-151, 1978.

[Bell, 1975] J. S. Bell. On wave packet reduction in the Coleman-Hepp model. *Helvetica Physica Acta*, 48: 93-98, 1975.

[Bell, 1987] J. S. Bell. *Speakable and Unspeakable in Quantum Mechanics.* Cambridge: Cambridge University Press, 1987.

[Bell, 2001] J. S. Bell. *John S. Bell on the Foundations of Quantum Mechanics.* Singapore: World Scientific, 2001.

[Beller, 1999] M. Beller. *Quantum Dialogue.* Chicago: University of Chicago Press, 1999.

[Belot, 2005] G. Belot. The representation of time and change in mechanics. This volume, 2005.

[Belot and Earman, 1997] G. Belot and J. Earman. Chaos out of order: quantum mechanics, the correspondence principle and chaos. *Studies in History and Philosophy of Modern Physics*, 28B: 147-182, 1997.

[Beltrametti and Cassinelli, 1984] E. G. Beltrametti and G. Cassinelli. *The Logic of Quantum Mechanics.* Cambridge: Cambridge University Press, 1984.

[Benatti, 1993] F. Benatti. *Deterministic Chaos in Infinite Quantum Systems.* Berlin: Springer-Verlag, 1993.

[Bene and Dieks, 2002] G. Bene and D. Dieks. A perspectival version of the modal interpretation of quantum mechanics and the origin of macroscopic behavior. *Foundations of Physics*, 32: 645-671, 2002.

[Berezin, 1974] F. A. Berezin. Quantization. *Mathematical USSR Izvestia*, 8: 1109-1163, 1974.

[Berezin, 1975a] F. A. Berezin. Quantization in complex symmetric spaces. *Mathematical*

USSR Izvestia, 9: 341-379, 1975a.

[Berezin, 1975b] F. A. Berezin. General concept of quantization. *Communications in Mathematical Physics*, 40: 153-174, 1975b.

[Berry, 1977a] M. V. Berry. Semi-classical mechanics in phase space: a study of Wigner's function. *Philosophical Transactions of the Royal Society*, 287: 237-271, 1977a.

[Berry, 1977b] M. V. Berry. Regular and irregular semi-classical wavefunctions. *Journal of Physics*, A10: 2083-2091, 1977b.

[Berry *et al.*, 1979] M. V. Berry, N. L. Balazs, M. Tabor, and A. Voros. Quantum maps. *Annals of Physics*, 122: 26-63, 1979.

[Berry and Tabor, 1977] M. V. Berry and M. Tabor. Level clustering in the regular spectrum. *Proceedings of the Royal Society*, A356: 375-394, 1977.

[Berry and Keating, 1999] M. V. Berry, and J. P. Keating. The Riemann zeros and eigenvalue asymptotics. *SIAM Review*, 41: 236-266, 1999.

[Binz *et al.*, 2004] E. Binz, R. Honegger, and A. Rieckers. Field-theoretic Weyl quantization as a strict and continuous deformation quantization. *Annales Henri Poincaré*, 5: 327-346, 2004.

[Binz *et al.*, 1988] E. Binz, J. Śniatycki, and H. Fischer. *The Geometry of Classical Fields*. Amsterdam: North-Holland, 1988.

[Birkhoff and von Neumann, 1936] G. Birkhoff and J. von Neumann. The logic of quantum mechanics. *Annals of Mathematics*, (2) 37: 823-843, 1936.

[Bitbol, 1996] M. Bitbol. *Schrödinger's Philosophy of Quantum Mechanics*. Dordrecht: Kluwer Academic Publishers, 1996.

[Benatti, 1992] M. Bitbol and O. Darrigol (eds.). *Erwin Schrödinger: Philosophy and the Birth of Quantum Mechanics*. Dordrecht: Kluwer Academic Publishers, 1992.

[Blackadar, 1998] B. Blackadar. *K-Theory for Operator Algebras*. Second edition. Cambridge: Cambridge University Press, 1998.

[Blair Bolles, 2004] E. Blair Bolles. *Einstein Defiant: Genius versus Genius in the Quantum Revolution*. Washington: Joseph Henry Press, 2004.

[Blanchard, 1996] E. Blanchard. Deformations de C^*-algebras de Hopf. *Bulletin de la Sociétémathématique de France*, 124: 141-215, 1996.

[Blanchard and Olkiewicz, 2003] Ph. Blanchard and R. Olkiewicz. Decoherence induced transition from quantum to classical dynamics. *Reviews in Mathematical Physics*, 15: 217-243, 2003.

[Bohigas *et al.*, 1984] O. Bohigas, M. -J. Giannoni, and C. Schmit. Characterization of chaotic quantum spectra and universality of level fluctuation laws. *Physical Review Letters*, 52: 1-4, 1984.

[Blume-Kohout and Zurek, 2004] R. Blume-Kohout and W. H. Zurek. A simple example of "Quantum Darwinism": Redundant information storage in many-spin environments. *Foundations of Physics*, to appear, 2004. arXiv: quant-ph/0408147

[Blume-Kohout and Zurek, 2005] R. Blume-Kohout and W. H. Zurek. Quantum Darwinism: Entanglement, branches, and the emergent classicality of redundantly stored quantum information. *Physical Review* A, to appear, 2005. arXiv: quant-ph/0505031

[Bogoliubov, 1958] N. N. Bogoliubov. On a new method in the theory of superconductivity. *Nuovo Cimento*, 7: 794-805, 1958.

[Bohr, 1927] N. Bohr. The quantum postulate and the recent development of atomic theory. *Atti del Congress Internazionale dei Fisici* (*Como*, 1927). Reprinted in Bohr (1934), pages 52-91, 1927.

[Bohr, 1934] N. Bohr. *Atomic Theory and the Description of Nature.* Cambridge: Cambridge University Press, 1934.

[Bohr, 1935] N. Bohr. Can quantum-mechanical description of physical reality be considered complete? *Physical Review*, 48: 696-702, 1935.

[Bohr, 1937] N. Bohr. Causality and complementarity. *Philosophy of Science*, 4: 289-298, 1937.

[Bohr, 1948] N. Bohr. On the notions of causality and complementarity. *Dialectica*, 2: 312-319, 1948.

[Bohr, 1949] N. Bohr. Discussion with Einstein on epistemological problems in atomic physics. In P. A. Schlipp (ed.), *Albert Einstein: Philosopher-Scientist*, pages 201-241. La Salle: Open Court, 1949.

[Bohr, 1958] N. Bohr. *Atomic Physics and Human Knowlegde.* New York: Wiley, 1958.

[Bohr, 1985] N. Bohr. In J. Kalckar (ed.), *Collected Works. Vol.* 6: *Foundations of*

Quantum Physics I (1926-1932). Amsterdam: North-Holland, 1985.

[Bohr, 1996] N. Bohr. In J. Kalckar (ed.), *Collected Works. Vol. 7: Foundations of Quantum Physics II* (1933-1958). Amsterdam: North-Holland, 1996.

[Bona, 1980] P. Bona. A solvable model of particle detection in quantum theory. *Acta Facultatis Rerum Naturalium Universitatis Comenianae Physica*, XX: 65-94, 1980.

[Bona, 1988] P. Bona. The dynamics of a class of mean-field theories. *Journal of Mathematical Physics*, 29: 2223-2235, 1988.

[Bona, 1989] P. Bona. Equilibrium states of a class of mean-field theories. *Journal of Mathematical Physics*, 30: 2994-3007, 1989.

[Bona, 2000] P. Bona. Extended quantum mechanics. *Acta Physica Slovaca*, 50: 1-198, 2000.

[Bonechi and De Bièvre, 2000] F. Bonechi and S. De Bièvre. Exponential mixing and ln $\hbar$ time scales in quantized hyperbolic maps on the torus. *Communications in Mathematical Physics*, 211: 659-686, 2000.

[Bosse and Hartle, 2005] A. W. Bosse and J. B. Hartle. Representations of spacetime alternatives and their classical limits. 2005. arXiv: quant-ph/0503182.

[Brack and Bhaduri, 1997] M. Brack and R. K. Bhaduri. *Semiclassical Physics*. Boulder: Westview Press, 1997.

[Bratteli and Robinson, 1987] O. Bratteli and D. W. Robinson. *Operator Algebras and Quantum Statistical Mechanics. Vol. I: C^*-and W^*-Algebras, Symmetry Groups, Decomposition of States.* 2nd Ed. Berlin: Springer, 1987.

[Bratteli and Robinson, 1981] O. Bratteli and D. W. Robinson. *Operator Algebras and Quantum Statistical Mechanics. Vol. II: Equilibrium States, Models in Statistical Mechanics.* Berlin: Springer, 1981.

[Brezger *et al.*, 2002] B. Brezger, L. Hackermüller, S. Uttenthaler, J. Petschinka, M. Arndt, and A. Zeilinger. Matter-Wave Interferometer for Large Molecules. *Physical Review Letters*, 88, 100404, 2002.

[Breuer, 1994] T. Breuer. *Classical Observables, Measurement, and Quantum Mechanics.* Ph. D. Thesis, University of Cambridge, 1994.

[Bröcker and Werner, 1995] T. Bröcker and R. F. Werner. Mixed states with positive Wigner functions. *Journal of Mathematical Physics*, 36: 62-75, 1995.

[Brun and Hartle, 1999] T. A. Brun and J. B. Hartle. Classical dynamics of the quantum harmonic chain. *Physical Review*, D60, 123503-1-20, 1999.

[Brush, 2002] S. G. Brush. Cautious revolutionaries: Maxwell, Planck, Hubble. *American Journal of Physics*, 70: 119-127, 2002.

[Bub, 1988] J. Bub. How to Solve the Measurement Problem of Quantum Mechanics. *Foundations of Physics*, 18: 701-722, 1988.

[Bub, 1999] J. Bub. *Interpreting the Quantum World.* Cambridge: Cambridge University Press, 1999.

[Bub, 2004] J. Bub. Why the quantum? *Studies in History and Philosophy of Modern Physics*, 35B: 241-266, 2004.

[Busch *et al.*, 1998] P. Busch, M. Grabowski, and P. J. Lahti. *Operational Quantum Physics*, 2nd corrected ed. Berlin: Springer, 1998.

[Busch *et al.*, 1991] P. Busch, P. J. Lahti, and P. Mittelstaedt. *The Quantum Theory of Measurement.* Berlin: Springer, 1991.

[Butterfield, 2002] J. Butterfield. Some Worlds of Quantum Theory. In R. Russell, J. Polkinghorne *et al.* (ed.). *Quantum Mechanics* (Scientific Perspectives on Divine Action vol 5), pages 111-140. Rome: Vatican Observatory Publications, 2, 2002. arXiv: quant-ph/0105052; PITT-PHIL-SCI00000204.

[Butterfield, 2005] J. Butterfield. On symmetry, conserved quantities and symplectic reduction in classical mechanics. *This volume*, 2005.

[Büttner *et al.*, 2003] L. Büttner, J. Renn, and M. Schemmel. Exploring the limits of classical physics: Planck, Einstein, and the structure of a scientific revolution. *Studies in History and Philosophy of Modern Physics*, 34B: 37-60, 2003.

[Camilleri, 2005] K. Camilleri. *Heisenberg and Quantum Mechanics: The Evolution of a Philosophy of Nature.* Ph. D. Thesis, University of Melbourne, 2005.

[Cantoni, 1975] V. Cantoni. Generalized "transition probability". *Communications in Mathematical Physics*, 44: 125-128, 1975.

[Cantoni, 1977] V. Cantoni. The Riemannian structure on the states of quantum-like systems. *Communications in Mathematical Physics*, 56: 189-193, 1977.

[Carson, 2000] C. Carson. Continuities and discontinuities in Planck's *Akt der Verzweiflung*. *Annalen der Physik*, 9: 851-960, 2000.

[Cassidy, 1992] D. C. Cassidy. *Uncertainty: the Life and Science of Werner Heisenberg*. New York: Freeman, 1992.

[Castrigiano and Henrichs, 1980] D. P. L. Castrigiano and R. W. Henrichs. Systems of covariance and subrepresentations of induced representations. *Letters in Mathematical Physics*, 4: 169-175, 1980.

[Cattaneo, 1979] U. Cattaneo. On Mackey's imprimitivity theorem. *Commentari Mathematici Helvetici*, 54: 629-641, 1979.

[Caves *et al.*, 2002] C. M. Caves, C. A. Fuchs, and R. Schack. Unknown quantum states: the quantum de Finetti representation. Quantum information theory. *Journal of Mathematical Physics*, 43: 4537-4559, 2002.

[Charbonnel, 1986] A. M. Charbonnel. Localisation et développement asymptotique deséléments du spectre conjoint d' opérateurs psuedodifférentiels qui commutent. *Integral Equations Operator Theory*, 9: 502-536, 1986.

[Charbonnel, 1988] A. M. Charbonnel. Comportement semi-classiques du spectre conjoint d' opérateurs psuedodifférentiels qui commutent. *Asymptotic Analysis*, 1: 227-261, 1988.

[Charbonnel, 1992] A. M. Charbonnel. Comportement semi-classiques des systémes ergodiques. *Annales de l' Institut Henri Poincaré - Physique Théorique*, 56: 187-214, 1992.

[Chernoff, 1973] P. R. Chernoff. Essential self-adjointness of powers of generators of hyperbolic equations. *Journal of Functional Analysis*, 12: 401-414, 1973.

[Chernoff, 1995] P. R. Chernoff. Irreducible representations of infinite dimensional transformation groups and Lie algebras I. *Journal of Functional Analysis*, 130: 255-282, 1995.

[Chevalley, 1991] C. Chevalley. Introduction: Le dessin et la couleur. In E. Bauer and R. Omnès (eds.), *Niels Bohr: Physique Atomique et Connaissance Humaine.* (French translation of Bohr, 1958), pages 17-140. Paris: Gallimard, 1991.

[Chevalley, 1999] C. Chevalley. Why do we find Bohr obscure? In D. Greenberger, W. L. Reiter, and A. Zeilinger (eds.), *Epistemological and Experimental Perspectives on Quantum Physics*, pages 59-74. Dordrecht: Kluwer Academic Publishers, 1999.

[Chiorescu *et al.*, 2003] I. Chiorescu, Y. Nakamura, C. J. P. M. Harmans, and J. E. Mooij.

Coherent quantum dynamics of a superconducting flux qubit. *Science*, 299, Issue 5614: 1869-1871, 2003.

[Cirelli *et al.*, 1983] R. Cirelli, P. Lanzavecchia, and A. Mania. Normal pure states of the von Neumann algebra of bounded operators as a Kähler manifold. *Journal of Physics*, A16: 3829-3835, 1983.

[Cirelli *et al.*, 1990] R. Cirelli, A. Manià, and L. Pizzocchero. Quantum mechanics as an infinitedimensional Hamiltonian system with uncertainty structure. I, II. *Journal of Mathematical Physics*, 31: 2891-2897, 2898-2903, 1990.

[Colin de Verdière, 1973] Y. Colin de Verdière. Spectre du laplacien et longueurs des géodésiques périodiques. I, II. *Compositio Mathematica*, 27: 83-106, 159-184, 1973.

[Colin de Verdière, 1977] Y. Colin de Verdière. Quasi-modes sur les variétés Riemanniennes. *Inventiones Mathematicae* 43, 15-52, 1977.

[Colin de Verdière, 1985] Y. Colin de Verdière. Ergodicitéet fonctions propres du Laplacien. *Communications in Mathematical Physics*, 102: 497-502, 1985.

[Colin de Verdière, 1998] Y. Colin de Verdière. Une introduction la mcanique semi-classique. *l'Enseignement Mathematique*, 44(2): 23-51, 1998.

[Combescure, 1992] M. Combescure. The squeezed state approach of the semiclassical limit of the time-dependent Schrödinger equation. *Journal of Mathematical Physics*, 33: 3870-3880, 1992.

[Combescure *et al.*, 1999] M. Combescure, J. Ralston, and D. Robert. A proof of the Gutzwiller semiclassical trace formula using coherent states decomposition. *Communications in Mathematical Physics*, 202: 463-480, 1999.

[Combescure and Robert, 1997] M. Combescure and D. Robert. Semiclassical spreading of quantum wave packets and applications near unstable fixed points of the classical flow. *Asymptotic Analysis*, 14: 377-404, 1997.

[Corwin and Greenleaf, 1989] L. Corwin and F. P. Greenleaf. *Representations of Nilpotent Lie Groups and Their Applications*, Part I. Cambridge: Cambridge University Press, 1989.

[Cucchietti, 2004] F. M. Cucchietti. *The Loschmidt Echo in Classically Chaotic Systems: Quantum Chaos, Irreversibility and Decoherence.* Ph. D. Thesis, Universidad Nacional de Córdobo, 2004. arXiv: quant-ph/0410121

[Cushing, 1994] J. T. Cushing. *Quantum Mechanics: Historical Contingency and the Copenhagen Hegemony.* Chicago: University of Chicago Press, 1994.

[Cvitanovic *et al.*, 2005] P. Cvitanovic *et al.*. *Classical and Quantum Chaos*, 2005. http://ChaosBook.org.

[Cycon *et al.*, 1987] H. L. Cycon, R. G. Froese, W. Kirsch, and B. Simon. *Schrödinger Operators with Application to Quantum Mechanics and Global Geometry.* Berlin: Springer-Verlag, 1987.

[Darrigol, 1992] O. Darrigol. *From c-Numbers to q-Numbers.* Berkeley: University of California Press, 1992.

[Darrigol, 2001] O. Darrigol. The Historians' Disagreements over the Meaning of Planck's Quantum. *Centaurus*, 43: 219-239, 2001.

[Davidson, 2001] D. Davidson. *Subjective, Intersubjective, Objective.* Oxford: Clarendon Press, 2001.

[Davies, 1976] E. B. Davies. *Quantum Theory of Open Systems.* London: Academic Press, 1976.

[De Bièvre, 1992] S. De Bièvre. Oscillator eigenstates concentrated on classical trajectories. *Journal of Physics*, A25: 3399-3418, 1992.

[De Bièvre, 2001] S. De Bièvre. Quantum chaos: a brief first visit. *Contemporary Mathematics*, 289: 161-218, 2001.

[De Bièvre, 2003] S. De Bièvre. Local states of free bose fields. Lectures given at the Summer School on Large Coulomb Systems, Nordfjordeid, 2003.

[De Bièvre *et al.*, 1993] S. De Bièvre, M. Irac-Astaud, and J. C. Houard. Wave packets localised on closed classical trajectories. In W. F. Ames, E. M. Harrell, and J. V. Herod (eds.). *Differential Equations and Applications in Mathematical Physics*, pages 25-33. New York: Academic Press, 1993.

[De Muynck, 2002] W. M. De Muynck. *Foundations of Quantum Mechanics: an Empiricist Approach.* Dordrecht: Kluwer Academic Publishers, 2002.

[Devoret *et al.*, 2004] M. H. Devoret, A. Wallraff, and J. M. Martinis. Superconducting Qubits: A Short Review. 2004. arXiv: cond-mat/0411174

[Diacu and Holmes, 1996] F. Diacu and P. Holmes. *Celestial Encounters. The Origins of*

Chaos and Stability. Princeton: Princeton University Press, 1996.

[Dickson, 2005] M. Dickson. Non-relativistic quantum mechanics. *This Volume*, 2005.

[Dieks, 1989a] D. Dieks. Quantum mechanics without the projection postulate and its realistic interpretation. *Foundations of Physics*, 19: 1397-1423, 1989a.

[Dieks, 1989b] D. Dieks. Resolution of the measurement problem through decoherence of the quantum state. *Physics Letters*, 142A: 439-446, 1989b.

[Dimassi and Sjöstrand, 1999] M. Dimassi and J. Sjöstrand. *Spectral Asymptotics in the Semi-Classical Limit*. Cambridge: Cambridge University Press, 1999.

[Dirac, 1926] P. A. M. Dirac. The fundamental equations of quantum mechanics. *Proceedings of the Royal Society*, A109: 642-653, 1926.

[Dirac, 1930] P. A. M. Dirac. *The Principles of Quantum Mechanics*. Oxford: Clarendon Press, 1930.

[Dirac, 1964] P. A. M. Dirac. *Lectures on Quantum Mechanics*. New York: Belfer School of Science, Yeshiva University, 1964.

[Dixmier, 1977] J. Dixmier. *C^*-Algebras*. Amsterdam: North-Holland, 1977.

[Doebner and Tolar, 1975] H. D. Doebner and J. Tolar. Quantum mechanics on homogeneous spaces. *Journal of Mathematical Physics*, 16: 975-984, 1975.

[Dowker and Kent, 1996] F. Dowker and A. Kent. On the consistent histories Approach to Quantum Mechanics. *Journal of Statistical Physics*, 82: 1575-1646, 1996.

[Dubin *et al.*, 2000] D. A. Dubin, M. A. Hennings, and T. B. Smith. *Mathematical Aspects of Weyl Quantization and Phase*. Singapore: World Scientific, 2000.

[Duclos and Hogreve, 1993] P. Duclos and H. Hogreve. On the semiclassical localization of the quantum probability. *Journal of Mathematical Physics*, 34: 1681-1691, 1993.

[Duffield, 1990] N. G. Duffield. Classical and thermodynamic limits for generalized quantum spin systems. *Communications in Mathematical Physics*, 127: 27-39, 1990.

[Duffield and Werner, 1992a] N. G. Duffield and R. F. Werner. Classical Hamiltonian dynamics for quantum Hamiltonian mean-field limits. In A. Truman and I. M. Davies (eds.), *Stochastics and Quantum Mechanics: Swansea, Summer* 1990, pages 115-129. Singapore: World Scientific, 1992a.

[Duffield and Werner, 1992b] N. G. Duffield and R. F. Werner. On mean-field dynamical semigroups on C^*-algebras. *Reviews in Mathematical Physics*, 4: 383-424, 1992b.

[Duffield and Werner, 1992c] N. G. Duffield and R. F. Werner. Local dynamics of mean-field quantum systems. *Helvetica Physica Acta*, 65: 1016-1054, 1992c.

[Duffield *et al.*, 1992] N. G. Duffield, H. Roos, and R. F. Werner. Macroscopic limiting dynamics of a class of inhomogeneous mean field quantum systems. *Annales de l'Institut Henri Poincaré-Physique Théorique*, 56: 143-186, 1992.

[Duffner and Rieckers, 1988] E. Duffner and A. Rieckers. On the global quantum dynamics of multilattice systems with nonlinear classical effects. *Zeitschrift für Naturforschung*, A43: 521-532, 1988.

[Duistermaat, 1974] J. J. Duistermaat. Oscillatory integrals, Lagrange immersions and unfolding of singularities. *Communications in Pure and Applied Mathematics*, 27: 207-281, 1974.

[Duistermaat and Guillemin, 1975] J. J. Duistermaat and V. Guillemin. The spectrum of positive elliptic operators and periodic bicharacteristics. *Inventiones Mathematicae*, 29: 39-79, 1975.

[Duistermaat, 1996] J. J. Duistermaat. *Fourier Integral Operators.* Original Lecture Notes from 1973. Basel: Birkhäuser, 1996.

[Duval *et al.*, 1991] C. Duval, J. Elhadad, M. J. Gotay, J. Śniatycki, and G. M. Tuynman. Quantization and bosonic BRST theory. *Annals of Physics*, 206: 1-26, 1991.

[Earman, 1986] J. Earman. *A Primer on Determinism.* Dordrecht: Reidel, 1986.

[Earman, 2005] J. Earman. Aspects of determinism in modern physics. *This volume*, 2005.

[Earman, to appear] J. Earman. Essential self-adjointness: implications for determinism and the classical-quantum correspondence. *Synthese*, to appear.

[Echterhoff *et al.*, 2002] S. Echterhoff, S. Kaliszewski, J. Quigg, and I. Raeburn. A categorical approach to imprimitivity theorems for C^*-dynamical systems, 2002. arXiv: math. OA/0205322

[Eddington, 1920] A. S. Eddington. *Space, Time, and Gravitation: An Outline of the General Relativity Theory.* Cambridge: Cambridge University Press, 1920.

[Effros and Hahn, 1967] E. G. Effros and F. Hahn. Locally compact transformation groups and C^*-algebras. *Memoirs of the American Mathematical Society*, 75, 1967.

[Ehrenfest, 1927] P. Ehrenfest. Bemerkung über die angenäherte Gultigkeit der klassischen

Mechanik innerhalb der Quantenmechanik. *Zeitschrift für Physik*, 45, 455-457, 1927.

[Einstein, 1905] A. Einstein. Über einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtpunkt. *Annalen der Physik*, 17: 132-178, 1905.

[Einstein, 1917] A. Einstein. Zum Quantensatz von Sommerfeld und Epstein. *Verhandlungen der deutschen Physikalischen Geselschaft* (2), 19: 82-92, 1917.

[Einstein, 1949] A. Einstein. Remarks to the essays appearing in this collective volume. (Reply to criticisms). In P. A. Schilpp (ed.), *Albert Einstein: Philosopher-Scientist*, pages 663-688. La Salle: Open Court, 1949.

[Elliott *et al.*, 1993] G. A. Elliott, T. Natsume, and R. Nest. The Heisenberg group and *K*-theory. *K-Theory*, 7: 409-428, 1993.

[Emch and Knops, 1970] G. G. Emch and H. J. F. Knops. Pure thermodynamical phases as extremal KMS states. *Journal of Mathematical Physics*, 11: 3008-3018, 1970.

[Emch, 1984] G. G. Emch. *Mathematical and conceptual foundations of 20th-century physics.* Amsterdam: North-Holland, 1984.

[Emch and Liu, 2002] G. G. Emch and C. Liu. *The Logic of Thermostatistical Physics.* Berlin: Springer-Verlag, 2002.

[Emch *et al.*, 1994] G. G. Emch, H. Narnhofer, W. Thirring, and G. Sewell. Anosov actions on noncommutative algebras. *Journal of Mathematical Physics*, 35: 5582-5599, 1994.

[d'Espagnat, 1995] B. d'Espagnat. *Veiled Reality: An Analysis of Present-Day Quantum Mechanical Concepts.* Reading (MA): Addison-Wesley, 1995.

[Esposito *et al.*, 2004] G. Esposito, G. Marmo, and G. Sudarshan. *From Classical to Quantum Mechanics: An Introduction to the Formalism, Foundations and Applications.* Cambridge: Cambridge University Press, 2004.

[Enz, 2002] C. P. Enz. *No Time to be Brief: A Scientific Biography of Wolfgang Pauli.* Oxford: Oxford University Press, 2002.

[Everett, 1957] H. Everett III. "Relative state" formulation of quantum mechanics. *Reviews in Modern Physics*, 29: 454-462, 1957.

[Faye, 1991] J. Faye. *Niels Bohr: His Heritage and Legacy. An Anti-Realist View of Quantum Mechanics.* Dordrecht: Kluwer Academic Publishers, 1991.

[Faye, 2002] J. Faye. Copenhagen interpretation of quantum mechanics. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2002 Edition), 2002. http://plato. stanford. edu/archives/sum2002/entries/qm-copenhagen/

[Faye and Folse, 1994] J. Faye and H. Folse (eds.). *Niels Bohr and Contemporary Philosophy.* Dordrecht: Kluwer Academic Publishers, 1994.

[Fell and Doran, 1988] J. M. G. Fell and R. S. Doran. *Representations of* -Algebras, Locally Compact Groups and Banach * -Algebraic Bundles*, Vol. 2. Boston: Academic Press, 1988.

[Feyerabend, 1981] P. Feyerabend. Niels Bohr's world view. *Realism, Rationalism and Scientific Method: Philosophical Papers*, Vol. 1, pages 247-297. Cambridge: Cambridge University Press.

[Fleming and Butterfield, 2000] G. Fleming and J. Butterfield. Strange positions. In J. Butterfield and C. Pagonis (eds.), *From Physics to Philosophy*, pages 108-165. Cambridge: Cambridge University Press, 2000.

[Folse, 1985] H. J. Folse. *The Philosophy of Niels Bohr.* Amsterdam: North-Holland, 1985.

[Ford, 1988] J. Ford. Quantum chaos. Is there any? In H. Bai-Lin (ed.), *Directions in Chaos*, Vol. 2, pages 128-147. Singapore: World Scientific, 1988.

[Frasca, 2003] M. Frasca. General theorems on decoherence in the thermodynamic limit. *Physics Letters*, A308: 135-139, 2003.

[Frasca, 2004] M. Frasca. Fully polarized states and decoherence. 2004. arXiv: cond-mat/0403678

[Frigerio, 1974] A. Frigerio. Quasi-local observables and the problem of measurement in quantum mechanics. *Annales de l' Institut Henri Poincaré*, A3: 259-270, 1974.

[Fröhlich *et al.*, 2002] J. Fröhlich, T. -P. Tsai, and H. -T. Yau. On the point-particle (Newtonian) limit of the non-linear Hartree equation. *Communications in Mathematical Physics*, 225: 223-274, 2002.

[Gallavotti, 1983] G. Gallavotti. *The Elements of Mechanics.* Berlin: Springer-Verlag, 1983.

[Gallavotti *et al.*, 2004] G. Gallavotti, F. Bonetto, and G. Gentile. *Aspects of Ergodic, Qualitative and Statistical Theory of Motion.* New York: Springer, 2004.

[Gell-Mann and Hartle, 1990] M. Gell-Mann and J. B. Hartle. Quantum mechanics in the light of quantum cosmology. In W. H. Zurek (ed.) *Complexity, Entropy, and the Physics of Information*, pages 425-458. Reading, Addison-Wesley, 1990.

[Gell-Mann and Hartle, 1993] M. Gell-Mann and J. B. Hartle. Classical equations for quantum systems. *Physical Review*, D47: 3345-3382, 1993.

[Gérard and Leichtnam, 1993] P. Gérard and E. Leichtnam. Ergodic properties of eigenfunctions for the Dirichlet problem. *Duke Mathematical Journal*, 71: 559-607, 1993.

[Gerisch *et al.*, 1999] T. Gerisch, R. Münzner, and A. Rieckers. Global C^*-dynamics and its KMS states of weakly inhomogeneous bipolaronic superconductors. *Journal of Statistical Physics*, 97: 751-779, 1999.

[Gerisch *et al.*, 2003] T. Gerisch, R. Honegger, and A. Rieckers. Algebraic quantum theory of the Josephson microwave radiator. *Annales Henri Poincaré*, 4: 1051-1082, 2003.

[Geyer *et al.*, 1993] B. Geyer, H. Herwig, and H. Rechenberg (eds.). *Werner Heisenberg: Physiker und Philosoph.* Leipzig: Spektrum, 1993.

[Giulini, 2003] D. Giulini. Superselection rules and symmetries. In E. Joos *et al.* (eds.), *Decoherence and the Appearance of a Classical World in Quantum Theory*, pages 259-316. Berlin: Springer, 2003.

[Glimm and Jaffe, 1987] J. Glimm and A. Jaffe. *Quantum Physics. A Functional Integral Point of View.* New York: Springer-Verlag, 1987.

[Gotay, 1986] M. J. Gotay. Constraints, reduction, and quantization. *Journal of Mathematical Physics*, 27: 2051-2066, 1986.

[Gotay, 1999] M. J. Gotay. On the Groenewold-Van Hove problem for R^{2n}. *Journal of Mathematical Physics*, 40: 2107-2116, 1999.

[Gotay *et al.*, 1996] M. J. Gotay, H. B. Grundling, and G. M. Tuynman. Obstruction results in quantization theory. *Journal of Nonlinear Science*, 6: 469-498, 1996.

[Gotay *et al.*, 1978] M. J. Gotay, J. M. Nester, and G. Hinds. Presymplectic manifolds and the Dirac-Bergmann theory of constraints. *Journal of Mathematical Physics*, 19: 2388-2399, 1978.

[Götsch, 1992] J. Götsch. *Erwin Schrödinger's World View: The Dynamics of Knowlegde and Reality.* Dordrecht: Kluwer Academic Publishers, 1992.

[Govaerts, 1991] J. Govaerts. *Hamiltonian Quantization and Constrained Dynamics.* Leuven: Leuven University Press, 1991.

[Gracia-Bondía *et al.*, 2001] J. M. Gracia-Bondía, J. C. Várilly, and H. Figueroa. *Elements* *of Noncommutative Geometry*. Boston: Birkhäuser, 2001.

[Griesemer *et al.*, 2001] M. Griesemer, E. H. Lieb, and M. Loss. Ground states in non-relativistic quantum electrodynamics. *Inventiones Mathematicae*, 145: 557-595, 2001.

[Griffiths, 1984] R. B. Griffiths. Consistent histories and the interpretation of quantum mechanics. *Journal of Statistical Physics*, 36: 219-272, 1984.

[Griffiths, 2002] R. B. Griffiths. *Consistent Quantum Theory*. Cambridge: Cambridge University Press, 2002.

[Grigis and Sjöstrand, 1994] A. Grigis and J. Sjöstrand. *Microlocal Analysis for Differential Operators*. Cambridge: Cambridge University Press, 1994.

[Groenewold, 1946] H. J. Groenewold. On the principles of elementary quantum mechanics. *Physica*, 12: 405-460, 1946.

[Groenewold, 1952] H. J. Groenewold. Information in quantum measurements. *Proceedings Koninklijke Nederlandse Akademie van Wetenschappen*, B55: 219-227, 1952.

[Guhr *et al.*, 1998] T. Guhr, H. Müller-Groeling, and H. Weidenmüller. Random matrix theories in quantum physics: common concepts. *Physics Reports*, 299: 189-425, 1998.

[Guillemin *et al.*, 2002] V. Guillemin, V. Ginzburg, and Y. Karshon. *Moment Maps, Cobordisms, and Hamiltonian Group Actions*. Providence (RI): American Mathematical Society, 2002.

[Guillemin and Sternberg, 1977] V. Guillemin and S. Sternberg. *Geometric Asymptotics*. Providence (RI): American Mathematical Society, 1977.

[Guillemin and Sternberg, 1990] V. Guillemin and S. Sternberg. *Variations on a Theme by Kepler*. Providence (RI): American Mathematical Society, 1990.

[Guillemin and Uribe, 1989] V. Guillemin and A. Uribe. Circular symmetry and the trace formula. *Inventiones Mathematicae*, 96: 385-423, 1989.

[Gustafson and Sigal, 2003] S. J. Gustafson and I. M. Sigal. *Mathematical concepts of quantum mechanics*. Berlin: Springer, 2003.

[Gutzwiller, 1971] M. C. Gutzwiller. Periodic orbits and classical quantization conditions. *Journal of Mathematical Physics*, 12: 343-358, 1971.

[Gutzwiller, 1990] M. C. Gutzwiller. *Chaos in Classical and Quantum Mechanics*. New York: Springer-Verlag, 1990.

[Gutzwiller, 1992] M. C. Gutzwiller. Quantum chaos. *Scientific American*, 266: 78-84, 1992.

[Gutzwiller, 1998] M. C. Gutzwiller. Resource letter ICQM-1: The interplay between classical and quantum mechanics. *American Journal of Physics*, 66: 304-324, 1998.

[Haag, 1962] R. Haag. The mathematical structure of the Bardeen-Cooper-Schrieffer model. *Nuovo Cimento*, 25: 287-298, 1962.

[Haag *et al.*, 1970] R. Haag, R. Kadison, and D. Kastler. Nets of C^*-algebras and classification of states. *Communications in Mathematical Physics*, 16: 81-104, 1970.

[Haag, 1992] R. Haag. *Local Quantum Physics: Fields, Particles, Algebras.* Heidelberg: Springer-Verlag, 1992.

[Haake, 2001] F. Haake. *Quantum Signatures of Chaos.* Second Edition. New York: Springer-Verlag, 2001.

[Hagedorn, 1998] G. A. Hagedorn. Raising and lowering operators for semiclassical wave packets. *Annals of Physics*, 269: 77-104, 1998.

[Hagedorn and Joye, 1999] G. A. Hagedorn and A. Joye. Semiclassical dynamics with exponentially small error estimates. *Communications in Mathematical Physics*, 207: 439-465, 1999.

[Hagedorn, 2000] G. A. Hagedorn and A. Joye. Exponentially accurate semiclassical dynamics: propagation, localization, Ehrenfest times, scattering, and more general states. *Annales Henri Poincaré*, 1: 837-883, 2000.

[Halliwell, 1998] J. J. Halliwell. Decoherent histories and hydrodynamic equations. *Physical Review*, D58, 105015-1-12, 1998.

[Halliwell, 2000] J. J. Halliwell. The emergence of hydrodynamic equations from quantum theory: a decoherent histories analysis. *International Journal of Theoretical Physics*, 39: 1767-1777, 2000.

[Halliwell, 2004] J. J. Halliwell. Some recent developments in the decoherent histories approach to quantum theory. *Lecture Notes in Physics*, 633: 63-83, 2004.

[Halliwell, 2005] J. J. Halliwell. How the quantum Universe becomes classical. 2005. arXiv: quant-ph/0501119

[Halvorson, 2004] H. Halvorson. Complementarity of representations in quantum mechanics. *Studies in History and Philosophy of Modern Physics*, B35: 45-56, 2004.

[Halvorson, 2005] H. Halvorson. Algebraic quantum field theory. This volume, 2005.

[Halvorson and Clifton, 1999] H. Halvorson and R. Clifton. Maximal beable subalgebras of quantum-mechanical observables. *International Journal of Theoretical Physics*, 38: 2441-2484, 1999.

[Halvorson and Clifton, 2002] H. Halvorson and R. Clifton. Reconsidering Bohr's reply to epr. In T. Placek and J. Butterfield (eds.), *Non-locality and Modality*, pages 3-18. Dordrecht: Kluwer Academic Publishers, 2002.

[Hannabuss, 1984] K. C. Hannabuss. Dilations of a quantum measurement. *Helvetica Physica Acta*, 57: 610-620, 1984.

[Harrison and Wan, 1997] F. E. Harrison and K. K. Wan. Macroscopic quantum systems as measuring devices: dc SQUIDs and superselection rules. *Journal of Physics*, A30: 4731-4755, 1997.

[Hartle, 1995] J. B. Hartle. Spacetime quantum mechanics and the quantum mechanics of spacetime. *Gravitation et Quantifications* (*Les Houches*, 1992), page 285-480. Amsterdam: North-Holland, 1995.

[Hartle, 2005] J. B. Hartle. What connects different interpretations of quantum mechanics? In A. Elitzur, S. Dolev, and N. Kolenda (eds.). *Quo Vadis Quantum Mechanics*, pages 73-82. Heidelberg: Springer-Verlag, 2005. arXiv: quant-ph/0305089

[Hawkins, 2005] E. Hawkins. Quantization of multiply connected manifolds. *Communications in Mathematical Physics*, 255: 513-575, 2005.

[Heath and Sudderth, 1976] D. Heath and W. Sudderth. De Finetti's theorem on exchangeable variables. *American Statistics*, 30: 188-189, 1976.

[Heelan, 1965] P. Heelan. *Quantum Mechanics and Objectivity: A Study of the Physical Philosophy of Werner Heisenberg*. Den Haag: Martinus Nijhoff, 1965.

[Heilbron, 2000] Heilbron, J. *The Dilemmas of an Upright Man: Max Planck as a Spokesman for German Science*. Second Edition. Los Angeles: University of California Press, 2000.

[Heisenberg, 1925] W. Heisenberg. Über die quantentheoretische Umdeutung kinematischer und mechanischer Beziehungen. *Zeitschrift für Physik*, 33: 879-893, 1925.

[Heisenberg, 1927] W. Heisenberg. Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Zeitschrift für Physik*, 43: 172-198, 1927.

[Heisenberg, 1930] W. Heisenberg. *The Physical Principles of the Quantum Theory*. Chicago: University of Chicago Press, 1930.

[Heisenberg, 1942] W. Heisenberg. Ordnung der Wirklichkeit. In Heisenberg [1984a], pages 217-306, 1942. Also available at http://werner-heisenberg.unh.edu/Ordnung.htm.

[Heisenberg, 1958] W. Heisenberg. *Physics and Philosophy: The Revolution in Modern Science*. London: Allen and Unwin, 1958.

[Heisenberg, 1969] W. Heisenberg. *Der Teil und das Ganze: Gespräche im Umkreis der Atomphysik*. München: Piper, 1969. English translation as Heisenberg [1971].

[Heisenberg, 1971] W. Heisenberg. *Physics and Beyond*. New York: Harper and Row, 1971. Translation of Heisenberg [1969].

[Heisenberg, 1984a] W. Heisenberg. *Gesammelte Werke. Series C: Philosophical and Popular Writings, Vol I: Physik und Erkenntnis* 1927-1955. W. Blum, H.-P. Dürr, and H. Rechenberg (eds.). München: Piper, 1984a.

[Heisenberg, 1984b] W. Heisenberg. *Gesammelte Werke. Series C: Philosophical and Popular Writings, Vol II: Physik und Erkenntnis* 1956-1968. W. Blum, H.-P. Dürr, and H. Rechenberg (eds.). München: Piper, 1984b.

[Heisenberg, 1985] W. Heisenberg. *Gesammelte Werke. Series C: Philosophical and Popular Writings, Vol III: Physik und Erkenntnis* 1969-1976. W. Blum, H.-P. Dürr, and H. Rechenberg (eds.). München: Piper, 1985.

[Held, 1994] C. Held. The Meaning of complementarity. *Studies in History and Philosophy of Science*, 25: 871-893, 1994.

[Helffer, 1988] B. Helffer. *Semi-classical Analysis for the Schrödinger Operator and Applications. Lecture Notes in Mathematics*, 1336. Berlin: Springer-Verlag, 1988.

[Heller and Tomsovic, 1993] E. J. Heller and S. Tomsovic. Postmodern quantum mechanics. *Physics Today*, July: 38-46.

[Hendry, 1984] J. Hendry. *The Creation of Quantum Mechanics and the Bohr-Pauli Dialogue*. Dordrecht: D. Reidel, 1984.

[Henneaux and Teitelboim, 1992] M. Henneaux and C. Teitelboim. *Quantization of Gauge Systems*. Princeton: Princeton University Press, 1992.

[Hepp, 1972] K. Hepp. Quantum theory of measurement and macroscopic observables. *Helvetica Physica Acta*, 45: 237-248, 1972.

[Hepp, 1974] K. Hepp. The classical limit of quantum mechanical correlation functions. *Communications in Mathematical Physics*, 35: 265-277, 1974.

[Hepp and Lieb, 1974] K. Hepp and E. Lieb. Phase transitions in reservoir driven open systems with applications to lasers and superconductors. *Helvetica Physica Acta*, 46: 573-602, 1974.

[Higson, 1990] N. Higson. A primer on *KK*-theory. *Operator Theory: Operator Algebras and Applications. Proceedings Symposia in Pure Mathematical*, 51, Part 1, pages 239-283. Providence, RI: American Mathematical Society, 1990.

[Hillery *et al.*, 1984] M. Hillery, R. F. O'Connel, M. O. Scully, and E. P. Wigner. Distribution functions in physics-Fundamentals. *Physics Reports*, 106: 121-167, 1984.

[Hislop and Sigal, 1996] P. D. Hislop and I. M. Sigal. *Introduction to Spectral Theory. With Applications to Schrödinger Operators.* New York: Springer-Verlag, 1996.

[Hoffmann, 1810] E. T. A. Hoffmann. *Musikalische Novellen und Aufsätze.* Leipzig: Insel-Bücherei, 1810.

[Holevo, 1982] A. S. Holevo. *Probabilistic and Statistical Aspects of Quantum Theory.* Amsterdam: North-Holland Publishing Co, 1982.

[Hogreve *et al.*, 1983] H. Hogreve, J. Potthoff, and R. Schrader. Classical limits for quantum particles in external Yang-Mills potentials. *Communications in Mathematical Physics*, 91: 573-598, 1983.

[Honegger and Rieckers, 1994] R. Honegger and A. Rieckers. Quantized radiation states from the infinite Dicke model. *Publications of the Research Institute for Mathematical Sciences (Kyoto)*, 30: 111-138, 1994.

[Honner, 1987] J. Honner. *The Description of Nature: Niels Bohr and the Philosophy of Quantum Physics.* Oxford: Oxford University Press, 1987.

[Hooker, 1972] C. A. Hooker. The nature of quantum mechanical reality: Einstein versus Bohr. In J. Colodny (ed.). *Paradigms and Paradoxes: The Philosophical Challenges of the Quantum Domain*, pages 67-302. Pittsburgh: University of Pittsburgh Press, 1972.

[Hörmander, 1965] L. Hörmander. Pseudo-differential operators. *Communications in Pure Applied Mathematical*, 18: 501-517, 1965.

[Hörmander, 1979] L. Hörmander. The Weyl calculus of pseudo-differential operators. *Communications in Pure Applied Mathematical*, 32: 359-443, 1979.

[Hörmander, 1985a] L. Hörmander. *The Analysis of Linear Partial Differential Operators, Vol. III.* Berlin: Springer-Verlag, 1985a.

[Hörmander, 1985b] L. Hörmander. *The Analysis of Linear Partial Differential Operators, Vol. IV.* Berlin: Springer-Verlag, 1985b.

[Horowitz and Marolf, 1995] G. T. Horowitz and D. Marolf. Quantum probes of spacetime singularities. *Physical Review*, D52: 5670-5675, 1995.

[Hörz, 1968] H. Hörz. *Werner Heisenberg und die Philosophie.* Berin: VEB Deutscher Verlag der Wissenschaften, 1968.

[Howard, 1990] D. Howard. "Nicht sein kann was nicht sein darf", or the Prehistory of epr, 1909-1935: Einstein's early worries about the quantum mechanics of composite systems. In A. I. Miller (ed.), *Sixty-Two Years of Uncertainty*, pages 61-111. New York: Plenum, 1990.

[Howard, 1994] D. Howard. What makes a classical concept classical? Towards a reconstruction of Niels Bohr's philosophy of physics. In J. Faye and H. Folse (eds.), *Niels Bohr and Contemporary Philosophy*, pages 201-229. Dordrecht: Kluwer Academic Publishers, 1994.

[Howard, 2004] D. Howard. Who Invented the Copenhagen Interpretation? *Philosophy of Science*, 71: 669-682, 2004.

[Howe, 1980] R. Howe. Quantum mechanics and partial differential equations. *Journal of Functional Analysis*, 38: 188-254, 1980.

[Hudson, 1974] R. L. Hudson. When is the Wigner quasi-probability density non-negative? *Reports of Mathematical Physics*, 6: 249-252, 1974.

[Hudson and Moody, 1975/76] R. L. Hudson and G. R. Moody. Locally normal symmetric states and an analogue of de Finetti's theorem. Z. *Wahrscheinlichkeitstheorie und Verw. Gebiete*, 33: 343-351, 1975/76.

[Hunziker and Sigal, 2000] W. Hunziker and I. M. Sigal. The quantum *N*-body problem. *Journal of Mathematical Physics*, 41: 3448-3510, 2000.

[Husimi, 1940] K. Husimi. Some formal properties of the density matrix. *Progress of the Physical and Mathematical Society of Japan*, 22: 264-314, 1940.

[Isham, 1984] C. J. Isham. Topological and global aspects of quantum theory. *Relativity*, *Groups and Topology*, *II* (*Les Houches*, 1983), pages 1059-1290. Amsterdam: North-Holland, 1984.

[Isham, 1994] C. J. Isham. Quantum logic and the histories approach to quantum theory. *Journal of Mathematical Physics*, 35: 2157-2185, 1994.

[Isham, 1997] C. J. Isham. Topos theory and consistent histories: the internal logic of the set of all consistent sets. *International Journal of Theoretical Physics*, 36: 785-814, 1997.

[Isham and Butterfield, 2000] C. J. Isham and J. Butterfield. Some possible roles for topos theory in quantum theory and quantum gravity. *Foundations of Physics*, 30: 1707-1735, 2000.

[Isham and Linden, 1994] C. J. Isham and N. Linden. Quantum temporal logic and decoherence functionals in the histories approach to generalized quantum theory. *Journal of Mathematical Physics*, 35: 5452-5476, 1994.

[Isham and Linden, 1995] C. J. Isham and N. Linden. Continuous histories and the history group in generalized quantum theory. *Journal of Mathematical Physics*, 36: 5392-5408, 1995.

[Isham *et al.*, 1994] C. J. Isham, N. Linden, and S. Schreckenberg. The classification of decoherence functionals: an analog of Gleason's theorem. *Journal of Mathematical Physics*, 35: 6360-6370, 1994.

[Ivrii, 1998] V. Ivrii. *Microlocal Analysis and Precise Spectral Asymptotics*. New York: Springer-Verlag, 1998.

[Jalabert and Pastawski, 2001] R. O. Jalabert and H. M. Pastawski. Environment-Independent Decoherence Rate in Classically Chaotic Systems. *Physical Review Letters*, 86: 2490-2493, 2001.

[Jammer, 1966] M. Jammer. *The Conceptual Development of Quantum Mechanics*. New York: McGraw-Hill, 1966.

[Jammer, 1974] M. Jammer. *The Philosophy of Quantum Mechanics*. New York: Wiley, 1974.

[Janssens, 2005] B. Janssens. *Quantum Measurement: A Coherent Description*. M. Sc. Thesis, Radboud Universiteit Nijmegen, 2005. arXiv: quant-ph/0503009

[Jauch, 1968] J. M. Jauch. *Foundations of Quantum Mechanics*. Reading (MA): Addison-Wesley, 1968.

[Joos and Zeh, 1985] E. Joos and H. D. Zeh. The emergence of classical properties through interaction with the environment. *Zeitschrift für Physik*, B59: 223-243, 1985.

[Joos *et al.*, 2003] E. Joos, H. D. Zeh, C. Kiefer, D. Giulini, J. Kupsch, and I. -O. Stamatescu. *Decoherence and the Appearance of a Classical World in Quantum Theory.* Berlin: Springer-Verlag, 2003.

[Jorgensen and Moore, 1984] P. E. T. Jorgensen and R. T. Moore. *Operator Commutation Relations.* Dordrecht: Reidel, 1984.

[Kaplan and Heller, 1998a] L. Kaplan and E. J. Heller. Linear and nonlinear theory of eigenfunction scars. *Annals of Physics*, 264: 171-206, 1998a.

[Kaplan and Heller, 1998b] L. Kaplan and E. J. Heller. Weak quantum ergodicity. *Physica D*, 121: 1-18, 1998b.

[Kaplan, 1999] L. Kaplan. Scars in quantum chaotic wavefunctions. *Nonlinearity*, 12: R1-R40, 1999.

[Katok and Hasselblatt, 1995] A. Katok and B. Hasselblatt. *Introduction to the Modern Theory of Dynamical Systems.* Cambridge: Cambridge University Press, 1995.

[Kadison and Ringrose, 1983] R. V. Kadison and J. R. Ringrose. *Fundamentals of the theory of operator algebras. Vol.* 1: *Elementary Theory.* New York: Academic Press, 1983.

[Kadison and Ringrose, 1986] R. V. Kadison and J. R. Ringrose. *Fundamentals of the theory of operator algebras. Vol.* 2: *Advanced Theory.* New York: Academic Press, 1986.

[Karasev, 1989] M. V. Karasev. The Maslov quantization conditions in higher cohomology and analogs of notions developed in Lie theory for canonical fibre bundles of symplectic manifolds. I, II. *Selecta Mathematica Formerly Sovietica*, 8: 213-234, 235-258, 1989.

[Kent, 1990] A. Kent. Against Many-Worlds Interpretations. *International Journal of Modern Physics*, A5, 1745, 1990.

[Kent, 1997] A. Kent. Consistent sets yield contrary inferences in quantum theory. *Physical Review Letters*, 78: 2874-2877, 1997. Reply by R. B. Griffiths and J. B. Hartle (1998). *ibid.* 81, 1981. Reply to this reply by A. Kent (1998). *ibid.* 81, 1982.

[Kent, 1998] A. Kent. Quantum histories. *Physica Scripta*, T76: 78-84, 1998.

[Kent, 2000] A. Kent. Night thoughts of a quantum physicist. *Philosophical Transactions of the Royal Society of London*, A 358: 75-88, 2000.

[Kiefer, 2003] C. Kiefer. Consistent histories and decoherence. In E. Joos *et al.* (eds.), *Decoherence and the Appearance of a Classical World in Quantum Theory*, pages 227-258. Berlin: Springer-Verlag, 2003.

[Kirchberg and Wassermann, 1995] E. Kirchberg and S. Wassermann. Operations on continuous bundles of C^*-algebras. *Mathematische Annalen*, 303: 677-697, 1995.

[Kirillov, 1990] A. A. Kirillov. Geometric quantization. In V. I. Arnold and S. P. Novikov (eds.), *Dynamical Systems IV*, pages 137-172. Berlin: Springer-Verlag, 1990.

[Kirillov, 2004] A. A. Kirillov. *Lectures on the Orbit Method.* Providence, RI: American Mathematical Society, 2004.

[Klauder and Skagerstam, 1985] J. R. Klauder and B. -S. Skagerstam (eds.). *Coherent States.* Singapore: World Scientific, 1985.

[Klingenberg, 1982] W. Klingenberg. *Riemannian Geometry.* de Gruyter, Berlin, 1982.

[Kohn and Nirenberg, 1965] J. J. Kohn and L. Nirenberg. An algebra of pseudo-differential operators. *Communications in Pure and Applied Mathematics*, 18: 269-305, 1965.

[Koopman, 1931] B. O. Koopman. Hamiltonian systems and transformations in Hilbert space. *Proceedings of the National Academy of Sciences*, 18: 315-318, 1931.

[Kostant, 1970] B. Kostant. Quantization and unitary representations. *Lecture Notes in Mathematics*, 170: 87-208, 1970.

[Krishnaprasad and Marsden, 1987] P. S. Krishnaprasad and J. E. Marsden. Hamiltonian structures and stability for rigid bodies with flexible attachments. *Archive of Rational Mechanics and Analysis*, 98: 71-93, 1987.

[Kuhn, 1978] T. S. Kuhn. *Black-body Theory and the Quantum Discontinuity: 18941912.* New York: Oxford University Press, 1978.

[Kümmerer, 2002] B. Kümmerer. Quantum Markov processes. A. Buchleitner and K. Hornberger (eds.), *Coherent Evolution in Noisy Environment.* Lecture Notes in Physics, Vol. 611, pages, 139-198. Berlin: Springer-Verlag, 2002.

[Lahti and Mittelstaedt, 1987] P. Lahti and P. Mittelstaedt (eds.). *The Copenhagen Interpretation 60 Years After the Como Lecture.* Singapore: World Scientific, 1987.

[Landau and Lifshitz, 1977] L. D. Landau and E. M. Lifshitz. *Quantum Mechanics: Nonrelativistic Theory.* 3d Ed. Oxford: Pergamon Press, 1977.

[Landsman, 1990a] N. P. Landsman. Quantization and superselection sectors I. Transformation group C^*-algebras. *Reviews in Mathematical Physics*, 2: 45-72, 1990a.

[Landsman, 1990b] N. P. Landsman. Quantization and superselection sectors II. Dirac Monopole and Aharonov-Bohm effect. *Reviews in Mathematical Physics*, 2: 73-104, 1990b.

[Landsman, 1991] N. P. Landsman. Algebraic theory of superselection sectors and the measurement problem in quantum mechanics. *International Journal of Modern Physics*, A30: 5349-5371, 1991.

[Landsman, 1992] N. P. Landsman. Induced representations, gauge fields, and quantization on homogeneous spaces. *Reviews in Mathematical Physics*, 4: 503-528, 1992.

[Landsman, 1993] N. P. Landsman. Deformations of algebras of observables and the classical limit of quantum mechanics. *Reviews in Mathematical Physics*, 5: 775-806, 1993.

[Landsman, 1995] N. P. Landsman. Observation and superselection in quantum mechanics. *Studies in History and Philosophy of Modern Physics*, 26B: 45-73, 1995.

[Landsman, 1997] N. P. Landsman. Poisson spaces with a transition probability. *Reviews in Mathematical Physics*, 9: 29-57, 1997.

[Landsman, 1998] N. P. Landsman. *Mathematical Topics Between Classical and Quantum Mechanics*. New York: Springer-Verlag, 1998.

[Landsman, 1999a] N. P. Landsman. Quantum mechanics on phase space. *Studies in History and Philosophy of Modern Physics*, 30B: 287-305, 1999a.

[Landsman, 1999b] N. P. Landsman. Lie groupoid C^*-algebras and Weyl quantization. *Communications in Mathematical Physics*, 206: 367-381, 1999b.

[Landsman, 2001] N. P. Landsman. Quantized reduction as a tensor product. In N. P. Landsman, M. J. Pflaum, and M. Schlichenmaier (eds.). *Quantization of Singular Symplectic Quotients*, pages 137-180. Basel: Birkhäuser, 2001.

[Landsman, 2002] N. P. Landsman. Quantization as a functor. *Contemporary Mathematics*, 315: 9-24, 2002.

[Landsman, 2005] N. P. Landsman. Functorial quantization and the Guillemin-Sternberg conjecture. S. T. Ali, G. G. Emch, A. Odzijewicz, M. Schlichenmaier, and S. L. Woronowicz (eds). *Twenty Years of Bialowieza: A Mathematical Anthology*, pages 23-45. Singapore: World Scientific, 2005. arXiv: math-ph/0307059

[Landsman, 2006a] N. P. Landsman. Lie Groupoids and Lie algebroids in physics and noncommutative geometry. *Journal of Geom. Physics*, 56: 24-54, 2006a.

[Landsman, 2006b] N. P. Landsman. When champions meet: Rethinking the Bohr-Einstein debate. *Studies in History and Philosophy of Modern Physics*, 37: 212-241, 2006b. arXiv: quant-ph/0507220

[Landsman and Ramazan, 2001] N. P. Landsman and B. Ramazan. Quantization of Poisson algebras associated to Lie algebroids. *Contemporary Mathematics*, 282: 159-192, 2001.

[Laurikainen, 1988] K. V. Laurikainen. *Beyond the Atom: The Philosophical Thought of Wolfgang Pauli.* Berlin: Springer-Verlag, 1988.

[Lazutkin, 1993] V. F. Lazutkin. *KAM Theory and Semiclassical Approximations to Eigenfunctions.* Berlin: Springer-Verlag, 1993.

[Leggett, 2002] A. J. Leggett. Testing the limits of quantum mechanics: motivation, state of play, prospects. *Journal of Physics: Condensed Matter*, 14: R415-R451, 2002.

[Liboff, 1984] R. L. Liboff. The correspondence principle revisited. *Physics Today*, February: 50-55, 1984.

[Lieb, 1973] E. H. Lieb. The classical limit of quantum spin systems. *Communications in Mathematical Physics*, 31: 327-340, 1973.

[Littlejohn, 1986] R. G. Littlejohn. The semiclassical evolution of wave packets. *Physics Reports*, 138: 193-291, 1986.

[Ludwig, 1985] G. Ludwig. *An Axiomatic Basis for Quantum Mechanics. Volume 1: Derivation of Hilbert Space Structure.* Berlin: Springer-Verlag, 1985.

[Lugiewicz and Olkiewicz, 2002] P. Lugiewicz and R. Olkiewicz. Decoherence in infinite quantum systems. *Journal of Physics*, A35: 6695-6712, 2002.

[Lugiewicz and Olkiewicz, 2003] P. Lugiewicz and R. Olkiewicz. Classical properties of infinite quantum open systems. *Communications in Mathematical Physics*, 239: 241-259, 2003.

[Maassen, 2003] H. Maassen. Quantum probability applied to the damped harmonic oscillator. *Quantum Probability Communications*, Vol. XII (Grenoble, 1998), pages 23-58. River Edge, NJ: World Scientific Publishing, 2003.

[Mackey, 1962] G. W. Mackey. *The Mathematical Foundations of Quantum Mechanics.* New York: Benjamin, 1962.

[Mackey, 1968] G. W. Mackey. *Induced Representations of Groups and Quantum Mecha-*

nics. New York: W. A. Benjamin; Turin: Editore Boringhieri, 1968.

[Mackey, 1978] G. W. Mackey. *Unitary Group Representations in Physics, Probability, and Number Theory.* Reading, Mass.: Benjamin/Cummings Publishing Co, 1978.

[Mackey, 1992] G. W. Mackey. *The Scope and History of Commutative and Noncommutative Harmonic Analysis.* Providence, RI: American Mathematical Society, 1992.

[Majid, 1988] S. Majid. Hopf algebras for physics at the Planck scale. *Classical and Quantum Gravity*, 5: 1587-1606, 1988.

[Majid, 1990] S. Majid. Physics for algebraists: noncommutative and noncocommutative Hopf algebras by a bicrossproduct construction. *Journal of Algebra*, 130: 17-64, 1990.

[Marmo *et al.*, 2005] G. Marmo, G. Scolarici, A. Simoni, and F. Ventriglia. The quantum-classical transition: the fate of the complex structure. *International Journal of Geometric Methods in Physics*, 2: 127-145, 2005.

[Marsden, 1992] J. E. Marsden. *Lectures on Mechanics.* Cambridge: Cambridge University Press, 1992.

[Marsden and Ratiu, 1994] J. E. Marsden and T. S. Ratiu. *Introduction to Mechanics and Symmetry.* New York: Springer-Verlag, 1994.

[Marsden *et al.*, 1984] J. E. Marsden, T. Rat, iu, and A. Weinstein. Semidirect products and reduction in mechanics. *Transactions of the American Mathematical Society*, 281: 147-177, 1984.

[Marshall *et al.*, 2003] W. Marshall, C. Simon, R. Penrose, and D. Bouwmeester. Towards quantum superpositions of a mirror. *Physical Review Letters*, 91, 130401-1-4, 2003.

[Martinez, 2002] A. Martinez. *An Introduction to Semiclassical and Microlocal Analysis.* New York: Springer-Verlag, 2002.

[Maslov and Fedoriuk, 1981] V. P. Maslov and M. V. Fedoriuk. *Semi-Classical Approximation in Quantum Mechanics.* Dordrecht: Reidel.

[McCormmach, 1982] R. McCormmach. *Night Thoughts of a Classical Physicist.* Cambridge (MA): Harvard University Press, 1982.

[Mehra and and Rechenberg, 1982a] J. Mehra and H. Rechenberg. *The Historical Development of Quantum Theory. Vol. 1: The Quantum Theory of Planck, Einstein, Bohr, and Sommerfeld: Its Foundation and the Rise of Its Difficulties.* New York: Springer-Verlag, 1982a.

[Mehra and and Rechenberg, 1982b] J. Mehra and H. Rechenberg. *The Historical Development of Quantum Theory. Vol. 2: The Discovery of Quantum Mechanics.* New York: Springer-Verlag, 1982b.

[Mehra and and Rechenberg, 1982c] J. Mehra and H. Rechenberg. *The Historical Development of Quantum Theory. Vol. 3: The formulation of matrix mechanics and its modifications*, 1925-1926. New York: Springer-Verlag, 1982c.

[Mehra and and Rechenberg, 1982d] J. Mehra and H. Rechenberg. *The Historical Development of Quantum Theory. Vol. 4: The fundamental equations of quantum mechanics* 1925-1926. *The reception of the new quantum mechanics.* New York: Springer-Verlag, 1982d.

[Mehra and and Rechenberg, 1987] J. Mehra and H. Rechenberg. *The Historical Development of Quantum Theory. Vol. 5: Erwin Schrödinger and the Rise of Wave Mechanics.* New York: Springer-Verlag, 1987.

[Mehra and and Rechenberg, 2000] J. Mehra and H. Rechenberg. *The Historical Development of Quantum Theory. Vol. 6: The Completion of Quantum Mechanics* 1926-1941. *Part* 1: *The probabilistic Interpretation and the Empirical and Mathematical Foundation of Quantum Mechanics*, 1926-1936. New York: Springer-Verlag, 2000.

[Mehra and and Rechenberg, 2001] J. Mehra and H. Rechenberg. *The Historical Development of Quantum Theory. Vol. 6: The Completion of Quantum Mechanics* 1926-1941. *Part* 2: *The Conceptual Completion of Quantum Mechanics.* New York: Springer-Verlag, 2001.

[Meinrenken, 1998] E. Meinrenken. Symplectic surgery and the Spinc-Dirac operator. *Adv. Mathematical*, 134: 240-277, 1998.

[Meinrenken and Sjamaar, 1999] E. Meinrenken and R. Sjamaar. Singular reduction and quantization. *Topology*, 38: 699-762, 1999.

[Mermin, 2004] N. D. Mermin. What's wrong with this quantum world? *Physics Today*, 57 (2): 10, 2004.

[Mielnik, 1968] B. Mielnik. Geometry of quantum states. *Communications in Mathematical Physics*, 9: 55-80, 1968.

[Miller, 1984] A. I. Miller. *Imagery in Scientific Thought: Creating 20th-Century Physics.* Boston: Birkhäuser, 1984.

[Mirlin, 2000] A. D. Mirlin. Statistics of energy levels and eigenfunctions in disordered systems. *Physics Reports*, 326: 259-382, 2000.

[Mittelstaedt, 2004] P. Mittelstaedt. *The Interpretation of Quantum Mechanics and the Measurement Process.* Cambridge: Cambridge University Press, 2004.

[Moore, 1939] G. E. Moore. Proof of an external world. *Proceedings of the British Academy*, 25: 273-300, 1939. Reprinted in *Philosophical Papers* (George, Allen and Unwin, London, 1959) and in *Selected Writings* (Routledge, London, 1993).

[Moore, 1989] W. Moore. *Schrödinger: Life and Thought.* Cambridge: Cambridge University Press, 1989.

[Morchio and Strocchi, 1987] G. Morchio and F. Strocchi. Mathematical structures for longrange dynamics and symmetry breaking. *Journal of Mathematical Physics*, 28: 622-635, 1987.

[Muller, 1997] F. A. Muller. The equivalence myth of quantum mechanics I, II. *Studies in History and Philosophy of Modern Physics*, 28: 35-61, 219-247, 1997.

[Murdoch, 1987] D. Murdoch. *Niels Bohrs Philosophy of Physics.* Cambridge: Cambridge University Press, 1987.

[Nadirashvili *et al.*, 2001] N. Nadirashvili, J. Toth, and D. Yakobson. Geometric properties of eigenfunctions. *Russian Mathematical Surveys*, 56: 1085-1105, 2001.

[Nagy, 2000] G. Nagy. A deformation quantization procedure for C^*-algebras. *Journal of Operator Theory*, 44: 369-411, 2000.

[Narnhofer, 2001] H. Narnhofer. Quantum K-systems and their abelian models. *Foundations of Probability and Physics*, pages 274-302. River Edge, NJ: World Scientific, 2001.

[Natsume and Nest, 1999] T. Natsume and R. Nest. Topological approach to quantum surfaces. *Communications in Mathematical Physics*, 202: 65-87, 1999.

[Natsume *et al.*, 2003] T. Natsume, R. Nest, and P. Ingo. Strict quantizations of symplectic manifolds. *Letters in Mathematical Physics*, 66: 73-89, 2003.

[Nauenberg, 1989] M. Nauenberg. Quantum wave packets on Kepler elliptic orbits. *Physical Review*, A40: 1133-1136, 1989.

[Nauenberg *et al.*, 1994] M. Nauenberg, C. Stroud, and J. Yeazell. The classical limit of an atom. *Scientific American*, June, 24-29, 1994.

[Neumann von, 1931] J. von Neumann. Die Eindeutigkeit der Schrödingerschen Operatoren. *Mathematische Annalen*, 104: 570-578, 1931.

[Neumann von, 1932] J. von Neumann. *Mathematische Grundlagen der Quantenmechanik.* Berlin: Springer-Verlag. English translation (1955): *Mathematical Foundations of Quantum Mechanics.* Princeton: Princeton University Press, 1932.

[Neumann von, 1938] J. von Neumann. On infinite direct products. *Compositio Mathematica*, 6: 1-77, 1938.

[Neumann von, 1981] J. von Neumann. Continuous geometries with a transition probability. *Memoirs of the American Mathematical Society*, 252: 1-210. (I. S. Halperin (ed.). MS from 1937), 1981.

[Neumann, 1972] H. Neumann. Transformation properties of observables. *Helvetica PhysicaActa*, 25: 811-819, 1972.

[Nourrigat and Royer, 2004] J. Nourrigat and C. Royer. Thermodynamic limits for Hamiltonians defined as pseudodifferential operators. *Communications in Partial Differential Equations*, 29: 383-417, 2004.

[O'Connor *et al.*, 1992] P. W. O'Connor, S. Tomsovic, and E. J. Heller. Semiclassical dynamics in the strongly chaotic regime: breaking the log time barrier. *Physica*, D55: 340-357, 1992.

[Odzijewicz, 1992] A. Odzijewicz. Coherent states and geometric quantization. *Communications in Mathematical Physics* 150, 385-413, 1992.

[Odzijewicz and Ratiu, 2003] A. Odzijewicz and T. S. Ratiu. Banach Lie-Poisson spaces and reduction. *Communications in Mathematical Physics*, 243: 1-54, 2003.

[Olkiewicz, 1999a] R. Olkiewicz. Dynamical semigroups for interacting quantum and classical systems. *Journal of Mathematical Physics*, 40: 1300-1316, 1999a.

[Olkiewicz, 1999b] R. Olkiewicz. Environment-induced superselection rules in Markovian regime. *Communications in in Mathematical Physics*, 208: 245-265, 1999b.

[Olkiewicz, 2000] R. Olkiewicz. Structure of the algebra of effective observables in quantum mechanics. *Annals of Physics*, 286: 10-22, 2000.

[Ollivier *et al.*, 2004] H. Ollivier, D. Poulin, and W. H. Zurek. Environment as witness: selective proliferation of information and emergence of objectivity. 2004. arXiv: quant-ph/0408125

[Olshanetsky and Perelomov, 1981] M. A. Olshanetsky and A. M. Perelomov. Classical integrable finite-dimensional systems related to Lie algebras. *Physics Reports*, 71: 313-400, 1981.

[Olshanetsky and Perelomov, 1983] M. A. Olshanetsky and A. M. Perelomov. Quantum integrable systems related to Lie algebras. *Physics Reports*, 94: 313-404, 1983.

[Omnès, 1992] R. Omnès. Consistent interpretations of quantum mechanics. *Reviews of Modern Physics*, 64: 339-382, 1992.

[Omnès, 1994] R. Omnès. *The Interpretation of Quantum Mechanics.* Princeton: Princeton University Press, 1994.

[Omnès, 1997] R. Omnès. Quantum-classical correspondence using projection operators. *Journal of Mathematical Physics*, 38: 697-707, 1997.

[Omnès, 1999] R. Omnès. *Understanding Quantum Mechanics.* Princeton: Princeton University Press, 1999.

[Ørsted, 1979] B. Ørsted. Induced representations and a new proof of the imprimitivity theorem. *Journal of Functional Analysis*, 31: 355-359, 1979.

[Ozorio de Almeida, 1988] A. M. Ozorio de Almeida. *Hamiltonian Systems: Chaos and Quantization.* Cambridge: Cambridge University Press, 1988.

[Pais, 1982] A. Pais. *Subtle is the Lord: The Science and Life of Albert Einstein.* Oxford: Oxford University Press, 1982.

[Pais, 1991] A. Pais. *Niels Bohrs Times: In Physics, Philosophy, and Polity.* Oxford: Oxford University Press, 1991.

[Pais, 1997] A. Pais. *A Tale of Two Continents: A Physicist's Life in a Turbulent World.* Princeton: Princeton University Press, 1997.

[Pais, 2000] A. Pais. *The Genius of Science.* Oxford: Oxford University Press, 2000.

[Parthasarathy, 1992] K. R. Parthasarathy. *An Introduction to Quantum Stochastic Calculus.* Basel: Birkhäuser, 1992.

[Paul and Uribe, 1995] T. Paul and A. Uribe. The semi-classical trace formula and propagation of wave packets. *Journal of Functional Analysis*, 132: 192-249, 1995.

[Paul and Uribe, 1996] T. Paul and A. Uribe. On the pointwise behavior of semi-classical measures. *Communications in Mathematical Physics*, 175: 229-258, 1996.

[Paul and Uribe, 1998] T. Paul and A. Uribe. A Weighted trace formula near a hyperbolic trajectory and complex orbits. *Journal of Mathematical Physics*, 39: 4009-4015, 1998.

[Pauli, 1925] W. Pauli. Über den Einflus der Geschwindigkeitsabhängigkeit der Elektronenmasse auf den Zeemaneffekt. *Zeitschrift für Physik*, 31: 373-385, 1925.

[Pauli, 1933] W. Pauli. *Die allgemeinen Prinzipien der Wellenmechanik*. Flügge, S. (Ed.). *Handbuch der Physik*, Vol. V, Part I, 1933. Translated as W. Pauli, *General Principles of Quantum Mechanics*. Berlin: Springer-Verlag, 1980.

[Pauli, 1949] W. Pauli. Die philosophische Bedeutung der Idee der Komplementarität. 1949. Reprinted in K. von Meyenn (ed.), *Wolfang Pauli: Physik und Erkenntnistheorie*, pages 10-23. Braunschweig: Vieweg Verlag, 1984. English translation in Pauli, 1994.

[Pauli, 1979] W. Pauli. *Wissenschaftlicher Briefwechsel mit Bohr, Einstein, Heisenberg. Vol* 1: 1919-1929. A. Hermann, K. von Meyenn, and V. Weisskopf (eds.), New York: Springer-Verlag, 1979.

[Pauli, 1985] W. Pauli. *Wissenschaftlicher Briefwechsel mit Bohr, Einstein, Heisenberg. Vol* 2: 1930-1939. K. von Meyenn (ed.). New York: Springer-Verlag, 1985.

[Pauli, 1994] W. Pauli. *Writings on Physics and Philosophy*. C. P. Enz and K. von Meyenn (eds.). Berlin: Springer-Verlag, 1994.

[Paz and Zurek, 1999] J. P. Paz and W. H. Zurek. Quantum limit of decoherence: environment induced superselection of energy eigenstates. *Physical Review Letters*, 82: 5181-5185, 1999.

[Pedersen, 1979] G. K. Pedersen. *C^*-algebras and their Automorphism Groups*. London: Academic Press, 1979.

[Pedersen, 1989] G. K. Pedersen. *Analysis Now*. New York: Springer-Verlag, 1989.

[Perelomov, 1986] A. Perelomov. *Generalized Coherent States and their Applications*. Berlin: Springer-Verlag, 1986.

[Peres, 1984] A. Peres. Stability of quantum motion in chaotic and regular systems. *Physical Review*, A30: 1610-1615, 1984.

[Peres, 1995] A. Peres. *Quantum Theory: Concepts and Methods*. Dordrecht: Kluwer Academic Publishers, 1995.

[Petersen, 1963] A. Petersen. The Philosophy of Niels Bohr. *Bulletin of the Atomic Scientists*, 19: 8-14, 1963.

[Pitowsky, 1989] I. Pitowsky. *Quantum Probability-Quantum Logic.* Berlin: Springer-Verlag, 1989.

[Planck, 1906] M. Planck. *Vorlesungen Über die Theorie der Wärmestrahlung* Leipzig: J. A. Barth, 1906.

[Poincaré, 1892-1899] H. Poincaré. *Les Méthodes Nouvelles de la Méchanique Céleste.* Paris: Gauthier-Villars, 1892-1899.

[Popov, 2000] G. Popov. Invariant tori, effective stability, and quasimodes with exponentially small error terms. I and II. *Annales Henri Poincaré*, 1: 223-248 and 249-279, 2000.

[Poulin, 2004] D. Poulin. Macroscopic observables. 2004. arXiv: quant-ph/0403212.

[Poulsen, 1970] N. S. Poulsen. *Regularity Aspects of the Theory of Infinite-Dimensional Representations of Lie Groups.* Ph. D Thesis, MIT, 1970.

[Primas, 1983] H. Primas. *Chemistry, Quantum Mechanics and Reductionism.* Second Edition. Berlin: Springer-Verlag, 1983.

[Primas, 1997] H. Primas. The representation of facts in physical theories. In H. Atmanspacher and E. Ruhnau (eds.), *Time, Temporality, Now*, pages 241-263. Berlin: Springer-Verlag, 1997.

[Prugovecki, 1971] E. Prugovecki. *Quantum Mechanics in Hilbert Space.* New York: Academic Press, 1971.

[Puta, 1993] M. Puta. *Hamiltonian Dynamical Systems and Geometric Quantization.* Dordrecht: D. Reidel, 1993.

[Raggio, 1981] G. A. Raggio. *States and Composite Systems in W^*-algebras Quantum Mechanics.* Ph. D Thesis, ETH Zürich, 1981.

[Raggio, 1988] G. A. Raggio. A remark on Bell's inequality and decomposable normal states. *Letters in Mathematical Physics*, 15: 27-29, 1988.

[Raggio and Werner, 1989] G. A. Raggio and R. F. Werner. Quantum statistical mechanics of general mean field systems. *Helvetica Physica Acta*, 62: 980-1003, 1989.

[Raggio and Werner, 1991] G. A. Raggio and R. F. Werner. The Gibbs variational principle for inhomogeneous mean field systems. *Helvetica Physica Acta*, 64: 633-667, 1991.

[Raimond *et al.*, 2001] R. M. Raimond, M. Brune, and S. Haroche. Manipulating quantum entanglement with atoms and photons in a cavity. *Reviews of Modern Physics*, 73: 565-

582, 2001.

[Rédei, 1998] M. Rédei. *Quantum logic in algebraic approach.* Dordrecht: Kluwer Academic Publishers, 1998.

[Rédei and Stöltzner, 2001] M. Rédei and M. Stöltzner (eds.). *John von Neumann and the Foundations of Modern Physics.* Dordrecht: Kluwer Academic Publishers, 2001.

[Reed and Simon, 1972] M. Reed and B. Simon. *Methods of Modern Mathematical Physics. Vol I. Functional Analysis.* New York: Academic Press, 1972.

[Reed and Simon, 1975] M. Reed and B. Simon. *Methods of Modern Mathematical Physics. Vol II. Fourier Analysis, Self-adjointness.* New York: Academic Press, 1975.

[Reed and Simon, 1978] M. Reed and B. Simon. *Methods of Modern Mathematical Physics. Vol IV. Analysis of Operators.* New York: Academic Press, 1978.

[Reed and Simon, 1979] M. Reed and B. Simon. *Methods of Modern Mathematical Physics. Vol III. Scattering Theory.* New York: Academic Press, 1979.

[Reichl, 2004] L. E. Reichl. *The Transition to Chaos in Conservative Classical Systems: Quantum Manifestations.* Second Edition. New York: Springer-Verlag, 2004.

[Rieckers, 1984] A. Rieckers. On the classical part of the mean field dynamics for quantum lattice systems in grand canonical representations. *Journal of Mathematical Physics*, 25: 2593-2601, 1984.

[Rieffel, 1989a] M. A. Rieffel. Deformation quantization of Heisenberg manifolds. *Communications in Mathematical Physics* 122, 531-562, 1989a.

[Rieffel, 1989b] M. A. Rieffel. Continuous fields of C^*-algebras coming from group cocycles and actions. *Mathematical Annals*, 283: 631-643, 1989b.

[Rieffel, 1994] M. A. Rieffel. Quantization and C^*-algebras. *Contemporary Mathematics*, 167: 66-97, 1994.

[Riesz and Sz.-Nagy, 1990] F. Riesz and B. Sz.-Nagy. *Functional Analysis.* New York: Dover, 1990.

[Robert, 1987] D. Robert. *Autour de l'Approximation Semi-Classique.* Basel: Birkhäuser, 1987.

[Robert, 1992] D. Robert (ed.). *Mèthodes Semi-Classiques. Astérisque*, 207: 1-212, *ibid.*: 210, 1-384, 1992.

[Robert, 1998] D. Robert. Semi-classical approximation in quantum mechanics. A survey of old and recent mathematical results. *Helvetica Physica Acta*, 71: 44-116, 1998.

[Roberts, 1990] J. E. Roberts. Lectures on algebraic quantum field theory. In D. Kastler (ed.), *The Algebraic Theory of Superselection Sectors. Introduction and Recent Results*, pages 1-112. River Edge, NJ: World Scientific Publishing Co, 1990.

[Roberts and Roepstorff, 1969] J. E. Roberts and G. Roepstorff. Some basic concepts of algebraic quantum theory. *Communications in Mathematical Physics*, 11: 321-338, 1969.

[Robinett, 2004] R. W. Robinett. Quantum wave packet revival. *Physics Reports*, 392: 1-119, 2004.

[Robinson, 1994] D. Robinson. Can superselection rules solve the measurement problem? *British Journal for the Philosophy of Science*, 45: 79-93, 1994.

[Robinson, 1988a] S. L. Robinson. The semiclassical limit of quantum mechanics. I. Time evolution. *Journal of Mathematical Physics*, 29: 412-419, 1988a.

[Robinson, 1988b] S. L. Robinson. The semiclassical limit of quantum mechanics. II. Scattering theory. *Annales de l' Institut Henri Poincaré*, A48: 281-296, 1988b.

[Robinson, 1993] S. L. Robinson. Semiclassical mechanics for time-dependent Wigner functions. *Journal of Mathematical Physics*, 34: 2185-2205, 1993.

[Robson, 1996] M. A. Robson. Geometric quantization of reduced cotangent bundles. *Journal of Geometry and Physics*, 19: 207-245, 1996.

[Rosenfeld, 1967] L. Rosenfeld. Niels Bohr in the Thirties. Consolidation and extension of the conception of complementarity. In S. Rozental (ed.). *Niels Bohr: His Life and Work as Seen by His Friends and Colleagues*, pages 114-136. Amsterdam: North-Holland, 1967.

[Rudolph, 1996a] O. Rudolph. Consistent histories and operational quantum theory. *International Journal of Theoretical Physics*, 35: 1581-1636, 1996a.

[Rudolph, 1996b] O. Rudolph. On the consistent effect histories approach to quantum mechanics. *Journal of Mathematical Physics*, 37: 5368-5379, 1996b.

[Rudolph, 2000] O. Rudolph. The representation theory of decoherence functionals in history quantum theories. *International Journal of Theoretical Physics*, 39: 871-884, 2000.

[Rudolph and Wright, 1999] O. Rudolph and J. D. M. Wright. Homogeneous decoherence functionals in standard and history quantum mechanics. *Communications in Mathematical Phy-*

sics, 204: 249-267, 1999.

[Sarnak, 1999] P. Sarnak. Quantum chaos, symmetry and zeta functions. I. and II. *Quantum Chaos. Current Developments in Mathematics*, 1997 (*Cambridge*, *MA*), pages 127-144 and 145-159. Boston: International Press, 1999.

[Saunders, 1993] S. Saunders. Decoherence, relative states, and evolutionary adaptation. *Foundations of Physics*, 23: 1553-1585, 1993.

[Saunders, 1995] S. Saunders. Time, quantum mechanics, and decoherence. *Synthese*, 102: 235-266, 1995.

[Saunders, 2005] S. Saunders. Complementarity and scientific rationality. *Foundations of Physics*, 35: 417-447, 2005.

[Scheibe, 1973] E. Scheibe. *The Logical Analysis of Quantum Mechanics*. Oxford: Pergamon Press, 1973.

[Scheibe, 1991] E. Scheibe. J. v. Neumanns and J. S. Bells Theorem. Ein Vergleich. *Philosophia Naturalis*, 28: 35-53, 1991. English translation in E. Scheibe, *Between Rationalism and Empiricism: Selected Papers in the Philosophy of Physics*. New York: Springer-Verlag, 2001.

[Scheibe, 1999] E. Scheibe. *Die Reduktion Physikalischer Theorien. Ein Beitrag zur Einheit der Physik. Teil II: Inkommensurabilität und Grenzfallreduktion*. Berlin: Springer-Verlag, 1999.

[Schlosshauer, 2004] M. Schlosshauer. Decoherence, the measurement problem, and interpretations of quantum mechanics. *Reviews of Modern Physics*, 76: 1267-1306, 2004.

[Schmüdgen, 1990] K. Schmüdgen. *Unbounded Operator Algebras and Representation Theory*. Basel: Birkhäuser Verlag, 1990.

[Schrödinger, 1926a] E. Schrödinger. Quantisierung als Eigenwertproblem. I. -IV. *Annalen der Physik*, 79: 361-76, 489-527, *ibid.* 80: 437-90, *ibid.* 81: 109-39, 1926a. English translation in Schrödinger (1928).

[Schrödinger, 1926b] E. Schrödinger. Der stetige Übergang von der Mikro-zur Makromekanik. *Die Naturwissenschaften*, 14: 664-668, 1926b. English translation in Schrödinger(1928).

[Schrödinger, 1926c] E. Schrödinger. Über das Verhaltnis der Heisenberg-Born-Jordanschen Quantenmechanik zu der meinen. *Annalen der Physik*, 79: 734-56, 1926c. English translation in Schrödinger (1928).

[Schrödinger, 1928] E. Schrödinger. *Collected Papers on Wave Mechanics.* London: Blackie and Son, 1928.

[Schroeck, 1996] F. E. Schroeck Jr.. *Quantum Mechanics on Phase Space.* Dordrecht: Kluwer Academic Publishers, 1996.

[Scutaru, 1977] H. Scutaru. Coherent states and induced representations. *Letters in Mathematical Physics*, 2: 101-107, 1977.

[Segal, 1960] I. E. Segal. Quantization of nonlinear systems. *Journal of Mathematical Physics*, 1: 468-488, 1960.

[Sewell, 1986] G. L. Sewell. *Quantum Theory of Collective Phenomena.* New York: Oxford University Press, 1986.

[Sewell, 2002] G. L. Sewell. *Quantum Mechanics and its Emergent Macrophysics.* Princeton: Princeton University Press, 2002.

[Simon, 1976] B. Simon. Quantum dynamics: from automorphism to Hamiltonian. *Studies in Mathematical Phyiscs: Essays in Honour of Valentine Bargmann*, pp. 327-350. Lieb, E. H., Simon, B. and Wightman, A. S. (Eds.). Princeton: Princeton University Press, 1976.

[Simon, 1980] B. Simon. The classical limit of quantum partition functions. *Communications in Mathematical Physics*, 71: 247-276, 1980.

[Simon, 2000] B. Simon. Schrödinger operators in the twentieth century. *Journal of Mathematical Physics*, 41: 3523-3555, 2000.

[Śniatycki, 1980] J. Śniatycki. *Geometric Quantization and Quantum Mechanics.* Berlin: Springer-Verlag, 1980.

[Snirelman, 1974] A. I. Snirelman. Ergodic properties of eigenfunctions. *Uspekhi Mathematical Nauk*, 29: 181-182, 1974.

[Souriau, 1969] J.-M. Souriau. *Structure des systèmes dynamiques.* Paris: Dunod, 1969. Translated as J.-M. Souriau (1997).

[Souriau, 1997] J.-M. Souriau. *Structure of Dynamical Systems. A Symplectic View of Physics.* Boston: Birkhäuser, 1997.

[Stapp, 1972] H. P. Stapp. The Copenhagen Interpretation. *American Journal of Physics*, 40: 1098-1116, 1972.

[Steiner, 1998] M. Steiner. *The Applicability of Mathematics as a Philosophical Problem.* Cambridge (MA): Harvard University Press, 1998.

[Stinespring, 1955] W. Stinespring. Positive functions on C^*-algebras. *Proceedings of the American Mathematical Society*, 6: 211-216, 1955.

[Størmer, 1969] E. Størmer. Symmetric states of infinite tensor products of C^*-algebras. *Jornal of Functional Analysis*, 3: 48-68, 1969.

[Strawson, 1959] P. F. Strawson. *Individuals: An Essay in Descriptive Metaphysics.* London: Methuen, 1959.

[Streater, 2000] R. F. Streater. Classical and quantum probability. *Journal of Mathematical Physics*, 41: 3556-3603, 2000.

[Strichartz, 1983] R. S. Strichartz. Analysis of the Laplacian on a complete Riemannian manifold. *Journal of Functional Analysis*, 52: 48-79, 1983.

[Strocchi, 1985] F. Strocchi. *Elements of Quantum mechanics of Infinite Systems.* Singapore: World Scientific, 1985.

[Strunz *et al.*, 2003] W. T. Strunz, F. Haake, and D. Braun. Universality of decoherence for macroscopic quantum superpositions. *Physical Review*, A 67: 022101-022114, 2003.

[Summers and Werner, 1987] S. J. Summers and R. Werner. Bells inequalities and quantum field theory, I, II. *Journal of Mathematical Physics*, 28: 2440-2447, 2448-2456, 1987.

[Sundermeyer, 1982] K. Sundermeyer. *Constrained Dynamics.* Berlin: Springer-Verlag, 1982.

[Takesaki, 2003] M. Takesaki. *Theory of Operator Algebras. Vols. I-III.* New York: Springer-Verlag, 2003.

[Thirring and Wehrl, 1967] W. Thirring and A. Wehrl. On the mathematical structure of the BCS model. I *Communications in Mathematical Physics*, 4: 303-314, 1967.

[Thirring, 1968] W. Thirring. On the mathematical structure of the BCS model. II. *Communications in Mathematical Physics*, 7: 181-199, 1968.

[Thirring, 1981] W. Thirring. *A Course in Mathematical Physics. Vol. 3: Quantum Mechanics of Atoms and Molecules.* New York: Springer-Verlag, 1981.

[Thirring, 1983] W. Thirring. *A Course in Mathematical Physics. Vol. 4: Quantum Mechanics of Large Systems.* New York: Springer-Verlag, 1983.

[Tomsovic and Heller, 1993] S. Tomsovic and E. J. Heller. Long-time semiclassical dynamics of chaos: The stadium billiard. *Physical Review*, E47: 282-299.

[Tomsovic and Heller, 2002] S. Tomsovic and E. J. Heller. Comment on "Ehrenfest times for classically chaotic systems". *Physical Review*, E65, 035208-1-2, 2002.

[Toth, 1996] J. A. Toth. Eigenfunction localization in the quantized rigid body. *Journal of Differential Geometry*, 43: 844-858, 1996.

[Toth, 1999] J. A. Toth. On the small-scale mass concentration of modes. *Communications in Mathematical Physics*, 206: 409-428, 1999.

[Toth and Zelditch, 2002] J. A. Toth and S. Zelditch. Riemannian manifolds with uniformly bounded eigenfunctions. *Duke Mathematical Journal*, 111: 97-132, 2002.

[Toth and Zelditch, 2003a] J. A. Toth and S. Zelditch. L^p norms of eigenfunctions in the completely integrable case. *Annales Henri Poincaré*, 4: 343-368, 2003a.

[Toth and Zelditch, 2003b] J. A. Toth and S. Zelditch. Norms of modes and quasi-modes revisited. *Contemporary Mathematics*, 320: 435-458, 2003b.

[Tuynman, 1987] G. M. Tuynman. Quantization: towards a comparison between methods. *Journal of Mathematical Physics*, 28: 2829-2840, 1987.

[Tuynman, 1998] G. M. Tuynman. Prequantization is irreducible. *Indagationes Mathematicae (New Series)*, 9: 607-618, 1998.

[Unnerstall, 1990a] T. Unnerstall. Phase-spaces and dynamical descriptions of infinite mean-field quantum systems. *Journal of Mathematical Physics*, 31: 680-688, 1990a.

[Unnerstall, 1990b] T. Unnerstall. Schrödinger dynamics and physical folia of infinite mean-field quantum systems. *Communications in Mathematical Physics*, 130: 237-255, 1990b.

[Vaisman, 1991] I. Vaisman, I. On the geometric quantization of Poisson manifolds. *Journal of Mathematical Physics*, 32: 3339-3345, 1991.

[van Fraassen, 1991] B. C. van Fraassen. *Quantum Mechanics: An Empiricist View*. Oxford: Oxford University Press, 1991.

[van Hove, 1951] L. van Hove. Sur certaines représentations unitaires d'un groupe infini de transformations. *Memoires de l'Académie Royale de Belgique, Classe des Sciences*, 26: 61-102, 1951.

[Van Vleck, 1928] J. H. Van Vleck. The Correspondence Principle in the Statistical Interpretation of Quantum Mechanics. *Proceedings of the National Academy of Sciences*, 14: 178-188, 1928.

[van der Waerden, 1967] B. L. van der Waerden (ed.). *Sources of Quantum Mechanics*. Amsterdam: North-Holland, 1967.

[van Kampen, 1954] N. van Kampen. Quantum statistics of irreversible processes. *Physica* 20, 603-622, 1954.

[van Kampen, 1988] N. van Kampen. Ten theorems about quantum mechanical measurements. *Physica*, A153: 97-113, 1988.

[van Kampen, 1993] N. van Kampen. Macroscopic systems in quantum mechanics. *Physica*, A194: 542-550, 1993.

[Vanicek and Heller, 2003] J. Vanicek and E. J. Heller. Semiclassical evaluation of quantum fidelity. *Physical Review*, E68, 056208-1-5, 2003.

[Vergne, 1994] M. Vergne. Geometric quantization and equivariant cohomology. *First European Congress in Mathematics*, *Vol.* 1, pages 249-295. Boston: Birkhäuser, 1994.

[Vermaas, 2000] P. Vermaas. *A Philosopher's Understanding of Quantum Mechanics: Possibilities and Impossibilities of a Modal Interpretation*. Cambridge: Cambridge University Press, 2000.

[Vey, 1975] J. Vey. Déformation du crochet de Poisson sur une variété symplectique. *Commentarii Mathematici Helvetici*, 50: 421-454, 1975.

[Voros, 1979] A. Voros. Semi-classical ergodicity of quantum eigenstates in the Wigner representation. *Stochastic Behaviour in Classical and Quantum Hamiltonian Systems. Lecture Notes in Physics*, 93: 326-333, 1979.

[Wallace, 2002] D. Wallace. Worlds in the Everett interpretation. *Studies in History and Philosophy of Modern Physics*, 33B: 637-661, 2002.

[Wallace, 2003] D. Wallace. Everett and structure. *Studies in History and Philosophy of Modern Physics*, 34B: 87-105, 2003.

[Wan and Fountain, 1998] K. K. Wan and R. H. Fountain. Quantization by parts, maximal symmetric operators, and quantum circuits. *International Journal of Theoretical Physics*, 37: 2153-2186, 1998.

[Wan *et al.*, 1999] K. K. Wan, J. Bradshaw, C. Trueman, and F. E. Harrison. Classical systems, standard quantum systems, and mixed quantum systems in Hilbert space. *Foundations of Physics*, 28: 1739-1783, 1998.

[Wang, 1986] X.-P. Wang. Approximation semi-classique de l'equation de Heisenberg. *Communications in Mathematical Physics*, 104: 77-86, 1986.

[Wegge-Olsen, 1993] N. E. Wegge-Olsen. *K-theory and C^*-algebras.* Oxford: Oxford University Press, 1993.

[Weinstein, 1983] A. Weinstein. The local structure of Poisson manifolds. *Journal of Differential Geometry*, 18: 523-557, 1983.

[Werner, 1995] R. F. Werner. The classical limit of quantum theory. 1995. arXiv: quant-ph/9504016.

[Werner, 1989] R. F. Werner. Quantum states with Einstein-Podolsky-Rosen correlations admitting a hidden-variable model. *Physical Review*, A40: 42774281, 1989.

[Weyl, 1931] H. Weyl. *The Theory of Groups and Quantum Mechanics.* New York: Dover, 1931.

[Wheeler and Zurek, 1983] J. A. Wheeler and W. H. Zurek (eds.). *Quantum Theory and Measurement.* Princeton: Princeton University Press, 1983.

[Whitten-Wolfe and Emch, 1976] B. Whitten-Wolfe and G. G. Emch. A mechanical quantum measuring process. *Helvetica Physica Acta*, 49: 45-55, 1976.

[Wick *et al.*, 1952] C. G. Wick, A. S. Wightman, and E. P. Wigner. The intrinsic parity of elementary particles. *Physical Review*, 88: 101-105, 1952.

[Wightman, 1962] A. S. Wightman. On the localizability of quantum mechanical systems. *Reviews of Modern Physics*, 34: 845-872, 1962.

[Wigner, 1932] E. P. Wigner. On the quantum correction for thermodynamic equilibrium. *Physical Review*, 40: 749-759, 1932.

[Wigner, 1939] E. P. Wigner. Unitary representations of the inhomogeneous Lorentz group. *Annals of Mathematics*, 40: 149-204, 1939.

[Wigner, 1963] E. P. Wigner. The problem of measurement. *American Journal of Physics*, 31: 6-15, 1963.

[Williams, 2003] F. L. Williams. *Topics in Quantum Mechanics.* Boston: Birkhäuser, 2003.

[Woodhouse, 1992] N. M. J. Woodhouse. *Geometric Quantization.* Second edition. Oxford: The Clarendon Press, 1992.

[Yajima, 1979] K. Yajima. The quasi-classical limit of quantum scattering theory. *Communications in Mathematical Physics*, 69: 101-129, 1979.

[Zaslavsky, 1981] G. M. Zaslavsky. Stochasticity in quantum systems. *Physics Reports*, 80: 157-250, 1981.

[Zeh, 1970] H. D. Zeh. On the interpretation of measurement in quantum theory. *Foundations of Physics*, 1: 69-76, 1970.

[Zelditch, 1987] S. Zelditch. Uniform distribution of eigenfunctions on compact hyperbolic surfaces. *Duke Mathematical J.*, 55: 919-941, 1987.

[Zelditch, 1990] S. Zelditch. Quantum transition amplitudes for ergodic and for completely integrable systems. *Journal of Functional Analysis*, 94: 415-436, 1990.

[Zelditch, 1991] S. Zelditch. Mean Lindelöf hypothesis and equidistribution of cusp forms and Eisenstein series. *Journal of Functional Analysis*, 97, 1-49, 1991.

[Zelditch, 1992a] S. Zelditch. On a "quantum chaos" theorem of R. Schrader and M. Taylor. *Journal of Functional Analysis*, 109: 1-21, 1992a.

[Zelditch, 1992b] S. Zelditch. Quantum ergodicity on the sphere. *Communications in Mathematical Physics*, 146: 61-71, 1992b.

[Zelditch, 1996a] S. Zelditch. Quantum dynamics from the semiclassical viewpoint. Lectures at the Centre E. Borel, 1996a. Available at http://mathnt.mat.jhu.edu/zelditch/Preprints/preprints.html.

[Zelditch, 1996b] S. Zelditch. Quantum mixing. *Journal of Functional Analysis*, 140: 68-86, 1996b.

[Zelditch, 1996c] S. Zelditch. Quantum ergodicity of C^* dynamical systems. *Communications in Mathematical Physics*, 177: 507-528, 1996c.

[Zelditch and Zworski, 1996] S. Zelditch and M. Zworski. Ergodicity of eigenfunctions for ergodic billiards. *Communications in Mathematical Physics*, 175: 673-682, 1996.

[Zurek, 1981] W. H. Zurek. Pointer basis of quantum apparatus: into what mixture does the wave packet collapse? *Physical Review*, D24: 1516-1525, 1981.

[Zurek, 1982] W. H. Zurek. Environment-induced superselections rules. *Physical Review*, D26: 1862-1880, 1982.

[Zurek, 1991] W. H. Zurek. Decoherence and the transition from quantum to classical. *Physics Today*, 44 (10): 36-44, 1991.

[Zurek, 1993] W. H. Zurek. Negotiating the tricky border between quantum and classical. *Physics Today*, 46 (4): 13-15, 81-90, 1993.

[Zurek, 2003] W. H. Zurek. Decoherence, einselection, and the quantum origins of the classical. *Reviews of Modern Physics*, 75: 715-775, 2003.

[Zurek, 2005] W. H. Zurek. Probabilities from entanglement, Born's rule $p_k = |\psi|^2$ from envariance. *Physical Review*, A 71, 052105, 2005.

[Zurek *et al.*, v] W. H. Zurek, S. Habib, and J. P. Paz. Coherent states via decoherence. *Physical Review Letters*, 70: 1187-1190, 1993.

[Zurek and Paz, 1995] W. H. Zurek and J. P. Paz. Why we don't need quantum planetary dynamics: decoherence and the correspondence principle for chaotic systems. In D. H. Feng *et al.* (eds.), *Proceedings of the Fourth Drexel Meeting*. New York: Plenum, 1995. arXiv: quant-ph/9612037

第六章　量子信息与量子计算

杰弗瑞·巴布

1. 引言

对量子信息这个主题的探讨可以追溯到关于量子力学基础概念问题的那场争论。

这场争论确实是从玻尔(Bohr)与爱因斯坦(Einstein)关于量子态解释的辩论开始的，具体地说就是关于对所谓“纠缠态”的解释，其展示了存在于分离量子系统之间的特有的非定域性统计关系。这方面的争论可参见文献[Bohr，1949，283]，以及爱因斯坦在该文献同一卷上对玻尔的回应[Schilpp(谢尔博)，1949]。爱因斯坦将量子力学简单地看作是一个不完备的理论。针对处于独特纠缠态的一对量子系统中粒子关系的特定约束集的基础，爱因斯坦、波多尔斯基和罗森(Einstein，Podolsky and Rosen，写为 EPR)在一篇影响深远的文章[Einstein *et al.*，1935]中指出：纠缠现象与所有物理理论都应该遵守的分离性与定域性这个简单的事实原理相冲突，除非我们认为量子态是一种不完备的描述。

玻尔的观点——正如他定义的“完备性”——最终成为了哥本哈根正统解释的基石，之后又由海森堡(Heisenberg)、泡利

(Pauli)、冯·诺伊曼(Von Neumann)、狄拉克(Dirac)、惠勒(Wheeler)等人对该解释进行了重塑，详细讨论请参见[Howard(霍华德)，2004]以及本书的第五章。正如泡利在与马克思·玻恩(Born)的通信[Born，1971，218]中所说，“外部观测者”这类在经典物理学中的描述，是被量子现象的本质所排斥的，并且对于一个事件的量子描述(在理论上)是非常完备的。由于任何有关量子理论的应用都要求“分离”观测者与被观测对象，或者说“分离”宏观测量仪器与被测系统，因此对量子系统的描述是在一个特定意义的情境下进行的，这个相关背景是由整体的宏观实验安排所决定的。例如，“位置测量背景”提供了关于位置的信息，但同时在理论上排除了在同一时刻获得动量信息的可能性，因为在这个背景下无法得到关于动量的事实：动量的值是不确定的。哥本哈根解释与爱因斯坦的实在论相违背，他的“哲学偏见”——泡利在给玻尔的信中如是说[Born，1971，221]——植根于爱因斯坦与玻尔争论的核心，即从经典力学到量子力学过渡的真正意义。

20世纪90年代，信息的量子理论得到发展，是建立在对纠缠的合理认识基础上的，纠缠不再被认为是给过去那种只涉及哲学家的物理学制造障碍的一个微小问题，而是确实能被用来开发非经典通讯渠道，执行经典世界中做不到的信息处理任务的重要资源。在一篇由两部分组成的评论中，薛定谔(Schrödinger)[1935，555]将纠缠视为“量子理论的典型特征，使量子理论整体从经典思想中脱离了出来”。这也导致了大量研究的涌现，物理学家和计算机科学家们的这些研究集中于将信息理论的思想应用于量子计算(即开发纠缠这种资源应用于量子计算机的设计，从而使其能够有效地执行特殊的计算任务)，应用于量子通讯(“纠缠辅助”的新形式，如量子隐形传态)，以及应用于量子密码(密码协议由量子力学的定律来保证绝对安全，以防止被窃听与欺骗，甚至使用量子计算机都不能破解)。

此方面的里程碑有：贝尔(Bell)的研究[1964]将EPR佯谬的争论转向证明爱因斯坦对于分离性与定域性的假设——该假设在经典物理中成立，并且是EPR不完备性争论的核心内容——与处于EPR纠缠态的分离系统的确定性量子统计关系(EPR实验没有明确地考虑这种关系)之间是不相容的；之后的实验[Aspect *et al.*(亚斯贝克等)，1981；1982]通过某些设置将两个分离系统之间

任何可能的非超光速、经典的通讯完全排除掉，从而确证了这种非经典关系的存在。

到了20世纪80年代，大量的学者，如威斯纳(Wiesner)、本内特(Bennett)和布莱森德(Brassard)[Wiesner，1983；Bennett and Brassard，1984；Bennett *et al.*，1982]都指出，量子力学测量过程中的性质可以被开发利用，用以消除密码传递过程中被窃听的可能性，尤其是在密钥分配中——在这个过程中，信息传递的双方，爱丽丝(Alice)和鲍伯(Bob)起初并不共享任何信息，而最后他们各自拥有一个随机密钥来用于他们之间保密信息的传送。任何第三方，如夏娃(Eve)，都不能得到爱丽丝和鲍伯通讯中的任何信息，密钥的建立使除爱丽丝和鲍伯之外的夏娃的干涉能够被发现，因为夏娃的测量必然会破坏通讯信道中系统的量子态。

本内特[1973]揭示了如何设计一个可用于任何计算的可逆通用图灵机，它需要在设计量子计算机时遵循态的幺正变化(因此是可逆的)，贝尼奥夫(Benioff)[1980]发展了量子计算机的哈密顿(Hamiltonian)模型。费曼(Feynman)[1982]考虑了利用量子资源有效模拟物理系统演化的问题(并指出一个量子过程的经典模拟会导致资源的指数级消耗)，其中就涉及了量子计算的观点，不过还是多伊奇(Deutsh)[1985；1989]最先定义了一个通用量子计算机的本质特征，并且设计了第一个真正意义上的量子算法。

紧随着多伊奇对量子逻辑门和量子网络的研究之后，又产生了一些量子算法，它们在执行计算任务时比任何已知的经典算法都更加有效，或者在某些情况下比一切经典算法都有效。这些算法中最引人关注的就是计算正整数 $N=pq$ 的两个质因子的肖尔(Shor)算法[1994；1997]，它拥有比目前最好的经典算法快指数倍的运算速度。由于质因子分解是已被广泛应用的公钥加密系统的基础(目前普遍应用于互联网间银行和商业交易的通讯)，因此肖尔的结果具有重要的实践意义。

在下面的几节中，我会详细介绍量子信息、量子通讯、量子密码学以及量子计算中的一些理论的发展现状。在结束语部分，我考虑了这样的问题：从量子信息的角度做出的这种审视是否能为解决量子力学的基础问题提供一种新方法，这里说的基础问题也正是爱因斯坦和玻尔之间争议的焦点。

本文在很大程度上受惠于迈克·尼尔斯(Michael Nielsen)和伊萨克·庄(Isaac Chuang)启蒙性和综合性的著作《量子计算与量子信息》(*Quantum Computation and Quantum information*)[2000]，以及一些有洞察力的论文，如撒度·波佩斯库(Sandu Popescu)与丹尼尔·罗赫里奇(Daniel Rohrlich)[1998]的《纠缠的乐趣》(*The Joy of Entanglement*)，理查德·约萨(Richard Jozsa)[1998]的《量子信息及其特性》(*Quantum Information and its Properties*)，安德鲁·斯蒂恩(Andrew Steane)[1998]的《量子计算》(*Quantum computing*)。

2. 经典信息

2.1　经典信息压缩与香农(Shannon)熵

在这一节中，我们主要回顾经典信息理论的基本概念。在2.1中，我介绍了关于信息源的香农熵概念，以及香农信源编码理论(无噪信道编码理论)中的信息压缩方面的基本概念。在2.2中，我主要定义了一些与香农有噪信道编码定理相关的信息理论概念。

经典信息理论最初是为了解决电子信号通信中的一些问题才发展起来的。香农最初的突破性论文《通信的数学理论》(*A Mathematical Theory of Communication*)[Shannon, 1948]是建立在奈奎斯特(Nyquist)[1924]与哈特利(Hartley)[1928]等人在20世纪20年代的前期研究基础之上的。其中讨论的最基本问题是关于信息表示方法的，而要表示的这些信息是从由信息源随机组成的信息总体中挑选出来的，通过这种对信息的表示，来确保信息在某个电子线路如有噪电报线路上的有效传输。

一套通信设备由发送机或者说信息源、(很可能是有噪的)信道以及接收器组成。信息源以字母符号序列来编码信息，香农将这些符号序列数学化地表征为独立同分布随机变量的值组成的序列。在接下来的理想化中，信息源被认为是稳定的，也就是可以(粗略地)认为任意符号(或n元符号组)出现在一个(非常长的)序列中某个给定位置的概率与出现在该序列其他位置的概率是相同的，这里的位置应该被视为与所有可能序列的系综有关的；信息源还是遍历的，也就是说这种“总体平均”概率应该等价于“时间平均”概率，这里的时间平均是

指一个符号(或 n 元符号组)出现在一个(非常长)序列中的概率。

香农考虑的基础性问题是如何对存储由一个源产生的信息所需的最小物理资源进行量化，从而保证信息能够无损失地通过信道传输并且在接收器中被重建。香农信源编码理论(无噪信道编码理论)回答了这个问题。

为了探讨这个理论背后的思想，我们考虑这样一个信息源，它可以产生由有限的字母符号 a_1，a_2，…，a_k 构成的长序列(信息)，这些符号出现的概率分别被指定为 p_1，p_2，…，p_k。一个给定的符号序列被表征为一个独立同分布随机变量的值 X_1，X_2，…的序列。一个长度为 n 的典型序列，规模为 n，可容纳近似 $p_i n$ 个符号 a_i，其中 $i = 1$，…，n。因此一个足够长的典型序列(假设是相互独立的)，其概率为：

$$p(x_1, x_2, \cdots, x_n) = p(x_1)p(x_2)\cdots p(x_n) \approx p_1^{p_1 n} p_2^{p_2 n} \cdots p_n^{p_n n} \tag{1}$$

两边取对数(在信息理论中习惯于以 2 为底)，得到：

$$\log p(x_1, \cdots, x_n) \approx n\sum_i p_i \log p_i := -nH(X) \tag{2}$$

其中 $H(X) := -\sum_i p_i \log p_i$ 为信息源的香农熵。

我们可以从多种角度来思考香农意义上的信息。我们可以将 $-\log p_i$ 这个关于 p_i 的减函数(其最小值为 0，当 i 取某些值而 $p_i = 1$ 时)看作是对某些信息的量度，这些信息与识别由一个信息源产生的符号 a_i 有关。$H(X) = -\sum_i p_i \log p_i$ 则为平均信息增益，或者信息增益的期望值，它与如何确定随机变量 X 的值有关。从另一个角度看，我们可以将熵视为在确定 X 的值之前对其总的不确定性的量度。一个可以产生两个可区分符号(且出现概率相同)其中之一的源，如抛一次硬币，就具有 1 比特的香农熵：确定产生了哪个符号，或者消除了关于产生哪个符号的不确定性，与这 1 比特的信息总量相关。①如果我们已经知道哪个符号将被产生(因此概率为 1 或 0)，那么熵则为 0：没有不确定性，就没有信息增益。

由于

① 术语“比特”(二进制数)在香农熵的意义下是用来表示经典信息的基本单元，或者是一个表征基础经典信息源可能输出的基本双态经典系统。

$$p(x_1,\ \cdots,\ x_n)=2^{-nH(X)} \tag{3}$$

对于足够长的典型序列成立，以及所有典型长度为 n 的序列其概率都小于 1，那么最多有 $2^{nH(X)}$ 个典型序列。事实上，如果 p_i 并不全部相同，那么当 $n\to\infty$，典型序列则由呈指数级减小的(同概率典型序列)集合 T 构成，其包含于全部序列的集中，但是由于当 $n\to\infty$，源产生的非典型序列的概率会趋向于 0，因此典型序列集的概率就近似于 1。因此每个典型的 n-序列都可以编码为 $nH(X)$ 个二进制数或者说 $nH(X)$ 比特中的一个，然后再通过信道被发送至接收器，在接收器中原始序列再按照 1 对 1 的反向编码图被重构出来。(只是对于某些罕见的非典型序列，如序列中的每个值都被编码为零字符串时，这种重构有很小的概率会失败。)

注意如果概率 p_i 都相同(对于任意 i，有 $p_i=1/k$)，则 $H(X)=\log k$，当存在某个 $p_j=1$ 时(因此 $p_i=0$ 当 $i\neq j$ 时)，则 $H(X)=0$(由于 $0\log 0=\lim_{x\to 0} x\log x=0$)。很容易得到：

$$0\leqslant H(X)\leqslant \log k \tag{4}$$

如果我们将 k 个符号都编码为一个对应的二进制数，也就是说，编码为 0 和 1 组成的字符串，那么我们就需要 $\log k$ 比特的二进制数($2^{\log k}=k$)。因此香农的分析表明，由一个随机源产生的信息能够被压缩，具体就是：如果将 k 个符号中的每一个符号 a_i 都编码为一个对应的 0 和 1 字符串，那么长度为 n 的序列($n\to\infty$，非典型序列出现的概率趋向于 0)就可以用 $nH(X)$ 比特而非 $n\log k$ 比特的数据来实现编码且不会丢失任何信息，这就是一种压缩，因为除了同概率分布的情况外，$nH(X)<n\log k$ 都成立。

更精确地，令 $\overline{X}=\dfrac{1}{n}(X_1+X_2+\cdots+X_n)$，这里 X_1，X_2，$\cdots$，X_n 为 n 个均值为 $\langle X\rangle$ 且方差有限的独立同分布随机变量，根据弱大数定理，对于任意的 ϵ，$\delta>0$：

$$\Pr(|\overline{X}-\langle X\rangle|\geqslant\delta)<\epsilon \tag{5}$$

对于足够大的 n 成立。

考虑一个随机变量 X，在字母表 $\mathcal{X}$ 中取值为 x 的概率为 $p(x)=\Pr(X=x)$，

$x \in \mathcal{X}$①。令：

$$Z = -\log p(X) \tag{6}$$

为 X 的函数，当 X 取值为 x 时，其值为 $-\log p(x)$。那么：

$$< Z > = -\sum_{x \in \mathcal{X}} p(x) \log p(x) = H(X) \tag{7}$$

且对于独立同分布随机变量序列 X_1，X_2，…，X_n：

$$\begin{aligned} -\frac{1}{n}\log p(X_1,\cdots,X_n) &= -\frac{1}{n}\sum_i \log p(X_i) \\ &= \frac{1}{n}(Z_1 + \cdots + Z_n) \\ &= \overline{Z} \end{aligned} \tag{8}$$

因此，根据弱大数定理，对于 ϵ，$\delta > 0$ 以及足够大的 n：

$$\Pr(|\overline{Z} - \langle Z \rangle| \geqslant \delta) < \epsilon \tag{9}$$

也就是说：

$$\Pr\left(\left| -\frac{1}{n}\log p(X_1, \cdots, X_n) - H(X) \right| \geqslant \delta\right) < \epsilon \tag{10}$$

或者等价地：

$$\Pr\left(\left| -\frac{1}{n}\log p(X_1, \cdots, X_n) - H(X) \right| < \delta\right) \geqslant 1 - \epsilon \tag{11}$$

因此，其概率大于或等于 $1-\epsilon$：

$$-n(H(X) + \delta) < \log p(X_1, \cdots, X_n) < -n(H(X) - \delta) \tag{12}$$

一个随机变量 X_1，…，X_n取值为$(x_1, \cdots, x_n) \in \mathcal{X}^n$，长度为 n 的 δ 型序列可以定义为由字母表$\mathcal{X}$中的符号组成的序列，其满足：

$$2^{-n(H(X)+\delta)} \leqslant p(x_1, \cdots, x_n) \leqslant 2^{-n(H(X)-\delta)} \tag{13}$$

将长度为 n 的 δ 型序列的集记作 $T_\delta^{(n)}$，$T_\delta^{(n)}$ 中的序列数记作 $|T_\delta^{(n)}|$。那么，对

① 注意 $p(x)$ 是 $p_X(x)$ 的简写，因此 $p(x)$ 和 $p(y)$ 分别对应两个不同的随机变量。表达式 $\Pr(X \in S) = \sum_{x \in S} p(x)$ 表示随机变量 X 在集合 S 中取某个值的概率，$\Pr(X = x)$ 则表示 X 取值为 x 的概率。$p(x_1, x_2, \cdots, x_n)$ 表示随机变量序列 X_1，X_2，…，X_n 取值为 $(x_1, x_2, \cdots, x_n)$ 的概率。这里的讨论请参见卡沃尔和托马斯（Cover and Thoms）[1991]，我采用了该文中的表示法。

于足够大的n：

$$\Pr(\{X_1,\ \cdots,\ X_n\}\in T_\delta^{(n)})\geqslant 1-\epsilon \tag{14}$$

且：

$$(1-\epsilon)2^{n(H(X)-\delta)}\leqslant |T_\delta^{(n)}|\leqslant 2^{n(H(X)+\delta)} \tag{15}$$

因此，粗略地说，$T^{(n)}$包含2^{nH}个等概率序列，每个序列的概率为2^{-nH}。

香农的信源编码定理对于典型序列给出的上述结果表明，对于由一个独立同分布随机变量源产生的符号，每个符号占用$H(X)$比特的压缩率是最优的。源产生符号为x_1, x_2, …, x_n的长度为n的序列，其概率为$p(x_1, x_2, \cdots, x_n)=p(x_1)p(x_2)\cdots p(x_n)$，这里的符号都是从字符表$\mathcal{X}$中选取的。如果$\mathcal{X}$中一共有$k$个符号，那么这个$n$序列就可以由$n\log k$比特的数据序列表征。假设存在一个“分组编码”的压缩表将每个“分组”或者说长度为n的序列（足够大的n）编码为nR比特的短序列，这里$0\leqslant R\leqslant \log k$，并且假设接收器有一个解压缩表来将$nR$比特的序列译码为$n$个符号的序列。这里提到的就是压缩率为$R$的压缩/解压缩表。

信源编码定理表述为：

如果信息源的香农熵为$H(X)$，那么当且仅当$R\geqslant H(X)$时，存在一个可靠的压缩率为R的压缩/解压缩表，这里说压缩表是可靠的意思是随着$n\to\infty$该压缩表译码出原序列的概率趋向于1。

对于可靠通信，我们希望通过对符号序列压缩与解压缩后可以得到原序列，但一般来说，接收器接收到的nR编码比特序列被译码为信源生成的初始n-序列的概率是固定的，为$q(x_1,\ \cdots,\ x_n)$。对于长度为n的分组，其压缩/解压缩表的平均保真度①被定义为：

$$F_n=\sum_{\text{所有的}n\text{序列}}p(x_1,\cdots,x_n)q(x_1,\cdots,x_n) \tag{16}$$

如果所有的概率$q(x_1,\ \cdots,\ x_n)$都为1，$F_n=1$；否则，$F_n<1$。用保真度来作为译码可靠性的一种量度，则信源编码定理表述为：

① 注意这里的保真度定义与尼尔斯和庄[2000，400]提出的保真度定义不同，其是作为一种分布之间的“距离测量”，表示两个概率分布$\{p_x\}$与$\{q_x\}$之间的逼真度。该定义为$F_{NC}(p_x,q_x):=\sum_x\sqrt{p_xq_x}$，因此当$p_x=q_x$时，$F_{NC}(p_x,\ q_x)=1$。

对于任意ϵ，$\delta>0$：（1）存在一个压缩/解压缩表，其对于信息源产生的长度为n(n足够大)的序列中的每个字符采用$H(X)+\delta$比特的数据来传输，并保证原序列在接收端以$F_n>1-\epsilon$的保真度被解压；（2）任何压缩/解压缩表，如果其对于信息源产生的长度为n(n足够大)的序列中的每个字符采用$H(X)-\delta$比特的数据来传输，其保真度为$F_n<\epsilon$。

举一个压缩的简单例子，考虑一个信息源可以产生由四个字母符号a_1，a_2，a_3，a_4组成的符号序列，其概率分别为1/2，1/4，1/8，1/8。每个符号可以用一个二进制数表示为：

a_1: 00

a_2: 01

a_3: 10

a_4: 11

因此，如果不压缩的话，每个符号需要两比特的空间来存储信息源的输出。信息源的香农熵为$H(X)=-\frac{1}{2}\log\frac{1}{2}-\frac{1}{4}\log\frac{1}{4}-\frac{1}{8}\log\frac{1}{8}-\frac{1}{8}\log\frac{1}{8}=\frac{7}{4}$。香农信源编码定理告诉我们存在一个压缩表编码每个符号只需7/4比特空间而非两比特，并且这个压缩表是最优的。这个最优表是由下面的编码方式来实现的：

a_1: 0

a_2: 10

a_3: 110

a_4: 111

因此，一个压缩序列的平均长度为：每个符号$\frac{1}{2}\cdot 1+\frac{1}{4}\cdot 2+\frac{1}{8}\cdot 3+\frac{1}{8}\cdot 3=\frac{7}{4}$比特。

香农信源编码定理的意义在于其表明存在一个最优和最有效的方式来压缩信源产生的信息(设定了一定的理想化条件)，以这种方式信息在接收器中能够被可靠地重建。既然一条消息可以抽象为由一个随机信息源产生的可辨识符号组成的序列，那么这条信息中唯一与可靠的压缩或解压缩有关的性质就是与具体符号相关联的概率序列：物理系统通过其所处的状态使信息的表达具体化，

其性质与这里的压缩定义无关(只要保证不同的状态是明确可区分的)，而是承载信息的内容或意义。香农熵 $H(X)$ 是对最小物理资源的量度，通过每字符占用比特的平均数这样的定义描述，这是可靠地存储一个信息源输出的充分必要条件。在此意义下，香农熵是对每个符号所包含信息的总和的量度，这些符号都是由同一个信息源产生的。

香农的信息量度的本质定义是压缩率：作为一种物理资源的信息是可以被压缩的，并且一个信息源产生的信息总量是由信息的最优压缩率来量度的。

2.2 条件熵、交互信息、信道容量

以上分析的前提是假定信息在信源和接收器之间是通过无噪信道传输的。现在我转向介绍与有噪信道相关的一些概念，并陈述香农有噪信道编码定理。

一个信道将由随机变量 X 的值构成的输入端信息映射到由随机变量 Y 的值构成的输出端信息，如果信道是有噪的，这种映射通常不是 1 对 1 的。考虑条件概率 $p(y\mid x)$，对于任意 x, y，给定一个输入值 x 可以得到一个输出值 y。我们可以通过下式由概率 $p(x)$ 计算 $p(y)$：

$$p(y) = \sum_x p(y\mid x)p(x)$$

并且对于任意 x, y，我们也可以通过贝叶斯法则由概率 $p(y\mid x)$ 和 $p(x)$ 来计算 $p(x\mid y)$，因此对于所有 x 及确定的 y，条件分布 $p(x\mid y)$ 的香农熵为 $H(X\mid Y=y)$，那么

$$H(X\mid Y) = \sum_y p(y)H(X\mid Y=y) \tag{17}$$

被称为条件熵。它是对于所有 y 来说，$H(X\mid Y=y)$ 的期望值。如果我们将 $H(X)$，分布 $\{p(x): x\in\mathcal{X}\}$ 的熵看作对 X 值的不确定性的量度，那么 $H(X\mid Y=y)$ 为给定 Y 的值为 y 时，X 值的不确定性的量度，$H(X\mid Y)$ 则为给定 Y 的值时，X 值的平均不确定性的量度。

换个说法，长度为 $n(n\to\infty)$ 的输入序列的数量与给定的输出序列相一致，为 $2^{nH(X\mid Y)}$，也就是说，$H(X\mid Y)$ 是为了将输入序列 X 从给定的 Y 序列中识别出来，增加的额外信息中每个符号所需的平均比特数。这是由 $2^{nH(X\mid Y)}$ 个典型序列对 (x, y) 决定的，其中相关熵 $H(X, Y)$ 通过联合概率 $p(x, y)$ 来计算。因此有

$$\frac{2^{nH(X,Y)}}{2^{nH(Y)}}=2^{n(H(X,Y)-H(Y))}=2^{nH(X|Y)} \tag{18}$$

个典型序列 X 与给定序列 Y 相关。

“链式法则”等式：

$$H(X,\ Y)=H(X)+H(Y|\ X)=H(Y)+H(X|\ Y)=H(Y,\ X) \tag{19}$$

可由量的对数定义直接得到：

$$\begin{aligned} H(X:Y) &:= -\sum_{x,y}p(x,y)\log p(x,y) \\ &= -\sum_{x,y}p(x)p(y|\ x)\log(p(x)p(y|\ x)) \\ &= -\sum_{x,y}p(x)p(y|\ x)\log p(x) - \sum_{x,y}p(x)p(y|\ x)\log p(y|\ x) \\ &= -\sum_{x}p(x)\log p(x) + \sum_{x}p(x)\Big(-\sum_{y}p(y|\ x)\log p(y|\ x)\Big) \\ &= H(X)+H(Y|\ X) \end{aligned} \tag{20}$$

注意 $H(X|\ Y)\neq H(Y|\ X)$。

交互信息是衡量通过确定一个 Y 的值我们得到的关于 X 的平均信息量，也就是说，一个随机变量中包含另一个变量的总信息量，或者说通过测量另外一个变量而使一个随机变量的不确定性降低的程度。

交互信息可以由相对熵来定义，相对熵是对诸如两个概率分布间距之类间隔的量度(虽然它不是一个真的量度，由于它是非对称的并且不满足三角不等式)。分布 $p(x)$ 与 $q(x)$ 之间的相对熵可以定义为：

$$D(p\,\|\,q)\ =\ \sum_{x\in A}p(x)\log\frac{p(x)}{q(x)} \tag{21}$$

由此交互信息被定义为：

$$\cdots\cdots\cdots\cdots\ H(X:Y)\ =\ D(p(x,y)\,\|\,p(x)p(y)) \tag{22}$$
$$=\ \sum_{x}\sum_{y}p(x,y)\log\frac{p(x,y)}{p(x)p(y)}$$

其满足下式：

$$H(X:\ Y)=H(X)-H(X|\ Y)=H(Y)-H(Y|\ X) \tag{23}$$

也就是说，两个随机变量的交互信息表征的是通过测量其中一个变量而得到的另外一个变量的信息平均量：两个随机变量其中一个的初始不确定性，与测量另一

个变量的值之后其残留的不确定性之间的差异。另外，既然 $H(X, Y) = H(X) + H(Y \mid X)$，并且有：

$$H(X:Y) = H(X) + H(Y) - H(X,Y) \tag{24}$$

也即两个平均变量的交互信息是对它们包含多少相同信息的量度：由测量香农熵得到的两个随机变量信息内容的总和(这里交互信息被计算了两次)，减去它们的交互信息。注意正如我们所期望的那样，$H(X:X) = H(X)$。

对于有噪信道来说，假设 X 表征信道的输入端而 Y 表征输出端，则 $H(X:Y)$ 就表示通过确定输出端 Y 的值而得到的关于输入端 X 的平均信息总量。信道容量 C 被定义为关于全部输入分布的 $H(X:Y)$ 的上确界。

或许有些令人惊讶，但香农信道编码定理表明，多达 C 比特的信息能够以任意低的出错率通过一个有噪信道传输。准确地说，就是指：

对于一个熵 $H \leqslant C$ 的信息源，存在一个最优编码以使得由此源产生的长度为 n 的序列能够在信道上准确地传输：当 $n \to \infty$ 时，出错率趋于零。当我们试图在此信道上传输超过 C 比特的信息时，出错率将趋于 1。

这意味着存在两种方式来提升一个有噪信道如电话线的传输率。一种是通过使用更快的线缆来提高信道容量，另一种是加强信息处理(数据压缩)。

3. 量子信息

我们在第二节中探讨的有关信息的物理概念，在深层的意义上已彻底从经典力学领域转变到了量子力学领域中。这一节的目的就是指出这种转变的本质。在 3.1 节中，我将介绍一些与量子信息相关的量子力学核心概念：纠缠态、施密特(Schmidt)分解、表征纯态和混合态的密度算符系统、混合态的“纯化”、广义量子测量——正算子值测量(POVMs)、由量子操作表示的开放系统演化。在本文中我始终假定在有限维希尔伯特(Hilbert)空间开展讨论(以避开在无限维希尔伯特空间中进行函数分析时需要应对的一些技术问题)。事实上，这里的讨论也并不失一般性，因为无论对于经典还是量子信息源产生的信息，都被认为是由一些有限字母表中的符号序列构成，这些符号序列是我们依据一组经典或量子态的有限集来给出的。此外，所有与讨论经典信息与量子信息之

间的差异相关的概念性论文都是在有限维希尔伯特空间下展开的。在 3.2 节中，我介绍香农熵的冯·诺伊曼推广，以及量子信息的相关概念。在 3.3 节和 3.4 节中，我讨论一些可以将量子信息与经典信息区分开的特殊性质。其中 3.3 节主要考虑由“不可克隆”定理给出的关于复制量子信息的限制，3.4 节主要处理由霍尔夫(Holevo)边界决定的量子信息的有限存取性。最后，3.5 节主要介绍量子信息中压缩率概念的应用，并概述了舒马赫(Schumacher)所做的香农源编码定理在量子信息中的扩展，区别了针对量子信息的“可见”与“非可见”压缩。

3.1 量子力学相关概念

纠缠态

考虑一个量子系统 Q，它是一个混合系统 QE 的一部分，这里 E 代表环境，尽管 E 也可以是任意量子系统，而 Q 是它的子系统。QE 的纯态被表示为张量积希尔伯特空间$\mathcal{H}^Q \otimes \mathcal{H}^E$中的射线或单位向量。一个广义的 QE 纯态可以由下式表示：

$$|\Psi\rangle = \sum c_{ij} |q_i\rangle |e_j\rangle \tag{25}$$

这里$|q_i\rangle \in \mathcal{H}^Q$是空间$\mathcal{H}^Q$中的正交归一态的一个完备集(一组基)，$|e_j\rangle \in \mathcal{H}^E$是$\mathcal{H}^E$中的一组基。如果系数 c_{ij}使得$|\Psi\rangle$不能展开为$|Q\rangle|E\rangle$的直积态，那么$|\Psi\rangle$就被称为叠加态。

对于 QE 系统中的任意态$|\Psi\rangle$，总是存在正交归一基$|i\rangle \in \mathcal{H}^Q$，$|j\rangle \in \mathcal{H}^E$使得$|\Psi\rangle$可以表示为一个双正交相关式，如：

$$|\Psi\rangle = \sum_i \sqrt{p_i} |i\rangle |i\rangle \tag{26}$$

其中系数$\sqrt{p_i}$为非负实数，且 $\sum p_i = 1$ 。这个表达式被称为施密特分解式。当且仅当 p_i 各不相同时施密特分解式是唯一的。

双正交 EPR 态①就是一个例子：

$$|\Psi\rangle = (|0\rangle|1\rangle - |1\rangle|0>)/\sqrt{2} \tag{27}$$

① 爱因斯坦、波多尔斯基和罗森提出的比位置和动量关系更复杂的纠缠态。自旋的例子参见玻尔[1951，pp. 611—623]。

是两个自旋为1/2粒子的单态（拥有正系数的施密特形式可以通过在基矢的定义中消除相对相位来得到）。在这个单态中，$|0\rangle$和$|1\rangle$可以被用来表示z方向上自旋的两个本征态，但是既然态是对称的，那么对于任意方向上的自旋，$|\Psi\rangle$都会保持其形式不变。EPR论证采取了这样的事实，即对两个相互分离任意远的粒子在同一方向上的自旋进行测量，得到的结果在任意自旋方向上都将是完全反关联的。贝尔的反驳采用的事实是：当测得一个粒子的自旋与z轴夹角为θ_1，而另一个粒子的自旋与z轴夹角为θ_2时，测得两个粒子具有同样结果（都为1或都为0）的概率为$\frac{1}{2}\sin^2(\theta_1-\theta_2)$。由此断定当$\theta_1-\theta_2=\pi$时，测得的结果是完全相关联的，当$\theta_1-\theta_2=2\pi/3$时，测得的结果中有3/4是相同的。另一方面，从贝尔不等式，也即从爱因斯坦关于可分离性和定域性实在论的假设来讲，我们知道当$\theta_1-\theta_2=2\pi/3$时，结果的关联性不应该超过2/3。进一步的讨论请参见迪克森（Dickson）的论文。

这意味着一个量子系统的动力学演化导致的结果是会出现一个能表示相互关联的信息的态，它无法用经典计算机来模拟。也就是说，经典计算机不能通过编程来实现下面的工作：对任意一对处于不同位置的输入角度θ_1、θ_2来说，在这两个位置输出一对值（0或1），使得当$\theta_1-\theta_2=\pi$时，这对值是完全相关联的；当$\theta_1=\theta_2$时，它们是完全不相关联的；当$\theta_1-\theta_2=2\pi/3$时，它们关联性为75%；并且每一次执行的相应时间，即给定输入和得到输出之间的间隔时间，要小于光从一个位置传播到另一个位置所需的时间。

注意下面的四个态：

$$|00\rangle=\frac{1}{\sqrt{2}}(|0\rangle|0\rangle+|1\rangle|1\rangle) \tag{28}$$

$$|01\rangle=\frac{1}{\sqrt{2}}(|0\rangle|1\rangle+|1\rangle|0\rangle) \tag{29}$$

$$|10\rangle=\frac{1}{\sqrt{2}}(|0\rangle|0\rangle-|1\rangle|1\rangle) \tag{30}$$

$$|11\rangle=\frac{1}{\sqrt{2}}(|0\rangle|1\rangle-|1\rangle|0\rangle) \tag{31}$$

形成一个2×2维希尔伯特空间中的正交归一基，我们称之为贝尔基。任意一

个贝尔态都可以通过一个定域幺正变换 X、Y 或 Z，变为其他任意的贝尔态，这里的 X、Y、Z 为泡利自旋矩阵：

$$X=\sigma_x=|0\rangle\langle 1| + |1\rangle\langle 0| = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{32}$$

$$Y=\sigma_y=i|0\rangle\langle 1| - i|1\rangle\langle 0| = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \tag{33}$$

$$Z=\sigma_z=|0\rangle\langle 0| + |1\rangle\langle 1| = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{34}$$

例如：

$$X\otimes I\cdot\frac{1}{\sqrt{2}}(|0\rangle\langle 1-|1\rangle|0\rangle)=\frac{1}{\sqrt{2}}(|0\rangle\langle 0|-1\rangle|1\rangle) \tag{35}$$

如果 QE 是一个封闭系统，处于下面的纠缠纯态：

$$|\Psi\rangle = \sum_i \sqrt{p_i}|i\rangle|i\rangle \tag{36}$$

那么在施密特分解中，$\mathcal{H}^Q$空间中的任意 Q 可观测量 A 的期望值可以通过下式得到：

$$\begin{aligned}\langle A\rangle &= \mathrm{Tr}(|\Psi\rangle\langle\Psi|A\otimes I) \\ &= \mathrm{Tr}_Q(\mathrm{Tr}_E(|\Psi\rangle\langle\Psi|A)) \\ &= \mathrm{Tr}_Q(\sum_i p_i|i\rangle\langle i|A) \\ &= \mathrm{Tr}_Q(\rho A)\end{aligned} \tag{37}$$

对于$\mathcal{H}^Q$上的任意正交归一基，这里的 $\mathrm{Tr}_Q()=\sum_q\langle q_i|\cdot|q_i\rangle$ 是$\mathcal{H}^Q$空间上的偏迹，$\mathrm{Tr}_E()$是$\mathcal{H}^E$空间上的偏迹，$\rho=\sum_i p_i|i\rangle\langle i|\in\mathcal{H}^Q$ 是开放系统 Q 的约化密度算子，它是一个具有单位迹的正算子。既然由公式(37)得到密度算符 ρ 代表所有 Q 可观测量的统计，那么 ρ 可以用来表征系统 Q 的量子态。

如果 QE 是一个纠缠纯态，那么开放系统 Q 就处于一个混合态 ρ，也就是说 $\rho\neq\rho^2$；对于纯态来讲，ρ 是一个投影算符并且 $\rho=\rho^2$。由密度算符 $\rho=\sum_{p_i}|i\rangle\langle i|$ 表示的混合态可以被视为具有先验概率 p_i 的纯态 $|i\rangle$的混合，但是这种表示不是唯一的——即使组成混合态的纯态是相互正交的。例如，在二维希尔伯特空

间$\mathcal{H}_2$中，正交归一态$|0\rangle$和$|1\rangle$的等权混合态，与任意一对正交归一态如态$\frac{1}{\sqrt{2}}(|0\rangle+|1\rangle)$，$\frac{1}{\sqrt{2}}(|0\rangle-|1\rangle)$的等权混合态，或者任意非正交归一态分别呈120°角的如态$|0\rangle$，$\frac{1}{2}|0\rangle+\frac{\sqrt{3}}{2}|1\rangle$，$\frac{1}{2}|0\rangle-\frac{\sqrt{3}}{2}|1\rangle$的等权混合态，或$\mathcal{H}_2$上所有可能态的一致连续分布等一样，具有相同的统计属性，因此也具有相同的密度算符$\rho=I/2$。

更一般地，对于正交归一态$|e_i\rangle\in\mathcal{H}^E$的任意基，纠缠态$|\Psi\rangle$可以展开为：

$$|\Psi\rangle=\sum_{ij}c_{ij}|q_i\rangle|e_j\rangle=\sum_j\sqrt{w_j}|r_j\rangle|e_j\rangle \tag{38}$$

这里归一化态$|r_j\rangle=\sum_i\frac{c_{ij}}{\sqrt{w_j}}|q_i\rangle$是$|e_j\rangle$（$\sqrt{w_j}=\sum_j|c_{ij}|^2$）的相对态。注意态$|r_j\rangle$并不是普遍正交的。既然$|e_j\rangle$是正交的，我们可以将表示系统$Q$的态的密度算符展开为：

$$\rho=\sum_i w_i|r_i\rangle\langle r_i| \tag{39}$$

实际上，对本征态为$|e_i\rangle$的E可观测量进行测量，会使混合系统QE以概率w_i塌缩到态$|r_i\rangle|e_i\rangle$中的一个上，对本征态为$|i\rangle$——上面式(36)中施密特分解的正交态——的E可观测量进行测量，会使混合系统QE以概率p_i塌缩到态$|i\rangle|i\rangle$中的一个上。既然Q和E在空间上可以很大程度地彼此分离，那么E上的测量将不会影响任意Q可观测量的统计性；否则E上的测量将会允许Q和E之间超光速的信号传递。由此可以断定，混合态ρ可以看作是正交态$|i\rangle$（ρ的本征态）以权重p_i进行混合，或者是非正交相对态$|r_j\rangle$以权重w_j通过无限多的方式进行混合，这两种认识取决于在$\mathcal{H}^E$中我们选取基：

$$\rho=\sum_i p_i|i\rangle\langle i|=\sum_j w_j|r_j\rangle\langle r_j| \tag{40}$$

并且这些具有相同密度算符ρ的不同混合方式，在物理上是不可区分的。

注意任意混合态密度算符$\rho\in\mathcal{H}^Q$可以通过添加一个适当的辅助系统E来被"纯化"，在此意义下ρ就是纯态$|\Psi\rangle=\mathcal{H}^Q\otimes\mathcal{H}^E$对$\mathcal{H}^E$的偏迹。一个混合态的纯化无疑不是唯一的，而是取决于$\mathcal{H}^E$中$|\Psi\rangle$的选择。哈格斯顿—约萨—伍特

斯(Hughston-Jozsa-Wootters)定理[Hughston *et al.*, 1993]表明，对纯态 $|r_i\rangle$ 以权重 w_i 构成的任意混合态，其中 $\rho = \sum_j w_j |r_j\rangle\langle r_j|$，都存在 ρ 的纯化，以及对系统 E 进行的适当的测量使得 Q 塌缩到混合态 ρ 上。因此在 E 中的观测者可以在远距离外将 Q 制备到与密度算符 ρ 相一致的任意混合态上(当然所有这些混合态在物理上是不可区分的)。相似的结果已经由薛定谔[1936]，杰恩斯(Jaynes)[1957]以及吉辛(Gisin)[1989]所证明。对超有限冯·诺伊曼代数的推广请参见霍尔沃森(Halvorson)[2004]。

测量

标准的冯·诺伊曼"真—假"测量是与投影算符相关的，因此标准可观测量在谱表征中被表示为对投影算符进行叠加的总和，带有表示可观测量本征值的系数。这样的测量是对经典物理系统中属性测量的一种量子类比。在经典视角下，我们认为系统的一个属性是与系统态空间(相空间)中的一个子集相对应的，确定系统是否具有某种性质就等同于确定系统的态是否包含在相应的子集中。在量子力学中，与相空间中的子集对应的是希尔伯特空间中的线性闭子空间。就像经典系统中可观测量(力学量)不同的可能值对应于一个相互独立、完全穷尽集(该集的子集覆盖经典态空间)的子集一样，一个量子可观测量的不同可能值也对应于一个相互独立(即正交)、完全穷尽集(该集的子空间覆盖量子态空间)的子空间。更进一步的讨论请参阅本书迪克森的论文，以及[Mackey(麦基)，1963]和[Bub(巴布)，1997]。

在量子力学中，特别是量子信息理论中(其中，任何读取由量子态编码的量子信息的过程都会用到量子测量)，考虑比投影测量更广义的测量类型是有意义的，这里的投影测量与可观测量值的确定性相关。并且讨论广义测量与广义可观测量是一种普遍的现象。但事实上，这样的术语虽然富有启发性，但更容易让人产生误解，因为广义测量并不能显示出量子系统是否具有相应的某类广义属性。在一定程度上，这种推广的着力点是在于发现量子态与经典态在信息的表达和操作过程中存在的不同点。

为了澄清这种观点，我将转述尼尔斯和庄在此方面所做的杰出论证[2000，§2.2.3—2.2.6]。量子测量可以完全广义地特征化一类特定的相互作用，这

类相互作用存在于两个量子系统 Q（被测系统）和 M（测量系统）之间。我们假设 Q 最初处于态 $|\psi\rangle$，而 M 最初处于标准态 $|0\rangle$，这里 $|m\rangle$ 是空间 $\mathcal{H}^M$ 上"指针"本征态的一个正交归一基。这样相互作用就可以通过希尔伯特空间 $\mathcal{H}^Q\otimes\mathcal{H}^M$ 上的幺正变换 U 来定义，从而得到下式的变换：

$$|\psi\rangle|0\rangle \xrightarrow{U} \sum_m M_m|\psi\rangle|m\rangle \tag{41}$$

这里 $\{M_m\}$ 是一个定义在 $\mathcal{H}^Q$ 上的线性算符（Kraus 算符）集，满足完备性条件：

$$\sum_m M_m^{\dagger}M_m = I \tag{42}$$

符号 $\dagger$ 是代表伴随共轭或厄米（Hermitian）共轭。完备性条件保证了这一演化是幺正的，因为它保证 U 会保持内积不变，也即：

$$\begin{aligned}|\phi\rangle\langle 0|U^{\dagger}U|\psi\rangle|0\rangle &= \sum_{m,m'}\langle m|\langle\phi|M_m^{\dagger}M_{m'}|\psi\rangle|m'\rangle \\ &= \sum_m\langle\phi|M^{\dagger}M|\psi\rangle \\ &= \langle\phi|\psi\rangle\end{aligned} \tag{43}$$

该式表明，由上面式（41）定义的 U，对任意直积态 $|\psi\rangle|0\rangle$（其中 $|\psi\rangle\in\mathcal{H}^Q$），可以展开为一个希尔伯特空间 $\mathcal{H}^Q\otimes\mathcal{H}^M$ 上的幺正算符。相应地，任何定义在系统 Q 所在希尔伯特空间上且满足完备性条件的线性算符集 $\{M_m\}$，可以在上述广义的意义下定义一次测量，其指数 m 标示测量的可能结果，且任意这样的集都会被看作是一个测量算符的集。

如果我们现在在 M 上执行一次标准投影测量，以确定指针可观测量的值 m，其中指针可观测量由投影算符定义：

$$P_m = I_Q\otimes|m\rangle\langle m|$$

那么，由式（37）①可知，得到结果 m 的概率为：

$$\begin{aligned}p(m) &= \langle 0|\langle\psi|U^{\dagger}P_mU|\psi\rangle|0\rangle \\ &= \sum_{m'm''}\langle m'|\langle\psi|M_{m'}^{\dagger}(I_Q\otimes|m\rangle\langle m|)M_{m''}|\psi\rangle|m''\rangle \\ &= \sum_{m'm''}\langle\psi|M_{m'}^{\dagger}\langle m'|m\rangle\langle m|m''\rangle M_{m''}|\psi\rangle\end{aligned} \tag{44}$$

---

① 投影算符（具有本征值 0 和 1 的幂等可观测量）的期望值，等于得到本征值 1 的概率。这里本征值 1 对应于结果 m。

$$= \langle \psi | M_m^{\dagger} M_m | \psi \rangle$$

并且，更普遍地，如果 Q 的初态是一个混合态 ρ，那么：

$$p(m) = \mathrm{Tr}_Q(M_\rho M^{\dagger}) \tag{45}$$

当完成 M 上的测量且得到结果为 m 时，QM 的终态为：

$$\frac{P_m U | \psi \rangle | 0 \rangle}{\sqrt{\langle \psi | U^{\dagger} P U | \psi \rangle}} = \frac{M_m | \psi \rangle | m \rangle}{\sqrt{\langle \psi | M_m^{\dagger} M_m | \psi \rangle}} \tag{46}$$

因此，M 的终态为 $| m \rangle$，Q 的终态为：

$$\frac{M_m | \psi \rangle}{\sqrt{\langle \psi | M_m^{\dagger} M_m | \psi \rangle}}$$

并且，更普遍地，如果 Q 的初态为混合态 ρ，那么 Q 的终态为：

$$\frac{M_m \rho M_m^{\dagger}}{\sqrt{\mathrm{Tr}_Q(M_m \rho M_m^{\dagger})}}$$

注意测量的这种广义定义包含了标准投影测量的情况。在这种情况下，$\{M_m\} = \{P_m\}$，其中 $\{P_m\}$ 是投影算符的集，它由可以用一个自伴算符表示的标准量子可观测量的谱测量来定义。该定义还包含了对那些与正算子取值测量(POVM)相关的“广义可观测量”的测量。令：

$$E_m = M_m^{\dagger} M_m \tag{47}$$

那么集 $\{M_m\}$ 就定义了一个正算符(“作用”)的集，以使得：

$$\sum E_m = I \tag{48}$$

式(48)可被看作定义了一个不要求投影取值测量(PVM)满足正交条件

$$P_m P_{m'} = \delta_{mm'} P_m \tag{49}$$

的“单位元分解”，在这种意义下，POVM 可以被看作 PVM 的推广。注意对于一次 POVM：

$$p(m) = \langle \psi | E_m | \psi \rangle \tag{50}$$

给定一个正算符的集 $\{E_m\}$，以使得 $\sum E_m = I$，测量算符 M_m 可以由下式定义：

$$M_m = U \sqrt{E_m} \tag{51}$$

这里 U 是一个幺正矩阵，该式表明：

$$\sum_m M_m^{\dagger} M_m = \sum E_m = I \tag{52}$$

当然，作为一个特例，我们可以取 $U=1$ 且 $M_m=\sqrt{E_m}$。反过来，给定一个测量算符的集$\{M_m\}$，就存在一个幺正算符 U_m，使得 $M_m=U_m\sqrt{E_m}$，这里$\{E_m\}$是一个 POVM，这可以从[Nielsen and Chuang, 2000, Theorem 2.3, p.78]中直接得到，参见[Nielsen and Chuang, 2000, Exercise 2.63, p.92]。

除了投影测量的表征情况外，我们可能还想知道的是，为什么挑选出这样的幺正变换是有意义的，为什么在广义的情况下这样的过程能被称为对 Q 的测量。接下来的例子会回答上述问题，参见[Nielsen and Chuang, 2000, 92]。假设我们知道存在于二维希尔伯特空间中的一个系统处于下面两个非正交态其中的一个：

$$|\psi_1\rangle=|0\rangle$$

$$|\psi_2\rangle=\frac{1}{\sqrt{2}}(|0\rangle+|1\rangle)$$

我们不可能通过量子测量来确定地区分这两个态，即使在上述的广义意义上也不行。这里所谓的“确定”是指完全可靠的确定，出错的概率为零。

接下来，假设存在这样的测量，由两个满足完备性条件的测量算符 M_1、M_2 来定义。当态为$|\psi_1\rangle$时，我们用

$$p(1)=\langle\psi_1|M_1^\dagger M_1|\psi_1\rangle=1 \tag{53}$$

来表示确定性；当态为$|\psi_2\rangle$时，我们用

$$p(2)=\langle\psi_2|M_2^\dagger M_2|\psi_2\rangle=1 \tag{54}$$

来表示确定性。由完备性条件得到：

$$\langle\psi_1|M_1^\dagger M_1+M_2^\dagger M_2|\psi_1\rangle=1 \tag{55}$$

由此可得：$\langle\psi_1|M_2^\dagger M_2|\psi_1\rangle=0$，也即 $M_2|\psi_1\rangle=M_2|0\rangle=0$。因此：

$$M_2|\psi_2\rangle=M_2\frac{1}{\sqrt{2}}(|0\rangle+|1\rangle)=\frac{1}{\sqrt{2}}M_2|1\rangle \tag{56}$$

且：

$$p(2)=\langle\psi_2|M_2^\dagger M_2|\psi_2\rangle=\frac{1}{2}\langle 1|M_2^\dagger M_2|1\rangle \tag{57}$$

但是由完备性条件我们可知：

$$\langle 1|M_2^\dagger M_2|1\rangle\leqslant\langle 1|M_1^\dagger M_1+M_2^\dagger M_2|1\rangle=\langle 1|1\rangle=1 \tag{58}$$

由此得到：

$$p(2) \leqslant \frac{1}{2} \tag{59}$$

而这是与式(54)相矛盾的。

然而在一般意义上我们可以实施一次测量并得到三个可能结果，这将允许我们在某段时间内准确地确定状态，也就是说，当出现可能结果中的前两个时，我们能够确定状态，当出现第三个可能结果时，我们不能对状态做出任何判断。

以下是这三个算符：

$$E_1 = \frac{\sqrt{2}}{1+\sqrt{2}} \frac{(|0\rangle - |1\rangle)(\langle 0| - \langle 1|)}{2} \tag{60}$$

$$E_2 = \frac{\sqrt{2}}{1+\sqrt{2}} |1\rangle\langle 1|$$

$$E_3 = I - E_1 - E_2$$

它们都是半正定算子，且满足 $E_1 + E_2 + E_3 = I$，因此它们可以定义一个 POVM 测量。事实上，E_1、E_2、E_3 分别是下面三个态的投影算符的倍数：

$$|\phi_1\rangle = |\psi_2\rangle^{\perp} \tag{61}$$

$$|\phi_2\rangle = |\psi_1\rangle^{\perp}$$

$$|\phi_3\rangle = \frac{(1+\sqrt{2})|0\rangle + |1\rangle}{\sqrt{2\sqrt{2}(1+\sqrt{2})}}$$

其系数分别为$\frac{\sqrt{2}}{1+\sqrt{2}}$，$\frac{\sqrt{2}}{1+\sqrt{2}}$，$\frac{1}{1+\sqrt{2}}$。测量涉及系统 M，其存在三个正交指针态 $|1\rangle$，$|2\rangle$，$|3\rangle$。适当的幺正相互作用 U 导致对输入态 $|\psi\rangle$ 进行如下变换：

$$|\psi\rangle|0\rangle \xrightarrow{U} \sum_m M_m|\psi\rangle|m\rangle \tag{62}$$

这里 $M_m = \sqrt{E_m}$。

如果输入态 $|\psi_1\rangle = 0$，我们得到的变换为：

$$|\psi_1\rangle|0\rangle \xrightarrow{U} \sqrt{E_1}|0\rangle|1\rangle + \sqrt{E_3}|0\rangle|3\rangle \tag{63}$$

$$=\alpha|\phi_1\rangle|1\rangle+\beta|\phi_3\rangle|3\rangle$$

因为$|\sqrt{E_2}|\psi_1\rangle=\sqrt{E_2}|0\rangle=0$。如果输入态$|\psi_2\rangle=\frac{1}{\sqrt{2}}(|0\rangle+|1\rangle)$，我们得到的变换为：

$$|\psi_2\rangle|0\rangle\xrightarrow{U}\sqrt{E_2}\frac{|0\rangle+|1\rangle}{\sqrt{2}}|2\rangle+\sqrt{E_3}\frac{|0\rangle+|1\rangle}{\sqrt{2}}|3\rangle \tag{64}$$

$$=\gamma|\phi_2\rangle|2\rangle+\delta|\phi_3\rangle|3\rangle$$

因为$\sqrt{E_1}|\psi_2\rangle=\sqrt{E_1}\frac{|0\rangle+|1\rangle}{\sqrt{2}}=0$，且这里$\alpha$，$\beta$，$\gamma$，$\delta$都是实系数。

我们发现，对M的指针态进行投影测量，若其结果为$m=1$，则可以确定地表明输入态为$|\psi_1\rangle=|0\rangle$。在这种情况下，测量将会导致系统Q塌缩在态$|\phi_1\rangle$上。若其结果为$m=2$，则可以确定地表明输入态为$|\psi_2\rangle=\frac{1}{\sqrt{2}}(|0\rangle+|1\rangle$，且测量将会导致系统$Q$塌缩在态$|\phi_2\rangle$上。若其结果为$m=3$，则输入态既是$|\psi_1\rangle=|0\rangle$也是$|\psi_2\rangle=\frac{1}{\sqrt{2}}(|0\rangle+|1\rangle)$，且系统$Q$塌缩在态$|\phi_3\rangle$上。

量子操作

当一个闭系统QE在幺正算符下演化时，可以将Q看作是在量子操作下的演化，也就是说，一个完全正线性映射：

$$\mathcal{E}: \rho\to\rho' \tag{65}$$

这里：

$$\mathcal{E}(\rho)=\mathrm{Tr}_E(U\rho\otimes\rho_E U^\dagger) \tag{66}$$

参见[Nielsen and Chuang，2000，356ff]。在$\mathcal{E}(\sum_i p_i\rho_i)=\sum_i p_i\,\mathcal{E}(p_i)$的意义下，映射$\mathcal{E}$是线性(或凸线性)的，在映射$\mathcal{E}$是从正算符到正算符的映射的意义下，它是正定的，且在$\mathcal{E}\otimes I$是一个从$\mathcal{H}^Q$扩展到希尔伯特空间$\mathcal{H}^Q\otimes\mathcal{H}^E$上的正映射的意义下，则它是完备正定的，这里$\mathcal{H}^Q\otimes\mathcal{H}^E$是与$Q$和其任意附属系统$E$构成的联合系统相关的。

希尔伯特空间$\mathcal{H}^Q$上的每次量子操作(即完备正线性映射)都可以被表示为一个(但不是唯一的)扩展了的希尔伯特空间$\mathcal{H}^Q\otimes\mathcal{H}^E$上的幺正演化，也即：

$$\mathcal{E}(\rho) = \mathrm{Tr}_E(U(\rho \otimes \rho_E)U^\dagger) \tag{67}$$

这里 ρ_E 是附属系统 E(我们可以将其看作 Q 所处的环境)的一个经过恰当选择的初态。我们有足够的理由将 ρ_E 看作是一个纯态，也即 $|0\rangle\langle 0|$，既然 E 的一个混合态总是能够通过扩大希尔伯特空间来纯化(即增加一个附属系统)。因此由一次量子操作描述的系统 Q 的演化总是能够被模型化为系统 QE 的幺正演化，其中 E 为一个初始的纯态。

并且，希尔伯特空间 $\mathcal{H}^Q$ 上的每次量子操作总是存在一个并非唯一内在于 $\mathcal{H}^Q$ 的算符直和表示：

$$\mathcal{E}(\rho) = \sum_i E_i \rho E_i^\dagger \tag{68}$$

其中 $E_i = \langle i|U|0\rangle$，对于 E 的一些正交归一基 $\{|i\rangle\}$ 来说。参见[Nielsen and Chuang，2000，Theorem 8.1，p. 368]。如果操作是保迹(trace-preserving)(或不可选择)的，那么 $\sum_i E_i^\dagger E_i = I$。对于非保迹(或可选择)的操作，$\sum_i E_i^\dagger E_i \leqslant I$。对应的情况是 QE 上的一次测量的结果被(选择性地)计入变换 $\mathcal{E} \to \mathcal{E}(\rho)$ 中。

如果 Q 和 E 之间不存在相互作用，那么 $\epsilon(\rho) = U_Q \rho U_Q^\dagger$，$U_Q U_Q^\dagger = I$，也就是说，一共就只有一个算符。在这种情况下，$U = U_Q \otimes U_E$ 且：

$$\mathcal{E}(\rho) = \mathrm{Tr}_E(U_Q \otimes U_E(\rho \otimes |0\rangle\langle 0|)U_Q^\dagger \otimes U_E^\dagger) \tag{69}$$

$$= U_Q \rho U_Q^\dagger \tag{70}$$

因此幺正演化是量子操作算符直和表示的一个特例，当然，另一个特例是变换 $\mathcal{E} \to \mathcal{E}(\rho)$，它发生在量子测量过程中，这里 $E_i = M_i$。一次保迹操作对应于一次不可选择测量：

$$\mathcal{E}(\rho) = \sum_i M_i \rho M_i^\dagger \tag{71}$$

而一次非保迹操作对应于一次可选择测量，这里态“塌缩”到相应的测量结果：

$$M_i \rho M_i^\dagger / \mathrm{Tr}(M_i \rho M_i^\dagger) \tag{72}$$

将算符直和表示的是应用于可能是不同的输入希尔伯特空间和输出希尔伯特空间之间的量子操作，就可以描述为下面的一般情形：一个处于未知初态 ρ 的量子系统与处于标准态的其他系统进行幺正相互作用，之后这个混合系统的一部分就被废弃了，而末态为 ρ'。我们就把从 $\rho \to \rho'$ 的变换定义为量子操作。因此，一次量子操作完全可以一般性地表示一个封闭系统的幺正演化，一个与其

环境发生相互作用的开放系统的非幺正演化，以及由幺正相互作用和选择性测量或非选择性测量的一次结合导致的演化。

如我们所知，关于扩大化的希尔伯特空间的一个基本观点就是：在扩大化的希尔伯特空间中，每个态都可以被纯化，每次测量都可以成为理想化的，每次演化都可以是幺正的。①

3.2 冯·诺伊曼熵

在本节中，我将定义一个量子混合态的冯·诺伊曼熵(用来描述经典信息源概率分布的香农熵的冯·诺伊曼推广)，条件熵和交互信息的相关概念。

香农意义上的信息是一类可计量的资源，它与符号状态下(适当理想化的)随机源的输出相关，这里包含这些态的系统的物理性质与和源相关的经典信息的总量无关。与随机源相关的信息的量通过其最优压缩率来定义，并且由香农熵来给出。随机源输出的一些性质能够被最优化地精简，这个事实最终反映了一个可计量资源对信息源的贡献多少。

信息在物理上被表示为物理系统的状态。经典信息和量子信息的本质区别在于经典状态和量子态的可辨识属性不同。正如我们接下来会看到的那样，只有正交量子态的集是完全可辨识的(也即出错率为零)，而不同经典状态的集也是完全可辨识的(经典状态的集由相空间中不相交的单元素子集表示，因此它们也是正交的，因为相交空间中的子集在某种意义上类似于希尔伯特空间中的正交子空间)。

经典信息是由一组可区分的状态——经典系统的状态或正交量子态——表示的那类信息，因此可以被看作是量子信息的一个子类，在量子信息中态既可以是可区分的也可以是不可区分的。量子信息理论的预设观点是将香农的压缩率概念扩展为关于量子态的一个随机源，其中量子态既可以是可区分的也可以是不可区分的。为此我们需要对适当测量进行定义，从而使香农熵的概念得以

---

① 这个观点是由约翰·斯莫林(John Smolin)创造的，这个公式的提出归功于本·舒马赫。请参见他的论著《关于量子信息理论的讲座笔记》(*Lecture Notes on Quantum Information Theory*)[1998]。

一般化，该测量是对量子态——混合态——概率分布中的信息进行的测量。

考虑处于纠缠态$|\Psi\rangle$的系统QE。其子系统Q处于一个混合态ρ，通常可以展开为：

$$\rho = \sum_i p_i |i\rangle\langle i| \tag{73}$$

其中p_i为ρ的本征值，纯态$|i\rangle$为ρ的正交本征态。上式是ρ的谱表示，任意密度算符——正算符(因此是厄米算符)——也可以通过这种方式展开。当且仅当p_i都是不同时这种表示是唯一的。如果p_i中的一些是相同的，那么存在一个关于ρ的唯一表示，可视为具有不同为系数p_i值的投影算符的总和，不过其中一些投影算符会投影到多维子空间上。

既然ρ有单位迹，$\sum p_i = 1$，且ρ的谱表示代表了一个正交的经典概率分布，那么它就是可区分的纯态。如果我们测量一个本征态为$|i\rangle$的Q可观测量，那么测量结果就会与一个随机变量X的值相关联，这里$\Pr(X=i)=p_i$。那么：

$$H(X) = -\sum p_i \log p_i \tag{74}$$

就是测量结果概率分布的香农熵。

现在：

$$-\mathrm{Tr}(\rho\log\rho) = -\sum p_i \log p_i \tag{75}$$

因为$\rho\log\rho$的本征值为$p_i\log p_i$，并且算符的迹是本征值的和，因此对于密度算符为ρ的任意混合量子态，香农熵的自然推广就是冯·诺伊曼熵①：

$$S := -\mathrm{Tr}(\rho\log\rho) \tag{76}$$

它与对ρ的本征基的测量的香农熵是相一致的。对于一个完备的混合态$\rho = I/d$，这里$\mathcal{H}^Q$的维度为d，ρ的d个本征值全都等于$1/d$且$S = \log d$。这是S在d维希尔伯特空间中的最大值。当且仅当ρ为纯态时，冯·诺伊曼熵S为零，是最小值，这里ρ的本征值为1或0。因此$0 \leqslant S \leqslant \log d$，这里$d$为$\mathcal{H}^Q$的维度。

回顾前面的内容，我们是把香农熵视为对平均信息量的测量，这些信息通过标识一个已知随机源产生的态来获得。或者，香农熵表示由信息源产生的信

---

① 冯·诺伊曼最先在热力学理论的基础中定义了这个量，参见[1955, 379]。

息的最优压缩率。冯·诺伊曼熵一般来说并不表示通过标识一个可表征为混合态的随机源产生的量子态的方式而获得的信息量，因为一个混合态中的非正交量子态是不能被确定地识别的。然而，我们将会在3.5节中提到，冯·诺伊曼熵可以通过舒马赫信源编码定理解释为压缩率。量子信息中的舒马赫信源编码定理[Schumacher, 1995]，是经典信息中香农信源编码定理的推广。对于一个基本的双态量子系统，它所在的二维希尔伯特空间被看作可以表征一个基本量子信源的输出，$S=1$ 代表两个正交态的等权分布（也就是说，代表密度算符 $\rho = I/2$），因此舒马赫将量子信息的基本单元视为“量子比特”。类比于术语“比特”，术语“量子比特”指的是就冯·诺伊曼熵而言的量子信息的基本单元，或是表示一个基本的双态量子系统，它被认为可以表征一个基本量子信息源的可能输出。

通过考虑条件熵和交互信息的量子概念（对比于2.2节），由冯·诺伊曼熵 S 测定的量子信息和由香农熵 H 测定的经典信息之间的差别就会呈现出来，特别是与量子信息相关的不可及性（inaccessibility）这个奇怪的特性。

对于一个复合系统 AB，类似于香农熵的对应概念，可对比于方程（19）、（23）和（24），条件冯·诺伊曼熵和交互信息可以由相关熵 $S(AB) = -\mathrm{Tr}(\rho^{AB}\log\rho^{AB})$ 来定义：

$$S(A \mid B) = S(A, B) - S(B) \tag{77}$$

$$S(A : B) = S(A) - S(A \mid B) \tag{78}$$

$$= S(B) - S(B \mid A) \tag{79}$$

$$= S(A) + S(B) - S(A, B) \tag{80}$$

相关熵满足子可加性不等式（Subadditivity inequality）：

$$S(A, B) \leqslant S(A) + S(B) \tag{81}$$

等号成立的条件是当且仅当 A 和 B 是非关联的，即 $\rho^{AB} = \rho^{A} \otimes \rho^{B}$。

这里，$S(A \mid B)$ 可以取负值，而条件香农熵只能为正数或零。考虑纠缠态为 $|\Psi\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$ 的例子。当 $|\Psi\rangle$ 为纯态时，$S(A, B) = 0$，但是 $S(A) = S(B) = 1$。因此 $S(A \mid B) = S(A, B) - S(A) = -1$。事实上，对于复合系统 AB 的纯态 $|\Psi\rangle$ 来说，当且仅当 $|\Psi\rangle$ 为纠缠态时 $S(A \mid B) < 0$。

对处于直积态的$\rho\otimes\sigma$复合系统AB来说，由相关熵的定义可得：

$$S(A,\ B)=S(\rho\otimes\sigma)=S(\rho)+S(\sigma)=S(A)+S(B) \tag{82}$$

如果AB处于纯态$|\Psi\rangle$，由施密特分解定理可知$|\Psi\rangle$被表示为：

$$|\Psi\rangle = \sum_i \sqrt{p_i}|i\rangle\langle i| \tag{83}$$

由此可得：

$$\rho_A = \mathrm{Tr}_B(|\psi\rangle\langle\psi|) = \sum_i |i\rangle\langle i| \tag{84}$$

$$\rho_B = \mathrm{Tr}_A(|\psi\rangle\langle\psi|) = \sum_i |i\rangle\langle i|$$

因此：

$$S(A) = S(B) = -\sum_i p_i \log p_i \tag{85}$$

考虑由态ρ_i以权重p_i混合构成的混合态。可知下式：

$$S(\sum_i p_i\rho_i) \leqslant H(p_i) + \sum_i p_i S(\rho_i) \tag{86}$$

当且仅当ρ_i在正交子空间中有支集(support)时，等号成立，参见[Nielsen and Chuang, 2000, Theorem 11.10, p. 518]。熵$H(p_i)$被描述为混合态ρ的预备熵。

如果态ρ_i为纯态，那么$S(\rho)\leqslant H(p_i)$。例如，假设$\mathcal{H}^Q$为二维空间，且$p_1=p_2=1/2$，那么$H(p_i)=1$。因此，如果我们有一个经典信息源能够以相同概率生成符号1和2，那么信息的压缩将是不可能的。然而，如果符号1和2被编码为非正交量子态$|r_1\rangle$和$|r_2\rangle$，则$S(\rho)<1$。我们在3.5节中将会提到，根据舒马赫信源编码定理，既然$S(\rho)<1$，量子压缩就是可能的。也就是说，根据信息源产生的每个量子态，我们可以用$S<1$个量子比特来可靠地传播量子比特的长序列。

值得注意的是，如果AB处于由态$\rho_i\otimes|i\rangle\langle i|$按权重$p_i$构成的混合态上，这里$\rho_i$可以是任意密度算符，不必一定是正交的，那么根据式(86)、式(82)，以及$S(|i\rangle\langle i|)=0$可得：

$$\begin{aligned} S(\sum_i p_i\rho_i\otimes|i\rangle\langle i|) &= H(p_i) + \sum_i p_i S(\rho_i\otimes|i\rangle\langle i|) \\ &= H(p_i) + \sum_i p_i S(\rho_i). \end{aligned} \tag{87}$$

由态ρ_i以权重p_i混合构成的混合态的冯·诺伊曼熵$\sum p_i\rho_i$，是分布态的

凸函数，即：

$$S(\sum_i p_i\rho_i) \geqslant \sum_i p_i S(\rho_i) \tag{88}$$

为了理解上式，考虑一个复合系统 AB，它处于下面的态：

$$\rho^{AB} = \sum p_i\rho_i \otimes |i\rangle\langle i| \tag{89}$$

我们知道：

$$S(A) = S(\sum_i p_i\rho_i) \tag{90}$$

$$S(B) = S(\sum_i p_i |i\rangle\langle i|) = H(p_i) \tag{91}$$

以及：

$$S(A,B) = H(p_i) + \sum_i p_i S(\rho_i) \tag{92}$$

可由式(87)得到。由于子可加性 $S(A)+S(B) \geqslant S(A,B)$，因此：

$$S(\sum_i p_i\rho_i) \geqslant \sum_i p_i S(\rho_i) \tag{93}$$

结果是投影测量总是会使熵增加，也就是说，如果 $\rho' = \sum_i P_i\rho P_i$，那么 $S(\rho') \geqslant S(\rho)$，但是广义的测量能使熵减少。例如，考虑对一个量子比特的整体测量，该量子比特处于初态 ρ，广义的测量由测量算符 $M_1 = |0\rangle\langle 0|$ 和 $M_2 = |0\rangle\langle 1|$ 来定义。注意这两个算符确实能定义广义的测量，因为$M_1^\dagger M_1 + M_2^\dagger M_2 = |0\rangle\langle 0| + |1\rangle\langle 1| = I$。测量之后有：

$$\begin{aligned}\rho' &= |0\rangle\langle 0|\rho|0\rangle\langle 0| + |0\rangle\langle 1|\rho|1\rangle\langle 0| \\ &= \mathrm{Tr}(\rho)|1\rangle\langle 1| \\ &= |1\rangle\langle 1|\end{aligned} \tag{94}$$

因此：

$$S(\rho') = 0 \leqslant S(\rho)$$

3.3 “不可克隆”定理

在 3.1 节中我们指出两个非正交量子态不能通过任何测量确定地区分开来。“不可克隆”定理则决定了非正交量子态不能被复制。为了对其进行说明，我们假设存在一个设备 D，能够复制系统 Q 的任意量子态，这里系统 Q 的态处于空间$\mathcal{H}^Q$中。假设设备 D 的初始预备状态为 $|0\rangle \in \mathcal{H}^D$。那么对于输入态

$\{|i\rangle\}$的任意正交归一集，我们要求：

$$|i\rangle|0\rangle \xrightarrow{U} |i\rangle|i\rangle \tag{95}$$

这里 U 是执行复制过程的幺正变换。根据线性要求，对于任意输入态 $\sum_i c_i|i\rangle$ 来说，有：

$$\left(\sum_i c_i|i\rangle\right)|0\rangle \xrightarrow{U} \sum_i c_i|i\rangle|i\rangle \tag{96}$$

但是对于复制来说，我们要求：

$$\left(\sum_i c_i|i\rangle\right)|0\rangle \xrightarrow{U} \left(\sum_i c_i|i\rangle\right)\left(\sum_i c_i|i\rangle\right) \tag{97}$$

且：

$$\sum_i c_i|i\rangle|i\rangle \neq \left(\sum_i c_i|i\rangle\right)\left(\sum_i c_i|i\rangle\right) = \sum_{ij} c_i c_j|i\rangle|j\rangle \tag{98}$$

除非 $c_i c_j = \delta_{ij}$，这意味着该设备不能复制不属于正交归一集 $|i\rangle$ 的任何态。

或者，我们可能会注意到如果两个态 $|\psi\rangle$ 和 $|\phi\rangle$ 不能被复制，那么：

$$|\psi\rangle|0\rangle \xrightarrow{U} |\psi\rangle|\psi\rangle \tag{99}$$

$$|\phi\rangle|0\rangle \xrightarrow{U} |\phi\rangle|\phi\rangle \tag{100}$$

既然幺正变换保持内积不变，那么我们要求：

$$\langle\psi|\phi\rangle = \langle\psi|\phi\rangle\langle\psi|\phi\rangle \tag{101}$$

当且仅当 $\langle\psi|\phi\rangle = 1$ 或 0 时，这是可能的。也就是说，如果克隆是可能的话，那么态要么是同一的要么是正交的。

狄克斯(Dieks)[1982]和伍特斯与朱瑞克(Wootters and Zurek)[1982]分别独立地证明了“不可克隆”定理。这个结果的一个非常重要的推广是关于混合态的，是由巴纳姆(Barnum)、科夫斯(Caves)、福斯(Fuchs)、约萨和舒马赫[1996a]完成的。在一个克隆过程中，系统 B 的一个预备态 σ 和系统 A 中将要被克隆的态 ρ 被转变为 ρ 的两个复制品。在更一般的传播过程中，预备态 σ 和将要被传播的态 ρ 被转变为 AB 的一个新态 ω，这里边界态 ω 涉及 A 和 B 时都是 ρ，也即：

$$\rho_A = \mathrm{Tr}_B(\omega) = \rho \tag{102}$$

$$\rho_B = \mathrm{Tr}_A(\omega) = \rho \tag{103}$$

“不可克隆”定理表明，当且仅当态是相互正交的纯态的集时能够被克隆。“不可传播”定理表明，量子态的任意集能被传播当且仅当它们由相互对易的密度算符表示。在经典信息中，从正式意义上讲，由于所有纯态都是正交的，并且所有算符(它们在相空间中表示真值函数)都是对易的，因此克隆和传播都是可能的。注意对于纯态来说，传播可以简化为克隆。

当然，制造一个有特定目的的设备来克隆一个给定(已知)的量子态 $|\psi\rangle$ 总是可能的，因为这不过就是一台制备态 $|\psi\rangle$ 的设备。“不可克隆”定理，从另外一种观点看，其实就是量子测量问题的一种陈述(参见第7节)：测量。从经典意义上讲，就是将第一个系统中的态在第二个系统中制作一个拷贝(或者更一般地，一个可以表示第一个系统中态的“指针态”)，但是在量子力学中测量是不可能的，除了对输入态的正交集的测量。

引出式(99)—式(101)的那个论题通过修改可表述为：我们不可能在没有干扰非正交态的情况下得到有关其特征的任何信息。假设设备 D 是一个测量仪器，它记录了关于输入态特征的一些信息，换言之，设备的输出态会随输入态 $|\psi\rangle$ 和 $|\phi\rangle$ 的不同而不同，且该设备不会改变输入态。即：

$$|\psi\rangle|0\rangle \xrightarrow{U} |\psi\rangle|\psi'\rangle \tag{104}$$

$$|\phi\rangle|0\rangle \xrightarrow{U} |\phi\rangle|\phi'\rangle \tag{105}$$

由上面的式子可知：

$$\langle\psi|\phi\rangle = \langle\psi|\phi\rangle\langle\psi'|\phi'\rangle \tag{106}$$

因此：

$$\langle\psi'|\phi'\rangle = 1 \tag{107}$$

因为当 $|\psi\rangle$ 和 $|\phi\rangle$ 非正交时，$\langle\psi|\phi\rangle \neq 0$。换句话说，如果没有对非正交输入态造成影响的话，就不可能获得关于态的性质的信息。因此，比如说，一个窃听者(夏娃)，假如她不对这个态进行干扰的话，不会得到有关用于爱丽丝和鲍伯之间通讯的非正交量子态的性质的任何信息，这就意味着在量子信息中被动的窃听是不可能的。

我们说当且仅当它们是相互正交的纯态的集可被克隆，就等同于说当且仅

当它们是相互正交的纯态的集可被确定地区别开来。因为如果我们能将一对量子态$|\psi\rangle$和$|\phi\rangle$区别开来，那么我们只要利用专门制备$|\psi\rangle$和$|\phi\rangle$的设备，来简单地将它们制备出来就实现了复制它们。并且如果我们能复制量子态的话，我们就可以为每个态制备任意多的拷贝。因为在极限$n\to\infty$下，直积态$|\psi\rangle^{\otimes n}$和$|\phi\rangle^{\otimes n}$都是正交的，这些态无疑是可区别的，克隆态$|\psi\rangle$和$|\phi\rangle$的可能性为区别它们提供了一种方法。

还需要指出的是，根据一个类似的观点，克隆会导致与同一个密度算符相关的不同混合态变得不可区分。量子比特的态$|\uparrow_z\rangle=|0\rangle$，$|\downarrow_z\rangle=|1\rangle$以相同权重混合形成的态（自旋可观测量$Z=\sigma_z$的纠缠态），与量子比特的态$|\uparrow_x\rangle=\frac{1}{\sqrt{2}}(|0\rangle+|1\rangle)$，$|\downarrow_x\rangle=\frac{1}{\sqrt{2}}(|0\rangle-|1\rangle)$以相同权重混合形成的态（自旋可观测量$X=\sigma_x$的纠缠态）具有相同的密度算符，$I/2$。由于克隆态$|\uparrow_z\rangle^{\otimes n}$，$|\downarrow_z\rangle^{\otimes n}$与克隆态$|\uparrow_x\rangle^{\otimes n}$，$|\downarrow_x\rangle^{\otimes n}$是无法区分的，因此克隆会导致两个混合态也无法区分。

此外，超光速的信号传输也变得可能。我们假设爱丽丝和鲍伯共同占有一个纠缠态$\frac{1}{\sqrt{2}}(|0\rangle|1\rangle-|1\rangle|0\rangle)$。如果爱丽丝在她的量子比特上对$Z$或$X$进行测量，就会使得鲍伯的量子比特处于混合态$\frac{1}{2}|\uparrow_x\rangle\langle\uparrow_x|+\frac{1}{2}|\downarrow_x\rangle\langle\downarrow_x|$或混合态$\frac{1}{2}|\uparrow_z\rangle\langle\uparrow_z|+\frac{1}{2}|\downarrow_z\rangle\langle\downarrow_z|$。如果鲍伯能够通过克隆区分出这两个混合态，且假如他区分这两个混合态所用的时间短于光在爱丽丝和鲍伯之间传播所需的时间，那么他就能确定爱丽丝测量的到底是X还是Z，这样的话，一比特的信息就会以超光速的方式从爱丽丝传输给鲍伯。

3.4 可访问的信息

量子态之所以能够被用来执行新类型的信息处理任务，是因为量子态与经典状态相比，具有不同的可区分的属性。当然，这里关心的不仅仅是量子态可区分性的缺失，还有量子态所特有的可区分性。量子信息受限制的可访问性反映出了其不可区分性。

为了精确地把握可访问性的概念，我们在香农意义上考虑一个经典信息源，其香农熵为 $H(X)$。假设该源产生的符号被表示为(字母表 x 中)一个随机变量X 的值 x，其概率为 p_x，然后这些符号被编码为量子态 ρ_x，$x \in X$。式(22)、式(23)和式(24)中定义的交互信息 $H(X:Y)$ 是用来衡量我们所获得的信息的多少的，这种信息是指在对给定的量子态进行测量时，得到的结果 Y 的基中存在关于随机变量 X 的值的平均信息。可访问信息定义为全部可能的测量：

$$\text{Sup } H(X:Y) \tag{108}$$

交互信息的 Holevo 边界给定了可访问信息的上界：

$$H(X:Y) \leqslant S(\rho) - \sum_x p_x S(\rho_x) \tag{109}$$

这里 $\rho = \sum_x p_x \rho_x$，且测量结果 Y 是由正算子取值测量(POVM) $\{E_y\}$ 得到的。由式(86)可得 $S(\rho) - \sum_x p_x S(\rho_x) \leqslant H(X)$，当且仅当态 ρ_x 有正交支集(support)时，等号成立，因此我们可知：

$$H(X:Y) \leqslant H(X) \tag{110}$$

注意 X 能与 Y 区分开来，当且仅当 $H(X:Y) = H(X)$ 时。如果态 ρ_x 为正交纯态，那么理论上就存在可以将态区分开的一次测量，且对此次测量来说 $H(X:Y) = H(X)$。在这种情况下，可访问信息与量子预备态的熵 $H(X)$ 是相同的。但是如果态是非正交的，那么 $H(X:Y) < H(X)$，且即使在一般意义上，也不存在能够确定地识别出 X 的测量。

特别需要注意的是，如果 X 的值被编码为一个量子比特的纯态，那么 $H(X:Y) \leqslant S(\rho)$ 且 $S(\rho) \leqslant 1$。由此可知，通过测量最多能够从 1 个量子比特中提取 1 比特的信息。如果 X 有 k 个等概率值，那么 $H(X) = \log k$。爱丽丝能够通过等权重混合 k 个非正交纯态的方式将这 k 个值编码到一个量子比特中，但是鲍伯最多只能提取出 1 比特关于 X 的值的信息。对于一个与 n 维希尔伯特空间有关的 n 态量子系统，$S(\rho) \leqslant \log n$。因此，即使爱丽丝能够(通过制备非正交态的混合态的方式)将任意大量的信息编码到这个 n 态量子系统中，鲍伯通过测量从该态中获取的信息最多也只能是 $\log n$，这与 n 态经典系统中所能编码和提取的最大信息量是相同的。因此，看起来由 Holevo 边界确定的量子信息的不可访问性会阻碍我们试图利用量子信息执行非经典的信息处理任务。在下一节

中我们会发现，情况并不是这样的。令人惊讶的是，量子信息的不可访问性竟然能被用于完成超越经典信息范围的信息处理任务。

关于 Holevo 边界的精彩推导过程（下面会介绍其本质思想），请参见[Nielsen and Chuang, 2000, Theorem 12.1, p. 531]。其基本观点如下。假设爱丽丝将一个熵为 $H(X)$ 的量子信号源产生的可区别的符号编码为量子态 ρ_x（不需要是正交的）。这就是说，爱丽丝拥有一个量子系统 P，即制备装置，其正交归一指针基 $|x\rangle$ 相当于由源产生的概率为 p_x 的随机变量 X 的值。制备过程将指针态 $|x\rangle$ 和量子系统 Q 的态 ρ_x 产生关联，以使得制备完成后 P 和 Q 的末态为：

$$\rho^{PQ}\sum_x p_x|x\rangle\langle x|\otimes\rho_x \tag{111}$$

爱丽丝将系统 Q 发送给鲍伯，鲍伯通过测量 Q 的态试图确定随机变量 X 的值。P，Q，以及鲍伯的测量仪器 M 的初态为：

$$\rho^{PQM}=\sum_x p_x|x\rangle\langle x|\otimes\rho_x\otimes|0\rangle\langle 0| \tag{112}$$

这里 $|0\rangle\langle 0|$ 是 M 的初始预备态。鲍伯的测量可由希尔伯特空间 $\mathcal{H}^Q\otimes\mathcal{H}^M$ 中的一个量子操作 $\mathcal{E}$ 来描述，该操作将与 $\mathcal{H}^Q$ 上的 POVM $\{E_y\}$ 相关的一个值 y，存储为 M 的指示态 $|y\rangle$，也就是说，由下式，可知 $\mathcal{E}$ 被定义为任意 $\sigma\in\mathcal{H}^Q$ 的态，且初始预备态 $|0\rangle\in\mathcal{H}^M$：

$$\sigma\otimes|0\rangle\langle 0|\xrightarrow{\mathcal{E}}\sum\sqrt{E_y}\,\sigma\sqrt{E_y}\otimes|y\rangle\langle y| \tag{113}$$

回顾式(78)—式(80)中量子交互信息的定义。我们有：

$$S(P:Q)=S(P:Q,\ M)^8 \tag{114}$$

因为 M 最初是与 PQ 无关联的，且：

$$S(P':Q',M')\leqslant S(P:Q,M) \tag{115}$$

因为量子操作永远不能增加交互信息（这里专指应用 $\mathcal{E}$ 后的指针态）[Nielsen and Chuang, 2000, Theorem 11.15, p. 522]。最后：

$$S(P':Q',M') \tag{116}$$

因为缩减系统永远不能增加交互信息[Nielsen and Chuang, 2000, Theorem 11.15, p. 522]，所以：

$$S(P':M') \leqslant S(P:Q) \tag{117}$$

通过一些代数运算，我们得到的上式就是对 Holevo 边界的表述，换言之，式(117)可以简化为式(109)。

为了明白这一点，我们通过式(111)注意到：

$$\rho^{PQ} = \sum_x p_x \mid x\rangle\langle x \mid \otimes \rho_x \tag{118}$$

因此 $S(P) = H(p_x), S(Q) = S(\sum_x p_x\rho_x) = S(\rho)$，且由式(87)：

$$S(P,Q) = H(p_x) + \sum_x p_x S(\rho_x) \tag{119}$$

因为态 $\mid x\rangle\langle x \mid \otimes\rho_x$ 在$\mathcal{H}^P \otimes \mathcal{H}^Q$中的正交子空间上有支集。由此可知：

$$\begin{aligned} S(P:Q) &= S(P) + S(Q) - S(P,Q) \\ &= S(\rho) - \sum_x p_x S(\rho_x) \end{aligned} \tag{120}$$

这是 Holevo 边界的右边界。

其左边界为：

$$\rho^{P'M'} = \mathrm{Tr}_{Q'}(\rho^{P'Q'M'}) \tag{121}$$

$$= \mathrm{Tr}_{Q'}\left(\sum_{xy} p_x \mid x\rangle\langle x \mid \otimes \sqrt{E_y}\rho_x \sqrt{E_y} \otimes \mid y\rangle\langle y \mid\right) \tag{122}$$

$$= \sum_{xy} p_x \mathrm{Tr}(E_y \rho_x E_y) \mid x\rangle\langle x \mid \otimes \mid y\rangle\langle y \mid \tag{123}$$

$$= \sum_{xy} p(x,y) \mid x\rangle\langle x \mid \otimes \mid y\rangle\langle y \mid \tag{124}$$

因为 $p(x, y) = p_x p(y \mid x) = p_x \mathrm{Tr}(\rho_x E_y) = p_x \mathrm{Tr}(\sqrt{E_y}\rho_x \sqrt{E_y})$，所以 $S(P':M') = H(X:Y)$。

Holevo 边界限制了将经典比特表示为量子比特。从另一个方面理解，Holevo 边界描述了将经典比特编码为量子比特所要消耗的资源：一个量子比特就是必要且充分的。那么我们能用经典比特表示量子比特吗？如果能的话，一个量子比特要消耗多少比特呢？这个问题的回答[Barnum *et al.*, 2001]是：一个产生非正交信号态的量子源能被以任意高保真度压缩为每信号 α 个量子比特加每信号大量的经典比特，当且仅当 α 至少应该与源的冯·诺伊曼熵 S 一样大时。这就意味着一般的量子源不能被分为经典部分和量子部分，量子信息不能被替换为大量的经典信息。

3.5 量子信息压缩

在3.2节中我们指出，香农信源编码定理(无噪信道编码定理)以及典型序列的核心概念能够被推广到量子信息源。这最早是由约萨和舒马赫[1994]以及舒马赫[1995]指出的，也可参见[Barnum *et al.*, 1996b]。

对于一个经典信息比特源，这个源的输出由一个随机变量X给定，其两个值x_1、x_2输出的概率分别为p_1、p_2，由该源产生的信息的香农熵为$H(X)=H(p_1, p_2)$。因此，由香农信源编码定理，通过每信号使用$H(X)$个比特(当$p_1 \neq p_2$时，少于每信号1比特)，信息就能够以任意低的出错率被压缩和传播给接收者。

现在假设源以概率p_1、p_2产生量子态$|\psi_1\rangle$、$|\psi_2\rangle$。混合态$\rho=p_1|\psi_1\rangle\langle\psi_1|+p_2|\psi_2\rangle\langle\psi_2|$的香农熵为$S(\rho)$。香农信源编码定理的舒马赫推广表明编码为混合态$\rho$的量子信息通过每信号使用$S(\rho)$比特，其中当量子态是非正交的时，$S(\rho)<1$，就能够以任意低的出错率被压缩和传播给接收者。

注意这里考虑的信号是量子比特态。舒马赫定理表明通过以每信号低于1量子比特的方式发送信号，我们就能够确定地传输由该源产生的量子比特态的序列。我们还要注意到，当量子比特态为非正交时$S(\rho)<H(p_1, p_2)$时，由量子比特态的序列表示的量子信息因而能够被压缩得超过与制备ρ的熵(即处于量子比特态的随机变量的香农熵)相关的经典信息的经典极限。

由于混合态中的个体态一般来说是不可区分的，因此存在两类适用于量子信息但不能应用于经典信息的压缩。在不可见压缩(blind compression)中，由源产生的量子态的序列通过一个压缩表得以压缩，该压缩表只由量子态的特性及其概率决定，也就是说，输入到压缩表的是与分布相关的密度算符。在可见压缩中，假定由源产生的每个个体量子态的特性是已知的，也就是说，输入到压缩表的是由源产生的序列中的个体量子态，且态的压缩是基于每个态的概率分布的。

举个例子如下。上述量子比特源(以概率p_1, p_2产生$|\psi_1\rangle$, $|\psi_2\rangle$)的一个低效率的可见压缩表会将量子态标记的经典信息压缩到每信号$H(p_1, p_2)$比特发送给接收者，接收者将经典信息解压缩后还原为初始的量子比特态。根

据舒马赫定理，（对于非正交量子比特态来说）这个压缩表不是最优的，因为 $S(\rho)<H(p_1, p_2)$。当然，舒马赫定理规定的是每量子信号 $S(\rho)$ 量子比特这样的压缩率，而这里应用的是香农定理，因此规定了每经典信号 $H(p_1, p_2)$ 比特这样的压缩率。但是要注意传播一个经典比特所需的物理资源与传播一个量子比特（两个正交量子态中的一个）是相同的。还要注意（非正交）量子比特态自身的发送，也需要每信号一量子比特的资源，且不会将序列中态的特性表达给传送者。因此，如果我们的目的是传播与量子比特态序列相关联的量子信息，那么标记在序列中的关于个体态的经典信息实际上是多余的（如果我们考虑一个产生 n 量子比特态的源，那么其经典信息会被限制在每信号 $\log n$）。

值得注意的是，舒马赫定理表明与量子纯态序列相关的量子信息的最优压缩率是每信号 $S(\rho)$ 量子比特，对于不可见压缩和可见压缩都一样。

为了明白其大概含义，我们考虑一个以概率 p_1，p_2 产生（可能是非正交的）量子比特 $|\psi_1\rangle$，$|\psi_2\rangle$ 的源。其概率分布的密度算符为 $\rho=p_1|\psi_1\rangle\langle\psi_1|+p_2|\psi_2\rangle\langle\psi_2|$。

该源产生的一个 n 序列态可以表示为空间 $\mathcal{H}_2^{\otimes n}$ 中的态：

$$|\Psi_{i_1\cdots i_n}\rangle=|\psi_{i_1}\rangle\cdots|\psi_{i_n}\rangle \tag{125}$$

这样的态出现的概率为 $p_{i_1\cdots i_n}=p_{i_1}\cdots p_{i_n}$。这些 n 序列存在于整个 2^n 维希尔伯特空间 $\mathcal{H}_2^{\otimes n}$ 上，但是当 $n\to\infty$ 时，在一个"特定子空间"上找到一个 n 序列的概率（在对由该源产生的一个 n 序列进行测量时其投影算符出现在该子空间上的概率）趋向于1。更确切地说，对于任意 ϵ，$\delta>0$，存在一个子空间 $\mathcal{T}_\delta^{(n)}$，其维度介于 $2^{n(S(\rho)-\delta)}$ 与 $2^{n(S(\rho)+\delta)}$ 之间，其投影算符为 $P_\delta^{(n)}$，使得：

$$\sum_{\text{所有序列}} p_{i_1\cdots i_n}\mathrm{Tr}(|\Psi_{i_1\cdots i_n}\rangle\langle\Psi_{i_1\cdots i_n}|P_\delta^{(n)})=\mathrm{Tr}(\rho^{\otimes n}P_\delta^{(n)})\geqslant 1-\epsilon \tag{126}$$

这里 $\rho^{\otimes n}=\rho\otimes\rho\cdots\rho$，为 ρ 自身的 n 重张量积，就是由该源产生的 n 序列态的密度算符：

$$\rho^{\otimes n}=\sum_{\text{所有}n\text{序列}} p_{i_1\cdots i_n}|\Psi_{i_1\cdots i_n}\rangle\langle\Psi_{i_1\cdots i_n}| \tag{127}$$

$$=\sum_{\text{所有}n\text{序列}} p_{i_1}\cdots p_{i_n}|\psi_{i_1}\rangle\langle\psi_{i_1}|\otimes\cdots\otimes|\psi_{i_n}\rangle\langle\psi_{i_n}| \tag{128}$$

这里每个态 $|\psi_{ij}\rangle$ 都是 d 维希尔伯特空间上 k 个可能态中的一个。我们回顾一下，对所有可能的测量来说，n 序列态的统计特性是由 $\rho^{\otimes n}$ 给定的，且不依赖于

$\rho^{\otimes n}$如何具体表示为态的混合。由于对于一个量子比特源，$S(\rho)\leqslant 1$，因此，$\mathcal{H}_2^{\otimes n}$中$\mathcal{T}_\delta^{(n)}$的维度会随$n\to\infty$而指数级降低，也就是说对于大数$n$来说，$\mathcal{H}_2^{\otimes n}$的大小是其上特定子空间的指数倍。

注意这并不意味着几乎所有由该源产生的n序列态都会位于特定的子空间中。而是说，几乎所有由该源产生的n序列是如此，使得在序列上对$P_\delta^{(n)}$的测量会产生结果1，也就是说，在对投影到特定子空间上的投影算符进行测量时，由该源产生的n序列会得到“肯定的”结果。因此，在此意义下，通过测量会发现，由该源产生的大部分序列位于特定子空间中，且对于维度低于$2^{n(S(\rho)-\delta)}$的任意子空间$\mathcal{V}$来说，当n足够大时，在$\mathcal{V}$中找到由该源产生的n序列的平均概率低于之前指定的ϵ。

现在考虑一般情况。信号源产生k个(不必须是正交的)态$|\psi_1\rangle$，…，$|\psi_k\rangle\in\mathcal{H}_d$，其概率分别为$p_1$，…，$p_k$。这里与该源相关的密度算符为$\rho=\sum_{i=1}^{k}p_i|\psi_i\rangle\langle\psi_i|$。长度为$n$的序列存在于$d^n=2^{n\log d}$维的子空间中，且特定子空间$\mathcal{T}_\delta^{(n)}$的维度介于$2^{n(S(\rho)-\delta)}$与$2^{n(S(\rho)+\delta)}$之间，因为$S(\rho)\leqslant\log d$，所以$\mathcal{H}_d^{\otimes n}$有该特定子空间的指数倍大。

对于遵循香农定理的压缩，我们将ρ用谱表示记为：

$$\rho=\sum_x p(x)|x\rangle\langle x| \tag{129}$$

这里$\{p(x)\}$为ρ的非零本征值的集，$\{|x\rangle\}$为ρ的本征态的正交归一集。如果ρ有本征值$p(x)$和本征态$|x\rangle$，那么$\rho^{\otimes n}$就有本征值$p(x_1)p(x_2)\cdots p(x_n)$和本征态$|x_1\rangle|x_2\rangle\cdots|x_n\rangle$。

在下式意义下，类似于序列x_1，x_2，…，x_n是一个δ型序列，一个δ型态被定义为态$|x_1\rangle|x_2\rangle\cdots|x_n\rangle$：

$$2^{-n(S(\rho)+\delta)}<p(x_1\cdots x_n)<2^{-n(S(\rho)-\delta)} \tag{130}$$

δ型子空间$\mathcal{T}_\delta^{(n)}$是所有δ型态所占据的一个子空间。投影到$\mathcal{T}_\delta^{(n)}$上的投影算符可以表示为：

$$P_\delta^{(n)}=\sum_{\delta\text{型的态}}|x_1\rangle\langle x_1|\otimes|x_2\rangle\langle x_2|\cdots|x_n\rangle\langle x_n| \tag{131}$$

那么，选定一个$\delta>0$，对于任意$\epsilon>0$且n足够大，有：

$$\mathrm{Tr}(P_\delta^{(n)}\rho^{\otimes n})\geqslant 1-\epsilon \tag{132}$$

且$\mathcal{T}_\delta^{(n)}$（$=\mathrm{Tr}(P_\delta^{(n)})$）的维度满足：

$$(1-\epsilon)2^{n(S(\rho)-\delta)} \leqslant \dim \mathcal{T}_\delta^{(n)} \leqslant 2^{n(S(\rho)+\delta)} \tag{133}$$

也就是说，$\mathcal{T}_\delta^{(n)}$的维度大概为$2^{nS(\rho)}$，随着$n\to\infty$，$H_2^{\otimes n}$的维度将会是其指数倍。

由此可知，密度算符$\rho^{\otimes n}$可由在特定子空间具有支集的密度算符$\tilde{\rho}^{\otimes n}$代替。在谱分布中取$\rho^{\otimes n}$，这里矩阵被对角化为$2^{n\log d}$个本征值$p(x_1\cdots x_n)=p(x_1)\cdots p(x_n)$，然后用0替换所有与典型序列相符的$p(x_1\cdots x_n)$。

在考虑量子信息的压缩/解压缩表之前，我们需要就精确度方面对这类压缩表的可靠性进行测量，就像在经典信息中我们所做的那样。下面的定义将会推广第2节中关于精确度的经典概念[Jozsa，1998，70]。如果$|\psi\rangle$为任意纯量子态，ρ为任意混合态，那么ρ和$|\psi\rangle$之间的逼真度为：

$$F(\rho|\psi\rangle) = \mathrm{Tr}((\rho|\psi\rangle\langle\psi|)) = \langle\psi|\rho|\psi\rangle \tag{134}$$

这是投影算符$|\psi\rangle\langle\psi|$在态ρ上的一次测量产生结果为1的概率，也即ρ在测量时为$|\psi\rangle$的概率。注意对于纯态$\rho=|\psi\rangle\langle\psi|$，$F(|\phi\rangle,|\psi\rangle)=|\langle\psi|\phi\rangle|^2$。对于$\rho$和$\sigma$的纯化态$|\psi\rangle$和$|\phi\rangle$，两个混合态$\rho$和$\sigma$之间的逼真度①为：

$$F(\rho,\sigma) = \max|\langle\psi|\phi\rangle|^2 = (Tr(\sqrt{\rho^{1/2}\sigma\rho^{1/2}}))^2 \tag{135}$$

由上式可以看出，$F(\rho,\sigma)$相对于ρ和σ是对称的。

就一个以先验概率$p_{i_1\cdots i_n}=p_{i_1}\cdots p_{i_n}$产生出$n$序列量子态$|\Psi_{i_1\cdots i_n}\rangle=|\psi_{i1}\rangle\cdots|\psi_{i_n}\rangle$的源来说，压缩/解压缩表一般将会产生一个混合态$\rho_{i_1\cdots i_n}$。对于一个$n$序列量子态，其压缩/解压缩表的平均逼真度被定义为：

$$F_n = \sum_{\text{所有}n\text{序列}} p_{i_1\cdots i_n}\mathrm{Tr}(\rho_{i_1\cdots i_n}|\Psi_{i_1\cdots i_n}\rangle\langle\Psi_{i_1\cdots i_n}|) \tag{136}$$

针对以概率p_1，…，p_n产生量子态$|\psi_1\rangle\cdots|\psi_n\rangle\in\mathcal{H}_d$的量子源（因此该源的输出相对应的密度算符为$\rho=\sum p_i(|\psi_i\rangle\langle\psi_i|)$，舒马赫量子信源编码定理（或者量子无噪信道编码定理）表明：

---

① 注意尼尔斯和庄将逼真度$F(\rho,\sigma)$定义为这里所定义值的平方根。如果ρ和σ是对易的，那么它们就能够在同样的基下被对角化。那么这个定义就可以简化为关于两个概率分布之间的经典逼真度的定义，这两个概率分布由ρ和σ本征值定义。

> 对于任意ϵ，$\delta>0$：(1)存在一个压缩/解压缩表，使用每量子态$S(\rho)+\delta$量子比特的数据，就能够传输由源产生的长度为n的序列，并使其在接收端能以$F_n>1-\epsilon$的保真率被解压缩，假设n足够大的话；(2)任何使用每量子态$S(\rho)-\delta$量子比特数据的压缩/解压缩表，传输长度为n的序列的话，保真率为$F_n<\epsilon$，假设n足够大的话。

这类量子信息源的一个压缩/解压缩表将会服从以下规则。传送者应用$\mathcal{H}_d^{\otimes n}$(维度为$d^n=2^{n\log d}$)中的一个幺正算符$U$，它将特定子空间中的任意一个态映射到一个由$n\log d$个量子比特组成的序列的线性叠加上，这里除了第一个$nS(\rho)$量子比特之外其他量子比特都处于态$|0\rangle$上，然后将第一个$nS(\rho)$量子比特传送给接收者。因此传送者将$n\log d$个量子比特压缩为$nS(\rho)$个量子比特。接收者添加$n\log d-nS(\rho)$个$|0\rangle$态的量子比特，然后应用幺正变换$U^{-1}$。一般来说，开始的$nS(\rho)$量子比特会与后面的$n\log d-nS(\rho)$量子比特之间存在轻微的纠缠，因此用态$|0\rangle$替换后面的这些量子比特会产生一个混合态$\tilde{\rho}_n$。态$U^{-1}\tilde{\rho}_n$会以大于$1-\epsilon$的逼真度被识别为初始态$|\Psi_{i_1\cdots i_n}\rangle$。

4. 借助于纠缠实现的量子传输

在这一节中我将说明如何将纠缠作为一个信道来实现量子信息的可靠传输。我将讨论两类借助于纠缠实现的传输：4.1中的量子隐形传态(teleportation)和4.2中的量子稠密编码(dense coding)。

4.1 量子隐形传态

在第一节中我们提到，薛定谔在一篇分为两部分的针对爱因斯坦—波多尔斯基—罗森争论[Einstein *et al.*, 1935]的评论[Schrödinger, 1935; 1936]中引入术语“纠缠”来描述EPR态的特殊非定域关系。薛定谔认为纠缠态是有问题的，因为这会导致他所谓的“远程控制”成为可能，他只把“远程控制”看作是希尔伯特空间的数学产物，而不是作为一种实际可能性存在。结果是，量子隐形传态正是两个分离系统之间存在远程控制的实验证实。这种现象最早由本内特、布莱森德、科瑞普(Crépeau)、约萨、佩雷斯(Peres)和伍特斯在论文

[1993]中提出，之后由几个团队分别以多种不同方式在实验上得到确认[Bouwmeester *et al.*（博瓦米斯特等），1997；Boschi（博斯基等），1998；Furasawa *et al.*（弗拉斯瓦等），1998；Nielsen *et al.*，1998]。

在 1935 年的论文中，薛定谔考虑具有唯一双正交分解的纠缠纯态，它与 EPR 态类似，但其双正交分解不是唯一的。他表明对一个系统进行适当的测量，就能确定与其纠缠的遥远系统的量子(纯)态，且这个态依赖于我们选择哪个可观测量进行测量，而不仅仅与测量结果相关。在第二篇论文中，他指出一个“熟练的实验员”通过其中一个系统执行一次适当的定域性测量，就能够“操控”远距离外的系统，使它处于由它的约化密度算符表示的纯态构成的任意混合态上。因此这个远距离的系统就能够(概率性地，取决于定域性测量的结果)被操控到其约化密度算符支集的任意纯态上，某个纯态出现的概率是非零的，且只依赖于该纯态自身。对于远距离系统的线性无关态的混合来说，通过在一个合适的基上执行一次定域性的标准投影值测量，也能实现操控。如果态是线性相关的，那么实验员可以执行一次广义测量(与 POVM 有关)，这等同于通过添加附属空间的方式扩大实验者的希尔伯特空间，以使得扩大了的希尔伯特空间的维度与线性无关态所处的数量相同。正如 3.1 中指出的那样，薛定谔的分析预示了后来哈格斯顿、约萨和伍特斯[1993]的结果。

假设爱丽丝和鲍伯，两方通讯协议中的传统主人公，各自拥有处于纠缠态

$$|\Psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle_A|1\rangle_B - |1\rangle_A|0\rangle_B) \tag{137}$$

的两个量子比特中的一个。鲍伯的量子比特分离后处于混合态$\rho_B = I/2$，它可以被理解为正交态$|0\rangle_B$和$|1\rangle_B$的等权混合，或者等价地视为其他无穷多的混合，举一个其中的特例，就是四个非正交的归一态的等权混合：

$$|\phi_1\rangle_B = \alpha|0\rangle_B + \beta|1\rangle_B$$
$$|\phi_2\rangle_B = \alpha|0\rangle_B - \beta|1\rangle_B$$
$$|\phi_3\rangle_B = \beta|0\rangle_B + \alpha|1\rangle_B$$
$$|\phi_4\rangle_B = \beta|0\rangle_B - \alpha|1\rangle_B$$

即：

$$\rho_B = I/2 = \frac{1}{4}(|\phi_1\rangle\langle\phi_1| + |\phi_2\rangle\langle\phi_2| + |\phi_3\rangle\langle\phi_3| + |\phi_4\rangle\langle\phi_4|) \quad (138)$$

如果爱丽丝在她的量子比特 A 上对具有本征态 $|0\rangle_A$ 和 $|1\rangle_A$ 的可观测量进行测量，且鲍伯在他的量子比特 B 上对相应的可观测量进行测量，那么爱丽丝的结果就会与鲍伯的结果形成反关联(0 对 1，1 对 0)。或者，如果爱丽丝准备了一个附属量子比特 A'，其处于态 $|\phi_1\rangle_{A'} = \alpha|0\rangle_{A'} + \beta|1\rangle_{A'}$，并在她拥有的量子比特对 $A' + A$ 上测量具有以下本征态的可观测量：

$$|1\rangle = (|0\rangle_{A'}|1\rangle_A - |1\rangle_{A'}|0\rangle_A)/\sqrt{2} \quad (139)$$

$$|2\rangle = (|0\rangle_{A'}|1\rangle_A + |1\rangle_{A'}|0\rangle_A)/\sqrt{2} \quad (140)$$

$$|3\rangle = (|0\rangle_{A'}|0\rangle_A - |1\rangle_{A'}|1\rangle_A)/\sqrt{2} \quad (141)$$

$$|4\rangle = (|0\rangle_{A'}|0\rangle_A + |1\rangle_{A'}|1\rangle_A)/\sqrt{2} \quad (142)$$

贝尔态定义了 $\mathcal{H}^{A'} \otimes \mathcal{H}^A$ 中的这些贝尔基。她会以相同的概率 1/4 得到四个结果 1，2，3，4 中的一个，且这四个结果分别与鲍伯的四个态 $|\phi_1\rangle_B$，$|\phi_2\rangle_B$，$|\phi_3\rangle_B$，$|\phi_4\rangle_B$ 相关。也就是说当爱丽丝报告她得到了结果 i 时，如果鲍伯查看他的粒子是否处于态 $|\phi_i\rangle_B$ 上的话，他就会发现事实上结果总是会遵从上面的规则。这是因为：

$$|\phi_1\rangle_{A'}|\Psi\rangle = \frac{1}{2}(-|1\rangle|\phi_1\rangle_B -$$

$$|2\rangle|\phi_2\rangle_B + |3\rangle|\phi_3\rangle_B - |4\rangle|\phi_4\rangle_B) \quad (143)$$

在此意义下，爱丽丝能够通过进行适当的定域性测量，来操控鲍伯的粒子，使其处于同他的密度算符 $\rho_B = I/2$ 相协调的任意混合态上。

薛定谔认为纠缠是有问题的正是由于上述意义下远距操控的可能性：

> 相当令人不安的是，该理论将会允许系统任凭实验者的摆布而被操控或引导至这个或那个典型态，尽管他并没有权利这样做。

现在，正是这种概率意义上的远距操控使得量子隐形传态成为可能。假设爱丽丝和鲍伯共同占有处于纠缠态(137)的一对量子比特，而爱丽丝得到了一个处于未知态 $|\phi_1\rangle$ 的量子比特 A'，她想将其发送给鲍伯。不存在一个程序能

使得爱丽丝确定未知态的性质，并且即使她能做到，她将要发送给鲍伯的，以便使他能制备态$|\phi_1\rangle$的经典信息数量也很可能是无穷的。因为对一般归一化的量子比特态$\alpha|0\rangle+\beta|1\rangle$进行精确说明需要两个实参数(独立参数的数量从四个减为两个是因为$|\alpha|^2+|\beta|^2=1$，且总相位是不相干的)。爱丽丝能够将量子比特本身发送给鲍伯，但是在可能存在干扰的传输过程中，量子比特态中的量子信息很可能会塌缩。

或者，爱丽丝仅仅需要两比特的经典信息，就能以完美的确定性成功地将未知量子态$|\phi_1\rangle$传送给鲍伯。爱丽丝所要做的事是，在贝尔基上测量她所拥有的2量子比特系统$A'+A$。根据她测量所得到的结果，即以相同概率出现的$i=1$，2，3，4，鲍伯的量子比特会被操控而呈现为态$|\phi_1\rangle_B$，$|\phi_2\rangle_B$，$|\phi_3\rangle_B$，$|\phi_4\rangle_B$中的一个。如果爱丽丝将她的测量结果传送给鲍伯(需要两比特经典信息的传输)，鲍伯就能在他的希尔伯特空间中应用四个局域幺正变换中的一个，以得到态$|\phi_1\rangle_B$：

$i=1$：什么都没做，即应用恒等变换I；

$i=2$：应用变换σ_z；

$i=3$：应用变换σ_x；

$i=4$：应用变换$i\sigma_y$。

这里σ_x，σ_y，σ_z是泡利自旋矩阵。

导致能够实现将量子态$|\phi_1\rangle$从爱丽丝传输给鲍伯，且A'并没有真的由爱丽丝处移动到鲍伯处这种技术出现的，是共享纠缠态赋予爱丽丝的一种能力，这种能力将四个测量结果(每个出现的概率为1/4)与鲍伯处的四个态一一对应起来，这四个态是鲍伯拥有的混合态的一种特殊分解。鲍伯的操作完成后，A'的态的传输就结束了，而鲍伯的操作依赖于爱丽丝发送给他的与其测量结果有关的两比特经典信息。根据传输协议，粒子A'的态在爱丽丝的测量中被破坏，通过鲍伯的操作，它又被重建为鲍伯所拥有的粒子的态——事实上，系统A和A'最终作为爱丽丝测量的结果而处于一个纠缠态。注意如果A'的态$|\phi_1\rangle$没有被破坏，那么该态就会存在两个备份，而这是违反量子“不可克隆”定理的。因此无论是爱丽丝还是鲍伯，或者其他的当事人，都不能得到关于被传送态特征的任何信息，这是因为如果在其他量子系统的态中记录这种信息，那么实际上

就是对这个被传送态中信息的部分复制。

共享纠缠为量子通讯提供了一个安全可靠的信道。这对于粒子之间按加密协议进行的量子信息的传送来说可能是有用的，对于量子计算机中处理元件之间量子信息的传递也是有意义的。由两个粒子构成的纠缠态的一个性质就是纠缠是不受二者之间环境中的噪声影响的。因此，通过隐形传态完成的量子信息通讯的可靠性取决于它所需要的经典通讯的可靠性，而经典通讯则可以通过众所周知的纠错码技术来消除噪声的影响。由两个粒子构成的纠缠态也不受它们相对空间位置变化的影响。因此爱丽丝甚至不需要知道鲍伯的位置就可以隐形传输一个量子态给他，只要广播两比特的信息即可。

4.2 量子稠密编码

由 Holevo 边界(参见 3.4)可知，通过将一个量子比特的量子态中的信息进行编码的方式来对其进行可靠的传输，其中所需的经典信息的最大量为 1 比特，即使任意大量的经典信息能够在一个量子比特的态中被编码(通过将符号编码为非正交量子态的方式)。量子稠密编码是一个程序，最早由本内特和威斯纳[1992]发现，是为了利用纠缠来使一个量子比特所能传播的经典信息量翻倍。

我们还是考虑贝尔态：

$$|1\rangle = (|0\rangle|1\rangle - |1\rangle|0\rangle)/\sqrt{2} \tag{144}$$

$$|2\rangle = (|0\rangle|1\rangle + |1\rangle|0\rangle)/\sqrt{2} \tag{145}$$

$$|3\rangle = (|0\rangle|0\rangle - |1\rangle|1\rangle)/\sqrt{2} \tag{146}$$

$$|4\rangle = (|0\rangle|0\rangle + |1\rangle|1\rangle)/\sqrt{2} \tag{147}$$

假设爱丽丝和鲍伯共同占有一对量子比特，其处于态：

$$|1\rangle = (|0\rangle_A|1\rangle_B - |1\rangle_A|0\rangle_B)/\sqrt{2} \tag{148}$$

存在四个由 $\mathcal{H}^A$ 中的幺正变换定义的操作：

$$U_1 = I \tag{149}$$

$$U_2 = \sigma_x \tag{150}$$

$$U_3 = \sigma_x \tag{151}$$

$$U_4 = i\sigma_y \tag{152}$$

通过对爱丽丝所拥有的量子比特执行四个局域操作其中的一个，她就能将量子比特对的态$|1\rangle$变换为任意贝尔态。例如：

$$I|1\rangle = |1\rangle \tag{153}$$

$$\sigma_z|1\rangle = |2\rangle \tag{154}$$

$$\sigma_x|1\rangle = |3\rangle \tag{155}$$

$$i\sigma_y|1\rangle = |4\rangle \tag{156}$$

因此为了向鲍伯传输两个经典比特，爱丽丝对她的量子比特应用上述四个操作中的一个，然后把该量子比特发送给鲍伯。接着鲍伯对处于贝尔基的两个量子比特执行一次测量。因为它们是正交态，所以他能分辨出是哪个态，并确定爱丽丝的操作。

5. 量子密码学

在过去的几年中，量子密码学或许是量子信息理论应用得最成功的领域。其主要成果就是关于密钥分配的多种安全协议的出现，它们最初起源于本内特和布莱森德[1984]的研究，以及迈耶斯(Mayers)[1996b；1997]、卢和周(Lo and Chau)[1998]提出的一个重要的“不可行”定理：无条件安全的双方量子比特承诺是不可能的。量子比特承诺定理推广了原先由迈耶斯[1996a]与卢和周[1997]提出的仅限于单向通讯的协议，并且它适用于量子、经典及量子与经典混合的系统(因为经典信息，正如我们所看到的，可以被视为受到一定约束的量子信息)。对双方系统的限制排除了涉及可信第三方以及可信的信道属性的系统，且限制系统必须是完全基于量子力学原理的，就排除了利用狭义相对论信号约束的系统，以及可能涉及时间机器或黑洞热力学的系统，等等。

在5.1节中，我将说明量子密钥分配的安全性是怎样依赖于量子信息的性质——不能克隆、无干扰情况下得不到信息、纠缠——来阻止窃听者秘密地获取双方之间量子通信的信息的，也就是说完全不被察觉的窃听对量子通信来说在原理上是不可能的。在5.2节中，我主要讨论量子比特承诺，并说明为什么无条件安全的量子比特承诺是不可能的。

5.1 密钥分配

量子密钥分配协议

在量子密钥分配协议中，爱丽丝和鲍伯双方最初并不共享信息，他们的目的是通过量子及经典信道来交换信息，以便最终能共享可用于加密的密钥，以这样的方式来确保窃听者夏娃任何企图获得密钥信息的努力能够以非零的概率被检测到。

一次性密码本为爱丽丝和鲍伯利用经典信息通讯提供了一个完全安全的方式，但这也是双方能实现完全安全经典通讯的唯一方式。一次性密码本，本质上是一个随机比特序列。如果爱丽丝和鲍伯都拥有一次性密码本的拷贝，那么爱丽丝可以安全地将信息传递给鲍伯，通过将信息转换为 n 位二进制数(根据爱丽丝和鲍伯双方都知道的一些方案)，并将这个二进制数构成的比特序列与一次性密码本中从首位开始的长度为 n 的比特序列进行(逐位模2)相加的方式。爱丽丝把加密后的序列发送给鲍伯，然后鲍伯可以使用来自于他自己的一次性密码本的拷贝中的相同比特序列进行解密。由于加密信息是随机的，如果没有一次性密码本的拷贝的话，夏娃不可能解密出信息。该方案安全性的本质在于，一旦信息被发送并解密，用于对信息进行加密的 n 个随机比特就会被废弃，每条不同的信息都用到一个唯一的随机序列，正因为如此，该术语为“一次性密码本”。

这个过程保证了绝对的保密性，假设只需并且只有爱丽丝和鲍伯拥有一个任意长的一次性密码本的拷贝。但是，这意味着，为了双方能进行秘密的通讯，他们必须共享一个秘密：随机密钥。密钥分配问题是关于最初如何安全地分配密匙，而不令其在传输过程中被秘密截获和复制的问题；密钥存储问题是关于如何安全地存储密钥，而不令其被偷偷复制的问题。我们希望有一个针对被动窃听来说能保证安全的程序，以便爱丽丝和鲍伯可以确信，他们之间的通讯确实是保密的。

量子密码学的核心思想一是利用非正交量子态的不可辨别性，如同在3.3节中我们看到的那样，它能够确保夏娃在获得关于这种态的性质的任何信息后，都会引起一些干扰，使得爱丽丝和鲍伯可以检测到；二是利用“不可克隆”

定理，它使得夏娃无法复制和存储爱丽丝和鲍伯之间的量子通讯，以便其在后来做出分析（做出这种分析，也许会使用到截获的爱丽丝和鲍伯之间的经典通讯）。

在本内特和布莱森德[1984]最初提出现在被称为BB84的协议后，一系列种类繁多的量子密钥分配方案已经出现。其核心理念是爱丽丝发送给鲍伯一个量子比特序列，这些量子比特由态$|0\rangle$，$|1\rangle$，$|+\rangle$，$|-\rangle$等概率地制备，其中$|0\rangle$，$|1\rangle$为一对正交态，$|+\rangle$，$|-\rangle$为一对正交态，它们之间是相互非正交的。鲍伯测量这些随机地可能在基$|0\rangle$，$|1\rangle$也可能在基$|+\rangle$，$|-\rangle$上的每一个量子比特。根据他测量的结果，他公开宣称对于每个量子比特序列，他使用的是哪个基，而爱丽丝则公开宣称她是用这些基中的哪一个来制备量子比特的。然后爱丽丝和鲍伯删除掉那些与他们的基不一致的量子比特。由于鲍伯通过测量得到的态与爱丽丝准备的态是相同的，因此爱丽丝和鲍伯通过剩下的量子比特共享了一个随机密钥。然后，他们可以牺牲这些量子比特中的一部分来检测窃听行为。爱丽丝公开宣布她制备的量子比特的态，而鲍伯检查他的测量结果来做出确认。如果他们在足够数量的态（取决于预期的错误率）的基础上得到一致的结果，那么他们就能得出没有窃听行为的结论，然后使用剩余的部分作为密钥。如果他们的结果不一致，那么他们就能得出量子比特被窃听干扰的结论，在这种情况下，他们放弃所有的量子比特，并重新开始整个程序。实际的协议与这一理念还存在进一步的细微差别，这里完全安全的密钥是从“天然密钥”中提取的，而实际中的密钥是通过纠错和保密增强的技术手段得到的。

BB84方案解决了密钥分配的问题，在这个意义上，爱丽丝和鲍伯最初并不共享秘密，而最终他们可以通过一个密钥分配协议来共享一个密钥，这个协议能够以任意高的可靠性来排除被窃听的可能（因为被牺牲掉作为检查窃听行为的量子比特序列的长度可以为任意长）。但显然，它并没有解决密钥存储问题，因为密钥分配协议的结果被存储为经典信息，而这是会遭到被动窃听的。

埃克特（Ekert）[1991]提出的一个方案允许爱丽丝和鲍伯通过对两个纠缠量子比特进行测量来创建一个共享的随机密钥。假设爱丽丝和鲍伯共享由两个

量子比特构成的一个纠缠纯态，比如贝尔态$\frac{1}{\sqrt{2}}(|0\rangle|1\rangle-|1\rangle|0\rangle)$的多个拷贝(可能由产生爱丽丝和鲍伯之间纠缠对的一个共同的源所发出)。爱丽丝和鲍伯通过对他们的量子比特分别进行测量，发现他们有三个可观测量是一致的，这里的测量是随机选取的，并且对于每个量子比特都是独立的。在对适当数量的纠缠对进行一系列测量之后，爱丽丝和鲍伯公布他们的测量结果的方向，并将测量结果分为两组：一组是他们测得的自旋在不同方向上的，另一组是自旋在相同的方向上的。他们公开给出第一组测量的结果，并以此来确定单态没有受到窃听的干扰。从本质上讲，他们计算一个相关系数：窃听者夏娃任何监测粒子的企图，都会干扰纠缠态，并产生一个贝尔不等式范围内的相关系数，因此，可以与纠缠态的相关系数区别开来。如果爱丽丝和鲍伯发现没有窃听行为发生，那么他们就会用第二组相反的相关测量结果作为密钥。

通过前选择和后选择产生的量子密钥分配

埃克特的方案解决密钥分配问题的同时也解决了密钥存储问题，因为针对每条信息会从存储的纠缠态中产生一个新的密钥，并且我们不知道处于纠缠态的密匙的任何信息。在这里，我也描述一个会涉及量子纠缠态的密钥分配协议[Bub，2001b]，但是与测试窃听行为的类型不同。不同于贝尔定理基础上的统计测试，这里的测试会利用关于测量结果的条件陈述，这些测量结果由前选择和后选择量子态所产生。

前选择和后选择量子态的特殊性质首先是由阿哈罗诺夫(Aharonov)、伯格曼(Bergmann)、勒波维茨(Lebowitz)[1964]指出的。如果：

(1)爱丽丝在时刻t_1制备了一个处于特定态$|\mathrm{pre}\rangle$的系统；

(2)鲍伯在时刻t_2测量该系统的某个可观测量M；

(3)爱丽丝在时刻t_3测量一个$|\mathrm{post}\rangle$为其本征值的可观测量，并用$|\mathrm{post}\rangle$表示后选择。

那么爱丽丝就可以为鲍伯在时刻t_2测量M的结果分配概率，这个概率分别与t_1时刻的态$|\mathrm{pre}\rangle$和t_3时刻的态$|\mathrm{post}\rangle$有如下关系[Aharonov *et al.*，1964；Vaidman *et al.*（魏德曼等），1987]：

$$\mathrm{prob}(q_k) = \frac{|\langle \mathrm{pre}|P_k|\mathrm{post}\rangle|^2}{\sum_i |\langle \mathrm{pre}|P_i|\mathrm{post}\rangle|^2} \tag{157}$$

其中，P_i为M在第i个本征空间上的投影算符。注意，式(157)——下面一般称其为“ABL规则”(阿哈罗诺夫—伯格曼—勒波维茨规则)——在态$|\mathrm{pre}\rangle$和$|\mathrm{post}\rangle$可以互换的意义上，是时间对称的。

如果M对爱丽丝来说是未知的，那么她可以用ABL规则为各种假设的M测量的结果进行概率分配。与通常的前选择态的波恩规则相比，ABL规则的一个有趣的特点是——对于适当选出的可观测量M，M'，…以及态$|\mathrm{pre}\rangle$和$|\mathrm{post}\rangle$——为一系列相互非对易的可观测量的结果赋予单位概率是可能的。也就是说，爱丽丝可以断言以下形式的条件陈述：“如果鲍伯测量M，那么其结果必然为m_i，如果鲍伯测量M'，那么其结果必然为m'_j，…，”其中M，M'，…为相互非对易的可观测量。由于鲍伯最多只能测量这些非对易可观测量中的一个，因此爱丽丝的条件信息当然不会与量子力学发生矛盾：她只知道可观测量M的本征值为m_i，如果她知道鲍伯事实上测量的是M的话。

魏德曼，阿哈罗诺夫，阿尔贝[1987]讨论过这种类型的一个案例，其中对自旋为$\frac{1}{2}$的粒子的三个自旋量$X=\sigma_x$，$Y=\sigma_y$，$Z=\sigma_z$中任一个的测量结果都可以由一个适当的前选择和后选择推导出来。爱丽丝制备的贝尔态为：

$$|\mathrm{pre}\rangle = \frac{1}{\sqrt{2}}(|\uparrow_z\rangle_A|\uparrow_z\rangle_C + |\downarrow_z\rangle_A|\downarrow_z\rangle_C \tag{158}$$

这里$|\uparrow_z\rangle$和$|\downarrow_z\rangle$表示σ_z的本征态。爱丽丝发送粒子中的一个——信道粒子，由下标C表示——给鲍伯并保留了其附属粒子，记作A。鲍伯在信道粒子上测量X，Y或Z中的一个，并将信道粒子传回给爱丽丝。然后爱丽丝测量粒子对上的一个可观测量R，其中R具有本征态(下标A和C被略去)：

$$|r_1\rangle = \frac{1}{\sqrt{2}}|\uparrow_z\rangle|\uparrow_z\rangle + \frac{1}{2}(|\uparrow_z\rangle|\downarrow_z\rangle e^{i\pi/4} + |\downarrow_z\rangle|\uparrow_z\rangle e^{-i\pi/4}) \tag{159}$$

$$|r_2\rangle = \frac{1}{\sqrt{2}}|\uparrow_z\rangle|\uparrow_z\rangle - \frac{1}{2}(|\uparrow_z\rangle|\downarrow_z\rangle e^{i\pi/4} + |\downarrow_z\rangle|\uparrow_z\rangle e^{-i\pi/4}) \tag{160}$$

$$|r_3\rangle = \frac{1}{\sqrt{2}}|\downarrow_z\rangle|\downarrow_z\rangle + \frac{1}{2}(|\uparrow_z\rangle|\downarrow_z\rangle e^{-i\pi/4} + |\downarrow_z\rangle|\uparrow_z\rangle e^{i\pi/4}) \quad (161)$$

$$|r_4\rangle = \frac{1}{\sqrt{2}}|\downarrow_z\rangle|\downarrow_z\rangle - \frac{1}{2}(|\uparrow_z\rangle|\downarrow_z\rangle e^{-i\pi/4} + |\downarrow_z\rangle|\uparrow_z\rangle e^{i\pi/4}) \quad (162)$$

注意：

$$|\mathrm{pre}\rangle = \frac{1}{\sqrt{2}}(|\uparrow_z\rangle|\uparrow_z\rangle + |\downarrow_z\rangle|\downarrow_z\rangle) \quad (163)$$

$$= \frac{1}{\sqrt{2}}(|\uparrow_x\rangle|\uparrow_x\rangle + |\downarrow_x\rangle|\downarrow_x\rangle) \quad (164)$$

$$= \frac{1}{\sqrt{2}}(|\uparrow_y\rangle|\downarrow_y\rangle + |\downarrow_y\rangle|\uparrow_y\rangle) \quad (165)$$

$$= \frac{1}{2}(|r_1\rangle + |r_2\rangle + |r_3\rangle + |r_4\rangle) \quad (166)$$

现在，爱丽丝可以通过 ABL 规则分配值给鲍伯的自旋测量结果，无论是鲍伯基于后选择 $|r_1\rangle$，$|r_2\rangle$，$|r_3\rangle$，$|r_4\rangle$ 测量的是 X，Y 还是 Z，如下表 1（其中，0 表示结果↑和 1 表示结果↓）[Vaidman *et al.*（维德曼等），1987]：

	σ_x	σ_y	σ_z
r_1	0	0	0
r_2	1	1	0
r_3	0	1	1
r_4	1	0	1

表 1：与 R 的本征值相关的 σ_x，σ_y，σ_z 的测量结果

我们可以利用这个案例来使爱丽丝和鲍伯通过下面的方式共享一个保密的随机密匙。爱丽丝准备式(158)中贝尔态 $|\mathrm{pre}\rangle$ 的一定数量的拷贝(数量取决于密钥的长度和所要求的保密级别)。她将信道粒子按顺序发送给鲍伯，并保留附属粒子。鲍伯随机在信道粒子上测量 X 或 Z，然后再按顺序将粒子返回给爱丽丝。然后爱丽丝在附属粒子和信道粒子对上测量可观测量 R，并将这个序列划分为两个子序列：子序列 S_{14}，表示她得到的结果为 r_1 或 r_4；子序列 S_{23}，表示她得到的结果 r_2 或 r_3。这一系列操作可以在量子线路上实现，参见[Metzger

(梅茨格)，2000]。

为了确定信道粒子没有被夏娃监测，爱丽丝现在公开宣布(广播)子序列S_{23}的指标。从表1中显然能看到，对于这个子序列，她可以做出如下形式的条件陈述："对于信道粒子i，如果X被测量，那么其结果为1(0)；如果Z被测量，那么其结果为0(1)。"这取决于她对R测量的结果是r_2还是r_3。并且她也公开宣布这些结果。对于某个指标i，如果这些结果中的一个与鲍伯的结果不符，那么夏娃一定监测了第i个信道粒子(当然，相符也并不保证粒子没有被监测)。

因为假设夏娃在一个信道粒子上测量了一个与鲍伯不同的可观测量的自旋分量，随后爱丽丝测量R得到了本征值r_2或r_3中的一个。鲍伯的测量结果，不是0就是1，将会仅与这两个本征值中的一个相容，假设夏娃没有干扰的话。但在夏娃的测量后，这两个本征值都将可能是爱丽丝测量的结果。因此，爱丽丝针对鲍伯关于子序列S_{23}的测量结果的回溯性诠释并不一定对应于鲍伯的记录。事实上，我们可以表明，如果夏娃对信道粒子随机测量X或Z，或者如果她对信道粒子测量可观测量X，Y或Z中的特定一个(对每个粒子都测量相同的可观测量)，那么在子序列S_{23}上能侦测到这种行为的概率是3/8。

在子序列S_{14}，鲍伯的测量结果0和1对应于爱丽丝对R测量的结果r_1和r_4。如果根据他们关于子序列S_{23}的公共通信，爱丽丝和鲍伯一致认为夏娃没有对信道粒子进行监测，那么他们使用子序列S_{14}来定义一个共享的原始密钥。

注意即使爱丽丝的回溯性诠释和鲍伯的记录之间仅存在一处不一致的地方也足以显示信道粒子被夏娃监测了。这不同于埃克特协议中的窃听测试。还要注意的是夏娃只能截获信道粒子，而不是粒子对。因此，夏娃不可能做到的是，用她自己的粒子替换掉所有信道粒子，并通过一些幺正变换使原始信道粒子与其附属粒子产生纠缠而成为一个系统，然后等爱丽丝和鲍伯完成公开通讯后再进行测量。夏娃也没有办法确保爱丽丝和鲍伯的结果是否一致，如果她不能访问粒子对，或者没有关于鲍伯测量的信息的话。

如上文所述的密钥分配协议解决了密钥分配问题但没解决密匙存储问题。如果正如协议所要求的那样，鲍伯随机测量X或Z，并在爱丽丝测量R之前记录下关于自旋测量的确定结果，那么鲍伯的测量记录——是作为经典信息来存

储的——原则上就可以被夏娃复制且不被发现。在这种情况下，根据爱丽丝和鲍伯之间的公共通信，来验证量子通信信道的完整性，夏娃就能够知道原始密钥(它包含在这些信息中)。

要解决密钥存储问题，该协议应该按下面的方式进行修改。不同于原先真实地在每个信道粒子中进行随机选择并测量自旋可观测量中的一个且同时记录测量结果，鲍伯应该将随机选择和自旋观测保持在“量子层面上”，直到爱丽丝公布她对 R 测量得到的子序列 S_{23} 的指标。为了做到这一点，鲍伯需要通过幺正变换，使信道粒子的量子态与他引进的两个附属粒子的态相互纠缠，来扩大希尔伯特空间。其中一个粒子与所选择的可观测量或“量子骰子(quantum die)”可观测量 D 的两个本征态 $|d_X\rangle$, $|d_Z\rangle$ 所存在的希尔伯特空间有关，另一个粒子与指针可观测量 P 的两个本征态 $|p_\uparrow\rangle$，$|p_\downarrow\rangle$ 所存在的希尔伯特空间有关，可参见在 5.2 节中关于如何在扩大了的希尔伯特空间上实现幺正变换的讨论。

根据修改后的协议(假设存储纠缠态的能力没有限制)，爱丽丝和鲍伯共享很大数量的关于 4 个粒子的纠缠态的拷贝。当他们想建立一个特定长度的随机密钥时，爱丽丝在她拥有的适当数量的粒子对上测量 R，并公布子序列 S_{23} 的指标。在爱丽丝公布子序列 S_{23} 的指标之前，无论是爱丽丝还是鲍伯都不存储任何经典信息。因此，没有任何信息可以让夏娃复制。在爱丽丝公布子序列 S_{23} 的指标后，鲍伯根据这些指标在他的附属粒子上测量可观测量 D 和 P，并公布本征值 $|p_\uparrow\rangle$ 或 $|p_\downarrow\rangle$ 作为他对 X 或 Z 测量的结果，具体选择哪个取决于 D 的本征值。如果爱丽丝和鲍伯确定夏娃没有窃听，那么鲍伯在子序列 S_{14} 中对他的附属粒子测量 C 和 P。我们很容易看到，ABL 规则适用于这种情况，就像它适用于鲍伯原先在艾丽斯测量 R 之前，就做出随机选择，并记录他对 X 或 Z 测量的确定结果的情况一样。事实上，如果这两种情况对于爱丽丝来说是不等价的话——如果爱丽丝能够从她对 R 的测量中辨别出鲍伯究竟是做出了随机选择并执行了自旋测量，还是仅仅使这种行为“保持在量子层面”的话——那么这种区别就可以被用来实现超光速的信号传输。

5.2 比特承诺

历史回顾

根据比特承诺协议，其中一方，爱丽丝提供一个加密比特给另一方鲍伯。从加密比特中获得的信息不足以使鲍伯确定该比特的值，但是假设爱丽丝会公布该比特的值，那么连同她在后续阶段提供的进一步信息，鲍伯就足以确信，该协议不会允许爱丽丝通过随意公布0或1的方式来进行欺诈。

为了阐明这个观点，我们假设爱丽丝声称其具有预测选举结果的能力。为了保证她的声明不会透露有价值的信息(给可能是潜在的雇主，鲍伯)，她建议采取下面的证明方式。她提议在选举前一个月，通过在一个便笺上写下0(代表“输”)或1(代表“赢”)的方式，将她关于某个特定候选人能否赢得选举的预言记录下来。然后，她将这个便笺锁在一个保险箱内，并将保险箱交给鲍伯保管，但自己留下钥匙。选举结束后，她将公布她所写的数字，并将钥匙交给鲍伯，鲍伯可以打开保险箱阅读便笺，从而证明她确实提前就做出了承诺。

显然，此过程的安全性依赖于保险箱的坚固程度或锁被设计的精巧程度。更一般地，爱丽丝可以给鲍伯发送(加密的)信息，来给出一个经典不可兼析取的真值(相当于她做出0或1的承诺)，且仅当信息表明两个析取项中一个为真时(因为一个经典不可兼析取为真，当且仅当其中的一个析取项为真而另一个为假)。经典力学的原理不能阻止鲍伯提取这些信息，所以经典比特承诺协议的安全性只与计算的复杂性相关。

现在的问题是，是否存在一个等价的量子过程是无条件安全的，即由物理定律(量子理论)保证的可证实的安全性防止爱丽丝或鲍伯作弊。需要注意的是鲍伯可以作弊，如果他在爱丽丝揭示她的承诺之前，能获取一些相关信息的话，这将使他在复制爱丽丝的协议时具有优势。爱丽丝也可以作弊，如果直到最后阶段，即当被要求揭示其承诺时，她事实上才做出承诺的话；或者她能在最后阶段改变她的承诺，并且只以非常低的概率会被发现的话。

本内特和布莱森德[1984]最初提出了一种量子比特承诺协议。其基本思路是将0和1承诺与两个由相同密度算符表示的不同混合态联系起来。如他们在文中所指出的，爱丽丝可以通过“EPR攻击”或欺骗策略来作弊：她准备一些量

子比特的纠缠对，保留下每个粒子对中的一个(附属粒子)，而将第二个量子比特(信道粒子)发送给鲍伯。以这种方式，她可以假装发送两个等价的混合态中的一个给鲍伯，并且在开始阶段就可以任意显示两个中的任一个比特，通过将鲍伯的粒子有效地转变为所需混合态的方式，这可以利用适当测量来实现。鲍伯无法检测到这种欺骗策略。

在后来的文章中，布莱森德、科瑞普、约萨和朗格卢瓦(Langlois)[1993]提出了一种量子比特承诺协议，他们声称这是绝对安全的。BCJL 方案最初由迈耶斯[1996a]证明是不安全的。随后，迈耶斯[1996b；1997]及卢和周[1997；1998]分别独立表明，本内特和布莱森德在[1984]中的见解可以扩展为一个证据，以论证 EPR 作弊策略的广义版本总是适用的，如果引入额外的附属粒子，并以适当的方式来扩大希尔伯特空间的话。

绝对安全的量子比特承诺是不可能的，这对由量子密码学家组成的团体来说多少有些令人惊讶，并会造成意义深远的后果。事实上，毫不夸张地说，就我们对量子力学的理解来说，量子比特承诺定理的意义与贝尔定理[Bell，1964]相当。布莱森德和福斯[Brassard，2000；Fuchs，1997；Fuchs，2000；Fuchs and Jacobs，2002]猜测，量子力学可以由两个与量子信息有关的假设推导出来：安全密钥分配的可能性与安全比特承诺的不可能性。在第 7 节中我们将介绍这对于量子力学的基础意味着什么。

也许是由于证明过于简单，而断言的普适性却很强，因此后来的很多文献都不断挑战量子比特承诺定理，它们认为对其进行的证明并没有涵盖所有可能被用来实现量子比特承诺的程序，参见[Yuen(元朗)，2005]。看起来一个普遍感觉是该定理“太好了以至于不是真的”，认为该定理必然会存在一个漏洞。

事实上，并不存在漏洞。虽然肯特(Kent)[1999a；1999b]已经证明如何在处于确定分离位置的爱丽丝和鲍伯之间，利用同步于通信的相对论信号约束来实现安全的经典比特承诺协议，而且哈迪和肯特(Hardy and Kent)[2004]以及阿哈罗诺夫、塔士玛(Tashma)、瓦齐拉尼(Vazirani)和姚(Yao)[2005]已经研究了“欺骗敏感性”的安全性或量子比特承诺的“弱”版本，但是这些结果并不与量子比特承诺定理相冲突。在通常理解的比特承诺协议中，在介于协议的承诺阶段结束之后和爱丽丝揭示比特值之前的这个阶段，存在一个任意长的时间

间隔，在此期间没有信息交换。肯特别出新裁的方案，事实上涉及了介于承诺阶段和揭示阶段之间的第三阶段，在此阶段鲍伯在某一位置与爱丽丝分别在所在位置处定期交换信息，直到爱丽丝在某一位置选择公布原来承诺的比特。在公布的那一刻，协议还尚未完成，因为在鲍伯于某个对应位置获得证实承诺所需的所有信息之前，还需要进一步公布一系列爱丽丝的位置和相对应的鲍伯的位置。如果一个比特承诺协议被理解为在承诺阶段和公布阶段之间需要任意长的空闲时间(在这段时间中协议没有被执行任何一步)的话，那么量子比特承诺定理就包含了需要利用狭义相对论信号约束的那些协议。①

一个重要见解

理解量子比特承诺定理的决定性见解在于，认识到在量子比特承诺协议中，任何需要爱丽丝或鲍伯做出明确选择的步骤(是否执行很多种测量中的一个，或是否实施很多个幺正变换中的一个)，总是能够由一个广义上的 EPR 欺骗策略所替换，如果我们假设爱丽丝和鲍伯都配备了量子计算机的话。也就是说，对确定可能性进行的经典析取——这样或那样的操作——(也许对于作弊者来说在更方便的时候)总是能够由量子纠缠及其后续测量来替换，其中某个可能性会变得明确。从本质上讲，经典析取能够被量子析取所替代。这种欺骗策略不会被检测到。同样，可以“在量子水平上”实施一个测量而不被检测到：不是实施测量并获得很多可能结果中的那个确定结果，而是在扩大了的希尔伯特空间上实施一个合适的幺正变换，在该空间中系统是与一个“指针”附属粒子以适当的方式相纠缠的，且获得确定结果的过程会被延迟。关键的一点是通过扩大希尔伯特空间，能够保持爱丽丝和鲍伯之间的一系列交流处于量子层面上，直到最后爱丽丝公布她的承诺时，才变为经典信息交换。

任何量子比特承诺方案都将涉及爱丽丝和鲍伯之间的一系列交流，在交流过程中，特定数量 n 的量子系统——“信道粒子”——会在二者之间传递，并被施加各种可能是随机选择的量子操作(幺正变换或测量等)。这些操作总是可以被替换为信道与一个或多个附属粒子的纠缠，而不会被发现。这些附属粒子对测量来说充当着“指针”粒子，或对随机选择来说作为“骰子”粒子。实际上，

① 非常感谢多米尼克·迈耶斯帮我澄清了这一要点。

这就是(广义的)EPR 欺骗策略。

举例说明如下假设在量子比特承诺协议的某个特定阶段，鲍伯需要对他从爱丽丝处接收到的每个信道粒子随机选择测量两个可观测量 X, Y 中的一个。为简单起见，假设 X 和 Y 各自具有两个本征值，x_1，x_2和 y_1，y_2。记录下测量结果后，鲍伯需要将信道粒子返回给爱丽丝。当爱丽丝接收到第 i 个信道颗粒时，她按次序向鲍伯发送下一个信道粒子。我们可以假设鲍伯记录下的测量结果构成了信息的一部分，这些信息可以使他确认爱丽丝的承诺，一旦她将这些信息(连同进一步的信息)进行公示，那么他就不需要将他的测量结果报告给爱丽丝，直到在协议的最后阶段，她公布她的承诺。

鲍伯可以不遵守协议，而是构建一个设备将信道粒子的输入态 $|\psi\rangle_C$ 与他引入的两个附属粒子的初始态 $|d_0\rangle_B$ 和 $|p_0\rangle_B$ 纠缠起来，第一个是作为对应于随机选择的“量子骰子”，而第二个是作为对应于测量的“量子指针”。假设鲍伯构造这种设备(事实上就是一个具有特殊目的的量子计算机)的能力仅受到量子力学规律的限制。

纠缠由幺正变换通过以下的方式实现:① 定义两个幺正变换，U_X 和 U_Y，它们分别在“在量子层面上”执行对信道粒子所在希尔伯特空间$\mathcal{H}_C$和鲍伯的指针附属粒子所在的希尔伯特空间$\mathcal{H}_{B_p}$的张量积的 X 测量和 Y 测量:

$$
\begin{gathered}
|x_1\rangle_C|p_0\rangle_B \xrightarrow{U_X} |x_1\rangle_C|p_1\rangle_B \\
|x_2\rangle_C|p_0\rangle_B \xrightarrow{U_X} |x_2\rangle_C|p_2\rangle_B
\end{gathered} \tag{167}
$$

以及:

$$
\begin{gathered}
|y_1\rangle_C|p_0\rangle_B \xrightarrow{U_Y} |y_1\rangle_C|p_1\rangle_B \\
|y_2\rangle_C||p_0\rangle_B \xrightarrow{U_Y} |y_2\rangle_C|p_2\rangle_B
\end{gathered} \tag{168}
$$

得到:

---

① 注意，我们假设信道粒子为纯态，这并不失一般性。如果信道粒子与爱丽丝的附属粒子相纠缠，那么该设备就通过变换 $I\otimes$……实现纠缠，在这里 I 为爱丽丝的附属粒子所存在的希尔伯特空间中的单位算符。

$$|\psi\rangle_C|p_0\rangle_B \xrightarrow{U_X} \langle x_1|\psi\rangle|x_1\rangle_C|p_1\rangle_B + \langle x_2|\psi\rangle|x_2\rangle_C|p_2\rangle_B \quad (169)$$

以及：

$$|\psi\rangle_C|p_0\rangle_B \xrightarrow{U_Y} \langle y_1|\psi\rangle|y_1\rangle_C|p_1\rangle_B + \langle y_2|\psi\rangle|y_2\rangle_C|p_2\rangle_B \quad (170)$$

类似地，随机选择也可以由对鲍伯的骰子附属粒子所在的希尔伯特空间$\mathcal{H}_{B_D}$,和希尔伯特空间$\mathcal{H}_C\otimes\mathcal{H}_{B_P}$的张量积执行的幺正变换$V$来定义。假设$|d_X\rangle$和$|d_Y\rangle$为$\mathcal{H}_{B_D}$上的两个正交态，且$|d_0\rangle=\frac{1}{\sqrt{2}}|d_X\rangle+\frac{1}{\sqrt{2}}|d_Y\rangle$，那么(去除明显的下标)$V$被定义为：

$$|d_X\rangle\otimes|\psi\rangle|p_0\rangle \xrightarrow{V} |d_X\rangle\otimes U_X|\psi\rangle|p_0\rangle \quad (171)$$

$$|d_Y\rangle\otimes|\psi\rangle|p_0\rangle \xrightarrow{V} |d_Y\rangle\otimes U_Y|\psi\rangle|p_0\rangle$$

可得：

$$|d_0\rangle\otimes|\psi\rangle|p_0\rangle \xrightarrow{V} \frac{1}{\sqrt{2}}|d_X\rangle\otimes U_X|\psi\rangle|p_0\rangle+\frac{1}{\sqrt{2}}|d_Y\rangle\otimes U_Y|\psi\rangle|p_0\rangle \quad (172)$$

这里张量积的符号被选择性地引入，用来表明U_x和U_y在$\mathcal{H}_C\otimes\mathcal{H}_{B_P}$上是确定的。

如果鲍伯实际上是在随机选择可观测量X或Y，并进行测量得到了一个特定的本征值，那么对于信道粒子来说爱丽丝的密度算符将是：

$$\frac{1}{2}(|\langle x_1|\psi\rangle|^2|x_1\rangle\langle x_1|+|\langle x_2|\psi\rangle|^2|x_2\rangle\langle x_2|)+$$

$$\frac{1}{2}(|\langle y_1|\psi\rangle|^2|y_1\rangle\langle y_1|+|\langle y_2|\psi\rangle|^2|y_2\rangle\langle y_2|) \quad (173)$$

假设爱丽丝不知道鲍伯选择什么样的可观测量进行测量，也不知道他得到了什么结果。但是，对于式(172)产生的态来说，这恰恰是通过追溯鲍伯的附属粒子所产生的那个相同的密度算符。换句话说，对爱丽丝来说，信道粒子的密度算符是相同的，无论鲍伯随机选择哪个可观测量来进行测量且实际执行了测量，也无论他是否用他那两个能在扩大的希尔伯特空间上产生式(172)变换的附属粒子实现了EPR欺骗策略。

如果鲍伯被要求在最后报告他进行了什么样的测量，以及他得到了什么样

的结果，那么在那个阶段他可以为本征态$|d_X\rangle$和$|d_Y\rangle$测量骰子附属粒子，为本征态$|p_1\rangle$或$|p_2\rangle$测量指针附属粒子。实际上，如果我们考虑两种测量的多个可能结果构成的系综，那么鲍伯将需要把通过追溯其附属粒子所产生的“不适当的”混合态修改为“适当的”混合态。但是适当的混合态和不适当的混合态之间的区别是爱丽丝检测不到的，因为她并不了解鲍伯的附属粒子，并且只有通过测量由信道粒子和鲍伯的附属粒子组成的复合系统，爱丽丝才能确定该信道粒子是与附属粒子相纠缠的。

事实上，如果区分适当和不适当的混合态是可能的话，那么超光速的信号传送就是可能的：通过监测她所拥有的信道粒子，爱丽丝就能在瞬间知道鲍伯是否对他的附属粒子进行了测量。注意，这会使鲍伯或爱丽丝谁先测量变得没有区别，因为测量是对不同希尔伯特空间中的可观测量进行的，因此是对易的。

显然，同样的论证也适用于在比特承诺协议的某个阶段鲍伯需要在一些幺正操作之间做出选择的情况。或许不那么明显的情况是，如果鲍伯需要视之前对第i个粒子做出的测量或视之前选择的操作而决定对第$i+1$个信道粒子执行一次测量或选择某个操作的时候，EPR 欺骗策略也是可能的。当然，如果鲍伯同时拥有所有信道粒子的话，他可以在整个序列上使其与附属粒子发生纠缠，从而把它们视为单个复合系统。但是，即使鲍伯一次只能访问一个信道粒子（他需要在执行测量或其他操作之后将其返回给爱丽丝，然后她再发送下一个信道粒子给他），他也总是能使第$i+1$个信道粒子与他曾用来和第i个粒子形成纠缠的附属粒子产生纠缠。

例如，假设鲍伯收到了序列中的两个信道粒子。他应该随机决定测量第一个粒子上的X或Y，在执行测量后将粒子返回给爱丽丝。爱丽丝收到第一个粒子后，她再发送给鲍伯第二个粒子。如果鲍伯在第一个粒子上测量X，并得到结果x_1，那么他应该在第二个粒子上测量X；如果他得到结果x_2，那么他应该在第二个粒子上测量Y。如果他在第一个粒子上测量Y并得到结果y_1，他应该对第二个粒子执行幺正变换U_1；如果他得到结果y_2，他应该执行幺正变换U_2。在执行完所需的操作后，他应该将第二个粒子返回给爱丽丝。

乍看起来，似乎鲍伯在对第二个粒子执行协议之前，必须确实对第一个粒

子进行测量，并得到一个特定的结果，如果考虑到他一次只能访问一个信道粒子的话，这样的话 EPR 欺骗策略就不能实现。但事实并非如此。鲍伯的策略是：他对第一个信道粒子应用上面讨论过的针对两个可选择测量的 EPR 策略。对于第二个信道粒子，他在他的附属粒子和信道粒子所在的希尔伯特空间构成的张量积上实施下面的幺正变换，这里第二个信道粒子的态被表示为 $|\phi\rangle$，而相应于第二个信道粒子的指针附属粒子的态表示为 $|q_0\rangle$（这里并不需要第二个骰子粒子）：

$$
\begin{aligned}
&|d_X\rangle|p_1\rangle|\phi\rangle|q_0\rangle \xrightarrow{U_C} |d_X\rangle|p_1\rangle\otimes U_X|\phi\rangle|q_0\rangle\\
&|d_X\rangle|p_2\rangle|\phi\rangle|q_0\rangle \xrightarrow{U_C} |d_X\rangle|p_2\rangle\otimes U_Y|\phi\rangle|q_0\rangle\\
&|d_Y\rangle|p_1\rangle|\phi\rangle|q_0\rangle \xrightarrow{U_C} |d_Y\rangle|p_1\rangle\otimes U_1|\phi\rangle|q_0\rangle\\
&|d_Y\rangle|p_2\rangle|\phi\rangle|q_0\rangle \xrightarrow{U_C} |d_Y\rangle|p_2\rangle\otimes U_2|\phi\rangle|q_0\rangle
\end{aligned}
\qquad (174)
$$

量子比特承诺定理的证明

由于 EPR 欺骗策略总是可以在不被察觉的情况下应用，那么量子比特承诺定理的证明过程就要假设在承诺阶段的最后，由爱丽丝的附属粒子、n 信道粒子以及鲍伯的附属粒子组成的复合系统，在希尔伯特空间 $\mathcal{H}_A\otimes\mathcal{H}_B$ 中将由一些复合纠缠态 $|0\rangle$ 或 $|1\rangle$ 表示，具体如何表示取决于爱丽丝的承诺[①]，这里 $\mathcal{H}_A$ 是在这个阶段爱丽丝所拥有的粒子（爱丽丝的附属粒子和爱丽丝保留的信道粒子，如果有的话）所在的希尔伯特空间，$\mathcal{H}_B$ 是鲍伯在这个阶段所拥有的粒子（鲍伯的附属粒子和鲍伯保留的信道粒子，如果有的话）所在的希尔伯特空间。

现在，描述鲍伯所能选择的两个承诺的信息的密度算符 $W_B(0)$ 和 $W_B(1)$，是通过描绘 $\mathcal{H}_A$ 中的态 $|0\rangle$ 或 $|1\rangle$ 得到的。如果这两个密度算符相同，那么在没有来自爱丽丝的进一步信息的情况下，鲍伯将无法区分 $|0\rangle$ 态和 $|1\rangle$ 态。在这种情况下，该协议被称为是“隐藏的”。应用双正交分解定理建立起来的证明过程是，如果 $W_B(0)=W_B(1)$，则存在 $\mathcal{H}_A$ 中的一个幺正变换，它能将 $|0\rangle$ 转

① 更准确地说，取决于爱丽丝打算公布 0 还是 1——因为我们假设当需要时，爱丽丝会采取 EPR 欺骗策略。

变为$|1\rangle$。也就是说，如果该协议是“隐藏的”，那么它对爱丽丝将不会是“有约束力的”：她总是能够遵循协议(应用EPR策略进行适当的替换)建立态$|0\rangle$。在最后阶段，当她被要求公布她的承诺时，她可以视具体情况而选择公布替代的承诺，通过在她自己的希尔伯特空间上应用适当的幺正变换将$|0\rangle$转变为$|1\rangle$，但鲍伯却不能检测到这一举动。因此，或者鲍伯可以通过在爱丽丝揭示她的承诺之前得到一些有关她的选择的信息进行欺骗，或者爱丽丝可以作弊。

证明的要点可以简述如下。

在施密特分解中，态$|0\rangle$和$|1\rangle$可表示为：

$$|0\rangle = \sum_i \sqrt{p_i}\,|a_i\rangle|b_i\rangle \tag{175}$$

$$|1\rangle = \sum_j \sqrt{p'_j}\,|a'_j\rangle|b'_j\rangle$$

这里$\{|a_i\rangle\}$，$\{|a'_j\rangle\}$为$\mathcal{H}_A$上态的两个正交归一集，$\{|b_i\rangle\}$，$\{|b'_j\rangle\}$为$\mathcal{H}_B$上的两个正交归一集。

密度算符$W_B(0)$和$W_B(1)$定义为：

$$W_B(0) = \mathrm{Tr}_A|0\rangle\langle 0| = \sum_i p_i|b_i\rangle\langle b_i| \tag{176}$$

$$W_B(1) = \mathrm{Tr}_A|1\rangle\langle 1| = \sum_j p'_j|b'_j\rangle\langle b'_j|$$

当且仅当$W_B(0)=W_B(1)$时，鲍伯不能欺骗。现在由谱定理，可知分解式

$$W_B(0) = \sum_i p_i|b_i\rangle\langle b_i|$$

$$W_B(1) = \sum_j p'_j|b'_j\rangle\langle b'_j|$$

在非退化的情况下是唯一的，这里p_i都是不同的，p'_j也都是不同的。条件式$W_B(0)=W_B(1)$意味着对于所有的k：

$$p_i = p'_i \tag{177}$$

$$|b_i\rangle = \langle b'_i|$$

因此：

$$|0\rangle = \sum_i \sqrt{p_i}\,|a_i\rangle|b_i\rangle \tag{178}$$

$$|1\rangle = \sum_i \sqrt{p_i}\,|a'_i\rangle|b_i\rangle$$

由此可知，存在一个幺正变换 $U\in\mathcal{H}_A$，使得：

$$\{|a_k\rangle\}\xrightarrow{U}\{|a'_k\rangle\} \tag{179}$$

因此：

$$|0\rangle\xrightarrow{U}|1\rangle \tag{180}$$

正如我们在 5.2 节中指出的，爱丽丝不是通过幺正变换将 $|0\rangle$ 转变为 $|1\rangle$，而是可以制备态 $|0\rangle$ 并对两个基中的一个进行测量，来实现同样的效果，具体测量哪一个取决于她打算公布 0 还是 1。

退化的情况也能以相似的方式处理。假设 $p_1=p_2=p'_1=p'_2=p$，那么 $|b_1\rangle$，$|b_2\rangle$ 和 $|b'_1\rangle$，$|b'_2\rangle$ 存在于 $\mathcal{H}_B$ 上的同一个子空间 $\mathcal{H}$ 上。因此（当 $k>2$ 时假设系数是不同的）：

$$|0\rangle=\sqrt{p}(|a_1\rangle|b_1\rangle+|a_2\rangle|b_2\rangle)+\sum_{k>2}\sqrt{p_k}|a_k\rangle|b_k\rangle \tag{181}$$

$$\begin{aligned}|1\rangle&=\sqrt{p}(|a'_1\rangle|b'_1\rangle+|a'_2\rangle|b'_2\rangle)+\sum_{k>2}\sqrt{p_k}|a'_k\rangle|b_k\rangle\\&=\sqrt{p}(|a''_1\rangle|b_1\rangle+|a''_2\rangle|b_2\rangle)+\sum_{k>2}\sqrt{p_k}|a'_k\rangle|b_k\rangle\end{aligned}$$

这里 $|a''_1\rangle,|a''_2\rangle$ 为生成 $\mathcal{H}$ 的正交归一态。由于 $\{|a''_1\rangle,|a''_2\rangle,|a_3\rangle,\cdots\}$ 为 $\mathcal{H}_A$ 上的正交归一集，因此存在一个幺正变换将 $\{|a_k\rangle;k=1,2,3,\cdots\}$ 转变为 $\{|a''_1\rangle,|a''_2\rangle,|a'_3\rangle,\cdots\}$，因此将 $|0\rangle$ 变为 $|1\rangle$。

将定理扩展到非理想情况，这里 $W_B(0)\approx W_B(1)$，这将使得鲍伯只能够以很小的概率区分出承诺已经被替换，因此爱丽丝有一个相对大的概率能成功地完成欺骗：存在一个 $\mathcal{H}_A$ 上的幺正变换 U，它能将 $W_B(0)$ 变得足够接近 $W_B(1)$，爱丽丝能够随意公布她真假承诺中的任一个，而鲍伯仅能以相当小的概率检测到这一行为。

该定理是如何起作用的：一个例子

阿舍·佩雷斯（Asher Peres）提出的例子（私人通信）是说明该定理如何运作的一个完美例证（我对这个例子的分析在很大程度上是得益于阿德里安·肯特和多米尼克·迈耶斯的帮助）。

假设爱丽丝发送给鲍伯一个处于等权混合量子比特态的信道粒子 C，如果她的承诺为 0 的话，该等权混合量子比特态为：

$$|c_0\rangle = |0\rangle \tag{182}$$

$$|c_2\rangle = -\frac{1}{2}|0\rangle + \frac{\sqrt{3}}{2}|1\rangle \tag{183}$$

$$|c_4\rangle = -\frac{1}{2}|0\rangle - \frac{\sqrt{3}}{2}|1\rangle \tag{184}$$

而如果她的承诺为 1 的话，量子比特的等权混合态为：

$$|c_1\rangle = |1\rangle \tag{185}$$

$$|c_3\rangle = \frac{\sqrt{3}}{2}|0\rangle - \frac{1}{2}|1\rangle \tag{186}$$

$$|c_5\rangle = -\frac{\sqrt{3}}{2}|0\rangle - \frac{1}{2}|1\rangle \tag{187}$$

注意这两个混合态有相同的密度算符：

$$\rho_0 = \rho_1 = I/2 \tag{188}$$

假设爱丽丝试图通过准备系统 AC 的纠缠态的方式来使 EPR 欺骗策略生效：

$$|0\rangle = \frac{1}{\sqrt{3}}(|a_0\rangle|c_0\rangle + |a_2\rangle|c_2\rangle + |a_4\rangle|c_4\rangle) \tag{189}$$

这里 $\{|a_0\rangle, |a_2\rangle, |a_4\rangle\}$ 为适当的附属系统 A 所在的三维希尔伯特空间 $\mathcal{H}_A$ 中的一个正交归一基。如果爱丽丝能将态 $|0\rangle$ 转变为态：

$$|1\rangle = \frac{1}{\sqrt{3}}(|a_1\rangle|c_1\rangle + |a_3\rangle|c_3\rangle + |a_5\rangle|c_5\rangle) \tag{190}$$

这里 $\{|a_1\rangle, |a_3\rangle, |a_5\rangle\}$ 为 $\mathcal{H}^A$ 的另一个正交归一基，那么通过 $\mathcal{H}^A$ 上的一个局域幺正变换，她能够将作出承诺的行为推迟到揭晓阶段。在这个阶段，如果她想要承诺为 0，那么她可以测量本征态为 $\{|a_0\rangle, |a_2\rangle, |a_4\rangle\}$ 的可观测量。如果她想要承诺为 1，她可以实施局域幺正变换将 $|0\rangle$ 变为 $|1\rangle$，并测量本征态为 $\{|a_1\rangle, |a_3\rangle, |a_5\rangle\}$ 的可观测量。

现在，$|0\rangle$ 可以被表示为：

$$|0\rangle = \frac{1}{\sqrt{3}}\left(|a_0\rangle\frac{|c_3\rangle - |c_5\rangle}{\sqrt{3}} + |a_2\rangle\frac{|c_1\rangle - |c_3\rangle}{\sqrt{3}} + |a_4\rangle\frac{|c_5\rangle - |c_1\rangle}{\sqrt{3}}\right) \tag{191}$$

$$= \frac{1}{\sqrt{3}}\left(\frac{|a_2\rangle - |a_4\rangle}{\sqrt{3}}|c_1\rangle + \frac{|a_0\rangle - |a_2\rangle}{\sqrt{3}}|c_3\rangle + \frac{|a_4\rangle - |a_0\rangle}{\sqrt{3}}|c_5\rangle\right) \tag{192}$$

在$|0\rangle$的这种表示中，$\mathcal{H}^A$上的系数态$\frac{|a_2\rangle-|a_4\rangle}{\sqrt{3}}$，$\frac{|a_0\rangle-|a_2\rangle}{\sqrt{3}}$，$\frac{|a_4\rangle-|a_0\rangle}{\sqrt{3}}$不是正交的，事实上它们是共面的：

$$|a_0\rangle-|a_2\rangle=-(|a_2\rangle-|a_4\rangle)-(|a_4\rangle-|a_0\rangle) \tag{193}$$

因此看起来不存在适当的幺正变换能将$|0\rangle$变为$|1\rangle$，因此 EPR 策略是行不通的！

当然，情况并不是这样。我们会看到存在这样的一个幺正变换，注意$|0\rangle$和$|1\rangle$能够展开为施密特分解式：

$$|0\rangle=\frac{1}{\sqrt{2}}\left(\frac{2|a_0\rangle-|a_2\rangle-|a_4\rangle}{\sqrt{6}}|c_0\rangle+\frac{|a_2\rangle-|a_4\rangle}{\sqrt{2}}|c_1\rangle\right) \tag{194}$$

$$|1\rangle=\frac{1}{\sqrt{2}}\left(\frac{|a_3\rangle-|a_5\rangle}{\sqrt{2}}|c_0\rangle+\frac{-2|a_1\rangle+|a_3\rangle+|a_5\rangle}{\sqrt{6}}|c_1\rangle\right) \tag{195}$$

显然，现在存在$\mathcal{H}^A$上的一个幺正变换 U，使得：

$$|0\rangle\xrightarrow{U}|1\rangle \tag{196}$$

由此可得：

$$\{|a_0\rangle,|a_2\rangle,|a_4\rangle\}\xrightarrow{U}\{|a_0'\rangle,|a_2'\rangle,|a_4'\rangle\} \tag{197}$$

这里$\{|a_0'\rangle,|a_2'\rangle,|a_4'\rangle\}$为$\mathcal{H}^A$上的一个基，因此：

$$|1\rangle=\frac{1}{\sqrt{3}}(|a_0'\rangle|c_0\rangle+|a_2'\rangle|c_2\rangle+|a_4'\rangle|c_4\rangle) \tag{198}$$

$$=\frac{1}{\sqrt{3}}(|a_1\rangle|c_1\rangle+|a_3\rangle|c_3\rangle+|a_5\rangle|c_5\rangle) \tag{199}$$

因此爱丽丝能通过准备状态$|1\rangle$，并对于承诺为 0 来说在基$\{|a_0'\rangle,|a_2'\rangle,|a_4'\rangle\}$上进行测量，或者对于承诺为 1 来说在基$\{|a_1\rangle,|a_3\rangle,|a_5\rangle\}$上进行测量，来执行 EPR 欺骗策略。同样地，她当然可以准备态$|0\rangle$并在两个不同的基上测量，因为将$|0\rangle$变为$|1\rangle$的幺正变换也能够将基$\{|a_1\rangle,|a_3\rangle,|a_5\rangle\}$转换为基$\{|a_1''\rangle,|a_3''\rangle,|a_5''\rangle\}$，因此：

$$|0\rangle=\frac{1}{\sqrt{3}}(|a_0\rangle|c_0\rangle+|a_2\rangle|c_2\rangle+|a_4\rangle|c_4\rangle) \tag{200}$$

$$=\frac{1}{\sqrt{3}}(|a_1''\rangle|c_1\rangle+|a_3''\rangle|c_3\rangle+|a_5''\rangle|c_5\rangle) \tag{201}$$

通过计算可得：

$$| a_1'' \rangle = \frac{1}{3}(| a_0 \rangle + (1 + \sqrt{3}) | a_2 \rangle + (1 - \sqrt{3}) | a_4 \rangle) \tag{202}$$

$$| a_3'' \rangle = \frac{1}{3}((1 + \sqrt{3}) | a_0 \rangle + (1 - \sqrt{3}) | a_2 \rangle + | a_4 \rangle) \tag{203}$$

$$| a_5'' \rangle = \frac{1}{3}((1 - \sqrt{3}) | a_0 \rangle + | a_2 \rangle + (1 + \sqrt{3}) | a_4 \rangle) \tag{204}$$

实际上，如果爱丽丝准备的纠缠态为 $| 0\rangle$ 并在基 $\{| a_0\rangle, | a_2\rangle, | a_4\rangle\}$ 上测量附属粒子 A，那么她会将信道粒子引导入非正交混合态 $\{| c_0\rangle, | c_2\rangle, | c_4\rangle\}$ 中。如果她在基 $\{| a_1''\rangle, | a_3''\rangle, | a_5''\rangle\}$ 上测量，那么她会将信道粒子引导入非正交混合态 $\{| c_1\rangle, | c_3\rangle, | c_5\rangle\}$ 中。

因此得到结论，爱丽丝不需要执行任何幺正变换就能够执行 EPR 欺骗策略——她可以简单地将信道粒子与一个适当的附属粒子相纠缠，并在开始阶段实施两个测量中的一个，这取决于她的承诺是什么。这表明，该定理所要求的幺正变换其实并不是必需的。如果欺骗策略是可行的话，即如果在欺骗策略中，爱丽丝在开始阶段既可以通过对纠缠态进行测量得到结果为 0 的承诺，也可以通过在她的希尔伯特空间中利用一个局部幺正变换将纠缠态转换为另外的态，然后再对转变后的态进行测量得到结果为 1 的承诺，这样的话一个同样好的欺骗策略就是可行的，即爱丽丝为两个承诺都各自准备一个纠缠态，并在开始阶段在两个基中的某一个上进行测量，具体选择哪一个取决于她的承诺。

最后的担忧

上述数学证明的核心是施密特分解定理。但其本质认识是扩大希尔伯特空间且在不被察觉的情况下实施 EPR 策略是可行的。

这就提出了以下问题：假设因为 $W_B(0) = W_B(1)$ 而使得鲍伯不能作弊，因而就可由该定理知道存在 $\mathcal{H}_A$ 上的幺正变换 U 能将 $| 0\rangle$ 变为 $| 1\rangle$，那么是否存在这样一个协议，在其中爱丽丝也不能作弊，这是因为尽管存在一个适当的么正变换 U，但她不知道应该应用什么样的幺正变换？确实存在这样的情况，但只有当 U 依赖于鲍伯的操作，而爱丽丝并不知道才是可能的。但随后鲍伯将必须做出明确的选择，或通过测量得到一个确定的结果，而他总是可以通过应用 EPR 策略，使得他在不被察觉的情况下避免这样做。

这就引出了一个更深层的问题：我们怎么知道，实施EPR策略从来不会对欺骗者不利呢？如果是这样的话，鲍伯可能在某种情况下会选择避开EPR策略，因为这对他将会是不利的。是否存在这样的一个比特承诺协议，在该协议的某个阶段鲍伯应用EPR策略会给爱丽丝而非鲍伯带来优势，同时遵守协议将确保任何一方都不可以欺骗？如果有这样的协议，那么鲍伯，实际上，将被迫遵守协议并避免EPR策略，从而无条件安全的比特承诺将会成为可能。

事实上，按照该定理这样的协议是不可能的，参见[Bub，2001a]，假设存在这样的协议，也就是说，假设鲍伯应用EPR策略，使得 $W_B(0) = W_B(1)$，因而由该定理可知，在爱丽丝的希尔伯特空间中存在一个幺正变换 U 能将 $|0\rangle$ 变为 $|1\rangle$。爱丽丝一定知道这个 U，因为只有鲍伯实施EPR策略来违背协议才能唯一确定 U，然而鲍伯所采用的EPR策略会使得所有在量子层面上的析取都保持线性叠加。同时也可以假设，如果鲍伯忠实地遵循协议(以至于对可能操作或可能测量结果的每一个析取都做出一个确定的选择)，那么 $W_B(0) = W_B(1)$，但爱丽丝的希尔伯特空间中的幺正变换，会令她随鲍伯的选择或测量结果而将 $|0\rangle$ 变为 $|1\rangle$，尽管对于爱丽丝来说她并不知道鲍伯的选择。

需要注意的一点是，爱丽丝的希尔伯特空间中可用的信息是不变的，无论在爱丽丝应用幺正变换将 $|0\rangle$ 变为 $|1\rangle$ 之前，鲍伯是否遵守协议并做出确定的选择或得到确定的测量结果，或者他是否通过EPR策略违背了协议，在该EPR策略中，鲍伯在爱丽丝应用变换 U 之前对他的附属粒子实现相应的纠缠，并将选择和测量结果保持在量子层面上，然后在协议的最后阶段，通过测量他的附属粒子来确定这些选择和测量结果。这对于爱丽丝来说是没有区别的，因为鲍伯对其附属粒子的测量和爱丽丝可能实施的任何测量或操作，发生在不同的希尔伯特空间，所以它们之间是对易的。如果爱丽丝的密度算符(通过追溯鲍伯的附属粒子来得到)——用来描述爱丽丝在她所在的那部分宇宙中所能实施的测量的统计性质——会由于鲍伯究竟是执行了所要求的测量还是通过与附属粒子实施相应的纠缠以保持可选择性存在于量子层面上而有所不同的话，那么我们就可能把这种不同用于超光速的信号传输上。鲍伯对其附属粒子的实际测量——将纠缠态中的可选择性变为确定性——会瞬间改变爱丽丝所在的那部分宇宙中的可用信息。

因此我们可以断定，在我们假设的比特承诺协议中，爱丽丝的希尔伯特空间中将$|0\rangle$变为$|1\rangle$的幺正变换U在真实的情况下和在欺骗情况下必须是相同的。但是，我们假设在真实的情况下变换对爱丽丝来说是未知的，它取决于鲍伯的测量结果，而变换在欺骗的情况下是唯一的，且对爱丽丝来说是已知的。因此并不存在这样的协议，即通过EPR策略来违背协议不会比遵守协议更不利于鲍伯。

就该定理而言，该证明可以如下方式进行公式化表述：欺骗策略会产生$\mathcal{H}_A\otimes\mathcal{H}_B$上两个可供选择的纯态$|0\rangle_c$和$|1\rangle_c$中的一个（'$c$'代表欺骗策略）。由于$\mathcal{H}_B$上的约化密度算符

$$W_B^{(c)}(0) = \mathrm{Tr}_A|0\rangle\langle 0|_C \tag{205}$$

$$W_B^{(c)}(1) = \mathrm{Tr}_A|1\rangle\langle 1|_C$$

必须被假设为是相同的，即：

$$W_B^{(c)}(0) = W_B^{(c)}(1) \tag{206}$$

态$|0\rangle_c$和$|1\rangle_c$可以由双正交分解表示为：

$$|0\rangle_C = \sum_i \sqrt{p_i}\,|a_i\rangle\langle b_i| \tag{207}$$

$$|1\rangle_C = \sum_i \sqrt{p_i}\,|a_i'\rangle\langle b_i|$$

同时$\mathcal{H}_A$上的约化密度算符

$$W_A^{(c)}(0) = \mathrm{Tr}_B|0\rangle\langle 0|_C = \sum_i p_i|a_i\rangle\langle a_i| \tag{208}$$

$$W_A^{(c)}(1) = \mathrm{Tr}_B|1\rangle\langle 1|_C = \sum_i p_i|a_i'\rangle\langle a_i'|$$

是不相等的，即：

$$W_A^{(c)}(0) \neq W_A^{(c)}(1) \tag{209}$$

由此断定存在一个由谱表示$W_A^{(c)}(0)$和$W_A^{(c)}(1)$定义的幺正算符$U_C\in\mathcal{H}_A$：

$$\{|a_i\rangle\} \xrightarrow{U_C} \{|a_i'\rangle\} \tag{210}$$

使得：

$$|0\rangle_C \xrightarrow{U_C} |1\rangle_C \tag{211}$$

遵守协议方案会产生$\mathcal{H}_A\otimes\mathcal{H}_B$上两个可供选择的纯态$|0\rangle_h$和$|1\rangle_h$中的一个（$h$

代表遵守协议)，这里状态对$\{|0\rangle_h, |1\rangle_h\}$取决于鲍伯的选择及其测量结果。

在欺骗方案中，通过假设$\mathcal{H}_B$上的约化密度算符 $W_B^{(h)}(0)$和 $W_B^{(h)}(1)$是相同的：

$$W_B^{(h)}(0) = W_B^{(h)}(1) \tag{212}$$

这就导致必然存在一个幺正算符 $U_h \in \mathcal{H}_A$，使得：

$$|0\rangle_h \xrightarrow{U_h} |1\rangle_h \tag{213}$$

这里 U_h 取决于鲍伯的选择及其测量结果。

现在，遵守协议方案和欺骗方案之间的区别在$\mathcal{H}_A$上是检测不到的，这意味着$\mathcal{H}_A$上的约化密度算符在遵守协议和欺骗方案两种情况下是相同的：

$$W_A^{(h)}(0) = W_A^{(c)}(0) \tag{214}$$

$$W_A^{(h)}(1) = W_A^{(c)}(1)$$

由于 U_h 通过谱表示 $W_A^{(h)}(0)$和 $W_A^{(h)}(1)$来定义，因此 $U_h = U_C$。但是我们假设 U_h 取决于鲍伯的选择和测量结果，而 U_C 是由鲍伯的 EPR 策略唯一地定义的，在该策略中没有确定的选择或测量结果。结论如下：不可能存在这样的比特承诺协议，使得如果鲍伯遵循协议的话，无论爱丽丝还是鲍伯都不能欺骗；或者如果鲍伯通过 EPR 策略违背协议的话，爱丽丝也能欺骗；事实上如果在遵循协议的情况下，鲍伯和爱丽丝都不能欺骗的话，那么在欺骗的情况下，鲍伯而非爱丽丝一定能够欺骗。

类似的论证还可以排除掉这样的协议，即如果鲍伯遵循协议，则任何一方都不可以欺骗(同上)，但如果鲍伯实施 EPR 策略，那么就会导致 $W_B(0) \approx W_B(1)$，所以鲍伯有一定的概率能欺骗成功，但爱丽丝有比鲍伯更大的概率能欺骗成功。再次指出，当鲍伯实施 EPR 策略时使爱丽丝有一定概率欺骗成功的幺正变换 U_c，则在鲍伯遵循协议的情况下必然保证爱丽丝能欺骗成功。但前面的假设是，如果鲍伯遵守协议则爱丽丝不能欺骗，因为在这种情况下幺正变换 U_h 取决于鲍伯的选择和测量结果，这对于爱丽丝是未知的。因此就可以断定，并不存在这样的协议。

因此，该定理的证明并没有漏洞。(在该定理意义上的)绝对安全的量子比特承诺确实是不可能的。

6. 量子计算

6.1 丘奇—图灵论题和计算复杂性

计算的经典理论关心的是什么能够被计算的问题，以及计算的效率如何的问题。

阿伦索·丘奇(Alonzo Church)、库尔特·哥德尔(Kurt Gödel)等人关于可计算性的多种形式概念能够被证明是等价于阿兰·图灵(Alan Turing)用图灵机来定义的可计算性的概念，参见[Lewis and Papadimitriou(刘易斯和博派德米特)，1981]。图灵机是一个抽象的计算设备，它能够处于由可能态构成的有限集中的一个可能态上。它有一个由用于存储信息的连续单元格(在每个单元格上存储0，1或空白)构成的很可能是无穷的磁带，以及一个活动磁头用于读取单元格上的信息。依赖于单元格上的符号以及机器的状态，磁带上的符号会发生变化，从而改变状态，将单元格移动到左边或右边，直到计算结束它最终停机。图灵机T上的一个程序(比如，为执行寻找整数的质因数算法而编写的一个程序)是一个有限字符串——它可以被表示为一个二进制数$b(T)$——用来为每个态和每个符号指派新的态和新的符号，并指示磁头的移动。图灵表明，存在一个通用图灵机U，可以模拟任意图灵机T的程序，最多以多项式衰减。也就是说，如果我们以$b(T)$和T的输入初始化U，那么U就可以执行与T一样的计算，为了模拟T的一步运算，U所需要耗费的步骤数是$b(T)$的一个多项式函数。丘奇—图灵论题将可计算函数的类等同于可以由通用图灵机计算的函数的类。与之等价地，我们可以用判定问题——即答案为是或否的问题(例如判定某个给定的数字是否为质数的问题)——来阐述丘奇—图灵论题。

从直观上看，某些计算比另一些更困难，或者说某些运算会花费比另一些运算更多的时间。一个算法的计算复杂性是通过一个图灵机执行该算法所需的步骤数来度量的。对于一个判定问题，如果存在一个算法能在多项式时间内解决该问题，即解决该问题所需的步骤数是输入值的规模n(用于存储输入值的比特的数量)的多项式函数，那么它被称为是属于复杂性类$\mathbf{P}$，即它是容易的或者说易于处理的。如果不存在一个多项式时间算法能解决某个问题，那么该问

题就是难的或者说难于处理的。如果解决某个问题应用的大多数有效算法所需步骤数都是输入值的规模 n 的指数函数，那么这个问题就被称为是属于复杂性类 **EXP** 的。这里的步骤数针对的是最坏情况下的运行时间 τ，对于一个多项式时间算法来说，它有 $\mathcal{O}(n^k)$ 阶，对于一个指数时间算法来说，它有 $\mathcal{O}(2^n)$ 阶。

注意对于在某些范围内的输入规模来说，指数时间算法也可能比多项式算法更有效率，因此要小心地理解上面的术语。考虑下面的这个例子［Barenco（巴林科），1998，145］：$\tau_P(n)=10^{-23}n^{1000}+10^{23}/n\approx\mathcal{O}(n^{1000})$，因为对于足够大的 n 来说，包含多项式的项将会占支配地位，即对于一个确定的参数 c 来说，$\tau_P(n)<cn^{1000}$，而 $\tau_E(n)=10^{23}n^{1000}+10^{-23}2^n\approx\mathcal{O}(2^n)$，因为对于足够大的 n 来说，包含指数的项将会占支配地位，即对于一个确定的参数 c 来说，$\tau_E(n)<c2^n$。但是当 n 足够小时，$\tau_E(n)<\tau_P(n)$。

上面定义的图灵机是一个确定性的机器。一个非确定性图灵机或者说概率图灵机针对每个态和每个符号，会在多种变换中做出一个随机的选择（即产生一个新的态、一个新的符号和机头的移动）。对于每个选择的序列，变化的序列对应于一个确定性图灵机执行的步骤的序列。如果这些机器中任意一个停机，我们就认为完成了计算。显然，一个非确定性图灵机不能计算确定性图灵机不能计算的函数，但是一个普遍的信念（而非证明）是，某些特定问题用非确定性图灵机解决会比用任何确定性图灵机解决更有效。复杂性类 **NP** 就是这些问题的类，它们能在多项式时间内解决，如果利用非确定性图灵机的话。这就等价于使用确定性图灵机在多项式时间内能解决的问题的类。例如，我们相信（而非证明）将整数分解为其质因数的因子分解问题是一个“难”问题：不存在针对此问题的已知的多项式时间算法。然而，检验一个候选的质因数是否确实是某个整数的质因数的问题却能在多项式时间内解决，因此，因数分解性是一个 **NP** 问题。

很明显 $\mathbf{P}\subseteq\mathbf{NP}$，但是在复杂性理论中是否有 $\mathbf{P}=\mathbf{NP}$ 是一个开放性问题。如果能够证明所有 **NP** 问题都存在一种解法，实现该解法所需的步骤数是解决 **NP** 完全问题所需步骤数的多项式函数，那么 **NP** 问题就是 **NP** 完全的。因此如果一个 **NP** 完全问题能在多项式时间内解决，那么所有的 **NP** 问题都能在多项式时间内解决，即 $\mathbf{P}=\mathbf{NP}$。决定布尔函数 $f\{0,1\}^n\to\{0,1\}$ 能否被满足（即是否存在一个输入值的集使得函数取值为 1，或者等价地是否存在一个针对按布尔逻辑构成

的复合命题中的原子命题的真值指派，以使得复合命题为真)的问题就是一个 NP 完全问题。可因数分解性问题是一个还不能被视为 NP 完全的 NP 问题。

既然图灵机能够以最多为多项式函数倍的速度模拟任何经典计算设备，那么复杂性的类对于任何计算模型都是一样的。例如，一个电子线路计算机通过将存储在输入寄存器中的数据经由布尔电路变换为输出寄存器中的数据来计算函数的值，其中输入寄存器中的数据表示函数的输入值，输出寄存器中的数据表示函数的输出值，布尔电路由通过线路相连的基础布尔门构成。基础布尔门有 1 比特门(如非门，它将 0 变为 1，或者相反)和 2 比特门(如与门，当且仅当 2 个输入比特都为 1 时，结果为 1，否则为 0)，并且可以证明这些门的组合构成一个“通用集”，能够满足 n 比特的任意变换。事实上，十六个 2 比特布尔门中的一个，即与非门(NAND gate 或 NOT AND gate)自身就构成了一个通用集，在与非门中，当且仅当 2 个输入比特都为 1 时，它取值为 0。

在量子计算机的线路模型中，寄存器存储量子比特，这些量子比特由基本的幺正门操控。可以证明[Nielsen and Chuang, 2000, 188]一个由单量子比特和双量子比特幺正门构成的集合——CNOT 门、哈达玛门、相位门和 $\pi/8$ 门——就是一个通用集，在此意义下由这些门连接组合起来的一个量子线路能够以任意精确度模拟 n 量子比特的任意幺正变换。CNOT(C 代表控制)门有两个输入量子比特，其中一个为“控制”量子比特，另一个为“受控”量子比特。这个门的作用为，当且仅当控制量子比特为 $|1\rangle$ 时，使受控量子比特反转。剩下的三个门为单量子比特门。哈达玛门将 $|0\rangle$ 转变为 $(|0\rangle + |1\rangle)/\sqrt{2}$，将 $|1\rangle$ 转变为 $(|0\rangle - |1\rangle)/\sqrt{2}$，有时它被认为是“非(NOT)的平方根”门，因为连续使用两次可将 $|0\rangle$ 转变为 $|1\rangle$，反之亦然。相位门保持 $|0\rangle$ 不变，将 $|1\rangle$ 转变为 $i|1\rangle$。$\pi/8$ 门保持 $|0\rangle$ 不变，将 $|1\rangle$ 转变为 $e^{i\pi/4}|1\rangle$。进一步的讨论以及 $\pi/8$ 门的命名原因请参见[Nielsen and Chuang, 2000, 174]。

量子计算还存在其他模型。在罗森德福和布里格尔(Raussendorf and Briegel)[2001b; 2001a]的“团簇态”(cluster state)或“单向”量子计算机中，需要制备一个由多个量子比特构成的确定的纠缠态(所谓的“团簇态”)，需要独立于计算，然后在该态上实施一系列单量子比特测量，在此过程中，下一次选择测

量什么样的可观测量取决于上次测量的结果。该过程不涉及任何幺正变换。引人注意的是，由量子幺正变换和测量构成的任意量子线路，都能够被一个团簇态计算机利用同样的资源在相同的时间内模拟[Jozsa，2005；Nielson，2003；Nielson，2005]。

一个有意思的问题是，量子计算机是否能够执行图灵机不可能完成的计算任务，或者能够比任何图灵机更有效率地执行某些任务。由于图灵机是由它所执行的程序定义的，程序又可以由一组有限的符号集来确定，因此只有可数多个图灵机。但是有无数多个关于自然数的函数，因此存在无数多个不可计算的函数，即不可由图灵机计算的函数。量子计算机不能计算非图灵可计算的函数，因为图灵机能够以任意的精确度——尽管有时候是无效率的，以指数级衰减的速度[Feynman，1982]——模拟任何系统的动力学演化，无论经典的还是量子的。但是有一些计算任务，量子计算可以利用纠缠来执行，那对于任何图灵机来说都是不可能的。回顾3.1中关于贝尔针对EPR佯谬的辩护的讨论：量子计算机而非经典计算机，能执行这样的任务，即迅速为位于不同地点的一对输入赋予一对值(0或1)，这对输入之间存在违反贝尔不等式的相互关系，而响应时间要短于光在两地间传播所需的时间。

量子计算机研究中当前关注的问题是量子计算机在计算某些图灵可计算问题时是否能比任何图灵机都更有效率。在下一节中，我将讨论一些量子算法，相较于任意的经典算法，或者说相较于已知的任意经典算法，它们能达到指数级加速运算的效果。这些量子算法中最引人瞩目的是肖尔的因数分解算法，以及与之有关的解决离散对数问题①的算法。

因数分解问题在密码学中有重要的实际应用价值。公共密钥分配协议如RSA[Rivest *et al.*(李维斯特等)，1978](广泛应用于互联网间的商业交易和银行与金融机构之间的交易等)依赖于因子分解这个“难”问题。普雷希尔(Preskill)[2005]曾指出，当前如果使用一个由数百个工作站构成的网络系统，并执行已知的最好的经典因数分解算法(数字筛除算法)，那么一个130位整数

① 对于一个给定的质数 p 和一个与 p 互质的整数 q，x 的离散对数就是使 $q^r = x \bmod p$ 成立的整数 r。进一步讨论请参见[Nielsen and Chuang，2000，238]。

的65位质因数能够在一个月左右被分解出来。他估算出一个400位整数的因子分解将会耗时10^{10}年，那正是宇宙的年龄。为了明白RSA协议背后的思想，我们假设爱丽丝想要发送一个秘密的信息给鲍伯。鲍伯的公共密钥由两个大数s和c组成。爱丽丝按$e = m^s \bmod c$将(二进制下的)信息m加密，并将加密信息发送给鲍伯。鲍伯通过$e^t \bmod c$将信息解密，这里t是一个只有鲍伯知道的整数。$m = e^t \bmod c$中的整数t很容易由s以及c的质因数确定，但是$c = pq$是只有鲍伯知道的两个大质数的乘积，对于窃听者夏娃来说，她只有将数c分解为它的质因数，才能读取信息。这个方案的巧妙之处在于这样的事实，即爱丽丝和鲍伯之间不需要分配密钥：鲍伯的密钥$\{s, c\}$是公共的，且允许任何人发送加密信息给他。如果能够构建一个量子计算机以执行肖尔算法，那么依赖于大数因数分解困难性的密钥分配协议将会是不安全的。

6.2 量子算法

在以下的三部分中，我关注的是涉及多伊奇异或(XOR)算法[Deutsch，1985]、西蒙的周期寻找算法(period-finding algorithm)[Simon，1994；1997]以及肖尔因数分解算法[Shor，1994；1997]的信息处理过程，主要比较作为经典计算基础的布尔逻辑和由希尔伯特空间中投影几何表示的非布尔逻辑之间的区别，在后者中，希尔伯特空间中的子空间结构取代了经典逻辑中的集合论结构。这三个算法最后都包含相似的几何公式化。

从根本上讲，这三个算法都包含一个关于函数的整体性质的结论，即析取性。并集被表示为一个恰当的希尔伯特空间的子空间，除了交点和重叠部分之外，其他可能的并集被表示为与其正交的子空间。真正的并集由通过测量得到的态矢量所在的子空间确定。一般来说算法不得不运行多次，因为态有可能出现在重叠区。这些量子计算的本质特征就在于不需要确定析取项的真值就能将真正的并集从那些可能的并集中分辨出来。在经典计算中，不先确定析取项的真值就不可能得到真正的并集。更一般地，量子计算机不需要计算在量子力学中属于是多余的那些信息就能计算函数的整体性质，而对于经典计算来说这些信息则是必需的。

除了上面提到的三个以外，还有其他的量子算法，如格罗夫排序算法，它

相对于任意的经典算法实现了平方加速的作用。进一步讨论请参见[Nielsen and Chuang, 2000]和[Jozsa, 1999]。

多伊奇 XOR 算法和多伊奇—约萨算法

令 $B=(0, 1)$为一个布尔代数(模为 2 的整数加法群)。在多伊奇 XOR 问题[Deutsch, 1985]中，我们得到一个“黑箱”或启示程序(oracle)用来计算函数 $f\colon B\rightarrow B$，我们必须确定这个函数是一个“常数”函数(对不同输入取相同的值)还是一个“变量”函数(对不同输入取不同的值)。在经典计算中，唯一的方法就是我们要分别针对 0 和 1 应用这个启示程序两次，然后比较输出的结果。

在对布尔函数的量子计算中，幺正变换 $U_f\colon |x\rangle|y\rangle\rightarrow|x\rangle|y\oplus f(x)\rangle$就相当于将输入值和相应输出值相联系起来的“黑箱”。[①] 计算按照以下程序进行：输入和输出寄存器是在标准基上初始化态为 $|0\rangle\langle 0|$ 的单量子比特寄存器。应用一次哈达玛变换到输入寄存器，产生一个相当于两个可能输入值 0 或 1 的线性叠加态，然后应用变换 U_f 到两个寄存器上，实现以下变换：

$$|0\rangle|0\rangle\xrightarrow{H}\frac{1}{\sqrt{2}}(|0\rangle+|1\rangle)|0\rangle \tag{215}$$

$$\xrightarrow{U_f}\frac{1}{\sqrt{2}}(|0\rangle|f(0)\rangle+|1\rangle|f(1)\rangle) \tag{216}$$

如果函数是常数型函数，那么两个寄存器最终的复合态为以下两个正交态中的一个：

$$|c_1\rangle=\frac{1}{\sqrt{2}}(|0\rangle|0\rangle+|1\rangle|0\rangle) \tag{217}$$

$$|c_2\rangle=\frac{1}{\sqrt{2}}(|0\rangle|1\rangle+|1\rangle|1\rangle) \tag{218}$$

如果函数是平衡型函数，那么最终的混合态为以下两个正交态中的一个：

---

① 注意用来计算函数的两个量子寄存器在幺正变换中不是一一对应的。对于函数 f，不同的输入值 x 或 y 被表示为正交态 $|x\rangle$和 $|y\rangle$。因此对于 $x\neq y$，如果 $f(x)=f(y)$，那么变换 $W_f\colon |x\rangle\rightarrow|f(x)\rangle$可能不是幺正的，因为正交态 $|x\rangle$和 $|y\rangle$根据 W_f 将必然会映射到同一个态。幺正变换的能力，即在计算不可逆函数时保持可逆性，可以通过为函数的每一个输出值保留一个输入值的方式来达到。

$$|b_1\rangle = \frac{1}{\sqrt{2}}(|0\rangle|0\rangle + |1\rangle|1\rangle) \tag{219}$$

$$|b_2\rangle = \frac{1}{\sqrt{2}}(|0\rangle|1\rangle + |1\rangle|0\rangle) \tag{220}$$

态$|c_1\rangle$，$|c_2\rangle$和$|b_1\rangle$，$|b_2\rangle$分别构成$\mathcal{H}^2\otimes\mathcal{H}^2$上的两个面$P_c$，$P_b$，它们由以下投影算符表示：

$$P_c = P_{|c_1\rangle} + P_{|c_2\rangle} \tag{221}$$

$$P_b = P_{|b_1\rangle} + P_{|b_2\rangle} \tag{222}$$

除了有一条交线之外，这两个面是正交的，因此它们的投影算符是对易的。这条交线由以下矢量①表示：

$$\frac{1}{2}(|00\rangle + |01\rangle + |10\rangle + |11\rangle) = \frac{1}{\sqrt{2}}(|c_1\rangle + |c_2\rangle) = \frac{1}{\sqrt{2}}(|b_1\rangle + |b_2\rangle) \tag{223}$$

在由态$|0'\rangle = H|0\rangle$，$|1'\rangle = H|1\rangle$共轭的“初始”(prime)基上，交线为态$|0'\rangle|0'\rangle$，“常数”面由$|0'\rangle|0'\rangle$，$|0'\rangle|1'\rangle$构成，“平衡”面由$|0'\rangle|0'\rangle$，$|1'\rangle|1'\rangle$构成。注意：

$$|0'\rangle|1'\rangle = \frac{1}{\sqrt{2}}(|c_1\rangle - |c_2\rangle) \tag{224}$$

$$|1'\rangle|1'\rangle = \frac{1}{\sqrt{2}}(|b_1\rangle - |b_2\rangle) \tag{225}$$

在算法的一般公式化中，决定函数f为常数函数还是平衡函数的方法是测量处于典型基上的输出寄存器。如果结果为$0'$（得到这个结果的概率为1/2，无论态最终是处于常数平面还是平衡平面），计算是没有结果的，关于函数f我们没有得到任何信息。如果结果为$1'$，那么我们再测量输入寄存器。如果输入寄存器测量的结果为$0'$，那么函数为常数函数；如果结果为$1'$，那么函数为平衡函数。

另外一种方法——这是说与西蒙算法和肖尔算法相比较——是测量处于本征态$|0'0'\rangle$，$|0'1'\rangle$，$|1'0'\rangle$，$|1'1'\rangle$的可观测量。终态处于正交于矢量

① 这里$|00\rangle = |0\rangle|0\rangle$，其他同理。

$|1'0'\rangle$ 的三维子空间，可能处于常数平面或平衡平面。如果态处于常数平面，我们将会以 1/2 的概率得到结果 0'0'（因为终态位于 $\pi/4$ 角时得到 $|0'0'\rangle$），在这种情况下计算是没有结果的，或者以 1/2 的概率得到结果 0'1'。如果态处于平衡平面，我们也将会以 1/2 的概率得到结果 0'0'，在这种情况下，计算是没有结果的，或者以 1/2 的概率得到结果 1'1'。因此，无论是哪种情况，我们都能够以 1/2 的概率，区分出两种算法，它们分别运行于由下列两个"常数"平面和"平衡"平面表示的量子并集：

$$P_c = P_{|0'0'\rangle} \vee P_{|0'1'\rangle} \tag{226}$$

$$P_b = P_{|0'0'\rangle} \vee P_{|1'1'\rangle} \tag{227}$$

且我们不需要找出计算中并集的真值(即不需要知道在"常数"情况下是否函数将 0 映射到 0 的同时将 1 映射到 0，还是将 0 映射到 1 的同时将 1 映射到 1，在"平衡"情况下也一样)。注意我们也能将哈达玛变换应用到两个寄存器的终态，并在计算基上进行测量，因为 $|0'0'\rangle \xrightarrow{H} |00\rangle$,等等。

在执行相同的计算任务时，多伊奇 XOR 算法是第一个体现出相对于经典算法具有加速性的量子算法。然而，应用该算法在计算时有一半的概率会失败，因此只有当该算法成功时，才能说相对于经典算法，该算法提高了计算效率，即使是这样，这种提高也是有限的：即运行一次量子算法对应于运行两次经典算法。下面介绍的多伊奇算法的变异避免了这种不足。见[Cleve *et al.*(克里夫等)，1998]。

首先我们分别将两个寄存器初始化为 $|0\rangle$ 和 $|1\rangle$（代替初始化为 $|0\rangle$ 和 $|0\rangle$），然后对两个寄存器都应用哈达玛变换，将会产生以下变换：

$$|0\rangle|1\rangle \xrightarrow{H} \frac{|0\rangle+|1\rangle}{\sqrt{2}}\frac{|0\rangle-|1\rangle}{\sqrt{2}} \tag{228}$$

因为：

$$U_f|x\rangle|y\rangle = |x\rangle|y \oplus f(x)\rangle \tag{229}$$

因此：

$$U_f|x\rangle\frac{|0\rangle-|1\rangle}{\sqrt{2}} = \begin{cases} |x\rangle\dfrac{|0\rangle-|1\rangle}{\sqrt{2}} & \text{如果}\quad f(x)=0 \\ -|x\rangle\dfrac{|0\rangle-|1\rangle}{\sqrt{2}} & \text{如果}\quad f(x)=1 \end{cases} \tag{230}$$

它可以被表示为：

$$U_f|x\rangle\frac{|0\rangle-|1\rangle}{\sqrt{2}}=(-1)^{f(x)}|x\rangle\frac{|0\rangle-|1\rangle}{\sqrt{2}} \tag{231}$$

注意现在函数值呈现为输入寄存器终态的一个相位，这种现象被称为“相位返程”。对于输入态 $1/\sqrt{2}(|0\rangle+|1\rangle)$，我们有：

$$U_f\frac{|0\rangle+|1\rangle}{\sqrt{2}}\frac{|0\rangle-|1\rangle}{\sqrt{2}}=\frac{(-1)^{f(0)}|0\rangle+(-1)^{f(1)}|1\rangle}{\sqrt{2}}\frac{|0\rangle-|1\rangle}{\sqrt{2}} \tag{232}$$

它可以被表示为：

$$U_f\frac{|0\rangle+|1\rangle}{\sqrt{2}}\frac{|0\rangle-|1\rangle}{\sqrt{2}} \tag{233}$$

$$=\begin{cases}\pm\dfrac{|0\rangle+|1\rangle}{\sqrt{2}}\dfrac{|0\rangle-|1\rangle}{\sqrt{2}}=\pm|0'\rangle|1'\rangle & \text{如果}\quad f(0)=f(1)\\ \pm\dfrac{|0\rangle-|1\rangle}{\sqrt{2}}\dfrac{|0\rangle-|1\rangle}{\sqrt{2}}=\pm|1'\rangle|1'\rangle & \text{如果}\quad f(0)\neq f(1)\end{cases}$$

前述算法中两个寄存器的终态最后成为常数平面上两个正交态中的一个或平衡平面上两个正交态中的一个，现在所用的算法与这样的情况不同，该算法中终态最后成为常数平面上的 $\pm|0'1'\rangle$，或平衡平面上的 $\pm|1'1'\rangle$，这些态能够被区分因为它们是正交的。因此仅仅运行一次算法，我们就能毫无疑问地确定函数是常数函数还是平衡函数。事实上，我们可以通过在初始基上简单地测量输入寄存器就能够区分这两种可能性。需要注意的是，我们需要在输入寄存器上实施一次哈达玛变换(将 $|0'\rangle$ 变为 $|0\rangle$，将 $|1'\rangle$ 变为 $|1\rangle$)，才能通过在初始基上测量输入寄存器区分这两种可能性。还要注意输出寄存器的态是没有变化的：在这个过程结束后，它处于态 $|1'\rangle=H|1\rangle$，如式(228)，且没有被测量过。

多伊奇 XOR 问题可以被推广，用来确定一个布尔函数 $f:B^n\rightarrow B$ 是常数函数还是平衡函数的问题(“多伊奇问题”)，前提是这个函数要么是常数函数要么是平衡函数。在这里“平衡”意味着函数取值为 0 或 1 的次数相同，即各 2^{n-1} 次。运行一次多伊奇—约萨算法[1992]就能确定 f 是常数函数还是平衡函数。

我们首先将 n 量子比特输入寄存器设置为 $|0\rangle$(态 $|0\cdots0\rangle=|0\rangle\cdots|0\rangle$

的简写)，将单量子比特输出寄存器设置为$|1\rangle$。然后，根据幺正变换 U_f，我们在输入寄存器上应用一个 n 重哈达玛变换，在输出寄存器上应用一个哈达玛变换，最后在输入寄存器上再应用一个 n 重哈达玛变换。

首先要注意：

$$H|x\rangle=\frac{1}{\sqrt{2}}\sum_{y\in\{0,1\}}(-1)^{xy}|y\rangle \tag{234}$$

因此：

$$H^{\otimes n}|x_1,\cdots,x_n\rangle=\frac{1}{\sqrt{2^n}}\sum_{y_1,\cdots,y_n}(-1)^{x_1y_1+\cdots+x_ny_n}|y_1,\cdots,y_n\rangle \tag{235}$$

上式可以表示为：

$$H^{\otimes n}|x\rangle=\frac{1}{\sqrt{2^n}}\sum_{y\in\{0,1\}}(-1)^{x\cdot y}|y\rangle \tag{236}$$

其中 $x\cdot y$ 为 x 和 y 的模为 2 逐位内积。

幺正变换(哈达玛变换，U_f)使得：

$$|0\rangle^{\otimes n}|1\rangle\xrightarrow{H}\sum_{x\in\{0,1\}^n}\frac{|x\rangle}{\sqrt{2^n}}\frac{|0\rangle-|1\rangle}{\sqrt{2}} \tag{237}$$

$$\xrightarrow{U_f}\sum_{x}\frac{(-1)^{f(x)}}{\sqrt{2^n}}|x\rangle\frac{|0\rangle-|1\rangle}{\sqrt{2}} \tag{238}$$

$$\xrightarrow{H}\sum_{y}\sum_{x}\frac{(-1)^{x\cdot y+f(x)}}{\sqrt{2^n}}\frac{|y\rangle}{\sqrt{2^n}}\frac{|0\rangle-|1\rangle}{\sqrt{2}} \tag{239}$$

现在考虑输入寄存器的态：

$$\sum_{y}\sum_{x}\frac{(-1)^{x\cdot y+f(x)}}{2^n}|y\rangle=\sum_{x}\frac{(-1)^{f(x)}}{2^n}|0\cdots0\rangle+\cdots \tag{240}$$

注意线性叠加中态$|0\cdots0\rangle$的系数(振幅)为 $\sum_x\frac{(-1)^{f(x)}}{2^n}$。如果 f 为常数函数，系数为 ±1，因此其他项的系数必须为 0。如果 f 为平衡函数，对于一半的 x 的值，$f(x)=0$；对于另一半，$f(x)=1$，因此$|0\cdots0\rangle$的正负系数都消为 0。换句话说，如果 f 为常数函数，输入寄存器的态为 $\pm|0\cdots0\rangle$，如果 f 为平衡函数，态处于正交子空间中。

这就是描述该算法如何运行的一般方式，在一定程度上这会使其几何图景

变得模糊。为了简单起见，现在考虑 $n=2$ 的情况。在完成变换 U_f 之后，但在完成最终的哈达玛变换之前，输入寄存器的态为：

$$\pm\frac{1}{2}(|00\rangle+|01\rangle+|10\rangle+|11\rangle) \tag{241}$$

如果 f 为常数函数时，或者态为如下形式：

$$\frac{1}{2}(\pm|00\rangle\pm|01\rangle\pm|10\rangle\pm|11\rangle) \tag{242}$$

如果 f 为平衡函数时，这里的系数情况是两个为 $+1$，两个为 -1。显然，存在三个（不同）相互正交的平衡态，且它们都正交于常数态。因此，这三个平衡态位于与常数态正交的三维子空间中，因此能够与平衡态相区别开来。最终的哈达玛变换将常数态转换为：

$$\pm\frac{1}{2}(|00\rangle+|01\rangle+|10\rangle+|11\rangle)\xrightarrow{H}\pm|00\rangle \tag{243}$$

将这三个平衡态转换为正交于 $|00\rangle$ 的三维子空间中的态。因此，要确定函数是常数函数还是平衡函数，我们只需测量输入寄存器以检查它是否处于态 $|00\rangle$。

西蒙算法

这个算法要解决的问题是找到周期函数 f: $B^n\to B^n$ 的周期，即满足以下关系的一个布尔函数的周期：

$$f(x_i)=f(x_j)\text{，当且仅当对于所有的 } x_i,x_j\in B^n \text{ 有 } x_j=x_i\oplus r \tag{244}$$

注意由于 $x\oplus r\oplus r=x$，因此该函数是二对一的。

西蒙算法能有效地解决该问题，相对于经典算法具有指数加速的能力，参见[Simon，1994；1997]。该算法执行起来与多伊奇—约萨算法一样，最初输入寄存器和输出寄存器被设定为以下计算基上的态 $|0\cdots0\rangle|0\rangle$：

$$|0\cdots0\rangle|0\rangle\xrightarrow{H}\frac{1}{\sqrt{2^n}}\sum_{x=0}^{2^n-1}|x\rangle|0\rangle \tag{245}$$

$$\xrightarrow{U_f}\frac{1}{\sqrt{2^n}}\sum_{x}|x\rangle|f(x)\rangle \tag{246}$$

$$=\frac{1}{\sqrt{2^{n-1}}}\sum_{x_i}\frac{|x_i\rangle+|x_i\otimes r\rangle}{\sqrt{2}}|f(x_i)\rangle \tag{247}$$

这里 U_f 如下式，是执行布尔函数的幺正变换：

$$U_f: |x\rangle|y\rangle \to |x\rangle|y \oplus f(x)\rangle \tag{248}$$

理解该算法是如何执行的一般方法是，考虑如果我们测量输出寄存器而保持输入寄存器的态不变①会怎么样，其中输入态形式为：

$$\frac{|x_i\rangle + |x_i \oplus r\rangle}{\sqrt{2}} \tag{249}$$

该态包含了信息 r，但它出现的方式却是与我们不想要的随机补偿 x_i 相加在一起，其中 x_i 取决于测量的结果。对态的直接测量会等概率地产生任意的 $x \in B^n$，这对于 r 不能提供任何信息。

现在我们应用哈达玛变换：

$$\frac{|x_i\rangle + |x_i \oplus r\rangle}{\sqrt{2}} \xrightarrow{H} \frac{1}{\sqrt{2^n}} \sum_{y \in B^n} \frac{(-1)^{x_i \cdot y} + (-1)^{(x_i \oplus r) \cdot y}}{\sqrt{2}} |y\rangle \tag{250}$$

$$= \sum_{y: r \cdot y = 0} \frac{(-1)^{x_i \cdot y}}{\sqrt{2^{n-1}}} |y\rangle \tag{251}$$

这里会出现后面的等号是因为如果 $r \cdot y = 1$，各项会发生破坏性的干涉。最后，我们测量处于计算基上的输入寄存器，会（等概率地）得到一个值 y，使得 $r \cdot y = 0$。然后我们重复该算法足够多次，从而找到足够多的值 y_i 使得满足 $r \cdot y_1 = 0, \cdots, r \cdot y_k = 0$ 的 r 能被确定。

为了在几何上理解该算法是如何进行的，我们考虑 $n = 2$ 的情况。周期 r 的可能值为01，10，11，对应于经过幺正变换 U_f 后的输入寄存器和输出寄存器的态为：

$$r = 01: \frac{1}{2}(|00\rangle + |01\rangle)|f(00)\rangle + \frac{1}{2}(|10\rangle + |11\rangle)|f(10)\rangle$$

$$r = 10: \frac{1}{2}(|00\rangle + |10\rangle)|f(00)\rangle + \frac{1}{2}(|01\rangle + |11\rangle)|f(01)\rangle$$

$$r = 11: \frac{1}{2}(|00\rangle + |11\rangle)|f(00)\rangle + \frac{1}{2}(|01\rangle + |10\rangle)|f(01)\rangle$$

---

① 这里对输出寄存器的测量是便于概念化的一个教学策略。只有输入寄存器才是被实际测量的。输入寄存器处于一个混合态，我们可以将其看作是与对输出寄存器测量得到的结果分布相关的混合态。

注意这种情况能够被简化为与多伊奇 XOR 算法中一样的几何结构。对于 $r=10$，输入寄存器的态为 $|c_1\rangle=\frac{1}{\sqrt{2}}(|00\rangle+|10\rangle)$ 或 $|c_2\rangle=\frac{1}{\sqrt{2}}(|01\rangle+|11\rangle)$，对于 $r=11$，输入寄存器的态为 $|b_1\rangle=\frac{1}{\sqrt{2}}(|00\rangle+|11\rangle)$ 或 $|b_2\rangle=\frac{1}{\sqrt{2}}(|01\rangle+|10\rangle)$，具体取决于对输出寄存器测量的结果。因此这三个可能的周期与 $\mathcal{H}^2\otimes\mathcal{H}^2$ 中的三个平面有关，它们相当于多伊奇 XOR 算法中的常数平面和平衡平面，再加上第三个平面，这三个平面相交于矢量 $|00\rangle$ 代表的直线。在应用哈达玛变换得到的原始基上，这三个平面分别为：

$r=01$：平面为 $|0'0'\rangle$，$|1'0'\rangle$

$r=10$：平面为 $|0'0'\rangle$，$|0'1'\rangle$（相当于“常数”平面）

$r=11$：平面为 $|0'0'\rangle$，$|1'1'\rangle$（相当于“平衡”平面）

我们可以简单地测量初始基上的输入寄存器以找到周期。或者，我们可以应用一次哈达玛变换（这意味着在 r 平面的上述表示中放弃使用初始基）并在计算基上进行测量。除了由矢量 $|00\rangle$ 构成的交线外，这三个平面是正交的。因此通过测量处于本征态 $|00\rangle,|01\rangle,|10\rangle,|11\rangle$ 的可观测量，这三个可能的周期就能够被区分开来，除非测量会使寄存器的态投影（“塌缩”）到交线对应的态 $|00\rangle$ 上（发生的概率为 1/2）。因此一般来说算法不得不重复多次，直到我们发现一个不是 00 的结果。

在 $n=2$ 的情况下西蒙算法可以简化为多伊奇 XOR 算法。那么在其他情况下呢？我们可以考虑 $n=3$ 的情况来理解一般情况会怎么样。现在存在七个可能的周期：001，010，011，100，101，110，111。对于周期 $r=001$，经过幺正变换 U_f 后两个寄存器的态为：

$$\frac{1}{2\sqrt{2}}(|000\rangle+|001\rangle)|f(000)\rangle+\frac{1}{2\sqrt{2}}(|010\rangle+|011\rangle)|f(010)\rangle+$$

$$\frac{1}{2\sqrt{2}}(|100\rangle+|101\rangle)|f(100)\rangle+\frac{1}{2\sqrt{2}}(|110\rangle+|111\rangle)|f(110)\rangle$$

如果我们测量输出寄存器，输入寄存器的态为以下四个态中的一个，具体取决于测量的结果：

$$\frac{1}{\sqrt{2}}(|000\rangle+|001\rangle)=\frac{1}{2}(|0'0'0'\rangle+|0'1'0'\rangle+|1'0'0'\rangle+|1'1'0'\rangle)$$

$$\frac{1}{\sqrt{2}}(|010\rangle+|011\rangle)=\frac{1}{2}|0'0'0'\rangle-|0'1'0'\rangle+|1'0'0'\rangle-|1'1'0'\rangle)$$

$$\frac{1}{\sqrt{2}}(|100\rangle+|101\rangle)=\frac{1}{2}(|0'0'0'\rangle+|0'1'0'\rangle-|1'0'0'\rangle-|1'1'0'\rangle)$$

$$\frac{1}{\sqrt{2}}(|110\rangle+|111\rangle)=\frac{1}{2}(|0'0'0'\rangle-|0'1'0'\rangle-|1'0'0'\rangle+|1'1'0'\rangle)$$

应用哈达玛变换意味着放弃初始基，因此如果周期为 $r=001$，那么输入寄存器最终处于由矢量 $|000\rangle,|010\rangle,|100\rangle,|110\rangle$ 构成的 $\mathcal{H}^2\otimes\mathcal{H}^2\otimes\mathcal{H}^2$ 的四维子空间上。

将同样的分析应用于其他六个可能周期的情况中。相应的子空间由以下矢量构成：

$$r=001:|000\rangle,|010\rangle,|100\rangle,|110\rangle$$
$$r=010:|000\rangle,|001\rangle,|100\rangle,|101\rangle$$
$$r=011:|000\rangle,|011\rangle,|100\rangle,|111\rangle$$
$$r=100:|000\rangle,|001\rangle,|010\rangle,|011\rangle$$
$$r=101:|000\rangle,|010\rangle,|101\rangle,|111\rangle$$
$$r=110:|000\rangle,|001\rangle,|110\rangle,|111\rangle$$
$$r=111:|000\rangle,|011\rangle,|101\rangle,|110\rangle$$

除了一些在二维平面中的交叉区域外，这些子空间都是正交的。通过计算基上的测量就能确定周期。重复测量最终将产生足够多的值，就可以确定末态会落在七个可能子空间中的哪一个上。在这种情况下($n=3$)，通过研究上面的列表能明确知道，不同于000的两个值就足以确定子空间，而且它们仅仅是满足 $y_i\cdot r=0$ 的值 y_i。注意子空间对应于量子析取。因此通过西蒙算法确定了函数的周期就意味着明确了这七个可能的并集中哪个是真的，也就是说，我们不需要确定并集的真值就能确定哪个子空间包含了这个态。

肖尔算法

肖尔因子分解算法利用了这样的事实，即一个正整数 $N=pq$ 的质因数可以通过确定一个函数周期的方式得到，该函数为 $f(x)=a^x \bmod N$，对于任意的 a

$<N$, a 与 N 互质，即除了 1 以外 a 与 N 没有共同的质因数。函数 $f(x)$ 的周期 r 取决于 a 和 N。一旦我们知道了周期，如果 r 为偶数且 $a^{r/2} \neq -1 \bmod N$，我们就能将 N 分解，这种情况出现的概率大于 1/2，如果 a 是随机选定的话（否则的话我们重新选择一个 a）。N 的质因数就是 $a^{r/2} \pm 1$ 与 N 的最大公因数，它能够通过欧几里得（Euclidpan）算法在多项式时间内找到。更多数论的结果请参见［Nielsen and Chuang，2000，Appendix 4］。因此将一个复合数 N 分解为两个质因数的积的问题就转化为寻找一个特定周期函数 $f:Z_S \to Z_N$ 的周期，这里 Z_N 是以 n 为模的整数的加法群（而不是像西蒙算法中的 B^n，即布尔代数 B 的 n 重笛卡尔积）。注意 $f(x+r) = f(x)$，当 $x + r \leqslant s$ 时。如果 r 能整除 s，那么函数 f 为周期函数，否则 f 为近似周期函数。

首先考虑这个算法的一般形式，如同我们通常进行公式化那样。我们先将输入寄存器（s 个量子比特）初始化为态 $|0\rangle \in \mathcal{H}^s$，将输出寄存器初始化为态 $|0\rangle \in \mathcal{H}^N$。将一个 s 重哈达玛变换应用到输入寄存器，按照幺正变换 U_f，执行函数 $f(x) = a^x \bmod N$：

$$|0\rangle|0\rangle \xrightarrow{H} \frac{1}{\sqrt{s}}\sum_{x=0}^{s-1}|x\rangle|0\rangle \tag{252}$$

$$\xrightarrow{U_f} \frac{1}{\sqrt{s}}\sum_{x=0}^{s-1}|x\rangle|x + a^x \bmod N\rangle \tag{253}$$

然后我们在计算基上测量输出寄存器①，得到一个由以下形式表示的输入寄存器的态：

$$\frac{1}{\sqrt{s/r}}\sum_{j=0}^{s/r-1}|x_i + j_r\rangle \tag{254}$$

这就是 r 能整除 s 的情况。值 x_i 为偏移，它取决于对输出寄存器测量的结果 i。上式中的总和由满足 $f(x_i + jr) = i$ 的 j 确定。当 r 不能整除 s 时，分析起来会稍微有些复杂，对此的讨论请参见［Barenco，1998，164］和［Jozsa，1997b］。由于态符号包含随机偏移量，因此对符号的随机测量不会产生关于周期的任何

① 如同在西蒙算法中那样，该测量完全是假设的，引入该测量是为了分析的方便，只有输入寄存器是确实被测量了。

信息。

现在对输入寄存器应用一个模为 s 的同余的整数进行的离散傅里叶变换，即下式表示的幺正变换：

$$|x\rangle \xrightarrow{U_{DFT_s}} \frac{1}{\sqrt{s}}\sum_{y=0}^{s-1} e^{2\pi i\frac{xy}{s}}|y\rangle, \text{对 } x \in Z_s \tag{255}$$

它将产生下述变换：

$$\frac{1}{\sqrt{\frac{s}{r}}}\sum_{j=0}^{\frac{s}{r}-1}|x_i+jr\rangle \xrightarrow{U_{DFT_s}} \frac{1}{\sqrt{r}}\sum_{k=0}^{r-1} e^{2\pi i\frac{x_i k}{r}}|ks/r\rangle \tag{256}$$

使偏移量转变为一个相因数，并将周期颠倒为 s/r 的倍数。在计算基上对输入寄存器进行测量将生成 $c=ks/r$。运行该算法数次，直到得到一个与 r 互质的 k 的值。将 c/s 约减到最小项得到满足 k/r 的 k 和 r。

因为在应用该算法之前并不知道 r 的值，因此我们当然不会知道哪次测量的结果会产生一个与 r 互质的 k 值。现在的想法是运行该算法，将 c/s 约减到最小项，得到 r 的一个候选值，因此就得到一个 N 的质因数的候选值，然后用它去除 N 以做出判断。即使我们获得了一个与 r 互质的 k 值，我们选择的 a 的某些值也会导致不能产生 N 的质因数，在这种情况下我们随机地选择一个 a 的值然后用该值运行该算法。关键在于这些步骤都是有效率的，也即是说能在多项式时间内完成的，因此我们只需重复运算多项式倍的次数，就能以任意给定的概率 $p<1$ 确定一个质因数，这个算法是一个多项式时间算法，它相对于已知的任何经典算法都能达到指数级加速的效果。

为了在几何上理解该算法函数是怎么运行的，我们考虑 $N=15$，$a=7$，$s=64$ 的情况，该情况在[Barenco，1998，160]中讨论过。在这种情况下，函数 $f(x)=a^x \bmod 15$ 为：

$$7^0 \bmod 15 = 1$$

$$7^1 \bmod 15 = 7$$

$$7^2 \bmod 15 = 4$$

$$7^3 \bmod 15 = 13$$

$$7^4 \bmod 15 = 1$$

$$\cdots$$

$$7^{63} \bmod 15 = 13$$

显然周期为 $r=4$。[①] 应用幺正变换 $U_f = a^x \bmod N$ 后，两个寄存器的态分别为：

$$\begin{aligned}&\frac{1}{8}(|0\rangle|1\rangle+|1\rangle|7\rangle+|2\rangle|4\rangle+|3\rangle|13\rangle\\&+|4\rangle|1\rangle+|5\rangle|7\rangle+|6\rangle|4\rangle+|7\rangle|13\rangle\\&\cdots\\&+|60\rangle|1\rangle+|61\rangle|7\rangle+|62\rangle|4\rangle+|63\rangle|13\rangle)\end{aligned} \tag{257}$$

这就是 $a=7$，$s=64$ 情况下的态(253)。这个态可以展开为：

$$\begin{aligned}&\frac{1}{4}(|0\rangle+|4\rangle+|8\rangle+\cdots+|60\rangle)|1\rangle\\&+\frac{1}{4}(|1\rangle+|5\rangle+|9\rangle+\cdots+|61\rangle)|7\rangle\\&+\frac{1}{4}(|2\rangle+|6\rangle+|10\rangle+\cdots+|62\rangle)|4\rangle\\&+\frac{1}{4}(|3\rangle+|7\rangle+|11\rangle+\cdots+|63\rangle)|13\rangle\end{aligned} \tag{258}$$

如果我们测量输出寄存器，我们就会(等概率地)得到输入寄存器中的四个态中的一个，具体是哪个取决于测量的结果，这四个态分别为：1，7，4 或 13：

$$\frac{1}{4}(|0\rangle+|4\rangle+|8\rangle+\cdots+|60\rangle) \tag{259}$$

$$\frac{1}{4}(|1\rangle+|5\rangle+|9\rangle+\cdots+|61\rangle) \tag{260}$$

$$\frac{1}{4}(|2\rangle+|6\rangle+|10\rangle+\cdots+|62\rangle) \tag{261}$$

$$\frac{1}{4}(|3\rangle+|7\rangle+|11\rangle+\cdots+|63\rangle) \tag{262}$$

这些态分别为伴随量是 $x_1=0, x_7=1, x_4=2, x_{13}=3$ 时的态(254)。应用离散傅里叶变换会产生：

---

① 15 的质因数 3 和 5 分别是 $a^{r/2}-1=48$ 和 15 以及 $a^{r/2}+1=50$ 和 15 的最大公因数。

$$x_1 = 0 : \frac{1}{2}(|0\rangle + |16\rangle + |32\rangle + |48\rangle)$$

$$x_7 = 1 : \frac{1}{2}(|0\rangle + i|16\rangle - |32\rangle - i|48\rangle)$$

$$x_4 = 2 : \frac{1}{2}(|0\rangle - |16\rangle + |32\rangle - |48\rangle)$$

$$x_{13} = 3 : \frac{1}{2}(|0\rangle - i|16\rangle - |32\rangle + i|48\rangle)$$

这就是态(256)。(这里 $s=64$，$r=4$；$\sqrt{\frac{s}{r}}=4$，$\frac{s}{r}-1=15$。)因此对于周期 $r=4$，输入寄存器的态最终处于由矢量 $|0\rangle$，$|16\rangle$，$|32\rangle$，$|48\rangle$ 构成的四维子空间中。

现在考虑所有满足 $f(x) = a^x \bmod 15$ 的可能的偶数周期 r，这里 a 与 15 互质。a 的可能值为 2，4，8，11，13，14，对应的周期分别为 4，2，4，2，4，2。因此我们只需要考虑 $r=2$。① 注意同周期下 a 的不同值影响输出寄存器的符号，例如对于 $a=2$，符号为 $|1\rangle$，$|2\rangle$，$|4\rangle$，$|8\rangle$，而非 $|1\rangle$，$|7\rangle$，$|4\rangle$，$|13\rangle$。

对于 $r=2$，如果我们测量输出寄存器，我们将会(等概率地)得到对应于输入寄存器中的两个态中的一个，具体是哪一个取决于测量的结果(比如说这两个态为 a 和 b)：

$$|0\rangle + |2\rangle + |4\rangle + \cdots + |62\rangle \tag{263}$$

$$|1\rangle + |3\rangle + |5\rangle + \cdots + |63\rangle \tag{264}$$

应用离散傅里叶变换之后，这两个态转换为：

$$x_a = 0 : |0\rangle + |32\rangle$$

$$x_b = 1 : |0\rangle - |32\rangle$$

在这种情况下，当 $r=2$ 时由 $|0\rangle$，$|32\rangle$ 构成的二维子空间 $\mathcal{V}_{r=2}$ 就包含于 $r=4$ 时的四维子空间 $\mathcal{V}_{r=4}$ 中。要想通过测量可靠地区分 $r=4$ 还是 $r=2$，也即区

---

① 除了 $a=14$ 之外的每个 a 的值都能产生 15 的正确的质因数。当 $a=14$ 时，该方法不可行：$r=2$，因此 $a^{\frac{r}{2}} = -1 \bmod 15$。

分输入寄存器的末态是在$\mathcal{V}_{r=4}$中还是$\mathcal{V}_{r=2}$中，只有当末态在$\mathcal{V}_{r=4}-\mathcal{V}_{r=2}$中时才可能实现，$\mathcal{V}_{r=4}$的这部分空间与$\mathcal{V}_{r=2}$正交。如果末态最终在$\mathcal{V}_{r=2}$中会怎么样呢?

肖尔算法是一个随机算法。正如上面我们提到的，该算法能产生一个周期r的候选值，因此就产生了N的一个可能的质因数，可以通过看它能否整除N就能(在多项式时间内)验证这个质因数的正确性。在计算基上测量输入寄存器得到一个结果$c=ks/r$。其中k的值通过对输出寄存器的测量等概率地得到。我们不断重复这个算法直到得到与r互质的k的值。在上述情况下将c/s约减到最小项得到满足k/r的k和r。

例如，假设我们选取$a=7$，在这种情况下$r=4$(对此我们并不知道)。与r互质的k的值为$k=1$或$k=3$(对此我们也不知道，因为k的值取决于r的值)。将c/s约减到最小项分别得到1/4和3/4，二者都可以产生正确的周期。从几何学的角度看，k的这些值相当于在计算基上测量后得到了为$|16\rangle$和$|48\rangle$的态，它们都可以被区分出是处于$\mathcal{V}_{r=4}$还是$\mathcal{V}_{r=2}$。

假设我们选取一个使$r=2$的a的值，此时得到值$c=32$。与r互质的k的唯一值为$k=1$。将c/s约减到最小项为1/2，就会产生正确的周期，因而产生N的正确的质因数。但是当$a=7$，$r=4$，$k=2$时，也会得到$c=32$，但却不会产生正确的周期，因而也不会产生N的正确的质因数。从几何上理解：$k=1$，$r=2$，与$k=2$，$r=4$对应于同一个态，$|32\rangle$。一旦我们(通过将$c/s=32/64$约减到最小项)得到候选周期$r=2$，我们就将N的质因数视为$a\pm1$与N的最大公因数，然后通过整除N来进行确证。如果$a=7$，这样计算出的质因数是不正确的。如果$a=2$，则这样计算出的质因数是正确的。

我们看到，通过尝试用一个可能的质因数来整除N的方式获得了额外的信息，说明肖尔随机算法同样意味着我们可以不确定并集的真值，就能确定两个并集中哪个是真的，也即哪个子空间包含了量子态。

6.3 加速性从何而来?

严格地说，量子计算机相对于经典计算机表现出非凡的高效率究竟依赖的是什么特性呢?在西蒙算法的案例中，对于所有经典算法，它体现出的加速性是指数级的。肖尔算法也体现出了相对于已知的任意经典算法指数级的加

速性。

对此斯蒂恩(Steane)[1998]评论道:

周期发现算法乍看就像是一个魔术把戏:我们并不清楚量子计算机是怎样产生出周期的,就像我们不知道兔子是怎么从帽子中变出来的一样……我认为最重要的性质包含于[$|\psi\rangle = \frac{1}{s}\sum_{x=0}^{s-1}|x\rangle|f(x)\rangle$]中。它们不仅包含已经提到的量子并行,还包含量子纠缠以及量子干涉。通过x与y寄存器之间的纠缠[$|\psi\rangle$],$f(x)$的每个值都与x的每个值联系在一起。当对y寄存器进行测量时,"魔术"发生了,x寄存器产生了一个特殊的态[$\frac{1}{s/r}\sum_{j=0}^{s/r-1}|x_i+jr\rangle$],而这正是纠缠所导致的,也可参见[Jozsa, 1997a]。最后的傅里叶变换可以看作是x寄存器中各种叠加态之间的干涉(对应于衍射光栅的功能)。

在经典光场中,或就此问题而言在水波中,干涉效应都能被用于实现计算目的,因此干涉自身在本质上不是量子特性。确切地说,指数级数量的干涉态以及纠缠,是经典系统中不存在的性质。

约萨[1997a]指出,经典复合系统的态空间(相空间)是其子系统态空间的笛卡尔积,而量子复合系统的态空间是其子系统态空间的张量积。对于n量子比特来说,其量子态空间为2^n维。因此一个一般态所需的信息随着n的增大指数性地增加,即使我们将对振幅的描述限制在有限的精确度之下,叠加一般来说也会具有$\mathcal{O}(2^n)$个组成部分。对于由n个二能级子系统组成的经典复合系统来说,可能态的数量随着n的增大呈指数增长,但是描述一个一般态所需的信息只是描述一个二能级子系统所需信息的n倍,也就是说,信息只是随着n的增大呈线性增加,因为复合系统的态只是乘积态。

更正式地,约萨和林登(Jozsa and Linden)[2002]表明:一个作用于纯态的量子算法能够相对于经典算法实现指数级加速的效果,仅当该量子算法涉及那种会随着输入规模的增大而无限制地增长的多粒子纠缠时。类似地,维达尔(Vidal)

[2003]也指出，经典计算机能够模拟由 n 量子比特构成的纯态的演化，只要计算资源随着 n 的增大呈线性增加，且在多粒子纠缠情况中则按指数增加。

我们在6.2节中讨论的量子计算的本质特征在于，它们可以从可能的析取项中选择一个代表函数整体性质的析取项，而不需要计算这些析取项的真值，这些真值在量子计算中属于冗余信息，但在经典计算中是必要的信息。注意量子析取被表示为输入和输出寄存器所在的张量积希尔伯特空间中的纠缠态所在的某个子空间。这类似于我们在5.2节中讨论过的涉及关键可观测量的程序，其中关键的可观测量是证明量子比特承诺定理的基础。由式(162)—(172)描述的一系列操作有效地构成了量子计算，在这些式子中，信道粒子与附属系统纠缠在一起，且随后附属系统也会被测量。

量子计算的第一阶段需要创造一个态，在这个态中函数的每个输入值都与对应的输出值相关联，这被称为“量子并行”，且有时被用作量子计算加速的资源。其思想是对于函数所有可能的值来说，量子计算类似于一个大规模的经典并行计算。这是多伊奇的观点，他认为并行计算发生在平行宇宙间。关于此观点的评论请参见[Steane，2003]，他的认识类似于这里提到的观点。当然，我们是得不到所有这些不同值的，计算基上的一次测量只能(随机)产生一个相互关联的输入输出对。我们还需要做进一步的处理，其中就包括6.2节中三个算法都会用到的最后的离散傅里叶变换。将量子算法的高效性归因于输入寄存器中由傅里叶变换产生的干涉是不正确的。傅里叶变换的作用只不过是使得在计算基上的测量能显示出表示目标析取值的哪个子空间是包含了这个态的。

那么，我们就可能想知道，为什么傅里叶变换是如此必不可少的。当然，我们可以简单地在某个不同的基上执行一次等价的测量。但要注意的是，我们将不得不进行一次计算以确定这个基。这就会产生某个问题，即如何精确地评估量子算法相对于与其竞争的经典算法的加速性。作出这种评估时，应该将什么样的相关计算步骤计入量子计算的步骤中？既然幺正变换的任意序列都等价于一个单一的幺正变换，而且进行一个幺正变换之后在一个特定基上进行测量等价于简单地在另一个不同基上执行一次测量，那么任意的量子计算就都能简化为仅仅一个步骤：在一个特殊基上进行一次测量！

当然，这个观点很难得以阐明，因为至少难度如基本计算那样的计算是不

得不需要执行以确定必需的基，但它确实表明究竟要把什么步骤计入到量子计算中是需要某些约定的。公认的约定是要求量子计算中的幺正变换应该由构成一个通用集的基础量子门(例如6.1节中讨论过的CNOT门、哈达玛门、相位门以及$\pi/8$门)来建构，并且将每个这样的门看作是一个步骤。另外，所有测量都要求在计算基上进行，而且这要被看作是额外的步骤。在多伊奇—约萨算法，西蒙算法和肖尔算法中，最后的离散傅里叶变换是必不可少的，这使得我们能在计算基上进行测量从而完成算法，而且利用基础幺正门傅里叶变换才能有效地执行正是这些算法的重要特征。我们说量子算法比经典算法快指数倍，就等于说，对于量子算法，我们按上述方法计算出的步骤数是输入规模(存储输入数据所需的量子比特数)的多项式函数，而对于经典算法，所需的步骤数将随着输入规模(存储输入数据所需的比特数)的增大而呈指数增加。

7. 量子信息视角下的量子力学基础

将经典信息理论扩展到量子领域，能否为我们在第一节中提到的由玻尔—爱因斯坦争论引出的量子力学基本问题，尤其是测量问题的解决，带来希望的曙光呢？对于这个问题，量子信息与量子计算领域的研究经常表现出将会给出一个肯定的答案，并预示了量子基础未来的发展方向。目前已经有不少学者提出了很多完全解决这些问题(总的来说，五花八门)的想法，具体参阅福斯[2001b；2002a；2002b；2001a]以及布吕克纳和泽林格(Brukner and Zeilinger)[2001；2002]。[Timpson(蒂姆森)，2004]则对布吕克纳—泽林格的贡献做出了详细的分析性批判，还应该注意的文献有哈尔(Hall)[2000]以及布吕克纳和泽林格[2000]中对他们的回应。在本节中我局限于讨论由克里夫顿(Clifton)、巴布以及霍尔沃森（CBH)[2003]提出的基于信息论约束下的量子力学表征理论的意义。

7.1 CBH 表征理论

CBH理论表明我们可以从三条基本信息论约束出发将一个用量子方式描述的物理系统的基本运动学性质推导出来，这三条约束如下：

· 在对两个物理系统中的一个实施测量之前，它们之间不存在超光速信息传输；

· 信息不可能以未知物理状态的方式得以完全传播（对于纯态，这意味着“非克隆性”）；

· 不存在无条件安全的信息通信（因此在理论上欺骗不属于该理论的讨论范围）。

更准确地说，CBH 理论是通过 C^* 代数这种一般架构使信息论的约束条件得以形式化，C^* 代数是对物理理论的数学抽象描述，具体地讲，包括有关波和粒子的所有经典理论以及各类量子理论，其中也包含量子场论（以及这些理论的交叉理论，例如与超选择定则相关的理论等）。在这个架构下，CBH 理论表明以上三条信息论约束条件限定了三个物理条件，通过它们来确定一个理论是否属于广义的量子理论。具体地说，信息论约束条件限定了以下三点：

· 属于不同物理系统的可观测量的代数形式相互对易，通常称为微观因果性，或者参见萨默斯（Summers）的定义[Summers，1990]，运动学无关性；

· 任何单个系统中可观测量的代数形式都是非对易的；

· 物理世界是非定域的，两个类空孤立系统之间可以存在纠缠态，即使它们保持相互分离。

CBH 理论还部分证明了逆推法，而将非定域性和比特承诺留作开放性问题。这些问题的解决请参见霍尔沃森[2004]，可以说，CBH 理论是一套以信息论约束术语对量子理论进行表示的理论。

这里需要注意的是，C^* 代数框架并不局限于讨论以希尔伯特空间中符合幺正动力学的系统为基础的量子力学，而是足够广义地包括了具有无限自由度系统的各种情况，比如量子场论和热力学极限下的量子统计力学（此时微系统的数量和它们占据的空间都趋于无穷，而由二者的比率定义的密度却仍为常数）。C^* 代数框架甚至还可以应用于弯曲时空中的量子场论，因此它也可以适用于对异类现象如霍金辐射（黑洞蒸发）的量子理论描述，参见[Wald（瓦尔德），1984]。斯通—冯·诺伊曼定理，保证了具有有限自由度系统间的典型交换关系存在一个幺正表示（取决于幺正等价），但是在这里是不成立的，典型交换关系存在着多种幺正等价表示。

当然，我们也可以考虑弱数学结构，但是看起来 C^* 代数体系能够满足所有迄今为止在经验中获得成功的物理理论，包括相空间理论和希尔伯特空间理论[Landsman(兰兹曼)，1998]，还包括基于流形的理论[Connes(孔耐)，1994]。关于这部分内容更多的讨论请参见霍尔沃森和巴布［2005］，还有霍尔沃森（本书第八章）、埃姆什(Emch)(本书第八章)和兰兹曼(本书第五章)的相关内容。

C^* 代数本质上是希尔伯特空间中算符代数结构的一种抽象概括。严格地讲，(单位) C^* 代数是关于包含一致性复杂数的巴拿赫(Banach) * 代数，这里对合操作 * 及其规范遵循 $\|A^*A\| = \|A\|^2$。因此希尔伯特空间 H 中所有有界算符的代数 $\mathfrak{B}(\mathcal{H})$ 是 C^* 代数，它们都具有 * 这样的伴随算符并遵循 $\|\cdot\|$ 这样的标准操作规范。

在标准量子理论中，如 3.1 节中所讨论的，$\mathfrak{B}(\mathcal{H})$ 上的一个态通过 $\mathcal{H}$ 上的一个密度算符来定义：所有通过 $\mathfrak{B}(\mathcal{H})$ 上自伴算符 A 表征的可观测量的一个期望值泛函 $\tilde{\rho}(A) = \mathrm{Tr}(\rho A)$ 。通过 ρ 来给出的 $\tilde{\rho}(A)$ 这个定义产生了一个正归一化线性泛函。因此，一般地，C^* 代数 $\mathfrak{C}$ 中的一个态被定义为任意的正归一化线性泛函 $\tilde{\rho}:\mathfrak{C}\to\mathbb{C}$ 。纯态可以由以下条件定义：如果 $\tilde{\rho} = \lambda\tilde{\rho}_1 + (1-\lambda)\tilde{\rho}_2$ ，其中 $\lambda \in (0,1)$ ，那么 $\tilde{\rho} = \tilde{\rho}_1 = \tilde{\rho}_2$ ；否则就属于混态。在下文中，我们将 $\tilde{\rho}$ 记作 ρ，但是要注意当标准量子力学的密度算符是 $\mathfrak{C} = \mathfrak{B}(\mathcal{H})$ 的一个元素时，一个 C^* 代数态 ρ 是 $\mathfrak{C}$ 上的一个正交归一化线性泛函。

由盖里森(Gleason)定理[Gleason，1957]可知，在此意义下的每一个 C^* 代数 $\mathfrak{C} = \mathfrak{B}(\mathcal{H})$ 中的 C^* 代数态都是由 $\mathfrak{B}(\mathcal{H})$ 的一个密度算符给出的。然而，由于可数可加性并不是由态的 C^* 代数概念预先假设的(因此盖里森定理并不是普遍成立的)，因此在 $\mathfrak{B}(\mathcal{H})$ 中的纯态并不能由 $\mathcal{H}$ 的向量来表示。事实上，如果 A 是一个 C^* 代数 $\mathfrak{A}$，且 $a \in \mathrm{sp}(A)$ ，那么总是存在 $\mathfrak{A}$ 中的一个纯态 ρ，为 A 赋予一个离散自由的值 a [Kadison and Ringrose(卡迪孙和林罗斯)，1997，Ex. 4. 6. 31]。既然这是成立的，即使我们只考虑希尔伯特空间中一个自伴算符连续谱上的一点，而不考虑任何对应的本征矢，那么就存在 C^* 代数意义上，$\mathfrak{B}(\mathcal{H})$ 中的一个纯态不可能是矢量态(事实上，不能表示为任何密度算符 $\mathcal{H}$)。

正如我们在 3.1 节中提到的，由幺正的相互作用以及选择性的或非选择性

的测量导致的量子系统一般演化可以由量子操作来描述，也即一个完全的正线性映射。由此可知，一个完全的正线性映射 T: $\mathfrak{C}\rightarrow\mathfrak{C}(0\leqslant T(I)\leqslant I)$ 被用来描述由可观测量的 C^* 代数表示系统的一般演化。如果 $T(I)<I$，映射或者说操作 T 被称为选择性的；如 $T(I)=I$，它是非选择性的。回想前面的内容可知，由投影算符 P 表示的一些幂等可观测量的是—否测量就是选择性操作的例子。这里，对于 C^* 代数 $\mathfrak{C}$ 中的所有 A，$T(A)=PAP$，而 ρ^T 这一转换(塌缩)态是在态 ρ 上测量 P 得到的值，并忽略了系综中对 P 不产生本征值 1 的其他元素。因此，当 $\rho(T(I))\neq 0$ 时，$\rho^T(A)=\rho(T(A))/\rho(T(I))$，否则 $\rho^T=0$。在海森堡绘景中由一个幺正算符 $U\in\mathfrak{C}$ 引导的时间演化是一个非选择操作的例子。在这里，$T(A)=UAU^{-1}$。与之相似，对一个拥有谱测度 $\{P_i\}$ 的可观测量 O 的测量，由于其不能选择结果，因此也是非选择操作的例子，这里 $T(A)=\sum_{i=1}^{n}P_iAP_i$。在一个存在于有限维希尔伯特空间系统里的标准量子理论中，即 3.1 节中的公式(67)，任何完备的线性映射都可以被视作一个更大系统中的幺正映射被限制于一个局域系统中。

C^* 代数 $\mathfrak{C}$ 的一种表征是一个任意映射 $\pi:\mathfrak{C}\rightarrow\mathfrak{B}(\mathcal{H})$ 保持 $\mathfrak{C}$ 的线性、积以及 * 结构。当 π 为一一映射时，这种表征是可靠的，这里 $\pi(\mathfrak{C})$ 是 $\mathfrak{C}$ 的同态复制。盖尔范德—奈马克(Gelfand-Naimark)定理表明在适当的希尔伯特空间 H 上，每一个 C^* 代数都具有一个固定的可靠表征作为 $\mathfrak{B}(\mathcal{H})$ 的一个标准闭 * 子代数。如上指示，在拥有无限自由度系统的案例(如量子场论)中，存在由交换关系定义的可观测量的 C^* 代数的这种等价表征。

每个经典相空间理论定义一个交换 C^* 代数。例如，由 n 个粒子组成的经典系统的可观测量——相空间 $\mathbb{R}^{6n}$ 上的连续实函数——能够被表征为 $\mathbb{R}^{6n}$ 上全部连续复函数 f 的 C^* 代数 $\mathfrak{B}(\mathbb{R}^{6n})$ 的自伴元素。相空间 $\mathbb{R}^{6n}$ 在局域上是紧致的，并且通过增加仅仅"在无限的情形下"这一点就可以变为紧致的，或者我们可以简单地考虑 $\mathbb{R}^{6n}$ 的一个有界子集(因此是紧致的)。系统的统计态由 $\mathbb{R}^{6n}$ 上的概率测量 μ 来给定，而纯态对应于粒子的最完备信息，由 $\mathbb{R}^{6n}$ 的每个具体点来给出。C^* 代数意义上的系统态 ρ 是对应于 μ 的期望泛函，由 $\rho(f)=\int_{\mathbb{R}^{6n}}fd\mu$ 定义。相反地[*Kadison and Ringrose*, 1997, *Thm.* 4.4.3]，每个交换 C^* 代数 $\mathfrak{C}$ 同构于由

$\mathfrak{C}$ 的纯态定义的局域紧致豪斯道夫空间(Hausdorff space) X 上全部连续复函数的集 $C(X)$。如果 $\mathfrak{C}$ 有一个乘法单位，那么"相空间"是紧致的。在 C 的这种"函数表征"上，同构性将 $C \in \mathfrak{C}$ 映射到函数 $\hat{C}$（C 的盖尔范德变换），它在任意 ρ 上的值就是由 ρ 赋予 C 的(离散自由)值。因此在每个抽象交换 C^* 代数"背后"，都存在一个经典相空间理论，由相空间 X 上其函数表征来定义。这个表征定理(及其逆定理)证明仅当其代数为可交换的时，就能够将一个 C^* 代数理论描述为经典的。

如上所述，CBH 将量子理论等同为非交换 C^* 代数的一个确定子类，这里运动学独立性条件由不同系统的可观测量的代数来满足，类空可分离系统的态被特征化为与纠缠相关联的非定域性的类。

为了弄清楚信息理论约束的特征和意义，我们考虑由两个子系统 A 和 B 组成的复合量子系统 AB。为了例子的简明性，我们假设系统是不可辨别的，因此它们的 C^* 代数 $\mathfrak{A}$ 和 $\mathfrak{B}$ 是同构的。分量系统 A 和 B 的可观测量可以各自表示为 $\mathfrak{A}$ 和 $\mathfrak{B}$ 的自伴元素。用 $\mathfrak{A} \vee \mathfrak{B}$ 表示由 $\mathfrak{A}$ 和 $\mathfrak{B}$ 产生的 C^* 代数。A，B 以及 AB 的物理态由它们相应(编码了全部可观测量的期望值)代数的正规范化线性泛函给出。为了保证该例子能够兼容 A 和 B 在物理上是两个不同的系统这样的情况，CBH 假定 $\mathfrak{A}$ 和 $\mathfrak{B}$ 的任意态都是相容的，即对于 $\mathfrak{A}$ 的任意态 ρ_A 和 $\mathfrak{B}$ 的任意态 ρ_B 来说，总会存在一个 $\mathfrak{A} \vee \mathfrak{B}$ 的态 ρ 使得 $\rho|_{\mathfrak{A}} = \rho_A$，$\rho|_{\mathfrak{B}} = \rho_B$。

"在测量中不存在超光速的信息转移"这句话的意义是：当爱丽丝和鲍伯进行所谓的定域测量时，爱丽丝的测量对于鲍伯测量结果的统计规律没有任何影响，反之亦然。也就是说，仅仅执行一个定域性测量，不能输送任何信息给一个物理上相异的系统，以至于对于那个相异系统来说，在测量结果的期望值方面，测量前后一切"看起来都相同"。CBH 表明这条约束是有效的：如果 $\mathfrak{A}$ 和 $\mathfrak{B}$ 在如前所述的意义上看是在物理上相异的，也就是说，$\mathfrak{A}$ 的每个元素和 $\mathfrak{B}$ 的每个元素是两两对易的，那么它们在运动学上是独立的系统[2003, Thm. 1]。更精确地说，当 $(T^*\rho)|_{\mathfrak{B}} = \rho|_{\mathfrak{B}}$ 对于 $\mathfrak{B}$ 的所有态 ρ 成立时，$\mathfrak{A} \vee \mathfrak{B}$ 上的操作 T 不能给鲍伯输送任何信息，这里 T^* 指由 T 导出的态的映射，也即 $\mathfrak{A} \vee \mathfrak{B}$ 的正线性泛函。显然地，$\mathfrak{A}$ 和 $\mathfrak{B}$ 的运动学独立性要求爱丽丝的定域性测量不能输送任何信

息给鲍伯，也即，对于$B \in \mathfrak{B}$, $T(B) = \sum_{i=1}^{n} E_i^{1/2} B E_i^{1/2} = B$，如果$T$执行的是$\mathfrak{A}$中的一个 POVM 的话。CBH 证明如果爱丽丝不能通过定域性测量输送任何信息给鲍伯的话，那么$\mathfrak{A}$和$\mathfrak{B}$就是运动学独立的。

“不传播”条件目前保证了单独的代数$\mathfrak{A}$和$\mathfrak{B}$是非交换的。回顾一下对于纯态来说，传播简化为克隆，而在初等量子力学中，一般而言克隆和传播都是不可能的(参见 3.2)。CBH 表明，对于经典系统来说，克隆和传播通常是可能的，也就是说，在可交换的情况下，存在一个通用的传播映射，该映射克隆任意一对输入的纯态，或者传播任意一对输入的混合态[Clifton *et al.*, 2003, Thm. 2]。相反地，如果任意两个态能够被(完美地)传播，那么任意两个纯态就能被克隆；如果一个C^*代数的两个纯态能被克隆，那么它们必然是正交的。因此，如果任意两个态能够被传播，那么所有纯态都是正交的，由此可以推出这个代数是可交换的。

量子力学中的干涉现象是量子可观测量(或者等价的，量子叠加态)非交换性的物理表现。因此，一个未知物理态中的信息不可能完美地传播，或者说一个未知纯态中的信息不可能被克隆或复制，这是干涉现象在信息理论中的体现。

现在，如果$\mathfrak{A}$和$\mathfrak{B}$是非交换的，但却是相互变换了的，那么就说明它们生成的C^*代数$\mathfrak{A} \vee \mathfrak{B}$中存在非定域纠缠态，参见[Landau(朗道), 1987; Summers, 1990; Bacciagaluppi(巴奇伽鲁皮), 1994]，以及[Halvorson, 2004]。更确切地说，在保证纠缠态能够存在的范围内，非定域纠缠态是存在的。因此，纠缠——薛定谔[1935, 555]将其视为“量子力学的典型特征，使物理学从整体上背离经典阵营的本质”，如我们在 4.1 中所提到的——看起来常常发生于包含可观测量的非交换代数的任意理论中。也就是说，看起来一旦我们假设“在测量中不存在超光速的信息转移”以及“不传播性”，那么能够被容许的物理理论的类型就被限定为这样的理论，其描述的物理系统表现出干涉和纠缠的性质。但是依据物理干涉来判断的话，这个结论就稍显武断了，因为纠缠态的推导取决于C^*代数工具的形式化特性。此外，我们不能确定处于纠缠态的两个系统在空间上相互分离时能够无限期地保持纠缠性，而量子纠缠的情况确实是

这样。但是这却是欺骗策略阻碍安全比特承诺时所需要的，因为爱丽丝将会保留一对纠缠系统中的一个，而将另一个发送给鲍伯，他们二者间的空间分离程度从理论上来说与协议的实施是无关的。在量子理论的信息论表述中，“复合系统的纠缠态能够被实例化，而且被非定域性地实例化，以使得随着子系统在空间中分离，复合系统的纠缠仍能保持”这样的事实，应该表现出是遵守某些信息论原理的。“零比特承诺”约束的作用就是要保证远距离纠缠的持续性，即保证一定程度上非定域纠缠态的存在——因此它提供的是非定域性，而不仅仅是“整体性”。

如5.2中所述，绝对安全的量子比特承诺是不可能的，因为通过引进额外的辅助粒子以及以恰当的方式扩大希尔伯特空间的方法，一个广义视角下的EPR欺骗策略总是能够得以实施。那就是说，对于一个由两个(分离)子系统组成的可用C^*代数表示为$\mathfrak{B}(\mathcal{H}_1)\otimes\mathfrak{B}(\mathcal{H}_2)$的量子系统，通过在用$\mathfrak{B}(\mathcal{H}_1)$表征的系统上实施一个适当的广义测量，$\mathfrak{B}(\mathcal{H}_2)$上的任意混合态就能推广为混合系统上的一个纠缠态。这就是薛定谔所谓的“远程作用”，并且他发现这样的话在物理上就是反直觉的，以至于他猜测实验证据最终应该表明这不过是理论造成的，而不是真实存在于我们的世界中(后来证明该认识是错误)[1936，451]。他指出当一个复合系统是分离的时，该系统的纠缠态应该几乎在瞬间就塌缩到一个混合态。① 复合系统的态之间应该存在着一些联系，但远程作用被认为是不可能的。

> 看起来值得我们注意的是：通过一个非常简单的假设，EPR佯谬能够被避免，即如果分离后的状态由纠缠态 $\overline{\Psi}(x,y)=\sum_k a_k g_k(x)f_k(y)$ 来描述的话，但需要在额外的条件下——复常数 a_k 之间的位相关系在分离过程的结果中消失。这意味着不只是部分，而且是整个系统都应该处于一个混合态而非纯态。我们不应该排除通过适当测量第二个系统来决定第一个系统状态的可能性，反之亦然。但是应该彻底消除实验者对他没有涉及的那个系统状态的影响。

---

① 弗里(Furry)在1936年也提出类似的可能性并拒绝接受这种可能性。

薛定谔将量子力学中伴随着非对易出现的干涉现象看作是没有问题的，因为他认为这正反映了粒子具有波动性这个事实。但是他不相信在我们所处的世界中，物理系统以纠缠态的方式表现出非定域性，因为这种状态将会允许爱丽丝操纵鲍伯的系统，使其成为与鲍伯的约化密度算符相适应的纯态的任意混合，并且他不希望实验证实这一点。当然，在 1935 年，薛定谔的猜想是否会发生还只是个实验命题。现在我们知道这个猜想是错误的。大量的实验证据，包括对贝尔不等式确凿无疑的多次违反[Aspect *et al.*，1981；Aspect *et al.*，1982]，以及量子通讯的实现[Bouwmeester *et al.*，1997；Boschi *et al.*，1998；Furasawa *et al.*，1998；Nielsen，1998]，都对此给出了证明。在我们的讨论中与薛定谔猜想相关的是：他提出了一种类量子世界的可能性，其中包含干涉但不存在非定域性纠缠。那么在信息理论基础下我们能排除这种可能性吗？

现在虽然绝对安全的比特承诺不存在于经典系统(其中可观测量的代数是对易的)中的可能性，不亚于不存在于量子系统中的可能性，但是关于在不可对易背景下比特承诺协议的不可靠性所需做的讨论却完全不同于经典可对易背景下的讨论。正如我们在 5.2 节中看到的，一个经典比特承诺协议的安全性是与计算复杂性相关的，而且不能是无条件的。

与之形成对比的是，正如薛定谔所推测的那样，如果我们在生活的世界中可观测量的代数是不对易的，而复合物理系统不能存在于非定域纠缠态上的话，当爱丽丝向鲍伯发送两个与相同的密度算符相关联的混合态中的一个以建立承诺时，她事实上是向鲍伯发送了一个证据，以证明不可兼析取的真值并不是基于特定析取的选择。(鲍伯的约化密度算符是与两个混合态模糊相关的，因此具有不可兼析取“0 或 1”的真值表。)非对易性允许不同混合态可能与相同密度算符相关联。阻止以这种方式应用混合态的模糊性以执行绝对安全比特承诺协议的事实是爱丽丝和鲍伯之间非定域性纠缠态确实存在。这就允许爱丽丝通过准备一个适当的纠缠态以替代混合态中的一个来实施欺骗，在这里鲍伯的约化密度算符是与那个混合态相同的。那么爱丽丝就可以任意地远程操纵鲍伯的系统，使其成为与两个承诺分别相关的两个混合态中的任意一个。

因此，在一个非对易理论中允许绝对安全比特承诺的要求就是物理上非定域纠缠态不存在，或者随着系统分离，纠缠自发消失。这样，我们就能将薛定

谓的讨论与安全比特承诺在我们的世界中是否可能存在相关联起来。实际上，薛定谔提出了一种可能性：我们生活在一种类量子的世界中，在其中绝对安全比特承诺是可能的！由此可见，绝对安全比特承诺的不可能性要求：对于爱丽丝和鲍伯能够遵循一些(比特承诺)协议准备的任意混合态来说，总是存在一个与其相对应的在物理上由爱丽丝和鲍伯的粒子所拥有的纠缠态，并且当粒子分离后它们仍能无限期地保持纠缠。

总结如下：CBH 定理的内容就是量子理论——一个可观测量和状态满足运动学独立性、不可对易性和非定域性的 C^* 代数理论——能够通过三条信息论约束得以表述，即不存在通过测量实现的超光速信息传输；不能(完美地)传播；以及不存在(绝对安全的)比特承诺。

7.2 作为信息理论的量子力学

让我们重新思考在第一节中提到过的爱因斯坦的观点[①]，量子力学是不完备的。本质上，爱因斯坦做出这个断言的论据是基于这样的要求，即一个完备的物理理论需要遵循一定的实在论原理(本质上就是定域性原理和可分离性原理)，这发展为这样的一种要求：在空间上分离的系统之间的统计相关性应该有一个普遍的因果解释，而这里的因果性对不同系统来说应该是一致的。爱因斯坦、波多尔斯基和罗森的论文[1935]发表大概 30 年后，约翰·贝尔[1964]表明空间上分离粒子的 EPR 纠缠态之间的统计相关性与一般因果关系意义下的经典概率分布的任何解释都是不一致的，这里的一般因果关系来自于粒子分离之前的相互关联。但是量子力学允许这种不能约化为一般因果关系的关联存在这个事实是量子理论的一个优点。正是纠缠态的这种非经典关联使得量子计算相比经典计算能指数级地提升计算速度，使得绝对安全密钥的分配成为可能，但使得绝对安全量子比特承诺成为不可能，并且它成为了一些特殊现象的基础，如量子传输以及其他依赖非经典纠缠辅助实现的通讯协议。

当爱因斯坦为不完备性所做的辩护失败后，其实在另一种意义上，量子力

① 这里的讨论部分源于论文[Bub，2004]和[Bub，2005]，但是这里的论证与前述论文有些不同之处。

学还可能会被称为是不完备的，并且也与纠缠态相关。在一次典型(理想化)的量子力学测量中，假设存在这样一种相互作用，即一个确定量子态的量子比特，其可观测量的两个可能值，即0和1，与一个宏观可观测量的两种可能状态，即p_0和p_1产生联系，形成的终态是一个纠缠态，它是态$|0\rangle|p_0\rangle$和态$|1\rangle|p_1\rangle$的一个线性叠加态，其系数可由量子比特的初量子态推出。为了戏剧性描述这个问题，薛定谔[1936]考虑了一个这样的例子，假设$|p_0\rangle$和$|p_1\rangle$分别代表密闭盒子中的一只猫活着或死了的两种状态，只有在测量完成之后观测者才能将盒子打开。以标准方式，我们将一个系统的量子态与那些具有确定性(确定的)真或假的命题以及那些没有确定真值的命题相关联起来，那么有关复合系统(微观系统+猫)的一些关联命题在纠缠态上为真，但是涉及断言猫是活的(同时量子比特的观测值为0)或者断言猫是死的(同时量子比特的观测值为1)的命题时，却没有一个决定性的真值。此外，如果我们假设量子命题形成一个同构于复合系统的希尔伯特空间的子空间——量子态和可观测量的表征空间——的代数结构的话，那么很容易就会导致一个形式矛盾，就是对应于纠缠态的关联命题是真的这个假设与猫只能是确定的生或死这一个事实之间的矛盾。① 薛定谔认为这种事情是荒谬的：假设量子力学要求我们接受在观测者打开盒子看之前，猫(一个宏观系统)保持既是活的也是死的(不具有确定的宏观特性)这种状态，在这种情况下，纠缠态非线性和随机地“塌缩”到一个可以表示猫的一个确定态和微观系统的一个确定态有关的结果上，其概率由微观系统的初量子态的系数决定。爱因斯坦[1967，39]在给薛定谔的一封信中对其观点表示赞同，并评论道：“如果这是真的话，那么物理学可能就会成为只有售货员和工程师感兴趣的学科，全部事情都只不过是令人沮丧的拙劣工作。”

这就是标准的量子力学“测量问题”。确实，问题的形式是高度理想化的，但是基础问题正是以这样的方式提出的(量子力学用这种方式通过纠缠态来表征关联)，并且即使降低理想化程度这个基础问题也不会完全消失(尽管由于考虑实验仪器以及整个实验环境的宏观属性，这个问题多少有些改变)。请参

① 这遵循下面会提到的巴布—克里夫顿定理。由一致性和EPR态定义的确定性量子命题的子格是最大化的：增加任何命题都会导致矛盾。

见迪克森在本书第四章的观点，以及［Bub，1997］中的进一步讨论。在以后的讨论中我将这个问题——理论的薛定谔不完备性——表述为薛定谔问题。相对于前面提到的关于概率的问题，我们认为这是一个关于真实性(或者属性的实例化)的问题。

在概率问题被公式化之前，让我们思考一下力学从经典过渡到量子的历程。1925年海森堡的矩阵力学是量子力学的最初形式，产生于“旧量子论”即为了满足普朗克量子假设而对经典力学做出的那些修补之后。本质上讲，海森堡是通过用一些不对易的数学表征量——矩阵——来代替一些经典力学如位置与动量等变量的方式来修正经典力学的运动学的。在这之后不久，薛定谔提出了量子力学的波动力学表示法，并证明这两种理论在形式上是等价的。用“波粒二象性”的视角来理解从经典力学到量子力学的转变是更为普遍的一种方式，这种观点认为量子系统如电子，与经典系统如岩石不同，它会在某些特定的环境下表现出波的特性，而在另一些特定的环境中则表现出粒子的特性。这种图像与其说阐明了理论，不如说它让理论更加令人费解。如果我们考虑海森堡的进一步研究的意义的话，那么在概念方面我们就会看得更加清晰，并且也使我们能从更广义的意义上理解物质及其属性。

海森堡将经典力学中的力学变量如位置、动量、角动量及能量等的交换代数替换为非交换代数。其中一些力学变量只能取值为0或1，并且能与其属性相对应。例如，我们可以用一个粒子的某个力学变量的值为1来代表其确实位于空间中某个区域的这种属性，相反则设其值为0。像位置这个力学量就对应于这个二值力学变量集或者说物理属性集。在粒子的位置这个案例中，对于全空间R，我们可以表征与处于区间R中的某个特定区域中的粒子相关联的某种属性。对于全空间R，如果我们知道粒子是否位于那个特定区域的话，那么我们就知道了关于粒子的位置特性，反之则不知道。一个经典系统的双值力学变量或者说属性就形成了一个布尔代数体系，它是力学变量交换代数的一个子代数。

将力学变量的交换代数替换为非交换代数就等价于将一个双值力学变量或属性的布尔代数替换为非布尔代数。物理系统表征的经典模式与量子力学模式的本质不同就在于经典系统的属性可以被表征为布尔代数或者说布尔格的形

式。每一个布尔格都与一个集合的子集的格同构的。[①] 我们说一个经典系统的属性形成了一个布尔格就意味着在说这些属性可以由一个集合(经典力学的相空间或者说态空间)的子集来表示。我们说一个物理系统有确定的属性就意味着将系统与表征空间中的一个特定集合联系了起来，在这个表征空间中，空间的全部元素——集合中的所有点——表示系统所有可能的状态。一个状态挑选一系列的集合表示系统处于该态时具有的属性，表征这个态的那些空间中的点属于这些集合。描述态空间中状态运动的定律就构成了经典力学的动力学。根据状态随时间的变化，由状态选择的属性的集合也发生变化。更详细的阐述请参见［Hughes(休斯)，1995］和[Bub，1997]。

因此，从经典力学到量子力学的转变意味着将表示属性的方式如布尔格表示(即一个集合的子集表示)用一类适当的非布尔格表示代替。狄拉克和冯·诺伊曼将薛定谔的等价证明进一步发展为一套表象理论，将量子系统的属性表征为线性复矢量空间即希尔伯特空间的子空间。上述问题中的非布尔格就成了这个空间的子空间对应的格。不同于用一个集合的子集表征经典系统的属性，量子力学将其属性表示为一个线性空间的子空间，该线性空间可以是线或平面，也可以是超平面，例如一个投影几何空间。从代数学上讲，在从经典力学到量子力学的转变过程中，这是一个重要的结构变化——尽管它还有更深刻的意义：显而易见的事实是，量子系统的态空间是复数而非实数的希尔伯特空间，它所反映的物理现象是与包含不同相关相位的叠加态的概率相关的。

我们可以等价地通过探讨命题来代替探讨属性。(我们认为一个给定的属性是被实现了的，当且仅当其对应的命题为真。)在一个布尔命题结构中，存在结构上的二值同态，对应于命题的真值赋予。事实上，相空间的每一个

---

① 格是一个部分有序集合，其中的任意一对元素都有一个与顺序相关的最大的下边界(下确界)和最小的上边界(上确界)，一个最小元素(记为0)和一个最大元素(记为1)。布尔格是一个有补分配格，也就是说每一个元素都有一个补(格类似于集合理论中的补集)，且分配律在下确界和上确界和范围内成立。由一个集合 X 的子集来表示的布尔格中的部分排序对应于在集合中由包含来定义的部分排序，因此下确界对应于集合的交集，上确界对应于集合的并集，0 对应于空集，1 对应于全集 x。由代数加(+)和积(·)操作定义的一个布尔代数，等价于由部分有序结构定义的布尔格。

点——代表一个经典状态——定义一个表征命题的子集上的真值赋予：包含点的每个子集表示一个真命题或者由系统实现了的一个属性，不包含点的每个子集表示一个假命题或者未由系统实现了的一个属性。因此，一个经典状态对应于一个完备的命题真值赋予，或者说系统属性的一个最大化相容“列表”，而所有可能的状态对应于全部可能的最大化相容“列表”。

在这样的经典属性结构上，我们可以引入概率作为对表征属性的子集的测量。既然每个相空间的点定义一个真值赋予，那么一个属性的概率就是针对将值1(真)赋予一个属性的真值赋予集进行的测量——实际上，我们(在测量理论的意义上)“计算”的是用特征实现(或者对应命题为真)来描述的状态的相对数量，而这个数量表示属性的概率。因此，将一个属性的概率解释为对我们关于这个属性是否被实现了的无知程度的测量，就显得有意义了。被表征为相空间上的点的经典状态的概率分布有时被称为“混合态”，以区别于“纯态”，即对应于相空间上的点态。

用一个希尔伯特空间的子空间表征的量子属性结构存在着一个问题，因为这种结构中不存在二值同态(二维希尔伯特空间上的一些特殊案例除外)。如果我们将希尔伯特空间的子空间结构严格视作量子力学的结构特征，如同经典力学中的布尔属性结构或命题结构那样的话，在希尔伯特空间的子空间的格上不存在二值同态就意味着有关量子系统的属性作为一个整体不能被区分为两个集合：被系统实现了的属性和未被系统实现的属性，也就是说整体命题不能被划分为真命题和假命题。(当然，可能还有其他方式将命题与希尔伯特空间的特性进行关联，而且也还有其他方式来分配真值，包括多值真值分配和语境真值分配。归根结底，这里所涉及的问题是，我们将什么样的改变看作是从经典力学到量子力学的过程中显著的结构性变化。)

这就导致了这种情况的出现：在标准解释中，由希尔伯特空间中的一个一维子空间——子空间结构中的最小元素——表示的一个量子纯态为量子命题定义了一种真值赋予方式，就像一个经典纯态为经典命题定义了一种真值赋予方式一样。具体地说，就是在标准解释中，一个量子纯态决定了由包含该纯态的子空间表示的命题为真，而由那些与该纯态正交的子空间表示的命题则为假。(注意这里的正交类似于子空间结构中的补集或者说否定，集合论意义下的一

个子空间的补集一般不是一个子空间。)

然而，这两种情况有一个很重要的区别。在经典态的情况下，由相空间子集表示的每一个可能属性都可以选择要么被系统实现，要么不被实现；等价地，命题不为真就为假。但是在量子态的情况下，由希尔伯特空间的子空间表示的属性不能被划分到两个相互独立却又能全面覆盖的两个集合中：一些命题不能被赋值。只有那些由包含该量子态的子空间表示的命题能够被赋值为“真”，且只有由正交于该量子态的子空间表示的命题能够被赋值为“假”。这意味着由那些与表示该量子态的射线呈非平行或非正交夹角的子空间所表示的命题不能为其对应的量子态赋予真值，而其对应的属性必须被视为非决定性的或不确定的：根据这种理论，这些属性是否被实现就成为不可知的事实。

这导致的结果是只有一种方法能为量子属性赋予(广义)概率值，那就是说，在量子属性的非布尔格的布尔子格上概率测量能满足一般柯尔莫哥洛夫(Kolmogorov)公理的权重。这就是盖里森定理的内容[Gleason, 1957]。对于一个量子态ρ，它的一个属性p由一个投影算符P来表示，可以被赋予一个概率$\mathrm{Tr}(\rho P)$。如果ρ是一个纯态$\rho = |\psi\rangle\langle\psi|$，那么$p$的概率则为$|\langle\psi_p|\psi\rangle|^2$，这里$|\psi_p\rangle$是$|\psi\rangle$在子空间$P$上的正交投影，也就是说$p$的概率就是射线$|\psi\rangle$与子空间$P$的夹角的余弦的平方。这意味着，由包含该量子态的子空间表示的属性的概率为1；由正交于该量子态的子空间表示的属性的概率为0；而其余的属性，即由那些与量子态呈非平行或非正交夹角的子空间表示的属性，其概率介于0和1之间。因此量子概率并不是由对真值赋予表的测量而得到的，也不能以常见的无知程度解释来理解。

那么现在问题就变为：这个“角概率”(或者说“角权重”更好)的意义是什么？正统的回答是用量子态来为一个系统的某个属性赋予概率，可以被理解为在测量过程中找到这个属性的概率是为了确定我们能否测得这个属性。再多做一些思考就会发现这种方案是很有问题的。当系统由一个量子态表征，而这个量子态的一个属性的概率为1/2时，我们认为这个属性就是非决定论的。这时物理学家就会说，认为这个属性是处于该量子态上的系统这种做法就成为“无意义的”了。但有时候我们还是要设计一些实验以确定这个属性是否被系统实现了。而在这些实验的测量中，得到“肯定的”答案和“否定的”答案的概率分

别是 1/2。很明显，量子力学中的一个测量过程并不是直接意义上确定一个属性是否被实现的过程。无法解释的是，在一个给定的量子态下，一个既没有被系统实现也没有不被系统实现的属性，量子测量会使它以一定的概率呈现，或者以一定的概率不呈现；或者等价地说，通过一个适当的测量过程，一个既不是真也不是假的命题就会以一定的概率为真，或者以一定的概率为假。

概率问题(对照于真值问题即薛定谔问题)是指当我们在理解“角权重”时，在哪些意义(相关频率？倾向性？主观贝叶斯博弈概率？)下，我们将其解释为概率的问题，而那些意义是不能约化为对量子力学的一种纯工具主义(工具主义认为理论就是一种能够异常精确地提供预言的工具)解释的。(关于此方面内容，请回顾爱因斯坦关于“在哥本哈根解释下物理学会成为只有售货员和工程师感兴趣的学科”的评述。)这个问题之所以被提出是因为只有以独一无二的方式概率才能被引入到量子力学中，还因为在哥本哈根解释中，测量与观测具有十分神秘的定义。

在经典理论中，我们通过测量来发现我们所未知的事物，但是在理论上测量不会改变事物本身(甚至即使造成被测事物的改变，也只不过是事物从一种状态到另一种状态的变化或者扰动，这种变化或扰动也是建立在经典理论自身基础上的)。然而在量子力学中，看上去测量为事物带来了一些不确定的东西，而不仅仅是之前的未知的东西，就是说，一个既非真也非假的命题在测量过程中变成了真的，并且根据量子理论，即使我们对客体、变化以及干涉做出最深刻的假设，这种不确定性发生的方式仍然是令人费解的。

现在，我们已经知道如何解决薛定谔问题了，也就是说，我们已经知道的解决该问题而对量子力学做出修正的所有可能的方式。这个问题之所以被提出，是由于量子理论的线性动力学的缘故，它导致了作为交叉测量结果的某一纠缠态的产生，它还将纠缠态解释为一系列事件的一个态，其中一些事件命题是确定为真的，一些是为假的，其余的是不确定的。或者我们通过某种方式改变线性动力学，或者我们保持其不变，并认为真值及不确定性与量子态之间的关系是非正交性的。两种选择都已经通过多种方式被非常详尽地研究过了，我们已经理解了薛定谔问题的解空间，以及选取其中一个特殊解的重要性。

“塌缩”理论可以说是由吉拉迪(Ghirardi)、瑞米尼(Rimini)和韦伯(Weber)(GRW)发展起来并由佩尔(Pearle)加以扩展的理论[Ghirardi, 2002]，它通过修改量子理论的线性动力学解决了上述问题。详细介绍请参见本书第四章迪克森的论文。在已修改过的理论中，存在一定的、非常小的概率，使得一个粒子的波函数在与特定宽度的峰值高斯数相乘后，会发生自发的“塌缩”。对于包含很多粒子的一个宏观系统来说，这个概率会在非常短的时间间隔内接近于1。实际上，这种塌缩的解决方案是通过增加不可控噪声来修改标准量子力学的线性动力学的。当达到介观和宏观层次时，修改后的动力学随机项就会变得非常重要，它们会趋向于使空间中的波函数局域化。因此，如果测量过程中的相互作用导致波函数的塌缩几乎在瞬间发生，并且被测可观测量与一个局域的宏观指针量的位置也随之塌缩的话，那么这种涉及设备(或者猫)的宏观部分的测量操作就能够与基本量子程序区分开来。

“非塌缩”解决方案则受到了一些肯定“不可行”定理的约束，这些定理限制了在非常普遍的有关可观测量代数的假设条件下对属性或者可观测量的值进行的真值赋予[Kochen and Specker, 1967]，或者限制了在与量子概率相关的真值分布的假设条件下对可观测量的值进行的真值赋予[Bell, 1964]。由巴布和克里夫顿[1996]提出的一个定理表明，如果我们假定有确定值的可观测量的集合存在一个确定的结构(在本质上允许将量子概率约化为对一些分布的经典测量，这些分布由值或属性不同的可能集合来定义)，并且一个测量过程中的指针可观测量属于这个有确定值的可观测量的集合，那么这些理论的类——所谓的“模态解释”——就是极其特殊的。更精确地，与任意单个可观测量 R 相关的子格是一个布尔格$\mathcal{B}$，并且一个量子态 $|\psi\rangle$ 就定义一个$\mathcal{B}$上的经典概率测量，是发生在以下的意义上，即由 $|\psi\rangle$ 赋给$\mathcal{B}$上元素的所有单个或联合概率被约化为在柯尔莫哥洛夫概率空间上进行的测量，这里的柯尔莫哥洛夫概率空间是由$\mathcal{B}$上的二值同态的相空间 X 来定义的。巴布—克里夫顿定理描述了与可观测量 R 以及给定的量子态 $|\psi\rangle$ 相关的任意布尔子格的最大格延伸$\mathcal{L}$，基于$\mathcal{L}$是一个正交格①的假设，以及基于保持 R 和 $|\psi\rangle$ 的格同构的不变量(不变量的概率由 $|\psi\rangle$ 赋给$\mathcal{L}$上元素)，能够被简单地约化为在柯尔莫哥洛夫概率空间上

① 也就是说，对于$\mathcal{L}$的每个元素都存在一个正交补。

进行的测量的假设，这里的柯尔莫哥洛夫概率空间是由$\mathcal{L}$上的二值同态的相空间Y来定义的。在此意义下，该定理描述了一个量子命题结构的经典性质的极限。结果就是不同的模态解释是分别与不同的“确定子格”$\mathcal{L}$相关联的，也就是跟不同的“首选可观测量”R相关联的。对于标准量子力学，R是单位元，“确定子格”$\mathcal{L}$则由全部量子属性构成，这些属性由态$|\psi\rangle$所在的子空间（$|\psi\rangle$赋予属性的概率为1）和与$|\psi\rangle$正交的子空间（$|\psi\rangle$赋予属性的概率为0）来表征。玻姆（Bohm）的隐变量理论可以被理解为一种模态解释，其中首选可观测量为位形空间中的位置。有关玻姆理论的更多内容请参见本书第四章以及[Goldstein（戈德斯坦），2001]。

关于薛定谔测量问题的另一种“非塌缩”解决方案为埃弗雷特（Everett）解释[Everett，1957]。更多相关内容请参见本书第四章。文献中有很多种类的埃弗雷特解释，最常见的一种可以表述为：一次测量的所有可能结果在某种指称意义上都可以被视为是真实的，它们分别与非定域纠缠态中的不同项相关（这取决于希尔伯特空间中的首选基是如何选取的），根据视角的不同，这些项又被理解为与不同的世界或者不同的想法相关联。埃弗雷特解释最精确的构想应该说是桑德斯—华莱士（Saunders-Wallace）的版本[Saunders，1998；Wallace，2003]。这里的首选基通过消相干（下面会对此进行详述）来选取，而概率则被理解为贝叶斯意义上的合理推测概率，通过最初源于多伊奇的判别—推理论证得来[Deutsch，1999]。

总结如下：薛定谔测量问题的任何解决方案要么涉及改变理论的线性动力学（“塌缩”理论），要么就在原有特性之外再额外地选取一些可观测量，以使得每一个量子态都有一个确定的值，并修改标准理论中的某些说法，即关于每个量子态的一些属性都具有真、假以及不确定这三种状态（模态解释，“非塌缩”隐变量理论），以便在测量相互作用（这里的测量会将微观被测量的某个可能值与宏观的某个指针状态相联系起来）完成之后，宏观的指针状态及其所对应的微观被测量最终有确定的值。还有一种选择（埃弗雷特解释），就是我们可以通过其他的一些量子力学解释来假设每个测量结果在某种表示的意义上（这涉及不同的世界、不同的想法或者纠缠态的不同分支等认识）都是确定的。

如果我们掌握了这些方案在可观的细节上是怎么样的，那么我们就理解了

他们是怎么修改量子力学的了。之所以我们说研究这些方案是有实际意义的课题，是因为在此过程中我们会更加了解量子力学，或许进一步研究这些方案的基础概念我们还会学到更多。但是在这里需要指出的一点是，所有那些(以不同方式通过引入额外的反而会混乱我们对现象的理解的结构性质)用来解决关于量子力学测量造成的“真值问题”的方案，都会涉及与纠缠相关的信息理论的应用，例如量子传输，与经典密码协议相关的某种量子密码协议的可能性或不可能性，与经典算法相关的量子算法的指数加速性等。

让我们重新返回来考虑关于量子力学解释的玻尔—爱因斯坦争论。有人或许会说爱因斯坦(以及说薛定谔)和玻尔之间的分歧在于他们对范·弗拉森(Van Fraassen)[1991, 4]所谓的“基本问题：世界怎么可能像量子力学描述的那样运行?”的回答不同。这其实是一种误解。爱因斯坦的回答其实是：这个世界不可能像量子力学描述的那样运行，除非量子理论是一个不完整的理论(因此理论的完善——也许是爱因斯坦提出的广受欢迎的统一场理论——大概能回答这个问题)。但是玻尔的互补性解释并不打算回答这个问题。更确切地说，互补性解释应该被理解为对另一个问题的回答：为什么世界必须按量子力学描述的那样运行?

为了明确这两个问题的区别，我们可以考虑爱因斯坦针对“原理性”理论与“建构性”理论所做的区分。爱因斯坦在发表于伦敦《时代》杂志 1919 年 11 月 28 日[1919]上的关于狭义和广义相对论的意义的文章中提出了这二者之间的区别：

> 我们可以对物理学中的各种各样的定义进行一下区分。它们中的大多数是建构性的。它们试图通过一些相对简单的形式结构来建构一幅关于更加复杂现象的图景。以这种方式，气体的运动理论设法将机械过程、热过程以及扩散过程约化为分子运动——也即从分子运动的假设来构建气体运动学。当我们说我们能够成功地理解一组自然过程时，总是意味着我们找到了一个建构性的理论来说明该过程。
>
> 与这类非常重要的理论相并列地存在着的是另一类理论，我称之为“原理性理论”。这类理论应用分析方法，而不是生成方式。组成

其基础和出发点的不是假设性的建构元素，而是经验性的客观发现，自然过程的一般特性，以及可表述为数学公理化标准的原理(理论中包含的自然过程及理论描述必须满足这些原理)。这样一来，分析性的热力学所试图做的就是从永动机不存在这样的普遍性事实出发去演绎出独立事件需要满足的定理。

爱因斯坦的观点是相对论应该被理解为一个原理性理论。在他的“自传式笔记”[1949，pp. 51—52]中他又重申了这个论题，他强调一开始他是试图找到一个建构性的理论来解释关于物质和能量的已知性质，但最终他确信这个问题的解决方案存在于一个原理性定理中，即在所有惯性参考系下，真空中的光速应该保持不变，并且对所有(机械的及电磁的)物理定律来说，惯性系都是等价的：

早在1900年之后的那段时间，也就是普朗克完成那开拓性的工作之后，这类想法就使我清晰地认识到，无论是机械的还是电磁的理论都不是绝对正确的(除了在极限情况下)。不久之后，我就对在已知事实的基础上通过建构性的方式来发现真实规律的这种努力丧失了信心。随着时间的推移，对此我越来越感到更强烈的失望，我越来越坚定地认为只有找到一个统一的形式化原理才能引导我们找到确定的结果。对此我找到的近在眼前的例子就是热力学。其中的普遍性原理可以表述为这个定理：自然规律决定了(第一类或第二类)永动机是不可能实现的。那么，这个普遍性原理是怎么被发现的呢？

在之后的叙述中[1949，57]他补充道：

狭义相对论的一般性原理包含于以下假设中：在洛伦兹(Lorentz)变换(从一个惯性系变换到其他任意惯性系)下，物理定律保持不变。这个原理是对所有自然规律的限制，就如同在热力学中永动机不存在的原理性限制一样。

爱因斯坦认为，在物理学中，我们需要区分这两类完全不同的理论。其中一类是将一个由相对复杂的现象构成的领域还原为由一些相对简单的元素组成的性质，比如在动力学理论中，我们将气体在机械和热方面的行为归纳到分子运动的规律上，这些基础性的元素组合起来就形成了建构性的理论。另一类理论则可以表述为“不可行”原理，它们给定了物理过程或物理事件的约束条件，比如热力学中的(“永动机不能实现”)。要了解更多关于爱因斯坦在此方面工作的启发性意义，请参阅马丁·克莱因(Martin Klein)[1967]对此的讨论。

狭义相对论是一个原理性的理论，通过两个基本原理来规范：不同惯性系对所有物理规律(既包括电磁现象也包括力学定律)都是等价的；在所有惯性系下真空中的光速是不变的。这两个原理与牛顿时空观下的几何学是不相容的，因为后者的惯性系变换是基于伽利略变换的。对该几何学的修订产生了闵可夫斯基(Minkowski)几何学，它的惯性系变换是基于洛伦兹变换的。狭义相对论表明在遵照洛伦兹变换的不同惯性系中，物理学定律是不变的。爱因斯坦将狭义相对论描述为“对自然定律的一个约束原理，就如同热力学中的永动机不存在这样的约束性原理一样”。(在广义相对论中，容许变换群包含时空流形的所有不同变换。)与之不同的是，洛伦兹定理[Lorentz，1909]却是一个建构性理论，它从以太的电磁学性质中推导出洛伦兹变换，并假设分子间的力通过以太来传输。

那么这个问题：

世界怎么可能会像量子理论描述的那样呢?

是由于我们将量子力学看作是一个建构性理论这样的难题引起的，因此适当的回答就是为解决这个难题而提出一些关于理论的建构性修补，或者证明量子力学这些在现象层面(干涉和纠缠现象)令人困惑的性质最终会从一个物理上没有问题的建构性理论中得到解答。

而另外一个问题：

世界为什么必须像量子理论描述的那样呢?

却并不从物理方法论和力学定律的方面去寻求一种对量子现象“由下至上”的解释。相反，这个问题将量子力学看作一个原理性理论，从现象出现的可能性中存在的可操作性约束视角出发，涉及的是理论“由上至下”的发展。在量子力学中，这种相关现象关系到了信息。

安德鲁·斯蒂恩在一篇对“量子计算”的回顾性论文[1998，119]中重点关注了以上两个问题所描绘的图景之间的变换：

> 从历史上看，大部分基础物理关心的是发现自然界中的基本粒子，并提出描述它们运动和相互作用的方程。现在看起来另外一类不同的任务或许同样重要：那就是发现自然界以什么样的方式允许或者阻止信息的表达和操作，而不是关注粒子的运动。

斯蒂恩以下面的建议作为此篇论文的结尾[1998，171]：

> 在最后，我想提出一个更加宽泛的理论性工作：总结出一系列类似于能量和动量守恒的原理，但要将之应用于信息，并且大量关于量子力学的问题应该要能从此类原理中推导出来。对这类观点的两个检验可能会是：EPR—贝尔关系是否会由于这样的原理而变得清晰，以及它们能否使得我们在应用诸如“测量”和“知识”这类术语时可以呈现出其显而易见的意义。

一个相似的正确变换隐含在惠勒“为什么是量子力学”这个提问中，这个问题是惠勒认为的“真正的大问题”［Wheeler，1998］中的一个。斯蒂恩的建议是将该问题的回答建立在如何理解量子力学来源于信息理论原理这个基础问题之上。沿着这种思路，吉尔斯·布莱森德和克瑞斯·福斯提出了一个更具体的建议。正如5.2节中提出的，布莱森德和福斯推测量子力学可能来源于信息理论约束，这种约束由确定的简单密码协议形成。具体地说，就是无条件安全密匙分配的可能性，以及无条件安全比特承诺的不可能性[Brassard，2000；Fuchs， 1997；Fuchs，2000；Fuchs and Jacobs，2002]。

由布莱森德—福斯推测所激发的 CBH 理论表明量子力学能够被看作是爱因斯坦意义上的原理性理论，它的原理就是信息理论约束。因此我们必须回答这个问题：世界为什么必须像量子理论描述的那样呢？干涉和非定域纠缠现象本来不该发生在这样的世界中，因为在此世界中信息的接受、通讯和处理过程中都有确定性约束。

作为对照，我们考虑作为现代物理学另一支柱（相对论）的情况。一个相对论性的理论是指一个具有确定对称性和不变性的理论，它由一组时空变换的术语来定义。按照爱因斯坦将狭义相对论看作一个原理性理论的构想，我们将这种不变性理解为某个事实的结果，这个事实就是在我们生活的世界中自然过程受限于以下的确定性约束：如赫尔曼·邦尼迪（Hermann Bondi）[1980]所表述的，粗略地讲，“光速不可能被超越”，以及“速度并不重要”（无论对电磁现象还是力学现象）。回顾爱因斯坦对狭义相对论的理解：“对自然定律的一个约束原理，就如同热力学中的永动机不存在这样的约束性原理一样。”没有爱因斯坦的分析，闵可夫斯基时空下的变换就是一个相当令人困惑的关于相对论性运动学和洛伦兹变换的运算法则而已，而且与牛顿时空下的运动学不相协调。爱因斯坦的分析为我们提供了一个将时空结构看作是闵可夫斯基时空的理由：我们将其视为是由狭义相对论两个原理的一致性所要求的。

在量子理论中，可观测量和状态都具有特定的典型代数结构。然而不同于相对论，量子力学是以一种方法或运算规则的方式产生的，它被用来计算由宏观测量仪器测量可观测量所得到的期待值。一个关于交换 C^* 代数的理论有一个相空间表征——并不一定是经典力学的相空间，但它却是这样的一个理论，在其中 C^* 代数的可观测量被替换为“可存在量（beables）”，此为贝尔的术语，参见[Bell，1987]，并且 C^* 代数的状态也被替换为可存在量的态以表示属性的完备目录（幂等量）。在这个例子中，将理论进行扩展以包含作为 C^* 代数统计源的测量仪器是可能的，这会使得它们不再是“黑箱”，而是由系统构成的，这个系统是由相空间理论中的属性和状态来表征的。也就是说，C^* 代数理论能够被一个“去观测者”的理论替代，该理论是有关于现象底层的物理过程的，借用泡利的术语[Born，1971，218]，就是该理论包含对测量仪器运行过程的说明。

注意这是依赖于表征定理的。在不可交换的例子中，我们只能保证关于 C^* 代数的一个希尔伯特空间表征的存在，但是对现象的“去观测者”描述是否可能，仍然是一个开放性问题。

试图解决薛定谔问题——真值问题——就等同于将量子力学视为一个失败的、或者说不完备的需要进行建构性修补的建构性理论。实际上，问题在于如何以一种理论架构的形式对量子信息做出解释——有关干涉和非定域性纠缠的令人困惑的性质的解释——并且在这种架构中从基本意义上看只有经典信息才是有意义的。如果我们将量子力学看作是关于信息的一个原理性理论，那么核心的基本问题就是概率问题。在这种图景下，问题就是在一个由信息理论约束特征化的量子世界里，如何解释经典信息的出现。

我们也许会抱怨将量子力学视为原理性理论就等同于简单地假设最终会有一个建构性理论比如 GRW 理论或者玻姆理论来解释量子力学。这将会等同于不承认原理性理论具有解释力这个观点。以这种理解来看，玻姆的建构性理论与量子力学的关系就如同洛伦兹变换与狭义相对论的关系一样。库欣(Cushing)[1998, 204]引用洛伦兹的话提出一个相似的抱怨“爱因斯坦简单地假设了我们推出的结果”。可参见 1916 年编著的《电子理论》(*The Theory of Electrons*)中的结论。

> 这里，我不会论及爱因斯坦通过应用这种(相对论)原理得到的众多非常有趣的结果。他的结果涉及电磁学和光学现象……与我们在前几页所得到的结果在主要方面是相一致的，主要的区别在于爱因斯坦从电磁领域的基础方程出发简单地假设了这些结果，但其中仍存在一些难题并且其假设也并不是完全令人满意的。通过这样做，他使我们在看待诸如迈克尔逊、瑞利以及布莱斯等人实验的否定结果时，并不认为是一些相反结果的偶然发生，而看作是一个更普遍更基本原理的体现，对此他一定会感到非常自豪。
>
> 是的，我认为目前也存在其他的一些断言赞同我上面介绍的理论所应用的模式。但我并不认同这种说法，而是将以太——可能是电磁场的能量和波动传播的载体——看作自身就具有一定程度的实体性，无论它看上去

与普通物质有多么的不同。在这种认识下，我们自然不会一开始就去假设无论物体是否穿越以太都不会造成什么不同，也不会利用相对于以太有固定位置的尺和钟去测量距离和时间间隔。

我们要注意洛伦兹的理论是受狭义相对论约束的，那就意味着以太作为电磁现象中的多余结构，在原理上必然是无法探测到的。因此这样的理论不可能包含比狭义相对论更多的经验内容。库欣[1998，193]还引用麦克斯韦(Maxwell)的话提出质疑，是否“承认我们现在还察觉不到的一种媒介的存在，并不比断言物体在一个不存在的空间中运动在哲学上更明智”。确实是这样的，但是如果我们不得不将其视为物理定律，并在原理上承认它，情况就会有所不同，假如我们生活在一切事件都由狭义相对论的两条原理约束的世界中，那么这种媒介就必然是不能被探测到的。

如果愿意的话，我们可以对量子现象提出一个建构性的解释，但是这样的一种解释，如果还是受信息论原理约束的话，那么它不会包含超越量子力学的经验内容。换一种方式来说，那就是如果我们试图提出一种解决薛定谔问题的方案，并且要比量子力学包含更多的经验内容的话，那么就必须违背一条或多条信息论约束。因此，关于量子现象的玻姆理论就类似于关于电磁场的以太理论一样。就像以太理论试图通过提出以太这种不同于一般力学系统的这一类特殊力学系统的方式来解释场的行为那样，玻姆的理论试图引入一个独特的场(量子势场或导向场)——它不同于一般的物理场——来解释量子现象。

这里关键的差别就是建构性理论和原理性理论之间的差别，从物理方法论和力学定理出发(由下至上)就形成了建构性理论，而从现象层面的可操作性约束出发(由上至下)就形成了原理性理论。建构性理论引入了可存在量和可存在量的态的代数，原理性理论则引入可观测量和可观测量的态的代数，它们从本质上来说就是概率性的。

看起来如果复制是不可能的话，那么可观测量的代数将会非平凡地和可存在量的代数相异。如果关于某个确定领域现象的一个建构性理论允许力学相互作用存在的话，也就是说由测量仪器指定的一个系统的某个可存在量，如果能

与由测量系统指定的另一个系统中的某个可存在量发生联系，并且不影响测量系统中其他可存在量的值的话，那么对于这个建构性理论，我们就可以通过力学相互作用，来确定被讨论的可存在量的值(在这种意义上，一个系统中的某个可存在量的值是被记录在了另一个系统的某个可存在量的值中的)。如果这是可能的话，那么通过将一系列测量相互作用联系起来，同时测量一个系统中任意数量可存在量的值就也是可能的，因此理论上确定任意可存在量的态的值就是可能的。因此，如果我们不能制造这样的设备的话，那么在这种意义下测量也一定是不可能的。由此得出结论，建构性理论中的“测量”绝不仅仅是在不干扰其他可存在量的情况下，对一个系统中某个可存在量的确定，并且在这样的一个理论中解答可观测量是什么这个问题，需要一种非平凡的分析。

玻姆也确实在他 1952 年发表的关于隐变量的包括两部分的论文的第二部分中给出了上述的那种分析，而杜瑞等人[Dürr *et al.*，2003]则给出了一个更细致和精密的分析，并就玻姆理论提出了他们的“玻姆力学”版本。正如我们可能期望的那样(给出平均分布假设，保证了玻姆理论从经验上与量子力学是不可区分的)，如果可存在量是一个构造空间中关于位置的函数(并形成一个交换代数)的话，那么理论中的可观测量则正是量子力学中的可观测量，并形成了一个非交换代数。

CBH 定理假设，对于我们关心的理论，可观测量形成一个 C^* 代数。CBH 定理的内容为：给出确定的信息论约束，可观测量和可观测量的态的 C^* 代数就能获得量子理论的某个特定形式特征。这个定理不涉及可存在量和可存在量的态，也不处理量子测量问题(薛定谔真值问题)，而是将其搁置。但是从这里采取的观点看，测量问题反映的是不能被复制的观测数据，而“测量问题的解决”在于从物理方法论和力学的角度提出方案，并对产生出标准量子力学中的可观测量和可观测量的态的测量进行分析。这样的理论为复制的不可能性提供了适当的解释。但是既然现在有大量的这类解释存在，并且——假定 CBH 信息论原理——在原理上没有经验性的约束条件可以区分这些解释，那么，看起来在深入讨论这个问题上它并没有什么成效。唯一的动机就是“解决薛定谔问题”的建构性理论，看起来不太可能在由其他理论和实验问题驱动的物理学取得基础性进展之后还能继续存在下去。

概率问题——将量子力学解释为关于信息的原理性理论的核心基础问题——可以通过这样的方式来解决：由信息论约束条件，我们得到一个关于相互关系(其导致不存在相空间解释)的非对易(非布尔)理论。我们可以定义，以一种独特的方式(根据盖里森定理)将“转移概率”或“转移权数”扩展为与非交换结构的特定结构性质相关联，即它们与表示量子“命题”的几何元素之间的角度相对应。问题是怎么将权重理解为表示概率，而不需要将问题约化为真值问题如何解决。

看起来我们需要考虑去相干现象，这可以参见兰兹曼在本书第五章、迪克森在本书第四章的论述以及[Zurek，2003；Olliver *et al.* （奥利弗等），2004]。发生在宏观系统与其环境之间的一个极快的自发相互作用的过程，它导致了量子干涉实际上在瞬间会被阻止。粗略地讲，所发生的就是宏观系统比如说薛定谔猫典型地以纠缠态的方式和环境——庞大数量的自由灰尘粒子，空气分子，光子，背景辐射等——产生关联，纠缠态呈现一种特定的结构，并与基态的优先集有关，基态会保持稳定，尽管相互作用会持续发生并且越来越多的粒子会加入进来。看起来好像环境会通过测量与优先态相关的属性的方式来“监控”系统，通过这种方式，这些属性的信息会冗余地存储在环境中。这种优先基的稳定性，或者说鲁棒性，和环境中信息的冗余性一起，允许我们以类似经典的方式，从相互关系的总体形式——一般而言，如宏观指示器、猫和信息采集器——中识别出特定的突现结构：展现这些结构之间存在的量子干涉所需的关联信息，实际上会消失在环境中。因此，信息论约束看起来与以下两个条件相一致：(1)保证作为经典信息源的测量仪器存在的条件；(2)当给定总体系统如环境中部分子系统的特征时，保证某类信息收集器，即具有使用测量仪器来检验量子力学能力的信息收集器存在的条件。也就是说，退相干为量子关联结构中经典信息的突现提供了解释。

如果类似上面关于退相干的描述是可接受的话，那么概率问题就会约化到在主观贝叶斯意义下如何说明通过这种信息收集器分配给测量结果的概率的问题，这正是盖里森广义转移概率问题。也就是说，我们需要表明，如果量子理论，在基本层面上是一个关于不存在相空间表征的相互关系的非对易理论的话，那么当关联信息在它们的环境中可以被忽略时，它也是一个关于信息收集

器概率行为的理论，以及在相互关系结构中关于特定的突现结构的理论。沿该方向所做的论证请参阅[Pitowsky(皮托斯基)，2002]。

从这里提出的观点看，该理论将证明测量结果不是确定的。在一定程度上，测量仪器是经典信息源，由信息源随机产生的一个特定可区分的事件(“符号”)的发生不在该理论研究的范围内。在这种意义下，测量仪器在某种程度上就像经典信息源那样起作用，在该理论中最终仍然是一个“黑盒子”。因此量子描述将不得不引入一个“分割”，以分界在给定的测量过程中我们所认为的基本测量仪器与仪器所显示的量子现象。但是这种“分割”将不再会是暂时的，或者神秘的，或者在某种程度上是有问题的，就像在哥本哈根解释中那样(请参见本书第五章兰兹曼的论文)。这里“分割”正是反映了这样的基本解释：量子力学是关于信息的表达和操作的理论，这种信息被我们所在世界中关于信息转移的可能性与不可能性所约束，而不是研究非经典波和粒子运动方式的理论。

致谢

本章完成于2005年任职于马里兰大学期间，得到校级研究委员会学期奖资助。感谢马里兰大学对我的研究资助，以及杰里米·巴特菲尔德对本章内容初稿所给出的大量非常有帮助的评论。

参考文献

[Aharonov *et al.*, 1964] Y. Aharonov, P. G. Bergmann, and J. L. Lebowitz. Time-symmetric quantum states. *Physical Review B*, 134: 1410-1416, 1964. Reprinted in *Quantum Theory and Measurement*, J. Archibald Wheeler and W. H. Zurek (eds.), Princeton University Press, Princeton, 680-686, 1983.

[Aharonov *et al.*, 2005] D. Aharonov, A. Ta-Shma, U. V. Vazirani, and A. C. Yao. Quantum bit escrow. In *STOC* 2000, *Proceedings of the Thirty Second Annual ACM Symposium on Theory Of Computing. arXiv e-print quant-ph/*0004017, 2005.

[Aspect *et al.*, 1981] A. Aspect, P. Grangier, and G. Roger. Experimental tests of realistic local theories via Bell's theorem. *Physical Review Letters*, 47: 460-477, 1981.

[Aspect *et al.* , 1982] A. Aspect, P. Grangier, and G. Roger. Experimental realization of EPR gedankenexperiment: a new violation of Bell's inequalities. *Physical Review Letters*, 49: 91-94, 1982.

[Bacciagaluppi, 1994] G. Bacciagaluppi. Separation theorems and Bell inequalities in algebraic quantum mechanics. In P. Busch, P. Lahti, and P. Mittelstaedt (eds.), *Symposium on the Foundations of Modern Physics* 1993: *Quantum Measurement, Irreversibility and the Physics of Information*, 29-37. World Scientific, Singapore, 1994.

[Barenco, 1998] A. Barenco. Quantum computation: an introduction. In Lo *et al.* [1998], 143-183.

[Barnum *et al.* , 1996a] H. Barnum, C. M. Caves, C. A. Fuchs, Richard Jozsa, and B. Schumacher. Noncommuting mixed states cannot be broadcast. *Physical Review Letters*, 76: 2828, 1996.

[Barnum *et al.* , 1996b] H. Barnum, C. A. Fuchs, R. Jozsa, and B. Schumacher. General fidelity limit for quantum channels. *Physical Review A*, 54: 4707, 1996.

[Barnum *et al.* , 2001] H. Barnum, P. Hayden, R. Jozsa, and A. Winter. On the reversible extraction of classical information from a quantum source. *Proceedings of the Royal Society (London) A*, 457: 2019-2039, 2001.

[Bell, 1964] J. S. Bell. On the Einstein-Podolsky-Rosen Paradox. *Physics*, 1: 195-200, 1964. Reprinted in John Stuart Bell, *Speakable and Unspeakable in Quantum Mechanics*, Cambridge University Press, Cambridge, 1989.

[Bell, 1987] J. S. Bell. Beables for quantum field theory. In *Speakable and Unspeakable in Quantum Mechanics*, 173-180. Cambridge University Press, Cambridge, 1987.

[Benioff, 1980] P. Benioff. The computer as a physical system: A microscopic quantum mechanical Hamiltonian model of computers as represented by turing machines. *Journal of Statistical Physics*, 22(5): 563-591, 1980.

[Bennett and Brassard, 1984] C. H. Bennett and G. Brassard. Quantum cryptography: public key distribution and coin tossing. In *Proceedings of IEEE international conference on computers, systems, and signal processing*, 175-179, New York, 1984. IEEE.

[Bennett and Wiesner, 1992] C. H. Bennett and S. J. Wiesner. Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states. *Physical Review Letters*, 69: 2881-

2884, 1992.

[Bennett *et al.*, 1982] C. H. Bennett, G. Brassard, S. Breidbart, and S. J. Wiesner. Quantum cryptography, or unforgeable subway tokens. In D. Chaum, R. L. Rivest, and A. T. Sherman (eds.), *Advances in Cryptology: Proceedings of Crypto* 82, 267-275, New York, 1982. Plenum Press.

[Bennett *et al.*, 1993] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. Wootters. Teleporting an unknown quantum state via dual classical and EPR channels. *Physical Review Letters*, 70: 1895-1899, 1993.

[Bennett, 1973] C. H. Bennett. Logical reversibility of computations. *IBM Journal of Research and Development*, 17: 525-532, 1973.

[Bohm, 1951] D. Bohm. *Quantum Theory*. Prentice Hall, Englewood Cliffs, NJ, 1951.

[Bohm, 1952] D. Bohm. A suggested interpretation of quantum theory in terms of "hidden" variables. i and ii. *Physical Review*, 85: 166-193, 1952.

[Bohr, 1949] N. Bohr. Discussion with Einstein on epistemological problems in modern physics. In P. A. Schilpp (ed.), *Albert Einstein: Philosopher-Scientist*, volume VII. The Library of Living Philosophers, Open Court, La Salle, IL, 1949.

[Bondi, 1980] H. Bondi. *Relativity and Common Sense*. Dover Publications, 1980.

[Born, 1971] M. Born. *The Born-Einstein Letters*. Walker and Co., London, 1971.

[Boschi *et al.*, 1998] D. Boschi, S. Branca, F. De Martini, L. Hardy, and S. Popescu. Experimental realization of teleporting an unknown pure quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Physical Review Letters*, 80: 1121-1125, 1998.

[Bouwmeester *et al.*, 1997] D. Bouwmeester, J.-W. Pan, K. Mattle, M. Eible, H. Weinfurter, and A. Zeilinger. Experimental quantum teleportation. *Nature*, 390: 575-579, 1997.

[Brassard *et al.*, 1993] G. Brassard, C. Crépeau, R. Jozsa, and D. Langlois. A quantum bit commitment scheme provably unbreakable by both parties. In *Proceedings of the* 34*th Annual IEEE Symposium and the Foundations of Computer Science*, 362-371, New York, November 1993. IEEE.

[Brassard, 2000] G. Brassard. Remarks on quantum foundations in the light of quantum cryptography. "Quantum Foundations in the Light of Quantum Information and Cryptography", Universitéde Montréal, May 17-19 (2000)., 2000.

[Brukner and Zeilinger, 2000] Č. Brukner and A. Zeilinger. Quantum measurement and Shannon information, a reply to M. J. W. Hall. *arXiv e-print quant-ph/*0008091, 2000.

[Brukner and Zeilinger, 2002] Č. Brukner and A. Zeilinger. Information and fundamental elements of the structure of quantum theory. In *Festschrift for C. F. v. Weizsäcker on the occasion of his* 90*th birthday. arXiv e-print quant-ph/*0212084. Unknown, 20002.

[Brukner and Zeilinger, 2001] Č. Brukner and A. Zeilinger. Conceptual inadequacy of the Shannon information in quantum measurement. *Physical Review A*, 63: 022113, 2001.

[Bub and Clifton, 1996] J. Bub and R. Clifton. A uniqueness theorem for "no collapse" interpretations of quantum mechanics. *Studies in the History and Philosophy of Modern Physics*, 27: 181-219, 1996.

[Bub, 1997] J. Bub. *Interpreting the Quantum World.* Cambridge University Press, Cambridge, 1997.

[Bub, 2001a] J. Bub. The bit commitment theorem. *Foundations of Physics*, 3: 735-756, 2001.

[Bub, 2001b] J. Bub. Secure key distribution via pre-and post-selected quantum states. *Physical Review A*, 63: 032309-032311, 2001.

[Bub, 2004] J. Bub. Why the quantum? *Studies in History and Philosophy of Modern Physics*, 35B: 241-266, 2004.

[Bub, 2005] J. Bub. Quantum mechanics is about quantum information. *Foundations of Physics*, 34: 541-560, 2005.

[Cleve *et al.*, 1998] R. Cleve, A. Ekert, C. Macchiavello, and M. Mosca. Quantum algorithms revisited. *Proceedings of the Royal Society A*, 454: 339-354, 1998.

[Clifton *et al.*, 2003] R. Clifton, J. Bub, and H. Halvorson. Characterizing quantum theory in terms of information-theoretic constraints. *Foundations of Physics*, 33: 1561-1591, 2003.

[Connes, 1994] A. Connes. *Noncommutative Geometry.* Academic Press, San Diego, 1994.

[Cover and Thomas, 1991] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* Wiley, New York, 1991.

[Cushing, 1998] J. T. Cushing. *Philosophical Concepts in Physics.* Cambridge University Press, Cambridge, 1998.

[Deutsch and Jozsa, 1992] D. Deutsch and R. Jozsa. Rapid solution of problems by quantum computation. *Proceedings of the Royal Society of London A*, 439: 553, 1992.

[Deutsch, 1985] D. Deutsch. Quantum theory, the Church-Turing principle and the universal quantum compute. *Proceedings of the Royal Society of London A*, 400: 497, 1985.

[Deutsch, 1989] D. Deutsch. Quantum computational networks. *Proceedings of the Royal Society of London A*, 400: 73, 1989.

[Deutsch, 1999] D. Deutsch. Quantum theory of probability and decisions. *Proceedings of the Royal Society of London A*, 455: 3129-3197, 1999.

[Dieks, 1982] D. Dieks. Communication by EPR devices. *Physics Letters A*, 92: 271-272, 1982.

[Dürr *et al.*, 2003] D. Dürr, S. Goldstein, and N. Zangh. Quantum equilibrium and the role of operators as obserables in quantum theory. *arXiv e-print quant-ph/*0308038, 2003.

[Einstein *et al.*, 1935] A. Einstein, B. Podolsky, and N. Rosen. Can quantum-mechanical description of physical reality be considered complete? *Physical Review*, 47: 777-780, 1935.

[Einstein, 1919] A. Einstein. What is the theory of relativity? *The London Times*, page 13, 1919. First published November 28, 1919. Also in A. Einstein, *Ideas and Opinions*, Bonanza Books, New York, 1954, pp. 227-232.

[Einstein, 1949] A. Einstein. Autobiographical notes. In P. Allen Schilpp (ed.), *Albert Einstein: Philosopher-Scientist*, 3-94. Open Court, La Salle, IL, 1949.

[Einstein, 1967] A. Einstein. Letter to Schrödinger dated december 22, 1950. In K. Przibram (ed.), *Letters on Wave Mechanics.* Philosophical Library, New York, 1967.

[Ekert, 1991] A Ekert. Quantum cryptography based on bell's theorem. *Physical Review Letters*, 67: 661, 1991.

[Everett, 1957] H. Everett. "relative state" formulation of quantum mechanics. *Reviews of ModernPhysics*, 29: 454-462, 1957.

[Feynman, 1982] R. P. Feynman. Simulating physics with computers. *International Journal of Theoretical Physics*, 21: 467, 1982.

[Fuchs and Jacobs, 2002] C. A. Fuchs and K. Jacobs. An information-theoretic tradeoff relation for finite-strength quantum measurements. *Physical Review A*, 63: 062305-062320, 2002.

[Fuchs, 1997] C. A. Fuchs. Information gain vs. state disturbance in quantum theory. *Fortschr. Phys.*, 46: 535-565, 1997. Reprinted in *Quantum Computation: Where Do We*

Want to Go Tomorrow? S. L. Braunstein (ed.), 229-259, Wiley VCH, Weinheim (1998).

[Fuchs, 2000] C. A. Fuchs. Just two nonorthogonal quantum states. In *Quantum Communication, Computing, and Measurement* 2, 11-16, Dordrecht, 2000. Kluwer.

[Fuchs, 2001a] C. A. Fuchs. Notes on a Paulian idea: foundational, historical, anecdotal and forward-looking thoughts on the quantum. *arXiv e-print quant-ph/*0105039, 2001.

[Fuchs, 2001b] C. A. Fuchs. Quantum foundations in the light of quantum information. In A. Gonis (ed.), *Proceedings of the NATO Advanced Research Workshop on Decoherence and its Implications in Quantum Computation and Information Transfer. arXiv e-print quantph/*0106166, 2001.

[Fuchs, 2002a] C. A. Fuchs. The anti-Växjo interpretation of quantum mechanics. *arXiv e-print quant-ph/*0204146, 2002.

[Fuchs, 2002b] C. A. Fuchs. Quantum mechanics as quantum information (and only a little more). *arXiv e-print quant-ph/*0205039, 2002.

[Furasawa *et al.*, 1998] A. Furasawa, J. L. Sorensen, S. L. Braunstein, C. A. Fuchs, H. J. Kimble, and E. S. Polzik. Unconditional quantum teleportation. *Science*, 282: 706-709, 1998.

[Furry, 1936] W. H. Furry. A note on the quantum mechanical theory of measurement. *Physical Review*, 49: 393-399, 1936.

[Ghirardi, 2002] G. -C. Ghirardi. Collapse theories. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, 2002. http: //plato. stanford. edu/entries/qm-collapse/.

[Gisin, 1989] N. Gisin. Stochastic quantum dynamics and relativity. *Helvetica Physica Acta*, 62: 363-371, 1989.

[Gleason, 1957] A. N. Gleason. Measures on the closed sub-spaces of hilbert spaces. *Journal of Mathematics and Mechanics*, 6: 885-893, 1957.

[Goldstein, 2001] S. Goldstein. Bohmian mechanics. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy*, 2001. http: //plato. stanford. edu/entries/qm-bohm.

[Grover, 1997] L. K. Grover. Quantum mechanics helps in searching for a needle in a haystack. *Physical Review Letters*, 79: 325, 1997.

[Hall, 2000] M. J. W. Hall. Comment on "Conceptual inadequacy of the Shannon information" by Č. Brukner and A. Zeilinger. *arXiv e-print quant-ph/*0007116, 2000.

[Halvorson and Bub, 2005] H. Halvorson and J. Bub. Can quantum cryptography imply quantum mechanics? reply to Smolin. *Quantum Information and Computation*, 5: 170-175, 2005.

[Halvorson, 2004] H. Halvorson. Remote preparation of arbitrary ensembles and quantum bit commitment. *Journal of Mathematical Physics*, 45: 4920-4931, 2004.

[Hardy and Kent, 2004] L. Hardy and A. Kent. Cheat sensitive quantum bit commitment. *Physical Review Letters*, 92: 157901, 2004. Longer version available at arXiv e-print quantph/9911043.

[Hartley, 1928] R. V. L. Hartley. Transmission of information. *Bell System Technical Journal*, 7: 53, 1928.

[Howard, 2004] D. Howard. Who invented the Copenhagen interpretation? A study in mythology. http: //www. nd. edu/ dhoward1/Papers. html, 2004.

[Hughes, 1995] R. I. G. Hughes. *The Structure and Interpretation of Quantum Mechanics*. Cambridge University Press, 1995.

[Hughston *et al.*, 1993] L. P. Hughston, R. Jozsa, and W. K. Wootters. A complete classification of quantum ensembles having a given density matrix. *Physics Letters A*, 183: 14-18, 1993.

[Jaynes, 1957] E. T. Jaynes. Information theory and statistical mechanics ii. *Physical Review*, 108: 171-190, 1957.

[Jozsa and Linden, 2002] R. Jozsa and N. Linden. On the role of entanglement in quantum computational speed-up. *arXiv e-print quant-ph/*0201143, 2002.

[Jozsa and Schumacher, 1994] R. Jozsa and B. Schumacher. A new proof of the quantum noiseless coding theorem. *Journal of Modern Optics*, 41: 2343-2349, 1994.

[Jozsa, 1997a] R. Jozsa. Entanglement and quantum computation. In S. A. Huggett, L. J. Mason, K. P. Tod, S. T. Tsou, and N. M. J. Woodhouse (eds.), *The Geometric Universe: Science, Geometry, and the Work of Roger Penrose*, 369-378. Oxford University Press, 1997.

[Jozsa, 1997b] R. Jozsa. Quantum algorithms and the fourier transform. *arXiv e-print quant-ph/*9707033, 1997.

[Jozsa, 1998] R. Jozsa. Quantum information and its properties. In Lo *et al.* [1998], 49-75.

[Jozsa, 1999] R. Jozsa. Searching in grover's algorithm. *arXiv e-print quant-ph/*9901021, 1999.

[Jozsa, 2005] R. Jozsa. An introduction to measurement based quantum computation. *arXiv e-print quant-ph/*0508124, 2005.

[Kadison and Ringrose, 1997] R. Kadison and J. Ringrose. *Fundamentals of the Theory of Operator Algebras.* American Mathematical Society, Providence, Rhode Island, 1997.

[Kent, 1999a] A. Kent. Secure classical bit commitment using fixed capacity communication channels. *arXiv e-print quant-ph/*9906103, 1999.

[Kent, 1999b] A. Kent. Unconditionally secure bit commitment. *Physical Review Letters*, 83: 1447-1450, 1999.

[Klein, 1967] M. J. Klein. Thermodynamics in Einstein's thought. *Science*, 157: 509-516, 1967.

[Kochen and Specker, 1967] S. Kochen and E. P. Specker. On the problem of hidden variables in quantum mechanics. *Journal of Mathematics and Mechanics*, 17: 59-87, 1967.

[Landau, 1987] L. J. Landau. On the violation of Bell's inequality in quantum theory. *Physics Letters A*, 120: 54-56, 1987.

[Landsman, 1998] N. Landsman. *Mathematical Topics Between Classical and Quantum Mechanics.* Springer, New York, 1998.

[Lewis and Papadimitriou, 1981] H. R. Lewis and C. H. Papadimitriou. *Elements of the theory of computation.* Prentice-Hall, Englewood Cliffs, New Jersey, 1981.

[Lo and Chau, 1997] H. -K. Lo and H. F. Chau. Is quantum bit commitment really possible? *Physical Review Letters*, 78: 3410-3413, 1997.

[Lo and Chau, 1998] H. -K. Lo and H. F. Chau. Why quantum bit commitment and ideal coin tossing are impossible. *Physica D*, 120: 177-187, 1998.

[Lo *et al.*, 1998] H. -K. Lo, S. Popescu, and T. Spiller (eds.) *Introduction to Quantum Computation and Information*, Singapore, 1998. World Scientific.

[Lorentz, 1909] H. A. Lorentz. *The Theory of Electrons.* Columbia University Press, New York, 1909.

[Mackey, 1963] G. Mackey. *Mathematical Foundations of Quantum Mechanics.* Benjamin, New York, 1963.

[Mayers, 1996a] D. Mayers. The trouble with quantum bit commitment. *arXiv e-print quant-*

*ph/*9603015, 1996.

[Mayers, 1996b] D. Mayers. Unconditionally secure quantum bit commitment is impossible. In *Proceedings of the Fourth Workshop on Physics and Computation*, 224-228, Boston, 1996. New England Complex System Institute.

[Mayers, 1997] D. Mayers. Unconditionally secure quantum bit commitment is impossible. *Physical Review Letters*, 78: 3414-3417, 1997.

[Metzger, 2000] S. Metzger. Spin-measurement retrodiction revisited. *arXiv e-print quantph/* 0006115, 2000.

[Nielsen and Chuang, 2000] M. A. Nielsen and I. Chuang. *Quantum Computation and Quantum Information.* Cambridge University Press, Cambridge, 2000.

[Nielsen *et al.*, 1998] M. A. Nielsen, E. Knill, and R. Laflamme. Complete quantum teleportation using nuclear magnetic resonance. *Nature*, 396: 52-55, 1998.

[Nielsen, 2003] M. A. Nielsen. Journal club notes on cluster-state quantum computation. http: //qinfo. org/qc-by-measurement/, 2003.

[Nielsen, 2005] M. A. Nielsen. Cluster-state quantum computation. *arXiv e-print quantph/* 0504097, 2005.

[Nyquist, 1924] H. Nyquist. Certain factors affecting telegraph speed. *Bell System Technical Journal*, 3: 324, 1924.

[Olliver *et al.*, 2004] H. Olliver, D. Poulin, and W. Zurek. Environment as a witness: selective proliferation of information and emergence of objectivity in a quantum universe. *Physical Review Letters*, 93: 22041, 2004.

[Pitowsky, 2002] I. Pitowsky. Betting on the outcomes of measurements: a Bayesian theory of quantum probability. *arXiv e-print quant-ph/*0208121, 2002.

[Popescu and Rohrlich, 1998] S. Popescu and D. Rohrlich. The joy of entanglement. In Lo *et al.* [1998], 29-48.

[Preskill, 2005] J. Preskill. Lecture notes for quantum computation. http: //www. theory. caltech. edu/people/preskill/ph229/, 2005.

[Raussendorf and Briegel, 2001a] R. Raussendorf and H. J. Briegel. A one-way quantum computer. *arXiv e-print quant-ph/*0208121, 2001.

[Raussendorf and Briegel, 2001b] R. Raussendorf and H. J. Briegel. Quantum computing via measurements only. *Physical Review Letters*, 86: 5188-5199, 2001.

[Rivest *et al.*, 1978] R. L. Rivest, A. Shamir, and L. M. Adleman. A method of obtaining digital signatures and public-key cryptosystems. *Communications of the Association for Computing Machinery*, 21: 120-126, 1978.

[Saunders, 1998] S. Saunders. Time, quantum mechanics, and probability. *Synthese*, 114: 373-404, 1998.

[Schilpp, 1949] P. A. Schilpp (ed.). *Albert Einstein: Philosopher-Scientist*, volume VII, La Salle, IL, 1949. The Library of Living Philosophers, Open Court.

[Schrödinger, 1935] E. Schrödinger. Discussion of probability relations between separated systems. *Proceedings of the Cambridge Philosophical Society*, 31: 555-563, 1935.

[Schrödinger, 1936] E. Schrödinger. Probability relations between separated systems. *Proceedings of the Cambridge Philosophical Society*, 32: 446-452, 1936.

[Schumacher, 1995] B. Schumacher. Quantum coding. *Physical Review A*, 51: 2738-2747, 1995.

[Schumacher, 1998] B. Schumacher. Lecture notes on quantum information. University of Innsbruck, 1998.

[Shannon, 1948] C. E. Shannon. A mathematical theory of computation. *Bell System Technical Journal*, 27: 379-423, 623-656, 1948.

[Shor, 1994] P. W. Shor. Algorithms for quantum computation; discrete logarithms and factoring. In*Proceedings, 35th Annual Symposium on Foundations of Computer Science*, Los Alamitos, CA, 1994. IEEE Press.

[Shor, 1997] P. W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Journal of Computation*, 26: 1484-1509, 1997.

[Simon, 1994] D. R. Simon. On the power of quantum computation. In *Proceedings, 35th Annual Symposium on Foundations of Computer Science*, 116-123, Los Alamitos, CA, 1994. IEEE Press.

[Simon, 1997] D. R. Simon. On the power of quantum computation. *SIAM Journal of Computation*, 26: 1474-1483, 1997.

[Steane, 1998] A. Steane. Quantum computing. *Reports on Progress in Physics*, 61: 117-173, 1998.

[Steane, 2003] A. Steane. A quantum computer only needs one universe. *Studies in History* *and Philosophy of Modern Physics*, 34B: 469-478, 2003.

[Summers, 1990] S. J. Summers. On the independence of local algebras in quantum field theory. *Reviews in Mathematical Physics*, 2: 201-247, 1990.

[Timpson, 2004] C. Timpson. *Quantum Information Theory and the Foundations of Quantum Mechanics.* PhD thesis, Oxford, 2004.

[Vaidman *et al.*, 1987] L. Vaidman, Y. Aharonov, and D. Z. Albert. How to ascertain the values of s_x, s_y, s_z of a spin-1/2 particle. *Phys. Rev. Letters*, 58: 1385, 1987.

[van Fraassen, 1991] B. van Fraassen. *Quantum Mechanics: An Empiricist View.* Clarendon Press, Oxford, 1991.

[Vidal, 2003] G. Vidal. Efficient classical simulation of slightly entangled quantum computations. *Physical Review Letters*, 91: 147902, 2003.

[von Neumann, 1955] J. von Neumann. *Mathematical foundations of quantum mechanics.* Princeton University Press, Princeton, 1955.

[Wald, 1984] R. M. Wald. *General Relativity.* University of Chicago Press, 1984.

[Wallace, 2003] D. Wallace. Everettian rationality: defending Deutsch's approach to probability in the Everett interpretation. *Studies in the History and Philosophy of Modern Physics*, 34: 415-439, 2003.

[Wheeler, 1998] J. A. Wheeler. *Geons, Black Holes, and Quantum Foam: A Life in Physics.* W. W. Norton, New York, 1998. With K. Ford.

[Wiesner, 1983] S. J. Wiesner. Conjugate coding. *SIGACT News*, 15: 17, 1983.

[Wooters and Zurek, 1982] W. K. Wooters and W. H. Zurek. A single quantum cannot be cloned. *Nature*, 299: 802-803, 1982.

[Yuen, 2005] H. P. Yuen. Unconditionally secure quantum bit commitment. *arXiv e-print quant-ph/*0505132, 2005.

[Zurek, 2003] W. H. Zurek. Decoherence, einselection, and the quantum origins of the classical. *Review of Modern Physics*, 75: 715, 2003.

第七章　量子场论的概念基础

杰拉德·特霍夫特

1. 量子化场概念导论

有那么一些宝贵的科学成就，它们取得的成功远远超过了它们应当取得的，量子场论就是其中之一，如果考虑到它是建立在明显不稳的逻辑基础上的话。所有已知的亚原子粒子似乎都以惊人的准确性遵循量子场论的一种范本规则，这一范本拥有很普通的名字“标准模型”。该模型的创立者们从未预想到会有这样的成功，人们能够理所当然地质问它成功在哪里。

虽然标准模型很成功，但是我们也早就意识到了它不可能是完全正确的。绝大多数的量子场论不是渐近自由的，这意味着它们不能够被延拓到任意小距离尺度。我们可以试图修正标准模型，但这并不能从根本上增进我们的理解，因为我们知道，在那些极其小距离的尺度下问题会变得关联起来，出现一种我们尚不能清晰描述的力：引力。它需要首先被理解。

也许这正是量子场论的真正优势：我们知道它的限度在哪，也知道这些限度很遥远。两个亚原子粒子间的引力是极其微弱的。只要我们舍弃引力，理论就很完美。而且，正如我将要解释的，其内在逻辑并不是完全不可靠的。

亚原子粒子都存在于自旋和作用都与普朗克(Planck)常数 $\hbar$ 相当的物理学领域中。很明显我们需要用量子力学来描述它们。因为在亚原子相互作用中有效的能量与这些粒子的静止质能 mc^2 相当，或常常大于后者，它们行进的速度常常接近于光速 c，所以相对论效应将同样重要。因而，在 20 世纪上半叶中提了出这样的问题：

"如何使量子力学与爱因斯坦(Einstein)的狭义相对论相一致?"

正如我们将要解释的，量子场论是对该问题的回答。

我们的第一直觉将是非常不同的，且确实如此[Pais(派斯)，1986；Crease and Mann(克瑞斯和曼)，1986]。我们要建立态的抽象希尔伯特(Hilbert)空间，每个态包含固定或可变数目的粒子。随后，我们要假设相互作用的一致方案。"一致"意味着什么？在量子力学中，我们学到了如何描述这样一个过程，在其中一定数量的粒子开始时互相远离，而后相互靠近，这是"入"态 $|\psi\rangle_{in}$；在相互作用发生之后，我们得到的结果是粒子都互相离开，即"出"态 $|\psi'\rangle_{out}$。从某个入态演化到给定的出态的概率是由量子力学的跃迁振幅描述的，即 ${}_{out}\langle\psi'|\psi\rangle_{in}$。由所有入态和出态形成的向量空间中所有这些振幅的集合称为散射矩阵。人们可以质问该如何以某种方式构建散射矩阵使得：(i)它在洛伦兹(Lorentz)变换下是不变的；(ii)它满足量子因果性的严格定律。对于"量子因果性"，我们指的是不存在一个可测量的效应会以超过光的速度行进的情形。在实践中，这意味着我们必须要求定域算符 $\mathcal{O}_i(\mathbf{x}, t)$ 的任意集合遵循对易规则，使得一旦向量 $(\mathbf{x}-\mathbf{x}', t-t')$ 是类空的，对易子 $[\mathcal{O}_i(\mathbf{x}, t), \mathcal{O}_j(\mathbf{x}', t')]$ 为零。我们可以表明，这意味着散射矩阵必须满足色散关系。

这事实上是物理学家们开始时关于这些问题的思考方式。但我们应该怎样构建这样一个散射矩阵？是否存在一些体系化的程序？

一个量子化的场可能是一些完全不同的东西，但它似乎允许构建一个满足洛伦兹不变性定律和因果性的相互作用媒介。定域算符可以由场来构建。之后我们全部所要做的就是建立相对论的协变场方程方案，如麦克斯韦(Maxwell's laws)定律。即使在这些方程中引入非线性项似乎也很好理解，且如果我们让这样的系统经历数学上定义明确的程序——"量子化"，我们将会得到上述问题求解的备选项。

认识到一个量子化的场中的能量来自于量子化的能量包，它在所有方面的行为都像是基本粒子，且反过来认识到场形式的算符在从由基本粒子组成的希尔伯特空间开始时就可以定义了，之后，人们发现量子化的场确实描述了亚原子粒子。随后，还发现了在量子化的场中，能够引入相互作用(基本上通过在场方程中增加非线性项)的方式是非常有限的。量子化要求所有的相互作用都能够以拉格朗日(Lagrangian)函数$\mathcal{L}$的形式给出；相对论要求该函数$\mathcal{L}$是洛伦兹不变的，且最为突出的是，自洽的量子场论给出了更多的限制，这些限制产生了写下所有可能相互作用完备列表的可能性。标准模型只是这些列表中的一个。

本篇关于量子场论的简短论文的篇幅有限，不允许对所有技术细节给出详细描述。取而代之的是，论文把重心放在概念问题上，这些概念问题产生于与此理论相关的无数疑惑和问题提出之际。大家能够通过本论文来学习量子场论，但对于许多的技术细节，必须查阅更多的文献[de Wit and Smith(德威特和史密斯)，1986；Aitchison and Hey(艾奇逊和海)，1989；Ryder(瑞德尔)，1985；Itzykson and Zuber(依捷克森和朱伯)，1980；Cheng and Li(郑与李)，1984]。我们也仅限于讨论量子场论在基本粒子物理学中的应用。这些方法以及类似的方法对于低温物理学中的许多案例也是有用的，但这里并不涉及。

2. 标量场

2.1 经典理论：费曼(Feynman)规则

这里，场被看作是一个物理变量，是时空坐标$x=(\mathbf{x}, t)$的函数。为了使我们的理论与狭义相对论相一致，我们必须确定一个场在齐次洛伦兹变换下是如何变换的：

$$x' = Lx \tag{1}$$

若一个场ϕ变换如下：

$$\phi'(x) = \phi(x') \tag{2}$$

则ϕ被称为标量场。反常(improper)洛伦兹变换(如宇称反射P和时间反演T)不那么重要，因为我们知道自然在这些变换下并不是完全不变的。

我们首先考虑实的标量场，之后很容易将它推广到用复数表示的那些场。

通过量子化，标量场出现能量包，其行为如同无自旋的玻色—爱因斯坦粒子，如 π^0，$\pi^{\pm}$ 和 η^0。概念上，标量场是最容易处理的，但在第 9 节中我们将指出其他种类的场能够真正增强理论内在一致性的原因。

洛伦兹不变的场方程往往采取如下形式①：

$$(\partial_\mu^2 - m_{(i)}^2)\phi_i = F_i(\phi); \ \partial_\mu^2 \equiv \vec{\partial}_x^2 - \partial_t^2 \tag{3}$$

这里，下标 i 代表标量场不同的可能种类，$F_i(\phi)$ 可以是场 $\phi_j(x)$ 的任意函数。但是，通常我们假定存在一个势函数 $V^{\text{int}}(\phi)$，使得 $F_i(\phi)$ 是 V^{int} 的梯度，此外我们假定 V^{int} 是一个多项式，其次数最多为 4：

$$V^{\text{int}}(\phi) = \frac{1}{6}g_{ijk}\phi_i\phi_j\phi_k + \frac{1}{24}\lambda_{ijk\ell}\ \phi_i\phi_j\phi_k\phi_\ell$$

$$F_i(\phi) = \frac{\partial V^{\text{int}}(\phi)}{\partial\phi_i} = \frac{1}{2}g_{ijk}\phi_j\phi_k + \frac{1}{6}\lambda_{ijk\ell}\ \phi_j\phi_k\phi_\ell \tag{4}$$

其中 g 与 λ 在其下标的所有置换下必须是完全对称的。这实际上是对 $F_i(\phi)$ 可采取的形式的限制。没有这一限制的话，我们就不会有守恒的能量，理论的量子化也将不可能。后面我们会看到为什么不允许多项式中出现更高次的项(第 7 节)。

为了理解这类方程经典解的一般结构，我们暂且在方程(3)的 $F_i(\phi)$ 上加上函数 $-J_i(x)$。之后，我们用 $J_i(x)$ 的幂展开该解：

$$(m_{(i)}^2 - \partial_\mu^2)\phi_i(x) = J_i(x) - \frac{\partial}{\partial\phi_i}V^{\text{int}}(\phi(x)) \tag{5}$$

$$\phi_i(x) = \phi_i^{(1)}(x) + \phi_i^{(2)}(x) + \psi_i^{(3)}(x) + \cdots$$

$$= \int d^4yG_{ij}(x-y)\left(J_i(y) - F_j(\phi^{(1)}(y) + \phi^{(2)}(y) + \phi^{(3)}(y) + \cdots)\right)$$

函数 $G_{ij}(x-y)$ 是下述方程的一个解：

$$(m_i^2 - \partial_\mu^2)G_{ij}(x-y) = \delta_{ij}\delta(x-y) \tag{6}$$

而 $\phi_i^{(2)}(x)$ 是 $J_j(y)$ 中的二次项，$\phi^{(3)}(x)$ 是其中的三次项，以此类推。把 $J_j(y)$ 中相同次幂的项相加我们发现求解场方程(2.1)的递归过程。在计算的最后，

① 我们采用求和约定：未置于括号间的重复指标自动求和。希腊符号 μ 是取 4 个值的洛伦兹指标，拉丁符号 $i, j, \cdots$ 是从 1 到 3。我们的度规约定为 $g_{\mu\nu} = \text{diag}(-1, 1, 1, 1)$。

我们可以设 $J_i(x)$ 等于零，或更好的是让 J 仅在远离粒子产生的区域不为零，从而 J 相互作用是下述机器的一个简化模型，该机器在很久以前制造了粒子。事实上，在量子理论中也能同样证明，可以很方便地把 J 看作是实验终端的粒子探测器的模型。

我们看到方程(2.1)的解可以写作是许多项的和。这些项中的每一个都可以用图式的形式表示，称为费曼图。在这些图中，我们把一个时空点表示成一个点，函数 $G_{ij}(x-y)$ 表示为连接 x 与 y 的一条线，下标 i 可以指示每一条线。一个点要么与一项 $J_j(y)$ 相关联，要么是一个与系数 g_{ijk} 相关联的三点顶角(three-point vertex)，又或是一个具有系数 $\lambda_{ijk\ell}$ 的四点顶角。图 1 描绘了一个典型的费曼图。

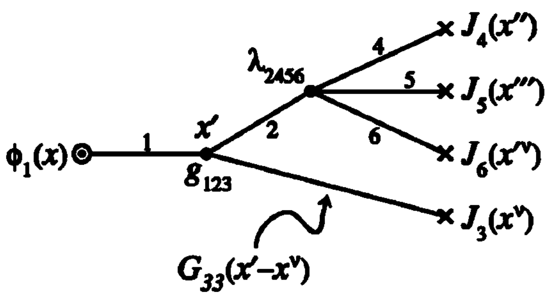


图 1　经典标量场的一个费曼图

我们来观察这些图的一般结构。图中有一些因子，如$\frac{1}{2}$，$\frac{1}{6}$等，它们可以很容易地从图的对称性中读出。经过构建，图中不存在闭圈，图只是简单地连接起来。在量子化的理论中将不会是这样。

这里提出一个重要的问题：格林(Green)函数 $G_{ij}(x-y)$ 并不是由方程(6)完全决定的，人们可以添加齐次方程解的任意组合：

$$(m_i^2-\partial_\mu^2)G_{ij}(x-y)=0 \tag{7}$$

在傅里叶空间中，这一任意性反映在下述事实中，即人们有在解中选择积分曲线 C 的自由：①

① 内积 $k \cdot x$ 代表$\vec{k}\cdot\vec{x}-k^0x^0$。

$$G_{ij}(x-y)=(2\pi)^{-4}\int_C d^4ke^{ik\cdot(x-y)}\frac{\delta_{ij}}{k^2+m_i^2} \tag{8}$$

我们的选择可以通过用无穷小虚数对极点(pole)的转换(shift)表现出来，在这之后我们选择沿着所有被积函数的实数轴作为围线(contour) C。一个合理的选择是：

$$G_{ij}^{+}(x-y)=(2\pi)^{-4}\int d^4ke^{ik\cdot(x-y)}\frac{\delta_{ij}}{\mathbf{k}^2-(k^0+i\varepsilon)^2+m_{(i)}^2} \tag{9}$$

其中 ε 是一个无穷小的正数。有了这些选择，在复平面 k^0 上的被积函数围线可以做转换，使得能够给 k^0 的虚部以一任意大的正值，且据此我们推断出，一旦时间差 x^0-y^0 是负的，则格林函数会为零。这样的格林函数，称为前向格林函数，能赋予我们的表达以想要的因果性结构，很显然不存在时间上反向传播的效应，或确实快于光的效应。

逆向选择 $G^-(x-y)$ 给我们以后向求解。但是，在量子化的理论中，我们将常常对其他的选择感兴趣，即费曼传播子(propagator)，它定义为：

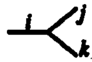
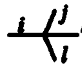
$$G_{ij}^{F}(x-y)=(2\pi)^{-4}\int d^4ke^{ik\cdot(x-y)}\frac{\delta_{ij}}{\mathbf{k}^2-k^{0^2}+m_i^2-i\varepsilon} \tag{10}$$

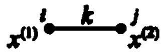
其中无穷小数 $\varepsilon>0$。

现在将用来得到解的完全展开的规则总结如下。

1)每一项都能够描绘为由线(称为传播子)连接的点(顶角 vertices)组成的图。终点，对应于我们希望知道场 ϕ 的点 x；其他的终点指关于对应点 $y^{(i)}$ 的因子 $J(y^{(i)})$，见图1。

2)不存在"闭圈"，即图必须简单连接起来(在量子理论中会不同)。

3)存在具有三叉(prongs)的顶角(3 顶角)，，每一个都与因子 g_{ijk} 相联系，且存在具有四叉的顶角(4 顶角)，，每一个有因子 λ_{ijkl}。

4)连接两点 $x^{(1)}$ 与 $x^{(2)}$ 的每条线，当在普通时空(位形空间)中时都与因子 $G_{ij}(x^{(1)}-x^{(2)})$ 相联系，或在动量空间中与因子：

$$\frac{\delta_{ij}}{k^2+m_i^2-i\varepsilon} \tag{11}$$

相联系(选择此 $i\varepsilon$ 的原因仅在量子化的理论中才突出)。

5)若处理的是位形空间，我们必须对每个顶角处的所有 x 值积分，除了那个 ϕ 被定义的顶角；若处理的是动量空间，我们必须对 k 值积分，在每个顶角处满足动量守恒的限制：$k_{out} = \sum_{in} k_{in}$。

6)一个"组合因子"(combinatorial factor)，对于经典理论而言，它是 $1/N$，其中 N 是不改变费曼图的源顶角(source vertices)置换的数目。

概括相互作用中更高次多项式的情形也不难，但现在不需要这样做。

2.2 自发对称破缺：戈德斯通模(Goldstone modes)

在经典理论中，哈密顿(Hamilton)密度是：

$$\mathcal{H}(\mathbf{x}, t) = \frac{1}{2}\dot{\phi}_i^2 + \frac{1}{2}(\vec{\partial}\phi_i)^2 + V(\phi); \quad V(\phi) = \frac{1}{2}m_i^2\phi_i^2 + V^{int}(\phi) \tag{12}$$

该理论在下述变换群下不变，如果 A 是正交的，且势函数 $V(\phi)$ 在该群下不变：

$$\phi_i'(x) = A_{ij}\phi_j(x) \tag{13}$$

最简单的例子是变换 $\phi \leftrightarrow -\phi$：

$$A = \pm 1; \quad V = V(\phi^2) = \frac{1}{2}a\phi^2 + \frac{\lambda}{24}\phi^4 \tag{14}$$

这里要考虑两种情形。

(i) $a > 0$。在此情形中，$\phi = 0$ 是 V 的绝对极小值。我们写作：

$$a = m^2 \tag{15}$$

且发现 m 事实上描述了粒子的质量。所有费曼图都有偶数根外(external)线。因为在量子理论中，这些线会与粒子相联系，我们发现拥有奇数个粒子的态绝不会演化为拥有偶数个粒子的态，反之亦然。若我们把量子数定义为 $Pc = (-1)^N$，其中 N 是 ϕ 粒子数，则我们发现 P_C 在相互作用过程中守恒。

(ii) $a < 0$。在此情形中，我们看到：

——试图用方程(15)来确定粒子的质量会产生奇怪的结果，即质量会是纯虚数。这样的客体，即"快子"(tachyon)的存在性未知，且它们可能很难与因果性相协调，此外还有：

——组态(configuration) $\phi = 0$ 并不对应于系统最低的能组态。当满足如下

条件时得到最低能：

$$\phi = \pm F；\ F^2 = -6a/\lambda \tag{16}$$

现在把势 V 改写为下式很容易：

$$V = \frac{1}{24}(\phi^2 - F^2)^2 - C \tag{17}$$

其中我们并不需要写出常数 C 的值，因为它并不出现在演化方程(2.1)中。现在有两个等价的真空态，也就是 V 的极小值。选择其中之一，我们引入新的场变量 ϕ 来表示：

$$\phi \equiv F + \tilde{\phi}$$

$$V = \frac{\lambda}{24}\tilde{\phi}^2(2F+\tilde{\phi})^2 = \frac{\lambda F^2}{6}\tilde{\phi}^2 + \frac{\lambda F^2}{6}\tilde{\phi}^3 + \frac{\lambda}{24}\tilde{\phi}^4 \tag{18}$$

且我们看到：

a)对于新的场 $\tilde{\phi}$，质量平方(mass-squared) $\widetilde{m}^2 = \lambda F^2/3$ 是正的；

b)出现了一个三叉顶角(three-prong vertex)，具有相关因子 λF。量子数 P_C 不再明显守恒。

这种现象称为“自发对称破缺”，它在量子场论中扮演着重要的角色。

接下来，我们考虑连续对称的情形。原型(prototype)案例是一个复数场的 $U(1)$ 对称。对称群由变换 $A(\theta)$ 组成，其中 θ 是一个角：

$$\Phi \equiv \frac{1}{\sqrt{2}}(\phi_1 + i\phi_2)；\ \Phi' = A(\theta)\Phi = e^{i\theta}\Phi \tag{19}$$

再者，在这些变换下不变的最普遍的势①为：

$$V(\Phi,\ \Phi^*) = a\Phi^*\Phi + \frac{1}{2}\lambda(\Phi^*\Phi)^2 - C \tag{20}$$

在 $U(1)$ 对称显然的那些情形中，我们可以改写费曼规则以直接应用于复场 Φ，注意到我们能够把势 V 写作两个独立变量 Φ 与 Φ^* 的实函数。有：

$$\partial_\mu^2\Phi = \frac{\partial V(\Phi,\ \Phi^*)}{\partial\Phi^*} \tag{21}$$

---

① 注意我们是如何调整组合因子的。这里选择的是在未来的计算中尽可能保持这些系数可预测的最自然的那些组合因子。

我们注意到费曼传播子可以写作其中拥有箭头的形式：一个指向点 x 的箭头，其中 x 是函数 $\Phi(x)$ 所要求的，它远离于点 x'，因子 $\Phi^*(x')$ 是在 x' 处从势 V 中提取的。在每个顶角处，进来的箭头与出去的箭头一样多，从而在相互作用过程中，随时间前向指向的线的总数与时间后向(backward)指向的线的差是守恒的。这是一个附加守恒量子数，解释为是“电荷”Q。根据诺特(Noether)定理，每个对称性都与这样一个守恒律相关联。

但是，若 $a<0$，这个 $U(1)$ 对称性就自发破缺了。则我们写作：

$$V=\frac{1}{2}\lambda(\Phi^*\Phi-F^2)^2-C;\ F^2\equiv -a/\lambda \tag{22}$$

这时，稳定的真空态在 Φ 值的复平面中形成一个闭合圆环(closed circle)。我们写作：

$$\Phi\equiv F+\tilde{\Phi};\ \tilde{\Phi}\equiv\frac{1}{\sqrt{2}}(\tilde{\phi}_1+i\tilde{\phi}_2) \tag{23}$$

$$\begin{aligned}V&=\frac{1}{2}\lambda(F(\tilde{\Phi}+\tilde{\Phi}^*)+\Phi^*\Phi)^2\\&=\lambda F\tilde{\phi}_1^2+\frac{\lambda F}{\sqrt{2}}\tilde{\phi}_1(\tilde{\phi}_1^2+\tilde{\phi}_2^2)+\frac{\lambda}{8}(\tilde{\phi}_1^2+\tilde{\phi}_2^2)^2\end{aligned}$$

关于该势的显著之处是，场 $\tilde{\phi}_2$ 的质量项消失了。$\tilde{\phi}_1$ 场的质量平方为 $\tilde{m}_1{}^2=2\lambda F$。有效场中的一个是无质量的这一事实是由一个连续对称的自发破缺引起的。一般而言，有戈德斯通定理：

若一个连续对称——它的对称群具有 N 个独立的生成元——自发破缺为一个剩余(residual)对称——它的群具有 N_1 个独立的生成元，则产生 $N-N_1$ 个无质量的有效场。

无质量场的传播子遵从没有 m^2 项的方程(6)，该 m^2 项给这些表达以“无穷大范围”：这样一个格林函数仅对大的类空或类时分离才会缓慢减小。这些无质量的振荡模式称为“戈德斯通模”。

2.3 经典理论的量子化

我们要如何“量子化”一个场理论呢？在量子力学的早期，我们学到的是“取经典系统的泊松(Poisson)括号，然后用对易子来替代它们”。在经典的表

达被算符替代时，我们需要不时地调整它们出现的顺序，但这些规则似乎留下了不必要的歧义。事实上，如果这样一个过程可能的话，我们会得到一个量子理论，它随着$\hbar$趋于零极限再现了最初的经典系统。同样，那些经典系统在其变换下不变的对称群，也常常在量子系统中重新出现。

但是，一个场理论拥有物理自由度的严格无限集（在三维空间中的每一点的场值，或傅里叶模的完备集）。通常经过“量子化”，会产生使理论没有意义的无穷大。因而我们不得不通过一些中间步骤，来更加仔细地构造“量子化”概念。现在因为问题的答案已经知晓，我们常常会忘了如何能够严密地得到这些答案，且为什么采取它们所具有的那种形式。论证的严格逻辑顺序是什么？

首先，期望每个经典场理论有一个量子力学的对应是不合理的。我们所希望的是，构建某个量子系统、其希尔伯特空间及其哈密顿量，使得在一个或更多特殊极限下，它能再现一个已知的经典理论。我们还要求理论的某些特征，如洛伦兹不变性和因果性，但最重要的是我们要求它的内在没有逻辑缺陷，从而允许我们在所有能想到的条件下，计算在这样一个系统中粒子是如何相互作用的。但是，我们将继续使用术语“量子化”，这意味着我们试图从某个给定的经典场论在 $\hbar\to 0$ 极限下构建一个量子理论。

通常，作者们会忘了提到该逻辑过程的第一步，也是非常重要的一步：用一个严格有限的理论来取代希望量子化的经典场理论。假定小于某一尺度的物理结构对我们而言并不重要，我们用离散但致密的点阵取代三维空间的连续统一体。在微分方程中，我们用有限差分比（finite ratios of differences）$\Delta/\Delta x^i$ 代替所有的偏微分 $\partial/\partial x^i$，其中 $\Delta\phi$ 表示 $\phi(x+\Delta x)-\phi(x)$。在傅里叶空间中，这意味着波数$\vec{k}$被限制在有限的范围内，即布里渊（Brillouin）区，从而对$\vec{k}$的积分能够永不发散。

如果该点阵足够致密，我们感兴趣的解便几乎不依赖于该点阵的细节，从而经典系统将恢复洛伦兹不变性，且光速将成为扰动速度的实际极限。如果有必要的话，我们也能够在三维空间中施加周期性边界条件，那样的话我们的系统是完全有限的。这种类型的有限系统允许旧版本意义的“量子化”：用对易子取代泊松括号。注意我们还没有离散化时间，故我们理论的哈密顿量具有如下

形式：

$$H = T + V \tag{24}$$
$$= \sum_{x^a}\prod_{a=1}^{3}(\Delta x^a)\left(\frac{1}{2}\sum_i(\partial\phi_i/\partial t)^2+\frac{1}{2}\sum_{i,a}\left(\frac{\Delta\phi_i}{\Delta x^a}\right)^2+V(\phi)\right)$$

场 $\phi_i(x)$ 相关联的正则动量为：

$$p^i(x) = (\partial\phi_i/\partial t)\prod_{a=1}^{3}(\Delta x^a) \tag{25}$$

并且，我们假定它们是满足下式的算符：

$$[\phi_i(x),\ \phi_j(x')]=0 \quad [p^i(x),\ p^j(x')]=0; \quad [\phi_i(x),\ p^j(x')]=i\delta_i^j\delta_{x,x'} \tag{26}$$

现在，我们需要等待并观察在无限致密空间点阵极限下会发生什么。是否会像经典理论那样，我们的量子调合物最终是洛伦兹不变的？我们如何在物理态上进行洛伦兹变换？这些问题被证明是非常需要予以回答的，但答案是已知的。我们首先需要一些有用的技术工具。

2.4 费曼路径积分

费曼路径积分常常被介绍为是一种“无穷维”积分。我们再次首先来看有限的情形。把广义坐标（这里是 ϕ_i 场）标示为 q_i，广义动量为 p_i。哈密顿量（2.3）是守恒类型的，空间元 $\prod_{a=1}^{3}(\Delta x^a)$ 充当着质量。为了之后的应用，我们需要一个更普遍的哈密顿量，其中也包含动量 p_i 的线性项：

$$H = T + V;\quad T = \sum_i\frac{(p_i - A_i(\mathbf{q}))^2}{2m_{(i)}};\quad V = V(\mathbf{q}) \tag{27}$$

原则上，我们维持坐标与动量的数目 n 有限，毋庸置疑在此情形下问题中的微分方程具有唯一的有限的解（假定函数 A_i 与 V 是足够光滑的，事实上我们处理的几乎都是多项式）。考虑位形态 $|\mathbf{q}\rangle$ 与动量态 $|\mathbf{p}\rangle$ 我们有：

$$\langle\mathbf{q}|\mathbf{q}'\rangle=\delta^n(\mathbf{q}-\mathbf{q}'),\ \langle\mathbf{p}|\mathbf{p}'\rangle=\delta^n(\mathbf{p}-\mathbf{p}');\ \langle\mathbf{q}|\mathbf{p}\rangle=(2\pi)^{-n/2}e^{ip_iq_i} \tag{28}$$

考虑到算符的顺序，我们写出动能：

$$T = \sum_i\frac{p_i^2-2A_ip_i+A_i^2}{2m_{(i)}}+iW(\mathbf{q}) \tag{29}$$

$$W(\mathbf{q}) = \sum_i \frac{[A_i(\mathbf{q}), p_i]}{2im_{(i)}} = \sum_i \frac{\partial_i A_i(\mathbf{q})}{2m_{(i)}}$$

这使我们能够快速地计算出矩阵元：

$$\langle \mathbf{q} | H | \mathbf{p} \rangle = \langle \mathbf{q} | \mathbf{p} \rangle (h(\mathbf{q}, \mathbf{p}) + iW(\mathbf{q})) \tag{30}$$

$$\langle \mathbf{p} | H | \mathbf{q} \rangle = \langle \mathbf{p} | \mathbf{q} \rangle (h(\mathbf{q}, \mathbf{p}) - iW(\mathbf{q})) \tag{31}$$

其中 $h(\mathbf{q}, \mathbf{p})$ 是经典的哈密顿量，它是两组变量 $\mathbf{q}$ 与 $\mathbf{p}$ 的函数。

对短的时间间隔 δt 而言，演化算符 $U(t, \delta t)$ 为：

$$U(t, \delta t) = e^{-iH(t)\delta t} = \mathbb{I} - iH\delta t + \mathcal{O}(\delta t)^2 \tag{32}$$

现在很容易得到在态 $\langle \mathbf{p} |$ 与 $| \mathbf{q} \rangle$ 之间的矩阵元：

$$\begin{aligned} \langle \mathbf{p} | U(t, \delta t) | \mathbf{q} \rangle &= \langle \mathbf{p} | \mathbf{q} \rangle - i\delta t \langle \mathbf{p} | H | \mathbf{q} \rangle + \mathcal{O}(\delta t)^2 \\ &= (2\pi)^{-n/2} e^{-ip_i q_i} \left(1 - i\delta t \{ h(\mathbf{q}, \mathbf{p}) - iW(\mathbf{q}) \} + \mathcal{O}(\delta t)^2 \right) \\ &= (2\pi)^{-n/2} \exp\left(-i\mathbf{p} \cdot \mathbf{q} - i\delta t \{ h(\mathbf{q}, \mathbf{p}) - iW(\mathbf{q}) \} + \mathcal{O}(\delta t)^2 \right) \end{aligned} \tag{33}$$

使得这一表达非常有用的原因是基于这样的事实，即它在极限 $\delta t \downarrow 0$ 时不会成为奇点(singular)。动量—动量与位置—位置矩阵元在该极限下会成为奇点。

接下来，我们考虑有限时间间隔 T。在该时间间隔内的演化算符形式上可以看作是许多个短时间隔 δt 演化算符的序列，有 $T = N\delta t$。利用闭包(closure)，在所有时间间隔上，同时在 $\mathbf{q}$ 空间和 $\mathbf{p}$ 空间中都有：

$$\mathbf{I} = \int d^n \mathbf{q} | \mathbf{q} \rangle \langle \mathbf{q} | = \int d^n \mathbf{p} | \mathbf{p} \rangle \langle \mathbf{p} | \tag{34}$$

我们可以写为：

$$\begin{aligned} | \psi(\mathbf{q}_N, T) \rangle = \langle \mathbf{q}_N | U(0, T) | \psi(0) = \int d^n \mathbf{q}_0 \int d^n \mathbf{p}_0 \cdots \int d^n \mathbf{q}_{N-1} \int d^n \mathbf{p}_{N-1} \\ \langle \mathbf{q}_N | \mathbf{p}_{N-1} \rangle \langle \mathbf{p}_{N-1} | U(t_{N-1}, \delta t) | \mathbf{q}_{N-1} \rangle \langle \mathbf{q}_{N-1} | \mathbf{p}_{N-2} \rangle \\ \langle \mathbf{p}_{N-2} | U(t_{N-2}, \delta t) | \mathbf{q}_{N-2} \rangle \cdots \langle \mathbf{p}_0 | U(0, \delta t) | \mathbf{q}_0 \rangle \langle \mathbf{q}_0 | \psi(0) \rangle \end{aligned} \tag{35}$$

代入方程(2.4)，我们看到：

$$\begin{aligned} | \psi(\mathbf{q}_N, T) \rangle = \left(\prod_{\tau=0}^{N-1} \int d^n \mathbf{q}_\tau \int d^n \mathbf{p}_\tau \frac{e^{-W(\mathbf{q}_\tau)\delta t}}{(2\pi)^n} \right) \times \\ \exp i \sum_{\tau=0}^{N-1} \delta t \left(\mathbf{p}_\tau \frac{\mathbf{q}_\tau + 1 - \mathbf{q}_\tau}{\delta t} - h(\mathbf{q}_\tau, \mathbf{p}_\tau, t_\tau) \right) \langle \mathbf{q}_0 | \psi(0) \rangle \end{aligned} \tag{36}$$

定义：

$$\dot{\mathbf{q}}_\tau \equiv \frac{\mathbf{q}_{\tau+1} - \mathbf{q}_\tau}{\delta t} \tag{37}$$

和

$$L(\mathbf{p},\ \mathbf{q},\ \dot{\mathbf{q}},\ t) = \mathbf{p} \cdot \dot{\mathbf{q}} - h(\mathbf{q},\ \mathbf{p},\ t) \tag{38}$$

以及测量：

$$\prod_{\tau=0}^{N-1} \int d^n \mathbf{q}_\tau \int d^n \mathbf{p}_\tau \frac{e^{-W(\mathbf{q}_\tau)\delta t}}{(2\pi)^n} \equiv \int \mathcal{D}\mathbf{q}\, \mathcal{D}\mathbf{p} \tag{39}$$

则我们得到一个表达式，它似乎很容易延拓到时间变量的无限细网格（fine grids）：

$$\langle \mathbf{q}_N \mid \psi(T) \rangle = \int \mathcal{D}\mathbf{q}\, \mathcal{D}\mathbf{p} (\exp i \sum_{\tau=0}^{N-1} \delta t\, L(\mathbf{p},\mathbf{q},\dot{\mathbf{q}},t)) \langle \mathbf{q}_0 \mid \psi(0) \rangle \tag{40}$$

在这些表达中，我们事实上允许在哈密顿量 H 与拉格朗日量 L 中的参数明确依赖于时间 t，从而揭示出这些表达式的物理结构。注意到：

$$L(\mathbf{p},\mathbf{q},\dot{\mathbf{q}},t) = -\sum_i \frac{(p_i - A_i - m_{(i)}\dot{q}_i)^2}{2m_{(i)}} - V(\mathbf{q}) + \sum_i (A_i \dot{q}_i + \frac{1}{2} m_{(i)} \dot{q}_i^2) \tag{41}$$

在所有动量变量上的积分很容易进行，在给定一些只依赖于质量 $m_{(i)}$ 的常数时：

$$\langle \mathbf{q}_N \mid \psi(T) \rangle = \int \mathcal{D}\mathbf{q} \exp(i \sum_{\tau=0}^{M-1} \delta t L(\mathbf{q},\dot{\mathbf{q}},t)) \langle \mathbf{q}_0 \mid \psi(0) \rangle \tag{42}$$

其中：

$$L(\mathbf{q},\dot{\mathbf{q}},t) = T - V; \quad T = \sum_i (\frac{1}{2} m_{(i)} \dot{q}_i^2 + A_i \dot{q}_i) \tag{43}$$

$$\mathcal{D}\mathbf{q} = e^{-\sum_\tau W(\mathbf{q}_\tau)\delta t} \prod_{\tau=0}^{N-1} \left(d^n \mathbf{q}_\tau \prod_i \left(\frac{m_{(i)}}{2\pi\delta t} \right)^{\frac{1}{2}} \right)$$

事实上，$L(\mathbf{q},\ \dot{\mathbf{q}},\ t)$是通过相对于 $\mathbf{p}$ 求 $L(\mathbf{p},\ \mathbf{q},\ \dot{\mathbf{q}},\ t)$ 的极值而从后者中得到的：

$$\frac{\partial}{\partial p_i} L(\mathbf{p},\ \mathbf{q},\ \dot{\mathbf{q}},\ t) = 0; \quad \dot{q}_i = \frac{\partial h(\mathbf{q},\ \mathbf{p},\ t)}{\partial p_i} \tag{44}$$

这恰好是在经典理论中拉格朗日量与哈密顿量之间的标准关系，故 L 确实是拉格朗日量。

若连续极限存在，方程(42)中的幂正好是经典作用量的 i 倍：

$$S=\int \mathrm{d}tL(\mathbf{q},\ \dot{\mathbf{q}},\ t) \tag{45}$$

假定方程(32)中的$\mathcal{O}(\delta t)^2$ 在极限下为零是很诱人的事，毕竟，它们只是乘了因子 $N\approx C/\delta t$。那样的话，方程(42)中的演化算符明显采用了从 $\mathbf{q}_0$ 到 $\mathbf{q}_N$ 所有路径的积分的形式，这就是费曼路径积分。在一个场理论情形下，我们考虑定义在空间中一个点阵上的场，因为路径积分是从时间变量的一个点阵开始的，我们最后处理的是时空中一个点阵。

> 场理论中的演化算符是这样来描述的：首先改写在时空中致密点阵上的理论。用相应的有限差分比取代偏微分，我们写出该理论的作用量 S 的表达式。通常，它可以写作是对拉格朗日密度$\mathcal{L}(\phi,\ \partial_\mu\phi)$的积分。理论的演化算符是通过对给定时空点(patch)上所有场组态 $\phi(\mathbf{x},\ t)$的 e^{iS}积分而得到的。积分测量是根据方程(43)来定义的。

时间微分中的线性项 A_i 并不在标量场理论中发挥作用，但在矢量场理论中发挥作用，且它们出现在测量(43)中这一事实常常被忽略。事实上，在绝大多数情形中，$W(\mathbf{q})$为零，但我们必须意识到在某些特殊情形中它可能会引起问题。我们眼下暂且忽略 W 项。

2.5　量子化理论的费曼规则

量子化场理论的费曼规则最初是通过对微扰理论的仔细分析而得到的。把量子哈密顿量 H 写作 $H=H_0+H^{\mathrm{int}}$，我们合并场中所有的双线性项和它们在 H_0 中的微商，并对 H_{int}的小值进行微扰展开。这产生了一组计算规则，非常类似于因为经典理论而得到的规则，见 2.1 节。现在大多数的这些规则(但并非所有的)能够从路径积分中精确地推导出来。

我们首先来推演这些规则以计算类型(42)那样的有限维积分。虽然通常我们的作用量并不包含变量 $q_i(t)$的线性项，但现在我们需要这样的项，所以如

果必要的话我们会手动添加，仅在计算结束时去掉它们。这里不需要明确标出时间变量 t，我们把它并入到指标 i 中。则作用量为：

$$\begin{aligned} S(\mathbf{q}) &= \sum_{\mathbf{x},t} \mathcal{L}(\mathbf{x},t) \\ &= J_i q_i - \frac{1}{2}M_{ij}q_iq_j - \frac{1}{6}A_{ijk}q_iq_jq_k - \frac{1}{24}B_{ijk\ell}\, q_iq_jq_kq_\ell \end{aligned} \tag{46}$$

为了计算 $\int d^N\mathbf{q}e^{iS(\mathbf{q})}$，我们在指数上只保留双线性部分（具有系数 M_{ij} 的项），然后展开所有其他项的指数：

$$\begin{aligned} {}_{out}\langle 0 \mid 0\rangle_{in} = C\int d^N\mathbf{q}\Big(\exp\Big(-\frac{1}{2}iM_{ij}q_iq_j\Big)\Big)\sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty}\sum_{m=0}^{\infty}\frac{1}{k!\ell!m!}\times \\ (iJ_{i_1}q_{i_1})\cdots(iJ_{i_k}q_{i_k})\Big(-\frac{i}{6}A_{i_1j_1k_1}q_{i_1}q_{j_1}q_{k_1}\Big)\cdots\Big(-\frac{i}{6}A_{i_\ell j_\ell k_\ell}\, q_{i_\ell}\, q_{j_\ell}\, q_{k_\ell}\Big) \\ \Big(-\frac{i}{24}B_{i_1j_1k_1\ell_1}q_{i_1}q_{j_1}q_{k_1}q_{\ell_1}\Big)\cdots\Big(-\frac{i}{24}B_{i_mj_mk_m\ell_m}q_{i_m}q_{j_m}q_{k_m}q_{\ell_m}\Big) \end{aligned} \tag{47}$$

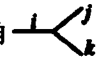
其中 C 是一个不依赖系数只依赖于它们维度的常数。

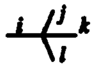
如果我们知道如何计算 J 项的话，就能够计算所有这些积分。但是既然我们知道如何精确地做高斯积分，我们就可以计算所有阶（orders）：

$$\begin{aligned} \int d^N\mathbf{q}\ \exp i\Big(-\frac{1}{2}M_{ij}q_iq_j + J_iq_i\Big) = \frac{(2\pi)^{\frac{N}{2}}}{(\det(M))^{\frac{1}{2}}}\exp\Big(\frac{1}{2}iJ_iM_{ij}^{-1}J_j\Big) = \\ C\sum_{k=0}^{\infty}\frac{1}{k!}\Big(\frac{1}{2}iJ_{i_1}M_{i_1j_1}^{-1}J_{j_1}\Big)\cdots\Big(\frac{1}{2}iJ_{i_k}M_{i_kj_k}^{-1}J_{j_k}\Big) \end{aligned} \tag{48}$$

这一表达告诉我们如何通过集中具有给定 J_i 的幂次的项在方程(47)中做积分。这一计算的结果可以简要总结如下。

1）每一项都可以描绘为一个由线（传播子）连接起来的点（顶角）组成的图。线可以终结于点 i，——×，它指的是因子 J_i。

2）存在具有三叉（3 顶角）的顶角 

，每一个叉都与一个因子 A_{ijk} 相联系，也存在具有四个叉（4 顶角）的顶角，

，每一叉有一个因子 $B_{ijk\ell}$。

3）连接两点 i 与 j 的每条线，都与一个因子 M_{ij}^{-1} 相联系。

4）但是，与经典理论相反，图中可能包含未连接的部分，或是多重连接的部分：闭圈。见图 2。

5)方程(47)中存在产生自系数如 $k!$ 的组合因子。我们可以从对这些因子的推导中获得经验，它们直接来自于图的对称结构。这些技术细节这里不做进一步展开。

显然当变量 q_i 被场 $\phi_i(\mathbf{x},\ t)$ 取代而作用量用一个场理论的作用量取代时，重新插入这些系数的$(\mathbf{x},\ t)$依赖性不会改变什么：

$$S=\int d^4x\,\mathcal{L}(\mathbf{x},\ t)$$

$$\mathcal{L}(\mathbf{x},\ t)=-\frac{1}{2}(\partial_\mu\phi_i)^2-\frac{1}{2}m_{(i)}^2\phi_i^2-V(\phi)+J_i(x)\phi_i(x) \tag{49}$$

这些规则与2.1节中的一样，唯一真正的区别在于，在量子理论中拥有闭圈的图有意义。这些图可以看作是经典场理论的“量子修正”。在第(4)点中提到的不连接的图，是由于技术原因才产生的，具体的原因我们不作进一步的阐述；在实际的计算中它们也常常被忽略掉。

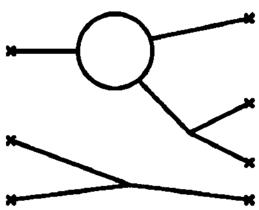


图2 量子化标量场的费曼图

但是，在一个地方我们做出了省略，即没有计算总的常数 C。这个忽略源自于对两个系数(一个是在测量中，一个来自于高斯积分)的相消(cancellation)，它们中的每一个都在无限密集网格(dense grid)极限下趋于无穷大。在大多数情形下，我们对这一系数(它指的是真空能)不感兴趣，但这意味着需要从这些费曼图中提取更多相关的物理信息。幸运的是，这一缺陷很容易消除。“源插入项(source insertions)”$J_i(x)\phi_i(x)$作为一个模型能够通用于产生和检测粒子。令$|0\rangle_{\rm in}$与$|0\rangle_{\rm out}$都是理论的真空态或基态。在早些时候，哈密顿量中的插入项 $-J(\mathbf{x},\ t)\phi$ 作用于这一真空态以使其激发到我们所感兴趣的初态。通过对 J 求微分，我们可以达到我们想要考虑的任意初态。类似地，在实验的终端，晚些时候 $J\phi$ 可以把我们希望检测的粒子态与最终的真空态联系起来。

简言之，对 $J(\mathbf{x},t)$ 求微分给我们所希望研究的任意矩阵元。这比我们想象的要容易得多：J_i 指的是 i 类型的粒子，且如果我们赋予其时空依赖性，这种时空依赖性与我们想要观测的粒子的波函数所拥有的相同（将它置于那个粒子的"质壳"上），则我们可以肯定的是，不会有来自多余粒子态的影响。我们只需要核对归一化，但这同样也不难：我们调整粒子 1 使它的振幅为 1，单个粒子不可能散射（它可以不稳定，但那是另一回事）。常数 C 总是不参与这些计算。

重要的一点是逆矩阵 M^{-1} 的模糊性。正如在经典情形中一样存在齐次解，所以如果我们处理的是动量空间，将会有如何在传播子的极点周围积分的问题。2.1 节中提到的 $i\varepsilon$ 解决方法现在很有必要。详述如下。考虑位置空间中的传播子，取定其极（pole）位置如下：

$$\int d^4k \frac{e^{i\mathbf{k}\cdot\mathbf{x}-ik^0t}}{m^2+\mathbf{k}^2-k^{0^2}-i\varepsilon};\varepsilon\downarrow 0 \tag{50}$$

极点在 $K^0=\pm(\sqrt{m^2+\mathbf{k}^2}-i\varepsilon)$ 处。现在考虑在 $t=-T+i\beta$ 时刻的这种传播子，其中 T 与 β 都大。因为 β 大，则在负 k^0 处对围线的选择是不重要的，因为那里的指数非常小。在正 k^0 处，我们选择围线以越过极点，从而 k^0 的虚部被选为正的。之后我们看到在负时间中指数迅速为零。简言之，当 T 与 β 都很大且是正的时，如果时间 t 趋于 $-T+i\beta$，我们的传播子趋于零。这对 $t\to+T-i\beta$ 也同样成立。事实上，我们想要使演化算符被这两个极限中的空图（empty diagram）所占据。写作：

$$\langle\psi\mid U(0,+T-i\beta)\mid\psi'\rangle=\sum_E\langle\psi\mid E\rangle\exp(-iET-\beta E)\langle E\mid\psi'\rangle \tag{51}$$

其中，$\mid E\rangle$ 是能量本征态。在大的 β 下，真空态应该处于支配地位。相反，若我们考虑时间的后向演化，则需要另外的 $i\varepsilon$ 解决方法。这样的话处理的将是适用于逆散射矩阵或散射矩阵复共轭的费曼规则。

现在，我们该补上识别具有入射和出射粒子的外围线（那些在图中伸出来的线）的方法了。对于一个入射粒子，我们采用源函数 $J(x)$ 来识别，其傅里叶分量发出正的能量 k^0。对于出射粒子，源发射负的 k^0。根据上面构造的规则，在傅里叶空间 $(k_\mu^2+m^2-i\varepsilon)^{-1}$ 中这些源会通过传播子与图中的其余部分连接起来。因为入射和出射粒子有 $K_\mu^2+m^2=0$，我们必须取极点处的留数。实践中，

这意味着我们必须移除掉外部传播子，该过程称为“割断(amputation)”。之后我们仍然需要建立一个归一化因子。利用光学定理，这一因子很容易通过核对散射矩阵的幺正性而得到。初看之下，这似乎只是一个简单的数值系数，但当自能修正影响到传播子的时候，在高阶时还是有些许复杂的。这些修正也从物理的散射矩阵中移除了不稳定的粒子。我们将在第6节中回到这一问题上来。完整的费曼规则列在4.5节中。

3. 旋量场

3.1 狄拉克(Dirac)方程

前一节中介绍的场只可用于描述自旋为0的粒子。在量子理论中，粒子会遇到小群(little group)的任意表征，小群是不影响粒子4动量的非齐次洛伦兹群的子群。对于普通空间中的有质量粒子，它是一个三向量的旋转群$SO(3)$。它可以通过一个$\geqslant 0$的整数或一个整数$+\frac{1}{2}$来表示，它表征了粒子的总自旋。因而，下一个要处理的是自旋$\frac{1}{2}$粒子。这样一个粒子的波函数有两个分向量，一个用于自旋向上，另一个用于自旋向下。因而，为了描述这一粒子的相对论理论，我们应该采用一个满足相对论协变场方程的二分量场。保罗·狄拉克第一个发现了关于自旋$\frac{1}{2}$的自由粒子的恰当相对论协变方程：

$$(m+\sum_{\mu}\gamma^{\mu}\partial_{\mu})\psi(x)=0 \tag{52}$$

但场$\psi(\mathbf{x},t)$有四个复分量。这里，γ^{μ}，$\mu=0,1,2,3$是4×4矩阵，满足：

$$\{\gamma^{\mu},\gamma^{\nu}\}=\gamma^{\mu}\gamma^{\nu}+\gamma^{\nu}\gamma^{\mu}=2g^{\mu\nu};\ \gamma_{\mu}^{\dagger}=g_{\mu\nu}\gamma^{\nu} \tag{53}$$

与标量场情形相反，这里狄拉克方程在时空微分上是一阶的。此外，人们可以对场施加一个“现实(reality)条件”，即马约拉纳(Majorana)条件，具有如下形式：

$$\psi(x)=C\psi^{*}(x),\ \gamma^{\mu}C=C(\gamma^{\mu})^{*},\ \mu=0,1,2,3 \tag{54}$$

这两个特征共同赋予狄拉克场与两个标量场相同的多重性(multiplicity)。通常,我们并不施加马约拉纳条件,从而狄拉克场是真正的复场,像两个复标量场一样拥有一个守恒的 $U(1)$ 电荷。

我们简要概括一下狄拉克方程的最主要的特征。4×4 狄拉克矩阵能够很简便地用泡利(Pauli)矩阵的两个对易集 σ_a 与 τ_a 表示。定义为:

$$\sigma_1=\begin{pmatrix}0&1\\1&0\end{pmatrix},\ \sigma_2=\begin{pmatrix}0&-i\\i&0\end{pmatrix},\ \sigma_3=\begin{pmatrix}1&0\\0&-1\end{pmatrix} \tag{55}$$

还有类似的 τ 矩阵,不同的是它们作用于不同的空间:一个狄拉克指标则被看作是一对($i\alpha$)指标 i 与 α,从而矩阵 σ_a 作用于第一个指标 i,矩阵 τ_A 作用于指标 α。我们有:

$$\sigma_a\sigma_b=\delta_{ab}+i\varepsilon_{abc}\sigma_c,\ \tau_A\tau_B=\delta_{AB}+i\varepsilon_{ABC}\tau_C,\ [\sigma_a,\ \tau_B]=0 \tag{56}$$

约定 $g_{\mu\nu}=\mathrm{diag}(-1,\ 1,\ 1,\ 1)$ 并定义:

$$\gamma^1=\sigma_1\tau_1,\ \gamma^2=\sigma_2\tau_1,\ \gamma^3=\sigma_3\tau_1,\ \gamma^0=-i\tau_3 \tag{57}$$

方程(54)中的矩阵 C 则为:

$$C=\gamma_2\gamma_4 \tag{58}$$

在非相对论极限下,狄拉克方程写作:

$$(m+i\gamma^\mu k^\mu)\psi\approx(m-i\gamma^0k^0)\psi\approx m(1-\tau_3)\psi=0 \tag{59}$$

从而四个场分量中仅有两个(那些具有 $\tau_3|\psi\rangle=|\psi\rangle$ 的分量)保留下来。仅仅因为洛伦兹不变性,在相对论粒子情形中也是这样。

3.2 费米—狄拉克(Fermi-Dirac)统计

这时候,我们可以试图继续基本的量子化方案了:给出系统的泊松括号,用对易子取代它们,把系统的哈密顿量重新写作算符形式,并求解得到薛定谔(Schrödinger)方程。

不幸的是,如果人们用的是普通(对易)数,这样并不奏效。与狄拉克方程相关的拉格朗日量将写为:

$$L=\int d^3\vec{x}\,\mathcal{L}(x);\quad \mathcal{L}(x)=-\bar{\psi}(x)\Big(m+\sum_{\mu=0}^{4}\gamma^\mu\partial_\mu\Big)\psi(x) \tag{60}$$

且正则过程将作为动量场给出:

$$p_\psi(\vec{x}) = \frac{\partial \mathcal{L}}{\partial(\partial_0\psi(\vec{x}))}\bar{\psi}(\vec{x})\gamma^0,\quad p_{\bar{\psi}}(\vec{x}) = 0 \tag{61}$$

据此，我们找到哈密顿量：

$$H = \int d^3\vec{x}\,\mathcal{H}(\vec{x});\quad \mathcal{H}(\vec{x}) = p_\psi\dot{\psi} - \mathcal{L}(\vec{x}) = \bar{\psi}(x)\Big(m + \sum_{i=1}^{3}\gamma^i\partial_i\Big)\psi(x) \tag{62}$$

这里，指标 i 是空间的，取值范围为1到3。但是，这并没有下界！这样的一个量子理论并不拥有一个真空态，因而也不适合作为大自然的一个模型。

为了更好地理解这一情形，我们把狄拉克方程化解为其基本的构架。在对角化它之后，我们发现拉格朗日量由具有如下形式的基本单元组成：

$$L = \bar{\psi}(i\partial_t\psi - M\psi);\quad p_\psi = i\bar{\psi};\quad H = \bar{\psi}M\psi \tag{63}$$

若我们采用普通数，则得到 H 下界的唯一方式是把 $\bar{\psi}$ 等同为 ψ。但是这样一来拉格朗日量的动能部分就会成为一个时间微商：

$$\bar{\psi}\partial_t\psi \to \frac{1}{2}\partial_t(\bar{\psi}\psi) \tag{64}$$

因而它并不对作用量有贡献。有人总结说，仅在反对易数空间(space of anti-commuting numbers)中，拉格朗日量(63)才有意义。因而，人们用反对易子取代 ψ 与 $\bar{\psi}$ 的泊松括号：

$$\{\bar{\psi},\psi\} \equiv \bar{\psi}\psi + \psi\bar{\psi} = 1;\quad \{\psi,\psi\} = 0;\quad \{\bar{\psi},\bar{\psi}\} = 0 \tag{65}$$

这一代数的基本表征是在"希尔伯特空间"中只包括两个态(真空态与单粒子态)，其中算符 ψ 和 $\bar{\psi}$ 是作为湮没算符(annihilators)和产生算符(creators)的：

$$\psi = \begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix};\quad \bar{\psi} = \begin{pmatrix}0 & 0\\ 1 & 0\end{pmatrix};\quad H = \begin{pmatrix}0 & 0\\ 0 & M\end{pmatrix} \tag{66}$$

回到非对角化的情形，我们可以保留拉格朗日量(60)与哈密顿量(62)，当对易规则(65)被下式取代时：

$$\begin{gathered}\{\bar{\psi}^i(\mathbf{x}),\ \psi_j(\mathbf{x}')\} = \delta^i_j\delta(\mathbf{x}-\mathbf{x}')\\ \{\psi_i(\mathbf{x}),\ \psi_j(\mathbf{x}')\} = 0;\quad \{\bar{\psi}^i(\mathbf{x}),\ \bar{\psi}^j(\mathbf{x}')\} = 0\end{gathered} \tag{67}$$

对易规则(67)把狄拉克粒子变为了费米子。似乎存在这样一个对任何洛伦兹不变的量子理论而言都是一致的条件，即整数自旋粒子必须是玻色子，具有整数 $+\frac{1}{2}$ 自旋的粒子必须是费米子。

3.3 反对易场的路径积分

现在我们推广路径积分概念以包括狄拉克场。这意味着我们要对反对易数（称为 θ_i）积分，其中 i 是一些指标（可能包括 $\mathbf{x}$）。它们是数，不是算符，因而所有的反对易子为零。考虑对变量 θ 的函数的泰勒展开。因为 $\theta^2=0$，这一展开仅有两个系数：

$$f(\theta)=f(0)+f'(0)\theta \tag{68}$$

所以，这是我们能够拥有的 θ 的最一般函数。一般认为人们应该通过如下假设来定义反对易数 θ 的积分：

$$\int \mathrm{d}\theta\, 1 \equiv 0; \quad d\theta\,\theta \equiv 1 \tag{69}$$

这样定义的原因在于，我们可以用与普通数积分相同的方式来操作这些表达式：

$$\int \mathrm{d}\theta f(\theta+\alpha) = \int \mathrm{d}\theta f(\theta); \quad \int \mathrm{d}\theta \frac{\partial f(\theta)}{\partial \theta} = 0 \text{，等等} \tag{70}$$

现在，考虑仅有一个费米自由度(66)的哈密顿量，我们将其写作：

$$H=M\,b^{\dagger}b; \quad \{b,\ b^{\dagger}\}=1; \quad b^2=(b^{\dagger})^2=0 \tag{71}$$

波函数写为 $\psi=\begin{pmatrix}\psi_0\\ \psi_1\end{pmatrix}$。定义如下关于 θ 的函数：

$$\psi(\theta)\equiv\psi_0\theta+\psi_1 \tag{72}$$

它现在被看作是我们的波函数。不难得到湮没算符 b 与产生算符 $b^{\dagger}$ 是如何作用于这些波函数的：

$$\text{若 } \phi=b\,\psi\text{，则 } \phi(\theta)=\theta\,\psi(\theta) \tag{73}$$

或：

$$b=\theta; \quad b^{\dagger}=\frac{\partial}{\partial\theta} \tag{74}$$

我们现在希望用一个路径积分来表达费米子波函数的演化，正如在 2.4 节中那样。考虑极短的时间间隔 δt。则忽略 $(\delta t)^2$ 阶的所有项，可得：

$$e^{-i\delta t H}\psi(\theta_1) = \psi_0\theta_1 + (1-iM\delta t)\psi_1$$

$$
\begin{aligned}
&= \int d\theta_0 (-\theta_1 + \theta_0 - iM\delta t\theta_0)(\psi_0\theta_0 + \psi_1) \\
&= \int d\theta_0 \int d\bar{\theta} \left(1 + \bar{\theta}(-\theta_1 + \theta_0 - iM\delta t\theta_0)\right)\psi(\theta_0) \qquad (75) \\
&= \int d\theta_0 \int d\bar{\theta}\, e^{\bar{\theta}(-\theta_1+\theta_0-iM\delta t\theta_0)}\psi(\theta_0)
\end{aligned}
$$

对许多无穷小时间间隔 $T=N\delta t$ 重复这一过程，得到形式表达：

$$
\begin{aligned}
\psi(\theta_T) &= \int d\theta_{T-1} d\bar{\theta}_{T-1} \cdots d\theta_0 d\bar{\theta}_0 \\
&\exp \sum_{\tau=0}^{N-1} \delta t \left(\bar{\theta}_\tau \left(\frac{-\theta_{\tau+1} + \theta_\tau}{\delta t} - iM\theta_\tau \right) \right) \psi(\theta_0) \qquad (76)
\end{aligned}
$$

指数变为：

$$
i \int dt\, L(t) \qquad (77)
$$

因而，正如在玻色子情形中，演化算符形式上是对所有(反对易)场 $\psi_i(\mathbf{x}, t)$ 的 e^{iS} 路径积分，其中场的作用量 S 是拉格朗日量 L 的时间积分，且事实上是拉格朗日密度 $\mathcal{L}(\mathbf{x}, t)$ 的时空积分。

在一些应用中，对狄拉克方程中边界条件的仔细考虑要求在作用量(77)上增加一个额外的边界项。这在我们当前的处理中并不重要。

3.4　狄拉克场的费曼规则

令 M_{ij} 为能够对角化的任意矩阵。通过方程(69)，我们找到积分：

$$
\prod_i \int d\theta_i \int d\bar{\theta}_i e^{\bar{\theta}_i M_i \theta_i} = \det_{ij}(M) \qquad (78)
$$

它很容易通过对 M 的对角化得到核实，写作：

$$
\int d\theta \int d\bar{\theta} e^{\bar{\theta} M \theta} = \int d\theta \int d\bar{\theta} (1 + \bar{\theta} M \theta) = M \qquad (79)
$$

因而，对反对易数的高斯积分给出类似于对对易数积分的结果，不同的是我们得到 $\det(M)$，而非 $C/\det(M)$。写作：

$$
\begin{aligned}
M &= M_0 + \delta M \\
\det(M) &= e^{\mathrm{Tr}(\log M)} \\
&= 1 + \mathrm{Tr}(\log M) + \frac{1}{2}(\mathrm{Tr}\log M)^2 + \cdots
\end{aligned}
$$

$$\operatorname{Tr}\log(M) = \operatorname{Tr}\log(M_0) + \operatorname{Tr}\log(1 + M_0^{-1}\delta M) \tag{80}$$

我们看到这可以通过切换该展开中所有奇数项的正负号从 $\det(M^{-1})$ 中得到。因为第 N 项对应于具有 N 个闭费米子圈的费曼图，通过每遇到一个闭费米子圈时切换一次正负号，我们可以推导出费曼规则能够从普通对易场的那些规则中读出。

我们有：

$$-\operatorname{Tr}\log M = -\operatorname{Tr}\log M_0 + \sum_{n=1}^{\infty}\frac{(-1)^n}{n}\operatorname{Tr}(M_0^{-1}\delta M)^n \tag{81}$$

这里，正如在玻色子情形中，$-M_0$ 是理论的传播子，且 δM 表征任意微扰的贡献。因而，如果包括可能的相互作用项的拉格朗日量为：

$$\mathcal{L} = -\bar{\psi}_i(m_{(i)} + \gamma^\mu\partial_\mu)\psi_i + \bar{\psi}_i g_{ij}(\phi)\psi_j \tag{82}$$

则在傅里叶空间中传播子为：

$$(m_{(i)} + i\gamma^\mu k_\mu)^{-1} = \frac{m_{(i)} - i\gamma^\mu k_\mu}{m_{(i)}^2 + k^2 - i\varepsilon} \tag{83}$$

而 $g_{ij}(\phi)$ 产生费曼图中的相互作用顶角。出于相同的原因，即真空态必定是具有最低能量的态，故 $i\varepsilon$ 项的选择如同玻色子理论中那样。

传播子的极点可以用于定义入射和出射粒子，通过在拉格朗日量中添加源项：

$$\delta\mathcal{L} = \bar{\eta}(x)\psi + \bar{\psi}\eta(x) \tag{84}$$

其中 $\eta(x)$ 与 $\bar{\eta}(x)$ 像反对易数那样保持不变。我们能够继续推导出关于自旋向上或向下的入射和出射粒子的精确规则，但把这一步推延到我们讨论 S 矩阵幺正特征的时候更为简便，那里很明确地需要这些规则，也是在那里我们发现这些态归一化的精确方式（第 6 节）。

注意到我们的拉格朗日量在反对易场中总保持为双线性的。这是因为我们坚持$\mathcal{L}$本身必须是一个对易数，且此外那些费米子场中的四次项有太高的维度。我们将在下一节中揭晓为什么要避免这样的项。

4. 规范场

4.1 可重整性

我们继续探索基本场，其洛伦兹协变场方程要能满足我们的量子化方案。原则上，这样的场会是庞加莱(Poincaré)代数的任意表征，即我们能考虑任意种类的张量场，$A_{\mu\nu\lambda}\cdots(\mathbf{x}, t)$。但是，已经得到证明具有多于一个洛伦兹指标的张量不能用。这是因为我们希望场的能量密度有下界，另外我们还希望相互作用的维度充分低，从而所有的耦合强度有零或正的质量维度。

如果一个理论所有的相互作用参数 λ_i(即我们需要对它做微扰展开的所有参数)都有一个为正或零的质量维度，则该理论称为"可重整的"。实践中，耦合系数的维度很容易建立，这将在第 7 节"重整化"中进一步说明。拥有质量维度小于零的耦合强度产生了对短尺度相互作用贡献不可接受的发散表示。人们想要囊括进来的一个主要案例是用度规 $g_{\mu\nu}(x)$ 描述的引力场，但其唯一可能的相互作用是引力的相互作用，它的耦合强度——牛顿常数 G_N——拥有不合适的维度。从而我们得到的不可重整化理论是许多研究的主题，但不属于本论文的范围，可参见罗韦利(Rovelli)的论文。

所以，这里要考虑的只是自旋为 1 的场 $A_\mu^a(x)$。这里，μ 是一个洛伦兹指标，与此同时，场类型的数目通过指标 $a=1, \cdots, N_V$ 来计数。这些场应该描述自旋为 1 的粒子的生成与湮灭。当静止时，这样的粒子将处于三个可能自旋态中的一个。然而，为了得到洛伦兹不变性，矢量场 A_μ 应该包括四个分量。因此，至少其中之一应该是非物理的，虽然人们可能会考虑接受一个与向量粒子相关的附加的无自旋粒子。因而更为重要的是这样的考虑，即在对应的经典理论中，能量应该有下界。

如此一来，这排除了对四矢量场的处理，就好像我们拥有的是四标量场一样，因为洛伦兹不变的产物有一个不确定的度规。我们能够为给出有下界的哈密顿量的矢量场构造一个拉格朗日量吗?

让我们来看这些矢量场之一的高动量极限。在那里拉格朗日量中仅可以留存下来的两项是:

$$\mathcal{L}=-\frac{1}{2}\alpha(\partial_\mu A_\nu)^2+\frac{1}{2}\beta\partial_\mu A_\mu\partial_\nu A_\nu \tag{85}$$

因为具有这一维度的其他项可以通过对作用量的部分积分约化为这两项，而质量项(没有偏微商的项)变得不那么重要。对正则动量场我们有：

$$E^i=\frac{\partial\mathcal{L}}{\partial\partial_0 A_i}=\alpha\partial_0 A_i \quad (i=1,\ 2,\ 3) \tag{86}$$

$$E^0=\frac{\partial\mathcal{L}}{\partial\partial_0 A_0}=(\beta-\alpha)\partial_0 A_0-\beta\partial_i A_i$$

现在，考虑哈密顿量矩阵$\mathcal{H}=E^\mu\partial_0 A_0-\mathcal{L}$。对于所有的场组态 $A_\mu(\mathbf{x},\ t)$，它必定要有下界。让我们首先考虑类空分量 A_i 和所有类空微商 ∂_i 相比于 $\partial_0 A_0$ 可忽略时的情形：

$$\mathcal{H}\to\frac{1}{2}(\beta-\alpha)(\partial_0 A_0)^2 \tag{87}$$

则，当 A_0 和所有时间微商可忽略时：

$$\mathcal{H}\to\frac{1}{2}\alpha(\partial_i A_j)^2-\frac{1}{2}\beta(\partial_i A_j)^2 \tag{88}$$

这些必定都有下界。式(87)要求$\beta\geqslant\alpha$，而式(88)要求 $\alpha\geqslant\beta$。我们得到结论 $\alpha=\beta$，它们都能够归一化为1。因为拉格朗日量中总的微商并没有意义，则我们能够把最初的拉格朗日量(85)重新写作：

$$\mathcal{L}\to-\frac{1}{4}F^a_{\mu\nu}F^a_{\mu\nu},\qquad F^a_{\mu\nu}=\partial_\mu A^a_\nu-\partial_\nu A^a_\mu \tag{89}$$

在认识到这是通常的 QED 的拉格朗日量后，我们知道它的能量矩阵有合适的下界。我们得到结论，每个矢量场理论必定有一个在高能量和高动量时接近方程(89)的拉格朗日量。

我们注意到，选择了 $\alpha=\beta$ 之后式(87)和式(88)都趋于零。事实上，任意可以写为时空梯度 $A^a_\mu=\partial_\mu\Lambda^a(\mathbf{x},\ t)$的场 A^a_μ，都有 $F^a_{\mu\nu}=0$，因而既不对拉格朗日量起作用，也不对哈密顿量起作用。这样的场可以任意的强，但仍然携带零能量。它们可以表示没有能量的粒子和力，这在正统量子场论中是不可接受的。我们该如何避免理论具有这样的特征呢？

确切地说有一种方法可以避免。我们必须保证下述类型的场置换：

$$A^a_\mu\to A^a_\mu+\partial_\mu\Lambda^a(x)+\cdots \tag{90}$$

完全不影响我们描述的所有物理态，这就是我们所谓的定域规范变换。我们必须坚持我们的理论在定域规范变换下不变。方程(90)中的椭圆数表明我们允许并不对拉格朗日量(89)的双线性部分起作用的附加项。因而，我们得到了杨—米尔斯(Yang-Mills)场理论。

4.2　杨—米尔斯方程

我们从上面得到的结论是，每个矢量场都与一个定域规范对称相联系。定域规范群的维度必须等同于 N_V，即出现的矢量场的数目。在矢量场之外，定域对称变换可能也会影响标量场与旋量场。简言之，矢量场必定是杨—米尔斯场。我们这里给出对杨—米尔斯理论的简要概括[Yang and Mills，1954]。

我们有一个定域李群，它在点 x 的元素为 $\Omega(x)$。令矩阵 T^a，$a=1$，…，N_V 为它的无穷小生成元：

$$\Omega(x) = \mathbb{I} + i\sum_a \Lambda^a(x)T^a; \quad T^a = (T^a)^\dagger \tag{91}$$

该群的特征是它的结构常数 f_{abc}：

$$[T^a, T^b] = if_{abc}T^c \tag{92}$$

正如在群理论中所熟知的，我们可以用 f_{abc} 是完全反对称的这样一种方式来选择 T^a 的归一化：

$$f_{abc} = -f_{bac} = f_{bca} \tag{93}$$

通常，旋量场 $\psi(x)$ 与标量场 $\phi(x)$ 以这样一种方式被引入，即它们作为规范群的表示(不可约集)来变换。则一个定域规范变换为：

$$\psi'(x) = \Omega(x)\psi(x); \qquad \phi'(x) = \Omega(x)\phi(x) \tag{94}$$

用无穷小形式来表示为：

$$\psi'(x) = \psi(x) + i\Lambda^a(x)T^a\psi(x) + \mathcal{O}(\Lambda)^2 \tag{95}$$

$\phi(x)$ 也类似。不可约表征的维度可以随场的种类不同而不同，因而标量场与旋量场常常形成了不同维度的规范矢量。在这里和之后的表示中，我们省略了标示场 ψ，Ω 与 T^a 不同分量的指标。

通过要求构建这些场的规范—协变梯度的可能性，就很方便地引入了我们的矢量场 $A_\mu^a(x)$：

$$D_\mu\psi(x)\equiv(\partial_\mu+igA_\mu^a(x)T^a)\psi(x) \tag{96}$$

其中 g 是可自由调整的耦合参数。对表示矢量场不同种类的重复指标 a 求和。通过要求变换规则：

$$(D_\mu\psi(X))'=\Omega(x)D_\mu\psi(x)=D_\mu\psi(x)+i\Lambda^a(x)T^aD_\mu\psi(x)+\mathcal{O}(\Lambda)^2 \tag{97}$$

人们很容易得到矢量场 $A_\mu^a(x)$ 的变换规则：

$$\begin{aligned}igA_\mu^{a\prime}(x)T^a&=\Omega(x)(\partial_\mu+igA_\mu^a(x)T^a)\Omega^{-1}(x)\\&=igA_\mu^a(x)T^a-i\partial_\mu\Lambda^a(x)T^a+g[T^a,\ T^b]\Lambda^a(x)A_\mu^b(x)\end{aligned} \tag{98}$$

忽略掉 $\mathcal{O}(\Lambda)^2$ 项，则在方程(92)下，它变为：

$$A_\mu^{a\prime}(x)=A_\mu^a(x)-\frac{1}{g}\partial_\mu\Lambda^a(x)+f_{abc}\Lambda^b(x)A_\mu^c(x) \tag{99}$$

如果我们保证所有用到的梯度都是协变梯度，我们就能够直接构建定域规范不变的标量和旋量场的拉格朗日量的一般表达：

$$\mathcal{L}_{\text{scalar}}^{\text{inv}}(x)=-\frac{1}{2}(D_\mu\phi)^2-V(\phi^2) \tag{100}$$

$$\mathcal{L}_{\text{Dirac}}^{\text{inv}}(x)=-\overline{\psi}(\gamma^\mu D_\mu+m)\psi \tag{101}$$

此外也能够构造其他可能的没有微商的不变定域相互作用项。

两个协变微商的对易子为：

$$\begin{aligned}&[D_\mu,\ D_\nu]\psi(x)=igF_{\mu\nu}^a(x)T^a\psi(x)\\&F_{\mu\nu}^a(x)=\partial_\mu A_\nu^a-\partial_\nu A_\mu^a+gf_{abc}A_\mu^bA_\nu^c\end{aligned} \tag{102}$$

不像 $A_\mu^a(x)$ 或 $A_\mu^a(x)$ 的直接梯度，这个杨—米尔斯场 $F_{\mu\nu}^a$ 的变换就像定域规范群的真正共轭表征那样：

$$F_{\mu\nu}^{a}{}'(x)=F_{\mu\nu}^a(x)+f_{abc}\Lambda^b(x)F_{\mu\nu}^c(x) \tag{103}$$

这允许我们为矢量场构建一个定域规范不变的拉格朗日量：

$$\mathcal{L}_{\text{YM}}^{\text{inv}}(x)=-\frac{1}{4}F_{\mu\nu}^a(x)F_{\mu\nu}^a(x) \tag{104}$$

场 $F_{\mu\nu}$ 的定义(102)中的结构常数 f_{abc} 意味着在杨—米尔斯拉格朗日量(104)中的相互作用量会出现。若 f_{abc} 不为零，则我们谈论的是一个非阿贝尔(Abelian)规范理论。

在费米子情形中存在一个重要的难题，即狄拉克矩阵 $\gamma^5\equiv\gamma^1\gamma^2\gamma^3\gamma^4$ 可以用

于展示手征(不对称)部分(chiral sectors):

$$\psi \equiv \psi_L + \psi_R; \quad \psi_L = \frac{1}{2}(1+\gamma^5)\psi; \quad \psi_R = \frac{1}{2}(1-\gamma^5)\psi \tag{105}$$

因为狄拉克拉格朗日量的动能部分可以按照下式分开:

$$\mathcal{L}_{\mathrm{Dirac}} = -\overline{\psi}_L(\gamma D)\psi_L - \overline{\psi}_R(\gamma D)\psi_R \tag{106}$$

所以我们可以选择左旋场(left-han-ded fields)ψ_L 在表征上和右旋场(right-han-ded)ψ_R 不同。然而，因为一个质量项将左右连起来了:

$$-m\overline{\psi}\psi = -m\overline{\psi}_L\psi_R - m\overline{\psi}_R\psi_L \tag{107}$$

这样的项将被禁止，因而这样的手征场必须是无质量的。其次，并非所有手征费米子的组合都是允许的，第 8 节中将会讨论到一个重要的限制。已证明场 ψ_L 描述了仅有左旋螺旋性的自旋$\frac{1}{2}$无质量粒子，而 $\overline{\psi}_L$ 描述了它们的反粒子，即仅具有右旋螺旋性。

4.3　对定域规范不变性的需求

在规范理论的早期，人们认为定域规范不变性会是一种“近似的”对称。或许人们可以为违背定域对称的矢量场添加质量项，但这使得模型看起来更像在粒子物理学中观测到的情形。但是，现在我们知道，这样的模型遇到了一个严重的缺陷，即它们是不可重整化的。原因在于，可重整化性要求我们的理论直到最微小的距离尺度上都是一致的。至少在原则上，一个质量项会把贡献于方程(99)中用$\Lambda(x)$描述的场组态转变为物理上可观测的场，拉格朗日量现在依赖于 $\Lambda(x)$。但是，因为 $\Lambda(x)$缺乏动力学项，激烈振荡的 Λ 场不携带巨大的能量，故它们不大可能会受到能量守恒的抑制。不受控制的短距离振荡是理论成为不可重整化的真正的物理原因。

类似地，时空度规不可控的短距离涨落引起了广义相对论的量子化版本(“量子引力”)的不可重整化。为了修正这样的理论需要更激烈的措施。

因为可重整化性为我们的理论提供了所需的连贯性，用方程(94)与(99)描

述的定域规范对称必须是所有量子场论的精确的(而非近似的)对称。① 显然，在亚原子世界中绝大多数矢量粒子都携带有质量，这一事实必须以某种其他方式得到说明。在这里用到的是布劳特—恩格勒—希格斯(Brout-Englert-Higgs)机制，见第5节。

4.4 规范固定

矢量场的纵向部分并不直接出现在杨—米尔斯拉格朗日量(104)中，这完全是因为它在形如式(99)的变换下的不变性。然而如若我们想要描述解的话，我们需要选择一个纵向分量。这就是为什么我们希望同时在经典与量子化的理论中对我们关于解的描述施加一些附加的限制，即所谓的规范条件。在电动力学中，我们常常施加诸如 $\partial_\mu A_\mu(x)=0$ 或 $A_0=0$ 的限制。在杨—米尔斯理论中，指标 a 的每一个值都需要这样一个限制。被置等于零的场 $C^a(x)$ 表示了一个规范固定项：

$$C^a(x)=0;\quad a=1,\cdots,N_V;\ \text{其中} \tag{108}$$

$$\text{要么 } C^a(x)=\partial_\mu A_\mu^a(x)\,(\text{费曼规范}) \tag{109}$$

$$\text{要么 } C^a(x)=A_0^a(x)\,(\text{类时规范}) \tag{110}$$

或是其他可能的规范选择。总可以找到一个满足这些条件之一的 $\Lambda^a(x)$。例如，为了得到费曼规范(109)，所有我们需要去做的就是在规范群的变分下取一个积分的极值：

$$\delta\int \mathrm{d}^4x(A_\mu^a(x))^2=0\rightarrow\partial_\mu A_\mu^a(x)=0 \tag{111}$$

对经典理论而言，施加该规范条件最为简洁的方式是在拉格朗日量上增加一个拉格朗日乘数项：

$$\mathcal{L}(x)=\mathcal{L}^{\text{inv}}(x)+\lambda^a(x)C^a(x) \tag{112}$$

其中 $C^a(x)$ 是任意可能的规范固定项，$\lambda^a(x)$ 为自由的动力学变量。这里，$\mathcal{L}^{\text{inv}}$ 代表拉格朗日量中所有规范不变的项的集合。则经典理论的欧拉—拉格朗日方

① 一个显然的例外是如下情形，在那里纵向分量完全退耦合，这发生在大质量的QED中。但即使在该情形中，也最好将纵光子看作是一个希格斯场，见第5节。

程必然地产生了加有限制的杨—米尔斯场方程，除了一个小细节：边界条件。因为$\mathcal{L}^{\text{inv}}$并不改变，改变规范变换之后人们发现 $D_\mu\lambda^a(x)=0$。我们需要施加更严格的方程 $\lambda^a(x)=0$，它是通过在我们系统的边界处施加 $\lambda^a(x)=0$ 而获得的。

替代性地，我们可以用下式取代不变的拉格朗日量：

$$\mathcal{L}(x)=\mathcal{L}^{\text{inv}}(x)-\frac{1}{2}(C^a(x))^2 \tag{113}$$

它具有这样的优点，即在部分积分之后双线性部分变得非常简单：$\mathcal{L}=-\frac{1}{4}F_{\mu\nu}F_{\mu\nu}-\frac{1}{2}(\partial_\mu A^\mu)^2\rightarrow-\frac{1}{2}(\partial_\mu A_\nu)^2$，从而矢量场可以被当作好像它只是 4 个标量那样来处理。再一次，改变规范变换 $\Lambda^a(x)$，人们发现有 $D_\mu C^a(x)=0$，通过加上恰当的边界条件，则它必须被更为严格的条件 $C^a(x)=0$ 所取代。

注意拉格朗日—哈密顿体系会给一些场分量的能量以错误正负号，我们应该继续采用施加规范限制之前推导的能量。若我们采用类时规范(110)，能量是对的，但理论似乎缺乏洛伦兹不变性。洛伦兹变换现在必须与规范变换共存。

在量子化的理论中该如何处理规范限制？这一问题已经由德维特(B. S. DeWitt)[1964；1967a；1967b]与法捷耶夫和波波夫(Faddeev and Popov)[1967；1969；1984]予以解决。规范限制被施加到函数积分的被积函数中：

$$Z=\int\mathcal{D}A(x)\int\mathcal{D}\phi(x)\cdots e^{i\int d^4x\mathcal{L}^{\text{inv}}(x)}\prod_{a,x}\delta(C^a(x))\Delta\{A,\phi\} \tag{114}$$

因而，我们仅对那些满足规范条件的场组态进行积分。$\Delta\{A,\phi\}$是一个雅可比(Jacobian)因子，稍后我们会讨论到它。形式上的德尔塔函数可以用拉格朗日乘数来代替：

$$\int\mathcal{D}\lambda^a(x)e^{i\int d^4x\lambda^a(x)C^a(x)} \tag{115}$$

事实上，如果$\lambda^a(x)$仅被简单地添加进理论动力学变量的列表中，费曼规则就能够明确地得到，正如在标量场和旋量场中那样。

但是，仍然有一个问题，似乎很难证明规范不变性。更精确地说，我们需要确定的是，如果我们转换到一个不同的规范固定函数 $C^a(x)$，理论的物理内

容——特别是散射矩阵——保持不变。这一困难与积分测度有关系，它不是规范不变的，除非我们在方程(114)中添加附加项 $\Delta\{A, \phi\}$。这一项与无穷小规范变换的体积元(volume)相关联。假定场组合 $C^a(x)$ 在一个规范变换下的变换为：

$$C^{a\prime}(x) = C^a(x) + \frac{\partial C^a(x)}{\partial \Lambda^b(x')}\Lambda^b(x') \tag{116}$$

则需要的体积元项是雅可比量：

$$\Delta\{A, \phi\} = \det\left(\frac{\partial C^a(x)}{\partial \Lambda^b(x')}\right) \tag{117}$$

通过采用3.4节中的评论，即在反对易变量上的高斯积分给出一个行列式，见方程(78)，就能很巧妙地计算出此行列式。所以，我们引入反对易标量场 η 与 $\overline{\eta}$，则写作：

$$(117) = \int \mathcal{D}\eta^a(x)\int \mathcal{D}\overline{\eta}^a(x)\exp\left(\overline{\eta}^a(x)\frac{\partial C^a(x)}{\partial \Lambda^b(x')}\eta^b(x')\right) \tag{118}$$

这就是作用量中所谓的法捷耶夫—波波夫项。把所有的这些项结合放一起，我们得到杨—米尔斯理论的下述作用量：

$$\mathcal{L}(x) = \mathcal{L}^{\text{inv}}(x) + \lambda^a(x)C^a(x) + \overline{\eta}^a(x)\frac{\partial C^a(x)}{\partial \Lambda^b(x')}\eta^b(x') \tag{119}$$

同样也可能找到经典拉格朗日量(113)的量子类似。首先，用 $C^a(x) - F^a(x)$ 取代 $C^a(x)$，其中 $F^a(x)$ 是函数积分(119)中一个固定的但依赖 x 的量。物理效应应该完全独立于 $F^a(x)$。因此，我们能够采用我们想要的任意权重因子对 $F^a(x)$ 进行函数积分。选择权重因子 $e^{-\frac{1}{2}\int d^4x(Fa(x))^2}$。拉格朗日乘数 $\lambda^a(x)$ 现在只不过要求 $C^a(x)$ 等同于 $F^a(x)$。我们最后得到有效拉格朗日量：①

$$\mathcal{L}(x) = \mathcal{L}^{\text{inv}}(x) - \frac{1}{2}(C^a(x))^2 + \overline{\eta}^a(x)\frac{\partial C^a(x)}{\partial \Lambda^b(x')}\eta^b(x') \tag{120}$$

这是规范理论中最常用到的拉格朗日量。与标量和旋量场的拉格朗日量相比，这里不是所有的场都表征物理粒子。矢量场的纵向部分和费米子的标量场 $\overline{\eta}$ 与 η 是“鬼场(ghost)”。

---

① 人们通常将方程(99)的因子 $1/g$ 合并到 $\overline{\eta}$ 场的定义中。

4.5　费曼规则

用微扰理论来计算散射矩阵元时需要用到费曼规则，该规则能够在规范固定的拉格朗日量(119)或(120)中直接读出。在这两种情况下，我们首先分离双线性部分，① 把拉格朗日量写作：

$$\mathcal{L} = -A_\alpha(x)\hat{M}_{\alpha\beta}A_\beta(x) - \overline{\psi}_\alpha(x)\hat{D}_{\alpha\beta}\psi_\beta(x) + \mathcal{L}^{\text{int}} \tag{121}$$

其中$\mathcal{L}^{\text{int}}$包含了所有的三线性和四线性项。这里，$A_\alpha(x)$是对所有玻色子(标量和矢量)场的缩写，$\overline{\psi}$ 和 ψ 是对狄拉克费米子和法捷耶夫—波波夫费米子的缩写。系数 $\hat{M}_{\alpha\beta}$，$\hat{D}_{\alpha\beta}$和三线性系数可能包含有梯度算符 $\partial/\partial x^\mu$。在傅里叶展开之后，它将变为一个因子 ik_μ。

——传播子 $\hat{P}_{\alpha\beta}$和 $\hat{P}^{\text{ferm}}_{\alpha\beta}$将会是系数 $\hat{M}-i\varepsilon$ 和 $\hat{D}-i\varepsilon$ 的逆，所以：

$$\text{若 } \hat{M}_{\alpha\beta} = (m_{(\alpha)} - \partial_\mu^2)\delta_{\alpha\beta}，\text{则 } \hat{P}_{\alpha\beta} = \frac{\delta_{\alpha\beta}}{m_{(\alpha)}^2 + k^2 - i\varepsilon}$$

$$\text{若 } \hat{D}_{\alpha\beta} = (m_{(\alpha)} + \gamma^\mu\partial_\mu)\delta_{\alpha\beta}，\text{则 } \hat{P}^{\text{ferm}}_{\alpha\beta} = \frac{(m_{(\alpha)} - i\gamma^\mu k_\mu)\delta_{\alpha\beta}}{m_{(\alpha)}^2 + k^2 - i\varepsilon} \tag{122}$$

——正如2.5 节中那样，顶角由$\mathcal{L}^{\text{int}}$的三线性和四线性项产生。若我们有类似于 $J^a(x)\phi_a(x)$，$\overline{\eta}^i(x)\psi_i(x)$或$\overline{\psi}^i(x)\eta_i(x)$的源项，则它们对应于以点结束的传播子，在那里动量 k 要与源的给定傅里叶分量相匹配。所有这些都能够很好地从如(47)的函数积分的形式展开中读出。

——对每个费米子的闭圈，存在一个总的负号。

——每个图都伴随有如 $1/k!$ 和$(2\pi)^{-4N}$的正则系数，其中 $k!$ 是图的内部对称群的维度，N 是圈积分的数目。这些系数能够通过对比函数积分与普通积分而得到。

——对每个外线都存在一个归一化系数，它依赖于为入射和出射粒子选择的波函数。我们将在第 6 节中再回到这一问题上。

---

① 人们可能决定将对拉格朗日量双线性部分的微小修正留待与高阶项一起处理，就好像它们是“两点顶角”。

注意，能够写为某(定域地确定的)场组态梯度的拉格朗日量中的任何项都能够用零来代替。这是因为(在足够仔细选择的边界条件下)这样的项并不对总的作用量 $S=\int \mathrm{d}^4x\,\mathcal{L}(x)$ 有贡献。

4.6 BRST 对称

正如读者可能注意到的，我们偏离了我们原先的计划，即把时空留在点阵中，仅在计算结束时才转到连续统极限。我们甚至都没有开始做计算，就已经建立起了费曼规则，好像场建立在时空连续统上。事实上，我们应该保持时空为离散的，从而函数积分只不过是空间中的普通积分，该空间具有很大但仍然有限的维度。但是在实践中，连续统要容易处理得多，所以通常我们不明确提到时空的有限尺寸网格(meshes)。

我们将在第 7 节中首次尝试构造连续统的极限。之后我们会看到拉格朗日量(120)中的系数必须被重整化。则出现了下述问题：

如果我们看到一个看似(120)的拉格朗日量，我们如何能够检查它的系数是一个真正规范理论所具有的那些系数呢?

这一问题的答案是，规范固定的拉格朗日量(4.35)和(4.36)拥有一种对称性。最初对所涉对称性的辨识并未成功，因为鬼场是费米子的，而规范固定项是玻色子的。在早些时候我们认为在规范固定项与鬼场项之间所需的关系必须通过检验来核实['t Hooft and Veltman(特霍夫特和韦尔特曼)，1972a]。但之后拜奇(Becchi)、罗埃特(Rouet)和斯道拉(Stora)[1975；1976]，还有秋京(Tyutin)[Tyutin，1975]独立发现了完备的答案。这种称为 BRST 的对称性是一种超对称。对于比式(119)更一般的拉格朗日量(120)，变换规则为：

$$A_\alpha'(x)=A_\alpha(x)+\bar{\varepsilon}\,\frac{\partial A_\alpha(x)}{\partial\Lambda^b(x')}\eta^b(x') \qquad (a) \qquad (123)$$

$$\eta^{a\prime}(x)=\eta^a(x)+\frac{1}{2}\bar{\varepsilon}f_{abc}\eta^b(x)\eta^c(x) \qquad (b)$$

$$\bar{\eta}^{a\prime}(x)=\bar{\eta}^a(x)+\bar{\varepsilon}C^a(x) \qquad (c)$$

其中反对易数 $\bar{\varepsilon}$ 是此(全域的)超对称变换的无穷小生成元。

在这一超对称变换下拉格朗日量(120)的不变性很容易检验，除了或许对

最后一项变分的取消不影响(123b):

$$\overline{\eta}^{a}\frac{\partial C^{a}}{\partial \Lambda^{b}}\frac{1}{2}\bar{\varepsilon} f_{bcd}\eta^{c}\eta^{d}+\overline{\eta}^{a}\%,\frac{\partial}{\partial \Lambda^{c}}\frac{\partial C^{a}}{\partial \Lambda^{d}}\eta^{c}\eta^{d}=\cdots \tag{124}$$

为规范限制函数 C^a 代入一些实例，人们发现这些项总是能被排除。式(124)为零的原因是基于这样的事实，即规范变换形成了一个群，该群意味着雅可比恒等式:

$$f_{asb}f_{scd}+f_{asc}f_{sdb}+f_{asd}f_{sbc}=0 \tag{125}$$

反过来却更加难以证明：若一个理论在形如式(123)的变换下不变(BRST 不变性)，则它是一个规范固定的定域规范理论。在实践中真正需要的是要表明鬼粒子并不对 S 矩阵起作用。这事实上源自于 BRST 不变性，通过所谓的斯拉夫诺夫—泰勒(Slavnov-Taylor)恒等式[Slavnov, 1972; Taylor, 1971]而得到，该恒等式是产生自此对称性的振幅之间的关系。

5. 布劳特—恩格勒—希格斯机制

按照上面描述的方式，杨—米尔斯规范理论似乎并不适用于描述具有自旋为 1 的有质量粒子。但在我们的方法中，通过假定拉格朗日量取式(85)的形式我们仅关注于矢量粒子的高能、高动量极限理论。质量项集中在红外，或低能区域。在这里，人们可能会注意到我们还没有穷尽所有的可能性。

我们需要施加精确的定域规范不变性，正如在 4.3 节中说明的。所以我们的理论必须沿着在 4.2 节中论述的那条线索来构建。所有的标量场和旋量场必须作为规范群的表征。这样一来，我们忽略了什么?

在我们对最普遍的、定域规范不变的拉格朗日量的描述中，潜在地假定了标量势函数 $V(\phi)$ 的最小值出现在 $\phi=0$，从而，正如在全域对称性中那样，粒子谱中的对称性很明显：物理粒子恰似完全定域对称群的表征。但是，正如我们在 2.2 节中一个全域对称性情形中看到的，势的最小值可能发生在其他 ϕ 值处。若这些值在规范群下不是不变的，则它们形成了群的一个非平凡表征，仅在规范群的一个子群下不变。若该不变子群完全非平凡的话，物理粒子正是形成了对不变子群的表征，但其余的对称性都难以发现。事实上，如果我们不考虑与矢量场的耦合， 会再次得到 2.2 节中描述的情形。正如我们在那里强调

的，粒子谱包含无质量的粒子，即戈德斯通玻色子。这些戈德斯通玻色子表征了与全域对称变换相关的场激发，它确实并不影响能量：所以没有质量。

但是，全域规范戈德斯通玻色子确实携带了一个动能项。因而，它们在以光速运动时带走了能量。这是因为全域对称性仅规定了如果场是时空无关的话，戈德斯通场不携带能量。

相反地，定域规范对称性要求，即使戈德斯通场确实依赖于时空的话也不携带有能量。因而，在定域对称情形下，戈德斯通模完全在理论的鬼场部分中，戈德斯通粒子则是非物理的。让我们在案例中看看它是如何发生的。

5.1 $SO(3)$情形

作为一个原型，我们取群 $SO(3)$ 为我们的定域规范群，且出于简单性的考虑我们忽略圈图的贡献，它表征了对场方程的高阶量子修正。令标量场 ϕ_a 在三维空间表征(3-representation)中。则拉格朗日量的不变部分为：

$$\mathcal{L}^{\text{inv}} = -\frac{1}{4}(F^a_{\mu\nu})^2 - \frac{1}{2}(D_\mu\phi_a)^2 - V(\phi); \qquad V(\phi) = \frac{1}{8}\lambda\left((\phi_a)^2 - F^2\right)^2 \quad (126)$$

这里，D_μ 表示协变微商：$D_\mu\phi_a = \partial_\mu\phi_a + g\epsilon_{abc}A^b_\mu\phi_c$。正如在 2.2 节中的方程(18)，我们如下来定义转换场 $\tilde{\phi}_a$：

$$\phi_a \equiv \tilde{\phi}_a + \begin{pmatrix} 0 \\ 0 \\ F \end{pmatrix}; \qquad V(\tilde{\phi}) = \frac{1}{2}\lambda F^2\tilde{\phi}_3^2 + \frac{1}{2}\lambda F\tilde{\phi}^2\tilde{\phi}_3 + \frac{1}{8}\lambda(\tilde{\phi}^2)^2 \quad (127)$$

ϕ 的动能项也必须进行转换：

$$D_\mu\phi_a = D_\mu\tilde{\phi}_a + gF\begin{pmatrix} A^2_\mu \\ -A^1_\mu \end{pmatrix}; \qquad -\frac{1}{2}(D_\mu\phi_a)^2 =$$

$$-\frac{1}{2}(D_\mu\tilde{\phi}_a)^2 - gF(A^2_\mu D_\mu\tilde{\phi}_1 - A^1_\mu D_\mu\tilde{\phi}_2) - \frac{1}{2}g^2F^2(A^{1\,2}_\mu + A^{2\,2}_\mu) \quad (128)$$

定义复场：

$$\tilde{\Phi} = \frac{1}{\sqrt{2}}(\tilde{\phi}_1 + i\,\tilde{\phi}_2); \qquad \mathcal{A}_\mu = \frac{1}{\sqrt{2}}(A^1_\mu + iA^2_\mu)$$

$$D_\mu\tilde{\Phi} = (\partial_\mu + iA^3_\mu)\tilde{\Phi} - i\,\mathcal{A}_\mu\tilde{\phi}_3 \quad (129)$$

我们看到拉格朗日量(126)变为：

$$\mathcal{L}^{\text{inv}} = -\frac{1}{4}(F^a_{\mu\nu})^2 - \frac{1}{2}(D_\mu\tilde{\phi}_3)^2 - D_\mu\tilde{\Phi}^* D_\mu\tilde{\Phi})$$

$$-\frac{1}{2}M_H^2\tilde{\phi}_3^2 - M_V^2\mathcal{A}^*_\mu\mathcal{A}_\mu + M_V\Im(\mathcal{A}^*_\mu D_\mu\tilde{\Phi}) - V^{\text{int}}(\tilde{\phi})$$

其中，$M_H=\sqrt{\lambda}F$；$M_V=gF$，且 V^{int}为势项的其余部分。$\Im$表示虚部。　　(130)

因而，标量场的"中性"分量即希格斯粒子，获得质量 M_H，见方程(127)，且矢量场的"荷"分量得到具有质量为 M_V 的质量项。去除(某些)戈德斯通玻色子并且为矢量粒子生成质量的机制，称为布劳特—恩格勒—希格斯(BEH)机制[Englert and Brout，1964；Higgs，1964b；Higgs，1964a；Higgs，1966]。在每个方面，矢量场的中性的、无质量的分量的行为就像是电磁场的矢量势，且复的矢量粒子带有电荷。

5.2　规范固定

如果人们想试着用4.5节中的规则直接从$\mathcal{L}^{\text{inv}}$得到费曼规则的话，他会发现描述拉格朗日量双线性部分的矩阵 $\hat{M}$ 没有逆，这是因为规范首先必须是固定的(fixed)。选择 $\partial_\mu A^a_\mu(x)=0$ 有这样的优势，即某些棘手的项$\Im(\mathcal{A}^*_\mu\partial_\mu\tilde{\Phi})$可以通过部分积分变为零。则(在动量空间中)矢量传播子很容易计算出是：

$$P^{ab}_{\mu\nu}(k)=\frac{\delta_{\mu\nu}-k_\mu k_\mu/(k^2-i\varepsilon)}{k^2+m^2_{(a)}-i\varepsilon}\delta_{ab} \tag{131}$$

其中，对于荷矢量场有 $m_{(a)}=M_V$，对于中性场为零。

这似乎描述了具有质量 $m_{(a)}$与附加横截性(transversality)限制的矢量粒子。但是，我们可以做得更漂亮。如果在规范固定的拉格朗日量(120)中，我们选择：

$$C^3=\partial_\mu A^3_\mu;\quad C^1=\partial_\mu A^1_\mu - M_V\tilde{\phi}_2;\quad C^2=\partial_\mu A^2_\mu + M_V\tilde{\phi}_1 \tag{132}$$

则我们发现标量—矢量混合项被抵消，但现在$(\partial_\mu A_\mu)^2$ 项也被抵消了，所以矢量传播子失去了它的 $k_\mu k_\nu$ 项。则向量传播子为：

$$P^{ab}_{\mu\nu}(k)=\frac{\delta_{\mu\nu}\delta_{ab}}{k^2+m^2_{(a)}-i\varepsilon} \tag{133}$$

且荷标量鬼场获得了质量 M_V，物理场$\tilde{\phi}_3$ 不受影响。

在此规范中计算法捷耶夫—波波夫鬼场的拉格朗日量是有益的。人们很容易发现它为：

$$\mathcal{L}^{\text{ghost}}=\bar{\eta}^{a}\partial^{2}\eta^{a}-M_{V}^{2}(\bar{\eta}^{1}\eta^{1}+\bar{\eta}^{2}\eta^{2})+\text{相互作用项} \tag{134}$$

正如可以通过更精确的计算所确证的，该理论具有一个质量为 M_V 的物理的、带荷的矢量粒子，一个中性的(无质量的)光子和一个质量为 M_H 的中性标量粒子。后者称为该理论的希格斯粒子。拉格朗日量中的所有其他场描述了鬼场。显然，在上面描述的规范中，所有“非物理”的带荷粒子、鬼场、矢量场的类空分量，还有戈德斯通玻色子，都有相同的质量 M_V，非物理的中性粒子的质量都为零。

人们将这一例子的对称性形态总结如下：定域规范群 $SO(3)$ 通过布劳特—恩格勒—希格斯机制破缺为其子群 $SO(2)$(围绕一固定轴的旋转，由 ϕ_a 的真空值形成)，或破缺为等价的 $U(1)$。因而，三个矢量玻色子中的两个获得了质量，而剩下一个无质量的 $U(1)$ 光子。与此同时，三个标量中的两个变为鬼粒子，第三个变为了希格斯粒子。

布劳特—恩格勒—希格斯机制并不改变粒子谱中独立物理态的总数目。在我们的例子中，三个标量粒子中的两个消失，但两个有质量的自旋 1 粒子现在每个都有三个自旋螺旋度，而无质量的光子仅有两个自旋螺旋度。

5.3 与其他场的耦合

在场定义中的转换(127)，给所有的相互作用以一个非对称表象。这就是为什么在文献中人们谈论“定域对称性的自发破缺”。事实上，这是某种名称的误用。在全域对称情形中，自发破缺指的是真空态是简并的。在一个全域对称变换之后，真空态被变换为一个物理上不等价的真空态，而后者不会在系统中被实现。无质量的戈德斯通玻色子的存在证明了这一点。在定域对称情形中，不会发生这类事情。因为在那里仅存在一个真空态，它总是在定域对称下不变的。这就是戈德斯通玻色子之所以是非物理的原因。事实上，所有的物理态在定域规范变换下都是形式不变的。当然，该规则明显的例外是 QED 中的荷粒子，但这是因为我们总是希望忽略它们与无限远处矢量势的相互作用。在现实中，由于荷粒子的长程相互作用，使得人们很难对它们进行充分讨论。

在所有这些观点中，最好不去说一个局域对称性是自发破缺的，而是谈论布劳特—恩格勒—希格斯机制[Englert and Brout，1964；Higgs，1964b]，即物理粒子谱并不形成局域对称群的一个表征这一现象。局域对称性只有通过把标量场转换回它们对称性的记号，即最初的场 ϕ 才能够被识别。局域对称性必须不被看作是物理态的一个特征，而是看作为我们对物理态描述方式的特征。

但是，如果我们对规范场耦合的小值做一个微扰展开的话，我们会发现在规范耦合为零处局域对称性自发地破缺了。所以，通过罗列规范群和它们破缺形成的子群来刻画我们的微扰描述仍然是相当有用的。

现在，让我们假定存在其他的场，如狄拉克费米子 ψ_i。在对称记号中，它们必定形成了局域规范群的一个表征。所以，我们有：

$$\mathcal{L}^{\text{Dirac}} = -\overline{\psi}^i(\gamma_\mu D_\mu + m_{(i)})\psi_i - \overline{\psi}^i g_Y t^a_{ij}\phi_a\psi_j \tag{135}$$

其中 D_μ 是恰当的协变微商，它包含了那些适用于给定表征的矩阵 T^a，参见式(95)和式(96)，且 g_Y 代表一个或更多的汤川(Yukawa)耦合参数。质量项 $m_{(i)}$ 和耦合系数 t^a_{ij}是规范群的不变张量，仅当费米子不是手征的时候质量才是允许的，见方程(106)之后的讨论。

这里，我们再一次从更显而易见的对称场 ϕ_a 开始，但物理场$\tilde{\phi}$是通过转换 $\phi_a = F_a + \tilde{\phi}_a$ 而获得的。因而，狄拉克拉格朗日量最低阶的双线性部分变为：

$$\mathcal{L}^{\text{Dirac}} \to -\overline{\psi}^i\left((\gamma_\mu\partial_\mu + m_{(i)})\delta_{ij} + g_Y t^a_{ij}F_a\right)\psi_j \tag{136}$$

特别地，若对称性确实作用于费米场的手征部分，则质量项 $m_{(i)}$ 被禁止，但不那么对称的第二项可能会产生质量，且在任何时候该质量都不同于费米子的质量。因而，不只是矢量粒子与标量粒子不再形成对最初定域规范群的表征，费米子也不能这样做。

5.4 标准模型

现在被称为“标准模型”的仅仅是希格斯理论的一个案例。规范群为 $SU(3)\times SU(2)\times U(1)$。这意味着矢量场集分解为三个群：8 相关于 $SU(3)$，3 相关于 $SU(2)$，最后 1 相关于 $U(1)$。标量场 ϕ_i 形成一个对这三个群中两个的二维复表征：它在 $SU(2)$下是二重态，在 $U(1)$下作为一个具有荷$\frac{1}{2}$的粒子在旋转。

用四个实场分量来表征希格斯标量，会发现布劳特—恩格勒—希格斯机制消除了其中的三个，留下一个中性的、物理的希格斯粒子。$SU(2)\times U(1)$破缺为一个对角化的子群$U(1)$。四个规范场中的三个获得了一个质量。留存下来的那个光子场是在对矢量场重新对角化之后而得到的，它是初始的$U(1)$场和$SU(2)$规范场三分量之一的线性组合。

$SU(3)$群并没有受到布劳特—恩格勒—希格斯机制的影响，所以人们可以期望所有的"物理的"粒子出现在$SU(3)$的表征中。取而代之会发生什么将在第11节中解释，即仅有场的规范不变性组合那样是可观测的，就像粒子一样在我们的探测器中那样。

标准模型中的费米子形成了三个"族"。在每个族中，我们看到相同的形态。左手场ψ_L在$SU(2)$下形成了二重态，在$SU(3)$下形成了一个三重态("夸克")和单态("轻子")的一个组合。右手分量ψ_R在$SU(3)$下形成了相同的表征，但在$SU(2)$下形成两个单态的对，所以它们并不耦合于$SU(2)$矢量场。左手$SU(2)$二重态的$U(1)$荷对轻子是$-\frac{1}{2}$，对夸克为$\frac{1}{6}$；右手单态的$U(1)$荷为-1和0(轻子)或$-\frac{1}{3}$和$\frac{2}{3}$(夸克)。

标准模型的结构应归功于拥有希格斯标量的不同种可能的汤川相互作用项。它们都具有$\bar{\psi}\phi\psi$形式，在完全规范群下不变，但因为存在三族费米子，每一族都有左手征和右手征分量，所以仍然存在相当多的这些项，每一项都描述了一个相互作用强度，我们理论中的原理并不对该强度的值做任何要求[Hoddeson *et al.*(霍德森等)，1997]。

6. 幺正性

正如我们在4.5节中看到的，费曼规则很明显是从我们对理论拉格朗日量的表达中得到的。更精确地，在那里得到的是一组用于可能的源插入项$J_i(x)$出现时从真空到真空的振幅的规则，源插入项包括非对易源η_i，$\bar{\eta}_i$。在像式(47)那样的高斯积分中总的乘法常数C通过下述要求被完全确定，即在没有源

的时候真空到真空的振幅应该为1，从而通过构建得到的散射矩阵应该最终是幺正的。

但是，实际上事情并没有那么简单。在实际计算中，人们常常遇到发散问题，这是无意义的表达。它发生在当人们很快做出向连续统极限转换的时候——要记得我们坚持时空首先要保持是离散的。S 矩阵的幺正性成为了我们是否正确地执行连续统极限的敏感标准。在我们开始为相对论的、量子化的粒子构建可行模型方案时，它曾是我们的主要要求之一。另一个要求是，色散关系的有效性可以用与幺正性相同的方式来处理，这两个概念将被表明是紧密相关的。下面描述的体系建立在库特考斯基(Cutkosky)等人的工作上，但经由韦尔特曼做出了很大的简化['t Hooft and Veltman，1994]。

本节的部分内容会非常技术化，初次阅读时可以跳过。

6.1　最大时间方程

我们开始于基本的费曼传播子$(k^2+m^2-i\varepsilon)^{-1}$和它在位形空间中的傅里叶变换，出于简单性忽略因子$(2\pi)^4$：

$$\Delta^F(x)=-i\int \mathrm{d}^4k\frac{e^{ikx}}{k^2+m^2-i\varepsilon},\quad x=x^{(1)}-x^{(2)} \tag{137}$$

此外，我们定义在壳(on-shell)传播子：

$$\Delta^{\pm}(x)=2\pi\int \mathrm{d}^4k e^{ikx}\delta(k^2+m^2)\theta(\pm k^0);\quad kx=\vec{k}\cdot\vec{x}-k^0x^0 \tag{138}$$

且 θ 是赫维赛德(Heaviside)阶梯(step)函数，当 $x\geqslant 0$ 时 $\theta(x)=1$，其他时候 $=0$。积分对闵可夫斯基(Minkowski)变量$\vec{k}$，k^0 进行。算符(138)将具有给定能量正负号的质壳上的粒子从 $x^{(2)}$ 传播到 $x^{(1)}$，或反向传播时，此能量正负号相反。我们有：

$$\Delta^+(x)=(\Delta^-(x))^*;\quad \Delta^+(x)=\Delta^-(-x) \tag{139}$$

我们的出发点是把传播子分解为前向和后向两部分：

$$\Delta^F(x)=\theta(x^0)\Delta^+(x)+\theta(-x^0)\Delta^-(x) \tag{140}$$

显然：

$$\Delta^{F*}(x)=\theta(x^0)\Delta^-(x)+\theta(-x^0)\Delta^+(x) \tag{141}$$

人们很容易通过形变在复的 k^0 平面内的围线积分来证明这一点。

现在考虑具有 n 个顶角的费曼图，在那里线与一给定的拓扑结构相伴随，该结构将保持固定。外线被假定为"截断的"：没有传播子与它们相伴随。费曼规则正如在2.5和4.5节中描述的那样被应用，则费曼图是我们对一个 S 矩阵元计算的部分。我们考虑在动量表征和位置表征中的图，在位置表征中我们得到的表达称为 $F(x^{(1)}, x^{(2)}, \cdots, x^{(n)})$。

接下来，我们引入与同一个图相关的一个表达，但在那里某些顶角有下划线：

$$F(x^{(1)}, x^{(2)}, \cdots, \underline{x}^{(i)}, \cdots, \underline{x}^{(j)}, \cdots, x^{(n)})$$

其中 $x^{(i)}$ 指的是当人们建立费曼规则时必须被积分涵盖的位置。计算这一新振幅的规则如下：

1)若 $x^{(i)}$ 和 $x^{(j)}$ 都没有下划线的话，用传播子 $\Delta^F(x^{(i)} - x^{(j)})$；

2)若 $x^{(i)}$ 而不是 $x^{(j)}$ 有下划线的话，用传播子 $\Delta^+(\underline{x}^{(i)} - x^{(j)})$；

3)若 $x^{(j)}$ 而不是 $x^{(i)}$ 有下划线的话，用传播子 $\Delta^-(x^{(i)} - \underline{x}^{(j)})$；

4)若 $x^{(i)}$ 与 $x^{(j)}$ 都有下划线的话，用传播子 $\Delta^{F*}(\underline{x}^{(i)} - \underline{x}^{(j)})$；

5)对每个下划线的顶角都要加上一个负号。

在所有其他方面，振幅的计算规则都不改变。

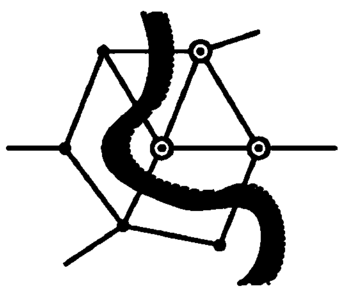


图3　下划线顶角时的图，这些顶角用小圆圈表示

现在我们得到了最大时间方程：

令 $x^{(k)}$ 为具有最大时间的坐标：$x^{(k)0} \geqslant x^{(i)0}$，$\forall i$。则：

$$F(x^{(1)}, x^{(2)}, \cdots, x^{(k)}, \cdots, \underline{x}^{(n)}) = -F(x^{(1)}, x^{(2)}, \cdots, \underline{x}^{(k)}, \cdots, \underline{x}^{(n)}) \quad (142)$$

在上述两项中，下划线在除 $x^{(k)}$ 以外的点处，或划线的方式不同。

人们很容易用方程(140)与(141)来证明这一点。此定理的一个结果是：

$$\sum_{\text{所有}2^n\text{个可能的划线处}} F(\{x^{(i)}\}) = 0 \tag{143}$$

我们现在表明这些是关于幺正方程或“光学定理”的图：

$$\sum_n S|n\rangle\langle n|S^\dagger = \mathbb{I} \tag{144}$$

矩阵 S 的图早些时候已经描述过。$S^\dagger$ 的图包括传播子的复共轭。因为函数积分中的顶角也都乘了因子 i，它们在 $S^\dagger$ 中必须都改变正负号。另外，e^{ikx}中的动量 k 也改变正负号。简而言之，用以计算 $S^\dagger$ 的图确实是有下划线的格林函数。注意到在动量空间中最大时间方程(142)不能够应用于单个顶角，因为虽然被积分涵盖，但具有最大时间的顶角改变了位置。然而，求和方程(143)是有效的。在方程(144)右手边的单位量(identity) $\mathbb{I}$ 来自于保留下来的一个结构，即完全没有顶角的图。

我们注意到，如果我们加上所有可能方式幺正性仍在的话，在这些方式中具有给定拓扑的图可以分为两个，正如在图 3 中描绘的。阴影线把 S 与 $S^\dagger$ 区分开来。

将 S 与 $S^\dagger$ 连接起来的线表征了方程(144)中的中间态 $|n\rangle$。它们在质壳上，且有正能量，这就是为什么我们在那里需要因子 $\delta(k^2+m^2)\theta(k^0)$ 的原因。如果一个传播子配备了一些额外的系数 R_{ij}：

$$P_{ij}(k) = \frac{-iR_{ij}(k)}{k^2+m^2-i\varepsilon} \tag{145}$$

则我们仍然能够采用相同的分解(140)，假定 R_{ij}是局域的话，它必定是 k 的一个有限多项式，写作：

$$R_{ij} = \sum_k f_i(k) f_j^*(k) \tag{146}$$

我们可以把因子 $f_i(k)$纳入 S 的定义中，假如 R_{ij}所有的本征值是非负的话。事实上，拉格朗日量中的动能项必须都有相同的正负号。

注意到我们不允许用拉格朗日量那些项的复共轭来代替它们，这意味着为了幺正性证明，必须强制要求拉格朗日密度是场的实函数。

这些方程的一个重要特征是它对于 k^0 是 θ 函数。它们保证了中间态仅在它们的总能量并不超过在给定频率内可用的能量时才起作用。

6.2 着衣(dressed)传播子

在前一节中，并不是对 $SS^{\dagger}$ 起作用的所有图都得到正确的处理。当自能图出现时会存在一种困难。若将自能团(blob)两端的线的其中一条用 $\Delta^{\pm}$ 来取代，则其他传播子 Δ^{F} 在那个狄拉克 δ 顶端放置一极点。在这种情形下，我们不得不用到一种更为复杂的描述。为了明白发生了什么，我们必须首先对传播子插入项进行几何级数求和，见图4(a)。我们得到了所谓的着衣传播子。在动量空间中，我们把图4(a)中单个团的贡献写为 $-i\delta M(k)$。它表征了所有不可约图的累积贡献，这些不可约图是具有两条外线的图，这两条外线在有人切断一条内线时不会分开。我们需要它的实部和虚部：$\delta M(k)\equiv\delta m^{2}(k)-i\Gamma(k)$。把完全传播子写为：

$$\begin{aligned}P^{\mathrm{dr}}(k)&=P^{0}(k)-P^{0}(k)i\delta M(k)P^{0}(k)+\cdots\\&=P^{0}(k)\sum_{n=0}^{\infty}(-i\delta M(k)P^{0}(k))^{n}\\&=\frac{P^{0}(k)}{1+i\delta M(k)P^{0}(k)}\end{aligned}\tag{147}$$

若 $P^{0}(k)=-i(M(k)-i\varepsilon)^{-1}$，则 $P^{\mathrm{dr}}(k)=-i(M(k)+\delta M(k)-i\varepsilon)^{-1}$ (148)

其中 $P^{0}(k)$为非微扰("裸")传播子。

若我们把着衣传播子(在动量空间)的实部定义为：

$$\Re(P^{\mathrm{dr}}(k))=\frac{\Gamma(k)}{(k^{2}+M+\delta m^{2})^{2}+\Gamma^{2}}=\pi\varrho(-k^{2})\tag{149}$$

则通过围线积分可得：

$$P^{\mathrm{dr}}(k)=\int_{0}^{\infty}dm^{2}\frac{\varrho(m^{2})}{k^{2}+m^{2}-i\varepsilon};\tag{150}$$

我们称它为传播子的契伦—雷曼(Källen-Lehmann)表征。稍后，将会表明若 $m^{2}<0$则会有$\varrho(m^{2})=0$。

现在最好的策略是把最大时间方程应用于整个着衣传播子。取代方程(140)和(141)，写为：

$$\begin{aligned}P^{\mathrm{dr}}(x)&=\theta(x^{0})\Delta_{\mathrm{dr}}^{+}(x)+\theta(-x^{0})\Delta_{\mathrm{dr}}^{-}(x)\\P^{\mathrm{dr}}(x)^{*}&=\theta(x^{0})\Delta_{\mathrm{dr}}^{-}(x)+\theta(-x^{0})\Delta_{\mathrm{dr}}^{+}(x)\end{aligned}\tag{151}$$

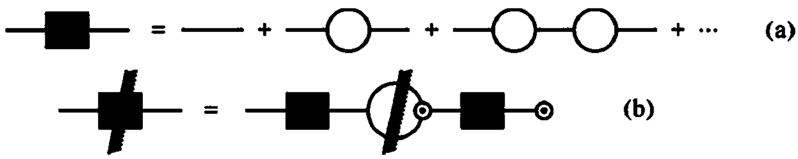


图 4　(a)作为几何级数的修饰传播子

(b)切断修饰传播子

则：

$$\Delta^{\pm}_{\mathrm{dr}}(k) = 2\pi\int \mathrm{d}^4 k\, e^{ikx}\,\varrho(-k^2)\,\theta(\pm k^0) \tag{152}$$

不可约图的虚部 $\Gamma(k)$本身能够通过应用截断规则而被再次找到。写下 $S = \mathbb{I} + iT$，我们发现所有相应于 $i(T - T^{\dagger}) + TT^{\dagger} = 0$ 的非平凡图的幺正性，和在图 4b 中描述的 $TT^{\dagger}$ 图。它们恰好是整个散射矩阵幺正性所需要的图，此散射矩阵涉及两个外部虚粒子通道中的单个虚粒子。

人们观察到函数$\varrho(-k^2)$必须是非负的，仅对于类时的 k 不为零。后者通过在 k^0 中的 θ 函数得以保证。仅仅是ϱ中的 θ 峰与出现在散射矩阵初态与终态中的稳定粒子相关联。具有确定宽度的共振通过它们稳定的衰变产物对散射矩阵的幺正性起作用。

6.3　入射和出射粒子的波函数

完整讨论中的许多技术细节需要太多的篇幅，所以我们不得不提纲挈领地讲。在我们处理矢量或旋量粒子时，传播子的留数 R_{ij}表征了粒子波函数绝对平方的和。我们在方程(145)中看到了它是如何形成的。例如，如果一个矢量粒子用下述传播子来描述：

$$P_{\mu\nu} = -i\frac{\delta_{\mu\nu} + k_{\mu}k_{\nu}/M^2}{k^2 + M^2 - i\varepsilon} \tag{153}$$

则我们首先看到计数器是 k 的多项式，正如所要求的那样，且如果我们继续回到质壳(mass shell) $k^2 = -M^2$ 上，则我们看到与 k_{μ}成比例的场分量突显出来。特别是，若我们取$k = (0, 0, 0, iM)$，则 $R_{ij} = \delta_{ij}$和它的类时分量为零，所以对于所描述的粒子确实存在三个独立的态。

对于费米子，裸传播子为：

$$P^{\mathrm{Dirac}} = -i\frac{m - i\gamma k}{k^2 + m^2 - i\varepsilon} \tag{154}$$

在把这与波函数的重整化联系起来之前，我们必须注意到所有的 γ^μ 都是厄米的，虽然 k_i 是实的，k_4 是虚的。我们观察到 $S^\dagger$ 的费曼规则类似于 S 的费曼规则，但是用 $-\gamma^4$ 取代了 γ^4。接下来，传播子中的箭头必须逆向。这导致对每一个矢量 k_μ 都有附加的负号，同时 γ^μ 被 $\gamma^{\mu\dagger}$ 所取代。总之，我们要求当 γ^4 保持不变时 $\gamma^i \to -\gamma^i$。这等于置换 $\gamma^\mu \to \gamma^4\gamma^\mu\gamma^4$。人们总结出，若所有费米线进入或离开有附加因子 γ^4 的图时，$S^\dagger$ 的费曼规则类似于 S 的费曼规则。这意味着在图中的外费米子波函数归一化为：

$$(m - i\gamma_i k^i + \gamma^4 k^0)\gamma^4 = \sum_{i=1}^{2} |\psi_i(k)\rangle\langle\psi_i(k)| \quad (k^0 > 0) \tag{155}$$

而对于反费米子，我们必须要求：

$$\gamma^4(m - i\gamma_i k^i - \gamma^4 k^0) = -\sum_{i=3}^{4} |\psi_i(k)\rangle\langle\psi_i(k)| \quad (k^0 > 0) \tag{156}$$

负号是必要的，因为(156)中的算符有两个负的本征值。人们总结出对于这些粒子的每一个闭圈而言，幺正性要求自旋 $\frac{1}{2}$ 粒子携带一个附加的负号。这导致了费米—狄拉克统计的必要性。再者，很重要的是，高阶修正中没有哪一项会影响这些投影算符的本征值的正负号，因为这些从未通过粒子波函数的重整化而调整。

本节的结论是在物理粒子态空间中的散射矩阵是幺正的，这一结论是意料之中的，因为我们的理论就是那样来构造的。仍然很重要的是，我们在这里明白了费曼图是以什么方式结合进而明确地产生幺正性的。

我们同样明白了，当我们由于规范固定过程有了鬼场时，幺正性变得很难去控制。则我们的矢量粒子有了传播子，在那里方程(153)被如下的表达所取代：

$$P_{\mu\nu}^{\mathrm{ren}} = \frac{-ig_{\mu\nu}}{k^2 + M^2 - i\varepsilon} \tag{157}$$

这里我们写作 $g_{\mu\nu}$ 而非 $\delta_{\mu\nu}$ 是为了强调我们的论证适用于闵可夫斯基空间，在那

里时间分量显然“有着错的正负号”。与此相关的场分量会对应于对散射概率起负贡献的粒子。为了更正这一点，人们不得不用 $-|n\rangle\langle n|$ 取代 $|n\rangle\langle n|$，这是不能够通过重整化态 $|n\rangle$ 而得到的。这里，我们用 BRST 关系来表明，所有的非物理态都可以被变换掉。实践中，我们采用这样的事实，即散射矩阵并不依赖于对规范固定函数 $C^{a}(x)$ 的选择，所以我们这样选择它以使所有的鬼粒子都有超过某临界值 Λ 的质量。在中间态中，它们的投影算符 $\Delta^{\pm}(k)$ 则仅在给定的频率内总能量超过 Λ 时才起作用。这意味着在中间态中没有鬼粒子，所以散射矩阵仅在物理粒子空间中才是幺正的——这在关于这些理论是内部一致性的论证中是绝不可少的步骤。所要求的规范固定函数 $C^{a}(x)$ 并不难构建，但它们的存在意义仅是为了完善这一形式论证。在实际的计算中使用它们是非常麻烦的。

6.4 色散关系

最大时间方程也可以用于导出非常重要的图的色散关系。这些意味着任意的图 D 都可以看作是两组图 D_i 与 $D_i^{\dagger}$ 的组合：

$$D = \sum_i \int_0^{\infty} \frac{\mathrm{d}k^0}{-k^0 - i\varepsilon} D_i(k^0) D_i^{\dagger}(k^0) \tag{158}$$

这里，$D_i(k^0)$ 和 $D_i^{\dagger}(k^0)$ 代表依赖于不同外部动量 k 的振幅，在那里对一个类时分量 k^0 积分。这是通过挑选图中的两点 $x^{(1)}$ 与 $x^{(2)}$，和它们的时间顺序而得到的。推导的细节超出了本论文的范围(虽然它们并不比本节中的前一个推导更难)。方程(158)可以通过具有更少闭圈的图用来表示具有闭圈的图，也可用于讨论重整化所需要的减除程序。

7. 重整化

为了对重整化概念进行真正的讨论，我们必须强调我们的出发点。首先，用点的致密点阵来取代空间连续统，且仅在所有计算结束的时候我们才试图取连续统极限。2.4 节中讲解过的路径积分过程意味着时间也可以用点阵来代替。在傅里叶空间中，时空点阵产生了能量和动量值的有限范围(布里渊区)，从而

所有紫外发散消失。如果我们还希望保证不会有红外发散，我们必须用一有限的方格(box)来取代时空体积元。常常需要这样做，因为软(soft)虚粒子，特别是光子的发散贡献会造成问题的复杂化。烦人的红外发散发生在有禁闭的理论中，这将在第11节中讨论。

对于理论中的紫外发散，我们要用到的方法如下。我们假定理论中所有可自由调整的物理常数，指的是如"裸"质量和粒子的电荷那样的"裸"参数，都应该加以仔细调整以与观测相一致，但调整过程可能严重依赖于时空点阵的网格尺寸 a。因而，在我们改变 a 的同时，我们允许理论中所有的裸参数(比如 λ)都依赖于 a，这些裸参数在 $a \to 0$ 时常常趋向于无穷大或零。若该过程与一个微扰展开相结合，比如说用一个小的耦合 g 展开，我们期待发现最低程度依赖于 a 的可观测特征，前提是该裸耦合 $g(a)$ 在极限 $a \downarrow 0$ 仍然很小。

这是使我们的理论有意义的一个重要条件。我们如何知道 $g(a)$ 是否趋于零？最简单的办法是看 g 的维度。一个场理论的所有参数都具有长度某次幂的量纲。这些维度通常依赖于时空的维度 n。很容易得到计算它们的规则：

——作用量 $S = \int d^n x\, \mathcal{L}(z)$ 是无量纲的；

——因而拉氏密度$\mathcal{L}$的维度是(长度)$^{-n} = m^n$，其中 m 是质量；

——场的维度可以由拉格朗日量中的动能项读出，因为它们没有包含更多的参数，一个标量场 ϕ 的维度为 $m^{(n-2)/2}$，费米子场 ψ 的维度为 $m^{(n-1)/2}$；

——规范耦合常数 g 拥有维度 $m^{(4-n)/2}$，形如 $\lambda\phi^k$ 的相互作用项中的耦合参数 λ 的维度为 $m^{n+k-nk/2}$，以此类推，等等。

> 一个理论称为幂次计数可重整化，若所有展开参数有正的或零质量维度。

这就是为什么在四维时空中，我们不能够接受标量中高于四次的相互作用的原因。实际上，在四维时空中，大多数展开参数的维度为零。在第9节中，我们将看到无量纲的耦合参数仍然依赖于 a 的大小，但仅仅是对数地依赖：

$$\lambda(a) \approx \lambda_0 + C\lambda_0^2 \log(a) + \text{更高阶项} \tag{159}$$

不考虑它是否在连续统极限下趋于零或无穷大，我们发现在连续统理论中，对裸参数 λ 的微扰更正是发散的，这并不值得担心。然而，若 λ 本身是一个我们希望用它来做微扰展开的小的参数，则如果它的裸值趋于无穷大会有明显的麻烦。事实上，我们会证明一般这样的理论是不一致的。

这里需要做出两个非常重要的评论。

——当所有耦合在连续统极限下真正地趋于零的情况时，理论可以构建。这些理论被称为是渐近自由的（第 9 节），且它们允许在紫外区进行精确近似。人们普遍相信，这样的理论可以通过它们在紫外区的微扰展开而以一种完全清晰的方式被定义。还有，它们允许对其所有的物理特征进行非常精确的计算。QCD 就是最好的案例。

——若一个理论不是渐近自由的，而是仅有小的耦合参数，则微扰展开形式上会发散，且连续统极限在形式上并不存在。但微扰展开的前 N 项是有意义的，那里 $N = \mathcal{O}(1/g)$。这意味着不可控的误差范围是指数级小的，为 $e^{-C/g}$ 或 e^{-C/g^2} 量级，它在实际中要远小于理论中的其他不确定量，所以它们几乎不会有实际的影响。因而，在这样的情形下，我们的理论有内在的不准确度，但它们是指数级衰减的。实践中，这样的理论仍然是非常有价值的。标准模型就是一个案例。

一种有用的办法是通过形式的级数展开把所有数值都代入理论，在级数展开中对参数展开，即该参数是理论的所有耦合参数共同拥有的一个因子，在形式上保持无穷小。那样的话，所有微扰系数都被唯一地确定，虽然人们对有关展开收敛或发散的第一手知识知之甚少。

在上述两种情形中，我们的理论是用它们的微扰展开来定义的。显然，微扰展开不只是计算的简便策略，它也是我们理论必不可少的要素。接下来我们研究重整化在微扰理论中是如何一步步地运作的。

在一个连接图中，令外线的数目为 E，传播子数目为 P，并令 V_n 为具有 n 个叉的顶角数。通过在每个传播子上画两个点，在每条外线上画一个点，人们发现点的数目为：

$$2P + E = \sum_n nV_n = 3V_3 + 4V_4 \tag{160}$$

对于树图(简单连接的图)，通过约化人们发现顶角的数目 V 为 $V=\sum_n V_n$：

$$V=P+1 \tag{161}$$

通过切断 L 传播子，一个具有 L 个闭圈的图(一个 L 重连接图)变为一个树图。因而，有：

$$P=V-1+L \tag{162}$$

结合式(160)与式(162)，有：

$$E+2L-2=\sum_n (n-2)V_n=V_3+2V_4 \tag{163}$$

因此，如果每个3顶角伴随有一个因子 g，每个4顶角伴随有因子 λ，且如果具有给定外线数目 E 的图表现为 $g^{2n}\lambda^k$，它必定拥有 $L=n+k+1-\frac{1}{2}E$ 个闭圈。从而微扰展开常常被认为是用闭圈数目表达的展开。

7.1 正规化(regularization)方案

在动量空间的一个树图中，不需要做积分——流经每个传播子的动量是通过入射和出射粒子的动量来确定的。但若有 L 个圈，人们需要在动量空间中做 $4L$ 次积分。这些积分常常在大动量时趋向于发散。

当然，若动量空间被截断的话，这些发散会停止，正如在有限点阵情形中那样。然而，因为我们的点阵不是洛伦兹不变的，且可能缺乏如规范不变性那样的其他对称性，因而发现其他修正理论的方式以使得紫外发散消失是有用的。这称为“正规化”。我们给出两个例子。

泡利—维拉斯(Pauli-Villars)正规化

假定具有上述形式的传播子以如下方式被取代：

$$\frac{A(k)}{k^2+m^2-i\varepsilon}\to\sum_i e_i\frac{A(k)}{k^2+\Lambda_i^2-i\varepsilon};\quad \sum_i e_i=0,\sum_i e_i\Lambda_i^2=0 \tag{164}$$

若我们取 $e_1=1$，$\Lambda_1=m$，同时所有其他 Λ_i 趋向于 $+\infty$，我们得到了最初的传播子。但在有限的 Λ_i 下，我们可以使所有的动量积分在无穷大时收敛，从而我们的理论是有限的。这就是(某种简化版本的)泡利—维拉斯正规化。

但是，新的传播子不能描述普通粒子。具有 $e_i<0$ 的传播子给幺正关系以错误的正负号！另一方面，$i\varepsilon$ 描述一如平常，所以这些粒子携带有正的能量。

在总能量小于 Λ_i 的其他频段下，“泡利—维拉斯鬼粒子”对幺正关系完全没有贡献。所以，在我们为所考虑的总能量设置极限的理论中，泡利—维拉斯正规化在物理上是可接受的。在实践中，我们将试图把所有鬼粒子的质量 Λ_i 变为无穷大。

维数正规化

维数正规化['t Hooft and Veltman, 1972b]由在 $4-\varepsilon$ 维中对所有环积分作形式化操作组成，其中 ε 可以是任意数(可能是复数)。只要 ε 是无理数，所有的积分都可以通过具有清晰描述的有限表达来取代，下面会给出解释。若 $\varepsilon=0$，人们也可以减除积分，但描述常常不是清晰的，所以可能会产生反常。这就是为什么在每当人们希望理解并控制的反常出现时维数正规化特别重要的原因。

同样重要的是应该认识到当 $\varepsilon\neq0$ 时积分可能会发散，但对于无理数 ε 会做出清晰的减除。这需要得到解释，但首先人们需要确定拥有非整数维意味着什么。这样一个定义仅在微扰或环展开的框架下才容易理解。考虑一个具有 L 个圈与 N 条外线的不可约图，在那里我们保持外动量 $p_{(1)}$，…，$p_{(N)}$不变。很显然根据理论的构造得到被积函数是 $L(4-\varepsilon)$ 变量的纯有理函数。在观察到外动量张成了某个 $N-1$ 维空间之后，我们现在采用这样的事实，即在其余维度上的积分是旋转不变的。在那里，我们把半径为 r 的 ℓ 维，欧几里得(Euclid)球面公式写为：

$$\int \mathrm{d}^{\ell}k\delta(k^2-r^2)=\frac{\pi^{\ell/2}}{\Gamma(\ell/2)}r^{\ell-2} \tag{165}$$

这里，Γ 代表欧拉的伽玛函数，对整数 z，$\Gamma(z)=(z-1)!$

正是在这点上，我们可以确定这一表达对 ℓ 的任意值——可能是复值——定义了积分。每当 ℓ 是一个正整数时该积分收敛于普通值。在应用方程(165)后，人们得到了对函数 $f(k)$ 的 s 个变量 k_μ 的积分，其中 s 为一个整数，但 $f(k)$ 包含了 k 多项式的 ε 依赖次幂。

一个积分的收敛或发散可以从简单的幂次计数论证中获悉，且一开始当 ε 接近于零的时候人们几乎看不到任何进展。然而，所得到的是红外发散($k_\mu\to0$)从紫外发散($k_\mu\to\infty$)中分离了出来，这允许我们清晰地定义积分中的“有限

部分”。

·所有积分 $\int d^s k f(k)$ 都被函数 $I(\{f(k)\})$ 所代替，该函数满足与普通积分相同的结合律：

$$I(\alpha f_1 + \beta f_2) = \alpha I(f_1) + \beta I(f_2)$$
$$I(\{f(k+q)\}) = I(\{f(k)\}) \tag{166}$$

·如果它收敛的话，$I(f) = \int d^s k f(k)$；

·当 $2p+s$ 不是一个整数时，如果 $f(k) = (k^2)^p$，则 $I(f) = 0$。

若我们从不是整数的 ε 开始，后一个条件通常是满足的。

这些规则足以代替我们通过一些有限的表达在费曼图中遇到的任意积分。但是，需要注意如果在 $2p+s$ 是整数时，特别是当它等于零的时候应用这些规则将会出现复杂的问题。在那种情形下，该表达在紫外与红外区都发散，所以在这里不能够用它来去除所有的发散——仅能够用这一个替代另外一个。结果就是，随着 $\varepsilon \to 0$ 我们的有限表达趋向无穷大。

确证维数正规化完全遵守幺正性和上面讨论的色散关系是重要的。因而，在一些“自然的”减除下，“维数正规化的”图对应于色散关系和幺正关系的解。

正规化方案的等价性

一般地，由上面讨论过的不同正规化方案给出的减除并不相同。在给定的阶上，它们都遵守形如式(158)的同一个色散关系。当我们问哪一个振幅可以加入到一个方案中以产生另一个方案，或两个方案之间的不同的振幅是什么——在消除掉这些在子图 $D_i(k^0)$ 被计算的那个阶上这些不同之后——的时候，我们发现了下述内容。这一不同必定是一个洛伦兹不变的表达，它仅来自于方程(158)中非物理的泡利—维拉斯鬼场的维数正规化的贡献。因为它们的大的质量，该方程中只有非常大的 k^0 值才起作用。则 p^0 依赖性必须约化为一个多项式的项(p 为固定外线的动量)，且由于洛伦兹不变性，该表达在 p_μ 的分量上都必须是多项式。这恰恰是通过在理论中的裸拉格朗日量中放置抵消(counter)项所能达到的。以这种方式，人们推导出不同正规子(regulator)因为在裸拉格朗日量中有不同的有效耦合而彼此不同。

人们喜欢哪一个正规子则是喜好问题。因为通常维数正规化完全遵守定域

规范不变性，[①] 且还因为它被证明在实践中是非常简便与有效的，人们常常偏好于它。但是，应该时刻谨记的是，维数正规化是某种数学技巧，物理表达仅在极限 $\varepsilon \to 0$ 下才有意义。

7.2　规范理论的重整化

利用前一节得到的结果，我们决定笼统地处理量子场论，而具体讨论规范理论。首先，我们正规化此理论，通过采用一个“点阵截断”，或泡利—维拉斯截断，或通过转向 $n=4-\varepsilon$ 维。所有这些过程都具有一个小的参数，如 ε，这样一来在极限 $\varepsilon \to 0$ 下就从形式上得到了物理理论。这些过程在下述意义上都是等价的，即通过在拉格朗日量中增加定域相互作用项，人们可以把一个方案的结果映射到其他方案的结果上。之后，我们重整化该理论。这意味着拉格朗日量中的所有参数都经过了有限的修正，不论这一修正在极限 $\varepsilon \to 0$ 下会发散到什么程度。若很好地选择这些抵消项的话，理论在该极限下保持有限且定义明确。特别地，我们应该有一个幺正的、因果的理论。

仅当理论是规范不变的时才能够保证幺正性。因此，人们喜欢那种能够全面维持规范不变性的正规子方案。而维数正规化常常能够满足这一条件。那样的话，重整化过程遵守 BRST 不变性，见 4.6 节。

8. 反常

接下来的这一节（再一次）对于学习量子场论来讲过于简略以至于不足以形成完整的教材。这里我们的目标是给出对特征的总结，这些特征对于理解相对论量子场论的一般结构是极其重要的。

对于一个给定的理论，如果不存在明显规范不变的正规化过程，则可能是由于下述原因：这样一个理论可能根本就不可重整化。原则上，可以检验如下。人们总是决定采用一个并不遵守他想要的对称性的正规化过程，假如该对

① 仅在一种情形下，即当存在阿德勒—贝尔—杰克反常（Adler-Bell-Jackiw anomalies）时会有问题，见第 8 节。

称性能够在正规子的物理可观测效应消失的极限下——如 $\varepsilon\to 0$ 或 $\Lambda_i\to\infty$，$i>0$——重建的话。若一个规范不变的正规子确实存在，但它还没有被明确构建时，则我们知道它与通过一组有限的抵消项建立起来的其他正规子不同。实际上，找到这样的抵消项并不难，仅仅把所有需要的项添加进来以恢复振幅的BRST不变性即可。

但是，在正规子未知的情况下，我们如何能够确定这样的项存在呢？BRST不变性要求斯拉诺夫—泰勒不等式的有效性，但它们似乎高估了减除项。这是我们最初在['t Hooft，1971]中对这一问题的叙述。事实上，这里确实会有冲突。若发生冲突的话，它被称为是一个反常(anomaly)[Jackiw，1995，Chapter. 1]。

事实上，幸运的是这种反常的发生率有限。这是因为大多数理论都找到了完全规范不变的正规子方法。维数正规化常常是有效的，它不起作用的一种情形是当存在手征费米子时的情形。经典上，人们可以把任意的费米子场分为左手与右手部分，正如在4.2节中提到的：

$$\psi(x)=P_+\psi_L(x)+P_-\psi_R(x);\qquad P_\pm=\frac{1}{2}(1\pm\gamma^5)$$

$$\gamma_5=\frac{1}{24}\varepsilon_{\mu\nu\alpha\beta}\gamma^\mu\gamma^\nu\gamma^\alpha\gamma^\beta \tag{167}$$

事实上，因为$(\gamma^5)^2=1$，算符$P_\pm$是真正的投影算符：$P_\pm^2=P_\pm$。

费米子的左手与右手部分，见方程(106)，可能分别是规范不变的，在规范变换下的变换不同。但是，这要求γ^5与所有其他γ^μ，$\mu=1$，…，n反对易。但是，正如我们在它们的定义——方程(167)中看到的，γ^5仅与四个，而不是所有n个γ^μ反对易。这就是为什么$-\varepsilon$剩下的维度的贡献将不再是规范变的原因。

贝尔和杰克[Bell and Jackiw，1969]以及阿德勒[Adler，1969；Adler and Bardeen，1969；Bardeen，1969]独立发现，满足所有对称性条件并具有期望维度的定域抵消项不存在，贝尔和杰克试图采用非常规正规子，但这被证明是不容许的。主要的过错在于三角图5(a)表示了两个光子场中轴矢流$\bar{\psi}\gamma^\mu\gamma^5\psi$的矩阵元，两个光子中的每一个都与矢量流$\bar{\psi}\gamma^\alpha\psi$相耦合。

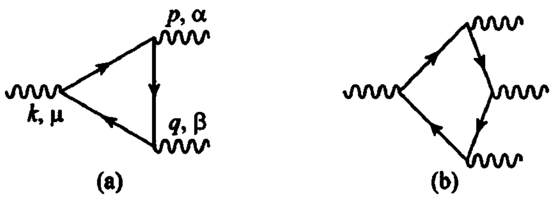


图 5 (a)反常的三角图，μ,α 与 β 都是偏振度，k,p 与 $q = k - q$ 为外部动量。(b)在非阿贝尔理论中的一个反常图

出于简单性，我们这里假定费米子是无质量的。我们称这个振幅为 $\Gamma_{\mu}^{\alpha,\beta}(p,\ q)$，它是线性发散的。在正规化的时候，存在两个抵消项或减除项，它们的系数在与振幅的有限部分正确组合下应该是确定的。只关心正确的量子数和维度，我们可发现两个量：

$$\delta_1\Gamma_{\mu}^{\alpha,\beta}(p,\ q) = \varepsilon_{\mu\alpha\beta\gamma}p_{\gamma}$$

$$\delta_2\Gamma_{\mu}^{\alpha,\beta}(p,\ q) = \varepsilon_{\mu\alpha\beta\gamma}q_{\gamma} \tag{168}$$

我们可以通过应用下述条件来确定它们的系数，即在光子场的规范变换下总的振幅不变。这意味着当任意两个光子为纵向的(longitudinal)(即 $A_{\mu} = \partial_{\mu}\Lambda$)时，该表达必须为零，它意味着：

$$p_{\alpha}\Gamma_{\mu}^{\alpha,\beta}(p,\ q) = 0; \qquad q_{\beta}\Gamma_{\mu}^{\alpha,\beta}(p,\ q) = 0 \tag{169}$$

因为：

$$p_{\alpha}\delta_1\Gamma_{\mu}^{\alpha,\beta}(p,\ q) = 0; \qquad q_{\beta}\delta_1\Gamma_{\mu}^{\alpha,\beta}(p,\ q) = A_{\mu,\alpha}$$

$$p_{\alpha}\delta_2\Gamma_{\mu}^{\alpha,\beta}(p,\ q) = A_{\mu,\beta}; \qquad q_{\beta}\delta_2\Gamma_{\mu}^{\alpha,\beta}(p,\ q) = 0$$

$$A_{\mu,\alpha} = \varepsilon_{\mu\alpha\beta\gamma}p_{\gamma}q_{\beta} \tag{170}$$

这确定了在 $\delta_1\Gamma$ 与 $\delta_2\Gamma$ 前面的系数。

现在当我们研究这一振幅相对于轴矢流是否同样是横向的时，我们遇到了一个惊喜。通过条件(170)确定的抵消项也在这里起作用：

$$k_{\mu}\delta_1\Gamma_{\mu}^{\alpha,\beta}(p,\ q) = -k_{\mu}\delta_2\Gamma_{\mu}^{\alpha,\beta}(p,\ q) = A_{\alpha,\beta} \tag{171}$$

但它们并不取消对有限部分的贡献。在相对于两个向量嵌入施加规范不变性之后，人们发现(在单个手征费米子情况下)：

$$k_{\mu}\Gamma_{\mu}^{\alpha,\beta}(p,\ q) = (4\pi^2)^{-1}\varepsilon_{\mu\alpha\beta\gamma}p_{\mu}q_{\gamma} \tag{172}$$

且它可以重写为一矢量流的有效发散特征：

$$\partial_\mu J_\mu^5 = -\frac{iLe^2}{8\pi^2}F_{\mu\nu}\tilde{F}_{\mu\nu} \tag{173}$$

其中，$\tilde{F}_{\mu\nu}=\frac{1}{2}\varepsilon_{\mu\nu\alpha\beta}F_{\alpha\beta}$，$L$ 是左手费米子数目与右手费米子数目之差，且假定光子与电荷 e 耦合。

关于这一点令人惊讶的是三角图本身——图 5a——似乎在所有置换下都是完全对称的，因为 γ^5 可以改变为任意其他的终点。在其三个终点中的两个上施加规范不变性意味着第三个不变性的破缺。

这一结果是非常重要的。它意味着正规化过程显然被赋予了对守恒律的连带违背。它也意味着不可能把三个规范玻色子与这样一个三角图耦合起来，因为它不能够以一种规范不变的方式来完成。但是，在绝大多数的理论中，我们同时有与左手和右手费米子的耦合。它们的贡献有相反的正负号，这意味着它们能够被抵消。因此，我们得到了对手征费米子规范理论的重要限制，即三角反常必须被抵消。

令左手手征费米子在规范群完全集的表征中，变换为：

$$\psi_L^i \to \psi_L^i + i\Lambda^a T_{Lj}^{ai}\psi_L^j \tag{174}$$

其中 Λ 为无穷小，T_L^a 为左手费米子的规范生成元。对于右手 ψ_R^i 也类似。定义：

$$d_L^{abc} = \mathrm{Tr}(T_L^a T_L^b T_L^c + T_L^b T_L^a T_L^c) \tag{175}$$

类似地有 d_R^{abc}。反常约束则为：

$$\sum d_L^{abc} = \sum d_R^{abc} \tag{176}$$

其中求和是对理论中的所有费米子种类进行的。在标准模型中，唯一有意义的是三个指标 d^{abc} 全部或其中之一指的是 $U(1)$ 群。很快就被证明了的是，夸克和轻子的 $U(1)$ 荷确实是以这样一种方式分布的，即以式(176)被完全确证的方式，但这仅发生在夸克代(generations)数目和轻子代的数目相等时。在 10.3 节中，我们将看到这一观测结果的物理意义。

注意，在非阿贝尔情形中，在具有 4 个外支线(leg)的图中仍然有反常，见图 5(b)。它们产生自 $F_{\mu\nu}\tilde{F}_{\mu\nu}$ 中的三线性项(四线性项取消)。仅在这些情形中正规化过程可能会违背规范不变性。在具有更多圈的图中，或具有更多外线的子图中，可以发现保留了规范不变性的正规化过程。

9. 渐近自由

9.1 重整化群

斯蒂克尔堡(Stueckelberg)和彼得曼(Peterman)[Stueckelberg and Peterman, 1953]于1953年注意到，虽然理论的微扰展开依赖于如何把拉格朗日量中的裸参数分为最低阶的参数和重整化需要的抵消项，但整个理论应该不依赖于此。他们把这解释为一种不变性，把从最低阶到更高阶修正的参数置换行为解释为一种群操作，人们得到“重整化群”。

仅在一个实例中这样的变换才有真正的影响，即当将在一个质量或距离标度上的理论与另一不同标度上的相同理论进行对比时。标度变换必定与抵消项的置换相关联。因而，物理学家们开始把标度变换概念看作是一个“重整化群变换”。

盖尔曼(Gell-Mann)和劳(Low)[Gell-Mann and Low, 1954]发现这一过程可以用于推导QED短距离行为。人们认为，有效的精细结构常数依赖于标度μ，用下述方程来描述：

$$\frac{\mu^2 \mathrm{d}}{\mathrm{d}\mu^2}\alpha(\mu)=\beta(\alpha)=\frac{\alpha^2 N}{3\pi}+\mathcal{O}(\alpha^3) \tag{177}$$

其中N是带电费米子类型的数目。只要$\alpha(\mu)$保持很小，从而$\mathcal{O}(\alpha^3)$项可以忽略，我们看到它的μ依赖为：

$$\alpha(\mu)=\frac{\alpha_0}{1-(\alpha_0 N/3\pi)\log(\mu^2/\mu_0^2)},\ 若\ \alpha(\mu_0)=\alpha_0 \tag{178}$$

当μ到达与$\exp(3\pi/2N\alpha_0)$相当的值时事情变得不可控制，但是至少在QED情形中(其中$\alpha_0\approx 1/137$)，这一质量标度如此之大以至于实际上不会出现问题。方程(178)中的极点称之为朗道(Landau)极点。朗道总结说，诸如QED的量子场论由于这一极点因而没有真正的连续统极限。但是，盖尔曼和劳怀疑$\beta(\alpha)$可能在α某个大的值处为零，所以在μ值很高时α接近这一值，但并不会超过这一稳定点。

在朗道极点或在其附近究竟发生了什么，并不能够单独通过做微扰展开来

确定，因为这将依赖于方程(177)中所有的高阶项，事实上，甚至都不知道量子场论是否能足够准确地重建以做出判断。然而，问题并不像它看起来那么重要，因为朗道极点将会超过普朗克质量，那样的话我们知道引力项将会占上风，更重要的将是先解决量子引力。

在函数$\beta(\lambda)$为负的理论中会出现完全不同的情形。很久以来人们认为这一情形从不会发生，除非耦合强度λ本身被给予错的正负号(即会导致经典理论的能量密度没有下界的正负号)，但后来证明仅有在只包含标量和旋量场的理论中才是这样。若理论中存在一个非阿贝尔的杨—米尔斯分量，则并不会产生负的β函数。在最简单的情形中，即在基本二重态表征中具有N_f个费米子的$SU(2)$规范理论中，β函数为：

$$\frac{\mu^2 \mathrm{d}}{\mathrm{d}\mu^2} g^2(\mu) = \beta(g^2) = \frac{N_f - 11}{24\pi^2} g^4(\mu) + O(g^6) \tag{179}$$

所以，只要$N_f < 11$时，随着$\mu \to \infty$我们有耦合强度$g(\mu)$对数地趋于零。这一特征称为渐近自由(asymptotic freedom)。在一个$SU(N_c)$规范理论中，β函数与$N_f - \frac{11}{2}N_c$成比例，所以在夸克味的当前数目$N_f = 6$时，QCD($N_c = 3$)是渐近自由的。与常常用到的表示一致，这里的下标c代表“色”，在QCD中，色的数目为$N_c = 3$。

9.2 β函数的代数

在拥有规范场、费米子和标量的理论中，情形更为复杂。人们已完成了β函数的一个普遍算法。一般结果最为紧凑的表示法可以通过如下方式给出，即写出整个拉氏密度$\mathcal{L}$在一个标度变换下是如何标度的。令拉格朗日量$\mathcal{L}$为：

$$\mathcal{L} = -\frac{1}{4} G^a_{\mu\nu} G^a_{\mu\nu} - \frac{1}{2}(D_\mu \phi_i)^2 - V(\phi) - \bar{\psi}_i \gamma_\mu D_\mu \psi_i$$

$$-\bar{\psi}_i \left(S_{ij}(\phi) + i\gamma_5 P_{ij}(\phi) \right) \psi_j \tag{180}$$

其中协变微商的定义如下：[①]

---

① T和U是厄米的，但因为ϕ是实的，则T的元素必须是虚的。

$$D_\mu\phi_i \equiv \partial_\mu\phi_i + iT^a_{ij}A^a_\mu\phi_j;\qquad D_\mu\phi_i = \partial_\mu\psi_i + iU^a_{ij}A^a_\mu\psi_j \tag{181}$$

结构常数f^{abc}的定义为：

$$[T^a,\ T^b] = -if^{abc}T_c \tag{182}$$

从而：

$$G^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + f^{abc}A^b_\mu A^c_\nu \tag{183}$$

我们把费米子分为右手和左手表征，从而有：

$$U^a = U^a_L P_L + U^a_R P_R;\qquad P_L = \frac{1+\gamma_5}{2},\qquad P_R = \frac{1-\gamma_5}{2} \tag{184}$$

函数$S(\phi)$和$P(\phi)$对ϕ至多是线性的，$V(\phi)$至多是四次的。拉格朗日量(180)是规范不变的部分，我们并不写出规范固定部分或鬼部分，最终的结果不依赖于这些细节。

一个代数计算的结果为：

$$16\pi^2\frac{\mu^2\mathrm{d}\mathcal{L}}{\mathrm{d}\mu^2} = G^a_{\mu\nu}G^b_{\mu\nu}\left[-\frac{11}{12}C^{ab}_1 + \frac{1}{24}C^{ab}_2 + \frac{1}{6}C^{ab}_3\right] - \Delta V - \bar{\psi}(\Delta S + i\gamma_5\Delta P)\psi \tag{185}$$

其中：

$$C^{ab}_1 = f^{apq}f^{bpq} \tag{186}$$

$$C^{ab}_2 = \mathrm{Tr}(T^aT^b) \tag{187}$$

$$C^{ab}_3 = \mathrm{Tr}(U^a_LU^b_L + U^a_RU^b_R) \tag{188}$$

$$\Delta V = \frac{1}{4}V^2_{ij} - \frac{3}{2}V_i(T^2\phi)_i + \frac{3}{4}(\phi T^aT^b\phi)^2 + \phi_iV_j\mathrm{Tr}(S_{,i}S_{,j} + P_{,i}P_{,j}) - \mathrm{Tr}(S^2+P^2)^2 + \mathrm{Tr}[S,\ P]^2 \tag{189}$$

其中：

$$V_i \equiv \frac{\partial V(\phi)}{\partial\phi_i};\ S_{,i} \equiv \frac{\partial S}{\partial\phi_i},\ \text{等等} \tag{190}$$

且记作$S + iP = W$，人们从下式中发现ΔS和ΔP：

$$\Delta W = \frac{1}{4}W_iW^*_iW + \frac{1}{4}WW^*_iW_i + W_iW^*W_i - \frac{3}{2}(U^2_RW) - \frac{3}{2}W(U_L)^2 + W_i\phi_j\mathrm{Tr}(S_iS_j + P_iP_j) \tag{191}$$

这一表达式并不包括关于场 ϕ_i 和 ψ_i 在标度下如何变换的信息。场不是直接可观测量。

这一代数表达可以用于找出耦合强度在动量的重新标度下一般是如何运作的。计算出渐近自由的条件，也就是所有耦合强度在无穷大动量下趋于零的条件，这是饶有趣味的练习。一般地，人们发现标量场仅在出现规范场和费米子时才存在；后者必须是对规范群的充分高度表征。

10. 拓扑扭曲

如果我们希望局限于那些在动量重新标度下对数地运作的耦合强度，例如方程(178)，则拉格朗日量(180)是所允许的最一般的拉格朗日量。这一理论的有效性区域涵盖了动量以指数增长的大值(原则上如果理论是渐近自由的话会涵盖所有的动量)。这个一般拉格朗日量的最显著特征是，它在拓扑上是高度非平凡的。局域稳定的场组态可能存在，其中有一些拓扑扭曲。特别是在布劳特—恩格勒—希格斯机制中这一点会更明确。这里的这些扭曲在经典层面已经见过了(即忽略量子效应)。

如果我们说一个标量场 ϕ_i 有一个真空期望值，则这意味着我们的微扰展开开始于真空中形如 $\phi_i=(F,\ 0,\ \cdots)$ 的场值，之后假定围绕这一值的场涨落 $\delta\phi$ 很小。人们假定势 $V(\phi)$ 在那里有最小值。这似乎会违背规范不变性，如果 ϕ_i 在定域规范变换下变换到彼此，但严格说来“定域规范对称的自发破缺”这一表达是不合适的，因为它同样也意味着我们选择了一个规范条件。但是，事实是如果我们在 $\phi_i=0$ 处微扰的话，物理粒子的谱被表明是全然不同的，所以这一“希格斯模式”不管怎样都是一个重要的概念。

10.1 涡旋

如果希格斯场仅有两个实分量，如当 $U(1)$ 破缺为单位元群时，人们可以考虑这样的组态，在那里当沿着一个闭合曲线时，该场做一个完全 360° 的扭曲。在曲线内部必定有一个零点。这些零点自身必定形成了一条曲线，它们消耗能量。这就是阿布里科索夫涡旋(Abrikosov vortex)。在它中心之外的地方，

人们会将 ϕ_i 变换回一个常数值，但这产生了一个矢量势 $A_\mu(x)$，满足：

$$\oint A_\mu \mathrm{d}x^\mu = \frac{2\pi}{e} \tag{192}$$

这意味着此旋涡携带了一个磁流的量，其量级精确地为 $2\pi/e$。显然，在这一模型中磁场线凝聚于局域稳定的旋涡中[Nielsen and Olesen(尼尔斯和奥利森)，1973]。这也发生在超导体的磁场中。

10.2 磁单极

如果希格斯场有三个实分量，会有类似的事情发生。那样的话，人们可以把$V(\phi)$最小值的 S_2 球面映射到三维空间中的球面上。在该球面中会有分离的零点。这些客体的行为像是局域稳定的粒子。如果试图把场局域地变换为一个常数值，人们会发现又出现了一个矢量势。

例如，如果在一个 $SU(2)$ 理论中，伴随表征(它有三个实分量)中的一个希格斯粒子将规范群破缺为 $U(1)$，则人们在球面内发现一个分离磁场源的矢量势。这意味着该源是具有磁荷 $g_m = \frac{4\pi}{e}$的一个磁单极，其中 e 是 $SU(2)$ 理论中原来的耦合强度。事实上，早在 1931 年狄拉克[1931；1948]已经得到了磁荷 g_m 和电荷 q 必须满足狄拉克量子化条件：

$$qg_m = 2\pi n \tag{193}$$

显然，对于该模型中的磁单极有 $n=2$。但是，很容易在基本表征中引入粒子，它有 $q = \frac{1}{2}e$，这些内容充实了狄拉克条件(193)。

狄拉克不能对他提出的磁单极的质量说太多有用的信息，但是在现在的理论中质量是可计算的。一般地，磁单极质量被证明为是一个普通粒子的质量除以规范耦合强度平方量级的数。

仔细分析存在的李群和它们可能自发破缺到一个或更多子群 $U(1)$ 的方式，揭示了一个普遍特征：仅当根本的规范群是紧致的且有一个紧覆盖群时，$U(1)$ 规范群中的电荷才被量子化，否则将不会禁止在 $U(1)$ 荷上加上任意的实数，且仅当根本规范群的覆盖群是紧致群时，才可以构造磁单极的解。显然，每当规范群结构为电荷的量子化提供一个有信服力的原因时，才能够保证磁单

极解的存在。因而，假定对于电子与质子的电荷元是相等的且量子化为 e 的倍数大，这自然拥有令人信服的理由，我们必须假定磁单极解必定存在。但是，在绝大多数的“大统一图像(Grand Unification Schemes)”中，相关的质量标度要比今天研究的粒子质量标度高出许多个数量级，因此，除以一个耦合强度平方得到的磁单极的质量甚至更大。

根据一个磁单极的希格斯场结构，人们得到系统在旋转下是不变的，如果旋转与规范旋转相联系在一起的话。它的一个结果是，当限制到一个磁单极时，具有半奇数同位旋的基本粒子产生了具有半奇数整数轨道角动量的束缚态[Hasenfratz and 't Hooft(哈森弗拉茨和特霍夫特)，1976]。关于此令人感到奇怪的是，这样的粒子应该发展出狄拉克统计。事实上，人们可以得出自旋与电荷和磁荷束缚态的统计都从玻色—爱因斯坦统计转换到费米—狄拉克统计，或是从费米—狄拉克统计转换到玻色—爱因斯坦统计[Goldhaber(歌德哈伯)，1976]，如果它们形成了狄拉克量子 n 的奇数值的话，见方程(193)。

10.3 瞬子

具有两个实分量的希格斯场产生了旋涡，具有三个分量的希格斯场产生磁单极，那么当希格斯场有四个实分量时我们会得到什么？例如，它可能会是这样的情形，即通过基本表征(两个复数 = 四个实分量)中的一个希格斯粒子，$SU(2)$ 自发破缺到单位元。人们找到的拓扑上稳定的对象为四维时空中的稳定点。它们表征事件，且涉及了它们的类粒子表象，(在欧几里得空间中)产生的解称为“瞬子(instanton)”。因为这个希格斯场在 $SU(2)$ 情形中完全破坏了规范对称性，所以人们可以论证说这个解在纯的规范理论中同样是拓扑稳定的，完全不需要希格斯机制。在远离原点的地方，矢量势场被描述为值 $A_\mu^a(x)=0$ 的一个定域规范旋转。这样来描述所涉的规范旋转 $\Omega(x)$，即通过注意到 $SU(2)$ 矩阵形成了一个 S_3 空间，是在四维中一个球体的三维表面。将此 S_3 一一映射到(欧几里得)时空中某个大区域的边界上，给出了一个瞬子的场组态。

贝拉文(Belavin)、波利考夫(Polyakov)、施瓦茨(Schwarz)和泰朴金(Tyupkin)[Belavin *et al.*，1975]，他们也是首次写下该解的人，注意到该解有一个非零值：

$$\int \mathrm{d}^4 x F^a_{\mu\nu} \widetilde{F}^a_{\mu\nu} = \frac{32\pi^2}{g^2} \tag{194}$$

被积函数是流(current)的散度：

$$F^a_{\mu\nu} \widetilde{F}^a_{\mu\nu} = \partial_\mu K_\mu; \quad K_\mu = 2\varepsilon_{\mu\nu\alpha\beta} A^a_\nu \left(\partial_\alpha A^a_\beta + \frac{1}{3} g f^{abc} A^b_\alpha A^c_\beta\right) \tag{195}$$

该流是所谓的陈—西蒙(Chern-Simon)流。但是，该流不是规范不变的，这就是为什么它在无穷大时非零的原因。在无穷大处用0来代替 A^a_μ 的规范变换 $\Omega(x)$ 下它会为零。通过利用陈—西蒙流非常容易得到方程(194)。恰巧瞬子也是下述方程的解：

$$F_{\mu\nu} = \widetilde{F}_{\mu\nu} \tag{196}$$

从而我们也发现作用量由 $-8\pi^2/g^2$ 给出。

在一个纯规范(没有费米子)理论中，瞬子可以解释为表征了隧穿跃迁(tunneling transitions)。在普通量子力学中，隧穿是指数衰减的跃迁。如果我们用一个虚量 $i\tau$ 取代时间 t，指数衰减就变为一个振荡表达。振荡指数是在虚时间中的经典跃迁作用量。人们也可以说，一个隧穿跃迁可以用一个经典力学的跃迁来描述，如果势 $V(\vec{q})$ 被 $2E - V(\vec{q})$ 取代，其中 E 是能量。经典作用量则表征了(指数衰减)隧穿跃迁指数中的量。

上面的代换正是通过用 $i\tau$ 取代时间 t 而得到的。在相对论的量子场论中，这也恰好是从闵可夫斯基时空到欧几里得时空的威克(Wick)旋转。简言之，瞬子表征了与衰减因子 $e^{-8\pi^2/g^2}$ 相关联的隧穿。

跃迁也可以通过在时间规范 $A_0 = 0$ 中构建一个规范理论而得到进一步的理解。在此规范中，存在一个在时间无关的规范变换 $\Lambda(\mathbf{x})$ 下的剩余不变性。因此，所有“物理的态”都作为这一定域规范群的表征。但是，通常我们局限于平凡表征 $\Omega \mid \psi\rangle = \mid \psi\rangle$，其中 $\Omega = e^{i\int \Lambda(\mathbf{x})\mathrm{d}^3\mathbf{x}}$，因为这一组态随时间是守恒的，且因为任何其他选择将违背洛伦兹不变性。然而，更详细的分析表明，人们仅需要将此限制施加于那些能够从单位元变换中连续达到的规范变换。通过把 $SU(2)$ 变换的 S_3 空间非平凡地映射到三维空间 $\mathbb{R}_3$ 而得到的变换并不是这样的。这些变换形成了一个离散集，用整数 $k = 0, \pm 1, \pm 2, \cdots$ 来刻画，写作：

$$\Omega_k(\mathbf{x}) = \Omega_1(\mathbf{x})^k, \quad \Omega_k \mid \psi\rangle = e^{i\theta k} \mid \psi\rangle \tag{197}$$

我们发现用瞬子描述的隧穿跃迁引起了理论中物理现象对θ指数地衰减的依赖性。因为，在宇称变换P下，角度θ变换为$-\theta$，一个非零的θ也意味着在强相互作用中明显的宇称(最终是PC)不守恒。

在有费米子出现的时候情况就完全不同了。由于手征反常，对于手征费米子$J^5_\mu(x)$的流，我们有方程(173)。手征费米子的总数$Q^5=\int d^3\mathbf{x}J^5_0(\mathbf{x})$由于一个瞬子而改变一个单位：$\Delta Q^5=\pm 1$。这可以通过注意到无质量的手征费米子的狄拉克方程在瞬子的欧几里得空间中有一个局域化的解来理解。在闵可夫斯基时空中，该解转变为描述一个手征费米子的态，该手征费米子要么消失在狄拉克海中，要么从狄拉克海中出现，从而粒子减去反粒子的数目对于每一个手征费米子种类而言都改变一个单位。如果左手和右手费米子以相同的方式耦合于规范场，如同在QCD中那样，则瞬子去除一个左手费米子并生成一个右手费米子，或换句话说，它翻转了手征性。这一$\Delta Q^5=\pm 2$事件恰好拥有戈德斯通玻色子质量项的量子数，则戈德斯通玻色子与手征荷(η粒子)的守恒相关联。这解释了为什么η粒子要比介子重许多，介子由于夸克的手征对称性失去了它们的大部分质量['t Hooft，1986]。

通过对瞬子的研究能够得出结论：QCD即强相互作用的理论极好地解释了所观察到的强子谱的对称结构，包括用以解释η质量的手征荷守恒的违背。

在弱电领域，同样有瞬子。我们现在看到在夸克与轻子领域中对反常的取消表明了弱电理论的一个重要特征：因为反常并不只遵守夸克领域的规范不变性，夸克被表明并不是精确守恒的。人们发现，瞬子引发了重子数不守恒事件：三个重子(一共是九个不同的夸克)可能会变为三个反轻子，或者反过来也成立。

11. 禁闭

标准模型中的一个重要组成部分是建立在规范群$SU(3)$上的强相互作用的规范理论。夸克是$SU(3)$基本表征中的费米子。观测到的强子粒子都是在$SU(3)$下规范不变的夸克和/或反夸克束缚态的组合。这里有一个重要的问题：

是一种什么力把这些夸克约束在一起？我们已经看到了涡旋解可以视为在磁单极之间引起了一种有意义的力的形态：在磁单极的希格斯理论中，这些磁单极会与阿布里科索夫涡旋绑定在一起。

事实上，这是一种约束力：每个磁单极必须是一个涡旋的终点，涡旋的另一个终点是具有相反磁荷的磁单极。事实上，禁闭将是绝对的：孤立的磁单极不可能存在。因此，曾经一度认为夸克必须是磁单极。然而，这将与夸克仅在很短的距离下有微弱的相互作用而磁荷总是非常强的发现不相符。一个更简洁的观点是，束缚力形成了电涡旋，而不是磁涡旋。电涡旋可以理解为是一个磁涡旋的对偶变换。它发生在当布劳特—恩格勒—希格斯机制影响自由移动的磁荷粒子时。进一步的分析论证和数值研究揭示出，这样的客体确实出现在 QCD 中，且希格斯机制很可能发生在这一领域中。我们将用文字简略解释这一情形。

11.1　极大阿贝尔规范

区分非阿贝尔规范理论与阿贝尔规范理论的一个特征是，规范选择的参照系，即规范条件可以单独从纯规范场角度局域地部分确定，注意到协变场强度 $G_{\mu\nu}$ 作为自伴表征来变换，人们可以选择规范使得这些分量之一，比如说 G_{12} 是对角化的。这样一来就去除了规范群的非贝尔部分，但对角部分也就是称为嘉当(Cartan)子群的部分保留了下来。以这种方式，一个非阿贝尔规范理论转变为一个阿贝尔理论。一个略微巧妙但同样有效的非定域规范为 $\sum_{i\neq j}(A_{\mu j}^{i})^{2}$ 是最小化的条件，这被称为极大阿贝尔规范。

但是，这样一个规范选择会产生奇点。当 G_{12} 的两个本征值相同时通常会出现奇点。不难弄明白，这些奇点表现为粒子，且这些粒子携带有关于子群的磁荷。一旦这些带磁荷的粒子经历布劳特—恩格勒—希格斯机制，就会发生绝对禁闭。

虽然这是解释夸克约束力绝对性质的首选图像，但是要注意到磁荷粒子并不需要直接与禁闭机制有密切关系。相反，它们是指示器(indicators)。我们从禁闭也发生在具有非常大的色数目 N_c 的理论中这一事实得到它，在极限 $N_c\to\infty$ 下，磁荷粒子似乎被抑制在微扰领域，但电涡旋仍然稳定。一个涡旋的强度

是由其有限的宽度确定的，这一宽度是由最轻的胶子态——“胶球(glueball)”来控制的。在与最轻胶球的质量倒数相比很大的距离尺度下，一个电涡旋不能被打破。

禁闭是一个凝聚相位，该相位是布劳特—恩格勒—希格斯相位的逻辑替代。但是，在某些情形下，这两个相位可以共同存在。这样共存的一个例子是标准模型中的$SU(2)$区域。通常，这一区域被看作是希格斯机制的一个原型，但碰巧的是标准模型的$SU(2)$区域可以完全像$SU(3)$区域一样来处理，即好像存在一个限制。为了明白这一点，我们必须看到，希格斯二重态场可以用于清晰地固定规范群的$SU(2)$区域。这意味着所有物理粒子都可以与规范不变源相联系，通过把它们看作是希格斯粒子规范不变的束缚态和模型中其他的基本二重态。例如，把希格斯二重态写作$\phi_a=\begin{pmatrix}F\\0\end{pmatrix}+\tilde{\phi}_a$，轻子二重态写作$\psi_a$，电子被看作是与“重子”场$\varepsilon^{ab}\phi_a\psi_b$相关联，微中子与$\phi^{*a}\psi_a$关联，$Z_0$玻色子与$\phi^{*a}D_\mu\phi_a$关联，以此类推。

在其中真正禁闭相位不同于希格斯相位的理论是那样一些理论，在其中希格斯场不像在$SU(2)$伴随表示中那样是对规范群的一对一表征。

12. 展望

量子场论作为对亚原子粒子准确与充分的描述已经达到一个相当高的水准。从纯数学的角度看，准确性上存在一些内在的局限，据此定义了需要的应用范围，但几乎在所有可想象的境况中，其内在的准确性要比实验所达到的高很多。这并不意味着我们在实际计算中能够达到这样的准确性，事实上实际计算常常会遇到技术限制，特别是当相互作用强的时候，比如在QCD中。这一领域仍然需要大量的技术改进。

12.1 自然性

正如我们今天所知道的，当标准模型用来推算超过大约1TeV的能量范围时，会遇到一个不属于数学性质而是物理性质的困难，也就是很难相信它表示

了真实的世界。当在一个非常精细的点阵上考虑时，裸拉格朗日量需要有这样的参数，即为了产生如希格斯粒子和弱矢量玻色子这样质量远小于1TeV的粒子，而必须被精确调整的参数。这一精细调节(fine-tuning)被认为是不自然的。在一个充分的物理理论中，不希望出现这样的巧合。在某种确定性下，人们可以说自然的基本定律必须允许对高能量有比具有这些精细调谐点阵更为优美的描述。通常期望的要么是一个新的对称性原理，要么是可能拥有物理场完全不同集合的新体系。

完全不同的体系的一个备选是所谓的艺彩(technicolour)理论，它是QCD的一个竞争者，但其典型的能量标度为TeV，而不是GeV。标准模型的夸克、轻子和希格斯粒子都将是该艺彩理论中的强子。不同的规范群会取代这里的$SU(3)$。但是，根据这一方案，将会得到一个新的强相互作用体系，在那里标准模型中区领域中用到的微扰展开不得不崩溃。随着精确测量和计算持续确证着这些微扰展开的可靠性，艺彩方案被认为是不太可能的。

12.2　超对称

一个首选的方案是对标准模型对称性的简单但巧妙的升级：超对称。这一对称性，它把费米子与玻色子归纳为单个多重态，并不真的改变理论的基本方面。但它在应用范围内对振幅的表达式做了相当的简化，不只是在微扰领域中，同样在许多情形下，它允许我们深入到理论的非微扰领域。关于超对称有非常多的文献，但它的一些方面仍然有些模糊。我们需要知道更多关于超对称的物理起源与意义，还有在今天实验观测可及的标准模型范围内引起其破缺——且使得大多数不可见——的机制。

12.3　微扰展开的重求和

对于耦合参数的任意值而言，量子场论中的微扰展开几乎肯定是发散的。戴森(Dyson)提出对该发散的一个简单论证[Dyson，1952]：想象在QED理论中存在一个边界ε，使得每当$|\alpha|<\varepsilon$时微扰展开都会收敛，其中α是精细结构常数。它对于α的某些负实数值是收敛的。但很容易断定，对于α的任意负值，真空将是不稳定的：真空涨落允许大量电子对产生，且因为像电荷一样相

互吸引，高负荷的电子云可能有负的能量。

具有渐近自由的理论允许以一种自然的方式去重新求微扰级数的和，先在高能时非常精确地求解理论，之后人们不得不对薛定谔方程积分以得到在低能时的物理振幅。目前还没有完成这一计划，因为对这些薛定谔方程的积分超出了我们当前的能力，但是人们猜测原则上它应该是可能的。如果能够建立一个紫外固定点的话，可能会允许对非渐近自由的理论进行更为精确的处理。

可以研究和预测微扰展开的发散程度。这里确实用到的是波莱尔(Borel)重求和方法。振幅为：

$$\Gamma(\lambda) = \sum_{n=1}^{\infty} a_n \lambda^n \tag{198}$$

可以改写为：

$$\Gamma(\lambda) = \int_0^{\infty} B(z) e^{-z/\lambda} dz$$

$$B(z) = \sum_{k=0}^{\infty} a_{k+1} z^k / k! \tag{199}$$

一般地，人们期望 $B(z)$ 的级数拥有一个有限的收敛半径。如果 $B(z)$ 能够解析地延拓到 $0 \leqslant z < \infty$ 区间，则它(重新)定义了我们的振幅。但是，一般地，人们可以得出 $B(z)$ 必须在实数轴上有奇点，例如在 z 对应于瞬子或瞬子对的作用量的地方。此外，人们期望得到与理论的红外和/或紫外发散相关联的奇点。有时，这些不同的奇点相互干涉。

12.4 广义相对论和超弦理论

但是，微扰展开的收敛或发散问题是否是物理相关的，关于这一问题仍然是不确定的。我们知道量子场论不能够包含关于亚原子世界的全部真理，引力肯定不会是可重整化的，所以在那些引力可与其他相互作用相比较的尺度上，即在所谓的普朗克尺度上需要一个全新的理论。当前最有希望能够发展成这样一个理论的是超弦理论。在该理论下，物理学家正开启新的篇章，在其中我们丢弃了传统的量子场论，正如在本章中描述的那样。在它当前的形式下，超弦理论似乎已变成大胆想法的集合，被称为 M 理论，其基础仍然非常不稳定。一些世界上最聪明的人在竞相将此理论变为某个能够用以提供可靠预言并能在教

科书中讲授的东西，但这尚未实现。

参考文献

[Adler, 1969] S. Adler. Axial-vector vertex in spinor electrodynamics. *Physical Review*, 177: 2426-2438, 1969.

[Adler and Bardeen, 1969] S. Adler and W. Bardeen. Absence of higher-order corrections in the anomalous axial-vector divergence equation. *Physical Review*, 182: 1517-1536, 1969.

[Aitchison and Hey, 1989] I. Aitchison and A. Hey. *Gauge Theories in Particle Physics.* Adam Hilger, 1989.

[Bardeen, 1969] W. Bardeen. Anomalous Ward identities in spinor field theories. *Physical Review*, 184: 1848-1859, 1969.

[Becchi *et al.*, 1975] C. Becchi, A. Rouet, and R. Stora. Renormalization of the Abelian Higgs-Kibble model. *Communications in Mathematical Physics*, 42: 127-162, 1975.

[Becchi *et al.*, 1976] C. Becchi, A. Rouet, and R. Stora. Renormalization of gauge theories. *Annals of Physics*, 98: 287-321, 1976.

[Belavin et al., 1975] A. Belavin, A. Polyakov, A. Schwartz, and Y. Tyupkin. Pseudoparticle solutions of the Yang-Mills equations. *Physics Letters*, 59 B: 85-87, 1975.

[Bell and Jackiw, 1969] J. Bell and R. Jackiw. A pcac puzzle: pi /sup 0/ to gamma gamma in the sigma-model. *Il Nuovo Cimento*, 60A: 47-61, v.

[Cheng and Li, 1984] T. Cheng and L. Li. *Gauge theory of elementary particle physics.* Clarendon Press, Oxford, 1984.

[Crease and Mann, 1986] R. Crease and C. Mann. *The Second Creation: makers of the revolution in twentieth-century physics.* Macmillan, New York, 1986.

[de Wit and Smith, 1986] B. de Wit and J. Smith. *Field Theory in Particle Physics.* North Holland, 1986.

[DeWitt, 1964] B. DeWitt. Theory of radiative corrections for non-Abelian gauge fields. *Physical Review Letters*, 12: 742-746, 1964.

[DeWitt, 1967a] B. DeWitt. Quantum theory of gravity. I. The canonical theory. *Physical Review*, 160: 1113-1148, 1967a.

[DeWitt, 1967b] B. DeWitt. Quantum theory of gravity. II. The manifestly covariant theory. *Physical Review*, 162: 1195-1239, 1967b.

[Dirac, 1931] P. Dirac. Quantised singularities in the electromagnetic field. *Proceedings of the Royal Society*, A133: 60-72, 1931.

[Dirac, 1948] P. Dirac. The theory of magnetic poles. *Physical Review*, 74: 817-830, 1948.

[Dyson, 1952] F. Dyson. Divergence of perturbation theory in quantum electrodynamics. *Physical Review*, 85: 631-632, 1952.

[Englert and Brout, 1964] F. Englert and R. Brout. Broken symmetry and the mass of gauge vector mesons. *Physical Review Letters*, 13: 321-323, 1964.

[Faddeev, 1969] L. Faddeev. The Feynman integral for singular Lagrangians. *Theoretical and Mathematical Physics*, 1: 3-18, 1969. Russian. English summary.

[Faddeev and Popov, 1967] L. Faddeev and V. Popov. Feynman diagrams for the Yang-Mills field. *Physics Letters*, 25B: 27-29, 1967.

[Faddeev and Takhtajan, 1984] L. Faddeev and L. Takhtajan. *Soviet Journal of Mathematics*, 24: 241, 1984.

[Gell-Mann and Low, 1954] M. Gell-Mann and F. Low. Quantum electrodynamics at small distances. *Physical Review*, 95: 1300-1312, 1954.

[Goldhaber, 1976] A. Goldhaber. Connection fo spin and statistics for charge-monopole composites. *Physical Review Letters*, 36: 1122-1125, 1976.

[Hasenfratz and 't Hooft, 1976] P. Hasenfratz and G. 't Hooft. Fermion-boson puzzle in a gauge theory. *Physical Review Letters*, 36: 1119-1122, 1976.

[Higgs, 1964a] P. Higgs. Broken symmetries and the masses of gauge bosons. *Physical Review Letters*, 13: 508-509, 1964a.

[Higgs, 1964b] P. Higgs. Broken symmetries, massless particles and gauge fields. *Physics Letters*, 12: 132-133, 1964b.

[Higgs, 1966] P. Higgs. Spontaneous symmetry breakdown without massless bosons. *Physical Review*, 145: 1156-1163, 1966.

[Hoddeson *et al.*, 1997] L. Hoddeson, L. Brown, M. Riordan, and M. Dresden (eds.). *The Rise of the Standard Model, Particle Physics in the* 1960*s and* 1970*s*. Cambridge University Press, 1997.

[Itzykson and Zuber, 1980] C. Itzykson and B. Zuber. *Quantum Field Theory.* McGraw Hill, New York, 1980.

[Jackiw, 1995] R. Jackiw. *Diverse topics in Theoretical and Mathematical Physics.* World Scientific, 1995.

[Nielsen and Olesen, 1973] H. Nielsen and P. Olesen. Vortex-line models for dual strings. *Nuclear Physics B*, 61: 45-61, 1973.

[Pais, 1986] A. Pais. *Inward bound: of matter and forces in the physical world.* Oxford University Press, 1986.

[Ryder, 1985] L. Ryder. *Quantum Field Theory.* Cambridge Univ. Press, 1985.

[Slavnov, 1972] A. Slavnov. Ward identities in gauge theories. *Theoretical and Mathematical Physics*, 10: 99-107, 1972.

[Stueckelberg and Peterman, 1953] E. Stueckelberg and A. Peterman. *Helvetica Physica Acta*, 26: 499-, 1953.

['t Hooft, 1971] G. 't Hooft. Renormalization of massless Yang-Mills fields. *Nuclear Physics B*, 33: 173-199, 1971.

['t Hooft, 1986] G. 't Hooft. How instantons solve the U(1) problem. *Physics Reports*, 142: 357-387, 1986.

['t Hooft and Veltman, 1972a] G. 't Hooft and M. Veltman.. Combinatorics of gauge fields. *Nuclear Physics B*, 50: 318-353, 1972a.

['t Hooft and Veltman, 1972b] G. 't Hooft and M. Veltman. Regularization and renormalization of gauge fields. *Nuclear Physics B*, 44: 189-213, 1972b.

['t Hooft and Veltman, 1994] G. 't Hooft and M. Veltman. *Under the Spell of the Gauge Principle*, volume 19 of *Advanced Series in Mathematical Physics.* World Scientific, Singapore, 1994.

[Taylor, 1971] J. Taylor. Ward identities and charge renormalization of the Yang-Mills field. *Nuclear Physics B*, 33: 436-444, 1971.

[Tyutin, 1975] I. Tyutin. Lebedev Preprint. FIAN 39. unpublished, 1975.

[Yang and Mills, 1954] C. Yang and R. Mills. Conservation of isotopic spin and isotopic gauge invariance. *Physical Review*, 96: 191-195, 1954.

第八章　代数量子场论

汉斯·霍尔沃森

导论

看见这一章的标题，人们可能猜测这个主题表示的方法对量子场论(QFT)来说会有些另类。就研究人数而言该方法的确有些另类：从事研究QFT的人中只有一小部分研究代数量子场论(AQFT)。然而，哲学家和那些对基础问题感兴趣的人，要研究这个“代数”方法具有特定的原因。

在分析传统的科学哲学中，研究一个理论T的基础最起码要澄清这个理论T所指示的对象。(而且，要把哲学跟社会学和历史学分开，T不能指向某些人群的活动。)在二十世纪早期，认为T的指示对象一定是某些最好为一阶形式语言公理的一个集合，人们很快就认识到并没有多少有意义的物理理论能按这样的方式形式化，而在很多情况下，我们无法置身于公理之中，这是费耶阿本德(Feyerabend)的说法。因此，这些标准被多少放宽了些，但也只是放宽到专业数学家团体对这些标准同时所能放宽的程度。在很多哲学家中存在一个不成文的前提，研究一个理论的基础要求这个理论具有一个数学描述(这个哲学家的前提肯定符合统计力学、狭义和广义相对论以及非

相对论量子力学的情况)。在任何情况下，不管拥有的数学描述是不是强制性的，拥有这样一个描述对于我们能够作出安全有效的推论具有巨大帮助。

因此，物理学哲学家们把他们的研究对象定为理论，这些理论对应着数学对象(或许是模型集)，但是不太清楚"量子场论"能够置于数学世界的什么地方。在缺乏某种数学上明确的有关 QFT 描述的情况下，物理学哲学家有两种意见：要么找到一种新方法解释任务，要么对量子场论的解释保持沉默。①

正是由于这个原因，AQFT 跟量子场论基础特别相关。简单讲，AQFT 是我们有关 QFT 处于数学世界什么位置的最好说法，因而是基础研究的自然起点。

1. 代数学前言

第 1 节为本章其余部分的数学前提提供一个简短概要。

1.1 冯·诺伊曼(von Neumann)代数

冯·诺伊曼代数的标准定义牵涉到拓扑学，而且随后就会通过冯·诺伊曼的二次交换定理(double commutant theorem)表明，这个拓扑学条件跟代数条件(定义 2 中的条件 2)是一致的。而就当前的目的来说，以代数条件为基础就足够了。

定义 1。设$\mathcal{H}$是个希尔伯特(Hilbert)空间，假设$\mathfrak{B}(\mathcal{H})$是$\mathcal{H}$上有界线性算子的集合，指的是对每一个 $A\in\mathfrak{B}(\mathcal{H})$存在最小非负数 $\|A\|$，使得对于所有的单位向量 $x\in\mathcal{H}$都有$\langle Ax, Ax\rangle^{1/2}\leq\|A\|$。接下来我们会用 $\|\cdot\|$ 同时表示$\mathcal{H}$上的范数和$\mathfrak{B}(\mathcal{H})$上的范数。我们利用并列 AB 表示$\mathfrak{B}(\mathcal{H})$的两个元素 A，B 的合成。对每个 $A\in\mathfrak{B}(\mathcal{H})$，我们设 A^* 表示$\mathfrak{B}(\mathcal{H})$的唯一元素，使得对所有的 x，$y\in\mathfrak{R}$，都有$\langle A^*x, y\rangle=\langle x, Ay\rangle$。

定义 2。设$\mathfrak{R}$是$\mathfrak{B}(\mathcal{H})$的 * 子代数，$\mathfrak{B}(\mathcal{H})$是作用在希尔伯特空间上的有界算子。如果：

① 对第一种意见，参看 Wallace(华莱士)有关论文。

1. $I\in\mathfrak{R}$

2. $(\mathfrak{R}')'=\mathfrak{R}$，其中$\mathfrak{R}'=\{B\in\mathfrak{B}(\mathcal{H})$：$[B,A]=0$，$\forall A\in\mathfrak{R}\}$

那么$\mathfrak{R}$是冯 · 诺伊曼代数。

定义 3。我们需要在$\mathcal{H}$上的有界线性算子集$\mathfrak{B}(\mathcal{H})$上面的四个标准拓扑。这些拓扑每一个都是通过半范数族来定义的，详情参见[Kadison and Ringrose（卡迪孙和林罗斯），1997，Chapters. 1，5]。

· $\mathfrak{B}(\mathcal{H})$上的一致拓扑是通过单范数定义的：

$$\|A\|=\sup\{\|Av\|:\ v\in\mathcal{H},\ \|v\|\leqslant 1\}$$

其中右边的范数是在$\mathcal{H}$上的给定向量范数，因此，算符A是当且仅当$(\|A_i-A\|)_{i\in\mathbb{N}}$收敛为0时序列$(A_i)_{i\in\mathbb{N}}$的一个极限点。

· 在$\mathfrak{B}(\mathcal{H})$上的弱拓扑是通过半范数族$\{p_{u,v}$：$u,\ v\in\mathcal{H}\}$定义的，其中：

$$p_{u,v}(A)=\langle u,\ Av\rangle$$

相应的拓扑一般不是第一可数的，因而$\mathfrak{B}(\mathcal{H})$的一个子集$S$的闭包一般也比$S$中所有序列极限点的集合大。更准确地说，$S$的闭包是$S$中一般化序列的极限点（网点）的集合，详情参见[Kadison and Ringrose，1997，Chapter 1]。$\mathfrak{B}(\mathcal{H})$中的网$(A_i)_{i\in\mathbb{N}}$弱收敛于$A$，当且仅当$(p_{u,v}(A_i))_{i\in\mathcal{I}}$收敛于对所有$u,\ v\in\mathcal{H}$的$p_{u,v}(A)$。

· 在$\mathfrak{B}(\mathcal{H})$上的强拓扑是通过半范数族$\{p_v$：$v\in\mathcal{H}\}$来定义的，其中：

$$p_v(A)=\|Av\|$$

因此，网$(A_i)_{i\in\mathcal{I}}$强收敛于A，当且仅当$(p_v(A_i))_{i\in\mathcal{I}}$对$v\in\mathcal{H}$收敛于$p_v(A)$。

· 在$\mathfrak{B}(\mathcal{H})$上的超弱拓扑是通过族$\{p_\rho$：$\rho\in\mathcal{I}(\mathcal{H})\}$来定义的，其中$\mathcal{I}(\mathcal{H})$是正的迹为1的$\mathcal{H}$上算符（"稠算符"）的集合，并且：

$$p_\rho(A)=\mathrm{Tr}(\rho A)$$

因此，网$(A_i)_{i\in\mathcal{I}}$超弱地收敛于A，仅只在情况$(\mathrm{Tr}(\rho A_i))_{i\in\mathcal{I}}$中对所有$\rho\in\mathcal{I}(\mathcal{H})$收敛于$\mathrm{Tr}(\rho A)$。

事实 4。这些拓扑顺序如下：

$$\begin{array}{ccccc} & \subsetneq & \text{正常} & \supsetneq & \\ \text{超弱} & & & & \text{强} \\ & \supsetneq & \text{弱} & \subsetneq & \end{array}$$

由于闭集正好是开集的补集，这就意味着弱闭集是超弱闭的，而超弱闭子集是正常闭包，甚至$\mathfrak{B}(\mathcal{H})$上的四个拓扑当且仅当$\mathcal{H}$是有限维时是拓扑结构一致的。

事实 5。如果 S 是$\mathfrak{B}(\mathcal{H})$有界的凸子集，那么 S 的弱、超弱和强闭合都是一样的。

事实 6。对于一个包含 I 的$\mathcal{H}$上的 $*$ 代数$\mathfrak{R}$，下面的情况是等价的：(i)$\mathfrak{R}$是弱闭的；(ii)$\mathfrak{R}''=\mathfrak{R}$。这就是冯·诺伊曼二次交换定理。

定义 7。设$\mathfrak{R}$是$\mathfrak{R}(\mathcal{H})$的子集，向量 $x\in\mathcal{H}$被称为对$\mathfrak{R}$循环，仅在情况$[\mathfrak{R}x]=\mathcal{H}$中，其中$\mathfrak{R}x=\{Ax:\ A\in\mathfrak{R}\}$，而$[\mathfrak{R}x]$是$\mathfrak{R}x$ 生成的封闭线性子空间。向量 $x\in\mathcal{H}$被称为对$\mathfrak{R}$分离，仅在情况 $Ax=0$ 和 $A\in\mathfrak{R}$蕴含 $A=0$ 时。

事实 8。设$\mathfrak{R}$是$\mathcal{H}$上的一个冯·诺伊曼代数，并设 $x\in\mathcal{H}$。那么当且仅当 x 对$\mathfrak{R}'$分离时，x 对$\mathfrak{R}$是循环的。

定义 9。冯·诺伊曼代数$\mathfrak{R}$的正常状态是一个超弱连续状态，我们用$\mathfrak{R}^*$表示 R 的正常状态空间。

1.2　C^*代数及其表征

定义 10。一个 C^*代数是由 $*$ 代数$\mathfrak{A}$ 和范数 $\|\cdot\|:\ \mathfrak{A}\to\mathbb{R}$ 构成的一个对，使得

$$\|AB\|\leqslant\|A\|\cdot\|B\|,\qquad \|A^*A\|=\|A\|^2$$

对所有 A，$B\in\mathfrak{A}$ 成立。我们通常用$\mathfrak{A}$ 表示代数及其范数。

在本章里我们仅仅会用到包含乘法单位元 I 的 C^* 代数。

定义 11。一个$\mathfrak{A}$ 上的状态 ω 是一个线性泛函，使得对所有的 $A\in\mathfrak{A}$ 有 $\omega(A^*A)\geqslant0$，并且 $\omega(I)=1$。

定义 12。一个$\mathfrak{A}$ 的状态 ω 叫作混合态的条件是 $\omega=\frac{1}{2}(\omega_1+\omega_2)$，而 $\omega_1\neq\omega_2$；否则 ω 叫作纯态。

定义 13。设$\mathfrak{A}$ 是一个 C^* 代数，$\mathfrak{A}$ 的一个表示是一对$(\mathcal{H},\ \pi)$，其中$\mathcal{H}$是希尔伯特空间，而 π 是$\mathfrak{A}$ 到$\mathfrak{B}(\mathcal{H})$的 $*$ 同态。如果 $\pi(\mathfrak{A})$ 在$\mathfrak{B}(\mathcal{H})$中是弱稠的，则表示$(\mathcal{H},\ \pi)$是不可约化的。如果 π 是同构的，则表示$(\mathcal{H},\ \pi)$是忠实表示。

定义 14。设($\mathcal{H}$, π)和($\mathcal{K}$, ϕ)是 C^* 代数$\mathfrak{A}$的表示，那么($\mathcal{H}$, π)和($\mathcal{K}$, ϕ)被说成是：

1. 酉等价的(unitarily equivalent)。如果存在酉 U: $\mathcal{H}\to\mathcal{K}$，使得对所有 $A\in\mathfrak{A}$ 都有 $U\pi(A)=\phi(A)U$。

2. 准等价的。如果冯·诺伊曼代数 $\pi(\mathfrak{A})''$和 $\phi(\mathfrak{A})''$是 * 同构的。

3. 不相交的。如果它们不是准等价的。

定义 15。一个表示($\mathcal{K}$, ϕ)在如下情况下被叫作是($\mathcal{H}$, π)的子表示，仅在存在同构 V: $\mathcal{K}\to\mathcal{H}$的情况下成立，使得对所有的 $A\in\mathfrak{A}$ 都有 $\pi(A)V=V\phi(A)$。

事实 16。两个表示是准等价的当且仅当它们具有酉等价的子表示。

著名的盖尔范德—奈马克—西格尔定理(Gelfand-Naimark-Segaltheorem，简称 GNS)表明每个 C^* 代数态都能用一个希尔伯特空间中的向量表示。

定理 17(GNS)。设 ω 是$\mathfrak{A}$的一个态，那么存在$\mathfrak{A}$的一个表示($\mathcal{H}$, π)，以及一个单位向量 $\Omega\in\mathcal{H}$，使得：

1. $\omega(A)=\langle\Omega, \pi(A)\Omega\rangle$，对于所有的 $A\in\mathfrak{A}$；

2. $\pi(\mathfrak{A})\Omega$ 在$\mathcal{H}$中是稠的。

而且，表征($\mathcal{H}$, π)是唯一一个(达到酉等价)满足这两个条件的。

由于后面我们需要涉及 GNS 结构的细节，我们在此概述一下其证明思路。

证明思路。我们通过$\mathfrak{A}$中的等价类元素建构希尔伯特空间$\mathcal{H}$，并且表示 π 是通过左乘作用得到的。特别是，在$\mathfrak{A}$上定义一个有界半双线性形式，即：

$$\langle A, B\rangle_{\omega}=\omega(A^*B), \qquad A, B\in\mathfrak{A}$$

设$\mathcal{H}_0$是通过范数 $\|A\|_{\omega}=\langle A, A\rangle_{\omega}^{1/2}$ 约化的$\mathfrak{A}$的商，假设$\mathcal{H}$是准希尔伯特空间$\mathcal{H}_0$的唯一完备，因此存在一个具有$\mathcal{H}$中 $j(\mathfrak{A})$稠的包含映射 J: $\mathfrak{A}\to\mathcal{H}$。定义$\mathcal{H}$上的算符 $\pi(A)$，即：

$$\pi(A)j(B)=j(AB), \qquad B\in\mathfrak{A}$$

一定能够证明 $\pi(A)$是确定的，而且可以唯一地推广到$\mathcal{H}$上的有界线性算子。也一定能够证明 π 是个 * 同态。最后，如果我们设 $\Omega=j(I)$，那么 Ω 显然对$\pi(\mathfrak{A})$是循环的。

定义 18。设 ω 是$\mathfrak{A}$的一个态，$\mathfrak{A}$的由 ω 约化的 GNS 表示($\mathcal{H}$, π)是不可约

化的，当且仅当ω是纯态。

注：关于C^*代数讨论的标准参考文献包括[Kadison and Ringrose, 1997]和[Takesaki(竹崎)，2002]。

1.3 冯·诺伊曼代数的类型划分

定义19。在冯·诺伊曼代数$\mathfrak{R}$中两个映射E，F被说成是等价的，写作$E\sim F$，当且仅当存在$V\in\mathfrak{R}$使得$V^*V=E$和$VV^*=F$。

注意20。如果我们再仔细点，我们应该在上述定义中用“等价模$\mathfrak{R}$”替代“等价”，并且类似地用“$\sim_{\mathfrak{R}}$”替代“$\sim$”，但是我们不指明$\mathfrak{R}$也没有什么影响。在前面定义中算符V被称为具有初始投影E和最终投影F的部分等距。

定义21。对于冯·诺伊曼代数$\mathfrak{R}_1$和$\mathfrak{R}_2$，另外假设$\mathfrak{R}_1\wedge\mathfrak{R}_2=\mathfrak{R}_1\cap\mathfrak{R}_2$。另外再假设$\mathfrak{R}_1\vee\mathfrak{R}_2$代表由$\mathfrak{R}_1$和$\mathfrak{R}_2$产生的冯·诺伊曼代数，即包含$\mathfrak{R}_1$和$\mathfrak{R}_2$的全部冯·诺伊曼代数的交集。

定义22。$Z(\mathfrak{R})=\mathfrak{R}\wedge\mathfrak{R}'$称为冯·诺伊曼代数$\mathfrak{R}$的核心。一个冯·诺伊曼代数$\mathfrak{R}$当且仅当$Z(\mathfrak{R})=\mathbb{C}I$时被叫作因子，等价地，$\mathfrak{R}\vee\mathfrak{R}'=\mathfrak{B}(\mathcal{H})$。一个投影$E\in Z(\mathfrak{R})$被称为在$\mathfrak{R}$里的一个中心投影。

定义23。设$E\in\mathfrak{R}$是个投影，并且假设$E\mathfrak{R}E=\{EAE: A\in\mathfrak{R}\}$。那么显然，$E\mathfrak{R}E$是$\mathfrak{R}$的线性子空间。而且，因为对于$A$，$B\in\mathfrak{R}$，有$AEB\in\mathfrak{R}$和$A^*\in\mathfrak{R}$，因此$E\mathfrak{R}E$在积作用下和$*$作用下是封闭的。也不难看到$E\mathfrak{R}E$是弱闭合的，从而是个$E\mathcal{H}$上的冯·诺伊曼代数。

定义24。设$\mathfrak{R}$是个冯·诺伊曼代数，一个投影$E\in\mathfrak{R}$被说成是：

1. 最小的，仅在$\mathfrak{R}$不包含E的严格子投影的情况下；
2. 阿贝尔的，仅在代数$E\mathfrak{R}E$是阿贝尔(Abelian)情况下；
3. 无限的，仅在存在投影$E_0\in\mathfrak{R}$使得$E_0<E$和$E\sim E_0$的情况下；
4. 有限的，仅在它不是无限的情况下；
5. 真正无限的，仅在E是无限的，而且对于每个$\mathfrak{R}$的中心投影P，要么$PE=0$，要么PE是无限的情况下。

事实25。我们有下面的投影关系：

$$\text{最小的} \Rightarrow \text{阿贝尔的} \Rightarrow \text{有限的}$$

$$\text{真正无限的} \Rightarrow \text{无限的} \Leftrightarrow \neg\ \text{有限的}$$

对这些因子，两行的第一个箭头都可以反过来。

现在我们给出了因子的莫雷—冯·诺伊曼类型(Murray-von Neumann type)分类，详情请见[Kadison and Ringrose, 1997, Chapter 7]和[Sunder(森德), 1987, Chapter 1]。

定义 26。冯·诺伊曼因子分为：

1. 类型Ⅰ，如果包含一个阿贝尔投影；
2. 类型Ⅱ，如果包含一个有限投影，而没有阿贝尔投影；
3. 类型Ⅲ，既不是类型Ⅰ也不是类型Ⅱ。

类型Ⅰ因子已经被莫雷和冯·诺伊曼完全分类为：对每一个基数 κ 存在唯一(达到同构)类型 I_κ 因子，即$\mathfrak{B}(\mathcal{H})$，其中$\mathcal{H}$是 κ 维希尔伯特空间。类型Ⅱ因子能够被进一步划分，按照单位投影是有限(类型II_1)还是无限(类型II_∞)。类型Ⅲ因子能够被再分成III_λ，其中 $\lambda \in [0, 1]$，尽管这个子分类的基础依赖于富田—竹崎(Tomita-Takesaki)模理论(参见 1.4 节)。

对于一般的冯·诺伊曼代数，类型划分应该只是有点更复杂：类型Ⅰ代数被定义为，具有阿贝尔投影 E，使得没有在 $Z(\mathfrak{R})$ 中的非平凡投影优于 E。类似地，类型Ⅱ代数被定义为，具有有限投影 E，使得没有在 $Z(\mathfrak{R})$ 中的非平凡投影优于 E。因此我们有：

命题 27。设$\mathfrak{R}$是个冯·诺伊曼代数，那么$\mathfrak{R} = \mathfrak{R}_{\mathrm{I}} \oplus \mathfrak{R}_{\mathrm{II}} \oplus \mathfrak{R}_{\mathrm{III}}$，这里$\mathfrak{R}_X$是对 $X = \mathrm{I}, \mathrm{II}, \mathrm{III}$ 类型的 X。

证明。参见[Kadison and Ringrose, 1997, Thm. 6. 5. 2]。

我们很快就会看到 QFT 中的局域代数是“典型的”类型Ⅲ，而这具有很多有意义的含义，类型Ⅲ代数不具有阿贝尔投影的事实，跟 3.3 节中的局域性问题是联系在一起的；类型III_1因子的状态空间是齐次的事实，也是跟 3.3 节中的局域性问题联系在一起的；类型Ⅲ代数不包含其态表征(即稠算子)的事实，是跟第 5 节中 QFT 的模态解释联系在一起的。

冯·诺伊曼代数的下述分类也是自然的，但是它打破了莫雷—冯·诺伊曼分类。

定义 28。一个冯·诺伊曼代数$\mathfrak{R}$被分为：

· 无限类型的，如果 I 在$\mathfrak{R}$中是无限的；

· 真正无限的，如果 I 在$\mathfrak{R}$中是真正无限的；

· 半无限的，如果在$\mathfrak{R}$中的中心投影 E_{III}（在命题 27 中定义的）是零。

有限因子包括类型I_n和类型II_1因子，无限因子包括类型I_∞因子和类型II_∞因子以及类型Ⅲ因子，有限和无限因子之间的划分是跟迹态的存在是一致的。

定义 29。冯·诺伊曼代数$\mathfrak{R}$上一个忠实的正规化迹是一个$\mathfrak{R}$上的态 ρ，使得：

1. ρ 是有迹的，即 $\rho(AB)=\rho(BA)$，其中所有 A，$B\in\mathfrak{R}$；

2. ρ 是忠实性的，即 $\rho(A^*A)=0$ 仅当 $A=0$。

事实 30。一个冯·诺伊曼代数$\mathfrak{R}$是有限的，当且仅当存在一个$\mathfrak{R}$上的忠实正常迹态 ρ，一个冯·诺伊曼因子$\mathfrak{R}$是半无限的，当且仅当存在$\mathfrak{R}$上的"忠实正常迹态"；但是我们并不限于定义的这个概念。

1.4　模理论

我们这里只是叙述而不证明有关富田—竹崎模理论的某些基本事实，这些事实为理解冯·诺伊曼类型Ⅲ的分类所必需，后者反过来对理解 AQFT 的数学结构是必不可少的。

定义 31。设$\mathfrak{R}$是作用在希尔伯特空间$\mathcal{H}$上的冯·诺伊曼代数，并且假定 $\Omega\in\mathcal{H}$对$\mathfrak{R}$是循环的和分离的。在此情况下，我们说($\mathfrak{R}$，Ω)处于标准形式。定义$\mathcal{H}$上 S_0 算符，即：

$$S_0A\Omega=A^*\Omega,\qquad A\in\mathfrak{R}$$

那么 S_0 扩展到$\mathcal{H}$上闭的反线性算子 S，设 $S=J\Delta^{1/2}$是 S 的极分解，因此 Δ 是正的（而一般是非有界的），而 J 是反酉的（回顾一个正算子具有$\mathbb{R}^+$内的谱）。我们称为模算子而 J 为跟($\mathfrak{R}$，Ω)对相关联的模共轭。

定理 32。设$\mathfrak{R}$是一个具有循环和分离向量 Ω 的冯·诺伊曼代数，那么$J\Omega=\Omega=\Delta\Omega$，并且：

$$\Delta^{it}\mathfrak{R}\Delta^{-it}=\mathfrak{R},\qquad \forall t\in\mathbb{R}$$

$$J\mathfrak{R}J=\mathfrak{R}'$$

证明。参见[Kadison and Ringrose, 1997, Thm. 9.2.9]或者[Sunder, 1987, Thm. 2.3.3]。

定义 33。设$(\mathfrak{R}, \Omega)$处于标准形式，并且假设ω是由Ω引入的$\mathfrak{R}$的态，对于每一个$t \in \mathbb{R}$，如下定义$\mathfrak{R}$的模自同构σ_t^ω：

$$\sigma_t^\omega(A) = \Delta^{it} A \Delta^{-it}, \qquad A \in \mathfrak{R}$$

对于所有$A \in \mathfrak{R}$。通过取$\gamma(A) = JA^* J$，来定义一个*反同构$\gamma: \mathfrak{R} \to \mathfrak{R}'$，对于所有$A \in \mathfrak{R}$。

定义 34。如果$\mathfrak{A}$是一个C^*代数，我们设$\mathrm{Inn}\,\mathfrak{A}$代表$\mathfrak{A}$的内自同构群；即$\alpha \in \mathrm{Inn}\,\mathfrak{A}$，当且仅当存在一个酉$U \in \mathfrak{A}$，使得对于所有$A \in \mathfrak{A}$都有$\alpha(A) = UAU^*$。

模算子Δ的谱给了模自同构群$(\sigma_t^\omega)_{t \in \mathbb{R}}$周期的粗糙测度；即$\Delta$的谱越小，自同构$\sigma_t^\omega$越接近恒等$\iota: \mathfrak{R} \to \mathfrak{R}$。在极端情况下，如果$\mathrm{sp}\Delta = \{1\}$，那么对于所有的$t \in \mathbb{R}$都有$\sigma_t^\omega = \iota$。相反，当$\Delta$达到$\mathbb{R}^+$，群$(\sigma_t^\omega)_{t \in \mathbb{R}}$倾向于是各态历经的（即没有固定点）。

定义 35。定义$\mathfrak{R}$的模谱$S(\mathfrak{R})$为：

$$S(\mathfrak{R}) = \bigcap_\omega \mathrm{sp}(\Delta_\omega)$$

其中ω跑遍$\mathfrak{R}$的忠实正常态的族，而Δ_ω是相应的模算子。

命题 36。设$\mathfrak{R}$是一个具有循环和分离向量Ω的冯·诺伊曼因子，那么下述说法是等价的：

1. $\mathfrak{R}$是半有限的。

2. 对于所有$t \in \mathbb{R}$，模自同构σ_t^ω是内部的；即存在一个酉$U \in \mathfrak{R}$，使得对于$A \in \mathfrak{R}$都有$\sigma_t^\omega(A) = UAU^*$。

3. $S(\mathfrak{R}) = \{1\}$。

证明。参见[Takesaki, 2003, 122]和[Sunder, 1987, 111]。

我们现在着手孔耐(Connes)的类型Ⅲ因子的子分类。这个子分类用了"权流的周期"（其中权是态概念的推广）。当然，为了绕过一些背景知识，我们使用了下述（可证明是等价的）定义。

定义 37。类型Ⅲ的$\mathfrak{R}$因子被认为是：

1. 类型III_0，如果$S(\mathfrak{R})=\{0, 1\}$。

2. 类型III_λ，$\lambda\in(0, 1)$，如果$S(\mathfrak{R})=\{\lambda^n: n\in\mathbb{Z}\}\cup\{0\}$。

3. 类型III_1，如果$S(\mathfrak{R})=\mathbb{R}^+$。

定义 37 中的条件并不具备明显的物理解释。也就是说，对$\lambda\neq\mu$，类型III_λ代数的物理（如果有的话）跟类型III_μ代数的物理如何不同并不明显。然而，孔耐和斯托默（Connes and Størmer）[1978]的一个结果，兑现了区分两个不同类型因子的某些重要意义。

定义 38。设$\mathfrak{R}$是一个冯·诺伊曼代数，并且假设$\mathfrak{R}_*$是其正常态空间，我们定义态轨迹空间的直径$d(\mathfrak{R})$为：

$$d(\mathfrak{R})=\sup\{\inf\{\|(\omega_1\circ\alpha)-\omega_2\|:\alpha\in\mathrm{Inn}\,\mathfrak{R}\}:\omega_1, \omega_2\in\mathfrak{R}_*\}$$

或者，设$[\omega]$代表$\{\omega\circ\alpha:\alpha\in\mathrm{Inn}\,\mathfrak{R}\}$的范数闭包（在内自同构下的态的轨迹），并且假设$K$代表正常态空间$\mathfrak{R}_*$的商，那么$d(\mathfrak{R})$是跟约化度量相关的$K$的直径：

$$\bar{d}([\omega_1], [\omega_2])=\inf\{\|\omega'_1-\omega'_2\|:\omega'_i\in[\omega_i]\}$$

显然$d(\mathfrak{R})\in[0, 2]$，而$d(\mathfrak{R})=0$，当且仅当每个态的轨迹在正常态空间中是稠的。如果$\mathfrak{R}$不是因子，那么存在态ω_1，ω_2使得对于所有$\alpha\in\mathrm{Inn}\,\mathfrak{R}$有$\|\omega_1\circ\alpha-\omega_2\|=2$，同样$d(\mathfrak{R})=2$也如此。对于类型$\mathrm{I}_n$因子，正常态之间的距离跟相应的稠算子的迹范数距离是一样的，在此情况下，我们有：

$$d(\mathfrak{R})=2\left(1-\frac{1}{n}\right)=\|\tau-\omega\|$$

其中τ是迹而ω是任何纯态。我们也有对于I_∞的因子和类型II的因子的$d(\mathfrak{R})=2$，见[Takesaki, 2003, 430]。

如果$d(\mathfrak{R})$给出了代数$\mathfrak{R}$是“如何不对易”的某种测度，那么类型III_1因子是最不对易的。

定义 39。一个冯·诺伊曼代数$\mathfrak{R}$被说成是可数可分解的，当且仅当在$\mathfrak{R}$中的相互正交投影算符的任何族是可数的。

命题 40。如果$\mathfrak{R}$是可数可分解的类型III_λ因子，那么：

$$d(\mathfrak{R})=2\frac{1-\lambda^{1/2}}{1+\lambda^{1/2}}$$

证明。参看[Connes and Størmer), 1978]和[Takesaki, 2003, 427]。

函数$f(\lambda)=2(1-\lambda^{1/2})/(1+\lambda^{1/2})$在[0, 1]是单调减少的，特别是，$f(1)=0$对于类型$\mathrm{III}_1$因子，任何正常态$\omega$的轨迹是在态空间中的范数稠密的，按照孔耐[1994, 473]这意味着“人们不能通过闭合和内部自同构下不变量的性质在类型III_1因子的两个态之间进行区分”。特别是，由于两个酉等价态必须被看成是“均匀混合的”，通过类型III_1因子的混合的方式作不出什么区分。

注：要概述模理论，参见[Summers(萨默斯), ND]或者[Connes, 1994]；需要全面的论述，参见[Takesaki, 2003]；如需模理论在QFT中的应用的详细阐述，则参见[Borchers(博彻斯), 2000]。

2. 可观测量代数网的结构

2.1 代数网，基本属性

AQFT靠分离出某些著名的QFT模型的结构性假设来进行，它对这些假设形式化，然后再用所谓“抽象却有效的废话”来导出这些假设的结果。

AQFT的基本形式是一个时空上的“局域可观测量代数网”，虽然这个形式化能够被应用到十分广泛的时空种类，在本章中我们只将注意力局限于闵可夫斯基(Minkowski)时空。

一个开放的闵可夫斯基时空对顶锥，是x点的因果性未来和y点的因果性过去对x的未来的交界。假设$\mathcal{K}$是开放闵可夫斯基时空对顶锥，并且假设$O\mapsto\mathfrak{A}(O)$是从$\mathcal{K}$到C^*代数的映射，我们假定我们所有的C^*代数都是有单位的，即有乘法单位元。我们假定C^*代数的集$\{\mathfrak{A}(O):O\in\mathcal{K}\}$(称为闵可夫斯基时空上的可观测量代数网)是在如下意义上的归纳系统：

如果$O_1\subseteq O_2$，那么存在一个嵌入(即一个等距*同态)$\alpha_{12}:\mathfrak{A}(O_1)\mapsto\mathfrak{A}(O_2)$。

假定41(保序性)。映射$O\mapsto\mathfrak{A}(O)$是个归纳系统。

保序性假设有时是受如下观念的激发，可观测量在区域O_1是可测量的，就更不用说在任何包含O_1的O_2区域是可以测量的。而保序公理也得到其应用的辩护，因为如果$\{\mathfrak{A}(O):O\in\mathcal{K}\}$是个归纳系统，那么存在一个通过所有局域

代数产生的归纳极限C^*代数$\mathfrak{A}$。我们称$\mathfrak{A}$为准局域代数，因为它包含能够被局域可观测量一致逼近的可观测量。

注释42。在某些时空里，对顶锥的集是没有方向的。在很多这样的情况下，有可能通过更复杂的方法定义准局域代数[Fredenhagen，1993]。

现在我们又回到AQFT的主要相对论假设。

假设43(微观因果性)。一个C^*代数的网$\mathfrak{A}$被认为是满足微观因果性的，当且仅当O_1，O_2是类空分离的对顶锥时，有$[\mathfrak{A}(O_1),\ \mathfrak{A}(O_2)]=\{0\}$。

这个假设被看成是相对论施加在时空结构上的约束的反映。

注释44。跟类空分离区域相关的量必须用对易算子表示，这并不是AQFT的教条。事实上，赋值于类空区域的费米场算子也会是反对易的，因此，AQFT没有必要在用$\mathfrak{A}(O)$的元素表征的可观测量和用"场算子"表示的不可观测量之间进行区分。有关这个区分更多的说法，参见7.2和下述DHR超选择理论。

在这一章我们不试图维护或怀疑微观因果性假设，当然，我们会简单讨论它跟第3节中局域性问题的关联。

2.2 真空态/表征的存在/唯一性

变换不变的态的存在性

这一节我们需要涉及真空态及其表示的存在性和唯一性。为此回顾一下，一个仿射空间(比如闵可夫斯基时空)是由符合某种性质的一个集合S、向量空间V以及映射$+:S\times V\to S$组成的三集子，在此情况下，V被称为是变换群。

假设45(变换协变性)。如果$\mathfrak{A}$是仿射空间上算子代数的网，那么我们假设存在$\mathfrak{A}$的自同构群Aut $\mathfrak{A}$中的变换群的忠实连续表征$x\mapsto\alpha_x$，并且对于对顶锥O和变换x有：

$$\alpha_x(\mathfrak{A}(O))=\mathfrak{A}(O+x)$$

注释46。对于闵可夫斯基时空的情况，变换群是庞加莱(Poincaré)群的子群。在很多有物理意义的案例中，$x\to\alpha_x$扩展到在$\mathfrak{A}$的自同构群Aut $\mathfrak{A}$中的完全庞加莱群的表征，但是我们只需要一个结果的事实(命题132)。

变换不变性在传统意义上被看成是真空态的必要条件。

事实 47。如果存在$\mathfrak{A}$上变换群的一个作用α，那么$\mathfrak{A}$上的变换不变的态存在。确实，由于变换群是阿贝尔群，因此它有一个变换平均值μ——即一个变换不变的正的线性泛函，它是作用在群G上的本质上有界可测（对于 Haar 测度）函数的代数$L^{\infty}(G)$上的。给定$\mathfrak{A}$的态ω，我们就能定义一个平均态ρ：

$$\rho(A) := \int \omega(\alpha_x A)\, d\mu(x)$$

态ρ是变换不变量。参见埃姆什（Emch）在本书撰写的内容。

注释 48。正在进行的论证不能用来说明洛伦兹（Lorentz）不变的态存在，洛伦兹群经不起检验，从而也不承认不变平均值。因此，我们不能使用这些一般方法证明洛伦兹不变的态存在。当然，在具体模型中（比如自由波色场和费米场）存在其他方法建立起这些态的存在性。

令G是一个在$\mathfrak{A}$上的由自同构作用的群，GNS 定理的推广证明了，$\mathfrak{A}$的一个G不变态ω，产生带有G的酉表征U的 GNS 希尔伯特空间$\mathcal{H}$，并且 GNS 向量Ω在$\mathcal{H}$上的G作用下是不变量。

事实 49。假设α是群G的一个通过$\mathfrak{A}$的自同构的强连续作用。如果ω是$\mathfrak{A}$的一个G不变态，那么由ω引入的$\mathfrak{A}$的 GNS 表示$(\mathcal{H}, \pi)$，是在如下意义上的G不变量，即有在$\mathfrak{B}(\mathcal{H})$的酉群中$G$的强连续表示$U$，使得：

1. $U(g)\pi(A)U(g)^* = \pi(\alpha_g(A))$，对所有的$A \in \mathfrak{A}$；
2. $U(g)\Omega = \Omega$，对所有的$g \in G$。

每个希尔伯特空间只有一个真空

注释 50。一旦我们考虑C^*代数的自同构群 Aut $\mathfrak{A}$，我们就取标准拓扑为在$\mathfrak{A}$上有界线性映射的集合$L(\mathfrak{A})$的强拓扑（看作巴拿赫空间）。那就是说，α_i收敛到α，当且仅当对于每个$A \in \mathfrak{A}$，$\alpha_i(A)$收敛到$\mathfrak{A}$上的范数中的$\alpha(A)$。

定义 51。我们用 GNS 表示定理（定理 17），把有关表征的术语（定义 14）转换成有关态的术语。因此，比如我们说两个态是不相交的，如果它们的 GNS 表示是不相交的。

一个真空态至少应该是变换不变量。甚至，在网$\mathfrak{A}$上的微观因果性假设蕴含着，任何两个可观测量在一个被转换成类空无限的“极限情况下”是对易的。那就是对于任何$A, B \in \mathfrak{A}$以及任何类空矢量x，有：

$$\lim_{t\to\infty}\|[\alpha_{tx}(A),B]\|=0$$

这最终意味着在如下意义上作为大的自同构群 G 作用在 $\mathfrak{A}$ 上。

如果 ω 是 G 不变态，而($\mathcal{H}$，π)是由 ω 引入的 $\mathfrak{A}$ 的 GNS 表示，那么对任意 $A\in\mathfrak{A}$，$\overline{\mathrm{conv}}\{\pi(\alpha_g(A)):g\in G\}$，与 $\pi(\mathfrak{A})'$具有非空交集。

其中我们用 $\overline{\mathrm{conv}}S$ 代表 S 的弱闭凸包。参见[Størmer，1970]的相关证明。需要注意的是，我们也希望同样的情况在非相对论设定下也是真实的，因为我们希望跟不相交空间区域相联系的可观测量对易(我们没有援引任何闵可夫斯基时空中的向量就是两个类空向量之和的事实)。

多亏了有关"C^*动力学系统"的广泛研究，知道很多关于在 G 作为 $\mathfrak{A}$ 的同构大群作用时的 G 不变态。特别是，G 不变态的集合是凸的和闭的(在弱*拓扑中)，因此这个集合有极限点，叫作极值不变的态。(显然，如果 $\mathfrak{A}$ 的纯态是 G 不变的，那么它就是极端不变量。)再者，我们也有下面的有关 G 不变态的不相交性的结果。

命题 52。假设 ω 是个 $\mathfrak{A}$ 的 G 不变态，设 $\mathcal{H}$ 是其 GNS 希尔伯特空间，并且假设 Ω 是 GNS 向量，那么下面说法是等价的：

1. ω 是在下述意义上是聚类：

$$\lim_{t\to\infty}\omega(\alpha_{tx}(A)B)=\omega(A)\omega(B)$$

2. ω 是极端不变量；

3. 如果一个 G 不变态 ρ 对于 ω 是准等价的，那么 $\rho=\omega$；换言之，不存在其他的 G 不变态准等价于 ω；

4. 由 Ω 生成的射线是 $\mathcal{H}$ 的唯一(达到标量积) G 不变子空间。

证明。参见[Størmer，1970]。对于相关细节，也参见[Emch，1972，183，287]和本章中第 3 节。

因此，如果一个(真空)态是聚类，那么不存在其他变换不变态是处于其叶线上(即准等价于那个态的态集合)。类似地，如果态是极值不变量(更不用说如果它是纯态)，那么它是在其叶线上的唯一变换不变态。

注释 53。分离真空的存在是跟自发性对称破缺相联系的，参见 10.7 节。

注释 54。命题 52 在[Emch，1972，248]"哈格(Hagg)定理"的证明中起到核心作用。特别是，极端的 G 不变态的唯一性等同于"真空极化"的非存在性。

2.3 瑞赫—施列德定理(Reeh-Schlieder Theorem)

我们已经假定真空态是变换不变的，但是我们希望真空态遵守反映理论的相对论性质的较强约束。特别是，事实49中定义的酉表征是被四个动量算子**P**无穷小地生成的。四个动量算子的想法可以在“SNAG(Stone-Naimark-Ambrose-Gelfand)定理”中弄明确，该定理将斯通定理推广到产生单参数酉群的自伴算子的存在性。我们要求能量在每一个洛伦兹参照系下都是正的，也即是**P**的谱处于未来的光锥中。

我们现在推广这个要求，通过从未来光锥中抽掉细节。未来光锥G_+具有下面性质：$G_+\cap(-G^+)=\{0\}$，其中$-G_+=\{-g:g\in G_+\}$，因此，谱条件只要求变换群的酉表征具有在进行加法逆运算下是非对称的集合里的谱。

假定55(谱条件)。设G是变换群，并且假设ω是$\mathfrak{A}$的G不变的态，我们说$(\mathfrak{A},\omega)$对满足谱条件，当且仅当存在一个G的子集G_+使得$G_+\cap(-G_+)=\{0\}$，并且在由ω引入的$\mathfrak{A}$的GNS表征$(\mathcal{H},\pi)$中，G的归纳酉表征的谱$\mathrm{sp}(U)$被包含在G_+中。

瑞赫—施列德定理表明，谱条件意味着真空向量Ω对于每个局域代数是循环的。对于这个定理，我们假定$\mathfrak{A}$上的变换不变真空态ω已经选定，并且$(\mathcal{H},\pi)$是由ω引入的$\mathfrak{A}$的GNS表征。然后我们通过

$$O\mapsto\mathfrak{R}(O)\equiv\pi(\mathfrak{A}(O))''$$

定义$\mathcal{H}$上的冯·诺伊曼代数的对应网$\mathfrak{R}$。

如果网$\mathfrak{A}$满足微观因果性，那么$\mathfrak{A}$也会。因为Ω对$\pi(\mathfrak{A})$是循环的，集合$\{\mathfrak{R}(O)\Omega:O\in\mathcal{K}\}$在$\mathcal{H}$中是稠的。

为了证明这个定理，我们需要一个附加假设。

假设56。网$O\mapsto\mathfrak{R}(O)$被认为是满足可加性，当且仅当对任何对顶锥O，集合$\{\mathfrak{R}(O+x):x\in G\}$产生作为$C^*$代数的$\mathfrak{R}$(这里再次看到$G$代表变换群)。

可加性假设有时在下面基础上得到辩护，理论上不应该存在最小长度标度，即任何可观测量都是由来自任意小区域的可观测量通过取积及求和等产生的。

定理57(瑞赫—施列德)。假定网$O\mapsto\mathfrak{R}(O)$满足谱条件和可加性。那么对于所有的对顶锥O，Ω是对$\mathfrak{R}(O)$循环的。如果网$\mathfrak{R}$也满足微观因果性，那么Ω

是对于任何局域代数分离的。

瑞赫—施列德定理曾经是相对论 QFT 基础中重要研究问题之一。在[Redhead(雷德黑德)，1995a；1995b]中，雷德黑德证明 RS 定理意味着真空态显示了非局域相关性。参见[Halvorson and Clifton(霍尔沃森和克里夫顿)，2000]。雷德黑德还指出由于真空是对每个局域代数分离的，任何局域事件具有发生在真空态的非零概率，特别是，不存在局域数字算子(由于它们具有作为本征矢的真空态)。最后，[Fleming(弗莱明)，2000]论证了 RS 定理意味着不良种类的非局域性，比非相对论量子力学中的非局域性差，因此也说明需要修正 AQFT 的正统形式。有一个可能的回应，参见[Halvorson，2001]。

由于谱条件的使用，看来可能是由于 RS 定理是个“纯粹相对论的结果”，跟非相对论 QM 或者 QFT 的情况不完全类似，参见[Saunders(桑德斯)，1992]。而且，我们可能希望，源于 RS 定理的相对论 QFT 的很多其他结果，有可能对非相对论理论失效。的确，非相对论 QFT 并不认同局域性数字算子。不过，一个谱条件方案，以及随后的一个 RS 定理的方案，已经被证明为非相对论理论所保留[Requardt(雷夸特)，1986]。

注：最初的瑞赫—施列德定理是按 QFT 的公理化方法来形式化的，并且可以在[Reeh and Schlieder，1961]中找到，定理的最新表述可以在[Horuzhy(何露西)，1990；D'Antoni(德安东尼)，1990；Baumgärtel and Wollenberg(保姆加特尔和沃伦伯格)，1992]和[Araki(荒木)，1999]中找到。

2.4 汇集性质(Funnel Property)

定义 58。假设$\mathfrak{R}_1$，$\mathfrak{R}_2$是$\mathcal{H}$上的冯·诺伊曼代数，使得$\mathfrak{R}_1 \in \mathfrak{R}_2$。如果存在循环并且对$\mathfrak{R}_1$，$\mathfrak{R}_2$和$\mathfrak{R}_1' \cap \mathfrak{R}_2$是分离的向量 $\Omega \in \mathcal{H}$，那么($\mathfrak{R}_1$，$\mathfrak{R}_2$)对被认为是冯·诺伊曼代数的一个标准包含。

注释 59。假设 $O \mapsto \mathfrak{R}(O)$是闵可夫斯基时空上的冯·诺伊曼代数网，假定 Ω 具有瑞赫—施列德性质，即对于每一个对顶锥 O，Ω 是对于$\mathfrak{R}(O)$循环的和分离的。那么，如果 O_1，O_2 是对顶锥，使得 O_1 的闭包 $\overline{O}_1$ 包含在 O_2 之中，则($\mathfrak{R}(O_1)$，$\mathfrak{R}(O_2)$)对是冯·诺伊曼代数的一个标准包含。

定义 60。设$\mathfrak{R}_1$，$\mathfrak{R}_2$是$\mathcal{H}$上的冯·诺伊曼代数，使得$\mathfrak{R}_1\subseteq\mathfrak{R}_2$。($\mathfrak{R}_1$，$\mathfrak{R}_2$)对被说成是分裂包含，如果存在类型 I 因子$\mathfrak{N}$使得$\mathfrak{R}_1\subseteq\mathfrak{N}\subseteq\mathfrak{R}_2$。

假设 61(汇集性质)。冯·诺伊曼代数的网 $O\mapsto\mathfrak{R}(O)$被说成是满足汇集性质，如果对任何具有包含在 O_2 之中 $\overline{O}_1$ 的对顶锥 O_1，O_2，($\mathfrak{R}(O_1)$，$\mathfrak{R}(O_2)$)对是一个分裂包含。

注释 62。一个类型 I 因子$\mathfrak{N}$是可数可分解的，当且仅当$\mathfrak{N}$是同构于$\mathcal{H}$可分的$\mathfrak{B}(\mathcal{H})$，此同构当且仅当$\mathfrak{N}$在弱拓扑中是可分的，参见[Kadison and Ringrose，1997，Exercise 5.7.7]。

在我们对超选择理论的讨论中(7.2—11 节)，在一个关键之处(命题 243)，我们不得不求助于真空希尔伯特空间是可分离的这个假定。这将是本章中唯一需要假设希尔伯特空间是可分离的地方。特别是可分离假设需要在两个超选择区的概念之间建立起对应关系，其中之一是物理上的动机，而另一个是数学上的用处。下面的结果是我们连接分离假设到具有(或许)更清楚物理意义的东西上的唯一企图。一般地，我们非常担心物理上对分离假设的保证，跟第 6 节和[Halvorson，2004]比较。

命题 63。设$\mathfrak{R}$是$\mathcal{H}$上的冯·诺伊曼代数网，并且假设 $\Omega\in\mathcal{H}$对所有局域代数是循环的和分离的。如果网满足汇集性质，那么$\mathcal{H}$是分离的。

证明。跟[Doplicher and Longo(多普利克尔和隆哥)，1984]中的命题 1.6 比较。设 O_1，O_2 是具有 $\overline{O}_1\subseteq O_2$ 的对顶锥，假设$\mathfrak{R}$类型 I 因子使得$\mathfrak{R}(O_1)\subseteq\mathfrak{N}\subseteq\mathfrak{R}(O_2)$，并且假设 ω 是 Ω 约化的$\mathfrak{R}$的态。回顾 $\mathfrak{N}$ 是同构于某些希尔伯特空间$\mathcal{K}$的$\mathfrak{B}(\mathcal{K})$，由于 $\mathfrak{N}\subseteq\mathfrak{R}(O_2)$和 Ω 对于$\mathfrak{R}(O_2)$分离，ω 是忠实的和正规的。因此$\mathcal{K}$是分离的，并且存在一个在 $\mathfrak{N}$ 中弱稠的可数集合$\mathfrak{R}_0$。由于$\mathfrak{R}(O_1)\subseteq\mathfrak{N}$，并且 Ω 对$\mathfrak{R}(O_1)$是循环的，从而得出$[\mathfrak{N}_0\Omega]=\mathfrak{N}\Omega]=\mathcal{H}$，因此$\mathcal{H}$是分离的。

如果人们想确证真空希尔伯特空间是分离的这个假设，命题 63 说明它足以确证汇集性质。存在具体模型，其中汇集性质确实是不能保持[Horuzhy，1990，23]。但是这些模型的物理意义是不清楚的，并且存在两个其他考虑可能有利于汇集性质：(i)在 3.3 节中，我们表明汇集性质具有非局域性的问题；(ii)巴克霍尔兹和威赫曼(Buchholz and Wichmann)[1986]论证汇集性质是 QFT 粒子解释的充分条件。当然，QFT 的解释想要批判性地检验巴克霍尔兹和威赫

曼的“粒子解释”概念。(跟4.5节相比较，其中粒子解释得到进一步讨论，跟6.2节比较，此处把非分离性希尔伯特空间和QFT的场解释关联起来了。)

注释64。对于自由场的汇集性质在[Buchholz，1974]有说明。

2.5 局域代数的类型

我们现在总结一下目前已知的局域代数类型的信息，限于局域可观测量代数网的物理上的相关表征。

定义65。设$\mathfrak{R}_1$和$\mathfrak{R}_2$是希尔伯特空间$\mathcal{H}$上的冯·诺伊曼代数网，我们说$\mathfrak{R}_1$和$\mathfrak{R}_2$是局域准等价的，当且仅当对于任一对顶锥O存在一个同构φ_o: $\mathfrak{R}_1(O)\to\mathfrak{R}_2(O)$。

注释66。虽然它不是一个AQFT的“公理”，还是有好的理由相信，物理(特别是基本粒子物理)关注的表征，局域地准等价于某些真空表征，其中一个真空表征是某些特征(比如可能是变换下不变的)态的GNS表征。例如局域准等价在按照多普利克尔—哈格—罗伯茨(Doplicher-Haag-Roberts)的选择标准(参见7.2节)和按照[Buchholz and Fredenhagen(巴克霍尔兹和弗雷登阿根)，1982]的选择标准的两个物理表象之间成立。因此，我们关于在真空表征中的局域代数的结构得到的任何结论，按理说也都能够对其他这些表征成立。

局域代数是真正无限的

有些相对简单的结果限制了局域代数类型的可能选择。对此，我们定义了重要的“性质B”，因为它是可行假设的结果(通过加法和谱条件)，也因为它使下面情形有意义，即没有变换群的情形(跟谱条件不一样)，同时还因为它是我们所需要得到各种结果的全部，特别是那些局域代数是真正无限的结果。

定义67。设$O\to\mathfrak{R}(O)$是在某些希尔伯特空间$\mathcal{H}$上的冯·诺伊曼代数网，我们认为网$\mathfrak{R}$满足性质B，当且仅当对于任何两个对顶锥O_1和O_2使得$\overline{O}_1\subseteq O_2$，如果$E\in\mathfrak{R}(O_1)$是非零投影，那么$E$在$\mathfrak{R}(O_2)$中等价于单位投影$I$，即存在等距$V\in\mathfrak{R}(O_2)$使得$VV^*=E$。

注释68。如果对于每一个O，代数$\mathfrak{R}(O)$是类型Ⅲ，那么网$\mathfrak{R}$满足性质B。

我们希望性质B对可观测代数网成立，因为它可以从出于物理上的弱可加性和谱条件假设得到。

命题69。设 $O \mapsto \mathfrak{R}(O)$ 是满足微观因果性、谱条件和弱可加性的冯·诺伊曼代数网，那么网 $O \mapsto$ 满足性质 B。

证明。对于原始证明参见[Borchers, 1967]，对于新近阐述参见[D'Antoni, 1990]。

假设70(非平凡性)。C^* 代数的网 $O \mapsto \mathfrak{A}(O)$ 被认为是满足非平凡性的，当且仅当每一个对顶锥 O 有 $\mathfrak{A}(O) \neq \mathbb{C}I$。

命题71。设 $O \mapsto \mathfrak{R}(O)$ 是满足微观因果性、性质 B 和非平凡性的冯·诺伊曼代数网，那么对任何对顶锥 O，冯·诺伊曼代数 $\mathfrak{R}(O)$ 和 $\mathfrak{R}(O')'$ 都是真正无限的。

证明。我首先证明 $\mathfrak{R}(O)$ 是真正无限的，即 $\mathfrak{R}(O)$ 中的每个中心投影都是无限的。令 C 是 $\mathfrak{R}(O)$ 中的中心投影，选择一个非平凡的其闭包包含在 O 内的对顶锥 O_1，那么通过性质 B，对于每一个非零投影 $E \in \mathfrak{R}(O_1)$，E 等价于以 $\mathfrak{R}(O)$ 为模数的 I。由于 $\mathfrak{R}(O_1) \neq \mathbb{C}I$，存在投影 $E \in \mathfrak{R}(O_1)$，使得 $E \sim (I-E) \sim I$ 以 $\mathfrak{R}(O)$ 为模数。然后可以得到 $EC \sim (I-E)C \sim C$ 以 $\mathfrak{R}(O)$ 为模数，显然 $EC = CEC \leq C$。如果 $EC = C$ 那么 $(I-E)C = 0$，这发生了矛盾。因此 $EC < C$ 和 $EC \sim C$ 以 $\mathfrak{R}(O)$ 为模数。那就是说，C 是一个 $\mathfrak{R}(O)$ 中的无限投影，而且 $\mathfrak{R}(O)$ 是真正无限的。由于微观因果性，$\mathfrak{R}(O_1) \subseteq \mathfrak{R}(O')'$，因此先前的论证也表明 $\mathfrak{R}(O')'$ 是真正无限的。

特别是，前述命题排除了类型 I_n 和类型 II_1 冯·诺伊曼代数。这个结论已经对非局域性问题具有重要意义；参见3.3节的命题102。然而，前述命题对局域代数可能是类型 I_∞ 因子留有余地，它也对局域代数可能是冯·诺伊曼代数的混合类型的直和留有余地。

局域代数是超有限的

我们将简要看看我们已经指出局域代数是类型Ⅲ的事实这样的最佳结果，该类型原本被认为是跟物理学没有相关性的无规则反常。然而，我们首先表明在某些物理上可能的条件下，局域代数被有限维的代数(即它们是“超有限的”)近似，这证明它们毕竟不是如此毫无规则。

定义72。设 $\mathfrak{R}$ 是冯·诺伊曼代数，那么 $\mathfrak{R}$ 被认为是超有限的，当且仅当存在一个在 $\mathfrak{R}$ 中的有限维冯·诺伊曼代数的族 $(\mathfrak{R}_a)_{a\in\Lambda}$，使得 $\mathfrak{R} = (\cup_{a\in\Lambda}\mathfrak{R}_a)''$。

超有限最后成为对数学目标而言极其有用的条件。确实，超有限是与正则条件期望的存在紧密相关的，参见[Kadison and Ringrose，1997，Chapter 8]，并且存在一个唯一的Ⅱ$_1$型超有限因子，和一个唯一的Ⅲ$_1$型超有限因子。从物理/基础观点看，人们也可能认为$\mathfrak{R}$的超有限的失败可能使得在代数$\mathfrak{R}$的元素跟只涉及有限个任务数的真实实验过程之间很难找到对应。

事实 73。每一种Ⅰ型冯·诺伊曼代数是超有限的，参见[Kadison and Ringrose，1997，Exercise 8.7.26]。

假设 74(内部/外部连续性)。冯·诺伊曼代数网 $O\mapsto\mathfrak{R}(O)$ 被说成是内部连续的，如果对于任何具有最后上界 O 的单调增加的网 $(O_a)_{a\in A}$，我们有：

$$\bigvee_{a\in A}\mathfrak{R}(O_a)=\mathfrak{R}(O)$$

其中$\mathfrak{R}_1\vee\mathfrak{R}_2$代表由$\mathfrak{R}_1$和$\mathfrak{R}_2$产生的冯·诺伊曼代数。外部连续性是通过采取一个增加的区域网以及相应冯·诺伊曼代数的交集定义的。

注释 75。从内部看网$\mathfrak{R}$是连续的这个条件得到满足，只要$\mathfrak{R}$是按照从下面的怀特曼场(Wightman fields)标准方法建构的“最小的”网。参见[Buchholz *et al.*，1987]，类似地，最大网满足外部的连续性。

命题 76。假定网 $O\mapsto\mathfrak{R}(O)$ 满足汇集性质及其内部或者外部连续性，那么对于每一个对顶锥 O，$\mathfrak{R}(O)$是超有限的。

证明概略。比较一下[Buchholz *et al.*，1987，134]。我们刚才看到了网是内部连续的情况，通过汇集性质存在加入$\mathfrak{R}(O_i)$和$\mathfrak{R}(O)$之间的Ⅰ型因子$\mathfrak{N}_i$，那么得出超有限因子的上升序列$\mathfrak{N}_i$的并集在$\mathfrak{R}(O)$内是稠的，因此$\mathfrak{R}(O)$是超有限的。

局域代数是类型Ⅲ$_1$因子

其积累超过了三十年的一系列结果揭示了相对论 QFT 的局域代数是Ⅲ型冯·诺伊曼代数，更具体点说是超有限Ⅲ$_1$型因子。我们在这一节总结了这些结果中的一些。第一个结果归功于隆哥[1979]，德里斯勒(Driessler)改进了一些早期结果。

命题 77。设$\mathfrak{R}$是一个作用在$\mathcal{H}$上的冯·诺伊曼代数，对$\mathfrak{R}$的分离单位向量 $\Omega\in\mathcal{H}$，G 是具有对偶 Γ 的局域紧阿贝尔群，而 U 是作用在$\mathcal{H}$上的G 的连续酉表征，使得 $U\Omega=\Omega$ 和射线$\mathbb{C}\Omega$ 是$\mathcal{H}$的唯一 $U(G)$ 不变子空间。假设存在子空间

$G_+ \subseteq G$ 和 $\Gamma_+ \subseteq \Gamma$ 使得：

1. 对于所有 $g \in G_+$，$G_+ \cup (-G_+) = G$ 和 $U(g)\mathfrak{R}U(g)^* \subseteq \mathfrak{R}$；

2. $\Gamma_+ \cap (-\Gamma_+) = \{0\}$ 和 $\mathrm{sp}(U) \subseteq \Gamma_+$。

那么或者 $\mathfrak{R} = \mathbb{C}I$，或者 $\mathfrak{R}$ 是 III_1 型因子。

证明概略。详情参见[Longo，1979，203]。设 ω 是由 $\omega(A) = \langle \Omega, A\Omega \rangle$ 给出的 $\mathfrak{R}$ 的态，这个结果的证明是要说明 $\mathfrak{R}_\omega = \mathbb{C}I$，其中 $\mathfrak{R}_\omega$ 是态 ω 的中心化。特别是，假设 E 是 $\mathfrak{R}_\omega$ 中的投影，并且通过下式定义函数 $f: G \to \mathbb{C}$：

$$f(g) = \langle \Omega, EU(g)E\Omega \rangle = \langle \Omega, EU(g)EU(-g)\Omega \rangle$$

使用在 $\mathrm{sp}(U)$ 上的约束，能够证明 f 是常数，从而对于所有的 $g \in G$ 有 $U(g)E\Omega = E\Omega$。因为 $\mathbb{C}\Omega$ 是唯一的 $U(G)$ 下不变的子空间，因此 $E\Omega = \Omega$，并且由于 Ω 对于 $\mathfrak{R}$ 是分离的，在 $E = 0$ 或者 $E = I$。

前述命题适用于 $\pi(\mathfrak{A}(W))''$ 形式的代数，其中 W 是嵌入的区域，并且 π 是准局域代数 $\mathfrak{A}$ 的真空表征。确实，我们能够取 G_+ 成为来自在 W 的顶点的类光变换的单参数半群，在此情形 $\mathbb{R} = G_+ \cup (-G_+)$。假设 ω 是 $\mathfrak{A}$ 上的变换不变态，使得 $(\mathfrak{A}, \omega)$ 满足谱条件（假设 55）。那么我们拥有下述说法，在 $\mathbb{R}^4$ 中的 G 的对偶群 Γ 也是类光线，从而谱条件意味着存在 Γ 的子集 Γ_+，即指向未来点的那些向量，使得 $\Gamma_+ \cap (-\Gamma_+) = \{0\}$。最后，正如我们在 2.2 节看到的，当 ω 是极端不变的，那么射线 $\mathbb{C}\Omega$ 是 $\mathcal{H}$ 的唯一的 $U(G)$ 不变子空间。

对于跟局域代数相关的结果，我们必须在网 $\mathfrak{R}$ 上增加一个进一步的条件。第一个结果[Buchholz *et al.*，1987]要求参考具有用检验函数涂抹过的无界算子的公理化 QFT，参见[Streater and Wightman（斯特赖特尔和怀特曼），1964]，即我们必须假设网 $\mathfrak{R}$ 源自基本的怀特曼场论，这种场论满足某种条件——渐进标量不变性。

回顾在公理化方法中，场本质上是形式 $\Phi(f)$ 的自伴算符，其中 f 是时空上的检验函数，这些检验函数的存在允许定义渐进标度不变性的概念。

定义 78。 设 $N: \mathbb{R}^+ \to \mathbb{R}^+$ 是个单调函数，那么检验函数的标度变换由 $f \mapsto f_\lambda$ 给出，其中 $f_\lambda(x) = N(\lambda)f(\lambda^{-1}x)$。假设 Φ_α 是一个产生网 $O \mapsto \mathfrak{R}(O)$ 的怀特曼场的集合。我们说这些场满足渐进标度不变性，当且仅当存在某些具有零真空期望值的某些场 Φ：

$$\langle \Omega,\ \Phi(f)\Omega\rangle = 0$$

并且对于一个合适的 $N(\lambda)$ 选择，标量场算符 $\Phi(f_\lambda)$ 具有下面性质：

1. 期望值 $\langle\Omega,\ \Phi(f_\lambda)^*\Phi(f_\lambda)\Omega\rangle$ 在 $\lambda\to 0$ 极限时对所有检验函数收敛，而对某些 f 是非零的；

2. 范数 $\|\Phi(f_\lambda)^*\Phi(f_\lambda)\Omega\|$ 和 $\|\Phi(f_\lambda)\Phi(f_\lambda)^*\Omega\|$ 在此极限有界。

当一个冯·诺伊曼代数网从具有渐进标度不变性的怀特曼理论中产生，则由此可得局域代数是超有限 III_1 型因子。

命题 79。[Buchholz *et al.*, 1987]设 $\mathfrak{R}$ 是一个满足微观因果性、谱条件和汇集性质的冯·诺伊曼代数，也假设 $\mathfrak{R}$ 能够由一个满足渐进标度不变性的基本怀特曼理论建构出来。那么对每一个对顶锥 O，$\mathfrak{R}(O)=\mathfrak{M}\overline{\otimes}\mathfrak{Z}$，其中 $\mathfrak{M}$ 是唯一的 III_1 型超有限因子而 $\mathfrak{Z}$ 是 $\mathfrak{R}(O)$ 的中心。

注释 80。在[Buchholz *et al.*, 1987]中，汇集性质由更基础的所谓“核性”假设中推导出来，该假设把有界加到了局域自由度上。

当然，人们希望一个对于 AQFT 更本质的结果，这样一个结果已在[Buchholz and Verch(巴克霍尔兹和沃奇), 1995]中得到，其中使用了缩放代数的方法，允许局域可观测量的网 $\mathfrak{A}$ 的短程(收缩)极限的计算。对于收缩代数的简单阐述，我们向读者推荐[Buchholz, 1998]。总之，除了关于这个网的基本假设，需要导出 III_1 型性质的只有附加假设是该网应具有非平凡收缩极限。

注释 81。在某些具体模型中，能够直接证明局域代数是唯一的 III_1 型超有限因子。比如，质量 $m=0$ 的自由玻色子场(在闵可夫斯基真空表征中)，局域代数是同构于嵌入区域的代数。因此命题 77 证明局域代数是 III_1 型超因子。甚至质量 $m>0$ 的自由玻色子场是局域准等价于 $m=0$ 的情况，从而其局域代数也是 III_1 型超有限因子，参见[Horuzhy, 1990, 254]。

III_1 型性质的导出是当代数学物理的最惊人和有趣的结果之一。但是这个结果的基础意义是什么？比方说如果局域代数是类型 $\mathrm{III}_{1/2}$，这个世界将会有什么不同呢？甚或本质上更加不同的，比如是类型 II_∞ 呢？举例来说，在熟悉的类型 I 代数上的态的结构跟类型 III 代数上的态的结构存在重要差异：由于类型 III 代数没有原子投影，并且纯正常态的支持投影是原子的，因此类型 III 代数没有纯正常态(但是对于类型 II 代数当然有相同的结果)。正如在[Clifton and Halvorson，2001b］和［Ruetsche（鲁切），2004］指出的，纯态的缺失是量子概率的解释的进一步的障碍（也参见3.4节。）

因外逊（Yngvason）［2005］给出几个有趣断言，是关于类型Ⅲ代数的概念重要性的，尤其是跟非局域性问题有关系的。第一，按照因外逊的观点，“类型Ⅰ直观”能够导致悖论，正如费米的二原子系统遇到的。不过，因外逊断言，如果我们适当运用类型Ⅲ代数模型化这些情形，这些悖论就会消失。第二，因外逊断言，类型III_1因子$\mathfrak{R}$的状态空间的均匀性能够被解释成，对于$\mathfrak{R}$上的任何两个态ω_1，ω_2，ω_2能够从ω_1（以任意好的精确度）通过酉运算得到制备。当然，这个运算是非选择性的，因而不会改变$\mathfrak{R}'$中可观测量的测量的统计性。因此，在某种意义上，具有类型Ⅲ代数的观测者比起类型Ⅰ代数的观测者更容易控制他的状态空间。

3. 非局域性及其AQFT中的开放系统

注释82。在这一节，我们遵守下面的符号约定：大写罗马字母表示代数，小写罗马字母表示算子，而1表示代数中的乘法单位元。

AQFT的基本假定为，可观测量$A(O_1)$和$A(O_2)$在O_1跟O_2是类空分离时是相互对易的。这个要求——我们称为“微观因果性”——有时也叫“爱因斯坦因果性”，因为在代数$A(O_1)$，$A(O_2)$的对易性跟相对论关于“超光速信号”禁令之间暗含关联。在此关联中蕴含着的一个说法是，如果对于$a\in A(O_1)$和$b\in A(O_2)$有$[a, b]\neq 0$，那么a的测量可能改变b的测量的统计性。

尽管存在非相对论QM根本不提时空这个事实，它还是涉足了超光速信号的相对论禁令。特别是，两个不同对象的状态空间是一个张量积$H_1\otimes H_2$，而它们的观测量联合代数是$B(H_1)\otimes B(H_2)$。在此张量积构建中我们把系统A的可观测量表示为一个简单张量$a\otimes 1$，而系统B的为$\mathbb{1}\otimes b$，因此，我们拥有一个微观因果性的方案。但是我们也有强的独立性质，比如，对于系统A的每一个状态φ_1，系统B的态φ_2，都存在$A\otimes B$的态φ，使得$\varphi|_A=\varphi_1$和$\varphi|_B=\varphi_2$。

在这一节，我们考察了两个局域代数$A(O_1)$，$A(O_2)$能够被想成表示实在的不同独立部分的范围。在3.1节和3.2节中，我们讨论微观因果性跟$A(O_1)$，

$A(O_2)$代数的其他独立假设之间的关系。在 3.3 节，我们总结了有关在 AQFT 中违背贝尔(Bell)不等式的一些结果。最后，在 3.4 节中，我们考察局域代数 $A(O)$是否能够从其环境的影响中孤立出来。

3.1　C^*和冯·诺伊曼代数的独立性

我们先来看看一对冯·诺伊曼代数和一对 C^* 代数之间的通常独立性概念。

定义 83。如果 e，f是希尔伯特空间$\mathcal{H}$上的投影算符，那么我们假设 $e\wedge f$ 代表到封闭子空间 $e(\mathcal{H})\cap f(\mathcal{H})$的投影。

事实 84。设 R 是一个作用在$\mathcal{H}$上的冯·诺伊曼代数，如果 e，$f\in R$ 那么 $e\wedge f\in R$。

定义 85(施列德性质)。假设 R_1，R_2 是作用在希尔伯特空间$\mathcal{H}$上的冯·诺伊曼代数，我们说(R_1, R_2)对满足施列德性质，当且仅当 $e\in R_1$ 和 $f\in R_2$ 是非零投影，那么 $e\wedge f\neq 0$。施列德性质意味着对于 $e\in R_1$，$f\in R_2$，如果 e，$f\neq 0$ 和 e，$f\neq \mathbb{1}$，那么：

$$e\wedge f\neq 0,\quad \neg e\wedge\neg f\neq 0,\quad e\wedge\neg f\neq 0,\quad \neg e\wedge f\neq 0$$

其中$\neg x=\mathbb{1}-x$ 是到 $x(\mathcal{H})$的正交元上的投影。因此如果“∧”是经典逻辑中的合取的类推，那么施列德性质是逻辑独立性的类推。

定义 86。如果 A，B 是某些 C^* 代数 C 的 C^* 子代数，我们令 $A\vee B$ 代表由 $A\cup B$ 产生的 C^* 代数。

定义 87(C^* 独立性)。设 A，B 是 C^* 代数，我们认为(A, B)对是 C^* 独立的，当且仅当对于 A 的任何态 ω_1 和 B 的任何态 ω_2，存在 $A\vee B$ 的态 ω，使得 $\omega|_A=\omega_1$ 和 $\omega|_B=\omega_2$。换言之，A 的每一个态跟 B 的每一个态兼容。

C^* 独立性假设具有明显的操作主义者的动机：如果爱丽丝(Alice)是在 O_1 处的观测者，而鲍勃(Bob)是在 O_2 处的观测者，那么 C^* 独立性等于是在断言爱丽丝对制备一个态的选择无法妨碍鲍勃制备一个态的能力。确实，[Summers and Buchholz, 2005]断言一个 C^* 独立性的失败可能被局域观测者所探测，另一方面，C^* 独立性也可能被看作是对象独立性概念的一种解释：

两个对象A，B 是真正独立的，当且仅当 A 的任何状态与 B 的任

何状态兼容，即关于A和B的状态预测之间不存在逻辑关系。

不幸的是，C^*独立性并不意味着微观因果性。

例子88。我们证明了C^*独立性并不意味着微观因果性。与[Napiórkowski（拿皮沃科斯基），1972]相比，考虑有限维的$*$代数$C(\mathbb{Z}_4)\oplus M_2$，其中$C(\mathbb{Z}_4)$是四维阿贝尔$*$代数，而$M_2$是$\mathbb{C}$上的$2\times 2$矩阵，$C(\mathbb{Z}_4)$的投影格子是拥有两个原子的布尔代数；因此在逻辑上包含独立元e_1，e_2。现在选择两个投影f_1，$f_2\in M_2$，使得$[f_1, f_2]\neq 0$，并且设R_i是投影$e_i\oplus f_i$产生的$C(\mathbb{Z}_4)\oplus M_2$的阿贝尔$*$子代数。

为了理解(R_1, R_2)是C^*独立的，设ω_i是在R_i上的态，并且设$\lambda_i=\omega_i(e_i\oplus f_i)$。由于$e_1$，$e_2$的逻辑独立性，存在$C(\mathbb{Z}_4)$的态$\rho$，使得$\rho(e_i)=\lambda_i$。那么在$C(\mathbb{Z}_4)\oplus M_2$上的$\rho\oplus 0$是$\omega_i$的常见扩展，因为：

$$(\rho\oplus 0)(e_i+f_i)=\rho(e_i)=\lambda_i$$

而且一个在$e_i\oplus f_i$上的状态值决定了其在R_i上的值。因此，(R_1, R_2)是C^*独立的。另一方面，$[e_1+f_1, e_2+f_2]=[f_1, f_2]\neq 0$，因此，$(R_1, R_2)$并不满足微观因果性。

在前面的例子中，代数R_1和R_2拥有一个共同的超选择区域，每一个都跟投影$p=\mathbb{1}\oplus 0$对易。然而，约化代数pR_ip不是C^*独立的。事实上，这个例子的分析判断可以推广成下面的结果。

命题89。设R_1和R_2是作用在希尔伯特空间$\mathcal{H}$上的冯·诺伊曼代数，如果对于任何一个投影$e\in Z(R_1\vee R_2)$，(eR_1e, eR_2e)对是C^*独立的，那么$[R_1, R_2]=\{0\}$。

证明。参看[Summers and Buchholz，2005]。

定义90（分裂性质（Split Property））。设R_1和R_2是$\mathcal{H}$上的冯·诺伊曼代数使得$R_1\subseteq R'_2$，那么(R_1, R_2)对被说成是满足分裂性质，当且仅当存在类型Ⅰ因子M使得$R_1\subseteq M\subseteq R'_2$。

评论91。（ⅰ）显然上述定义是等价于说(R_1, R'_2)是像定义60一样的“分裂嵌入”。

(ⅱ)如果(R_1，R_2)满足分裂性质，那么在某些十分标准的条件(比如 R_1 或者 R_2 是类型Ⅲ)下，存在一个在$\overline{R_1 \vee R_2}$和冯·诺伊曼代数张量积 $R_1 \overline{\otimes} R_2$ 之间的自然 * 同构 α；说 α 是"自然的"，我们的意思是它扩展了映射 $AB \mapsto A \otimes B$。再者，* 同构 α 是空间性的，即存在一个酉算子 u 使得 $\alpha(x) = uxu^*$。参见[Summers，1999，212]。

(ⅲ)另一方面，假定 R 是一个因子，使得 $R \cup R'$产生作为冯·诺伊曼代数的 $B(\mathcal{H})$，即$\overline{R \vee R'} = \mathcal{B}(\mathcal{H})$。那么 R'是跟 R 一样的相同类型(Ⅰ、Ⅱ或者Ⅲ)[Kadison and Ringrose，1997，Thm. 9. 1. 3]，从而冯·诺伊曼代数张量积 $R \overline{\otimes} R'$是跟 R 同样的类型[Kadison and Ringrose，1997，830]。因此如果 R 是类型Ⅱ或者Ⅲ，那么$\overline{R \vee R'}$确实比 $R_1 \overline{\otimes} R'$大而且是非同构的。

定义 92(W^*独立性)。设 R_1 和 R_2 是作用在$\mathcal{H}$上的冯·诺伊曼代数，(R_1，R_2)对被认为是 W^*独立的，当且仅当，对于 R_1 的每一个正常态 φ_1 和 R_2 的每个正常态 φ_2，存在 $R_1 \vee R_2$ 的一个正常态 φ 使得 $\varphi|_{R_i} = \varphi_i$。

根据 R_1 和 R_2 的相互交换性(即微观因果性)假设，我们具有下面的含义，参见[Summers，1990，222]：

分裂性质

$\Downarrow$

W^*独立性

$\Downarrow$

C^*独立性　$\Longleftrightarrow$　**施列德性质**

3.2　局域代数的独立性

现在我们考虑在与类空分离区域结合的代数对之间成立的那些独立性质，一般来说能够称为这种代数的独立性质的不多，为了有把握获得这样的结果，我们需要类空分离的更强的概念。

定义 93。对顶锥 O_1，O_2 被认为是严格类空分离的，当且仅当零的邻域 N 对于所有 $x \in N$ 都有 $O_1 + x$ 是和 O_2 类空分离的。

命题 94。假定网 $O \mapsto R(O)$满足微观因果性，弱可加性以及谱条件。如果 O_1 和 O_2 是严格类空分离的，那么($R(O_1)$，$R(O_2)$)满足施列德性质。

证明。参看[Schlieder, 1969]。

按照逻辑强度，下述概念居于类空分离和严格类空分离之间，而且，这个概念对没有变换群的时空有意义。

定义 95。 对顶锥 O_1，O_2 被认为是强类空分离的，当且仅当存在对顶锥 $\widetilde{O}_i$ 使得 $\overline{O}_i \subseteq \widetilde{O}_i$，并且 $\widetilde{O}_1$ 和 $\widetilde{O}_2$ 是类空的。

事实 96。 如果 O_1 和 O_2 是严格类空分离的，那么它们是强类空分离的。

当然，命题 94 的假设(微观因果性、可加性、谱条件)肯定是常常用来导出网的性质 B(命题 69)。因此，它或许阐明了从性质 B 到施列德性质的简单导出。(这样的结果也适用于背景——例如在弯曲时空上的 QFT——在那里谱条件没有意义。)

命题 97。 假定冯·诺伊曼代数网 $O \mapsto R(O)$ 满足微观因果性和性质 B，如果 O_1 和 O_2 是强类空分离的，那么$(R(O_1), R(O_2))$满足施列德性质。

证明。设 O_1 和 O_2 是强类空分离的，并且设 $e_i \in \Re(O_i)$ 是投影，那么存在区域 $\widetilde{O}_i$ 使得 $\overline{O}_i \subseteq \widetilde{O}_i$，而 $\widetilde{O}_1$ 对 $\widetilde{O}_2$ 是类空的。根据性质 B，存在等距算子 $v_i \in R(\widetilde{O}_i)$ 使得 $v_i v_i^* = e_i$，进而 $[v_1, v_2] = 0$，并因此 $e_1 e_2 = v_1 v_2 (v_1 v_2)^*$，但是 $v_1 v_2$ 是等距的，因此 $v_1 v_2 (v_1 v_2)^* \neq 0$。

分裂性质显然并不对$(R(W), R(W'))$成立，这里的 W 是嵌入区域，而 W' 是其因果补充。确实，由于 $R(W)$ 和 $R(W')$ 是 III_1 型因子，在 $R(W) \overline{\otimes} R(W')$ 跟 $\overline{R(W) \vee R(W)'} = B(H)$ 之间不可能存在 * 同构。不过，如果汇集性质为网 $O \mapsto R(O)$ 所有，那么$(R(O_1), R(O_2))$在 O_1 和 O_2 是严格类空分离对顶锥的时候满足分裂性质。

3.3 冯·诺伊曼代数之间的贝尔关联

我们首先定义对冯·诺伊曼代数的贝尔类型测量的一般化概念。

定义 98。 设 A 和 B 是某些 C^* 代数 C 的相互对易的 C^* 子代数，然后我们设

$$\mathbb{B}(A, B) \equiv \{(1/2)[a_1(b_1 + b_2) + a_2(b_1 - b_2)] : a_i = a_i^* \in A,\ b_i = b_i^* \in B,\ -\mathbb{1} \leqslant a_i,\ b_i \leqslant 1\}$$

$\mathbb{B}(A, B)$的元素被称为(A, B)的贝尔算子。

设r是(A, B)的贝尔算子，可能证明在C上每一个状态φ有$|\varphi(r)| \leqslant \sqrt{2}$[Summers and Werner(萨默斯和维纳)，1987]。也直接验证了一点，如果φ是可分离的态(即积态的一个混合)，那么$|\varphi(r)| \leqslant 1$。确实，只要$|\varphi(r)| \leqslant 1$在态φ中的贝尔测量关联能够通过局域隐变量模型再现[Summers and Werner，1987；Baez(贝兹)，1987]。

定义 99。$A \vee B$的一个态φ的贝尔关联系数通过

$$\beta(\varphi, A, B) = \sup\{|\varphi(r)| : r \in \mathbb{B}(A, B)\}$$

定义，如果$|\beta(\varphi, A, B)| > 1$，那么$\varphi$被认为是违背了贝尔不等式，否则是贝尔关联的。

一个简单运用就可以证明如果R_1是阿贝尔的冯·诺伊曼代数并且$R_1 \subseteq R'_2$，那么对于任何态φ，$\beta(\varphi, R_1, R_2) \leqslant 1$。就一种逆命题，朗道(Landau)[1987]证明如果R_1和R_2是非阿贝尔代数，使得$R_1 \subseteq R'_2$，并且，如果(R_1, R_2)满足施列德性质，那么存在某些态φ违反跟(R_1, R_2)极大相关的贝尔不等式。类似地，巴奇伽鲁皮(Bacciagaluppi)[1994]证明，如果A和B是C^*代数，那么，当A和B是非阿贝尔的，则某些态违背了$A \otimes B$的贝尔不等式。

当A和B具有更深一层的性质，那么我们能够导出更强的结果。就目前的目标而言，我们仅简单地把一对已知结果运用到AQFT的情况。参看文献[Summers，1990]中的更多细节。

命题 100。设R是一个作用在可分离希尔伯特空间$\mathcal{H}$上的III_1型因子，那么$B(\mathcal{H})$的每一个模态φ是通过(R, R')极大贝尔关联的，即$\beta(\varphi, R, R') = \sqrt{2}$。

证明。参见[Summers and Werner，1988；Summers and Werner，1995]。

注释 101。命题77告诉我们在非常一般的条件下，嵌入代数$R(W)$是III_1型因子。在此情况下，命题100告诉我们真空是通过$R(W)$，$R(W)'$极大贝尔关联的。

命题 102。假定R_1和R_2是在$\mathcal{H}$上的冯·诺伊曼代数使得$R_1 \subseteq R'_2$，并且(R_1, R_2)满足施列德性质。如果R_1和R_2是真正无限的，那么在$\mathcal{H}$中存在一个稠矢量集，导致(R_1, R_2)是贝尔关联态。

证明。参见[Halvorson and Clifton，2000]。

注释 103。如果在$\mathcal{H}$上的冯·诺伊曼代数的一个网$O \mapsto R(O)$满足属性 B 和非平凡性，那么命题 102 的假说适合于代数$R(O_1)$和$R(O_2)$，只要O_1和O_2是强类空分离的。

注释：对 AQFT 中局域代数独立性在 1990 年之前的成果的综合评论参看[Summers，1990]。对于最近成果，参看[Summers，1997；Florig and Summers（弗洛里希和萨默斯），1997；Rédei（瑞德），1998；Halvorson and Clifton，2000；Summers and Buchholz，2005]。

3.4 内在纠缠态

按照克里夫顿和霍尔沃森[2001b]，AQFT 中局域代数的Ⅲ型性质表明局域系统从其环境脱离纠缠是不可能的。为了明白这个论证，回想一下，它是一个标准(或许有些正当的)假定，可观测量的动力学演化T的一般形式，由C^*代数A的自伴元表示，是通过A的完备正(CP)线性映射T给定，使得$T(1)=1$的。(这样一个假定肯定是平凡的，比如说在量子信息理论中。)在此我们回顾一下相关定义。

定义 104。设A是一个C^*代数，如果对于每个$a\in A$都有$T(a^*a)\geqslant 0$，则A的一个线性映射T被认为是正的。如果对每一个$n\in\mathbb{N}$，映射$T\otimes \mathrm{id}_n$：$A\otimes M_n\rightarrow A\otimes M_n$在基本张量定义为

$$(T\otimes \mathrm{id}_n)(a\otimes b)=T(a)\otimes b$$

是正的，且T被认为是完全正的。这里的M_n是在$\mathbb{C}$上的$n\times n$矩阵的C^*代数。

注释 105。如果T：$\mathfrak{A}\rightarrow\mathfrak{A}$是正的并且$T(1)=\mathbb{1}$，那么对于$\mathfrak{A}$的每一个$\omega$，我们通过$T^*(\omega)(a)=\omega(T(a))$定义$T^*(\omega)$。因此$T^*$是状态空间到自身的仿射映射。

对于Ⅰ型因子，克劳斯定理[Kraus，1983]证明 CP 映射是“内部的”。

定理 106(克劳斯表示)。如果R是I_n型因子那么对于线性映射T：$R\rightarrow R$，下述说法是等价的：

1. T是完全正的，并且$T(\mathbb{1})=1$；

2. T是在形式$R\otimes B(H)$的代数上的一个自同构$x\mapsto uxu^*$的约束；

3. 存在正算子 a_1，…，$a_n \in R$ 使得 $\sum_{i=1}^{n} a_i = \mathbb{1}$，并且：

$$T(x) = \sum_{i=1} a_i^{1/2} x a_i^{1/2} \tag{1}$$

方程(1)的一个特殊情况是具有投影算子 e 和 $1-e$ 的卢德思规则(Lüders rule)：

$$T_e(x) = exe + (1-e)x(1-e)$$

再者，如果代数 R 是类型Ⅰ，我们能够选择 $e \in R$ 为一个阿贝尔投影，可以得到如下结果：

如果局域代数 R 是Ⅰ型因子，那么存在一个普遍的解纠缠(disentangle)运算 T_e。那就是说，不管初始状态是什么，运用 T_e 的结果是末态总是可分离的。

然而，假设 R 没有阿贝尔投影(比如 R 是类型Ⅲ)，那么对于每一个非零投影 $e \in R$，代数 eRe 和 $eR'e$ 都是非阿贝尔的，从而存在对于(eRe，$eR'e$)对的某些纠缠态 φ。这个纠缠态是在$\overline{R \vee R'}$上的某些态在运算$(Te)^*$之下的像。因此，运算 Te 并不解纠缠所有态。

这个有启发性的论证可以严格作为一个“证明”，认为没有 R 上的运算能够解纠缠$\overline{R \vee R'}$的态。详情参见[Clifton and Halvorson，2001b]。

注释 107。(i)克劳斯表示定理在其代表类型Ⅲ代数时无效。确实，克劳斯表示定理是斯坦斯普林分解定理的特殊情况[Stinespring，1955]。

(ii)一个在冯·诺伊曼代数上的 CP 运算通常也看成是超弱连续的。T 的连续性可能在下述基础上得到辩护，如果 T^* 把标准态映射到标准态是必要的。对于违反连续性要求，参看[Srinivas，1980]。

4. 粒子图景

相对论 QFT 的主要应用是在基础粒子物理方面，但是仍然不完全清楚基本粒子物理真正是关于粒子什么。实际上，除了 QFT 准许的通过福克(fock)空间实现的粒子解释的原初意义，还存在很多有关相对论 QFT 粒子本体论可能性的隐含意义。这一节重点从 AQFT 角度考察粒子观。

4.1 福克空间的粒子

我们对粒子的考察从有关如何给出 QFT 一个粒子解释的“母亲膝下的童话

故事”开始，参见[Teller(泰勒)，1995]对这个故事的哲学解释。这个故事始于特殊的希尔伯特空间，叫作福克空间。现在福克空间仅仅是另一种可分离的无限维希尔伯特空间(从而同构于它所有的可分离的无穷维“兄弟”)，但是关键在于按照粒子解释的流行说法把它描述出来。特别是，假设 H 是单粒子希尔伯特空间，即单个粒子的状态空间。现在要看我们的粒子是玻色子还是费米子，这些粒子对的状态空间是$E_S(H\otimes H)$还是 $E_a(H\otimes H)$，其中 E_S 是 $H\otimes H$ 上的$\Sigma_{H,H}$交换下映射到向量不变量上的投影，而 E_a 是到在$\Sigma_{H,H}$之下会变号的向量的投影。就目前的目标来说，我们忽略这些差异，只是简单地用 $H\otimes H$ 代表这个或那个可能性。现在沿着这个思路，对 n 个粒子，我们有希尔伯特空间 $H^n=H\otimes\cdots\otimes H$，等等。

在 H^n 中的一个态肯定是 n 个粒子的态，为了得到分离的态，我们使用在希尔伯特空间上的直积运算“$\oplus$”。因此，我们定义 H 上的福克空间$\mathcal{F}(H)$为无穷直积：

$$\mathcal{F}(H)=\mathbb{C}\oplus H\oplus(H\otimes H)\oplus(H\otimes H\otimes H)\oplus\cdots$$

因此，福克空间中的态矢量包括没有粒子的态(向量处于第一个被加数中)、存在一个粒子的态、存在两个粒子的态……乃至存在不同数目粒子的叠加态。

人们还有时间去担心粒子数能够被叠加到底意味着什么。但是那是“半个空杯子”的观点，从“半个满杯子”的观点看，粒子计数是有意义的。确实，正的(无限)算子

$$N=0\oplus1\oplus2\oplus3\oplus4\oplus\cdots$$

是允许我们讨论粒子数的模型的形式要素。

注释 108。在希尔伯特空间的范畴内，所有分离的希尔伯特空间都是同构的——与在福克空间跟单粒子空间之间没有什么差别。如果我们不细心，就可能会被“福克空间”的说法弄混淆。

如果我们移到适当的范畴，这样的混淆就消失。按照维格纳(Wigner)的分析[Wigner，1939]，一个粒子对应于庞加莱群的单位元$\mathcal{P}$的不可约化酉表示。那么单个粒子空间和福克空间是在$\mathcal{P}$的表示范畴内分离的对象。两个表示的基本希尔伯特空间都是分离性的(因此在希尔伯特空间是同构的)，但是两个表征

都肯定不是等价的(一个不可约化，另一个可约化)。

4.2 福克空间形成可观测量代数

福克空间的故事在通向 QFT 的代数方法中并没有被放弃。事实上，只要条件具备，对于可观测量代数的某些真空态，福克空间作为 GNS 希尔伯特空间就会出现。在提出有关朴素的福克空间故事的问题之前，我们简单描述一下它是如何出现的。我们这里只涉及对称的(即玻色子)情况，类似的处理也适合反对称的(即费米子)情况。

福克空间的代数重构源自于正则量子化的代数方案，假定 S 是实矢量空间(具有某些适当的拓扑学)，而且 σ 是 S 上的辛形式。那么，S 代表一个经典的相空间(参看巴特菲尔德在本书中所撰的章节)。外尔(Weyl)代数 $\mathfrak{A}[S,\ \sigma]$ 是由 $W(f)$ 形式的元产生的具体 C^* 代数，有 $f\in S$ 并且满足外尔—西格尔(Weyl-Segal)形式的正则对易关系：

$$W(f)W(g)=e^{-i\sigma(f,g)/2}W(f+g)$$

假定也存在针对 S 元的时空局域性的一些概念，即从闵可夫斯基时空中对顶锥到 S 的子空间的映射 $O\mapsto S(O)$。因此，如果某些约束得到满足，那么映射对

$$O\mapsto S(O)\mapsto \mathfrak{A}(O)\equiv C^*\{W(f):\ f\in S(O)\}$$

能够通过给出闵可夫斯基时空上的 C^* 代数网构成(这里的 C^*X 是由集合 X 产生的 C^* 代数)。

现在如果我们给出在 S 上的某些动力学，然后如果我们能够再一次地满足某些标准，定义一个在 $\mathfrak{A}[S,\ \sigma]$ 上的相应动力学的自同构群 α_t。那么存在一个唯一的动力学上稳定的 $\mathfrak{A}[S,\ \sigma]$ 的纯态 ω_0，并且我们考虑由 ω_0 约化的 $\mathfrak{A}[S,\ \sigma]$ 的 GNS 表示$(\mathcal{H},\ \pi)$。使我们高兴的是，我们发现单参数群 $\{\pi(W(f))\}_{t\in\mathbf{R}}$ 的无穷小产生子 $\Phi(f)$ 的情况正像老式福克空间方式的场算子。进而(现在不太严格地说)，如果我们定义算子为：

$$a(f)=2^{-1/2}(\Phi(f)+i\Phi(Jf))$$

$$a^*(f)=2^{-1/2}(\Phi(f)-i\Phi(Jf))$$

我们发现它们的表现类似粒子产生和湮灭算符(这里的 J 是在能够跟动力学兼容的 S 上唯一的“复杂结构”)。

特别是，通过运用它们到真空态 Ω，我们得到整个 GNS 希尔伯特空间$\mathcal{H}$。最终，如果我们取 S 的正交基$\{f_i\}$，那么其和：

$$\sum_{i=1}^{\infty} a^*(f_i)a(f_i)$$

是数字算符 N。因此，传统福克空间形式化体系就作为外尔代数状态的 GNS 表示的一个特殊情况出现。

注释 109。 $\mathfrak{A}$的闵可夫斯基真空表征($\mathcal{H}_0$，π_0)是庞加莱协变的，即庞加莱群在$\mathfrak{A}$上的自同构作用是靠$\mathcal{H}$上的酉算子 $U(a, \Lambda)$来完成的。当我们说$\mathcal{H}$同构于福克空间$\mathcal{F}(H)$，这并不是说$\mathcal{H}$是和福克空间$\mathcal{F}(H)$同构的这个平凡事实。不如说，我们认为庞加莱群的酉表示($\mathcal{H}$，U)是福克表征。

注：参看[Bratteli and Robinson(布拉特利和罗宾逊)，1997，Section 5.2]详细解释了从外尔代数角度重构福克空间。也请参看[Clifton and Halvorson，2001a]和[Halvorson，2001]的简单描述。

4.3 粒子解释的非唯一性

如果我们具有准局域代数$\mathfrak{A}$的表征($\mathcal{H}$，π)，使得$\mathcal{H}$同构于福克空间，那么我们能够使讨论粒子有意义。而且，这样的表征是存在的，比如，自由玻色场的闵可夫斯基真空态 ω_0 的 GNS 表征。因此，在最简单的情况下(比如在平坦时空上的自由场)，不存在涉及理论的粒子解释存在性的问题。

但是存在一个唯一性问题：存在$\mathfrak{A}$的酉不等价表征，其中每一个又是同构于福克空间的。而且，来自[Caiken(柴肯)，1967；Caiken，1968]的结果表明两个不等价的福克表征，对应于两个不能看作实在的相同描述记号变量的数字算符。的确，不存在把正确值赋予两个数字算符的$\mathfrak{A}$的态。因此，由两个福克表征提供的粒子解释是相互排斥的。

不等价福克表征的问题在[Clifton and Halvorson，2001a]中得到深入处理。就目前的目的来说，我们仅仅注意到，关于非唯一性的担心是跟对准局域 C^* 代数$\mathfrak{A}$的不等价表征的更一般性的担心连在一起的。但是，这个更一般性的问题，不参照超选择区域理论中的最新进展(参见 7.2 节及其之后的内容)，是不可能得到解决的。我们将在第 7 节再回到这个问题。

4.4 局域化粒子的问题

假定我们已经处理好上一小节提出的唯一性问题——比如，我们找到好的理由选择一个$\mathfrak{A}$中的特定福克表征($\mathcal{H}$，π)，从而我们具有一个作用在$\mathcal{H}$上的特定整体性数字算符 N。下一个问题是，是否相对论 QFT 跟局域化粒子的本体论是一致的——那就是说，讨论在空间的有限区域 O 中的粒子数目是否有意义。

正如在 2.3 节指出的，瑞赫—施列德(RS)定理意味着，AQFT 的局域代数并不包括湮灭真空的算符。因此，如果数字算符具有作为本征态的真空，那么就不存在局域数字算符。那或许足以说服大多数读者，局域化粒子在相对论 QFT 中是不可能的。尽管如此，还是有人努力绕过 RS 定理，其中最有名的是牛顿和维格纳的提议，该建议最近被重新应用于[Fleming，2000]。进一步，它能够独立于 AQFT 整个框架，而且在不使用 RS 定理的情况下，正能条件跟微观因果性结合排除局域数算符[Halvorson and Clifton，2002]。

尽管在相对论 QFT 中局域粒子有各种不可行结果，在高能物理中对实验的解释，似乎要求一个会引起探测器反应的某种东西的概念，而且“探测器”被很好地限制在时空的某些边界上。一个探测器对应于一个$\mathfrak{A}$中的正算符 C，并且是“完全可靠的”，只要它完全赋予真空态 0，即 $C\Omega=0$。因此，瑞赫—施列德定理意味着 C 不包括在任何局域代数当中。尽管如此，C 的近似局域化概念得以保留，即选择具有 $0\leqslant A\leqslant I$ 的某些 $A\in\mathfrak{A}(O)$，并且取：

$$C=\int f(x)\alpha_x(A)dx$$

其中 f 是光滑函数，它的傅里叶变换保留在正向光锥中的反面(函数 f 自动具有无限支持)。那么 $C\Omega=0$，以及函数 f 也能够得到选择，使得 C 在范数拓扑中对$\mathfrak{A}(O)$中的一个算子是“闭合的”。

近似局域化探测器的概念被广泛运用在哈格—吕埃勒(Hagg-Ruelle)散射理论及其新的进展中，这正是我们要讨论的。

4.5 一般化的粒子解释：散射理论及其他

为了在希尔伯特空间中具有作为粒子态解释的态，$\mathfrak{A}$的($\mathcal{K}$，π)表征一定

要是福克表征，这是不正确的。实际上，"散射理论"的一个主要任务，就是为定义粒子态提供一个标准——在缺少整体福克空间结构时。为了描述那样的散射实验，即在福克空间中不能描述，但是需要粒子态来描述输入态和输出态的时候，这些标准是必要的。

哈格和斯威卡(Hagg and Swieca)[1965]提议，通过局域化探测器寻找 n 粒子态，我们叫探测器标准：

> 一个至少具有 n 个粒子的态，是一个可能触发在空间中远距离分离的 n 个探测器的态。

哲学家可能担心探测器标准太过操作主义。事实上，有人断言探测器自身是由粒子构成，因此通过探测器来定义粒子可能是恶性循环。

如果我们试图给出一个粒子概念的分析，那么我们就需要解决这种担忧。当然，散射理论并不终止于探测器标准。实际上，其目的在于把探测器标准维系在某些粒子态更内在的定义上，粒子态的传统固有定义是维格纳的对称标准：

> 一个 n 粒子态(自旋 s_j 和质量 m_i)，是一个对应于庞加莱群的表征的张量积中的态。

因此，散射理论——如我们最初构想的——需要证明满足探测器标准的态对应于一个适当的庞加莱群表示。特别是，其目的是想证明存在等距 Ω^{in}，Ω^{out}，把福克空间$\mathcal{F}(\mathcal{H})$嵌入到$\mathcal{K}$，以及把给定的$\mathcal{F}(H)$上的庞加莱群的表征跟$\mathcal{K}$上的扭结起来。

在这些想法的基础上，人们已经做出详细模型，主要针对存在质量差异的情况。不幸的是，迄今为止还没有下述情况下的模型，其中$\mathcal{H}^{in}=\mathcal{H}^{out}$，这是理论具有 S 矩阵、定义在输入和输出态之间的跃迁概率的必要条件。(这里的$\mathcal{H}^{in}$是$\mathcal{K}$中福克空间的像，同构于 Ω^{in}，并且类似于$\mathcal{H}^{out}$。)

最近，巴克霍尔兹及其合作者断言维格纳的对称标准太过严格——即存在更一般化的粒子态定义。他们断言只有通过这种更一般的粒子态定义，我们才

能解决“粒子外(infraparticles)”的问题，其中有质量的粒子携带光子云，参见[Buchholz *et al.*, 1991]。

注：对于 AQFT 中散射理论进展的回顾，参见[Haag，1996，Chapter 6]和[Buchholz and Summers，ND]。

5. 在 AQFT 中值的确定性问题

非相对论 QM 的“测量问题”表明，理论的标准方法陷入进退维谷之境：要么(i)在需要解释测量结果时人们必须作出动力学(“塌缩”)专门的调整；要么(ii)跟表面现象相反，测量并没有结果，参见迪克森(Dickson)在本书中所撰写的章节。

对这个困境存在两个主要回应：一方面，有人提议我们放弃 QM 的幺正动力学，以便统计动力学精确预测我们测量结果的实验；另一方面，有人提议我们坚持量子态的幺正动力学，但是某些量(比如粒子位置)即便不能由量子态具体化也能够取值。参看 Dickson 在本书中所撰写的第 5.5 节，对可能回应有更细微差别的讨论。

两种方法——改变动力学的方法和带有附加值的方法——用来回应非相对论 QM 中的测量问题是完全成功的。但是一旦把量子力学跟相对论结合，两种方法都会出问题。特别是，附加值的方法，比如德布罗意—玻姆(de Broglie-Bohm)波导理论，似乎要求所选择的参照系来确定附加值的动力学，参见[Cushing(库欣)，1994，pp. 188—191，196—198]，[Holland(霍兰德)，1995]，以及[Bohm and Hiley(玻姆和希利)，1995，Chapters. 11，12]，并且在此情况下有可能测试洛伦兹不变性失败。

量子力学的“模态”解释，在精神实质上也类似于德布罗意—玻姆理论，但其出发点更加抽象，即关于赋予某些可观测量确定值这个问题上。按照[Bell，1987]，我们可以称之为理论的“披头士(beables)”。不是从直观物理动机出发对确定值(如粒子位置)进行选择，模态解释从数学动机出发选择量子态(比如密度算符)的谱分解为确定值。参看[Dieks and Vermaas(狄克斯和弗马斯)，1998；Vermaas，1999]对模态解释的评论，就动机而言参看[Clifton，1995]。

跟德布罗意理论不一样的是，模态解释并不明显违背相对论约束的精神实质和字面意义，比如洛伦兹不变性[Dickson and Clifton，1998，9]。因此，似乎有希望在AQFT的框架内发展一种模态解释，这是狄克斯[2000]提出AQFT模态解释的出发点。我们不去叙述狄克斯当初的说法，而是直接批判[Clifton，2000]，为读者提供进一步的详细阐述。

5.1 模态代数的克里夫顿—北岛(Klifton-Kitajima)分类

克里夫顿对AQFT模态解释的批评建立在一个值得注意的定理基础之上，它对跟态ρ相关的局域冯·诺伊曼代数$\mathfrak{R}(O)$的所有可能"模态子代数"进行分类。按照克里夫顿的观点(而模态解释似乎同意这一点)确定的局域可观测量的代数$\mathfrak{D}$，$\mathfrak{D}\subseteq\mathfrak{R}(O)$，应该满足跟$\mathfrak{R}(O)$的一个给定态$\rho$相关的下述约束：

定义110。设$\mathfrak{R}$是冯·诺伊曼代数，并且设ρ是$\mathfrak{R}$的一个态，那么$\mathfrak{R}$的一个冯·诺伊曼子代数$\mathfrak{D}$被认为是$(\mathfrak{R},\rho)$的模态代数，当且仅当：

1.(值的确定性)约束态$\rho|_{\mathfrak{D}}$是无弥散态的混合(定义：一个态是无弥散的，当且仅当它对每个投影算符赋值0或者1)；

2.(可定义性)$\mathfrak{D}$是让态ρ不变的在$\mathfrak{R}$全部对称下的左不变量；

3.(极值性)$\mathfrak{D}$是最大的，符合前两个条件。

最后一个要求容易用来排除唯一性的一般反例，比如，人们总可以找出单位元标量乘积的代数$\mathbb{C}I$。第二个要求试图阐明$\mathfrak{D}$是由状态ρ"挑出"(即通过它来定义)的。我们还剩下"对称性"的模糊性概念(而我们将在下一节回到这个问题)，但是克里夫顿使对称性符合$\mathfrak{R}$的$*$自同构，而这对主要结果(定理114)是必需的。

为了陈述这个结果，我们需要定义态的中心化子的概念。下面的命题建立起中心化子的两个可能定义的等效性。

命题111。设$\mathfrak{R}$是冯·诺伊曼代数，令ω是$\mathfrak{R}$的忠实的正规的态，并且假设σ_t^ω是$\mathfrak{R}$的模自同构群，那么下面两个集合是共存的。

1. $\{A\in\mathfrak{R}:\ \sigma_t^\omega(A)=A,\ \forall t\in\mathbb{R}\}$

2. $\{A\in\mathfrak{R}:\ \omega(AB)=\omega(BA),\ \forall B\in\mathfrak{R}\}$

命题 111 的证明依赖于模理论的所有工具，有关的详情我们向读者推荐[Takesaki，2003]。

定义 112。很明显，上述命题中定义的集合事实上是$\mathfrak{R}$的冯·诺伊曼子代数，我们称此子代数为$\mathfrak{R}$中的 ω 的中心化子，并且用$\mathfrak{R}_{\omega}$来表示它。

例子 113。设$\mathfrak{R}=\mathfrak{B}(\mathcal{H})$，并且设 ω 为$\mathfrak{R}$的忠实的正规的态，那么对于稠算子 $D\in\mathfrak{R}\omega$ 具有形式：

$$\omega(A)=\mathrm{Tr}(DA)，A\in\mathfrak{R}$$

特别是，如果$\mathfrak{R}_{\omega}=\{D\}'$，并且 $Z(\mathfrak{R}_{\omega})$是阿贝尔冯·诺伊曼代数$\{D\}''$。特别是，如果 ω 是 I_n 型因子的最大混合态，那么$\mathfrak{R}_{\omega}=\mathfrak{B}(\mathcal{H})$，并且 $Z(\mathfrak{R}_{\omega})=\mathbb{C}I$。

克里夫顿—北岛定理表明，存在唯一的$(\mathfrak{R}，\omega)$模代数，并且在 ω 是忠实的情况下，它是 $Z(\mathfrak{R}_{\omega})$，ω 的中心化子的中心。

定理 114(克里夫顿—北岛)。设$\mathfrak{R}$是作用在希尔伯特空间$\mathcal{H}$上的冯·诺伊曼代数，并且设 ω 是$\mathfrak{R}$的正规态。

1. 如果 ω 是忠实的，那么 $Z(\mathfrak{R}_{\omega})$是$(\mathfrak{R}，\omega)$的唯一模代数；

2. 一般地，$(\mathfrak{R}，\omega)$的唯一模代数是$\mathfrak{N}\oplus Z(\mathfrak{R}_{\omega})E$，其中 E 是满足 $\omega(E)=1$ 的$\mathfrak{R}$中最小投影，并且$\mathfrak{N}$是在$(I-E)(\mathcal{H})$上的所有有界算子的代数。

这个结果在[Clifton，2000]中就 ω 是忠实的情况有证明，一般情况的证明在[Kitajima，2004]中。

正如克里夫顿[2000]所指出的，定理 114 是 AQFT 模态解释的麻烦所在，因为在很多情况下代数 $Z(\mathfrak{R}_{\omega})$是平凡的。参看[Ruetsche and Earman(鲁切和厄尔曼)，2005]中对这一点的进一步发展。

1. 设 W 是闵可夫斯基时空的嵌入区域，并且设 Ω 是真空态，那么不存在模自同构群 σ_t^{ω} 的$\mathfrak{R}(W)$中的固定点。参见命题 77 的证明，以及[Driessler，1975]。因此，$\mathfrak{R}_{\omega}=\mathbb{C}I$，以及 $Z(\mathfrak{R}_{\omega})=\mathbb{C}I$。

2. 在相对论 QFT 中，局域代数是III_1 型超无限因子$\mathfrak{R}$(参见 2.5 节)。但是$\mathfrak{R}$具有遍历态——具有平凡中心化子态的稠集，对于所有这些态，$Z(\mathfrak{R}_{\omega})=\mathbb{C}I$。

因此，它作出了大量的区分，至少对模态解释的可行性局域代数是类型III_1。因为，如果局域代数要么是I_{∞}型要么是III_0型，那么对于模态解释将是好消息。

命题 115。设$\mathfrak{R}$是I_∞型因子，那么对于$\mathfrak{R}$的任何正规态ω，唯一模态代数$\mathfrak{D}_\omega$是非平凡的。

证明。我们有$\mathcal{D}_\omega = Z(\mathfrak{R}_\omega) = \{D\}''$，其中$D$是稠算子，即在$\mathfrak{R}$中的正算子，通过迹公式实现态$\omega$。而且，如果$\mathfrak{R}$是$\mathrm{I}_\infty$，$D$不可能是单位元的乘积。

命题 116。设$\mathfrak{R}$是类型III_0因子，那么对$\mathfrak{R}$的任何忠实的正规态ω，唯一模代数$\mathfrak{D}_\omega$是非平凡的。

证明。在[Takesaki，2003，402]命题 3.15 意味着，$\mathfrak{D}_\omega$没有原子投影，因而是无穷维的。

5.2 什么是 AQFT 中的对称?

在这里我们关注一个将克里夫顿—北岛定理应用于 AQFT 的问题：所牵涉的对称性可能太过自由，尤其对我们具有的时空上的代数网，而不是单个的冯·诺伊曼代数。克里夫顿的定理运用假定了任何$\mathfrak{R}$的自同构是对称的。然而，如果$\mathfrak{R} = \mathfrak{R}(O)$只是一个全网$O \mapsto \mathfrak{R}(O)$的一个代数，那么不清楚是否$\mathfrak{R}$的每个自同构是相关系统的对称性。我们需要的是网$O \mapsto \mathfrak{R}(O)$的对称性概念。

注释 117。一个部分有序集合$\mathcal{K}$能够被看作一个范畴，其中对x，$y \in \mathcal{K}$，如果$x \leqslant y$，那么$\mathrm{Hom}(x, y) = \{(x, y)\}$，否则$\mathrm{Hom}(x, y) = \emptyset$。设$C^*$是具有作为对象的$C^*$代数和具有$*$同态。关于这个概念，在闵可夫斯基时空上的$C^*$代数网，是一个函子$\mathfrak{A}$：$\mathcal{K} \to C^*$，其中$\mathcal{K}$是闵可夫斯基时空中对顶锥的范畴，通过包含排序，并且使得$\mathfrak{A}(\mathrm{Hom}(O_1, O_2))$在$\mathrm{Hom}(O_1, O_2)$不空时是等距的。

定义 118。设$\mathcal{K}$是部分有序集(比如在某些由包含排序的流形中的区域)。设$O \mapsto \mathfrak{A}(O)$和$O \mapsto \mathfrak{B}(O)$，是在$\mathcal{K}$上$C^*$代数的网。一个网同构$\alpha$：$\mathfrak{A} \to \mathfrak{B}$，是一个在函子之间的自然变换，也就是说，$\alpha$由下述同构集组成：

$$\{\alpha_O \colon \mathfrak{A}(O) \to \mathfrak{B}(O) \colon O \in \mathcal{K}\}$$

它在O中是自然的。换言之，对于每个$f \in \mathrm{Hom}(O_1, O_2)$，有$\alpha_{O_2} \circ \mathfrak{A}(f) \to \mathfrak{B}(f) \circ \alpha_{O_1}$，这意味着如下的交换图：

$$\begin{tikzcd}
\mathfrak{A}(O_1) \arrow[r, "\alpha_{O_1}"] \arrow[d, "\mathfrak{A}(f)"'] & \mathfrak{B}(O_1) \arrow[d, "\mathfrak{B}(f)"] \\
\mathfrak{A}(O_2) \arrow[r, "\alpha_{O_2}"'] & \mathfrak{B}(O_2)
\end{tikzcd}$$

事实 119。网自同构对应于准局域代数的自同构，保留每一个子代数整体不变性。为了明确说明这一点，设$\mathfrak{A}$代表从$\mathcal{K}$到 C^* 的函子，并且设$\mathfrak{B}$代表$\mathfrak{A}$的归纳极限，我们把$\mathfrak{A}(O)$等同于它在$\mathfrak{B}$中的像。那么 α 是$\mathfrak{A}$一个网自同构，当且仅当存在一个$\mathfrak{B}$的自同构β，使得：

$$\beta|_{\mathfrak{A}(O)}=\alpha_O$$

现在，给定一个具有归纳极限$\mathfrak{B}$的网$\mathfrak{A}$，那么应该把$\mathfrak{B}$的对称看作什么？

提议 120。$\mathfrak{A}$的对称对应于一个网自同构 α，即$\mathfrak{A}$的自然变换。也即是说，$\mathfrak{A}$的对称对应于一个准局域代数的自同构，保留每个局域子代数整体不变性。

这一个提议确实太过严格，因为它排除了时空基本对称约化的对称情况。但是如果$\mathcal{K}$由时空 M 区域的一个适当集合构成(即在时空对称性下是闭合的集合)，那么 M 的对称性会约化一个在$\mathcal{K}$上的保序双射 F。注意到由于 F 是函子，$\mathfrak{A}\circ F$ 也是函子。因此，我们来看下述自由化定义。

提议 121。一个网$\mathfrak{A}$的对称，由一个(F,α)对构成，其中 F 是一个$\mathcal{K}$的保序双射，并且 α 是从$\mathfrak{A}$到$\mathfrak{A}\circ F$ 的网同构(自然变换)。

如果我们接受这个提议，那么我们必须用下述修正条件取代克里夫顿的可定义性条件：

可定义性 2：给定 $O\in\mathcal{K}$，设$\mathcal{K}_0$是具有对象$\{O_0:\ O_0\leqslant O\}$的$\mathcal{K}$的满子范畴，并且设$\mathfrak{R}_O$代表冯・诺伊曼代数取值函子$\mathfrak{R}$到$\mathcal{K}_0$的限制。那么代数 $\mathfrak{D}$必定是左不变的，依据是保留在$\mathfrak{R}(O)$上的态 ρ 的所有$\mathfrak{R}_O$对称。

由于不是所有$\mathfrak{R}(O)$的自同构都是网$\mathfrak{R}_O$的对称，新的定义性条件比旧的要弱：通常不存在有 $\mathfrak{D}$作用的更多候选者。

克里夫顿—北岛定理在$\mathfrak{R}(O)$的对称性修正定义下并不适用。另一方面，我们不知道一个积极结果，表明$\mathfrak{R}(O)$的子代数的存在性和唯一性，它定义在态 ω 中，并且在保留 ω 的所有网自同构下是不变的。存在有建设性的建议，比如在[Doplicher and Longo，1984]的结果中。

命题 122。设$(\mathfrak{R}_1\subseteq\mathfrak{R}_2,\ \omega)$是一个冯・诺伊曼代数的标准分裂包含，那么存在一个唯一的 I 型因子$\mathfrak{N}$使得：(i)$\mathfrak{R}_1\subseteq\mathfrak{N}\subseteq\mathfrak{R}_2$；(ii)$\mathfrak{N}$在同时保留$\mathfrak{R}_1$和态 ω 的$\mathfrak{R}_2$的全部自同构下是不变的。

当然，代数$\mathfrak{N}$本身并不具有无弥散态，从而不可能是确定可观测量的代数。

然而，态$\omega \mid \mathfrak{R}$是正规的，并且由于$\mathfrak{R}$是类型Ⅰ因子，存在一个稠算子$D \in \mathfrak{R}$约化下述意义上的态；对于所有$A \in \mathfrak{R}$，存在$\omega(A) = \mathrm{Tr}(DA)$，那么假定$\mathfrak{R}_1$，由于某些原因，必定在$\mathfrak{R}_2$的对称下是左不变性的，代数$\mathfrak{D} = \{D\}''$就像是在态$\omega$中$\mathfrak{R}_2$里值确定的可观测量模态解释集的好候选者。

为了把命题122应用到具有$\mathfrak{R}_i = \mathfrak{R}(O_i)$以及$O_1 \subseteq O_2$的AQFT，我们可能不得不假定分裂性质得以保留。虽然分裂性质并非在每个模型中都保留，失去分裂性质意味着一种病态，如果存在某种物理上的病态情况，其中模态解释产生一个确定量的平凡集，这一点毫不奇怪。

注：对于近来讨论适用于相对论环境的模态解释，参见[Myrvold(迈耶弗德)，2002；Ruetsche and Earman，2005]。

6. 量子场和时空点

在QFT的标准/启发性的表述中，基础物理量(可观测量，或者更一般的量子场)是由时空点标记的算子：$\Phi(x)$，可参见特霍夫特('t Hooft)在本书撰写的内容。在这个事实的基础上，至少有一位哲学家[Teller，1995]按照一个算子场及其期望值这个观念描述了QFT的本体论。另一方面，QFT的数学方法(比如怀特曼方法)避免在这些点使用算子，有利于用检验函数$\Phi(f)$涂抹(smeared)在(时)空上的算子。按照阿伦采纽斯(Arnteznius)[2003]，这个事实支持了这样的观念，即时空没有类空事件，时空点上也就更加没有场的值。

随着QFT在数学上越来越严格，会形成一种直观印象，即不仅很难确定场在一点的值，而且也不可能这样做——这些量简直就不存在。可比较冯·诺伊曼对狄拉克(Dirac)δ函数的批评和类点局域化粒子的概念。)这个直觉有时被启发式或者操作主义论证加强——比如，玻尔和皮特森(Bohr and Petersen)[1950]认为不可能测量一点的场强的论证。例如，哈格[1996，58]断言，“某一点的量子场$\Phi(x)$不可能是真正可观测的。”甚至哲学家们有这样的说法，“场算子需要在空间里被‘涂抹’”[Huggett(胡格特)，2000，631]。

但是反对在点上的场算子的论证，常常把可测量性问题跟存在性问题混淆，很少提高到一个足以作出形而上学结论的精确水平。在这一节，我们评论

一些点上的场量存在与否的精确论证。我们将会看到这些结果无法明确排除点上的场量，但是它们澄清了必须作出的解释性取舍。

6.1　不可行定理

在下面三小节里，我们回顾时空点上的场算子不可行定理。

变换协变性排除了点上的算子

第一个不可行定理表明，如果存在变换群的一个连续酉表征，那么在任意固定时间 t，场位形算子 $\phi(x, t)$ 跟场动量算子 $\pi(x', t)$ 对易，即便是在这些算子跟同一点联系起来的时候。这个结果是个严重的问题，因为 $\phi(x, t)$ 跟 $\pi(x, t)$ 被假设成是正则共轭的，参见[Ryder(瑞德尔)，1996，131]和[Huggett，1999]：

$$[\phi(x, t), \pi(x', t)] = i\delta(x - x') \tag{2}$$

甚至，这个不良结果不可能是任何种类的量子力学跟相对论的“冲突”所致，因为这个不良结果也适用于非相对论理论。

定理 123。设 $\phi(x, t)$ 和 $\pi(y, t)$ 是算子场，或者是约束的或者是非约束和自伴的，使得 $x \neq y$ 时有：

$$[\phi(x, t), \pi(y, t)] = 0$$

在非约束情况下，我们的意思是 $\phi(x, t)$ 和 $\pi(y, t)$ 定义在共同的稠集 $\mathcal{D}$ 上，并且它们在这个集上对易。如果 $y \mapsto U(y)$ 是一个连续的变换群表征，使得对于所有的 x，$y \in \mathbb{R}^3$，有 $U(y)\pi(x, t)U(y)^* = \pi(x + y, t)$，那么，对于所有的 x，$y \in \mathbb{R}^3$，有：

$$[\phi(x, t), \pi(x, t)] = 0$$

证明。由于这个证明仅仅把场算子用到相同的时间切片上，我们禁止用来指代时间 t。先假定 $\phi(x)$ 和 $\pi(y)$ 是约束算子，在此情况下，映射：

$$f(y) := [\phi(x), \pi(x + y)] = [\phi(x), U(y)\pi(x)U(y)^*] \tag{3}$$

是从 $\mathbb{R}^3$ 到 $\mathcal{H}$ 上约束算子的弱算子连续函数。选择一个收敛到 0 的非零矢量的系列 $(y_n)_{n\in\mathbb{N}}$，由于 f 是连续的，并且对所有 $n \in \mathbb{N}$ 有 $f(y_n) = 0$，

$$[\phi(x), \pi(x)] = f(0) = \lim_{n\to\infty} f(y_n) = 0 \tag{4}$$

现在假设 $\phi(x)$ 和 $\pi(y)$ 是无界但自伴的。那么用其谱投影 $Es(x)$ 之一替代

$\pi(x)$，其中 S 是$\mathbb{R}$ 的布尔子集，而用其谱投影 $F_{S'}(y)$ 之一替代 $\pi(x)$，其中 S' 是$\mathbb{R}$ 的布尔子集。通过前面的论证，$E_S(x)$ 和 $F_{S'}(y)$ 是对易的，由于这对于所有的这种谱投影对都是正确的，因而 $\phi(x)$ 的谱投影成对地对易于 $\pi(x)$ 的谱投影。因此 $\phi(x)$ 和$\pi(x)$ 是定义在$\mathcal{H}$中的一个公共稠集$\mathcal{D}$上，并且它们在该稠集中相互对易。

庞加莱协变性排除了点上的算子

对于我们，下面有两个不可行定理，我们需要集焦在两个经典结果上。

定义 124。一个函数 f： $\mathbb{R}^n \to \mathbb{C}$被说成是正的类型，当且仅当对于每一个 c_1，…，$c_n \in \mathbb{C}$，以及 x_1，…，$x_n \in \mathbb{R}^n$，我们都有：

$$\sum_{i=1}^{n}\sum_{j=1}^{n}\bar{c}_j c_i f(x_i - x_j) \geqslant 0$$

定理 125（博克纳（Bochner））。设 f： $\mathbb{R}^n \to \mathbb{C}$是个正类型的连续函数，那么 f 的傅里叶变换是$\mathbb{R}^n$ 上的有界测度。

证明。对于博克纳定理的证明，参见[Rudin（鲁丁），1991，303]和[Folland（弗兰德），1995，95]。

注释 126。只有$\mathbb{R}^n$ 的群结构是博克纳的定理真正需要的。因此，把闵可夫斯基时空作为此语境中的$\mathbb{R}^4$，我们没有犯任何错误。

我们在后续结果中需要下面这个引理。

引理 127。设 f 是一个在$\mathbb{R}^n$ 上的连续的确定的正函数，那么 f 是常数 1 函数，当且仅当 f 的傅里叶变换是具有支集$\{0\}$的概率测度。

上面引理的证明是平凡的，具有支集$\{0\}$测量的傅里叶变换是由下式定义的函数 f：

$$f(x) = \int_{\mathbb{R}^4} e^{i(x\cdot p)} d\mu(p) = e^{i(x\cdot 0)} = 1$$

但是傅里叶变换是一个在$\mathbb{R}^n$ 上的复的拉东（Radon）测度跟在$\mathbb{R}^n$ 上的有界连续函数之间的双射。

定义 128。我们认为闵可夫斯基时空上的测度 μ 是洛伦兹不变量，当且仅当对于 M 的每一个布尔子集 S 有 $\mu(\Lambda(S)) = \mu(S)$，以及每一个齐次洛伦兹变换 Λ，其中$\Lambda(S) = \{\Lambda(x): x \in S\}$。

很明显，在闵可夫斯基时空上仅有的洛伦兹不变性概率测度，是在{0}支集的概率测度(齐次洛伦兹群的唯一固定点)。下面的结果是那个事实的“傅里叶变换的”版本。

引理129。 设M是闵可夫斯基时空，如果$f: M\to\mathbb{C}$是正类型的连续函数，使得对于每一个洛伦兹变换Λ都有$f(\Lambda x)=f(x)$，那么f是常数。

证明概略。通过博克纳的定理，如果$f: M\to\mathbb{C}$是正类型的连续函数，那么f是在M上有界测度μ的傅里叶变换。可以直接证明，如果f是洛伦兹不变性那么μ也如此。但是一个约束的洛伦兹不变性测度在{0}上被支持。通过引理127，μ的傅里叶变换是一个常数函数，因此，$f=1$是常数。

事实130。 设U是在希尔伯特空间$\mathcal{H}$上的变换群的酉表征，那么下面说法等价：

1. 表征U的谱是Δ;
2. 对每一个u，$v\in\mathcal{H}$，函数$f: \mathbb{R}^n\to\mathbb{C}$由下式给出：

$$f(x)=\langle u,\ U(x)v\rangle,\qquad x\in\mathbb{R}^n$$

在Δ中存在有支集的傅里叶变换。

最后，我们下面两个结果的核心引理如下。

引理131。 设A: $M\to\mathfrak{B}(\mathcal{H})$是一个值为函数的算子，并且设$U$是一个$\mathcal{H}$上的变换群的酉表征，使得对于所有$x\in M$，有$U(x)A(0)U(x)^*=A(-x)$。定义一个函数$f: M\to\mathbb{C}$，通过下式：

$$f(x)=\langle\Omega,\ A(X)A(0)\Omega\rangle=\langle\Omega,\ U(x)^*A(0)U(x)A(0)\Omega\rangle$$

如果f是常数，那么存在一个$c\in\mathbb{C}$，使得对于每个$x\in M$，有$A(x)\Omega=c\Omega$。

证明。设$\psi=A(0)\Omega$，那么$f(x)=f(0)$表示为：

$$\langle\psi,\ U(x)\psi\rangle=\langle\psi,\ \psi\rangle=\|\psi\|^2$$

但是我们也具有$\|\psi\|=\|U(x)\psi\|$，因为$U(x)$是幺正的，因此：

$$\langle\psi,\ U(x)\psi\rangle=\|\psi\|\cdot\|U(x)\psi\|$$

并且柯西—施瓦茨不等式(Cauchy-Schwartz inequality)意味着对所有的x都有$U(x)\psi=\psi$，即$U(x)A(0)\Omega=A(0)\Omega$。另外注意，尽管有$U(x)A(y)\Omega=U(x+y)A(0)\Omega=A(0)\Omega$。因此所有的矢量$A(x)\Omega$在变换群下是不变的。

现在，第二个不可行定理[Wizimirski（魏兹米塞斯基），1966]表明，不存在作用在闵可夫斯基时空上的有界算子的非平凡的庞加莱协变场。

定理132。假设A：$M\to\mathfrak{B}(\mathcal{H})$是函数取值算子，并且$U$是作用在$\mathcal{H}$上的庞加莱群的连续酉表征，使得：

1. 对于所有的$(y,\Lambda)\in\mathcal{P}$且$x\in M$，有$U(y,\Lambda)A(x)U(y,\Lambda)^*=A((\Lambda x)-y)$；

2. 存在唯一的（对于标量乘积）变换不变性矢量$\Omega\in\mathcal{H}$。

那么存在一个$c\in\mathbb{C}$，使得，对于所有$x\in M$，有$A(x)\Omega=c\Omega$。

注释133。(i)Ω的唯一性假设有可能无法保证，但是在某些十分标准的条件下，这个假设能够导出（参见2.2节）；(ii)这个定理得不出算子$A(x)$和$A(y)$之间对易关系的假定。

定理132的证明：定义函数f：$M\to\mathbb{C}$为：

$$f(x)=\langle\Omega,A(x)^*A(0)\Omega\rangle,\qquad x\in M$$

由条件2，我们有$U(x)\Omega=\Omega$，因此由条件1，我们有$A(x)^*=U(x)A(0)^*U(x)^*$，从而有：

$$f(x)=\langle A(0)\Omega,U(x)^*A(0)\Omega\rangle$$

这明显是正定的。而且，由于$x\mapsto U(x)^*$是弱连续的，f是连续的。

现在我们建立，对于所有$x\in M$和所有洛伦兹变换Λ，都有$f(\Lambda(x))=f(x)$。我们有：

$$\begin{aligned}
f(\Lambda x)&=\langle\Omega,A(\Lambda_x)^*A(0)\Omega\rangle\\
&=\langle\Omega,U(0,\Lambda)A^*(x)U(0,\Lambda)^{-1}A(0)\Omega\rangle\\
&=\langle U(0,\Lambda)^{-1}\Omega,A(x)^*U(0,\Lambda)^{-1}A(0)U(0,\Lambda)\Omega\rangle\\
&=\langle\Omega,A(x)^*A(\Lambda(0))\Omega\rangle\\
&=\langle\Omega,A(x)^*A(0)\Omega\rangle\\
&=f(x)
\end{aligned}$$

因此，引理129意味着f是常数，并且引理131意味着，存在一个$c\in\mathbb{C}$，使得对于所有$x\in M$都有$A(x)\Omega=c\,\Omega$。

微观因果性和谱条件排除位于点上的算子

最后一个不可行定理，源于怀特曼[1964]调用微观因果性和谱条件。参见[Horuzhy，1990，46]和[Baumgärtel and Wollenberg，1992，115]中可供选择的相关证明。

定理 134。假设 A: $M \to \mathfrak{B}(\mathcal{H})$ 是一个取值函数算子，并且 U 是在 $\mathcal{H}$ 上变换群的连续幺正表象，使得：

1. 当 x 和 y 是类空分离时，$[A(x), A(y)]=0$；
2. 对所有 x，$y \in M$，$U(x)A(y)U(x)^{*}=A(y-x)$；
3. U 满足谱条件；
4. 存在一个唯一的变换不变的矢量 $\Omega \in \mathcal{H}$。

那么，存在一个 $c \in \mathbb{C}$，使得对于所有 $x \in M$ 有 $A(x)\Omega = c\,\Omega$。

证明。同上面一样，定义 f ： $M \to \mathbb{C}$ 为：

$$f(x)=\langle \Omega, A(x)A(0)\Omega \rangle, \qquad x \in M$$

固定一个非零类空矢量 x，那么由条件 1 有：

$$U(x)^{*}A(0)U(x)A(0)=A(x)A(0)=A(0)A(x)=A(0)U(x)^{*}A(0)U(x)$$

因此，

$$\begin{aligned} F(x) &= \langle \Omega, U(x)^{*}A(0)U(x)A(0)\Omega \rangle = \langle \Omega, A(0)U(x)^{*}A(0)\Omega \rangle \\ &= \langle \Omega, A(0)U(-x)A(0)\Omega \rangle = f(-x) \end{aligned}$$

现在考虑由 $F(t)=f(tx)$ 给定的函数 F: $\mathbb{R} \to \mathbb{C}$，使得 $F(t)=F(-t)$。由条件 3，f 的傅里叶变换在是未来光锥里成立的。因此，F 的傅里叶变换在 $[0, +\infty)$ 成立。但是，由于 $F(t)=F(-t)$，F 的傅里叶变换在 $(-\infty, 0]$ 也成立。因此，F 的傅里叶变换是在 $\{0\}$ 质量点。由引理 129，F 是常数。最后，由于在 M 中的任何两点都能够用两个类空矢量关联起来，我们能够使用两次前面的方法证明 f 是常数。因此，通过 131 引理，存在 $c \in \mathbb{C}$，使得对所有 $x \in M$ 都有 $A(x)\Omega = c\,\Omega$。

推论 135。设 $O \mapsto \mathfrak{R}(O)$ 是一个不可约化的作用在希尔伯特空间 $\mathcal{H}$ 上的冯·诺伊曼代数网，并且设 U 是强连续酉表征，以实现在网 $\mathfrak{R}$ 上的变换群的作用。假定该网满足微观因果性(假设 43)。假设 U 满足谱条件，并且存在一个变换不变的矢量 $\Omega \in \mathcal{H}$。那么，对于每一个点 $x \in M$，有：

$$\bigcap_{\{O \in K: x \in O\}} \mathfrak{R}(O) = \mathbb{C}I$$

证明。固定 $x \in M$，并且固定对顶锥 $x \in O$，在

$$\bigcap_{\{O \in \mathcal{K}: x \in O\}} \mathfrak{R}(O)$$

里选择一个任意算子，由 $A(x)$ 表示。

而对于一般的 $y \in M$，定义：

$$A(y) = U(x-y)A(x)U(x-y)^{*}$$

使得映射 A: $M \to \mathfrak{B}(\mathcal{H})$ 自动满足定理 134 的条件 2。进而，由于网$\mathfrak{R}$满足微观因果性，并且酉群 U 实现$\mathfrak{R}$上变换，映射 A 满足定理 134 的条件 1。那么，结果是存在 $c \in \mathbb{C}$使得 $A(x) = cI$。因为 x 就是一个 M 的任意元素，结果得证。

6.2 可行定理(Go theorems)

为什么我们要关注如果 $\Phi(x)$ 不能用来表征希尔伯特空间上的任何非平凡算子？这对 QFT 的解释确实有什么意义吗？毕竟，对于任何 x 的领域 O，我们都能找到在 O 中成立的检验函数 f，并且我们也能用表示项“$\Phi(f)$”替代非表示项“$\Phi(x)$”。事实上，我们真的不能把“$\Phi(x)$”看成是序列 $\{\Phi(f_n)\}_{n=1}^{\infty}$ 的另一个名称吗？其中 $\{f_n\}_{n=1}^{\infty}$ 是收敛到 x 的 δ 函数的检验函数序列。更确切地说，似乎我们有可能努力定义伪算子 $\Phi(x)$ 的期望值如下：如果 ρ 是量子场的一个态，那么定义：

$$\rho(\Phi(x)) := \lim_{n \to \infty} \rho(\Phi(f_n)) \tag{5}$$

在这一节中，我们把这个想法落实在点上的场量的两个可行定理中。我们报告的第一个结果是来自于雷伯格(Rehberg)和沃伦伯格的工作[Rehberg and Wollenberg，1986；Wollenberg，1986]，也可参看[Fredenhagen and Hertel(弗雷登阿根和赫特尔)，1981]和[Bostelmann(博斯特尔曼)，2000；Bostelmann，2004]。这个结果证明，在怀特曼框架内，点上的一个量子场能够用一个半双线性形式来表征。第二个结果证明，如果我们把连续性的要求落实到我们变换群的表征上，那么点上的量子场能够用自伴算子表征。

作为半双线性形式的量子场

定义 136。 设$\mathcal{H}$是希尔伯特空间，一个$\mathcal{H}$上的半双线性形式是一个$\mathcal{H}$的线性子空间 $D(t)$ 和在第一论证中是反线性的及在第二论证中是线性的映射 t: $D(t) \times D(t) \to \mathbb{C}$。形式 t 被认为是稠确定的，当且仅当 $D(t)$ 在$\mathcal{H}$中是稠的。

形式 t 被认为是对称的，当且仅当对所有的 φ，$\psi \in D(t)$ 都有 $t(\varphi, \psi) = \overline{t(\psi, \varphi)}$。形式 t 被认为是正的，当且仅当对所有 $\psi \in D(t)$ 都有 $t(\psi, \psi) \geqslant 0$。

定义 137。如果 t 是 $\mathcal{H}$ 上的半双线形式，那么我们通过对所有 $\psi \in D(t)$ 的 $t(\psi) = t(\psi, \psi)$ 定义相关二次形式。一个正的二次形式 t 被认为是封闭的，当且仅当，对于 $D(t)$ 中的任何序列 $(\psi_n)_{n\in\mathbb{N}}$，如果 $\psi_n \to \psi$ 并且 $t(\psi_n - \psi_m) \to 0$，那么 $\psi \in D(t)$ 并且 $t(\psi_n - \psi) \to 0$。

注释 138。一个稠确定的、对称的半双线性形式，是表征一个物理量或者可观测量的第一候选者。由于 t 是对称的，对应的二次形式是实值的，因此，对每一个单位矢量 $\psi \in D(t)$，我们都可以说在态 ψ 中的 t 的"期望值"是 $t(\psi)$。的确，初看上去期望值映射 $t \mapsto t(\psi)$ 似乎具有跟算子相应的期望值映射完全相同的属性。

定理 139。设 $\Phi(\cdot)$ 是一个希尔伯特空间 $\mathcal{H}$ 上的怀特曼场，即 Φ 把一个检验函数空间 $\mathcal{S}(\mathbb{R}^4)$ 的元素，映射到具有某些共同稠定义域 $\mathcal{D}$ 的 $\mathcal{H}$ 上的无界算子。设 $(\delta_n)_{n\in\mathbb{N}}$ 是一个其支集收缩到 x 点的检验函数序列，那么对于每一个 u，$v \in \mathcal{D}$，序列

$$\langle u, \Phi(\delta_1)v\rangle, \langle u, \Phi(\delta_2)v\rangle, \langle u, \Phi(\delta_3)v\rangle, \cdots$$

收敛于一个有限复数，我们用 $\langle u, \Phi(x)v\rangle$ 表示。映射 $u, v \mapsto \langle u, \Phi(x)v\rangle$ 是具有定义域 $\mathcal{D}$ 的半双线性形式，我们用 $\Phi(x)$ 表示。

证明。参看[Baumgärtel and Wollenberg，1992，332]和[Rehberg and Wollenberg，1986；Wollenberg，1986]。

注释 140。人们自然会想得到这个定理的一个更纯粹代数形式的版本，这样的情况在标量代数[Buchholz and Verch，1995；Buchholz，1998]的背景下是可行的。

这个结果有两个原因让人吃惊。我们可能认为 $\Phi(x)$ 不是算子的理由是，因为期望值 $\langle u, \Phi(\delta_n)u\rangle$ 在检验函数 δ_n 收缩到一点时无限增长——即存在某种发散，但是定理 139 证明这个猜想是错的。阻碍 $\Phi(x)$ 成为一个算子的原因一定另有别情。

因此，我们有几个反对量子场作为算子（甚至是无界算子）的不可行定理，和一个关于作为半双线性形式量子场的可行定理。我们应该从这些明显矛盾的

结果中得出什么结论？我们能否说在点 x 存在或者说不存在一个场量吗？

为了回答这个问题我们需要更深入地思考希尔伯特空间上算子跟物理量之间的关系，为什么我们就认为物理量对应于算子？如果我们假定一个无界算子能够表示一个量，那么那个算子一定是自伴的(即 A 跟 A^* 一定在一个共同的稠定义域相一致)，或它的确足够满足一些更弱的条件吗？任何对称的半双线性形式确实具有表征物理量所必需的全部特征吗？为了搞清楚这些问题，先搞清楚半双线性形式跟算子之间的关系的数学性质可能有帮助。幸运的是在这个方向上有不少成果。

显然，每一个在$\mathcal{H}$上的线性(可能无界的)算子 T，通过下面方程定义一个定义域为 $D(T)$ 的半双线性形式：

$$t(\psi, \varphi) = \langle \psi, T\varphi \rangle \tag{6}$$

另一方面，仍然不太清楚的是，什么时候一个任意形式 t 通过方程(6)对应于一个算符。

定义 141。 一个$\mathcal{H}$上的半双线性形式 t 被认为是有界的，当且仅当存在一个 $n \in \mathbb{N}$，使得当任何具有 $\|\varphi\|$，$\|\psi\| \leqslant 1$ 的 φ，$\psi \in D(t)$ 的时候，都有 $|t(\varphi, \psi)| \leqslant n$。

命题 142。 在$\mathcal{H}$上的稠定的有界半双线性形式跟 $\mathfrak{B}(\mathcal{H})$ 的元素之间存在一一对应。特别是，如果 t 是$\mathcal{H}$上有界半双线性形式，那么存在一个唯一算子 $T \in \mathfrak{B}(\mathcal{H})$，使得对于所有 φ，$\psi \in \mathcal{H}$都有 $t(\varphi, \psi) = \langle \varphi, T\psi \rangle$。而且，$t$ 是对称的当且仅当 T 是自伴的。

证明。参看[Kadison and Ringrose，1997，Theorem 2.4.1]。

命题 143。 如果 t 是一个稠定的、正的、封闭的二次形式，那么就存在一个$\mathcal{H}$上唯一的正算符，使得对于所有φ，$\psi \in D(t)$都有 $T^{1/2}$ 的定义域是 $D(t)$ 并且

$$t(\varphi, \psi) = \langle T^{1/2}\varphi, T^{1/2}\psi \rangle$$

特别地，对所有 φ，$\psi \in D(t)$ 都有 $t(\varphi, \psi) = \langle \varphi, T\psi \rangle$。

注释 144。 上述命题在证明一个数字算符 N 能够被定义在外尔代数 $\mathfrak{A}[S, \sigma]$ 的一个表征中的时候是有用的。详情参见[Clifton and Halvorson，2001a]和[Bratteli and Robinson，1997，27]。

上述两个命题并不适用于半双线性形式 $\Phi(x)$，因为它既不是有界的也不

是正的。甚至，并不知道(对作者而言)当一个对称的半双线形式承认一个作为算子的表征时的特征的情况，虽然在此方向存在某些部分结果，参见[McIntosh(麦金托什)，1970]。显然 $\Phi(x)$不是算符，不清楚算符具有而 $\Phi(x)$缺少的特征，以及这些特征是不是一个数学对象表示一个量所必需的。相应地，$\Phi(x)$是否表示一个实在的要素也不清楚。

作为在不可分离的希尔伯特空间上算符的量子场

我们关于点上的量子场算符的第二个可行结果，实际上只是一个例子的概述。我们采取一个不可分离的希尔伯特空间 H，能够表示具有点位置的粒子的状态，可与[Halvorson，2004]对比。然后我们把标准的二次量子化程序——它并不依赖于可分离的单个粒子空间，用来获得福克空间$\mathcal{F}(H)$以及用$\mathbb{R}$中的点来标记的自伴场算符 $\phi(x)$，$\pi(x)$。

设 $H=l_2(\mathbb{R})$是在$\mathbb{R}$上的平方可求和序列的希尔伯特空间；即 $l_2(\mathbb{R})$的一个元素 f 是从$\mathbb{R}$映射到$\mathbb{C}$，使得除了可数的多个点之外 f 为零，并且 $\sum_{x\in\mathbb{R}}|f(x)|^2<\infty$。$l_2(\mathbb{R})$上的内积通过下式得到：

$$\langle f,g\rangle = \sum_{x\in\mathbb{R}} \overline{f(x)}g(x) \tag{7}$$

设$\mathcal{F}(H)$是 H 上的福克空间，对于每一个 $x\in\mathbb{R}$，我们设 $\delta_x\in l_2(\mathbb{R})$代表$(x)$的特征函数，集$\{\delta_x: x\in\mathbb{R}\}$是一个 $l_2(\mathbb{R})$的(不可数的无穷)正交基。对于任何 $x\in\mathbb{R}$，我们定义产生算符 $a(x)$为：

$$a(x)(f_1\otimes\cdots\otimes f_n) := \delta_x\otimes f_1\otimes\cdots\otimes f_n \tag{8}$$

跟在标准情况中一样，我们证明 $a^-(x)+ia^+(x)$和 $ia^+(x)-ia^-(x)$都是前封闭的，即这些算符的图像的封闭性是线性算符的图像，参见[Kadison and Ringrose，1997，155]，我们表示为：

$$\phi(x)=\overline{a^-(x)+ia^+(x)} \tag{9}$$

$$\pi(x)=\overline{a^+(x)-ia^-(x)} \tag{10}$$

那么可以得出 $\phi(x)$和 $\pi(x)$都是自伴的，并且在一个$\mathcal{F}(H)$里的稠定义域$\mathcal{D}$上，我们有：

$$[\pi(x),\ \phi(x')]=i\langle\delta_x,\ \delta'_x\rangle=i\delta_0(x-x') \tag{11}$$

其中 δ_0 是一个完备的合法数学对象，即在$\{0\}$处的概率测度得到支持。

考虑在 $l_2(\mathbb{R})$ 上变换群的(非连续的)表征 $x \mapsto V(x)$，在基本元素 $\{\delta_y: y \in \mathbb{R}\}$ 上通过下式定义：

$$V(x)\delta_y = \delta_{y-x} \tag{12}$$

设 Γ 是“福克函子”，即 Γ 映射一个在单粒子空间 H 上的酉算子 V 到 $\mathcal{F}(H)$ 上的对应算子：

$$I \oplus V \oplus (V \oplus V) \oplus \cdots$$

那么 $x \mapsto U(x) \coloneqq \Gamma(V(x))$ 是一个在 $\mathcal{F}(H)$ 上的变换群的非连续表示，并且：

$$U(x)^* \phi(y) U(x) = \Phi(y-x) \tag{13}$$

因此，$(\phi(\cdot), \pi(\cdot), \mathcal{F}(H), U)$ 是 $l_2(\mathcal{R})$ 之上的一个场系统，其中 $x \mapsto U(x)$ 是一个非连续酉表征。那么我们可能使用这个场系统，定义一个 $\mathcal{F}(H)$ 上的冯·诺伊曼代数网 $O \mapsto \Re(O)$。但是这个网当然不能满足谱条件，因为变换群的表征不是连续的。

刚才描述的这个模型可能不够宽，不足以用来描述真实物理情形。而且，没有好的理由来思考我们接下来要推广到相互作用理论情况的过程，其中局域化类点算子是场算子的积有意义所需的。然而，我们希望能够不通过某些不可行定理的技术假设就能得到证明，这是可以理解的。因此，在我们试图使用这些定理得出为何 QFT 必须被解释的结论之前，我们应该非常认真地思考。

6.3 QFT 的场解释

在第 4 节，我们看见存在 QFT 粒子解释的严重障碍。人们可能通过一个排除过程讨论我们应该采用的 QFT 的“场解释”，参见例如[Teller, 1995; Huggett, 2000]。但是，如果我们检验立足于本身优点的场解释，说它比粒子解释更好，这一点还是不清楚。

在构建正则自由理论(比如自由玻色场和费米场)时，人们从一个希尔伯特空间 H 开始，该空间能够被解释成要么是“单粒子空间”(即单个量子力学粒子的波函数的空间)，要么是一个经典场的位形空间。对应于这两种解释，有两种构建量子场论的希尔伯特空间的两种方法：

1. 二次量子化：量子场的希尔伯特空间是在 H 之上的福克空间(见 4.1 节)。

2. 场量子化：量子场的希尔伯特空间是从 H 到ℂ的“二次可积”函数的空间 $L_2(H, d)$，　相对于 H 上的异常分布 d。

在一个严格处理中，$L_2(H, d)$的元素不是真正的函数，详情参见[Baez *et al.*, 1992, Section 1.3]。由这两种方法建构的自由场论常常是酉等价的。不过，场量子化方法使其能更自然地给出场解释。的确，在 QFT 基础的最新研究中[Huggett, 2000]，人们为波函数的空间 $H := L_2(\mathbb{R}^{3n})$上的函数发现富有启发性的记法：

$$\Psi(\phi), \qquad \phi \in L_2(\mathbb{R}^{3n})$$

因此，似乎是量子场态能被解释成经典场位形的叠加，跟 n 粒子的波函数能够被解释成与 n 粒子的经典位形的叠加是相同的意义。

然而，这种方法存在困难。第一，$L_2(H, d)$上的场算符 $\Phi(x)$跟粒子系统的位置算符 Q_i 精确相似。那就是，存在一个作为相互对易算符的族$\{\Phi(x): x \in M\}$之上的概率分布的一个函数 $\Psi \in L_2(H, d)$的自然解释。但是算符 $\Phi(x)$的不可行定理不利于把 Ψ 解释成经典场位形之上的概率分布。更直接地讲，由于 d 把零测度赋予 H 中的点(即给单个的场位形)，H 的单个子集的特征函数——即决定场位形的——等同于 $L_2(H, d)$中的零矢量。那就是说，没有处于确定位形量子场的状态。

其原因在于前面的考虑，即在一点的场算符的不可行定理，按照数字算符的不可行定理破坏粒子解释的同样方式破坏了 QFT 的场解释。因此，我们应该小心建立在粒子解释问题基础上的场解释论证。

6.4 时间点?

前面的结果致力于回答在一个时空点上是否能够存在场算符的问题。假定我们承认不可能有这样的场算符，并且我们继续使用标准的数学上严格的方法，比方说自由玻色场，其中场算符是被检验函数在空间上被涂抹了的，参见[Araki, 1963]。在这个情况下，量没被联系在时空点上，而是联系在类点时间上。然而，有些说法认为通常这些量也会不得不在时间中不明确。比如按照哈格说法：

> 重整化理论提出在空间和时间中涂抹掉 Φ 是关键的，跟自由场的情况相反，在那里在一个固定时间的三维空间上的一次平均就足够。由于比较强的奇异性，人们不能假设在相同时间上有场的确定对易关系[Haag，1996，59]。

但是这种说法有待推敲——我们知道没有能够证明相互作用场一定在时间中被涂抹掉的定理。因此，目前我们还没有特别好的理由得出结论说时间是无意义的。

7. 不等价表象的问题

局域量子物理(按哈格的术语)的哲学，是 QFT 的理论部分(比如不可观测的场、规范群)不应该算作已知数据的部分。取而代之的是，可观测量代数的抽象网 $\mathfrak{A}$ 应该作为最原始的。按照[Ruetsche，2002]的术语，我们定义“代数扩张主义(Algebraic Imperialism)”立场为：

> 量子场论的物理内容藏在网 $O\mapsto\mathfrak{A}(O)$ 中，对应于物理对称(包括动力学)的 Aut($\mathfrak{A}$)的子群，以及在准局域代数 $\mathfrak{A}$ 上的态。一个 $\mathfrak{A}$ 的表象($\mathcal{H}$，π)可能有助于计算，但是没有本体论意义。

这样一个态度对于 QM 的传统希尔伯特空间形式化体系中那些偏见似乎是不可理解的。的确，哈密顿(Hamiltonian)量在哪儿？跃迁概率在哪？以及我们如何描述测量？代数形式化体系的抽象性和一般性似乎掏空了我们在物理理论中期望的大量内容。

然而，其中一些人担心抽象代数形式化体系缺少内容是没有事实根据的，实际上，GNS 定理(定理 17)证明所有我们不再需要的希尔伯特空间被隐藏在代数自身内部。而且，很多人在基本 QM 中学会说的词汇能够在这个纯粹抽象背景中被定义。比如，态之间跃迁概率的定义参看[Roberts and Roepstorff(罗伯茨和勒普斯托夫)，1968]，而对测量概率的定义参看[Wald(瓦尔德)，1994]。

但是 QFT 的传统词汇能够在代数背景中复制其所有部分，这一点并不正确——至少就表面来看是如此。比如，准局域代数并不包含一个数字算符，而且数字算符谱上的概率分布也不能通过$\mathfrak{A}$上的期望值来定义，参见[Clifton and Halvorson，2001a]。或许更糟的是，从可观测量代数网 $O \mapsto \mathfrak{A}(O)$ 出发，我们已经完全看不见不可观测场的存在，它通常类空分离且彼此并不对易。因此，我们似乎没有办法解释诸如自旋(场算符的对易关系)跟统计学之间关联的 QFT 的深层理论事实。

这样的担心迫使我们转向第二个关于表征问题的主要观点，鲁切[2002]所谓的希尔伯特空间保守主义(Hilbert Space Conservation)：

理论不是网 $O \mapsto \mathfrak{A}(O)$，而是该网加上一个具体的表征($\mathcal{H}$，$\pi$)。

事实上，希尔伯特空间保守主义可能被看成是主流(拉格朗日量)QFT 的大多数工作者的默认观点，因为抽象代数(及其表征)在那起不到核心作用。

虽然有许多实在论观点，但是保守主义观点面临许多认识论上的困难：我们怎样决定哪一个是正确表征呢？在此情况下，困难特别大，因为在数学上能够证明，在任何一个表征当中态的预言能够用任何另一个表征的态复制到任意高的精确度。① 那是由于这样的事实，因为$\mathfrak{A}$是简单的，菲尔(Fell)定理意味着在任何叶形线中的态都是在态空间中弱*稠的。

虽然如此，还是容易想到代数形式化体系会引起一个解释困难。那就是，容易想到如果我们坚持用 QFT 的老方法做研究，不等价表征的问题并不产生，从而也不会有这样的解释困境。因此，不等价表征向我们透露了基础重要性的某些东西，或者它们正是些数学玩意？

代数方法的有力论证包括了正则对易关系的不等价表征的存在，以及跟瑞德勒—弗林量子(Rindler-Fulling quanta)相关的物理效应。但是，这些论证因为各种原因被反对，比如，怀疑瑞德勒真空表征作为实在的描述次于闵可夫斯基

① 陈述这个问题的方法是有偏见的，而且依赖于取"表征的预言"为在抽象代数中可观测量的期望值的意思，如果我们也把在弱封闭 $\pi(\mathfrak{A})^-$ 中可观测量的期望值包括在内，并且包括在$\mathcal{H}$上的无边界算符的期望值，那么理论就会变得更加复杂。比较[Clifton and Halvorson，2001b]。

真空表象。因此，在下一节中，我们讨论代数方法的另一个有力论证——即超选择规则。正是在超选择规则的分析中代数方法最清楚地展示了它自身的优美、实用和基础重要性。

7.1 超选择规则

在一篇现在非常著名的论文中，威克(Wick)、怀特曼和维格纳[1952]论证存在具有状态空间 H 的物理系统，并且态矢量 ψ_1，$\psi_2 \in H$ 使得线性结合

$$2^{-1/2}(\psi_1 + e^{i\theta}\psi_2), \qquad \theta \in [0, 2\pi) \tag{14}$$

产生了“经验上不可区分的”态。一旦如此，威克等人说在 ψ_1 和 ψ_2 之间存在“超选择规则”；或者，ψ_1 和 ψ_2 处在不同“超选择截面”内，我们对“经验上不可区分的”加了引号，因为学术界一直不清楚方程(14)中状态之间关系的性质。方程(14)中的状态仅仅是经验上不可区分的吗，或者存在一种更强的意义上的这两个态是等价的？如果不可区分性是经验的，模态强度如何？这些状态预言了在所有物理上的可能世界中存在同样的经验现象了吗？或者由于某些特殊特征(比如初始条件)它们在我们这个世界中的不可区分性是什么？在这篇论文中，我们不打算解决这些有关超选择规则性质的重要问题。① 相反，我们满足于解释多普利克尔、哈格以及罗伯茨(DHR)使威克等人的概念在 AQFT 的语境中变得明确所作的提议。

对超选择规则的第一种方法涉及人为抹杀状态空间和可观测量代数之间区别，结果混淆了什么应该作为最终理论的状态和可观测量(或者量)。人们从具有产生纯态的单位向量，以及具有作为可观测量(或者量)的 $B(H)$ 的自伴元素的希尔伯特空间 H 开始，然后给出一个启发性论证，宣称超选择规则存在于某些子空间 H_1 中的态矢量和补充子空间中的态矢量 $H_2 := H_1^{\perp}$ 之间。在此论证的基础上，状态空间 H 被约化到 H_1 和 H_2 的结合上，那就是，在 H_1 和 H_2 中的一

---

① 超选择规则也有其基本意义，因为它们被看作有助于测量问题，参见[Beltrametti and Cassinelli(贝尔特拉梅蒂和卡西内利)，1981，74]、[Landsman(兰兹曼)，1995]、[van Fraassen，1991，pp. 264—272]以及更一般地是因为它们跟经典世界的突然出现有关[Giulini(朱利尼)，2003]。当然，我们不在这一章处理这些具体问题。

个矢量的线性结合不再被看成是理论的一个可能(纯态)，纯态矢量要么依赖于 H_1 要么依赖于 H_2。等价地，可观测量 $B(H)$ 的代数被约化到 $B(H_1)\oplus B(H_2)$。能够在方程(14)中的态之间作区分的在 $B(H_1\oplus H_2)$ 中的算符被降级到"不可观测量"的地位，因此，可观测量的代数实际上是 $B(H_1)\oplus B(H_2)$ 而不是 $B(H_1\oplus H_2)$。

现在，代数方法提供了两种反相关的超选择规则的方法。

1. 第一，我们能够以一种更加原则性的方式导出最初的"状态空间抹杀"方法：假设我们已知的是以算符作用在某些大希尔伯特空间 $\mathcal{H}$ 上的一些场，设 $\mathfrak{F}$ 代表场算子的代数。这里的 $\mathfrak{F}$ 类似于代数 $B(H_1\oplus H_2)$，而 $\mathcal{H}$ 类似于 $H_1\oplus H_2$。当然，在此情况下，我们没有把 $\mathcal{H}$ 先验地分解成一个直和。假定我们已经知道作用在场上的一个规范群 G，然后我们定义可观测量为规范不变场。设 $\mathfrak{A}$ 代表可观测量的代数，我们也定义物理纯态为在规范群下不可约化变换的那些 $\mathcal{H}$ 中的向量。一系列的数学结果(部分描述在第9节里)表明，$\mathcal{H}$ 分解成在规范群下不可约化变换的子空间中的一个直和 $\oplus\mathcal{H}_\xi$；而且每一个子空间 $\mathcal{H}_\xi$ 对应于可观测量 $\mathfrak{A}$ 的代数的一个不可约化表征。我们在第9节概述这个"从上到下"的超选择规则方法。

2. 替代从场代数 $\mathfrak{F}$ 开始而导出超选择结构(即可观测量的代数 $\mathfrak{A}$ 的那套物理上感兴趣的表征)，我们能够从 $\mathfrak{A}$ 开始进而考虑其物理表象的集合。$\mathfrak{A}$ 的一个"物理"表象是什么？按照多普利克尔、哈格以及罗伯茨提出的标准(DHR 选择标准)，物理表征是那些仅在局域区域观测上不同于真空表征的那些表征。在此情况下，我们仍然拥有超选择截面的概念，但是还是没有场概念或者规范群的概念。还不能马上知道我们是否具有足够的结构来解释现象。

然而，正是在这一点上开始了深层数学分析。首先，人们证明 DHR 表象的范畴明确对应着可观测代数 $\mathfrak{A}$ 的局域化可输运的自同态的集合 Δ(参看8.2节)。其次，人们证明了集合 Δ 自然具有对称张量 * 范畴的结构(参见第8节)。最后，多普利克尔—罗伯茨重构定理(Doplicher-Roberts Reconstruction Theorem)表明，不可观测场 $\mathfrak{F}$ 和规范群 G 能够从范畴 Δ 唯一地重构。

下面几节概述了我们在超选择规则研究中某些最重要的见解，以及这个分析跟关于不等价表征的作用的基础问题有何关系。总之，我们的结论认为不等价表征是不相关的，并且它们也不成其问题。相反，正是表象范畴的结构提供

了 QFT 的真正有趣的理论内容。

7.2 关于可观测量代数的最小假设

就我们的超选择理论讨论来说，我们需要的只是一个关于可观测量代数网假设相当减少后的集合。因此，我们现在实际上取消了我们在第 2 节中对网作的全部假设。我们从一个白板开始，并且只加上那些我们在下面几节里会需要的假设。

把$\mathfrak{A}$叫作一个“网”，我们假设如果 $O_1 \subseteq O_2$，那么$\mathfrak{A}(O_1) \subseteq \mathfrak{A}(O_2)$，但是我们并没有把这提高到一个假定的地位。

假设 145(微观因果性)。如果 O_1 和 O_2 是类空分离的，那么 $\{\mathfrak{A}(O_1), \mathfrak{A}(O_2)\} = \{0\}$。

假设 146(性质 B)。冯·诺伊曼代数网 $O \to \mathfrak{R}_0(O) \equiv \pi_0(\mathfrak{A}(O))''$满足性质 B，其中$(\mathcal{H}_0, \pi_0)$是由 ω_0 引入的$\mathfrak{A}$的 GNS 表象。

假设 147(对偶性)。$(\mathfrak{A}, \omega_0)$对满足哈格对偶性，即：

$$\pi_0(\mathfrak{A}(O'))' = \pi_0(\mathfrak{A}(O))''$$

对每一个对顶锥 O，其中$(\mathcal{H}_0, \pi_0)$是由 ω_0 引入的$\mathfrak{A}$的 GNS 表象。

假设 148(分离性)。真空希尔伯特空间$\mathcal{H}_0$是可分离的。

假设 149(非平凡性)。对每一个对顶锥 O，$\pi_0(\mathfrak{A}(O))$包含一个并不是单位元倍数的算符，即 $\pi_0(\mathfrak{A}(O)) \neq \mathbb{C}I$。

对这些假设有几条评论：(i)第一条假设是关于网$\mathfrak{A}$的，而其余假设适用于$(\mathfrak{A}, \omega_0)$对，其中$\mathfrak{A}$是准局域代数，而 ω_0 是某些固定态。(ii)对偶性假设说的是，不仅$\mathfrak{R}_0(O')$中的可观测量是兼容于$\mathfrak{R}_0(O)$中的可观测量，而且$\mathfrak{R}_0(O')$包含所有跟$\mathfrak{R}_0(O)$中的可观测量集合兼容的可观测量。我们也会在(关于 DHR 超选择理论的)下面两节中假设，网$\mathfrak{A}$满足跟某些特殊真空态 ω_0 有关系的对偶性，但是它不是因为这种网满足跟任何物理表象有关系的对偶性。事实上，一个表征满足对偶性，当且仅当那个截面具有正常的(玻色/费米)统计；并且任何表征满足对偶性，当且仅当规范群是阿贝尔群。(iii)真空截面里的对偶性等价于自发破缺规范对称的不存在。关于对称破缺的情况，我们将施加一个较弱的要求：本质对偶性，例如 10.7 节。(iv)可分离性假设仅仅只有一次被调用——为了

说明由局域场引入的全部超选择截面是强局域等价的(命题243)。

很明显，注意我们现在没有作出下述假设：(i)没有关于在代数$\mathfrak{A}$上的时空对称作用的假设(平移对称、洛伦兹对称)；(ii)没有对真空态ω_0是平移不变性的效应的假设；(iii)没有关于在真空希尔伯特空间上时空对称作用的假设；(iv)没有关于谱条件的假设。

8. 局域化的可输运自同态的范畴 Δ

在这一节我们研究可观测量代数$\mathfrak{A}$的局域化可输运自同态的范畴$\Delta(\mathfrak{A})$。由于这个研究的物理动机可能不是一开始就清楚的，我们停下来看看自同态跟表征之间的关系。

假设π_0是一些物理上重要的$\mathfrak{A}$的一个固定表征，比如真空表征。那么对于$\mathfrak{A}$的任何自同态ρ，组合$\pi_0\circ\rho$也是$\mathfrak{A}$的一个表征。因此，$\mathfrak{A}$的自同态自然对应于$\mathfrak{A}$的表征，并且我们通过研究$\mathfrak{A}$的自同态，希望能够获得$\mathfrak{A}$的表征的结构。不过，$\mathfrak{A}$的自同态的集合 End $\mathfrak{A}$比起$\mathfrak{A}$的表征的集合 Rep $\mathfrak{A}$具有更多内部结构——比如，存在一个 End $\mathfrak{A}$上的积(即组合)运算，并且某些自同态具有逆。因此，除了传统的等价概念和表征的不相交性，存在具有$\rho\in$ End $\mathfrak{A}$的形式$\pi_0\circ\rho$的表征集合之上的物理重要性的传统关系。

如果拉格朗日量 QFT 的问题在于只存在一个希尔伯特空间，那么 AQFT 的问题在于具有太多的希尔伯特空间！确实，并非所有的$\mathfrak{A}$的表征都是物理的。在第9节中，我们将从一个更传统的观点看这个问题，特别是我们从一个作用在希尔伯特空间$\mathcal{H}$算子的场代数$\mathfrak{F}$开始，以及从一个$\mathcal{H}$上酉算子的规范群G开始，我们可以假定G是某些比如$SU(2)$基础对称群表征的像。我们也假定$\mathcal{H}$包含一个真空态Ω，那么我们把可观测量代数$\mathfrak{A}$定义为规范不变性场。但是那样我们就再次处于 AQFT 的领域：我们具有一个在$\mathcal{H}$上$\mathfrak{A}$的不可约化表征，并且π的不可约化子表征是这样的超选择截面，它们能够通过局域(不可观测)场的作用从真空截面达到。不是$\mathfrak{A}$的所有表征都出现在π的分解中——那些并非是剩余结构。不过，所有出现在π的分解中的表征都是$\pi_0\circ\rho$形式的，具有ρ是一个来自范畴$\Delta(\mathfrak{A})$的自同态！因此，研究这些自同态的动机是它们对应于这

样的表征，它们通过作用在具有(不可观测)场的真空表征的方法，以这个传统的物理上动机明显的方式产生。①

还存在另一个研究DHR范畴的动机：我们要理解规范对称的实质，而DR重构定理提供了关键的见解。特别是，该定理表明DHR范畴是处于有紧致群的对偶性中(在数学上精确的意义上)。因此，哪里有紧致群，哪里就有DHR范畴，反之亦然。这个DHR范畴研究和紧致规范群的研究是同一的，或者，一个更有争议性的例子是，$\mathfrak{A}$的物理表征的范畴结构解释了为什么存在一个紧致规范群，参见[Roberts，1975]。

我们现在定义范畴$\Delta=\Delta(\mathfrak{A})$，并且揭示出它的一些自然结构。正如前面讲的，我们的范畴Δ的对象，应该是$\mathfrak{A}$的$*$自同态的集合End $\mathfrak{A}$的一个子集。

定义150。设ρ是一个$\mathfrak{A}$的$*$自同态，即ρ: $\mathfrak{A}\to\mathfrak{A}$是一个$*$同态(不一定是满射的)。设$O$是一个在(闵可夫斯基)时空中的一个对顶锥，那么$\rho$被认为在$O$中是局域化的，当且仅当，对所有$A\in\mathfrak{A}(O')$都有$\rho(A)=A$，其中$O'$是$O$的类空元素。我们认为$\rho$是局域化的，当且仅当，存在一个对顶锥$O$，在里面它是局域化的。

注释151。通过定义，一个局域化自同态满足$\rho(I)=I$，其中I是$\mathfrak{A}$中单位元。

定义152。如果ρ在O中是局域化的，那么ρ被认为是可输运的，当且仅当，对于任何其他对顶锥O_1，存在一个在O_1中局域化的自同态ρ_1和一个酉算子$U\in\mathfrak{A}$，使得对于所有$A\in\mathfrak{A}$都有$U\rho(A)=\rho_1(A)U$。

定义153。对于每一个对顶锥$O\in\mathcal{K}$，我们假设$\Delta(O)$代表在O中局域化的可输运自态射的集合，并且我们设$\Delta=\cup_{O\in\mathcal{K}}\Delta(O)$，$\Delta$的元素是DHR范畴的对象。

现在我们必须定义对象之间的箭头。

定义154。设ρ，$\rho'\in\Delta$，我们定义在ρ跟ρ'之间箭头的集合Hom(ρ，ρ')如

---

① DHR表征并不包括通过非局域场就能从真空场得到的那些，因而DHR超选择规则的范围并不包括长程力的理论。不过局域场已经够复杂了，而且对更一般的情况来说是一个好的练习。

下：

$$\mathrm{Hom}(\rho,\ \rho') := \{T \in \mathfrak{A}:\ T\rho(A) = \rho'(A)T,\ \ \forall A \in \mathfrak{A}\}$$

如果 $T \in \mathrm{Hom}(\rho,\ \rho')$ 并且 $S \in \mathrm{Hom}(\rho',\ \sigma)$，那么我们定义 $S \circ T = ST$，这在 $\mathrm{Hom}(\rho,\ \sigma)$ 是很明显的。

显然，单位元 $I \in \mathfrak{A}$ 函数是所有对象的单位元箭头，即对所有 $\rho \in \mathrm{Obj}(\Delta)$ 都有 $I = \mathrm{id}_\rho \in \mathrm{End}(\rho)$。我们偶尔也会用 I_ρ 来指我们把 I 看作 $\mathrm{End}(\rho)$ 的单位元。

引理 155。假定对于 $i=1$，2 有 $\rho_i \in \Delta(O_i)$，以及 $T \in \mathrm{Hom}(\rho_1,\ \rho_2)$，那么对任何包含 $O_1 \cup O_2$ 的对顶锥 O，我们具有 $T \in \mathfrak{A}(O)$。

证明。设 $B \in \mathfrak{A}(O')$，那么：

$$TB = T\rho_1(B) = \rho_2(B)T = BT$$

因此 $T \in \mathfrak{A}(O')'$，根据真空截面里的对偶性，$T \in \mathfrak{A}(O)$。

命题 156。借助上面所给的 hom 集合的定义，Δ 是一个范畴。

证明。直接得证。

因此，我们已经证明 Δ 是个范畴。在这一节剩下的部分，我们揭示更多 Δ 上的结构。我们首先证明 Δ 是 C^* 范畴，这涉及 Δ 有直和（$\oplus$运算）、子对象，以及 Δ 的 hom 集是具有 $*$ 运算的矢量空间和遵守适当类似于 C^* 代数范数性质的范数 $\|\cdot\|$。然后我们落实到 hom-集合上的范数，并且证明存在一个 Δ 上的积运算使得（Δ，$\otimes$，ι）是张量 $*$ 范畴。

定义 157。一个范畴$\mathcal{C}$被认为是复数域$\mathbb{C}$上的一个线性范畴，或者一个$\mathbb{C}$线性范畴，当且仅当对于所有X，$Y \in \mathrm{Obj}(\mathcal{C})$，$\mathrm{Hom}(X,\ Y)$是一个复矢量空间以及态射的复合，是双线性的。在我们谈到$\mathbb{C}$线性范畴时，我们设定所有函数应该是$\mathbb{C}$线性的。

定义 158。一个在$\mathbb{C}$线性范畴$\mathcal{C}$上的 $*$ 运算，是一个赋予一个箭头 $s \in \mathrm{Hom}(X,\ Y)$ 到另一个箭头 $s^* \in \mathrm{Hom}(Y,\ X)$ 的映射。这个映射不得不是反线性的、对合(involutive)的($s^{**}=s$)以及逆变($(s \circ t)^* = t^* \circ s^*$)的。一个 $*$ 运算是正的，当且仅当 $s^* \circ s = 0$ 蕴含 $s=0$。一个 $*$ 范畴是一个具有正 $*$ 运算的$\mathbb{C}$线性范畴。

注释 159。如果$\mathcal{C}$是一个 $*$ 范畴，那么对每一个 $X \in \mathrm{Obj}(\mathcal{C})$，$\mathrm{End}(X)$是一个 $*$ 代数。

定义 160。一个 $*$ 范畴被称为 C^* 范畴，当且仅当，对于所有 X，$Y \in$

$\mathrm{Obj}(\mathcal{C})$,存在一个 $\mathrm{Hom}(X, Y)$ 上范数 $\|\cdot\|_{X,Y}$，使得 $\langle\mathrm{Hom}(X, Y), \|\cdot\|_{X,Y}\rangle$ 是巴拿赫(*Banach*)空间并且：

$$\|s\circ t\|_{X,Z}\leqslant\|s\|_{Y,Z}\cdot\|t\|_{X,Y}\qquad\forall s\in\mathrm{Hom}(Y, Z),\ \forall t\in\mathrm{Hom}(X, Y)$$

$$\|s^*\circ s\|_{X,X}=\|s\|^2_{X,Y},\qquad\forall s\in\mathrm{Hom}(X, Y)$$

我们从 * 代数理论借来一些定义。

定义 161。设 C 是一个 * 范畴。一个箭头 $f\in\mathrm{Hom}(X, Y)$ 被说成是一个等距的，当且仅当 $f^*\circ f=\mathrm{id}_X$。一个箭头 $f\in\mathrm{Hom}(X, Y)$ 被说成是幺正的，当且仅当 f 和 f^* 是等距。如果 $p=p^*$ 和 $p\circ p=p$，一个箭头 $p\in\mathrm{End}(Y)=\mathrm{Hom}(Y, Y)$ 被认为是投影。

注释 162。如果 $s\in\mathrm{Hom}(Y, X)$ 是等距的，那么箭头 $p\equiv s\circ s^*\in\mathrm{End}(X)$ 是一个投影。

定义 163。设 $\mathcal{C}$ 是一个 * 范畴。如果 $X, Y\in\mathrm{Obj}(\mathcal{C})$，那么 X 被说成是 Y 的一个子对象，当且仅当存在一个等距 $f\in\mathrm{Hom}(X, Y)$。(粗略地说，存在一个 X 到 Y 的等距嵌入。) * 范畴 $\mathcal{C}$ 被说成是具有子对象的，当且仅当，对每一个 $Y\in\mathrm{Obj}(\mathcal{C})$ 和投射 $g\in\mathrm{End}(Y)$，存在一个 $X\in\mathrm{Obj}(\mathcal{C})$ 和一个等距 $f\in\mathrm{Hom}(X, Y)$，使得 $f\circ f^*=g$。* 范畴 $\mathcal{C}$ 被认为具有直和，当且仅当，任何在 $\mathcal{C}$ 中的两个对象 X，Y，存在一个 $\mathcal{C}$ 中的对象 Z 和等距 $f\in\mathrm{Hom}(X, Z)$，$g\in\mathrm{Hom}(Y, Z)$，使得 $f\circ f^*+g\circ g^*=\mathrm{id}_Z$。

我们以证明 DHR 范畴 Δ 是 * 范畴开始，即 hom 集合是 $\mathbb{C}$ 上的矢量空间，并且存在一个正的 * 运算。

引理 164。DHR 范畴 Δ 是 * 范畴。那就是，如果 $\rho, \sigma\in\mathrm{Obj}(\Delta)$，那么 $\mathrm{Hom}(\rho, \sigma)$ 是一个由 $\mathfrak{A}$(它是一个 $\mathbb{C}$ 上矢量空间)运算得到的 $\mathbb{C}$ 上的矢量空间，并且箭头的复合是双线性的。而且，由 $\mathfrak{A}$ 得到的 * 运算是非线性的、对合的、逆变的以及正的。

证明。完全直接得证。

命题 165。DHR 范畴 Δ 有直和。

证明。设 $\rho_1\in\Delta(O_1)$ 和 $\rho_2\in\Delta(O_2)$，选择对顶锥 O 使得 $(O_1\cup O_2)^-\subseteq O$。令 E 是一个 $\mathfrak{A}(O_1)$ 里的投影，根据性质 B，存在等距 $V_1, V_2\in\mathfrak{A}(O)$ 使得 $V_1V_1^*+V_2V_2^*=I$。定义 ρ: $\mathfrak{A}\to\mathfrak{A}$ 为：

$$\rho(A)=V_1\rho_1(A)V_1^*+V_2\rho_2(A)V_2^*,\qquad \forall A\in\mathfrak{A}$$

因为 $V_iV_j=\delta_{ij}I$ 和 $\sum_i V_iV_i^*=I$，因此 ρ 是一个态射。由于 ρ_1，ρ_2 是在 O 里局域化的，并且 V_1，$V_2\in\mathfrak{A}(O)$，因此 ρ 在 O 里是局域化的。

为了理解 ρ 是可输运的，设 $\tilde{O}$ 是另一个对顶锥。由于 ρ_i 是可输运的，存在在 $\tilde{O}$ 中局域化的自同态 ρ_i' 以及幺正算子 $U_i\in\mathrm{Hom}(\rho_i,\ \rho_i')$。像前面一样，选择 $\mathfrak{A}(\tilde{O})$ 里等距 V_1'，V_2'，并且设 $\rho'=V_1'\rho_1'V_1'^*+V_2'\rho_2'V_2'^*$，那么 ρ' 在 $\tilde{O}$ 局域化并且：

$$V_1'U_1V_1^*\in\mathrm{Hom}(\rho,\ \rho'),\qquad V_2'U_2V_2^*\in\mathrm{Hom}(\rho,\ \rho')$$

如果我们设 $W=V_1'U_1V_1^*+V_2'U_2V_2^*$，那么 $W\in\mathrm{Hom}(\rho,\ \rho')$，因为它是一个矢量空间。而且：

$$\begin{aligned}W^*W&=[V_1'U_1V_1^*+V_2'U_2V_2^*]^*[V_1'U_1V_1^*+V_2'U_2V_2^*]\\&=[V_1'U_1^*V_1'^*+V_2'U_2^*V_2'^*][V_1'U_1V_1^*+V_2'U_2V_2^*]\\&=V_1V_1^*+V_2V_2^*=I\end{aligned}$$

并且对 WW^* 也相似。因此 W 是 $\mathrm{Hom}(\rho,\ \rho')$ 里的一个幺正算符，证明 ρ 是可输运的。

定义 166。如果 ρ_1，$\rho_2\in\Delta$，那么我们用 $\rho_1\oplus\rho_2$ 代表它们的直和。

命题 167。DHR 范畴 Δ 具有子对象。

证明。设 $\rho\in\Delta(O)$，并且令 E 是一个在 $\mathrm{End}(\rho)$ 里的投影；即对所有 $A\in\mathfrak{A}$ 都有 $E\rho(A)=\rho(A)E$。那么对所有 $A\in\mathfrak{A}(O')$，有：

$$EA=E\rho(A)=\rho(A)E=AE$$

因此，根据真空截面 $E\in\mathfrak{A}(O)$ 里的对偶性，选择 O_1 使得 $O^-\subseteq O_1$。根据性质 B，存在一个等距 $V\in\mathfrak{A}(O_1)$ 使得 $VV^*=E$。现在定义 ρ'：$\mathfrak{A}\to\mathfrak{A}$ 为：

$$\rho'(A)=V^*\rho(A)V,\qquad \forall A\in\mathfrak{A}$$

等距 V 把 ρ' 嵌入到 ρ，实际上，

$$\rho'(A)V^*=V^*\rho(A)VV^*=V^*\rho(A)E=V^*\rho(A)$$

并且 V 是 $\mathrm{Hom}(\rho',\ \rho)$ 里的一个等距，使得 $VV^*=E\in\mathrm{End}(\rho)$。

为了理解 ρ' 是可输运的，假定 O_2 是一个任意对顶锥，选择一个对顶锥 O_3 使得 $O_3^-\subseteq O_2$。由于 ρ 是可输运的，存在一个在 O_3 里局域化的态射 σ 和一个幺正 $U\in\mathrm{Hom}(\rho,\ \sigma)$。由此得出 $U\mathrm{End}(\rho)U^*=\mathrm{End}(\sigma)$，因此 $E'=UEU^*$ 是一个

End(σ)里的投影。根据性质 B，存在一个等距 $V' \in \mathfrak{A}(O_1)$，使得 $V'V'^* = E'$。设 $\sigma' = V'^* \sigma V'$，显然 σ'在 O_1 中是局域化，并且 $W = V'^* UV \in \mathrm{Hom}(\rho', \sigma')$。最后，$W$ 是幺正的：

$$W^* W = V^* U^* V'V'^* UV = V^* U^* E' UV$$
$$= V^* EV = V^* VV^* V = I$$

并且对 WW^* 也相似，因此 σ'等价于 ρ'。因为 O_2 是一个任意对顶锥，$\rho' \in \Delta$。

定义 168。假设$\mathcal{C}$是一个$\mathbb{C}$线性范畴，一个$\mathcal{C}$中的对象 X 当它是非零的被认为是不可约化的，且 $\mathrm{End}(X) = \mathbb{C}\mathrm{id}_X$。

注释 169。设 ι 是$\mathfrak{A}$的自同态单位元，那么 $\iota \in \mathrm{Obj}(\Delta)$，并且由于$\mathfrak{A}$的真空表征是不可约化的，$\iota$ 是一个不可约化的对象。

我们现在定义一个在 DHR 范畴 Δ 上双函子$\otimes = (\otimes, \times)$，并且证明$(\Delta, \otimes, \iota)$是一个张量 * 范畴。但是首先我们回顾相关定义。

定义 170。一个在范畴$\mathcal{C}$上的双函子由两个映射 F: $\mathrm{Obj}(\mathcal{C}) \times \mathrm{Obj}(\mathcal{C}) \to \mathrm{Obj}(\mathcal{C})$和 F: $\mathrm{Hom}(\mathcal{C}) \times \mathrm{Hom}(\mathcal{C}) \to \mathrm{Hom}(\mathcal{C})$组成，使得对于 $s \in \mathrm{Hom}(X, Y)$和 $t \in \mathrm{Hom}(X', Y')$，有 $F(x, t) \in \mathrm{Hom}(F(X, X'), F(Y, Y'))$，以及：

$$F(s_1 \circ s_2, t) = F(s_1, t) \circ F(s_2, t)$$
$$F(s, t_1 \circ t_2) = F(s, t_1) \circ F(s, t_2)$$
$$F(\mathrm{id}_X, \mathrm{id}_{X'}) = \mathrm{id}_{F(X, X')}$$

如果$\mathcal{C}$是 * 范畴的，那么一个双函子 F 也要求是双函子并且和 * 运算对易。那就是，对于 $s_i \in \mathrm{Hom}(X, X')$、$t_i \in \mathrm{Hom}(Y, Y')$以及 $c \in \mathbb{C}$，我们有：

$$F(s_1 + s_2, t) = F(s_1, t) + F(s_2, t)$$
$$F(s, t_1 + t_2) = F(s, t_1) + F(s, t_2)$$
$$F(cs, t) = cF(s, t) = F(s, ct)$$

并且：

$$F(s, t)^* = F(s^*, t^*)$$

定义 171。设$\otimes = (\otimes, \times)$是一个在范畴$\mathcal{C}$上的双函子，并且设 $1 \in \mathrm{Obj}(\mathcal{C})$。那么$(\mathcal{C}, \otimes, 1)$被认为是一个张量范畴，当且仅当，$\otimes$是跟自然同构相关的，并且 1 是对于自然同构的一个双侧单位元，对象 1 被称为幺半群的单位元。准确说，

$\otimes$是“跟自然同构相关的”意味着，对任何 X，Y，$Z\in \mathrm{Obj}(\mathcal{C})$，存在一个对所有 X，Y，Z 是“中性”的同构 $\alpha_{X,Y,Z}$：$X\otimes(Y\otimes Z)\to(X\otimes Y)\otimes Z$，即如果 s: $X\to X'$，那么：

$$((s\otimes \mathrm{id}_Y)\otimes \mathrm{id}_Z)\circ \alpha_{X,Y,Z}=\alpha_{X',Y,Z}\circ(s\otimes(\mathrm{id}_Y\otimes \mathrm{id}_Z)) \tag{15}$$

并且对于 Y 和 Z 也一样。而且，α 要求满足五角形对易：

$$\begin{tikzcd}
 & & (X\otimes Y)\otimes(Z\otimes Z') \arrow[drr, "\alpha"] & & \\
X\otimes(Y\otimes(Z\otimes Z')) \arrow[urr, "\alpha"] \arrow[dr, "\mathrm{id}_X\otimes\alpha"'] & & & & ((X\otimes Y)\otimes Z)\otimes Z' \\
 & X\otimes((Y\otimes Z)\otimes Z') \arrow[rr, "\alpha"] & & (X\otimes(Y\otimes Z))\otimes Z' \arrow[ur, "\alpha\otimes\mathrm{id}_{Z'}"'] &
\end{tikzcd}$$

认为 $1\in\mathcal{C}$是对于自然同构的一个双侧单位元意味着，对于每一个对象 $X\in \mathrm{Obj}(\mathcal{C})$，存在同构 $\lambda_X\in \mathrm{Hom}(1\otimes X,\ X)$ 和 $\rho_X\in \mathrm{Hom}(X\otimes 1,\ X)$ 使得：

1. λ_X 和 ρ_X 在 X 中是自然的；即对任何 s: $X\to Y$，

$$s\circ\lambda_X=\lambda_Y\circ(\mathrm{id}_1\otimes s) \tag{16}$$

$$s\circ\rho_X=\rho_Y\circ(s\otimes \mathrm{id}_1) \tag{17}$$

换言之，下面两个图对易：

$$\begin{tikzcd}
1\otimes X \arrow[r, "\lambda_X"] \arrow[d, "\mathrm{id}_1\otimes s"'] & X \arrow[d, "s"] \\
1\otimes Y \arrow[r, "\lambda_Y"'] & Y
\end{tikzcd}
\qquad
\begin{tikzcd}
X\otimes 1 \arrow[r, "\rho_X"] \arrow[d, "s\otimes \mathrm{id}_1"'] & X \arrow[d, "s"] \\
Y\otimes 1 \arrow[r, "\rho_Y"'] & Y
\end{tikzcd}$$

2. λ_X 和 ρ_X 使三角形图对易：

$$\begin{tikzcd}
X\otimes(1\otimes Y) \arrow[r, "\alpha"] \arrow[dr, "\mathrm{id}_X\otimes\lambda_Y"'] & (X\otimes 1)\otimes Y \arrow[d, "\rho_X\otimes \mathrm{id}_Y"] \\
 & X\otimes Y
\end{tikzcd}$$

如果$\mathcal{C}$也是一个 * 范畴，有两个进一步的要求：(a)双函子⊗一定跟 + 和 * 的运算兼容(就像双函子的定义所要求的)；(b)幺半群的单位元 1 一定是不可归约的，即 $\mathrm{End}(1)=\mathbb{C}\mathrm{id}_1$。对于 C^* 范畴$\mathcal{C}$，我们另外要求 $\|s\times t\|_{X\otimes Y,X'\otimes Y'}\leqslant\|s\|_{X,X'}\cdot\|t\|_{Y,Y'}$。

麦克·雷(Mac Lane)的连贯性定理证明，我们完全能够忽略自然同构 α，λ 和 ρ，即是说，我们能够把 $X\otimes(Y\otimes Z)$ 和 $(X\otimes Y)\otimes Z$ 看作同样的对象，并且我们能够把 X，$1\otimes X$ 和 $X\otimes 1$ 看作同样的对象。为了更加明确，我们定义：

定义 172。如果 $\alpha_{X,Y,Z}$，λ_X，ρ_X 对于所有的 X，Y，$Z\in\mathrm{Obj}(\mathcal{C})$ 来说都是单位射，一个张量范畴$\mathcal{C}$被说成是严格的。

比如，矢量空间的张量范畴(Vect，⊗，ℂ)是不严格的，因为比如 $V\otimes\mathbb{C}$ 是不完全跟 V 一样的相同矢量空间。另一方面，可对易幺半群 M 能够被看成具有一个对象和对应于 M 元素箭头的严格张量范畴。连贯性定理还能够按下面表述。

定理 173(连贯性定理)。任何张量范畴等价于严格张量范畴。

证明。参见[Mac Lane，1998]。

定义 174。如果$\mathcal{C}$是一个张量范畴，那么我们设 $\mathcal{C}^{st}$ 代表它的严格性。

根据所做的这些定义，我们现在开始定义一个 Δ 上的双函子，并且证明它满足全部相关性质。我们在 Δ 里的对象的积⊗正好是自同态的合成。

命题 175。如果 ρ，$\sigma\in\mathrm{Obj}(\Delta)$，那么 $\rho\sigma\in\mathrm{Obj}(\Delta)$。

证明。显然，如果 ρ 在 O_1 中是局域化的，并且 σ 在 O_2 中局域化，那么 $\rho\sigma$ 是在包含 $O_1\cup O_2$ 的任何对顶锥中局域化的。

为了理解 ρ 和 σ 是可输运的，设 O_3 是一个任意对顶锥，由于 ρ 和 σ 是可输运的，存在 ρ'，$\sigma'\in\Delta(O_3)$ 和酉 $U\in\mathrm{Hom}(\rho,\rho')$ 以及 $V\in\mathrm{Hom}(\sigma,\sigma')$，那么 $\rho'\sigma'$ 在 O_3 中是局域化的，并且 $U\rho(V)$ 是一个在 $\mathrm{Hom}(\rho\sigma,\rho'\sigma')$ 中幺正的。因此，$\rho\sigma$ 是可输运的。

定义 176。通过 $\rho\otimes\sigma=\rho\sigma$，定义⊗：$\mathrm{Obj}(\Delta)\times\mathrm{Obj}(\Delta)\to\mathrm{Obj}(\Delta)$

箭头的积 × 稍为复杂些。

命题 177。如果 $S\in\mathrm{Hom}(\rho,\rho')$ 并且 $T\in\mathrm{Hom}(\sigma,\sigma')$，那么：

$$S\rho(T)\in\mathrm{Hom}(\rho\otimes\sigma,\rho'\otimes\sigma')$$

证明。由于 $S\rho(T)=\rho'(T)S$，因此对于任何 $A\in\mathfrak{A}$，有：

$$(S\rho(T))\rho\sigma(A)=S\rho(T\sigma(A))=\rho'(T\sigma(A))S=\rho'(\sigma'(A)T)S$$
$$=\rho'\sigma'(A)(\rho'(T)S)=\rho'\sigma'(A)(S\rho(T))$$

因此 $S\rho(T)\in \mathrm{Hom}(\rho\sigma,\ \rho'\sigma')$。

定义 178。 对于 $S\in\mathrm{Hom}(\rho,\ \rho')$ 和 $T\in\mathrm{Hom}(\sigma,\ \sigma')$，我们设 $S\times T=S\rho(T)\in\mathrm{Hom}(\rho\otimes\sigma,\ \rho'\otimes\sigma')$。定义 $\times$: $\mathrm{Hom}(\Delta)\times\mathrm{Hom}(\Delta)\rightarrow\mathrm{Hom}(\Delta)$。

在这一节剩下的部分，我们证明(Δ, $\otimes$, ι)是张量 * 范畴。

$\otimes$是在 Δ 上的双函子。

命题 179。 对于 S_1，S_2，T_1，$T_2\in\mathrm{Obj}(\Delta)$，如果 T_i 的源是 S_i 的目标(从而 $T_i\circ S_i$ 被定义)，那么：

$$(T_1\times T_2)\circ(S_1\times S_2)=(T_1\circ S_1)\times(T_2\circ S_2)$$

证明。直接运算得证。

我们现在必须检查 $\times$ 跟 * 是兼容的。

命题 180。 对于所有的 S，$T\in\mathrm{Hom}(\Delta)$，有：

$$(S\times T)^*=S^*\times T^*$$

证明。直接运算得证。

ι 是一个幺半群单位元。

对于每一个 $\rho\in\mathrm{Obj}(\Delta)$，$\iota\otimes\rho$ 一定是自然同构于 $\rho\otimes\iota$ 和 ρ(正如在幺半群单位图中描述的)。但是在目前的情况下，$\iota\otimes\rho=\rho\otimes\iota=\rho$，因此这个自然同构平凡地保持。

$\otimes$的自然结合性。

下一步，积运算$\otimes=(\otimes,\ \times)$一定对自然同构是结合的，正如在五角图形中所描述的。但是这在目前情况下是平凡的，因为结合性严格保持，那就是：

命题 181。 对于所有 ρ_1，ρ_2，$\rho_3\in\mathrm{Obj}(\Delta)$，有：

$$\rho_1\otimes(\rho_2\otimes\rho_3)=(\rho_1\otimes\rho_2)\otimes\rho_3$$

并且对于所有 T_1，T_2，$T_3\in\mathrm{Hom}(\Delta)$，有：

$$T_1\times(T_2\times T_3)=(T_1\times T_2)\times T_3$$

证明。第一个断言平凡地源自于自同态的复合是结合的这个事实，第二个断言能够通过直接运算得到证明。

引理 182。$(\Delta, \otimes, \iota)$是具有从$\mathfrak{A}$保留下来的范数的 C^* 张量范畴。

证明。我们必须证明 $\mathrm{Hom}(\rho, \sigma)$在$\mathfrak{A}$上的范数里是封闭的，而这直接源自于下面这个事实：

$$\mathrm{Hom}(\rho, \sigma) = \{T \in \mathfrak{A}: T\rho(A) = \sigma(A)T, \ \forall A \in \mathfrak{A}\}$$

显然 $\| s \circ t \| \leqslant \| s \| \ \| t \|$，而且：

$$\| S \times T \| = \| S \ \rho(T) \| \leqslant \| S \| \cdot \| \rho(T) \| \leqslant \| S \| \cdot \| T \|$$

为了这一点，我们已经证明：(i) Δ 是一个 C^* 范畴；(ii) Δ 是一个张量 * 范畴。下面五个小节并没有一个严格顺序。8.1 节表明如何定义$(\Delta, \otimes, \iota)$上的正则辫子 $\varepsilon_{\rho_1, \rho_2}$，使得它是一个“辫子”张量 * 范畴。然后在 8.2 节中我们证明了我们关于研究范畴 Δ 动机的说法：我们证明在 Δ 跟满足 DHR 选择标准的表征范畴之间的函子对应。然后我们挑出有关张量 * 范畴的一些技术性信息，它们对于相应表征的物理解释是本质性的。在 8.3 节中我们看到如何定义一个在张量 * 范畴里的对象的“维度”概念，并且我们定义了“共轭”对象的概念。在 8.4 节中我们回过来讨论时空对称跟 DHR 表象的关系。最后，在 8.5 节中我们给出了 Δ_f 的对象的内在统计分类，这对应于在玻色子和费米子或者玻色场跟费米场之间的直观区分。

8.1 Δ 是一个辫子张量 * 范畴

在这一节中我们于 Δ 上定义正则辫子；这使我们把握住如果我们改变积中的顺序所发生的一切，比如说$\rho \otimes \sigma$ 跟 $\sigma \otimes \rho$ 相比。我们也将看到在时空维度和这个辫子性质之间的重要关联：如果时空有三维以上，那么辫子是对称的。我们首先回顾一下重要定义。

定义 183。如果$(\mathcal{C}, \otimes, 1)$是张量范畴，那么一个$\mathcal{C}$上辫子是一个同构的族

$$\{C_{X,Y} \in \mathrm{Hom}(X \otimes Y, Y \otimes X): X, Y \in \mathrm{Obj}(\mathcal{C})\}$$

满足下面两个条件：

1. $c_{X,Y}$在 X 和 Y 中是自然的；即对任何 $f \in \mathrm{Hom}(X, X')$ 和 $g \in \mathrm{Hom}(Y, Y')$，有：

$$(g \times f) \circ c_{X,Y} = c_{X',Y'} \circ (f \times g) \tag{18}$$

2. $c_{X,Y}$使得下面两个六角形对易：

$$\begin{tikzcd}
(X\otimes Y)\otimes Z \arrow[r, "c_{X\otimes Y,Z}"] \arrow[d, "\alpha^{-1}"'] & Z\otimes(X\otimes Y) \arrow[d, "\alpha"] \\
X\otimes(Y\otimes Z) \arrow[d, "\mathrm{id}_X\otimes c_{Y,Z}"'] & (Z\otimes X)\otimes Y \arrow[d, "c_{Z,X}\otimes \mathrm{id}_Y"] \\
X\otimes(Z\otimes Y) \arrow[r, "\alpha"'] & (X\otimes Z)\otimes Y
\end{tikzcd}$$

$$\begin{tikzcd}
X\otimes(Y\otimes Z) \arrow[r, "c_{X,Y\otimes Z}"] \arrow[d, "\alpha"'] & (Y\otimes Z)\otimes X \arrow[d, "\alpha^{-1}"] \\
(X\otimes Y)\otimes Z \arrow[d, "c_{X,Y}\otimes \mathrm{id}_Z"'] & Y\otimes(Z\otimes X) \arrow[d, "\mathrm{id}_Y\otimes c_{Z,X}"] \\
(Y\otimes X)\otimes Z \arrow[r, "\alpha^{-1}"'] & Y\otimes(X\otimes Z)
\end{tikzcd}$$

那就是要阻止可结合性同构，$c_{X\otimes Y,Z}$通过 $c_{X,Y}$和 $c_{X,Z}$被表示成：

$$c_{X\otimes Y,Z}=(\mathrm{id}_Y\otimes c_{Z,X})^{-1}\circ(\mathrm{id}_X\otimes c_{Y,Z})$$

并且 $C_{X,Y\otimes Z}$通过 $c_{X,Y}$和 $c_{Z,X}$被表示成：

$$c_{X,Y\otimes Z}=(\mathrm{id}_Y\otimes c_{Z,X})^{-1}\circ(c_{X,Y}\otimes \mathrm{id}_Z)$$

定义 184。对于所有 X，$Y\in\mathrm{Obj}(\mathcal{C})$如果$(c_{X,Y})^{-1}=c_{Y,X}$，辫子 $c_{X,Y}$被称为对称的。

定义 185。一个具有特殊辫子（对称性）的张量范畴被称为一个辫子（对称的）张量范畴。

为了在 Δ 上找到我们的辫子，我们需要下面的技术性引理。

引理 186。如果$\rho\in\Delta(O_1)$和$\sigma\in\Delta(O_2)$，其中 O_1 和 O_2 是类空分离的，那么$\rho\sigma=\sigma\rho$。

证明。由于$\{\mathfrak{A}(O)$: $O_1\cup O_2\subseteq O\}$的并在$\mathfrak{A}$中是稠的，它足以说明$\rho\sigma(A)=\sigma\rho(A)$，每当$A\in\mathfrak{A}(O)$具有 $O_1\cup O_2\subseteq O$ 的时候。选定对 O 是类空的 O_3，O_4，并且使得 $O_1\cup O_3$ 对于 $O_2\cup O_4$ 是类空的。（这总是可以做到的，即使是在二维

时空中。)由于 ρ′，σ 是可输运的，存在分别在 O_3，O_4 中局域化的 ρ'，σ'，以及幺正算子$U_1 \in \mathrm{Hom}(\rho, \rho')$和 $U_2 \in \mathrm{Hom}(\sigma, \sigma')$。那么：

$$\sigma(A) = U_2\sigma'(A)U_2^* = U_2^* A U_2$$

而且，根据在真空截面里的对偶性 $U_2 \in \mathfrak{A}(O'_1)$。因此$\rho(U_2) = U_2$，并且：

$$\rho\sigma(A) = U_2 U_1 A U_1^* U_2^*$$

由于 $U_2U_1 = U_1U_2$，因此$\rho\sigma(A) = \sigma\rho(A)$。

我们不能够以简短方式定义辫子 $\varepsilon_{\rho_1,\rho_2}$。相反，我们先定义箭头：

$$\varepsilon_{\rho_1,\rho_2}(U_1, U_2) \in \mathrm{Hom}(\rho_1 \otimes \rho_2, \rho_2 \otimes \rho_1)$$

这依赖于"观测者态射"$\tilde{\rho}_1$，$\tilde{\rho}_2$ 的选择，和幺正扭结子 $U_i \in \mathrm{Hom}(\rho_i, \tilde{\rho}_i)$。然后我们会证明这个定义独立于观测者态射和幺正扭结子。(但是，有趣的是当时空是二维的时候，这个定义依赖于空间方向的选择。)

定义 187。假定$\rho_1 \in \Delta(O_1)$和$\rho_2 \in \Delta(O_2)$。设 $\tilde{O}_1$ 和 $\tilde{O}_2$ 是类空分离的对顶锥，由于ρ_1 和ρ_2 是可输运的，存在$\tilde{\rho}_i \in \Delta(\tilde{O}_i)$和酉算子 $U_i \in \mathrm{Hom}(\rho_1, \tilde{\rho}_i)$。因此 $U_1 \times U_2 \in \mathrm{Hom}(\rho_1 \otimes \rho_2, \tilde{\rho}_1 \otimes \tilde{\rho}_2)$和 $U_2^* \times U_1^* \in \mathrm{Hom}(\tilde{\rho}_2 \otimes \tilde{\rho}_1, \rho_2 \otimes \rho_1)$。由于$\tilde{O}_1$ 类空于 $\tilde{O}_2$，引理 186 意味着$\tilde{\rho}_1 \otimes \tilde{\rho}_2 = \tilde{\rho}_2 \otimes \tilde{\rho}_1$，因此，我们可以定义$\varepsilon_{\rho_1,\rho_2}(U_1, U_2) \in \mathrm{Hom} \in (\rho_1 \otimes \rho_2, \rho_2 \otimes \rho_1)$为：

$$\varepsilon_{\rho_1,\rho_2}(U_1, U_2) := (U_2 \times U_1)^* \circ (U_1 \times U_2) = \rho_2(U_1^*)U_2^* U_1 \rho_1(U_2) \qquad (19)$$

注释 188。由于自同态保证了幺正性，$\varepsilon_{\rho_1,\rho_2}(U_1, U_2)$是幺正的。

为了证明 $\varepsilon_{\rho_1,\rho_2}(U_1, U_2)$是独立于 U_1 和 U_2，我们需要下述引理，它表明 $\varepsilon_{\rho_1,\rho_2}(U_1, U_2)$在某些 U_1 和 U_2 的"干扰"下不发生改变。

引理 189。对于 $i = 1, 2$，设$\rho_i \in \Delta(O_i)$，假设 $\tilde{O}_1$ 和 $\tilde{O}_2$ 是类空分离的，假设$\tilde{\rho}_i \in \Delta(\tilde{O}_i)$以及假设 $U_i \in \mathrm{Hom}(\rho_i, \tilde{\rho}_i)$。那么 $\varepsilon_{\rho_1,\rho_2}(U_1, U_2)$只是在下述意义上依赖于 U_1，U_2 的邻域：如果 W_1，W_2 是幺正的，使得 $W_1 \in \mathfrak{A}(\tilde{O}_2')$，$W_2 \in \mathfrak{A}(\tilde{O}_1')$，并且 $W_1W_2 = W_2W_1$，那么：

$$\varepsilon_{\rho_1,\rho_2}(W_1U_1, W_2U_2) = \varepsilon_{\rho_1,\rho_2}(U_1, U_2)$$

证明。我们必须证明：

$$(W_2U_2 \times W_1U_1)^* \circ (W_1U_1 \times W_2U_2) = (U_2^* \times U_1^*) \circ (U_1 \times U_2) \qquad (20)$$

对于任何两个幺正算子，W_1，$W_2 \in \mathfrak{A}$，我们有：

$$W_iU_i \times W_jU_j = W_iU_i\rho_i(W_j)\rho_i(U_j) = W_i\rho_i'(W_j)(U_i \times U_j)$$

由于 $W_1 \in \mathfrak{A}(\tilde{O}_2')$ 和 $\tilde{\rho}_2$ 是在 $\tilde{O}_2$ 局域化，$\tilde{\rho}_2(W_1) = W_1$，并且类似地 $\tilde{\rho}_1(W_2) = W_2$。因此，方程(20)左边变成：

$$[(U_2 \times U_1)^* \tilde{\rho}_2(W_1^*) W_2^*][W_1\tilde{\rho}_1(W_2)(U_1 \times U_2)] = (U_2 \times U_1)^* W_1^* W_2^* W_1 W_2(U_1 \times U_2) = (U_2 \times U_1)^*(U_1 \times U_2)$$

其中我们在第二个等式用到 $W_1W_2 = W_2W_1$。

引理 190。设 $\rho_i \in \Delta(O_i)$，$i=1$，2，并且设 $T \in \mathrm{Hom}(\rho_1, p_2)$。那么对于包含 $O_1 \cup O_2$ 的任何对顶锥有 $T \in \mathfrak{A}(O)$。

证明。设 O 是包含 $O_1 \cup O_2$ 的一个对顶锥，并且设 $A \in \mathfrak{A}(O')$，那么 $\rho_1(A) = \rho_2(A) = A$，从而有：

$$TA = T\rho_1(A) = \rho_2(A)T = AT$$

因此 $T \in \mathfrak{A}(O')'$，并且由于真空截面里的对偶性有 $T \in \mathfrak{A}(O)$。

现在我们能够证明 $\varepsilon_{\rho_1,\rho_2}(U_1, U_2)$ 是只依赖于观测者态射的局域化区域。

命题 191。$\varepsilon_{\rho_1,\rho_2}(U_1, U_2)$ 是可以用 ρ_1，ρ_2 和区域 $\tilde{O}_1$，$\tilde{O}_2$ 的方式定义的，并且如果后者用对顶锥 $\tilde{\tilde{O}}_1$，$\tilde{\tilde{O}}_2$ 替代，使得 $\tilde{O}_1 \subseteq \tilde{\tilde{O}}_1$ 和 $\tilde{O}_2 \subseteq \tilde{\tilde{O}}_2$，那么情况完全不变。

证明。(1) 我们首先证明对于一个给定的类空分离的对顶锥的 $(\tilde{O}_1, \tilde{O}_2)$ 对，定义 $\varepsilon_{\rho_1,\rho_2}(U_1, U_2)$ 独立于观测者态射 $(\tilde{\rho}_1, \tilde{\rho}_2)$ 和幺正扭结子 (U_1, U_2)。因此，假设 $\tilde{\tilde{\rho}} \in \Delta(\tilde{O}_i)$ 和 $U_i' \in \mathrm{Hom}(\rho_i, \tilde{\tilde{\rho}}_i)$。假定 $W_i = U_i'U_i^* \in \mathrm{Hom}(\tilde{\rho}_i, \tilde{\tilde{\rho}}_i)$，使得 $U_i' = W_iU_i$。由于 W_i 在 $\tilde{O}_i$ 里有左右支集，$W_i \in \mathfrak{A}(\tilde{O}_i) \subseteq \mathfrak{A}(\tilde{O}_j)'$。因此，$W_1W_2 = W_2W_1$，并且引理 189 的假设被满足。所以 $\varepsilon_{\rho_1,\rho_2}(U_1, U_2) = \varepsilon_{\rho_1,\rho_2}(U_1', U_2')$。

(2) 现在设 $\tilde{\tilde{O}}_1$ 和 $\tilde{\tilde{O}}_2$ 是对顶锥，对 $i=1$，2 使得 $\tilde{\tilde{O}}_1 \perp \tilde{\tilde{O}}_2$，并且 $\tilde{O}_i \subseteq \tilde{\tilde{O}}_i$。选择 $\tilde{\tilde{\rho}} \in \Delta(\tilde{\tilde{O}}_i)$，和幺正 $U'_i \in \mathrm{Hom}(\rho_i, \tilde{\tilde{\rho}}_i)$，但是我们也有 $\tilde{\rho}_i \in \Delta(\tilde{O}_i) \in \Delta(\tilde{\tilde{O}}_i)$。并且证明的第一部分表明，对于固定支集领域 $(\tilde{\tilde{O}}_1, \tilde{\tilde{O}}_2)$，$\varepsilon_{\rho_1,\rho_2}$ 的定义是独立于观测者态射和幺正扭结子的选择。因此 $\varepsilon_{\rho_1,\rho_2}(U_1, U_2) = \varepsilon_{\rho_1,\rho_2}(U_1', U_2')$。

注释 192。我们总能选择观测者态射在严格类空分离区域里是局域化的。

确实，给定类空分离的 $\tilde{O}_1$，$\tilde{O}_2$，选择 $\tilde{\tilde{O}}_1$ 使得 $(\tilde{\tilde{O}}_1)^- \subseteq \tilde{O}_1$。然而该引理意味着（通过缝合 $\tilde{\tilde{O}}_1$ 和 $\tilde{O}_1$，并且设置 $\tilde{\tilde{O}}_2 = \tilde{O}_2$）我们从使用 $\tilde{\tilde{O}}_1$ 或者 $\tilde{O}_1$ 中获得同样的定义。更一般地，由于 ρ_1 是可输运的，区域 $\tilde{O}_i$ 能够被选择得任意小。

注释 193。上述注释说明一个 $\varepsilon_{\rho_1,\rho_2}(U_1, U_2)$ 的定义，总是等价于一个使用在严格类空分离区域里局域化的观测者态射的定义。那就是说，存在一个零的邻域 N，使得对于所有 $x \in N$ 都有 $\tilde{O}_1 + x \subseteq \tilde{O}_2'$。再者，由于 $\tilde{O}_1$ 和 $\tilde{O}_1 + x$ 被包含在对顶锥 $\tilde{\tilde{O}}_1 \subseteq \tilde{O}_2'$ 里，上述引理（用过两次）意味着 $(\tilde{O}_1, \tilde{O}_2)$ 对和 $(\tilde{O}_1 + x, \tilde{O}_2)$ 产生相同的 $\varepsilon_{\rho_1,\rho_2}$ 定义。

如果需要，也可以收缩 $\tilde{O}_2$，并且重复上述建构，从而我们看到对于任何矢量 x，$(\tilde{O}_1 + x, \tilde{O}_2 + x)$ 和 $(\tilde{O}_1, \tilde{O}_2)$ 产生相同的 $\varepsilon_{\rho_1,\rho_2}$ 定义。

注释 194。在接下来的内容里，对“一维时空”，我们指的是距零维时间的一维空间。在此情况下，一个对顶锥仅仅涉及一个开间隔，并且“类空分离”意味着不相交。

命题 195。对于维度为二或者更少的时空，$\varepsilon_{\rho_1,\rho_2}(U_1, U_2)$ 通过 ρ_1，ρ_2 以及 $\tilde{O}_1$ 相对 $\tilde{O}_2$ 的空间方向是可定义的。那就是说，$\varepsilon_{\rho_1,\rho_2}(U_1, U_2)$ 独立于 $(\tilde{O}_1, \tilde{O}_2)$ 的选择，遵守具有同样空间指向的约束。

证明。给定 $\tilde{O}_i$，$\tilde{\tilde{O}}_i$ 使得 $\tilde{O}_1 \perp \tilde{O}_2$，$\tilde{\tilde{O}}_1 \perp \tilde{\tilde{O}}_2$，并且 $\tilde{O}_1$ 对 $\tilde{O}_2$ 的取向就像 $\tilde{\tilde{O}}_1$ 对 $\tilde{\tilde{O}}_2$ 的取向。回顾 $(\tilde{O}_1, \tilde{O}_2)$ 的变换并不能改变 $\varepsilon_{\rho_1,\rho_2}(U_1, U_2)$；用一个对顶锥取代 $\tilde{O}_i$ 也不能改变，不管是包含它还是被包含在它里面，以及类空于 $\tilde{O}_j$。但是 $(\tilde{O}_1, \tilde{O}_2)$ 能够在一系列这样的移动中被 $(\tilde{\tilde{O}}_1, \tilde{\tilde{O}}_2)$ 所取代。

定义 196。对二维以下的时空，固定一个空间方向，并且用 $O_1 < O_2$ 决定了 O_1 处于 O_2 左边。

引理 197。对于二维以下的时空，如果 $\tilde{O}_1$ 相对 $\tilde{O}_2$ 的空间方向跟 $\tilde{\tilde{O}}_1$ 相对于 $\tilde{\tilde{O}}_2$ 的空间方向是相反的，那么：

$$\varepsilon_{\rho_1,\rho_2}(U_1, U_2) = [\varepsilon_{\rho_1,\rho_2}(U_2', U_1')]^*$$

证明。为了定义 $\varepsilon_{\rho_1,\rho_2}(U_1, U_2)$，我们可以选择 $\tilde{O}_1 = O_1$，$\tilde{O}_2 < O_1$，$\tilde{\rho}_1 = \rho_1$，

以及$U_1 = I_{\rho_1} = I \in \mathrm{Hom}(\rho_1, \rho_2)$。在此情况下，定义可以简化为：

$$\varepsilon_{\rho_1,\rho_2}(I, U_2) = U_2^* \rho_1(U_2) \qquad (\tilde{O}_2 < \tilde{O}_1)$$

使用同样的观测者态射，我们有：

$$\varepsilon_{\rho_2,\rho_1}(U_2, I) = \rho_1(U_2^*) U_2 \qquad (\tilde{O}_2 < \tilde{O}_1)$$

这后一个表述使用了相反的空间方向。由 $\varepsilon_{\rho_1,\rho_2}(U_1, U_2)$按空间方向(命题 195)的定义，我们可知 $\varepsilon_{\rho_1,\rho_2}(U_1, U_2) = [\varepsilon_{\rho_2,\rho_1}(U_2', U_1')]^*$，当对这两个定义采用相反的空间方向时。

定义 198(Δ 上的正则辫子)。对于二维以下的时空，我们得到具有 $\tilde{O}_2 < \tilde{O}_1$ 时 $\varepsilon_{\rho_1,\rho_2} = \varepsilon_{\rho_1,\rho_2}(U_1, U_2)$的约定。前面的引理表明，如果我们按相反的约定定义 $\bar{\varepsilon}_{\rho_1,\rho_2}$，那么 $\bar{\varepsilon}_{\rho_1,\rho_2} = (\varepsilon_{\rho_2,\rho_1})^*$。对于超过三维的时空，我们用类空分离的 $\tilde{O}_1$ 和 $\tilde{O}_2$ 定义 $\varepsilon_{\rho_1,\rho_2} = \varepsilon_{\rho_1,\rho_2}(U_1, U_2)$。

我们现在证明 $\varepsilon_{\rho_1,\rho_2}$是一个$(\Delta, \otimes, \iota)$上的辫子。

命题 199。$\varepsilon_{\rho,\sigma}$是一个在 DHR 范畴$(\Delta, \otimes, \iota)$上的辫子。

证明。

(1)我们首先证明 $\varepsilon_{\rho,\sigma}$在 ρ 和 σ 里是自然的。因为这一点，它足以证明如果 $T \in \mathrm{Hom}(\rho, \rho')$，那么：

$$(I_\sigma \times T) \circ \varepsilon_{\rho,\sigma} = \varepsilon_{\rho',\sigma} \circ (T \times I_\sigma) \tag{21}$$

$$\varepsilon_{\rho,\sigma} \circ (I_\sigma \times T^*) = (T^* \times I_\sigma) \circ \varepsilon_{\sigma,\rho'} \tag{22}$$

设 O_1，O_2，O_3 是对顶锥，使得$\rho \in \Delta(O_1)$，$\sigma \in \Delta(O_2)$以及$\rho' \in \Delta(O_3)$。选择对 O_i，$i = 1, 2, 3$ 是类空的对顶锥 O_4；如果时空维度小于三，选择 O_4 为所有三个的左边。选择$\sigma' \in \Delta(O_4)$和 $U \in \mathrm{Hom}(\sigma, \sigma')$，那么 $\varepsilon(\rho, \sigma) = U^* \rho(U)$ 和 $\varepsilon(\rho', \sigma) = U^* \rho'(U)$。由于 $T \in \mathrm{Hom}(\rho, \rho') \subseteq \mathfrak{A}(O_4')$，则 $\sigma'(T) = T$。因此：

$$\sigma(T) U^* \rho(U) = U^* \sigma'(T) \rho(U) = U^* T \rho(U) = U^* \rho'(U) T$$

这建立了方程 21，第二个方程可以通过一个类似的运算得以建立。

(2)现在我们证明 $\varepsilon_{\rho,\sigma}$得到六角形对易，由于 Δ 是严格单项的，我们能够删除可积性自同态。那就是，它足以证明：

$$\varepsilon_{\rho\otimes\sigma,\tau}=(\varepsilon_{\rho,\tau}\times I_{\sigma})\circ(I_{\rho}\times\varepsilon_{\sigma,\tau}) \tag{23}$$

$$\varepsilon_{\rho,\sigma\otimes\tau}=(I_{\sigma}\times\varepsilon_{\rho,\tau})\circ(\varepsilon_{\rho,\sigma}\times I_{\tau}) \tag{24}$$

选择$\tau'\in\Delta$使得τ'在一个区域被支持，该区域类空于ρ，σ，τ的支集领域；对一维或者二维的时空，选择τ'的支集领域左边。设$U\in\mathrm{Hom}(\tau,\ \tau')$，那么$\varepsilon_{\rho,\tau}=U^{*}\rho(U)$，$\varepsilon_{\sigma,\tau}=U^{*}\sigma(U)$以及$\varepsilon_{\rho\otimes\sigma,\tau}=U^{*}\rho\sigma(U)$。而且，

$$U^{*}\rho(U)\rho[U^{*}\sigma(U)]=U^{*}\rho[UU^{*}\sigma(U)]=U^{*}\rho\sigma(U)$$

建立起方程(23)。下一个方程也可以类似地证明。

命题200。对于二维以下的时空，$\varepsilon_{\rho_1,\rho_2}$是$(\Delta,\ \otimes,\ \iota)$上唯一的辫子，使得当$\rho_i\in\Delta(O_i)$和$O_2<O_1$的时候，$\varepsilon_{\rho_1,\rho_2}=I$。对于三维以上时空，$\varepsilon_{\rho_1,\rho_2}$是$(\Delta,\ \otimes,\ \iota)$上唯一的辫子，使得，一旦$\rho_i\in\Delta(O_i)$和$O_1$跟$O_2$类空分离的时候，$\varepsilon_{\rho_1,\rho_2}=I$。

证明。选择$\tilde{O}_2$处于O_1左边，我们能够设$\varepsilon_{\rho_1,\rho_2}=\varepsilon_{\rho_1,\rho_2}(I,\ U_2)=U_2^{*}\rho_1(U_2)$，其中$U_2\in\mathrm{Hom}(\rho_1,\ \rho_2')$。现在设$c_{\sigma_1,\sigma_2}$是$(\Delta,\ \otimes,\ \iota)$上另一个辫子，使得只要$\sigma_1$是局域化的，在$\sigma_2$的局域化区域的右边就有$c_{\sigma_1,\sigma_2}=I$。那么由于$c_{\rho_1,\rho_2}$在$\rho_1$和$\rho_2$中是自然的，并且$c_{\rho_1,\rho_2'}=I$，则：

$$c_{\rho_1,\rho_2}=(U_2^{*}\times I_{\rho_1})\circ c_{\rho_1,\rho_2'}\circ(I_{\rho_1}\times U_2)=U_2^{*}\rho_1(U_2)=\varepsilon_{\rho_1,\rho_2}$$

这个高维情况的证明在结构上是完全相同的。

命题201。对于三维以上的时空，$\varepsilon_{\rho_1,\rho_2}=(\varepsilon_{\rho_2,\rho_1})^{-1}$，因此$\varepsilon_{\rho_1,\rho_2}$是一个在$(\Delta,\ \otimes,\ \iota)$上的对称。

证明。我们首先证明$\varepsilon_{\rho_1,\rho_2}(U_1,\ U_2)$是独立于支集观测者态射的选择$(\tilde{O}_1,\ \tilde{O}_2)$。(比较命题195的证明。)由于时空至少具有三维，存在一个对顶锥的序列O_i，$i=1$，…，n，使得：$O_1=\tilde{O}_2$，对于每一个i，$O_i\cup O_{i+1}$是包含在一个类空于$\tilde{O}_1$的对顶锥里，并且O_n具有和$\tilde{O}_1$的空间相反的方向，这一点跟$\tilde{O}_2$一样。反复应用命题191，我们得出结论：$\varepsilon_{\rho_1,\rho_2}(U_1,\ U_2)=\varepsilon_{\rho_1,\rho_2}(U_1,\ U_2')$，其中$\tilde{\bar{\rho}}_1\in\Delta(O_n)$和$U_2'\in\mathrm{Hom}(\rho_2,\ \tilde{\bar{\rho}}_2)$。因此，$\varepsilon_{\rho_1,\rho_2}$并不依赖于$\tilde{O}_1$和$\tilde{O}_2$的相对空间方向。命题195说明，$\varepsilon_{\rho_1,\rho_2}(U_1,\ U_2)$能够仅仅通过它们的相对空间方向依赖于$(\tilde{O}_1,\ \tilde{O}_2)$。因此，$\varepsilon_{\rho_1,\rho_2}(U_1,\ U_2)$独立于$(\tilde{O}_1,\ \tilde{O}_2)$。

我们能够选择$\tilde{O}_1=O_1$，$\tilde{O}_2\perp O_1$，$\tilde{\rho}=\rho_1$以及$U_1=I_{\rho_1}=I\in\mathrm{Hom}(\rho_1,\ \rho_1)$，

使得：

$$\varepsilon_{\rho_1,\rho_2} = U^* \rho_1 (U_2)$$

但是从$(\tilde{O}_1, \tilde{O}_2)$的方向给定$\varepsilon_{\rho_1,\rho_2}$的独立性，我们也有：

$$\varepsilon_{\rho_2,\rho_1} = \rho_1 (U_2)^* U_2 = (\varepsilon_{\rho_1,\rho_2})^*$$

由于$\varepsilon_{\rho_1,\rho_2}$是幺正的，$\varepsilon_{\rho_1,\rho_2} = (\varepsilon_{\rho_1,\rho_2})^{-1}$。

注释 202。前面的命题是我们首次借助基础时空的维度的地方。一旦随后的结果依赖于维度的假定，我们就会清楚情况。

定义 203。假设$\varepsilon_\rho := \varepsilon(\rho, \rho) \in \mathrm{End}(\rho \otimes \rho)$。

8.2 局域化自同态跟表征之间的关系

在这一节定义的范畴Δ和Δ_f具有重要性质，它们的物理和哲学相关性确实不明显。因此我们把范畴Δ联系到某个网$\mathfrak{A}$的表征的范畴：

定义 204。设$O \mapsto \mathfrak{A}(O)$是一个可观测量的网，而$\pi_0$：$\mathfrak{A} \to B(H_0)$是一个真空表象。那么一个 DHR 表象（关于真空表象π_0），是一个 * 表象π：$\mathfrak{A} \to B(H)$，使得对于任何对顶锥O都有$\pi \mid \mathfrak{A}(O') \cong \pi_0 \mid \mathfrak{A}(O')$。换言之，限制在$\mathfrak{A}(O')$上，表象$\pi$和$\pi_0$是幺正等价的。其对象是具有有界扭结子算子的$\mathfrak{A}$的 DHR 表象的范畴，是用 DHR$(\mathfrak{A})$表示的，它显然是一个$C^*$范畴。

定义 205。设$\mathfrak{A}$是一个关于正能量表征U_0：$\mathcal{P} \to \mathcal{U}(H_0)$的庞加莱协变网。$\mathfrak{A}$的一个表象$(H, \pi)$被称为协变的（具有正能量），如果它具有一个庞加莱群的普遍覆盖的强连续幺正表征U_π：$\hat{\mathcal{P}} \to \mathcal{U}(H_0)$（具有 spec$P^\mu \subset V_+$），使得对于所有$h \in \hat{\mathcal{P}}$都有$\mathrm{Ad}U_\pi(h) \circ \pi = \pi \circ \alpha_h$，其中从记号$\alpha_h = \mathrm{Ad}U_0(h)$中略去了覆盖映射$\hat{\mathcal{P}} \to \mathcal{P}$。

注意这个定义蕴含着 DHR 表征的表征空间H，必定跟真空希尔伯特空间H_0具有同样的维度。

命题 206。存在C^*范畴的函子F：$\Delta \to DHR(\mathfrak{A})$，使得对于对象有$F(\rho) = \pi_0 \circ \rho$，以及对于态射的$s \in \mathrm{Hom}_\Delta(\rho, \sigma)$有$F(s) = \pi_0(s)$。这个函子是一个等价式。

证明。首先我们注意到这些定义有意义：$\rho \in \mathrm{Obj}(\Delta)$把$\mathfrak{A}$映射到自身，并因此能够由表象$\pi_0$构成，从而定义一个新的表象。而且，如果$S$是$\Delta$里的一

个箭头，那么引理 155 给出 $S \in \mathfrak{A}$，因此，$F(S) = \pi_0(S)$ 有意义。依据 $S \in \mathrm{Hom}_\Delta(\rho, \rho')$，我们有：

$$F(S)F(\rho)(A) = \pi_0(S)\pi_0(\rho(A)) = \pi_0(S\rho(A)) =$$
$$\pi_0(\rho'(A)S) = F(\rho')(A)F(S) \qquad \forall A \in \mathfrak{A}$$

因此 $F(S) \in \mathrm{Hom}(F(\rho), F(\rho'))$。由于 id_ρ 是 $\mathfrak{A}$ 的单位量，我们具有 $F(\mathrm{id}_\rho) = I_{H_0} = \mathrm{id}_{F(\rho)}$。性质 $F(S \circ T) = F(S) \circ F(T)$ 是明显的。由于 π_0 是忠实的，F 也是忠实的。我们必须证明表象 $F(\rho) = \pi_0 \circ \rho$ 满足 DHR 标准。由于 $\rho \in \Delta$ 是可输运的，对于任何对顶锥 O，都存在 O 中局域化的 $\rho_O \in \Delta$ 和一个酉 U_O: $\rho \to \rho_O$。由于 ρ_O 是在 O 中局域化的，因此表征 $F(\rho_O) = \pi_0 \circ \rho_O$ 跟 $\mathfrak{A}(O')$ 上的 π_0 是一致的。由于 $F(U_O)$: $F(\rho) \to F(\rho_O)$ 是幺正的，使我们具有：

$$F(\rho) \mid \mathfrak{A}(O') \cong F(\rho') \mid \mathfrak{A}(O') = \pi_0 \mid \mathfrak{A}(O')$$

这意味着 $F(\rho) = \pi_0 \circ \rho \in DHR(\mathfrak{A})$。现在设 ρ，$\rho' \in \mathrm{Obj}(\Delta)$ 和 $\tilde{S} \in \mathrm{Hom}(F(\rho), F(\rho'))$，如果 O 是一个包含 ρ，ρ' 的局域化区域的对顶锥，对所有 $A \in \mathfrak{A}(O')$ 有：

$$\cdots\cdots\cdots\cdots\cdots\tilde{S}\pi_0(A) = \tilde{S}\pi_0(\rho(A)) = \tilde{S}F(\rho)(A) = F(\rho')(A)S$$
$$= \pi_0(\rho'(A))\tilde{S} = \pi_0(A)\tilde{S}$$

因此，通过 π_0 的哈格对偶性，$\tilde{S} \in \pi_0(\mathfrak{A}(O'))' = \pi_0(\mathfrak{A}(O))$。因此，存在 $S \in \mathrm{Hom}_\Delta(\rho, \rho')$ 使得 $\tilde{S} = F(S)$，这证明函子 F 是满的。最后，设 $\pi \in DHR(\mathfrak{A})$ 是一个希尔伯特空间 H 上的 DHR 表征，选择任何对顶锥 O，那么 DHR 标准意味着幺正 U: $H \to H_0$ 的存在，使得对于所有 $A \in \mathfrak{A}(O')$ 都有 $U\pi(A) = \pi_0(A)U$。通过 $\pi'(\cdot) = U\pi(\cdot)U^*$，定义真空希尔伯特空间 H_0 上 $\mathfrak{A}$ 的一个新表征 π'。由这个定义，我们有对于所有 $A \in \mathfrak{A}(O')$ 的 $\pi'(A) = \pi_0(A)$。如果现在 $\hat{O}$ 是包含 O 的任何对顶锥，以及 $A \in \mathfrak{A}(\hat{O})'$ 和 $B \in \mathfrak{A}(\hat{O})'$，那么：

$$\pi'(B)\pi_0(A) = \pi'(BA) = \pi'(AB) = \pi'(A)\pi'(B) = \pi_0(A)\pi'(B)$$

通过 π_0 的哈格对偶性，意味着 $\pi'(\mathfrak{A}(\hat{O})) \subset \pi_0(\mathfrak{A}(\hat{O}'))' = \pi_0(\mathfrak{A}(O))$。因此 π' 把准局域代数 $\mathfrak{A}$ 映射到 $\pi_0(\mathfrak{A})$。由于 π_0 是内射，我们能够通过 $\rho = \pi_0^{-1} \circ \pi'$ 定义一个 $\mathfrak{A}$ 的自同态 ρ。通过建构，ρ 在 O 中是局域化的，以及我们有 $\pi' = \pi_0 \circ \rho = F(\rho)$。重复对一个不同对顶锥 $\hat{O}$ 的论证，我们看到 ρ 是可输运的，因此

$\rho \in \Delta$。由于 $\pi \cong \pi' = F(\rho)$，我们证明了任何 DHR 表象是幺正性地等价于形式 $F(\rho)$之一，其中 $\rho \in \Delta$。因此函子 F 本质上是满射的，并且，参看附录 A，也是一个范畴的等价。

注释 207。命题 206 具有双重的意义。一方面，它通过$\mathfrak{A}$的一类表象提供了范畴 Δ 的一个解释，如果人们承认范畴 DHR($\mathfrak{A}$)是值得研究的话，上述等价性就是一个有力工具。也就是说，它允许我们把 Δ 上对称性单项结构拖回 DHR($\mathfrak{A}$)——就像开始定义的只是一个范畴——诸如使函子 F: $\Delta \to \mathrm{DHR}(\mathfrak{A})$ 成对称张量 C^* 范畴的一个等价。但是一旦接受这一点，用范畴 Δ 就比用 DHR($\mathfrak{A}$)做事方便得多，因为 DHR($\mathfrak{A}$)上的张量结构会不严格。

至于 DHR 条件的物理动机，我们给出三个论证：

1. 靠一个增长中的对顶锥序列，我们的意思是对顶锥的一个系列 $O_1 \subset O_2 \subset \cdots$，使得 $\cup_i O_i = \mathbb{R}^d$(通常 d = 4)。在[Doplicher *et al.*, 1971]的一个附录里，下述结果(反过来是平凡的)得以证明：

定理 208。设 ω 是一个$\mathfrak{A}$上态，使得：

$$\lim_{n \to \infty} \| (\omega - \omega_0) \mid \mathfrak{A}(O_n{}') \| = 0$$

并且跟 ω 有关的 GNS 表征 π_ω 满足性质 B，那么存在一个对顶锥 O，使得 $\pi_\omega \mid \mathfrak{A}(O') \cong \pi_0 \mid \mathfrak{A}(O')$。

2. 在第 9 节我们会证明超选择截面满足 DHR 标准，该截面通过一个满足玻色—费米对易关系的场网跟真空截面连接在一起(确切的定义和描述参见第 9 节)。相反，在第 10 节我们会证明任何具有有限维的 DHR 表征都以这种方式出现。综合这些结果说明，DHR 超选择截面准确说是由(分次)局域场约化的那些。进一步讨论看 9—10 节。

3. 设(H, π)是$\mathfrak{A}$的庞加莱协变表征(在定义 205 的意义上)，使得 H 是可分离的，并且动量算符P^μ 的谱 $P^\mu \subset \mathbb{R}^d$ 具有一个孤立的质量壳$\{p \mid p^2 = m^2\}$在其底部，其中 $m > 0$(这样一个表象被称为有质量单粒子表象)，那么，就像[Buchholz and Fredenhagen, 1982]证明的，对于每一个"类空光锥"C，人们具有幺正等价 $\pi \mid \mathfrak{A}(C') \cong \pi_0 \mid \mathfrak{A}(C')$。对于比如在[Buchholz and Fredenhagen, 1982]的类空光锥的定义，尽管存在局域化的性质比起人们用 DHR 标准给予的要弱，局域化在类空光锥的表象范畴还是能够配置一个辫子单项结构，比如在

[Buchholz and Fredenhagen, 1982]，这个理论的纯粹表征理论部分在[Doplicher and Roberts, 1990, 4]。在此理论中，为了辫子是对称的，时空维度一定是$\geqslant 3+1$！其数学处理的技巧方面还要复杂，原因是：如果π是一个表征，使得$\pi \upharpoonright \mathfrak{A}(\mathcal{C}')=\pi_0 \upharpoonright \mathfrak{A}(\mathcal{C}')$，那么哈格对偶性意味着$\pi(\mathfrak{A}(\mathcal{C}))\subset\pi(\mathfrak{A}(\mathcal{C}))''$，但是由于弱封闭性，右手边并不包含在代数$\mathfrak{A}$中。我们在第10节讨论的场网的构建永远不能推广到局域化在类空光锥里的荷，比如在[Doplicher and Roberts, 1990, Section 5]。在所引结果的基础上，似乎光锥—局域化超选择截面在物理上比更严格的DHR截面动机还强烈。7—10小节里阐释的D(H)R理论作为技术上早期"数学实验"还是有用的。

8.3 在张量*范畴里的维度理论

对于任何张量*范畴，我们都能定义一个"共轭"概念。下面就是一个严格张量*范畴情况下的这个定义的简化方案。

定义209。 设$\mathcal{C}$是一个严格张量*范畴并且$X\in\mathrm{Obj}(\mathcal{C})$。共轭方程的一个解是一个三元$(\overline{X}, r, \overline{r})$，其中$\overline{X}\in\mathrm{Obj}(\mathcal{C})$，以及$r:\ 1\to\overline{X}\otimes X$，$\overline{r}:\ 1\to X\otimes\overline{X}$满足：

$$(\overline{r}^*\otimes\mathrm{id}_X)\circ(\mathrm{id}_X\otimes r)=\mathrm{id}_X$$

$$(r^*\otimes\mathrm{id}_{\overline{X}})\circ(\mathrm{id}_{\overline{X}}\otimes\overline{r})=\mathrm{id}_{\overline{X}}$$

一个严格张量*范畴$\mathcal{C}$具有共轭，如果存在一个对任何$X\in\mathcal{C}$的共轭方程的解。

例子210。 共轭定义是体现在紧致群的有限维表征的(严格)范畴$\mathrm{Rep}_f G$里。特别是，为人所熟知的，对于G的每一个表征(H, π)，存在G的一个共轭表征$(\overline{H}, \overline{\pi})$。存在共轭表象的几个不同结构，参见比如[Simon(西蒙), 1996, 30]。通过普遍性质的方式，$(\overline{H}, \overline{\pi})$是$G$的唯一不可约化表征，使得$(H\otimes\overline{H}, \pi\otimes\overline{\pi})$包含$G$的平凡表征的一个复制。

注释211。 假设$(\overline{X}, r, \overline{r})$是一个$X$的共轭，并且张量单位1是不可约化的，那么$r^*\circ r\in\mathrm{End}(1)=\mathbb{C}\mathrm{id}_1$。因此提到一个标量，$r$是一个等距，由此1是$\overline{X}\otimes X$的一个直和。而且，正如使用共轭方程所能够证明的，由$s\mapsto(\mathrm{id}_{\overline{X}}\otimes s)\circ r$定义的映射$\mathrm{End}(X)\to\mathrm{Hom}(1, \overline{X}\otimes X)$，是矢量空间的一个同构。因此，如果$X$

是不可约化的，那么直和 1 和 $\overline{X}\otimes X$ 里的乘数 1 一起出现。

定义 212。 设$\mathcal{C}$是一个张量 * 范畴，并且 $X\in \mathrm{Obj}(\mathcal{C})$，一个跟 X 相对的共轭方程解$(\overline{X}, r, \overline{r})$被称为是正规化的，如果：

$$r^{*}\circ r=\overline{r}^{*}\circ \overline{r}$$

并且是标准的，那么对于所有 $a\in \mathrm{End}(X)$ 有：

$$r^{*}\circ(\mathrm{id}_{\overline{X}}\otimes a)\circ r=\overline{r}^{*}\circ(a\otimes \mathrm{id}_{\overline{X}})\circ \overline{r}$$

注释 213。 如果 X，Y 具有(标准)共轭，那么 $X\otimes Y$ 和 $X\oplus Y$ 也具有(标准)共轭。如果一个对象具有一个共轭，那么它具有标准共轭。详情参看附录。

定义 214。 如果一个对象 $X\in \mathrm{Obj}(\mathcal{C})$具有一个标准共轭$(\overline{X}, r, \overline{r})$，我们定义其维度 $d(X)$为：

$$d(X)\,\mathrm{id}_{1}=r^{*}\circ r$$

如果一个对象 X 没有共轭，则我们形式上说 $d(X)=+\infty$。

注释 215。 对于所有 $X\in \mathrm{Obj}(\mathcal{C})$，$d(X)\geqslant 0$。而且，如果 X，$Y\in \mathrm{Obj}(\mathcal{C})$具有共轭，那么：

$$d(\overline{X})=d(X),\ d(X\otimes Y)=d(X)\cdot d(Y),\ d(X\oplus Y)=d(X)+d(Y)$$

并且 $d(1)=1$。参看附录对这些事实的讨论。

定义 216。 设 Δ 是 DHR 范畴，我们定义具有有限维的对象的完全子范畴 Δ_f 为：

$$\mathrm{Obj}(\Delta_f)=\{\rho\in \mathrm{Obj}(\Delta):\ d(\rho)<+\infty\}$$

注释 217。 根据定义，Δ_f 是一个具有共轭的范畴。它在张量积、直和以及子对象之下是封闭的。在具有共轭的任何 C^{*} 张量范畴下，任何对象的维度取值在$[1, \infty]$，并且在区间$[1, 2]$只可能出现形如 $2\cos(\pi/n)$，$n\geqslant 3$ 的值，参看[Longo and Roberts，1997]。在对称的 C^{*} 张量范畴里，全部维度都是整数，就像附录里已证明了的。

命题 218。 对于每一个 X，$Y\in \mathrm{Obj}(\Delta_f)$，$\mathrm{Hom}(X, Y)$是有限维矢量空间。每一个对象 $X\in \mathrm{Obj}(\Delta_f)$是不可约化对象的有限直和：即范畴 Δ_f 是半单的。

证明。参见附录。

注释 219。 存在一个由隆哥[1989]发现并且在[Longo and Roberts，1997]里得到进一步探讨的重要关联，它与 DHR 截面 $\rho\in\Delta$ 的维度跟子因子理论之间关系有关。在众多其他事件中，后者跟任何指标为$[M: N]\in[1, \infty]$因子的

嵌入 $N \subset M$ 结合在一起。为了应用这个理论到 AQFT，我们需要假设(或者证明)局域性的冯·诺伊曼代数 $\mathfrak{A}(O)$ 都是因子。(这是自动的，比如在共形协变理论里。)如果 $\rho \in \Delta$ 在 O 里的局域化，它把 $\mathfrak{A}(O)$ 的正规 * 同态限制到自身，产生一个嵌入 $\rho(\mathfrak{A}(O)) \subset \mathfrak{A}(O)$。这个子因子的指标是跟范畴上由下式确定的维度 $d(\rho)$ 关联的：

$$[\mathfrak{A}(O): \rho(\mathfrak{A}(O))] = d(\rho)^2 \tag{25}$$

隆哥的结果允许给出一个由 DHR 表征的(局域化自同态相关的)维度的最直接公式。即所有对 $\pi \cong \pi_0 \circ \rho$ 的自同态 $\rho \in \Delta$ 具有同样的范畴维度，从而写成 $d(\pi)$，并且对于任何对顶锥 O，我们有：

$$d(\pi) = [\pi(\mathfrak{A}(O'))': \pi(\mathfrak{A}(O))]^{1/2}$$

这可以看成：π 是幺正等价于表征 $\pi' = \pi_0 \circ \rho$，其中 $\rho \in \Delta$ 是在 O 里局域化的，那么嵌入 $\pi(\mathfrak{A}(O)) \subset \pi(\mathfrak{A}(O'))'$ 是幺正地等价于 $\pi'(\mathfrak{A}(O)) \subset \pi(\mathfrak{A}(O'))'$，等于 $\pi_0(\rho(\mathfrak{A}(O))) \subset \pi_0(\mathfrak{A}(O))$。现在这个说法是根据方程(25)和指标在幺正变换 $[UMU^*: UNU^*] = [M: N]$ 下是不变的事实。

似乎可以进行另一个评论：一个对象维度的范畴定义要求共轭对象的存在。另一方面，假设局域代数的因子性，(对一个在 O 中局域化的同态)表示 $[\mathfrak{A}(O): \rho(\mathfrak{A}(O))]$ 和(O 中其独立性来自温和加法公理的) $[\pi(\mathfrak{A}(O'))': \pi(\mathfrak{A}(O))]$ 并不以共轭的存在为前提。事实上，人们能够证明这些子因子指标的有限性，意味着共轭 DHR 表示的存在，参见[Guido and Longo(盖德和隆哥)，1992]。

8.4 协变表象

由于我们决定采用局域化可输运自同态的范畴 Δ 而不是直接使用 DHR 表象进行研究，我们需要下面的定义。

定义 220。设 $\mathfrak{A}$ 是一个具有协变真空表征 (H_0, π_0) 的庞加莱协变网。如果存在一个强的连续正能量表征 π_ρ: $\hat{\mathcal{P}} \to U(H_0)$，使得：

$$\mathrm{Ad}U_\rho(h) \circ \pi_0 \circ \rho = \pi_0 \circ \rho \circ \beta_h \quad \forall h \in \hat{\mathcal{P}} \tag{26}$$

那么一个自同态 $\rho \in \Delta(\mathfrak{A})$ 称为协变的。

由协变态射组成的 $\Delta(\mathfrak{A})$ 的全部子范畴由 $\Delta_c(\mathfrak{A})$ 表示。

注释 221。对于 $\rho\in\Delta$，$h\in\mathcal{P}$，我们定义 $\rho_h=\beta_h\circ\rho\circ\beta_h^{-1}$，如果 ρ 在对顶锥是局域化的，那么 ρ_h 在 hO 里是局域化的。如果 $\rho\in\Delta_c$，那么方程(26)能够被重新表述为：

$$\mathrm{Ad}(U(h)U_\rho(h)^*)\circ\pi_0\circ\rho=\pi_0\circ\beta_h\circ\rho\circ\beta_h^{-1}=\pi_0\circ\rho_h\qquad\forall h\in\hat{\mathcal{P}}$$

由于 ρ 和 ρ_h 两个都是局域化的，得出 $X_\rho(h)\equiv U(h)U_\rho(h)^*\in\mathrm{Hom}(\rho,\rho_h)$，因此 $X_\rho(h)\in\mathfrak{A}$。$\mathfrak{A}$ 值上循环是很方便的，因为像 $\rho(U(h))$ 这样的表述没有意义，然而 $\rho(X_\sigma(h))$ 有意义。它满足下面的循环方程：

$$\begin{aligned}X_\rho(gh)&=U(gh)U_\rho(gh)^*=U(g)U(h)U_\rho(h)^*U_\rho(g)^*\\&=\beta_g(U(h)U_\rho(h)^*)U(g)U_\rho(g)^*=\beta_g(X_\rho(h))X_\rho(g)\end{aligned}$$

同样的计算意味着，如果 $\rho\in\Delta$ 和 $h\mapsto X_\rho(h)\in\mathfrak{A}$ 满足对于所有的 g，$h\in\mathcal{P}$ 都有 $X_\rho(gh)=\beta_g(X_\rho(h))X_\rho(g)$，那么 $U_\rho(h):=X_\rho(h)^*U(h)$ 是 $\mathcal{P}$ 的表征并且保留方程(26)，即 $\rho\in\Delta_c$。

命题 222。Δ_c 是在张量积、直和和子对象下闭合的。

证明。设具有相关上循环 X_ρ，$X_{\rho'}$ 的 ρ，$\rho'\in\Delta_c$，那么：

$$X_{\rho\rho'}(h)=X_\rho(h)\otimes X_{\rho'}(h)=X_\rho(h)\rho(X_{\rho'}(h))\in\mathrm{Hom}(\rho\otimes\rho',\ \rho_h\otimes\rho'_h)\quad(27)$$

显然满足上述循环方程，因此 $\rho\rho'$ 是协变的。这里略去对直和以及子对象的证明，参看[Roberts, 1990]。

如果 $T\in\mathrm{Hom}(\rho,\rho')$，那么：

$$\beta_h(T)\rho_h(A)=\beta_h(T\rho\beta_h^{-1}(A))=\beta_h(\rho'\beta_h^{-1}(A)T)=\rho_h'(A)\beta_h(T)$$

因此 $\beta_h(T)\in\mathrm{Hom}(\rho_h,\ \rho_h')$。

现在我们探索有限维的一些结果。

命题 223。令 ρ，$\rho'\in\Delta_{fc}:=\Delta_f\cap\Delta_c$，那么：

(i) 如果 $T\in\mathrm{Hom}(\rho,\rho')$，那么对于所有 $h\in\hat{\mathcal{P}}$ 都有 $TU_\rho(h)=U_{\rho'}(h)T$；

(ii) 每一个 $\rho\in\Delta_{fc}$ 关于唯一的表象 U_ρ 都是协变的；

(iii) 如果 ρ，$\rho'\in\Delta_{fc}$ 并且 $T\in\mathrm{Hom}(\rho,\rho')$，那么

$$\beta_h(T)X_\rho(h)=X_{\rho'}(h)T\qquad\forall h\in\hat{\mathcal{P}}\qquad(28)$$

(iv) Δ_{fc} 在共轭下是封闭的。

证明。(i)对于 $h \in \hat{\mathcal{P}}$，定义 $T_h = U_{\rho'}(h)TU_\rho(h)^*$，对于任何 $A \in \mathfrak{A}$，我们有：

$$T_h\rho(A) = U_{\rho'}(h)TU_\rho(h)^*\rho(A) = U_{\rho'}(h)T\rho(\beta_h^{-1}(A))U_\rho(h)^*$$
$$= U_{\rho'}(h)\rho'(\beta_h^{-1}(A))TU_\rho(h)^* = \rho'(A)U_{\rho'}(h)TU_\rho(h)^* = T_h\rho'(A)$$

因此 $T_h \in \mathrm{Hom}(\rho, \rho')$。通过假设，$\rho$，$\rho'$具有共轭，从而由命题 218 可知 $\mathrm{Hom}(\rho, \rho')$是有限维的。因此$(h, T) \mapsto T_h$ 是庞加莱群$\mathcal{P}$的一个有限维表象。$TU_\rho(h) = U_{\rho'}(h)T$ 的说法等价于这个表征的平凡性，这个平凡性源自于$\hat{\mathcal{P}}$的有限维幺正表象的非存在性，只要人们产生一个 $\mathrm{Hom}(\rho, \rho')$上的正的确定的$\hat{\mathcal{P}}$不变性内积，对于最后一步请参看[Roberts，1990]。

(ii)把(i)用到$\rho' = \rho$，$U_{\rho'}(h) = \tilde{U}_\rho(h)$，$T = \mathrm{id}_\rho = 1_{H_0}$，有结论 $U_\rho = \tilde{U}_\rho$。

(iii)用(i)进行计算：

$$\beta_h(T)X_\rho(h) = (U(h)TU(h)^*)(U(h)U_\rho(h)^*) = U(h)TU_\rho(h^{-1})$$
$$= U(h)U_{\rho'}(h^{-1})T = X_{\rho'}(h)T,$$

(iv)参看[Roberts，1990]。

注释 224。在网$\mathfrak{A}$上弱附加假设之下，[Guido and Longo，1992，Theorem 5.2]已经证明，有限维的任何局域化自同态自动跟正能量协变！等效地，$\Delta_f \subset \Delta_c$，从而 $\Delta_{fc} = \Delta_f$。

8.5 辫子张量 * 范畴里的统计学

定义 225。设$(\mathcal{C}, \otimes, \mathbf{1})$是一个具有幺正辫子 $c_{X,Y}$的张量 * 范畴，并且设每一个$X \in \mathrm{Obj}(\mathcal{C})$具有一个共轭。对每一个 $X \in \mathrm{Obj}(\mathcal{C})$，我们定义 X 的扭(twist)$\Theta_X \in \mathrm{End}(X)$为：

$$\Theta_X = (\bar{r}^* \otimes \mathrm{id}_X) \circ (\mathrm{id}_{\bar{X}} \otimes c_{X,X}) \circ (\bar{r} \otimes \mathrm{id}_X)$$

其中$(\bar{X}, r, \bar{r})$是 X 的一个标准共轭。

注释 226。对于每一个 $X \in \mathrm{Obj}(\mathcal{C})$，$\Theta_X$ 是幺正的。当 X 是不可约化的、$\mathrm{End}(X) = \mathbb{C}\mathrm{id}_X$，由此 $\Theta_X = \omega_X \mathrm{id}_X$，其中 ω_X 是单位模的一个复数(叫作统计相位)。在此情况下 $c_{X,Y}$是对称的，那么$(c_{X,X})^* = c_{X,X}$，由此$(\Theta_X)^* = \Theta_X$。跟幺正

性一起，这意味着$(\Theta_X)^2=\mathrm{id}_X$。

定义227。设$(\mathcal{C},\otimes,1)$是一个具有幺正对称性$c_{X,Y}$的张量＊范畴，如果$X\in\mathrm{Obj}(\mathcal{C})$是不可约化的，如果$\omega_X=1$，那么我们说$X$是玻色子的对象，并且如果$\omega_X=-1$，那么我们说$X$是费米子的对象。

注释228。我们给出大量证据，因为我们的焦点在于有限维的DHR表象的范畴Δ_f上。

(i)作为对应于QFT里粒子类型的不可约化DHR表征的(幺正等价性类)的启发性解释，一个DHR表象ρ的共轭$\bar{\rho}$对应于反粒子。可能出现一个粒子是其自身的反粒子，即$\rho\cong\bar{\rho}$；但是反粒子的存在似乎是相对论量子场论的一个主要的部分。

(ii)承认上述意义的共轭的DHR截面根据算子代数方法(粗略)类似于具有有限多元素怀特曼场的AQFT。在怀特曼框架里[Streater and Wightman, 1964]，众所周知无限多元存在"病态"情况，PCT和自旋—统计学定理并不适合它们，并且在事实上是违反它们的。在代数QFT里，这些结果反映在这么一个事实里，即我们甚至不能够定义具有维度∞的玻色子和费米子对象，这是在其没有共轭的意义上。

(iii)在[Fredenhagen, 1981]已经被证明任何有质量的单粒子表征里，比如注释207(iii)，在[Buchholz and Fredenhagen, 1982]的结论中提出在类空光锥里是可局域化的，具有一个在光锥可局域化表征的C^*张量范畴里的共轭。因此，自然要求共轭也存在于对顶锥可局域化表征的更严格设定里。

(iv)正如在注释224里指出的，有限维的DHR自同态是自动协变的，只要人们接受这个结果所需要的网$\mathfrak{A}$上的附加条件。即便人们并不希望出现这个结果，对象的有限维度(通过hom集合的有限维)对命题223的证明是需要的，后者在第10节里对于把庞加莱作用从$\mathfrak{A}$提升到场论$\mathfrak{F}$是关键性的。

9. 从场到表象

在这一节我们对超选择规则采取"自上而下"思路。那就是，我们给定一个场代数$\mathfrak{F}$和具体作用在希尔伯特空间$\mathcal{H}$上的规范群G。然后我们定义可观测量

为$\mathfrak{F}$的规范不变的元素。$\mathcal{H}$上$\mathfrak{F}$的表征就会给我们$\mathfrak{A}$表征的一个优先集合，即那些可能是"通过局域场的作用从真空表征产生的"集合。我们在这一节里的数学目的是证明这些表征满足 DHR 选择标准，因此，所有的以传统方式——即通过对场真空的作用——出现的超选择截面，应按照 DHR 超选择理论来处理。但是要注意：我们只关注局域场。

定义 229。设ω_0是$\mathfrak{A}$上的一个态，并且设$(\mathcal{H}_0, \pi_0)$是对应的 GNS 表征。一个具有$(\mathfrak{A}, \omega_0)$规范对称的场系统是一个四元组$(\pi, \mathcal{H}, \mathfrak{F}, (G, k))$，其中$(\mathcal{H}, \pi)$是$\mathfrak{A}$的一个表征，$O \mapsto \mathfrak{F}(O)$是不可约化地作用在$\mathcal{H}$上的冯·诺伊曼代数的一个网，$G$是一个作用在$\mathcal{H}$上的幺正算子的强紧致群，$k$是$G$的一个核心元素使得$k^2 = e$，并且使得：

α)$(\mathcal{H}_0, \pi_0)$是$(\mathcal{H}, \pi)$的子表征，即存在一个等距$V: \mathcal{H}_0 \to \mathcal{H}$使得$V\pi_0 = \pi V$。

β)V把$\mathcal{H}_0$映射到$\mathcal{H}$的G不变性矢量的子空间。

γ)$U \in G$约化保留每一个$\mathfrak{F}(O)$整体上固定的自同构，并且$\pi(\mathfrak{A}(O))'' \subseteq \mathfrak{F}(O)$是在$G$对$\mathfrak{F}(O)$的作用之下不动点的集合。

δ)对于每一个$O \in \mathcal{K}$，$V(\mathcal{H}_0)$是对$\mathfrak{F}(O)$循环。

ε)场相对于可观测量是局域的，即只要O_1和O_2是类空分离的，那么$\mathfrak{F}(O_1)$和$\pi(\mathfrak{A}(O_2))$按元素对易。

几点关于场系统定义的评价：$\mathfrak{F}$是由局域代数$\{\mathfrak{F}(O): O \in \mathcal{K}\}$产生的事实，意味着$\mathfrak{F}$的元素代表局域场——即其激发态能够在有界时空区域内局域化的场。而且：

δ)是瑞赫—施列德条件：它说的是每一个局域区域O承担场的一个完全集，这是就这些局域场能够从真空截面达到每个截面的意义上而言的。但是注意条件(δ)只能保证$\mathcal{H}$里的截面能够从真空截面达到，一个强的完备性概念可能有赖于一些$\mathfrak{A}$物理截面的固有标准，并且可能要求所有这些截面被包含在$\mathcal{H}$里，参看定义 246。

γ)能够被解释为，对称群G是场的一个内部对称性群：它并不改变场算子的时空局域化区域。

ε)是相对局域性条件。由于场不必是可观测量，场代数不要求满足微观因

果性。当然，在通常情况下(即标准对易关系)，在一个时空区域局域化的场算子，要么对易要么反对易于在类空分离区域局域化的场算子。条件(ε)是标准(玻色/费米)对易关系要求的弱化。

由于 G 是作用在$\mathcal{H}$上幺正算子的一个紧致群，我们能够运用紧致群的幺正表征理论的全部工具，参看比如[Folland，1995]。特别是，$\mathcal{H}$分解成约化 G 作用的正交子空间的直和。因此在$\mathcal{H}_\xi$上 G 的约化幺正表象 U_ξ 是因子性的，即由算子$\{g \mid \mathcal{H}_\xi$: $g \in G\}$产生的冯·诺伊曼代数是一个因子。表象 U_ξ 分解成 G 的幺正等价不可约化表征的直和。因此，存在$\mathcal{H}$的一个特殊的直和分解：

$$\mathcal{H} = \bigoplus_\xi \mathcal{H}_\xi$$

其中子空间$\mathcal{H}_\xi$由$\mathcal{H}$中的矢量产生，这些矢量按照 G 的特征 ξ(不可约化表征的幺正等价类)变换。

在这一节里我们的主要目的是：

1. 证明子空间$\mathcal{H}_\xi$约化可观测量代数$\mathfrak{A}$的作用，因此，$\mathfrak{A}$在$\mathcal{H}$上的表象分解成子空间$\mathcal{H}_\xi$上的表征的直和$\oplus_\xi \pi_\xi$。

2. 证明$\mathfrak{A}$的每一个表征($\mathcal{H}_\xi$，π_ξ)是因子的，因此($\mathcal{H}_\xi$，π_ξ)的不可约化子表征是相互等价的。因此 G 的每一个特征 ξ 标记$\mathfrak{A}$的不可约化表征的一个等价类。

3. 证明 $V(\mathcal{H}_0)$确实是$\mathcal{H}$里的 G 不变量矢量的子空间。因此 G 的特征 $\mathbf{1}$ 标记了$\mathfrak{A}$的真空表征的等价类。

4. 证明($\mathcal{H}$，π)的每一个子表征是一个 DHR 表征。简短地说，能够通过局域场的应用从真空达到的截面对应于 DHR 表征(即模化任何区域而等价于真空表征的表征)。

至于目的(1)和(2)，足以证明 $\pi(\mathfrak{A})'' = G''$，因为冯·诺伊曼代数 $\pi(\mathfrak{A})''$ 和 G''分享共同的核心投影。

命题 230。如果(π，$\mathcal{H}$，$\mathfrak{F}$，(G，k))是具有($\mathfrak{A}$，ω_0)规范对称的场系统，那么$\pi(\mathfrak{A})' = G''$。

我们的概念会因此得到简化，如果我们用 g 和 $U(g)$大致表示$\mathcal{H}$上的 G 幺正群的元素。那就是说，我们把 $g \rightarrow U(g)$看作$\mathcal{H}$上的 G 的单位表征。

证明。定义 M: $\mathfrak{B}(\mathcal{H}) \rightarrow G'$为：

$$M(A) = \int_G U(g)AU(g)^* d\mu(g)$$

其中 μ 是 G 上的哈尔(Haar)测度。那么 M 是从 $\mathfrak{B}(\mathcal{H})$ 到 G' 范数的一个忠实的、正规的投影，由于 M 在 $\mathfrak{B}(\mathcal{H})$ 的单位球上是弱连续的，我们有：

$$G' = M(\mathfrak{B}(\mathcal{H})) = M(\overline{\mathfrak{F}}) = \overline{M(\mathfrak{F})} = \overline{\pi(\mathfrak{A})} \tag{29}$$

因此，$G'' = \pi(\mathfrak{A})'$。

从而有 $\mathfrak{A}$ 的表征 $(\mathcal{H}, \pi)$ 的因子子表征一一对应于 $\mathcal{H}$ 上的 G 的作用因子子表征。

注释 231。由于 G 是紧致的，G 的每一个不可约化表征是有限维的。设 $\hat{G}$ 是 G 的特征集合(不可约化表征的等价类)，并且对于 $\xi \in \hat{G}$，设 $d(\xi)$ 是基础希尔伯特空间的维度，那么上述结果给出 $\mathfrak{A}$ 的表征 $(\mathcal{H}, \pi)$ 的一个好的直观图像。对于每一个 $\xi \in \hat{G}$，选择因子表征 $(\mathcal{H}_\xi, \pi_\xi)$ 的一个不可约化子表征 $(\mathcal{H}_\rho, \pi_\rho)$，那么我们就有：

$$\pi(A) = \bigoplus_{\xi \in \hat{G}} d(\xi)\pi_\rho(A) = \bigoplus_{\xi \in \hat{G}} (\pi_\rho(A) \otimes I_\rho)$$

其中 $d(\xi)\pi_\rho(A) = \pi_\rho(A) \oplus \cdots \oplus \pi_\rho(A)$，$d(\xi)$ 倍以及 I_ρ 是 $d(\xi)$ 维希尔伯特空间上的单位元。

引理 232。设 $(\pi, \mathcal{H}, \mathfrak{F}, (G, k))$ 是一个具有 $(\mathfrak{A}, \omega_0)$ 规范对称的场系统，那么 $\mathcal{H}_0$ 对 $\mathfrak{F}$ 是分离的。

证明。设 $F \in \mathfrak{F}$，如果 $F\mathcal{H}_0 = \{0\}$，那么 $\mathscr{E}(F^*F)\mathcal{H}_0 = \{0\}$。由于 $\mathscr{E}(F^*F) \in \pi(\mathfrak{A})$ 和 π_0 是忠实的，就有 $\mathscr{E}(F^*F) = 0$。由于 $\mathscr{E}$ 是忠实的，那么 $F = 0$。因此，$\mathcal{H}_0$ 对 $\mathfrak{F}$ 是分离的。

为了获得关于场系统 $(\pi, \mathcal{H}, \mathfrak{F}, (G, k))$ 的进一步信息，我们统一场代数 $\mathfrak{F}$ 里的"G 的作用下的张量"。为了落实这个想法，暂时忘记 $\mathfrak{F}$ 具有积运算，而将它仅仅作为一个巴拿赫空间考虑。映射 $U \mapsto \mathrm{Ad}U$ 是一个在 $\mathrm{Aut}\,\mathfrak{F}$ 里的紧致群 G 的(强)连续表征，当然也是 $\mathfrak{F}$ 上可逆线性算子的子集。跟在希尔伯特空间 H 上的紧致李群的表示情况一样，$\mathfrak{F}$ 上 G 的表征分解成不相交表示的直和。一个算子 $F \in \mathfrak{F}$ 被认为是按照 $\mathfrak{F}$ 表示的 ρ 变换，当且仅当它是包含在支集 G 相应表征 $\mathfrak{F}$ 的线性子空间 H_ρ。事实上，我们将证明 $\mathfrak{F}$ 里的不可约化子空间具有特殊的

代数性质：它们具有支集 I。

引理 233。设$(\pi, \mathcal{H}, \mathfrak{F}, (G, k))$是一个具有$(\mathfrak{A}, \omega_0)$规范对称的场系统，并且假定$\mathfrak{A}$满足与$\omega_0$有关的性质 B。那么网 $O \mapsto \pi(\mathfrak{A}(O))''$满足性质 B。

证明。我们首先建立了 $\pi_0|_{\mathfrak{A}(O')}$准等价于每一个对顶锥 O 的 $\pi|_{\mathfrak{A}(O')}$。

由相对局域性条件(ε)，$\mathfrak{F}(O) \subseteq \pi(\mathfrak{A}(O'))'$。由瑞赫—施列德条件$(\delta)$，$\mathcal{H}_0$是$\mathfrak{F}(O)$的循环子空间。因此，

$$\mathcal{H} = [\mathfrak{F}(O)\mathcal{H}_0] \subseteq [\pi(\mathfrak{A}(O'))'\mathcal{H}_0]$$

令 E_0 是到$\mathcal{H}_0$上的正交投影。E_0 在 $\pi(\mathfrak{A}(O'))'$里的中心支集是投影到$[\pi(\mathfrak{A}(O'))'E_0(\mathcal{H})]$上的[Kadison and Ringrose，1997，Prop. 5.5.2]。因此 E_0 具有 $\pi(\mathfrak{A}(O'))'$里的中心支集 I，由此$(\pi_0|_{\mathfrak{A}(O')}, \mathcal{H}_0)$和$(\pi|_{\mathfrak{A}(O')}, \mathcal{H})$是准等价的[Kadison and Ringrose，1997，Thm. 10.3.3]。

设 O_1 是其闭合包含在 O 里的对顶锥，并且设 E 是在 $\pi(\mathfrak{A}(O_1))''$里的非零投影。选择跟 O 分离的对顶锥 O_2，前面论证表明存在一个从 $\pi_0(\mathfrak{A}(O_2'))''$到 $\pi(\mathfrak{A}(O_2'))''$的 $*$ 同态 φ，使得对于所有 $A \in \mathfrak{A}$都有 $\varphi(\pi_0(A)) = \pi(A)$。这个同态 φ 保留了网结构：对于包含在 O_2'里的任何对顶锥 O_3，有 $\varphi[\pi_0(\mathfrak{A}(O_3))] = \pi(\mathfrak{A}(O_3))$。再者，由于 φ 是超弱连续的[Kadison and Ringrose，1997，Cor. 7.1.16]，有 $\varphi[\pi_0(\mathfrak{A}(O_3))''] = \pi(\mathfrak{A}(O_3))''$。特别是，$\varphi(E)$是一个在 $\pi_0(\mathfrak{A}(O_1))''$里的投影。由于 π_0 的性质 B，存在一个等距 $V \in \pi_0(\mathfrak{A}(O))''$，使得 $VV^* = \varphi(E)$。因此，$W := \varphi^{-1}(V) \in \pi(\mathfrak{A}(O))''$是一个等距使得 $WW^* = E$。所以网 $O \mapsto \pi(\mathfrak{A}(O))''$满足性质 B。

定义 234。考虑$\mathfrak{F}$里元素的有序 n 元组$(F_1, F_2, \cdots, F_n)$，我们说这个 n 元组按照特征 ξ 在 G 作用下变换，当且仅当：

1. $F_i^* F_j = 0$ 如果 $i \neq j$；并且

2. $\alpha_g(F_i) = \sum_{j=1}^{n} u_{ij}^{\xi}(g) F_j$，其中 u_{ij}^{ξ}是一个 ξ 的矩阵元。即对于某些类 ξ 的 G 表示(H, ρ)，以及对于 H 的正交基$\{e_1, e_2, \cdots, e_n\}$，有 $u_{ij}(g) := \langle e_i, \rho(g)e_j\rangle_H$。

注释 235。如果$(F_1, F_2, \cdots, F_n)$是一个按照 ξ 变换的$\mathfrak{F}$中的张量，那么我们总能够用具有正交范围的部分等距 V_i 取代 F_i。确实，设 $V_i|F_i|$ 是 F_i 的极

化分解，当 $i\neq j$，$F_i^* F_j = 0$ 时，因而 F_i^* 和 F_j^* 具有正交的范围。回顾一下，如果 $F = V|F|$，那么 V 化零了 $r(F^*) = r(|F|)$ 的正交补，参看[Kadison and Ringrose，1997，Thm. 6.1.2]。因此 $V_j|F_i| = \delta_{ij}F_j$，并且：

$$(\sum_j u_{ij}^{\xi}(g)V_j)|F_i| = \sum_j u_{ij}^{\xi}(g)V_j|F_j| = \sum_j u_{ij}^{\xi}(g)F_j = F_i$$

由极化分解的唯一性，$\sum_j u_{ij}^{\xi}(g)V_j = V_i$，因此 $(V_1,\cdots,V_n)$ 是按照 ξ 变换的一个张量。

定义 236。给定 $\varphi,\psi\in\mathcal{H}_\xi$，定义一个映射 $M_{\varphi,\psi}^{\xi}$：$\mathfrak{F}\to\mathfrak{F}$ 为：

$$M_{\varphi,\psi}^{\xi}(F) = \int_G \langle\varphi, U(g)\psi\rangle \alpha_g(F)d\mu(g)$$

其中 μ 是 G 上哈尔测度。

事实 237。由于 μ 的不变性我们有 $\alpha_g \circ M_{\varphi,\psi}^{\xi}(F) = M_{U(g)\varphi,\psi}^{\xi}(F)$。

引理 238。设 $(F_1, F_2,\cdots,F_n)$ 是 $\mathfrak{F}(O)$ 里的一个张量，是作为类别 ξ 的一个幺正表征变换的，那么 $F_i(\mathcal{H}_0)\subseteq\mathcal{H}_\xi$，其中 $\mathcal{H}_\xi$ 是按照 ξ 变换的 $\mathcal{H}$ 的矢量子空间。

证明思路：设 $\varphi\in\mathcal{H}_0$，以及设 $g\in G$，那么：

$$U(g)[M_{\varphi,\psi}^{\xi}(F)\varphi] = U(g)M_{\varphi,\psi}^{\xi}(F)U(g)^*\varphi = M_{u(g)\varphi,\psi}^{\xi}(F)\varphi$$

然后使用 ξ 的矩阵元素进行直接计算得到这个结果。

引理 239。设 ξ 是 G 的一个特征，其非平凡地产生在 G 在 $\mathcal{H}$ 上作用的分解里，那么，对于每一个对顶锥 O，存在一个在 $\mathfrak{F}(O)$ 里的张量 $(F_1, F_2,\cdots,F_n)$，它是作为类别 ξ 的幺正表征来变换的。

证明思路。设 $(\psi_1,\psi_2,\cdots,\psi_n)$ 是一个从 $\mathcal{H}_\xi$ 的一个 G 不可约化子空间来的正交基。设 φ 是在此相同子空间里的单位矢量。由于 $\mathcal{H}_0$ 是 $\mathfrak{F}(O)$ 的循环，并且 $\mathfrak{F}(O)$ 是一个冯·诺伊曼代数，那么存在一个 $F\in\mathfrak{F}(O)$ 和一个矢量 $\varphi_0\in\mathcal{H}_0$，使得 $F_{\varphi_0} = \varphi$。设 $F_i = M_{\psi_i,\varphi}^{\xi}(F)$，然后人们证明 $(F_1, F_2,\cdots,F_n)$ 是所要求的张量。

引理 240。设 $F_1, F_2,\cdots,F_n\in\mathfrak{F}(O)$，使得 $(F_1, F_2,\cdots,F_n)$ 按照特征 ξ 变换，那么，如果 $\overline{O}\subseteq O_1$，存在 $X_1, X_2,\cdots,X_n\in\mathfrak{A}(O_1)$，使得 $(X_1, X_2,\cdots,X_n)$ 按照 ξ 变换，并且：

$$\sum_{i=1}^{n} X_i^* X_i = I$$

证明。首先用 V_1，V_2，…，V_n 取代 F_1，F_2，…，F_n，那么设 $V=\sum_{i=1}^{n}V_i$。由于 V_i 具有正交范围，V 是部分等距，并且 $V^*V=\sum_{i=1}^{n}V_i^*V_i$。一个直接计算表明对于 $g\in G$ 有 $\alpha_g(V^*V)=V^*V$。因此，$E=V^*V$ 是一个在 $\mathfrak{F}(O)\cap G'=\pi(\mathfrak{A}(O))''$ 里的投影。根据引理 233，$O\mapsto\pi(\mathfrak{A}(O))''$ 满足性质 B。因此，存在一个具有 $WW^*=E$ 的等距 $W\in\pi(\mathfrak{A}(O_1))''$，对于 $i=1$，2，…，n，设 $X_i=V_iW$，那么张量 $(X_i,\ \cdots,\ X_n)$ 按照 ξ 变换，并且：

$$\sum_{i=1}^{n}X_i^*X_i=W^*\left(\sum_{i=1}^{n}V_i^*V_i\right)W=I$$

引理 241。设 $\mathcal{H}_\xi\subseteq\mathcal{H}$ 是在 $\pi(\mathfrak{A})''$ 里的中心投影范围，那么对于每一个对顶锥 O，$\mathcal{H}_\xi$ 对于 $\mathfrak{F}(O)$ 是循环的。

证明。设 O_1 是一个对顶锥使得 $\overline{O}_1\subseteq O$，由瑞赫—施列德条件，$\mathcal{H}_0$ 对于 $\mathfrak{F}(O_1)$ 是循环的。根据引理 239，存在一个 $\mathfrak{F}(O_1)$ 里张量 $(F_1,\ F_2,\ \cdots,\ F_n)$，它是按照 G 的表征 $(H,\ \rho)$ 来变换的。由引理 240，存在一个按照同样方式变换的 $\mathfrak{F}(O)$ 里的张量 $(X_1,\ X_2,\ \cdots,\ X_n)$，并且使得 $\sum_{i=1}^{n}X_i^*X_i=I$。那么：

$$\mathfrak{F}(O)\ \mathcal{H}_0=\mathfrak{F}(O)\sum_{i=1}^{n}X_i^*X_i\mathcal{H}_0\subseteq\mathfrak{F}(O)\ \mathcal{H}_\xi$$

其中最后的引入依照了引理 238，因此 $\mathcal{H}_\xi$ 对 $\mathfrak{F}(O)$ 是循环的。

定义 242。设 $\mathrm{Rep}_{\mathfrak{F}}\mathfrak{A}$ 是 $\mathfrak{A}$ 的表征 $(\mathcal{H},\ \pi)$ 的子表示范畴。我们的意思是取 $\mathrm{Rep}_{\mathfrak{F}}\mathfrak{A}$ 为 $\mathfrak{A}$ 全部表征的范畴的子范畴，即在 $\mathrm{Rep}_{\mathfrak{F}}\mathfrak{A}$ 里的表征之间的 hom 集跟在更大范畴里的 hom 集一样。

命题 243。设 $(\pi,\ \mathcal{H},\ \mathfrak{F},\ (G,\ k))$ 是一个对 $(\mathfrak{A},\ \omega_0)$ 规范对称的场系统，那么存在一个忠实的函子 F: $\mathrm{Rep}_{\mathfrak{F}}\mathfrak{A}\to DHR(\mathfrak{A})$。

证明。假设 $(\mathcal{H}',\ \pi')$ 是一个 $\mathrm{Rep}_{\mathfrak{F}}\mathfrak{A}$ 的对象 V: $\mathcal{H}'\to\mathcal{H}$，使得 $V\pi'=\pi V$。我们随后把 $\mathcal{H}'$ 等同于 $\mathcal{H}$ 里的像，并且把 π' 处理成到 $\mathfrak{B}(\mathcal{H})$ 的映射。我们证明 $(\mathcal{H}',\ \pi')$ 是在 $DHR(\mathfrak{A})$ 里的；即对于任何对顶锥 O，$(\mathcal{H}',\ \pi'|_{\mathfrak{A}(O')})$ 是幺正等价于 $(\mathcal{H}_0,\ \pi_0|_{\mathfrak{A}(O')})$。

设 $\overline{\pi}=\pi|_{\mathfrak{A}(O')}$，由于 E_ι，$E_\xi\in\pi(\mathfrak{A})'\subseteq\pi'(\mathfrak{A}(O'))'$，$E_\iota$ 和 E_ξ 约化 $\overline{\pi}$。我们首先建立起 E_ι 和 E_ξ 具有在 $\pi(\mathfrak{A}(O'))^-$ 里同样的中心支集，它源自于 $E_\iota\overline{\pi}$ 和 $E_\xi\overline{\pi}$ 是准等价的[Kadison and Ringrose，1997，Thm. 10.3.3]。

由相对局域条件(ε)，有$\mathfrak{F}(O)\subseteq\pi(\mathfrak{A}(O'))'$；由瑞赫—施列德条件($\delta$)，$E_\iota\mathcal{H}$是$\mathfrak{F}(O)$的循环子群。因此，

$$\mathcal{H}=[\mathfrak{F}(O)E_\iota(\mathcal{H})]\subseteq[\pi(\mathfrak{A}(O)')'E_\iota(\mathcal{H})]$$

类似地，引理241意味着$E_\xi\mathcal{H}$是$\mathfrak{F}(O)$的循环子空间，由此$[\pi(\mathfrak{A}(O)')'E_\xi(\mathcal{H})]=\mathcal{H}$。而在$\pi(\mathfrak{A}(O)')'$里的中心支集$E_\iota$是到$[\pi(\mathfrak{A}(O)')'E_0(\mathcal{H})]$的投影，并且对$E_\xi$也类似[Kadison and Ringrose，1997，Prop. 5. 5. 2]。因此，E_ι和E_ξ具有在$\pi(\mathfrak{A}(O'))'$的中心支集I。因此$(\pi_0|_{\mathfrak{A}(O')},\mathcal{H}_0)$和$(\pi_\xi|_{\mathfrak{A}(O')},\mathcal{H}_\xi)$是准等价的，即存在$*$等距$\varphi$：$\pi_0(\mathfrak{A}(O'))\to\pi_\xi(\mathfrak{A}(O'))$，使得对于所有$A\in\mathfrak{A}(O')$都有$\varphi(\pi_0(A))=\pi_\xi(A)$。

前面的推理也表明(通过用一个类空分离的对顶锥取代O)，对于每一个对顶锥O，$(\pi_0|_{\mathfrak{A}(O)},\mathcal{H}_\xi)$对于$(\pi_\xi|_{\mathfrak{A}(O)},\mathcal{H}_\xi)$是准等价的。因此，特别是因为冯·诺伊曼代数的网$O\to\pi_0(\mathfrak{A}(O))''$满足性质B(根据假设)，因此网$O\to\pi_\xi(\mathfrak{A}(O))''$也满足此性质。

为了建立$(\pi_0|_{\mathfrak{A}(O')},\mathcal{H}_0)$和$(\pi_\xi|_{\mathfrak{A}(O')},\mathcal{H}_\xi)$是幺正等价的，我们将使用下述结果[Kadison and Ringrose，1997，Thm. 7. 2. 9]：

> 设$\mathfrak{R}_j$，$j=1$，2是作用在各自希尔伯特空间$\mathcal{H}_j$上的冯·诺伊曼代数，假设对于$j=1$，2，存在一个对$\mathfrak{R}_j$循环和分离的矢量$x_j\in\mathcal{H}_j$，如果α：$\mathfrak{R}_1\to\mathfrak{R}_2$是一个$*$同态，那么存在一个从$\mathcal{H}_1$到$\mathcal{H}_2$的幺正算子$U$，实现$\alpha$。

跟下面事实结合[Kadison and Ringrose，1997，Exercise 9. 6. 32]：

> 如果$\mathfrak{R}$是一个作用在分离希尔伯特空间$\mathcal{H}$上的冯·诺伊曼代数，并且，如果$\mathfrak{R}'$是真正无限的，那么存在一个对$\mathfrak{R}$循环和分离的矢量$x\in\mathcal{H}$。

由命题71，$\pi_0(\mathfrak{A}(O'))'$和$\pi_\xi(\mathfrak{A}(O'))'$是真正无限的，由于假设$\mathcal{H}_0$是分离的。因此，这足以证明$\mathcal{H}_\xi$是分离的。由于$\pi_\xi$是不可约化的，每一个非零矢

量$x \in \mathcal{H}_\xi$对$\pi_\xi(\mathfrak{A})$是循环的。因此，$\mathcal{H}_\xi$是对顶锥的一个增长中的序列O_n的$\pi_\xi(\mathfrak{A}(O_n))x$并的闭合。因此这足以证明$\pi_\xi(\mathfrak{A}(O))x$对于每一个$O \in \mathcal{K}$是分离的，由于$\mathcal{H}_0$是分离的，$\mathfrak{B}(\mathcal{H}_0)$的单位球是紧致可度量的[Kadison and Ringrose，1997，Thm. 5. 1. 3；Exercise 5. 7. 7]。由于$\pi_0(\mathfrak{A}(O))''$的单位球是$\mathfrak{B}(\mathcal{H}_0)$的单位球的闭合子集，它也是紧致可度量的。但是$\pi_\xi(\mathfrak{A}(O))''$是$*$同构的，因此弱同态于$\pi_0(\mathfrak{A}(O))''$。因此，$\pi_\xi(\mathfrak{A}(O))''$的单位球是紧致可度量的，因此在弱算子拓扑学里是可分离的。所以$\pi_\xi(\mathfrak{A}(O))''$是可分离的。

在命题206，已经证明存在忠实的、本质上满射的函子F'，在由DHR表示的范畴$DHR(\mathfrak{A})$到$\mathfrak{A}$的局域化可输运的态射的范畴Δ。因此，上述的命题意味着$F' \circ F$是一个从$\mathrm{Rep}_{\mathfrak{F}}\mathfrak{A}$到$\Delta$的忠实的函子。我们随后用恰当的$F$代替$F' \circ F$。

回顾Δ_f是具有共轭对象的Δ的完全子范畴。我们这一节要证明的最后一件事是$\mathrm{Rep}_{\mathfrak{F}}\mathfrak{A}$里的每一个对象在$F$之下同构于$\Delta_f$里的对象。即我们需要证明像对象具有共轭。

证明思路。人们证明在$\mathcal{H}$上的G的子表征在取共轭时是闭合的。通过注意到G在$\mathcal{H}$上的作用跟G在$\mathfrak{F}$上的作用之间的对应就能够证明。然后运用$\mathfrak{F}$是$*$-代数的事实。因此，对于π的每一个不可约化子表征π_ρ，存在π的一个不可约化子表征$\pi_{\bar{\rho}}$，证明$(F' \circ F)(\pi_{\bar{\rho}})$是$(F' \circ F)(\pi_\rho)$的一个共轭。因此，$F' \circ F$是一个从$\mathrm{Rep}_{\mathfrak{F}}\mathfrak{A}$到$\Delta_f$的忠实的函子。从而我们证明了：

$\mathfrak{A}$的每一个表示，来自于它被看作场代数的规范不变性部分，是具有$\rho \in \mathrm{Obj}(\Delta_f)$的形式$\pi_0 \circ \rho$的表征。因此，$\Delta_f$的研究始于场代数的方法产生的表示研究。

我们前面说过，在“正常”情况下，$\mathfrak{F}(O_1)$里的场算子要么对易于要么不对易于在$\mathfrak{F}(O_2)$里的场算子，只要O_1和O_2是类空分离的。更加明确一些，我们希望一个玻色场算子对易于另一个玻色场算子，也对易于费米场算子；并且我们希望一对费米场算子反对易。但是玻色场跟费米场算子之间呢？两者之间的区分是靠规范群G的特殊元k来进行的。

定义244。如果$\alpha_k(F) = F$，那么G被说成是一个玻色场算子，而如果$\alpha_k(F) = -F$，那么F被说成是费米场算子。

我们定义$\mathcal{H}$里的一个玻色截面为一个子空间$\mathcal{H}_\xi$，使得$U(k) \mid_{\mathcal{H}_\xi} = I$，而在$\mathcal{H}$里的一个费米截面定义成子空间$\mathcal{H}_\xi$，使得$U(k) \mid_{\mathcal{H}_\xi} = -I$。那么就有玻色场

算子从真空产生玻色截面，而费米场算子从真空产生费米截面。

我们现在能够使正规对易关系有如下意义：玻色场算子应该彼此对易，并且跟费米场算子对易。费米场算子应该彼此反对易。

定义 245。场$(\pi,(G,k),\mathfrak{F})$的一个局域算子代数系统，被认为是满足正常对易关系，当且仅当局域场代数满足分级的局域对易性：如果 O_1 和 O_2 是类空的，并且 $F_\sigma\in\mathfrak{F}(O_1)$，$F'_\sigma\in\mathfrak{F}(O_2)$，使得 $\alpha_k(F_\sigma)=\sigma F_\sigma$ 和 $\alpha_k(F'_\sigma)=\sigma F_\sigma$，其中$(\sigma=\pm)$，那么：

$$F_+F'_+=F'_+F_+,\quad F_+F'_-=F'_-F_+,\quad F_-F'_-=-F'_-F_-$$

10. 从表征到场

上一节导出了$\mathfrak{A}$的表征的性质，前提是这些表征都是从真空表征通过局域场在真空的作用下产生的。但是这样一种方法对于代数信奉者来说似乎只是启发性的。从代数信奉者的观点看，整个理论内容被包含在可观测量代数的抽象网$\mathfrak{A}$中。

一方面，代数信奉者可能是关于场和规范群的排斥论者。另一方面，他们可能断言场和规范群在物理上意义重大，但仅仅是因为它们能够从可观测量代数网“重新建构”。为了辩护后一个立场，信奉者可能需要完成下面的工作。

任务：按照数学本身的方式，从可观测量代数网努力重建 QFT 的全部工具，即场、规范群等。

这的确是一项过于理想化的任务！对人们来说，哲学家似乎已有定论，认为理论对经验来说总是不充分决定的；由此我们不应该指望能够隐藏在可观测量代数网里面的全部理论工具。但是有一个惊喜：这个任务已被实行并且被完成。DR 重建定理以一个完全牢固和确切的方式证明，DHR 范畴把所有需要重建的信息唯一地解码成场和规范群了。这一节就提供这个重构的细节。

定义 246。一个具有对$(\mathfrak{A},\omega_0)$规范对称$(\pi,\mathcal{H},\mathfrak{F},(G,K))$的场系统被认为是完备的，如果$\mathfrak{A}$的表象 π 包含了在$\mathfrak{A}$的 DHR 范畴 $DHR(\mathfrak{A})$里全部表征的复制。

定义 247。两个对$(\mathfrak{A},\omega_0)$具有规范对称$(\pi_1,\mathcal{H}_1,\mathfrak{F}_1,G_1)$和$(\pi_2,\mathcal{H}_2,$

$\mathfrak{F}_2$，G_2)的场系统，被认为是等价的，如果存在一个幺正算子 W: $\mathcal{H}_1 \to \mathcal{H}_2$，使得：

$$W\pi_1(A) = \pi_2(A)W, \qquad \forall A \in \mathfrak{A}$$

$$WU(G_1) = U(G_2)W$$

$$W\mathfrak{F}_1(O) = \mathfrak{F}_2(O)W, \text{对于每一个对顶锥 } O \text{ 而言}$$

多普里克尔—罗伯茨重建定理 设($\mathfrak{A}$，ω_0)是一个可观测量代数网，满足相对于特殊“真空”态 ω_0 的对偶性和性质 B，那么存在一个具有对($\mathfrak{A}$，ω_0)规范对称(π，$\mathcal{H}$，$\mathfrak{F}$，(G，K))的场系统，完备并且具有正规对易关系。任何对($\mathfrak{A}$，ω_0)完备的、正规场系统都等价于(π，$\mathcal{H}$，$\mathfrak{F}$，(G，K))。

重建定理的证明包含在[Doplicher and Roberts，1989]和[Doplicher and Roberts，1990]。在本文中，我们给出另一种证明，建立在德利涅(Deligne)的嵌入定理[Deligne，1990]，以及罗伯茨[Roberts，ND]所获得的结果之上，后者在得到重建定理的完备证明之前就已得到。

总体上，该定理首先证明——正如本质上在[Doplicher *et al.*，1971]建立的——DHR 超选择截面自然具有拥有共轭的辫子张量 * 范畴的结构——并且当时空维度是三或者更大时，我们就能够用“对称的”替代“辫子的”。由于，直到 20 世纪 80 年代后期，这第一个结果只仅仅是建议性的：已经知道在有限维(超)希尔伯特空间上紧致群 G 的表征的范畴 $\mathrm{Rep}_f G$，是一个拥有共轭的对称张量 * 范畴。因此，DHR 超选择截面的范畴似乎具有某些紧致的 G 的 $\mathrm{Rep}_f G$ 的全部结构。由经典的田中—克莱因(Tannaka-Krein)对偶性定理，有可能由 $\mathrm{Rep}_f G$ 重建 G。而且，罗伯茨[Roberts，ND]证明了条件性断言，如果超选择规则截面的范畴是等价于某些紧致 G 的范畴 $\mathrm{Rep}_f G$，那么场代数$\mathfrak{F}$可能被重建。

但是这里存在一个在超选择截面跟范畴 $\mathrm{Rep}_f G$ 之间的关键区别。范畴 $\mathrm{Rep}_f G$ 是“具体的”——它出现在希尔伯特空间范畴的嵌入，即遗忘函子及其可被看作结构集合的对象。需要构建一个场代数的这么一个嵌入的存在，这一点也是明确的，因为人们需要范畴里的对象具有“内在结构”，正如希尔伯特空间的范畴$\mathcal{H}$里的对象是结构的集合。嵌入定理的证明在附录里，在我们讲述它之前，我们需要一些关于“超数学结构”的预备定义。

10.1 超数学结构和嵌入定理

定义 248。一个超矢量空间，或者一个$\mathbb{Z}_2$分次矢量空间，是一个具有不同分解$V=V_+\oplus V_-$的矢量空间V，子空间V_+叫作偶空间，而V_-叫作奇空间，$V_+\cup V_-=:V_h$的元素被称为是齐次的。通过设定如果$v\in V_\pm$而$\omega(v)=\pm 1$，在齐次元素上定义奇偶函数。在两个超矢量空间之间的态射是一个线性映射$T: V\to W$，使得$T(V_\pm)\subseteq W_\pm$。我们设 SVect 代表超矢量空间的范畴，一个超希尔伯特空间是一个具有正定义内积的超矢量空间V，使得$V_-\perp V_+$。我们用$\mathcal{SH}$代表超希尔伯特空间的范畴。

我们现在定义把 SVect 变成一个对称张量范畴的运算。它直接证明两个超矢量空间之间态射的集合 Mor(V, W)是$B(V, W)$的一个线性子空间。因此，SVect 是一个线性范畴。

如果V和W是超矢量空间，那么它们的直和是具有偶的子空间$V_+\oplus W_+$和奇的子空间$V_-\oplus W_-$的矢量空间$V\oplus W$。我们定义 SVect 里的单项积为矢量空间$V\otimes W$，其偶的子空间和奇的子空间被定义为：

$$(V\otimes W)_\sigma=\bigoplus_{\sigma'\sigma''=\sigma}V_{\sigma'}\otimes V_{\sigma''}$$

其中$\sigma=\pm$，因此：

$$(V\otimes W)_+=(V_+\otimes W_+)\oplus(V_-\otimes W_-)$$

$$(V\otimes W)_-=(V_+\otimes W_-)\oplus(V_-\otimes W_+)$$

单项单位是$\mathbb{C}$，具有偶的子空间$\mathbb{C}$。

定义 249。对于两个超矢量空间V，W，我们定义对称同构

$$c_{V,W}:\ V\otimes W\to W\otimes V$$

是通过设定：

$$c_{V,W}(v\otimes w)=(-1)^{(1-\omega(v))(1-\omega(w))/4}\,\omega\otimes v,\qquad \forall_v\in V_h,\ \forall_w\in W_h$$

在齐次单张量上，然后再通过线性拓展。

命题 250。(Svect，$\otimes$，$\mathbb{C}$，$c_{V,W}$)和($\mathcal{SH}$，$\otimes$，$\mathbb{C}$，$c_{V,W}$)两者都是对称张量 * 范畴。

注释 251。由连贯性定理，$\mathcal{SH}$是等价于严格对称张量 * 范畴，后者也用$\mathcal{SH}$表示。

定义 252。一个超群是一个(G, k)对，其中G是一个群，而k是一个G里的核心元素使得$k \cdot k = e$。两个子群(G_1, k_1)和(G_2, k_2)之间的态射是群同态φ：$G_1 \to G_2$使得$\varphi(k_1) = k_2$。

定义 253。一个子群(G, k)的(幺正)表征π，是超希尔伯特空间$V = V_+ \oplus V_-$跟V上G的一个(幺正)表征π结合在一起，使得$\pi(k) \mid_{V_\pm} = \pm \mathrm{id}_{V_\pm}$。$(G, k)$的表征$\mathrm{Rep}(G, k)$形成一个对称张量＊范畴，具有张量积和源自$\mathcal{SH}$的对称，以及具有$\mathbb{C}$上的$(G, k)$的平凡表征的幺半群的单位。

注释 254。设$\mathcal{SH}_f$是有限维超希尔伯特空间的全部子范畴，对子群(G, k)我们用$\mathrm{Rep}_f(G, k)$代表(G, k)的有限维表征的全部子范畴。范畴$\mathcal{SH}_f$和$\mathrm{Rep}_f(G, k)$是半单的并且具有共轭(有关这个术语更多的内容参见附录)，也存在一个正则遗忘函子K：$\mathrm{Rep}_f(G, k) \to \mathcal{SH}_f$。

我们现在转而看嵌入定理的陈述，它要求场代数和规范群的重构。至于更多的超数学，我们建议读者参看[Varadarajan(瓦拉达拉杰)，2004；Deligne and Morgan(德利涅和摩尔根)，1999]。(但是要注意到 DHR 超选择理论不涉及彼此对称变换玻色子和费米子场意义上的超对称，我们的超群定义也是特别的。)

嵌入定理　设$\mathcal{SH}_f$是$\mathbb{C}$之上的有限维超希尔伯特空间的范畴，设$(\mathcal{C}, \otimes, 1, c_{X,Y})$是一个具有幺正对称$c_{X,Y}$、共轭、直和、子对象以及不可约化单项单位1的张量$C^*$范畴(这样一个范畴在附录叫作$STC^*$)，那么：

1. 存在一个忠实的对称张量＊函子E：$\mathcal{C} \to \mathcal{SH}_f$；

2. 存在一个紧致子群(G, k)，其中G是E的幺正自然单项变换的群，以及一个对称张量＊范畴的等价F：$\mathcal{C} \to \mathrm{Rep}_f(G, k)$使得$E = F \circ K$，其中$K$：$\mathrm{Rep}_f(G, k) \to \mathcal{SH}_f$是遗忘函子。

注释 255。嵌入定理的证明在附录 B，在其证明中我们假定张量范畴$\mathcal{C}$是严格的，并且我们也在工作中使用超希尔伯特空间的范畴的严格化$\mathcal{SH}$。在对称张量范畴的连贯性定理看来，严格化假定并不影响这个结果的一般性。不过，我们将要构建的张量函子F：$\mathcal{C} \to \mathcal{SH}_f$并不是严格的张量函子。在下面的场网的建构中，我们为了概念的简单性假设F是严格的。我们将在这一节结尾再评论这个问题。

10.2 场网和代数方面的建构

我们现在把嵌入定理应用到有限维的局域化可输运态射的 DHR 范畴 Δ_f 的情况。特别是，我们将证明，给定一个嵌入 E: $\Delta_f \to \mathcal{SH}_f$，有可能建构一个局域场代数系统($\pi$, $\mathcal{H}$, $\mathfrak{F}$, (G, k))。这个重建策略是基于没有出版的手稿[Roberts, ND]，它假设嵌入(或者纤维)函子的存在。嵌入函子的实际存在定理——虽然建立在田中和德利涅的工作基础上，但是结合了新近提出的简化——可以在附录里找到。

定义 256。作为一个集合，由三元组(A, ρ, ψ)的等价类组成的场代数$\mathfrak{F}_0$，具有 $A \in \mathfrak{A}$，$\rho \in \mathrm{Obj}(\Delta_f)$以及 $\psi \in E(\rho)$，对于 $T \in \mathrm{Hom}(\rho, \rho')$模的等价关系为：

$$(AT, \rho, \psi) = (A, \rho', E(T)\psi)$$

由于 $E(\lambda \mathrm{id}_\rho) = \lambda \mathrm{id}_{E(\rho)}$，我们具有($\lambda A$, ρ, ψ) = (A, ρ, $\lambda\psi$)。结果，我们并不在记号上区分三元组(A, ρ, ψ)和它的等价类。

命题 257。$\mathfrak{F}_0$ 是一个在如下运算下的复矢量空间：

$$\lambda(A, \rho, \psi) := (\lambda A, \rho, \psi), \qquad \lambda \in \mathbb{C} \tag{30}$$

以及

$$(A_1, \rho_1, \psi_1) + (A_2, \rho_2, \psi_2) := (A_1 W_1^* + A_2 W_2^*, \rho, E(W_1)\psi_1 + E(W_2)\psi_2) \tag{31}$$

其中 $\psi_i \in E(\rho_i)$和 $W_i \in \mathrm{Hom}(\rho_i, \rho)$是等距于：

$$W_1 W_1^* + W_2 W_2^* = \mathrm{id}_\rho \tag{32}$$

此外：

$$(A_1, \rho, \psi) + (A_2, \rho, \psi) = (A_1 + A_2, \rho, \psi)$$

$$(A, \rho, \psi_1) + (A, \rho, \psi_2) = (A, \rho, \psi_1 + \psi_2)$$

因此，认为$\mathfrak{A}$等同于{(A, ι, 1)：$A \in \mathfrak{A}$, $1 \in \mathbb{C} \equiv E(\iota)$}，则$\mathfrak{A}$就变成$\mathfrak{F}_0$的一个线性子空间，而认为 $E(\rho)$等同于{(I, ρ, ψ)：$\psi \in E(\rho)$}，$E(\rho)$就变成$\mathfrak{F}_0$ 的一个线性子空间。

证明。我们首先证明运算可以得到很好的定义。标量乘法可以定义，因为对于任何 $T \in \mathrm{Hom}(\rho, \rho')$有($\lambda A$, ρ', $E(T)\psi$) = ((λA)T, ρ, ψ) = (λ(AT),

ρ，ψ)。为了证明加法可定义，我们首先建立方程(31)是独立于 W_1 和 W_2 的选择。如果 $W_i' \in \mathrm{Hom}(\rho_i, \rho')$是另一个这样的选择，那么：

$$\begin{aligned}
&(A_1 W_1'^* + A_2 W_2'^*, \rho', E(W_1')\psi_1 + E(W_2')\psi_2) \\
&\quad = (A_1 W_1^* + A_2 W_2^* (W_1 W_1'^* + W_2 W_2'^*), \rho', E(W_1')\psi_1 + E(W_2')\psi_2) \\
&\quad = (A_1 W_1^* + A_2 W_2^*, \rho, E(W_1 W_1'^* + W_2 W_2'^*) E(W_1')\psi_1 + E(W_2')\psi_2) \\
&\quad = (A_1 W_1^* + A_2 W_2^*, \rho, E(W_1)\psi_1 + E(W_2)\psi_2)
\end{aligned}$$

为了理解加法是独立于等价类的，设 $T_i \in \mathrm{Hom}(\rho_i, \rho_i')$，$W_i$ 在 $\mathrm{Hom}(\rho_i, \rho)$里等距，以及 W_i'在 $\mathrm{Hom}(\rho_i', \rho)$里等距，那么：

$$\begin{aligned}
&(A_1 T_1, \rho_1, \psi_1) + (A_2 T_2, \rho_2, \psi_2) \\
&\quad = (A_1 T_1 W_1^* + A_2 T_2 W_2^*, \rho, E(W_1)\psi_1 + E(W_2)\psi_2) \\
&\quad = ((A_1 W_1'^* + A_2 W_2'^*)(W_1' T_1 W_1^* + W_2' T_2 W_2^*), \rho, E(W_1)\psi_1 + E(W_2)\psi_2) \\
&\quad = (A_1 W_1'^* + A_2 W_2'^*, \rho, E(W_1') E(T_1)\psi_1 + E(W_2') E(T_2)\psi_2) \\
&\quad = (A_1, \rho_1', E(T_1)\psi_1) + (A_2, \rho_2', E(T_2)\psi_2)
\end{aligned}$$

为了证明在第一个论证里的可加性，选择 $\sigma = \rho \oplus \rho$ 和 $W_i \in \mathrm{Hom}(\rho, \sigma)$相应的等距，那么：

$$\begin{aligned}
(A_1, \rho, \psi) + (A_2, \rho, \psi) &= (A_1 W_1^* + A_2 W_2^*, \sigma, (E(W_1) + E(W_2))\psi) \\
&= (A_1 W_1^* + A_2 W_2^*, \sigma, E(W_1 + W_2)\psi) \\
&= ((A_1 W_1^* + A_2 W_2^*)(W_1 + W_2), \rho, \psi) \\
&= (A_1 + A_2, \rho, \psi)
\end{aligned}$$

最后，证明在第二个论证里的可加性，选择 $\sigma = \rho \oplus \rho$ 和 $W_i \in \mathrm{Hom}(\rho, \sigma)$相应的等距，那么：

$$\begin{aligned}
(A, \rho, \psi_1) + (A, \rho, \psi_2) &= (A W_1^* + A W_2^*, \sigma, E(W_1)\psi_1 + E(W_2)\psi_2) \\
&= (A(W_1^* + W_2^*), \sigma, E(W_1)\psi_1 + E(W_2)\psi_2) \\
&= (A, \rho, E(W_1^* + W_2^*)(E(W_1)\psi_1 + E(W_2)\psi_2)) \\
&= (A, \rho, \psi_1 + \psi_2)
\end{aligned}$$

命题 258。如果我们定义下式，那么复线性空间$\mathfrak{F}_0$ 变成一个代数：

$$(A_1, \rho_1, \psi_1)(A_2, \rho_2, \psi_2) := (A_1 \rho_1(A_2), \rho_1 \otimes \rho_2, \psi_1 \otimes \psi_2) \tag{33}$$

其中 $\psi_1 \in E(\rho_i)$，$i = 1, 2$，而且，$\mathfrak{A}$是$\mathfrak{F}_0$ 的一个子代数，以及$(I, \iota, 1)$的等

价类是一个乘法单位元，其中I是$\mathfrak{A}$的乘法单位元，并且$1\in E(\iota)=\mathbb{C}$。

证明。我们首先证明方程(33)是在$\mathfrak{T}_0$上明确定义的，设$T_i\in \mathrm{Hom}(\rho_i, \rho_i')$。回顾$T_1\times T_2=\rho_1'(T_2)T_1$，我们具有：

$$\begin{aligned}
(A_1T_1, \rho_1, \psi_1)(A_2T_2, \rho_2, \psi_2) &= (A_1T_1\rho_1(A_2T_2), \rho_1\otimes\rho_2, \psi_1\otimes\psi_2)\\
&= (A_1\rho_1'(A_2T_2)T_1, \rho_1\otimes\rho_2, \psi_1\otimes\psi_2)\\
&= (A_1\rho_1'(A_2)\rho_1'(T_2)T_1, \rho_1\otimes\rho_2, \psi_1\otimes\psi_2)\\
&= (A_1\rho_1'(A_2)(T_1\times T_2), \rho_1\otimes\rho_2, \psi_1\otimes\psi_2)\\
&= (A_1\rho_1'(A_2), \rho_1'\otimes\rho_2', E(T_1\times T_2)(\psi_1\otimes\psi_2))\\
&= (A_1\rho_1'(A_2), \rho_1'\otimes\rho_2', E(T_1)\psi_1\otimes E(T_2)\psi_2)\\
&= (A_1, \rho_1', E(T_1)\psi_1)(A_2, \rho_2', E(T_2)\psi_2)
\end{aligned}$$

直接运算证明乘法具有可积性。对于分配性，设$W_i\in\mathrm{Hom}(\rho_i, \rho)$，那么：

$$\begin{aligned}
&[(A_1, \rho_1, \psi_1)+(A_2, \rho_2, \psi_2)](A_3, \rho_3, \psi_3)\\
&\quad= ((A_1W_1^*+A_2W_2^*)\rho(A_3), \rho\otimes\rho_3, (E(W_1)\psi_1+E(W_2)\psi_2)\otimes\psi_3)\\
&\quad= (A_1\rho_1(A_3)W_1^*+A_2\rho_2(A_3)W_2^*, \rho\otimes\rho_3,\\
&\qquad (E(W_1)\psi_1)\otimes\psi_3+(E(W_2)\psi_2)\otimes\psi_3)\\
&\quad= (A_1\rho_1(A_3)(W_1^*\times I_{\rho_3})+A_2\rho_2(A_3)(W_2^*\times I_{\rho_3}),\\
&\qquad \rho\otimes\rho_3, (E(W_1)\psi_1)\otimes\psi_3+(E(W_2)\psi_2)\otimes\psi_3))\\
&\quad= (A_1, \rho_1, \psi_1)(A_3, \rho_3, \psi_3)+(A_2, \rho_2, \psi_2)(A_3, \rho_3, \psi_3)
\end{aligned}$$

我们需要下述线性代数基本引理。

定义 259。如果H，H'是希尔伯特空间，并且$S\in\mathrm{Hom}(H\otimes H', \mathbb{C})$，那么我们通过下面的设定来定义一个反线性映射$\mathcal{J}S$: $H\to H'$为：

$$((\mathcal{J}S)x, x')=S(x\otimes x'), \qquad \forall x\in H, \forall x'\in H'$$

引理 260。

1. $\mathcal{J}$是反线性的：$\mathcal{J}(\lambda S)=\overline{\lambda}(\mathcal{J}S)$和$\mathcal{J}(S_1+S_2)=\mathcal{J}S_1+\mathcal{J}S_2$。
2. 如果$T\in\mathrm{Hom}(H', H)$，那么：

$$T\circ(\mathcal{J}S)=\mathcal{J}(S\circ(I_H\otimes T^*))$$
$$(\mathcal{J}S)\circ T=\mathcal{J}(S\circ(T\circ I_{H'}))$$

3. 如果$S'\in\mathrm{Hom}(H'\otimes H'', \mathbb{C})$，那么$(\mathcal{J}S')\circ(\mathcal{J}S)=(S\otimes I_{H''})\circ(I_H\otimes S'^*)$。

4. 设 $S_1 \in \mathrm{End}(H_1 \otimes H_1', \mathbb{C})$ 和 $S_2 \in \mathrm{End}(H_2 \otimes H_2', \mathbb{C})$，那么：

$$(\mathcal{J}S_2 \otimes \mathcal{J}S_1) \circ \Sigma_{H_1,H_2} = \mathcal{J}[S_1 \circ (I_{H_1} \otimes S_2 \otimes I_{H_1'})]$$

证明。直接得证，在基础线性代数中这是一个好练习。

注释 261。我们将把上述引理用到超希尔伯特空间。但是我们会取 $\Sigma_{H,H'}$ 为有限维希尔伯特空间的范畴 $\mathcal{H}_f$ 上的普通对称。

引理 262。设 $T \in \mathrm{Hom}(\rho, \rho')$，并且选择关于 ρ 和 ρ' 的共轭方程的解 $(\bar{\rho}, R, \overline{R})$ 和 $(\bar{\rho}', R', \overline{R}')$，那就是说，$R \in \mathrm{Hom}(\iota, \bar{\rho} \otimes \rho)$，$\overline{R} \in \mathrm{Hom}(\iota, \rho \otimes \bar{\rho})$，使得 $(\overline{R}^* \times I_\rho) \circ (I_\rho \times R) = I_\rho$，$(R^* \times I_{\bar{\rho}}) \circ (I_{\bar{\rho}} \times \overline{R}) = I_{\bar{\rho}}$，并且对 R' 和 $\overline{R}'$ 有类似情况。设：

$$\overline{T} := (I_{\bar{\rho}} \times \overline{R}'^*) \circ (I_{\bar{\rho}} \times T \times I_{\bar{\rho}'}) \circ (R \times I_{\bar{\rho}'}) = \bar{\rho}(\overline{R}'^* T)R$$

那么 $T \in \mathrm{Hom}(\bar{\rho}', \bar{\rho})$ 和：

$$(I_{\bar{\rho}} \times T) \circ R = (\overline{T} \times I_{\rho'}) \circ R' \tag{34}$$

$$(I_\rho \times \overline{T}^*) \circ \overline{R} = (T^* \times I_{\bar{\rho}'}) \circ \overline{R}' \tag{35}$$

证明。对方程(34)，我们有：

$$(\overline{T} \times I_{\rho'}) \circ R' = \overline{T}R' = \bar{\rho}(\overline{R}'^* T)RR' = \bar{\rho}(\overline{R}'^* T_\rho(R'))R$$

$$= \bar{\rho}(\overline{R}'^* \rho'(R')T)R$$

其中我们对第一个等式使用 $\times$ 的定义，对第二个等式使用 $\overline{T}$ 的定义，对第三个等式用 $R \in \mathrm{Hom}(\iota, \bar{\rho} \otimes \rho)$，以及对第四个等式用 $R \in \mathrm{Hom}(\rho, \rho')$。但是由共轭方程，$\overline{R}'^* \rho'(R') = (\overline{R}'^* \times I_{\rho'}) \circ (I_{\rho'} \times R') = I_{\rho'} = I$，由此有 $(\overline{T} \times I_{\rho'}) \circ R' = \bar{\rho}(T)R = (I_{\bar{\rho}} \times T) \circ R$。对于方程(35)，我们有：

$$(I_\rho \times \overline{T}^*) \circ \overline{R} = \rho(\overline{T}^*)\overline{R} = \rho(R^*)\rho\bar{\rho}(T^*\overline{R}')\overline{R} = \rho(R^*)\overline{R}T^*\overline{R}' \tag{36}$$

其中我们对第二个等式用了 $\overline{T}$ 的定义，而对第三个等式用了 $\overline{R} \in \mathrm{Hom}(\iota, \rho \otimes \bar{\rho})$。然后由共轭方程，有 $\rho(R^*)\overline{R} = (I_\rho \times R^*) \circ (\overline{R} \times I_\rho) = I_\rho$，由此有 $(I_\rho \times \overline{T}^*) \circ \overline{R} = T * \overline{R}' = (T^* \times I_{\bar{\rho}'}) \circ \overline{R}'$。

命题 263。代数 $\mathfrak{F}_0$ 变成一个 $*$ 代数，如果我们定义：

$$(A, \rho, \psi)^* := (R^* \bar{\rho}(A)^*, \bar{\rho}, \mathcal{J}E(\overline{R}^*)\psi) \tag{37}$$

其中 $\psi \in E(\rho)$，并且 $(\bar{\rho}, R, \overline{R})$ 是一个对 ρ 的共轭。

证明。我们首先证明 $*$ 的定义是独立于对 ρ 共轭的选择，为此，设 $(\bar{\rho}_1,$

R_1，$\overline{R}_1$)是任何其他选择，定义 $W \in \mathrm{Hom}(\overline{\rho}, \overline{\rho}_1)$ 为：

$$W := (R^* \times I_{\overline{\rho}_1}) \circ (I_{\overline{\rho}} \times \overline{R}_1) = R^* \overline{\rho}(\overline{R}_1) \tag{38}$$

我们由共轭方程有：

$$W^{-1} := (R_1^* \times I_{\overline{\rho}}) \circ (I_{\overline{\rho}_1} \times \overline{R}) = R_1^* \overline{\rho}_1(\overline{R})$$

再者，

$$\begin{aligned}(R_1^* \overline{\rho}_1(A)^*, \overline{\rho}_1, \mathcal{J}E(\overline{R}_1^*)\psi) &= (R^* W^{-1} \overline{\rho}_1(A)^*, \overline{\rho}_1, \mathcal{J}E(\overline{R}^*(I_\rho \times W^*))\psi) \\ &= (R^* \overline{\rho}(A)^*, \overline{\rho}, E(W^{-1})\mathcal{J}E(\overline{R}^*(I_\rho \times W^*))\psi) \\ &= (R^* \overline{\rho}(A)^*, \overline{\rho}, \mathcal{J}E(\overline{R}^*)\psi)\end{aligned}$$

其中我们对最后一个等式用了引理260.3。

为了搞清楚 $*$ 的定义跟等价类无关，假定 $T \in \mathrm{Hom}(\rho, \rho')$ 和 $\psi \in E(\rho)$，那么：

$$\begin{aligned}(AT, \rho, \psi)^* &= (R^* \overline{\rho}(T^* A^*), \overline{\rho}, \mathcal{J}E(\overline{R}^*)\psi) \\ &= (R'^* \overline{T}^* \overline{\rho}(A^*), \overline{\rho}, \mathcal{J}E(\overline{R}^*)\psi) \\ &= (R'^* \overline{\rho}'(A^*)\overline{T}^*, \overline{\rho}, \mathcal{J}E(\overline{R}^*)\psi) \\ &= (R'^* \overline{\rho}'(A^*), \overline{\rho}', E(\overline{T}^*)\mathcal{J}E(\overline{R}^*)\psi) \\ &= (R'^* \overline{\rho}'(A^*), \overline{\rho}', \mathcal{J}E(\overline{R}^* \circ (I_\rho \times \overline{T}))\psi) \\ &= (R'^* \overline{\rho}'(A^*), \overline{\rho}', \mathcal{J}E(\overline{R}'^* \circ (T \times I_{\overline{\rho}'}))\psi) \\ &= (R'^* \overline{\rho}'(A^*), \overline{\rho}', \mathcal{J}E(\overline{R}'^*)E(T)\psi) \\ &= (A, \rho', E(T)\psi)^*\end{aligned}$$

其中我们在第二个等式用了方程(34)，对第三个等式用了 $T^* \in \mathrm{Hom}(\overline{\rho}, \overline{\rho}')$ 的事实，第五式用了引理206.2，而第六式用了方程(35)。

我们证明 $*$ 是对合的：

$$\begin{aligned}(A, \rho, \psi)^{**} &= (R^* \overline{\rho}(A), \overline{\rho}, \mathcal{J}E(\overline{R}^*)\psi)^* \\ &= (\overline{R}^* \rho(\overline{\rho}(A)R), \rho, \mathcal{J}E(R^*)\mathcal{J}E(\overline{R}^*)\psi) \\ &= (A\overline{R}^* \rho(R), \rho, \mathcal{J}E(R^*)\mathcal{J}E(\overline{R}^*)\psi) \\ &= (A, \rho, E((R^* \times I_{\overline{\rho}})(I_{\overline{\rho}} \times \overline{R}))\psi) \\ &= (A, \rho, \psi)\end{aligned}$$

其中我们对倒数第二式用了引理260.3，以及对最后一个等式用了共轭方程。

为了证明 * 是反线性的，设 $W_i \in \mathrm{Hom}(\rho_i,\ \rho)$，那么：

$$[(A_1,\ \rho_1,\ \psi_1)+(A_2,\ \rho_2,\ \psi_1)]^*$$
$$=(A_1W_1^*+A_2W_2^*,\ \rho,\ E(W_1)\psi_1+E(W_2)\psi_2)^*$$
$$=(R^*\bar{\rho}(W_1A_1^*+W_2A_2^*),\ \bar{\rho},\ \mathcal{J}E(\bar{R}^*)(E(W_1)\psi_1+E(W_2)\psi_2)) \tag{39}$$

但是我们可以取：

$$R=(\overline{W}_1\times W_1)\circ R_1+(\overline{W}_2\times W_2)\circ R_2,\quad \bar{R}=(W_1\times\overline{W}_1)\circ\bar{R}_1+(W_2\times\overline{W}_2)\circ\bar{R}_2$$

其中 $\overline{W}_i\in\mathrm{Hom}(\bar{\rho}_i,\ \bar{\rho})$是等距，$\overline{W}_1\overline{W}_1^*+\overline{W}_2\overline{W}_2^*=I_{\bar{\rho}}$。那么方程(39)变成：

$$[(A_1,\ \rho_1,\ \psi_1)+(A_2,\ \rho_2,\ \psi_2)]^*$$
$$=(R_1^*\bar{\rho}_1(A_1^*)\overline{W}_1^*+R_2^*\bar{\rho}_2(A_2^*)\overline{W}_2^*,\ \bar{\rho},\ \mathcal{J}E(\bar{R}_1^*(I_{\rho_1}\times\overline{W}_1^*))\psi_1+\mathcal{J}E(\bar{R}_2^*(I_{\rho_2}\times\overline{W}_2^*))\psi_2)$$
$$=(R_1^*\bar{\rho}_1(A_1^*)\overline{W}_1^*+R_2^*\bar{\rho}_2(A_2^*)\overline{W}_2^*,\ \bar{\rho},\ E(\overline{W}_1)\mathcal{J}E(\bar{R}_1^*)\psi_1+E(\overline{W}_2)\mathcal{J}E(\bar{R}_2^*)\psi_2)$$
$$=(A_1,\ \rho_1,\ \psi_1)^*+(A_2,\ \rho_2,\ \psi_2)^*$$

对第二个等式用了引理260。

最后，我们证明了$[(A_1,\ \rho_1,\ \psi_1)(A_2,\ \rho_2,\ \psi_2)]^*=(A_2,\ \rho_2,\ \psi_2)^*(A_1,\ \rho_2,\ \psi_1)^*$。如果$\rho=\rho_1\otimes\rho_2$和$\rho'=\rho_1'\otimes\rho_2'$，那么我们可以取$R=(I_{\bar{\rho}_2}\times R_1\times I_{\rho_2})\circ R_2$和$\bar{R}=(I_{\rho_1}\times\bar{R}_2\times I_{\bar{\rho}_1})\circ\bar{R}_1$。因此：

$$[(A_1,\ \rho_1,\ \psi_1)(A_2,\ \rho_2,\ \psi_2)]^*=(A_1\rho_1(A_2),\ \rho_1\otimes\rho_2,\ \psi_1\otimes\psi_2)^*$$
$$=(R_2^*\bar{\rho}_2(R_1^*)\bar{\rho}_2\bar{\rho}_1(\bar{\rho}_1(A_2^*)A_1^*),\ \bar{\rho}_2\otimes\bar{\rho}_1,\ \mathcal{J}E(\bar{R}_1^*\circ(I_{\rho_1}\times\bar{R}_2^*\times I_{\bar{\rho}_1}))\psi_1\otimes\psi_2)$$
$$=(R_2^*\bar{\rho}_2(A_1^*)\bar{\rho}_2(R_1^*\bar{\rho}_1(A_1^*)),\ \bar{\rho}_2\otimes\bar{\rho}_1,\ \mathcal{J}E(\bar{R}_1^*\circ(I_{\rho_1}\times\bar{R}_2^*\times I_{\bar{\rho}_1}))\psi_1\otimes\psi_2)$$
$$=(R_2^*\bar{\rho}_2(A_2^*)\bar{\rho}_2(R_1^*\bar{\rho}_1(A_1^*)),\ \bar{\rho}_2\otimes\bar{\rho}_1,\ \mathcal{J}E(\bar{R}_2^*)\psi_2\otimes\mathcal{J}E(\bar{R}_1^*)\psi_1)$$
$$=(R_2^*\bar{\rho}_2(A_2^*),\ \bar{\rho}_2,\ \mathcal{J}E(\bar{R}_2^*)\psi_2)(R_1^*\bar{\rho}_1(A_1^*),\ \bar{\rho}_1,\ \mathcal{J}E(\bar{R}_1^*)\psi_1)$$
$$=(A_2,\ \rho_2,\ \psi_2)^*(A_1,\ \rho_1,\ \psi_1)^*$$

其中第三个等式源自于 $R^*\in\mathrm{Hom}(\bar{\rho}_1\otimes\rho_1,\ \iota)$的事实，而第四个式子源于引理260.4。

命题264。设 E：$\Delta_f\to\mathcal{SH}_f$，是从DHR范畴 Δ_f 到有限维超希尔伯特空间的严格化的范畴$\mathcal{SH}_f$的嵌入函子，那么公式

$$\alpha_g(A,\ \rho,\ \psi)=(A,\ \rho,\ g_\rho\psi) \qquad A\in\mathfrak{A},\ \psi\in E(\rho) \tag{40}$$

定义这样一个群同构 $g\mapsto\alpha_g$，从 E 的固有群 G 到 $\mathrm{Aut}_{\mathfrak{A}}\mathfrak{F}_0$，$\mathfrak{F}_0$ 的 * 自同态的群使$\mathfrak{A}$逐点固定。

证明。由于 g 是一个自然单项变换，$g_\iota=\mathrm{id}_{E(\iota)}=\mathrm{id}_{\mathbb{C}}$，对于任意 $g\in G$，α_g 是很好定义在$\mathfrak{F}_0$ 的，因为对于 $S\in\mathrm{Hom}(\rho,\ \rho')$，有：

$$\begin{aligned}\alpha_g(AS,\ \rho,\ \psi)&=(AS,\ \rho,\ g_\rho\psi)=(A,\ \rho',\ E(S)g_\rho\psi)\\&=(A,\ \rho',\ g(\rho')E(S)\psi)=\alpha_g(A,\ \rho',\ E(S)\psi)\end{aligned}$$

由于 $g_\iota=\mathrm{id}_{\mathbb{C}}$，$\alpha_g$ 使$\mathfrak{A}\subset\mathfrak{F}_0$ 逐点固定，每一个 g_ρ 是线性的，因此 α_g 是线性的。

$$\begin{aligned}&(A_1,\ \rho_1,\ g_{\rho_1}\psi_1)(A_2,\ \rho_2,\ g_{\rho_2}\psi_2)\\&=(A_1\rho_1(A_2),\ \rho_1\otimes\rho_2,\ (g_{\rho_1}\otimes g_{\rho_2})(\psi_1\otimes\psi_2))\end{aligned}$$

但是 $g_{\rho_1\otimes\rho_2}=g_{\rho_1}\otimes g_{\rho_2}$，所以：

$$\begin{aligned}&(A_1,\ \rho_1,\ g_{\rho_1}\psi_1)(A_2,\ \rho_2,\ g_{\rho_2}\psi_2)\\&=(A_1\rho_1(A_2),\ \rho_1\otimes\rho_2,\ g_{\rho_1\otimes\rho_2}(\psi_1\otimes\psi_2))\end{aligned}$$

因此，

$$\alpha_g(F_1)\alpha_g(F_2)=\alpha_g(F_1F_2) \tag{41}$$

为了证明 α_g 是一个 * 同构，回顾：

$$(A,\ \rho,\ g_\rho\psi)^*=(R^*\bar\rho(A)^*,\ \bar\rho,\ \mathcal{J}E(\bar R^*)g_\rho\psi) \tag{42}$$

如果 $\bar\psi\in E(\bar\rho)$，那么 $E(\bar R^*)(g_\rho\psi\otimes g_{\bar\rho}\bar\psi)=E(\bar R^*)((g_\rho\otimes g_{\bar\rho})(\psi\otimes\bar\psi))$，而且，

$$E(\bar R^*)(g_\rho\psi\otimes g_{\bar\rho}\bar\psi)=g_\iota E(\bar R^*)(\psi\otimes\bar\psi)=E(\bar R^*)(\psi\otimes\bar\psi)$$

因此 $g_{\bar\rho}^*\mathcal{J}E(\bar R^*)g_\rho=\mathcal{J}E(\bar R^*)$，并且由于 $g_{\bar\rho}$是幺正的，我们从式(42)可以得出：

$$\begin{aligned}(A,\ \rho,\ g_\rho\psi)^*&=(R^*\bar\rho(A)^*,\ \bar\rho,\ \mathcal{J}E(\bar R^*)g_\rho\psi)\\&=(R^*\bar\rho(A)^*,\ \bar\rho,\ g_{\bar\rho}\mathcal{J}E(\bar R^*)\psi)\end{aligned}$$

因此：

$$\alpha_g(F^*)=\alpha_g(F)^*, \qquad F\in\mathfrak{F} \tag{43}$$

方程(41)和(43)证明 α_g 是一个 * 同构，其逆明显是 $\alpha_{g^{-1}}$，因此 α_g 由方程(40)定义为 $\mathrm{Aut}_{\mathfrak{A}}\mathfrak{F}$的一个元。映射 $g\mapsto\alpha_g$ 显然是一个同态。

由于 G 是一个紧致群，对于任何 $g\neq e$，存在一个 $(H,\ \pi)\in \mathrm{Rep}_f G$，使得 $\pi(g)\neq \mathrm{id}_H$。由于函子 E 是一个等价，尤其是本质上是满射的，在那存在一个 $\rho\in \mathrm{Obj}(\Delta_f)$，使得 $E(\rho)$ 是同构于 $(H,\ \pi)$。因此存在 $\psi\in E(\rho)$，使得：

$$\pi(g)\psi = g_\rho\psi\neq\psi$$

定义 $F=(I,\ \rho,\ \psi)$，我们具有 $\alpha_g(F)\neq F$。这证明 $g\mapsto\alpha_g$ 的满射性。

仍然要证明 $G\mapsto \mathrm{Aut}_{\mathfrak{A}}\mathfrak{F}_0$ 是映射的，设 $\alpha\in \mathrm{Aut}_{\mathfrak{A}}\mathfrak{A}_0$，$A\in\mathfrak{A}$ 和 $\psi\in E(\rho)\subset\mathfrak{F}_0$，假设 $\Psi=(I,\ \rho,\ \psi)$，那么：

$$(\alpha(\Psi))A=\alpha(\Psi A)=\alpha(\rho(A)\Psi)=\rho(A)\alpha(\Psi)$$

容易证明，这意味着 $\alpha(\Psi)$ 是具有 $\psi'\in E(\rho)$ 的形式 $(I,\ \rho,\ \psi')$。因此 $\psi\mapsto\psi'$ 是 $E(\rho)$ 到我们用 g_ρ 表示的 $E(\rho)$ 的线性映射，然后还需要证明的是，$g=(g_\rho)_{\rho\in\Delta_f}$ 是 E 的幺半群的自然变换。对于 $S\in \mathrm{Hom}(\rho,\ \rho')$，我们具有：

$$(S,\ \rho,\ g_\rho\psi)=\alpha(S,\ \rho,\ \psi)=\alpha(I,\ \rho',\ E(S)\psi)=(I,\ \rho',\ g_{\rho'}E(S)\psi)$$

因此：

$$E(S)g_\rho\psi=g_{\rho'}E(S)\psi,\qquad \psi\in E(\rho)$$

即：

$$E(S)g_\rho=g_{\rho'}E(S)$$

并且 $g\in \mathrm{Nat}E$。为了检查单项性，选择任意的 $\psi_i\in E(\rho_i)$，并设 $\Psi_i=(I,\ \rho_i,\ \psi_i)$，那么：

$$\cdots\cdots\cdots\cdots g_{\rho_1\otimes\rho_2}(\psi_1\otimes\psi_2)=\alpha(\Psi_1\Psi_2)$$
$$=\alpha(\Psi_1)\alpha(\Psi_2)=(g_{\rho_1}\otimes g_{\rho_2})(\psi_1\psi_2)$$

因此，$g\in \mathrm{Nat}_\otimes E$。剩下来要证明的是 g 是幺正的，对于 ψ，$\psi'\in E(\rho)$ 和 $\Psi=(I,\ \rho,\ \psi)$，$\Psi'=(I,\ \rho,\ \psi')$，我们有：

$$\langle g_\rho\psi,\ g_\rho\psi'\rangle_{E(\rho)}I=\alpha(\Psi)^*\alpha(\Psi')=\alpha(\Psi^*\Psi')=\langle\psi,\ \psi'\rangle_{E(\rho)}I$$

其中第一和最后一个等式根据的是命题270。因此 g_ρ 对于每一个 $\rho\in \mathrm{Obj}(\Delta_f)$ 是幺正的。因此，任何 $\alpha\in \mathrm{Aut}_{\mathfrak{A}}\mathfrak{F}_0$ 具有 $g\in G=\mathrm{Nat}_\otimes E$ 的形式 α_g。

定义 265。给定一个对顶锥 O，我们定义 $\mathfrak{F}_0(O)$ 由那些 $\mathfrak{F}_0$ 里的元素 F 构成，使得存在 $A\in\mathfrak{A}(O)$，在 O 里局域化的 $\rho\in \mathrm{Obj}(\Delta_f)$ 以及具有 $F=(A,\ \rho,\ \psi)$ 的 $\psi\in E(\rho)$。

命题266。 $\mathfrak{F}_0(O)$是$\mathfrak{F}_0$的一个$*$子代数。

证明。设$F_1=(A_1, \rho_1, \psi_1)$和$F_2=(A_2, \rho_2, \psi_2)$是在$\mathfrak{F}_0(O)$里，因此，$A_i$能够从$\mathfrak{A}(O)$中被选择，而$\rho_i$能够在$O$里被选择局域化。由于$\rho_1(\mathfrak{A}(O))\subseteq \mathfrak{A}(O)$，因此：

$$F_1F_2=(A_1\rho_1(A_2), \rho_1\otimes\rho_2, \psi_1\otimes\psi_2)$$

也是在$\mathfrak{F}_0(O)$里。由于可输运性，$\bar{\rho}$能够在O里被选择局域化，并且在此情况下$\bar{\rho}\otimes\rho$在O里被局域化。由引理155，$R\in\mathfrak{A}(O)$，因此：

$$F^*=(R^*\bar{\rho}(A)^*, \bar{\rho}, \mathcal{J}E(\overline{R}^*)\psi)$$

是在$\mathfrak{F}_0(O)$里。类似地，在命题257里定义的加法之下$\mathfrak{F}_0(O)$是封闭的，因为ρ也能够在O里被选择局域化，从而等距W_1，W_2是在$\mathfrak{A}(O)$里（根据引理155）。

命题267。 在$\mathfrak{F}_0$上G的作用使$\mathfrak{F}_0(O)$整体上不动。

证明。如果$F\in\mathfrak{F}_0(O)$，那么，对某些$A\in\mathfrak{A}(O)$有$F(A, \rho, \psi)$，并且ρ在O里局域化，那么显然$\alpha_g(F)=(A, \rho, g_\rho\psi)$是在$\mathfrak{F}_0(O)$里。

注释268。 定义好超群(G, k)的作用之后，元素$k\in G$引入一个在$\mathfrak{F}_0$分级和局域代数$\mathfrak{F}_0(O)$上分级的$\mathbb{Z}_2$。

命题269。 场网$\mathfrak{F}_0$满足正常对易关系，那就是说，如果O_1和O_2是类空的，并且$F_i\in\mathfrak{F}(O_i)$使得：

$$\alpha_k(F_i)=\sigma_iF_i$$

那么：

$$F_1F_2=(-1)^{(1-\sigma_1)(1-\sigma_2)/4}F_2F_1$$

证明。选择具有$A_i\in\mathfrak{A}(O_i)$的$F_i=(A_i, \rho_i, \psi_i)$，并且$\rho_i$在$O_i$里局域化，那么$A_1A_2=A_2A_1$，$\rho_1(A_2)=A_2$，$\rho_2(A_1)=A_1$，$\varepsilon_{\rho_1,\rho_2}=id\rho_1\otimes\rho_2$。按照$G$作用在$\mathfrak{F}_0$的方式，我们有：

$$\sigma_i(A_i, \rho_i, \psi_i)=\alpha_k(A_i, \rho_i, \psi_i)=(A_i, \rho_i, k_{\rho_i}\psi_i)$$

由此有$k_{\rho_i}\psi_i=\sigma_i\psi_i$。那就是，$\psi_i$是齐次的，并且$\omega(\psi_i)=\sigma_i$。而且，由于$E$是一个对称函子$E(\varepsilon_{\rho_1,\rho_2})=\sum_{E(\rho_1),E(\rho_2)}$，其中$\sum_{H,H'}$是在$\mathcal{SH}_f$上对称的，从而有：

$$\sum_{H,H'}(\psi_1\otimes\psi_2)=(-1)^{(1-\sigma_1)(1-\sigma_2)/4}(\psi_2\otimes\psi_1)$$

因此：

$$F_1F_2=(A_1\rho_1(A_2), \rho_1\otimes\rho_2, \psi_1\otimes\psi_2)$$

$$
\begin{aligned}
&= (A_1A_2\varepsilon_{\rho_2,\rho_1},\rho_1\otimes\rho_2,\psi_1\otimes\psi_2)\\
&= (A_1A_2,\rho_2\otimes\rho_1,E(\varepsilon_{\rho_2,\rho_1})(\psi_1\otimes\psi_2))\\
&= (A_2\rho_2(A_1),\rho_2\otimes\rho_1,E(\varepsilon_{\rho_2,\rho_1})(\psi_1\otimes\psi_2))\\
&= (A_2\rho_2(A_1),\rho_2\otimes\rho_1,\textstyle\sum_{E(\rho_2),E(\rho_1)}(\psi_1\otimes\psi_2))\\
&= (-1)^{(1-\sigma_1)(1-\sigma_2)/4}(A_2\rho_2(A_1),\rho_2\otimes\rho_1,\psi_2\otimes\psi_1)\\
&= (-1)^{(1-\sigma_1)(1-\sigma_2)/4}F_2F_1
\end{aligned}
$$

命题270。 对于所有的 $\Psi=(I,\ \rho,\ \psi)$，$\Psi'=(I,\ \rho,\ \psi')$，其中 $\psi,\ \psi'\in E(\rho)$，我们有：

$$\Psi A=\rho(A)\Psi \tag{44}$$

$$\Psi^*\Psi'=\langle\psi,\ \psi'\rangle_{E(\rho)} \tag{45}$$

对于 $E(\rho)$ 的任何正交基 $\{\psi_i:\ i=1,\ \cdots,\ n\}$，我们有：

$$\sum_{i=1}^{n}\Psi_i\Psi_i^*=I \tag{46}$$

证明。

$$(I,\ \rho,\ \psi)(A,\ \iota,\ 1)=(\rho(A),\ \rho,\ \psi)=(\rho(A),\ \iota,\ 1)(I,\ \rho,\ \psi)$$

由此有(44)。对于(45)，我们检查：

$$
\begin{aligned}
(I,\ \rho,\ \psi)^*(I,\ \rho,\ \psi')&=(R^*,\ \bar\rho\otimes\rho,\ (\mathcal{J}E(\overline{R}^*)\psi)\otimes\psi')\\
&=(I,\ \iota,\ E(R^*)((\mathcal{J}E(\overline{R}^*)\psi)\otimes\psi'))
\end{aligned}
$$

由于 $\mathcal{J}E(\overline{R}^*)$：$E(\rho)\to E(\bar\rho)$ 和 $E(R^*)$：$E(\bar\rho)\otimes E(\rho)\to\mathbb{C}$，因此 $E(R^*)((\mathcal{J}E(\overline{R}^*)\psi)\otimes\psi')$ 是复数。事实上，由于 $\mathcal{J}$ 的定义和引理260.3，有：

$$
\begin{aligned}
E(R^*)((\mathcal{J}E(\overline{R}^*)\psi)\otimes\psi')&=\langle\mathcal{J}E(R^*)\circ\mathcal{J}E(\overline{R}^*)\psi,\ \psi'\rangle_{E(\rho)}\\
&=\langle\mathcal{J}E((R^*\times I_\rho)\circ(I_\rho\times\overline{R}^*))\psi,\ \psi'\rangle_{E(\rho)}\\
&=\langle\psi,\ \psi'\rangle_{E(\rho)}
\end{aligned}
$$

其中最后一个等式依据的是共轭方程。因此，联合前面两个方程我们有：

$$(I,\ \rho,\ \psi)^*(I,\ \rho,\ \psi')=(I,\ \iota,\ \langle\psi,\ \psi'\rangle_{E(\rho)})=\langle\psi,\ \psi'\rangle_{E(\rho)}(I,\ \iota,\ 1)$$

对于方程(46)，我们有：

$$
\begin{aligned}
\sum_i(I,\rho,\psi_i)(I,\rho,\psi_i)^*&=(\rho(R)^*,\rho\otimes\bar\rho,\textstyle\sum\psi_i\otimes\mathcal{J}E(\overline{R}^*)\psi_i)\\
&=(\rho(R)^*,\rho\otimes\bar\rho,E(\overline{R})1)\\
&=(\rho(R)^*\overline{R},\iota,1)=(I,\iota,1)
\end{aligned}
$$

其中第二个等式依据的是$\mathcal{J}$的定义，而最后一个等式依据的是共轭方程。

10.3 场网的完备化

我们现在构造一个*代数$\mathfrak{F}_0$的表征($\mathcal{H}$, π)，并且证明$\pi|_{\mathfrak{A}}$具有非平凡子表征，等价于由真空态ω_0约化的GNS表征。我们做到这些的依据是，从$\mathfrak{A}$到$\mathfrak{F}_0$扩展的状态ω_0，进而采用GNS表征。为了从$\mathfrak{A}$到$\mathfrak{F}_0$扩展状态ω_0，只要证明存在一个正的线性映射m: $\mathfrak{F}_0 \to \mathcal{A}$就可以。

注释271。设$\rho \in \mathrm{Obj}(\Delta_f)$，由于$\Delta_f$是半单的(参见命题218)，$\rho$是一个在$\mathrm{Obj}(\Delta_f)$里的不可约化对象的有限直和$\rho = \rho_1 \oplus \cdots \oplus \rho_n$。因此，存在一个把$P_\iota^\rho \in \mathrm{End}(\rho)$投影到在这个分解里的那些不可约化量的直和，后者是同构于ι的。

命题272。给定$(A, \rho, \psi) \in \mathfrak{F}_0$，定义：

$$m(A, \rho, \psi) := (AP_\iota^\rho, \rho, \psi) \tag{47}$$

那么m: $\mathfrak{F}_0 \to \mathfrak{A}$是一个从$\mathfrak{F}_0$到$\mathfrak{A}$忠实的正线性投影，而且：

$$m(AF) = Am(F) \qquad A \in \mathfrak{A}, \ F \in \mathfrak{F}_0 \tag{48}$$

证明。我们先来证明m是被明确定义了的，如果$T \in \mathrm{Hom}(\rho, \rho')$，那么$TP_\iota^\rho = P_\iota^{\rho'} TP_\iota^\rho = P_\iota^{\rho'} T$，因此：

$$\begin{aligned} m(AT, \rho, \psi) &= (ATP_\iota^\rho, \rho, \psi) = (AP_\iota^{\rho'} T, \rho, \psi) \\ &= (AP_\iota^{\rho'}, \rho', E(T)\psi) = m(A, \rho', E(T)\psi) \end{aligned}$$

就像所要求的那样。m显然是线性的并且满足方程(48)。我们现在证明m是正的。首先，由于ρ是有限维的，ρ包含真空表征的至多有限多个复制，因此，$P_\iota^\rho = \sum_i S_i S_i^*$，其中$S_i \in \mathrm{Hom}(\iota, \rho)$和$S_i^* S_i = \delta_{ij} \mathrm{id}_\iota$。因此：

$$m(A,\rho,\psi) = (AP_\iota^\rho,\rho,\psi) = \sum (AS_i,\iota,E(S_i^*)\psi)$$

然而，$E(S_i^*)\psi = \lambda_i 1$使得：

$$m(A,\rho,\psi) = \sum_i \lambda_i (AS_i,\iota,1) \in \mathfrak{A}$$

由于每一个$\rho \in \mathrm{Obj}(\Delta_f)$是不可约化对象的一个有限直和(命题218)，任何$F \in \mathfrak{F}_0$都可以写成一个有限和$F = \sum_i F_i$，$F_i = (A_i, \rho_i, \psi_i)$，其中$\psi_i \in E(\rho_i)$拥有$\rho_i$不可约化的并且两两不等价，从而可得：

$$m(F^* F) = \sum_{i,j} m(F_i^* F_j) = \sum_i m(F_i^* F_i)$$

因此，为了证明 m 是正的和忠实的，只需考虑 $m(F^*F)$，具有 $F=(A, \rho, \psi)$，$\psi\in E(\rho)$ 和 ρ 不可约化。在此情况下，有：

$$(A, \rho, \psi)^*(A, \rho, \psi)=(R^*\bar{\rho}(A^*A), \bar{\rho}\otimes\rho, \mathcal{J}E(\bar{R}^*)(\psi\otimes\psi))$$

用 $P_\iota^{\bar{\rho}\otimes\rho}=\|RR^*\|^{-1}RR^*=d(\rho)^{-1}RR^*$，我们得到：

$$\begin{aligned}d(\rho)m(F^*F)&=(R^*\bar{\rho}(A^*A)RR^*, \bar{\rho}\otimes\rho, \mathcal{J}E(\bar{R}^*)\psi\otimes\psi)\\&=(R^*\bar{\rho}(A^*A)R, \iota, E(R^*)\mathcal{J}E(\bar{R}^*)\psi\otimes\psi)\end{aligned}$$

现在，已知：

$$E(R^*)\mathcal{J}E(\bar{R}^*)(\psi\otimes\psi)=\langle\mathcal{J}E(R^*)\mathcal{J}E(\bar{R}^*)\psi, \psi\rangle_{E(\rho)}$$

故由引理 260 可得：

$$\begin{aligned}d(\rho)m(F^*F)&=R^*\bar{\rho}(A^*A)R\langle E((\bar{R}^*\times I_\rho)\circ(I_\rho\times R))\psi, \psi\rangle_{E(\rho)}\\&=R^*\bar{\rho}(A^*A)R\langle\psi, \psi\rangle_{E(\rho)}\end{aligned}$$

因此，$m(F^*F)\geqslant 0$ 和 $m(F^*F)=0$ 意味着 $\psi=0$ 或者 $\bar{\rho}(A)R=0$，但是 $\bar{\rho}(A)R=0$，仅当：

$$0=\bar{R}^*\rho\bar{\rho}(A)\rho(R)=A\bar{R}^*\rho(R)=A$$

因此 $m(F^*F)=0$ 意味着 $F=0$，并且 m 是从 $\mathfrak{F}_0$ 到 $\mathcal{A}$ 忠实的正线性投影。

引理 273。设 P_0^ρ 是在 $\mathrm{End}(E(\rho))$ 里投影，到 G 的子空间关于作用 $\pi_\rho(g)=g_\rho$ 的不变矢量，那么 $E(P_0^\rho)=P_0^\rho$。而且，条件期望值 m 是 G 不变量，即对所有的 $g\in G$ 和 $F\in\mathfrak{F}_0$ 有 $m(\alpha_g(F))=m(F)$。

证明。回顾一下，如果 (H, π) 是紧致群 G 的一个不可约化表征，并且 π 也不是平凡表象，那么 H 不包含 G 不变矢量。如果 $\rho=\oplus\rho_i$ 里有 ρ_i 不可约化，那么前面的观测，意味着 $E(\rho)$ 里的 G 不变矢量肯定是在 $E(P_i^\rho)$ 像里面的那些。因此 $E(P_i^\rho)=P_0^\rho$，意味着 $m(F)=\alpha_g(m(F))$。而且：

$$\begin{aligned}m\alpha_g(A, \rho, \psi)&=m(A, \rho, g_\rho\psi)=(AP_\iota^\rho, \rho, g_\rho\psi)=(A, \rho, P_0^\rho g_\rho\psi)\\&=(A, \rho, g_\rho P_0^\rho\psi)=(A, \rho, P_0^\rho\psi)\\&=(AP_l^\rho, \rho, \psi)=m(A, \rho, \psi)\end{aligned}$$

在命题 272 看来，$\omega_0\circ m$ 是在 * 代数 $\mathfrak{F}_0$ 上的一个忠实态。设 $(\mathcal{H}, \pi)$ 是 $\omega_0\circ m$ 约化的 $\mathfrak{F}$ 的 GNS 表征，设 $\mathfrak{F}$ 是 $\pi(\mathfrak{F}_0)$ 的范数封闭的，并且设 $\mathfrak{F}(O)$ 是 $\pi(\mathfrak{F}_0(O))$ 的弱封闭的。显然 $\mathfrak{F}$ 是网 $O\mapsto\mathfrak{F}(O)$ 的 C^* 归纳极限。由于 $\omega_0\circ m$ 通过引理 273 是 G 不变性的，存在 $\mathcal{H}$ 上 G 的一个幺正表征 U，实现 $\mathfrak{F}_0$ 的自同

构 α_g：

$$\pi(\alpha_g(F)) = U(g)\pi(F)U(g)^*, \qquad g\in G,\ F\in\mathfrak{F}_0$$

并因此拓展到$\mathfrak{F}$。由于 $g\mapsto\alpha_g$ 是单射的，U 是单射的。

定义 274。设 $\sigma\in\hat{G}$ 是 G 的一个不可约化特征，定义一个$\mathfrak{B}(\mathcal{H})$上的映射 $\mathscr{E}_\sigma$ 为：

$$\mathscr{E}_\sigma(A) = \int_G \overline{\sigma(g)}U(g)AU(g)^*d\mu(g)$$

其中 μ 是 $U(G)$ 上的哈尔测度。

注释 275。设 $F=(A,\rho,\psi)\in\mathfrak{F}_0$，由于 $U(g)$ 实现 α_g，我们有：

$$\begin{aligned}\mathscr{E}_\sigma(\pi(F)) &= \int_G \overline{\sigma(g)}\Big(\pi(\alpha_g(A,\rho,\psi))\Big)d\mu(g)\\ &= \int_G \overline{\sigma(g)}\pi(A,\rho,g_\rho\psi)d\mu(g) = \pi(A,\rho,P_\sigma^\rho\psi)\end{aligned} \tag{49}$$

其中 $P_\sigma^\rho\in\mathrm{End}(E(\rho))$是正交投影到子空间变换，按照不可约化表征 σ。由于 G 是紧致的，故$\mathscr{E}_\sigma$ 是强连续的。注意到$\mathscr{E}_0(\pi(F))=\pi[m(A,\rho,\psi)]$。

引理 276。$\mathfrak{F}_0(O)\mathfrak{A}=\mathfrak{F}_0$。

证明。设$(A,\rho,\psi)\in\mathfrak{F}_0$，由于 ρ 是可输运的，存在一个拥有 ρ'在 O 里局域化的幺正 $T\in\mathrm{Hom}(\rho,\rho')$，那么：

$(A,\rho,\psi)=(AT^*,\rho',E(T)\psi)=(AT^*,\iota,1)(I,\rho',E(T)\psi)=BF$，

其中 $B\in\mathfrak{A}$和 $F\in\mathfrak{F}_0(O)$。因此$\mathfrak{A}\mathfrak{F}_0(O)=\mathfrak{F}_0$。由于$\mathfrak{A}$，$\mathfrak{F}_0(O)$以及$\mathfrak{F}_0$ 是 * 代数，故$\mathfrak{F}_0(O)\mathfrak{A}=\mathfrak{F}_0$。

定理 277。$(\pi,\mathcal{H},\mathfrak{F},(G,k))$是具有正常对易关系的、拥有对$(\mathfrak{A},\omega_0)$规范对称的一个场系统（在定义 229 和定义 245 的意义上）。

证明。显然$\mathfrak{F}(O)$是$\mathfrak{F}$的一个 G 稳定冯·诺伊曼子代数，网 $O\mapsto\mathfrak{F}(O)$也满足正规对易关系。现在我们检查一遍定义 229 里的个别条件。

(γ)我们需要证明在 G 作用下的$\mathfrak{F}(O)$的不动点代数是 $\pi(\mathfrak{A}(O))$，首先要注意到$\mathcal{E}(\pi(\mathfrak{F}_0(O)))=\pi(m(\mathfrak{F}_0(O)))$。因此，

$$\begin{aligned}\mathfrak{F}(O)^G &= \mathscr{E}(\mathfrak{F}(O))\\ &= \mathscr{E}\Big(\overline{\pi(\mathfrak{F}_0(O))}\Big) = \overline{\mathscr{E}(\pi(\mathfrak{F}_0(O)))} = \overline{\pi(m(\mathfrak{F}_0(O)))}\\ &= \overline{\pi(\mathfrak{A}(O))}\end{aligned}$$

第三个等式根据的是$\mathscr{E}$的正规性，而最后一个等式是取决于 m 是从$\mathfrak{F}_0$ 到$\mathfrak{A}$的一个条件期望这一事实。

(δ)设j: $\mathfrak{F}_0 \to \mathcal{H}$是源自于$\omega_0 \circ m$ 的 GNS 表征的包含映射，由于$\overline{j(\mathfrak{A})} = \mathcal{H}_0$我们有：

$$\overline{\mathfrak{F}(O)\mathcal{H}_0} = \overline{\pi(\mathfrak{F}_0(O))\mathcal{H}_0} = \overline{\pi(\mathfrak{F}_0(O))j(\mathfrak{A})} = \overline{j(\mathfrak{F}_0(O)\mathfrak{A})} = \overline{j(\mathfrak{F}_0)} = \mathcal{H}$$

(ε)设 O_1 和 O_2 是类空分离的，$\mathfrak{F}_0$ 的子代数$\mathfrak{A}(O_1)$是在规范变换下逐点不变量。特别是，对于所有 $A \in \mathfrak{A}(O)$，有 $\alpha_k(A) = A$，即$\mathfrak{A}(O_1)$的元素是纯玻色子的。因此，由对易关系的正规性得到相对局域性(命题 269)。

现在我们断言 $\mathrm{Aut}_{\mathfrak{A}}\mathfrak{F} = G$，根据方程(49)，$\mathfrak{F}_\sigma(\pi(\mathfrak{F}_0))$是作为一个巴拿赫空间同构于$\mathfrak{A} \otimes P_\sigma^\rho E(\rho)$的，从而也是$\mathfrak{F}$的闭合子空间，由此可得：

$$\mathscr{E}_\sigma(\mathfrak{F}) = \mathscr{E}_\sigma(\overline{\pi(\mathfrak{F}_0)}) = \overline{\mathscr{E}_\sigma(\pi(\mathfrak{F}_0))} = \mathscr{E}_\sigma(\pi(\mathfrak{F}_0))$$

由于对任意 $F \in \mathfrak{F}$，我们都有 $F = \sum_{\sigma \in \hat{G}} \mathscr{E}_\sigma(F)$，以及$\mathscr{E}_\sigma(F) \in \pi(\mathfrak{F}_0)$，因此一个元素 $F \in \mathfrak{F}$是处于 $\pi(\mathfrak{F}_0)$里的，当且仅当，对于仅有的有限多个 $\sigma \in \hat{G}$有$\mathfrak{F}_\sigma(F) \neq 0$。结合 α 的线性性，这就意味着 $\alpha(\pi(\mathfrak{F}_0)) \subseteq \pi(\mathfrak{F}_0)$。因此存在一个 $g \in G$，使得 $\alpha|_{\pi(\mathfrak{F}_0)} = \alpha_g$(根据命题 264)。由于 α_g 是连续的，并且 $\pi(\mathfrak{F}_0)$在$\mathfrak{F}$里是稠的，故 α 是 α_g 到$\mathfrak{F}$的唯一扩展。

10.4 场网的庞加莱协变性

协变性考虑在第 7 节的 DHR 理论，或者在场网$\mathfrak{F}$里没有起到重要作用，现在我们证明后者是庞加莱协变的，如果基础 DHR 截面是如此的话。(回顾在特定环境下的评论 224，我们具有 $\Delta_{fc} = \Delta_f$。)

定理 278。如果在场网$\mathfrak{F}$的构造里，我们是从范畴 Δ_{fc}而不是 Δ_f出发，上面构造的场网在$\hat{P}$的自同态作用下是协变的，这个作用是靠$\mathfrak{F}$的 GNS 表征空间上的一个正能量表示实现的，它是对应于态 $\omega_0 \circ m$ 的。

证明。设$\beta_h = \mathrm{Ad}U(h)$是在$\mathfrak{A}$上的$\mathcal{P}$的作用，回顾注释 221 对于所有 $h \in \hat{\mathcal{P}}$都有 $\rho_h = \beta_h \circ \rho \circ \beta_h^{-1}$ 和 $X_\rho(h) \equiv U(h)U_\rho(h)^* \in \mathrm{Hom}(\rho, \rho_h)$。我们定义在$\mathfrak{F}_0$ 上的$\hat{\mathcal{P}}$的一个作用$\hat{\beta}$为：

$$\hat{\beta}_h((A, \rho, \psi)) \equiv (\beta_h(A), \rho_h, E(X_\rho(h))\psi)$$

$$=(\beta_h(A)X_\rho(h),\ \rho,\ \psi)=(U(h)AU_\rho(h)^*,\ \rho,\ \psi)\quad(50)$$

设ρ，$\rho'\in\Delta_{fc}$和$T\in\mathrm{Hom}(\rho,\ \rho')$，那么$\beta_h(T)\in\mathrm{Hom}(\rho_h,\ \rho_h')$，以及$TU_\rho(h)=U_{\rho'}(h)T$(参看8.4节)，因此：

$$\beta_h(T)X_\rho(h)=(U(h)TU(h)^*)(U(h)U_\rho(h)^*)=U(h)TU_\rho(h)^*$$
$$=U(h)U_{\rho'}(h)T=X_{\rho'}(h)T$$

使用这个方程，我们计算：

$$\begin{aligned}\hat{\beta}_k((AT,\ \rho,\ \psi))&=(\beta_h(AT),\ \rho_h,\ E(X_\rho(h))\psi)\\&=(\beta_h(A),\ \rho_h',\ E(\beta_h(T)X_\rho(h))\psi)\\&=(\beta_h(A),\ \rho_h',\ E(X_{\rho'}(h)T)\psi)\\&=\hat{\beta}_h((A,\ \rho',\ E(T)\psi)),\end{aligned}$$

因此$\hat{\beta}_g$被明确定义。设i：$A\mapsto(A,\ \iota,\ 1)$是在$\mathfrak{F}$里的$\mathfrak{A}$的包含。那么$\hat{\beta}_h\circ i=i\circ\beta_g$，因此$\hat{\beta}_g$拓展$\beta_g$。如果$F\in\mathfrak{F}(O)$，那么存在一个拥有$A\in\mathfrak{A}(O)$和$\rho\in\Delta(O)$的表征$F=(A,\ \rho,\ \psi)$。现在从定义明显有$\hat{\beta}_h(F)\in\mathfrak{F}(hO)$。从方程(50)的右手边明显有$g\mapsto\hat{\beta}_g$是一个群同态。现在，已知：

$$\begin{aligned}&\hat{\beta}_g((A_1,\ \rho_1,\ \psi_1)(A_2,\ \rho_2,\ \psi_2))=\hat{\beta}_g((A_1\rho_1(A_2),\ \rho_1\rho_2,\ \psi_1\otimes\psi_2))\\&=(U(h)A_1\rho_1(A_2)U_{\rho_1\rho_2}(h)^*,\ \rho_1\rho_2,\ \psi_1\otimes\psi_2)\\&=(\beta_h(A_1)\rho_{1,h}(\beta_h(A_2))U(h)U_{\rho_1\rho_2}(h)^*,\ \rho_1\rho_2,\ \psi_1\otimes\psi_2)\\&=(\beta_h(A_1)\rho_{1,h}(\beta_h(A_2))X_{\rho_1\rho_2}(h),\ \rho_1\rho_2,\ \psi_1\otimes\psi_2)\\&=(\beta_h(A_1)\rho_{1,h}(\beta_h(A_2))X_{\rho_1}(h)\rho_1(X_{\rho_2}(h)),\ \rho_1\rho_2,\ \psi_1\otimes\psi_2)\\&=(\beta_h(A_1)X_{\rho_1}(h)_{\rho_1}(\beta_h(A_2)X_{\rho_2}(h)),\ \rho_1\rho_2,\ \psi_1\otimes\psi_2)\\&=(U(h)A_1U_{\rho_1}(h)^*\rho_1(U(h)A_2U_{\rho_2}(h)^*),\ \rho_1\rho_2,\ \psi_1\otimes\psi_2)\\&=(U(h)A_1U_{\rho_1}(h)^*,\ \rho_1,\ \psi_1)(U(h)A_2U_{\rho_2}(h)^*,\ \rho_2,\ \psi_2)\\&=\hat{\beta}_g((A_1,\ \rho_1,\ \psi_1))\hat{\beta}_g((A_2,\ \rho_2,\ \psi_2))\end{aligned}$$

其中第五个等式是根据方程(27)得到的。因此$\hat{\beta}_g$是代数同态。

设$\rho\in\Delta_{fc}$，并且选择一个共轭($\bar{\rho}$，R，$\bar{R}$)。由于平凡态射ι是关于$X_\iota=\mathrm{id}_\iota$协变的，应用具有$T=R^*\in\mathrm{Hom}(\bar{\rho}\rho,\ \iota)$方程(50)，我们得到$R^*=\beta_h(R^*)$

$X_{\bar{\rho}\rho}(h)=\beta_h(R^*)X_{\bar{\rho}}(h)\bar{\rho}(X_\rho(h))$，其中再次用了方程(27)。这就等价于：

$$R^*\bar{\rho}(X_\rho(h)^*)=\beta_h(R^*)X_{\bar{\rho}}(h) \tag{51}$$

这在下面可能会用到。现在我们计算：

$$\begin{aligned}
(\hat{\beta}_h(A,\rho,\psi))^* &= (U(h)AU_\rho(h)^*,\rho,\psi)^* \\
&= (R^*\bar{\rho}(U(h)AU_\rho(h)^*)^*,\bar{\rho},\mathcal{J}E(\bar{R}^*)\psi) \\
&= (R^*\bar{\rho}(U_\rho(h)A^*U(h)^*),\bar{\rho},\mathcal{J}E(\bar{R}^*)\psi) \\
&= (R^*\bar{\rho}(U_\rho(h)U(h)^*\beta_h(A^*)),\bar{\rho},\mathcal{J}E(\bar{R}^*)\psi) \\
&= (R^*\bar{\rho}(X_\rho(h)^*\beta_h(A^*)),\bar{\rho},\mathcal{J}E(\bar{R}^*)\psi) \\
&= (\beta_h(R^*)X_{\bar{\rho}}(h)\bar{\rho}(\beta_h(A))^*,\bar{\rho},\mathcal{J}E(\bar{R}^*)\psi) \\
&= (U(h)R^*U_{\bar{\rho}}(h)^*\bar{\rho}(\beta_h(A))^*,\bar{\rho},\mathcal{J}E(\bar{R}^*)\psi) \\
&= (U(h)R^*\bar{\rho}(A)^*U_{\bar{\rho}}(h)^*,\bar{\rho},\mathcal{J}E(\bar{R}^*)\psi) \\
&= \hat{\beta}_h((R^*\bar{\rho}(A)^*,\bar{\rho},\mathcal{J}E(\bar{R}^*)\psi)) \\
&= \hat{\beta}_h((A,\rho,\psi)^*)
\end{aligned}$$

因此$\hat{\beta}_h$是一个 * 同态。在第六个等式里我们用了方程(51)。

根据下面式子：

$$\hat{\beta}_h((A,\rho,\psi))=(U(h)AU_\rho(h)^*,\rho,\psi)$$

$$\alpha_g((A,\rho,\psi))=(A,\rho,\pi_{E(\rho)}(g)\psi)$$

显然对所有 $g\in G$ 和 $h\in\hat{\mathcal{P}}$，具有$\hat{\beta}_h\circ\alpha_g=\alpha_g\circ\hat{\beta}_h$。根据$\pi\circ m=\varepsilon_0\circ\pi$，我们具有$\omega_0\circ\beta_h\circ m=\omega_0\circ m$。因此$\mathfrak{F}$的真空态是$\hat{\mathcal{P}}$不变性的，并且$\hat{\mathcal{P}}$是在 GNS 表征里幺正实现的。

10.5　场网的唯一性

在这一节里我们已经证明，给定纤维函子 E: $\Delta_f(\mathfrak{A})\to\mathcal{SH}$，存在一个具有完备性的正常对易关系的场网，即从真空产生$\Delta_f(\mathfrak{A})$里全部表征。我们叫这为罗伯茨场网并且用$\mathfrak{F}_E^R$表示。我们首先考虑在函子 E 上这个建构的独立性。

命题 279。设 E_1，E_2: $\Delta_f\to\mathcal{H}$是两个纤维函子，那么从它们建构的罗伯茨场网$\mathfrak{F}_{E_1}^R$，$\mathfrak{F}_{E_2}^R$是幺正等价的。

证明。由附录里的定理373，存在两个幺正的幺半群自然同构 α: $E_1 \to E_2$。在此基础上，我们通过γ: $(A, \rho, \psi) \to (A, \rho, \alpha_\rho \psi)$定义一个映射 γ: $\mathfrak{F}^R_{0,1} \to \mathfrak{F}^R_{0,2}$。由于 $\psi \in E_1(\rho)$和 $\alpha \in \mathrm{Hom}(E_1(\rho), E_2(\rho))$，使上述定义具有意义，$\gamma$ 被明确定义，对于 $T \in \mathrm{Hom}(\rho, \rho')$，我们有：

$$\begin{aligned}\gamma(AT, \rho, \psi) &= (AT, \rho, \alpha_\rho \psi)\\ &= (A, \rho', E_2(T) \circ \alpha_\rho \psi)\\ &= (A, \rho', \alpha_\rho \circ E_1(T)\psi)\\ &= \gamma(A, \rho', E_1(T)\psi)\end{aligned}$$

γ 是一个代数同构的根据在于：

$$\begin{aligned}\gamma((A_1, \rho_1, \psi_1))\gamma((A_2, \rho_2, \psi_2)) &= (A_1, \rho_1, \alpha_{\rho_1}\psi_1)(A_2, \rho_2, \alpha_{\rho_2}\psi_2)\\ &= (A_1\rho_1(A_2), \rho_1 \otimes \rho_2, \alpha_{\rho_1}\psi_1 \otimes \alpha_{\rho_2}\psi_2)\\ &= (A_1\rho_1(A_2), \rho_1 \otimes \rho_2, \alpha_{\rho_1 \otimes \rho_2}(\psi_1 \otimes \psi_2))\\ &= \gamma((A_1\rho_1(A_2), \rho_1 \otimes \rho_2, \psi_1 \otimes \psi_2))\\ &= \gamma((A_1, \rho_1, \psi_1)(A_2, \rho_2, \psi_2))\end{aligned}$$

其中我们已经使用了 α 的单项性 $\alpha_{\rho_1 \otimes \rho_2} = \alpha_{\rho_1} \otimes \alpha_{\rho_2}$，由于一个逆能够用自然同构 α^* 得到，γ 是场代数$\mathfrak{F}^R_{0,1}$跟$\mathfrak{F}^R_{0,2}$之间的一个同构，显然涉及局域结构，即$\mathfrak{F}^R_{0,1}(O)$映射到$\mathfrak{F}^R_{0,2}(O)$。

下一步我们断言 $m_2 \circ \gamma = \gamma \circ m_1$，其中，$m_1$，$m_2$ 是以前定义的投影，也就是：

$$m_2 \circ \gamma((A, \rho, \psi)) = m_2((A, \rho, \alpha_\rho \psi)) = (AP^\rho_\iota, \rho, \alpha_\rho \psi) =$$
$$\gamma((A P^\rho_\iota, \rho, \psi)) = \gamma \circ m_1((A, \rho, \psi))$$

这意味着$\mathfrak{F}^R_{0,1}$上的态 $\omega_0 \circ m_1$ 和 $\omega_0 \circ m_2 \circ \gamma$ 一致，由此同构 γ: $\mathfrak{F}^R_{0,1} \to \mathfrak{F}^R_{0,2}$拓展到在 GNS 表征里范数完备性的幺正等价。

为了研究一个任意完备范数正常场网$\mathfrak{F}$，而不是限定的形式$\overline{\mathfrak{F}^R}^{\|\cdot\|}$，我们使用下述命题。

命题280。设$\mathfrak{F}$是一个可观测量网$\mathfrak{A}$的完备正常场网，那么存在一个严格的张量算子 $E_\mathfrak{F}$: $\Delta_f(\mathfrak{A}) \to \mathcal{SH}_f$到有限维超希尔伯特空间的范畴上。就这些对象，$E_\mathfrak{F}$是由下面矢量空间给定的：

$$E_\mathfrak{F}(\rho) := \{F \in \mathfrak{F} \mid F\pi_0(A) = \pi_0(\rho(A))F \quad \forall A \in \mathfrak{A}\}$$

内积是由$\langle F, F'\rangle \mathbf{1} = F^* F'$给定的，而$\mathbb{Z}_2$分级通过$k \in G$的作用给定。对于不可约化的$\rho$，$\rho' \in \Delta_f$，我们具有$E(\varepsilon(\rho, \rho')) = \pm \sum_{i,j} \psi'_i \psi_j \psi_i'^* \psi_j^*$，其中$\{\psi_i, i=1, \cdots, d(\rho)\}$和$\{\psi'_i, i=1, \cdots, d(\rho')\}$是$E(\rho)$和$E(\rho')$各自的正交基，并且负号出现在当且仅当$\rho$和$\rho'$两者都是费米子的情况下。

证明。(在此证明中我们把$E_{\mathfrak{F}}$写成E)对于$s \in \mathrm{Hom}(\rho, \rho')$，我们定义$E(s) = \pi_0(s) \in \mathfrak{F}$。对于$F \in E(\rho)$，我们有：

$$\begin{aligned}\pi_0(s) F \pi_0(A) &= \pi_0(s) \pi_0(\rho(A)) F = \pi_0(s\rho(A)) F \\ &= \pi_0(\rho'(A)s) F = \pi_0(\rho'(A)) \pi_0(s) F\end{aligned}$$

对于所有$A \in \mathfrak{A}$成立，因此$\pi_0(s) F \in E(\rho')$和E是函子。如果F，$F' \in E(\rho)$，那么$F^* F' \in \mathfrak{F} \cap \mathfrak{A}' = \mathbb{C}\mathbf{1}$，允许我们定义$\langle F, F'\rangle \mathbf{1} = F^* F'$，设$s \in \mathrm{Hom}(\rho, \rho')$和$F \in E(\rho)$，$F' \in E(\rho')$，那么：

$$\langle F', E(s)F\rangle = \langle F', \pi_0(s)F\rangle = F'^* \pi_0(s) F = (\pi_0(s)^* F')^* F =$$

$$\langle \pi_0(s^*) F', F'\rangle = \langle E(s^*) F', F'\rangle$$

其中我们用到的π_0是$*$保持的，表明E是$*$保持的。由第9节我们具有$E(\rho) E(\rho') = E(\rho \otimes \rho')$。如果$S_i \in \mathrm{Hom}(\rho_i, \rho'_i)$，$F_i \in E(\rho_i)$，那么：

$$E(S_1 \times S_2) F_1 F_2 = \pi_0(S_1 \rho_1(S_2)) F_1 F_2 = \pi_0(S_1) F_1 \pi_0(S_2) F_2 \in E(\rho'_1 \otimes \rho'_2)$$

故$E(S_1 \times S_2) = E(S_1) \otimes E(S_2)$，因此$E$是一个严格张量算子。场网的完备性结合第9节的讨论意味着E是忠实的，并且满足$E(\rho) = d(\rho)$。(后者也可以由附录里的命题344得出。)最后，设$F \in E(\rho)$，$F' \in E(\rho')$是范数之一，现在设ρ，$\rho' \in \Delta_f$，并设ψ_i，$i=1, 2, \cdots, d(\rho)$和$\psi'_i = 1, 2, \cdots, d(\rho')$是和$E(\rho')$各自的正交基，那么：

$$\tilde{c}(\rho, \rho') = \sum_{i,j} \psi'_i \psi_j \psi_i'^* \psi_j^*$$

是在$\mathfrak{F}^G$里的，并且独立于所选择的基，而且$\tilde{c}(\rho, \rho') \in \mathrm{Hom}(\rho \otimes \rho', \rho' \otimes \rho)$。上述证明的$E$的函子性质意味着$\tilde{c}(\rho, \rho')$在两个论证里都是自然的，如果$\{\rho'' \in \Delta_f\}$和$\psi''_k$，$k=1, 2, \cdots, d(\rho'')$是一个在$E(\rho'')$里的正交基，那么$\{\psi'_j \psi''_k\}$是一个在$E(\rho' \otimes \rho'')$里的正交基，因此：

$$\begin{aligned}\tilde{c}(\rho, \rho' \otimes \rho'') &= \sum_{i,j,k} \psi'_j \psi''_k \psi_j \psi_k''^* \psi_j'^* \psi_i^* \\ &= \sum_{i,j,m,k,l} \psi'_m (\psi''_i \psi_j \psi_i''^* \psi_j^*) \psi_m'^* (\psi'_k \psi_l \psi_k'^* \psi_l^*)\end{aligned}$$

$$= \mathrm{id}_{\rho'} \otimes \tilde{c}(\rho,\rho'') \circ \tilde{c}(\rho,\rho') \otimes \mathrm{id}_{\rho}.$$

这是辫子关系之一。人们容易看出 $\tilde{c}(\rho, \rho')\tilde{c}(\rho', \rho) = \mathbf{1}$，因此 $\tilde{c}(\cdot, \cdot)$是对张量范畴 Δ_f 的对称，如果ρ和ρ'都是不可约化的，并且彼此局域化类空的，对应场的正常对易关系意味着 $\tilde{c}(\rho, \rho) = \pm 1$，其中负号出现在$\rho$和$\rho'$是费米子的情况下。由于，对不可归约的 ρ，ρ'定义 $c(\rho, \rho') = \pm c(\rho, \rho')$，其中我们是在当且仅当$\rho$和$\rho'$都是费米子的情况下取负号，并且通过自然性质扩展 c 到可归约对象。那么，$c(\rho, \rho') = 1$ 只要ρ，ρ'是局域化类空的。目前它是源自于命题200里 $E(\varepsilon(\rho, \rho')) = c(\rho, \rho')$的独特结果。因此 $E_{\mathfrak{F}}$是下面意义上的对称张量函子，即在它把Δ_f的ε映射到希尔伯特空间的范畴$\mathcal{H}$的对称 C 意义上。等效地，E 是进入具有对称 $\tilde{c}$ 的超希尔伯特空间范畴的一个对称张量函子。

因此每一个完备的正常场网$\mathfrak{F}$产生一个严格对称 * 保持纤维函子 $E_{\mathfrak{F}}$，用$\mathfrak{F}^R_{E_{\mathfrak{F}}}$表示跟后者结合的罗伯茨场网，我们的目标是构造同构$\mathfrak{F} \cong \mathfrak{F}^R_{E_{\mathfrak{F}}}$。

定理281。设$\mathfrak{F}$是一个对$\mathfrak{A}$的完备正规场网，并且 $E_{\mathfrak{F}}$: $\Delta_f \to \mathcal{SH}$是源自命题280的纤维函子，那么存在一个场网的幺正等价$\mathfrak{F}_{E_{\mathfrak{F}}} \to \mathfrak{F}$。

证明。根据命题280，存在一个对称 * 保持纤维函子 $E_{\mathfrak{F}}$: $\Delta_f \to \mathcal{SH}$，根据具体的田中定理(附录的定理377)，$E$ 的幺正的幺半群自然变换的紧致群 $G_{E_{\mathfrak{F}}}$是幺正地表示在 $E_{\mathfrak{F}}(\rho)$上。另一方面，伴随我们场网$\mathfrak{F}$的紧致群 G 也作用在这些空间上，提供一个同态 $G \to G_{E_{\mathfrak{F}}}$。这个同态是单射的，因为 G 是作为$\mathfrak{F}$所处的希尔伯特空间 H 上的酉群具体给定的。它也是满射的，否则 $\pi \mid$ 离前面竖线近就会包含那些不在 Δ_f 里的表示，这跟$\mathfrak{F}$是完全场网的假设相矛盾。因此所给定的群 G 能够被等同于一个源自于纤维函子 $E_{\mathfrak{F}}$的重构，对于任何 $\sigma \in \hat{F}$ 我们定义一个就像在定义274里那样的在$\mathfrak{F}$上的投影$\mathcal{E}_{\sigma}$，我们用$\mathfrak{F}_0$ 代表代数直和$\oplus_{\sigma \in \hat{G}} \mathcal{E}_{\sigma}(\mathfrak{F})$，这跟$\{F \in \mathfrak{F} \mid \mathcal{E}_{\sigma}(F) = 0$，对几乎所有 $\sigma \in \hat{G}\}$一样。

我们现在通过 γ: $(A, \rho, \psi) \mapsto \pi_0(A)\psi$ 定义一个映射 γ: $\mathfrak{F}^R_{E_{\mathfrak{F}},0} \to \mathfrak{F}$。乍看上去，这个公式有点奇怪，但是它具有完美的意义，因为 $\psi \in E_{\mathfrak{F}}(\rho)$，其中根据定义 $E_{\mathfrak{F}}(\rho)$是$\mathfrak{F}$的子空间。通常地，γ 是明确定义了的，因为，对于 $T \in \mathrm{Hom}(\rho, \rho')$，有：

$$\gamma((AT, \rho, \psi)) = \pi_0(AT)\psi = \pi_0(A)E_{\mathfrak{F}}(T)\psi = \gamma((A, \rho', E_{\mathfrak{F}}(T)\psi))$$

而且：

$$\begin{aligned}\gamma((A_1, \rho_1, \psi_1)(A_2, \rho_2, \psi_2)) &= \gamma((A_1\rho_1(A_2), \rho_1\rho_1, \psi_1 \otimes \psi_2)) \\ &= \pi_0(A_1\rho_1(A_2))\psi_1\psi_2 \\ &= \pi_0(A_1)\psi_1\pi_0(A_2)\psi_2 \\ &= \gamma((A_1, \rho_1, \psi_1))\gamma((A_2, \rho_2, \psi_2))\end{aligned}$$

其中我们用到$\psi_1 \in E_{\mathfrak{F}}(\rho_1) = \{F \in \mathfrak{F} \mid F\pi_0(A) = \pi_0(\rho(A))F\}$，因此$\gamma$是一个代数同态，跟$(A, \rho, \psi) = (A, \iota, 1)(\mathbf{1}, \rho, \psi)$结合起来，这就意味着$\gamma$是一个$*$同态，假若对$F = (\mathbf{1}, \rho, \psi)$有$\gamma(F^*) = \gamma(F)^*$。由于，使用定义在命题263里的$\mathfrak{F}^R$上$*$运算，我们有：

$$\gamma((\mathbf{1}, \rho, \psi)^*) = \gamma((R^*, \bar{\rho}, (\mathcal{J}E(\overline{R}^*))\psi)) = \pi_0(R^*)(\mathcal{J}E(\overline{R}^*)\psi)$$

另一方面，$\gamma((1, \rho, \psi))^* = \psi^*$，因此$\gamma$是一个$*$同态，假若对于所有$\psi \in E(\rho)$都具有$\psi^* = R^*(\mathcal{J}E(\overline{R}^*))\psi$。

由于，对于任意$\overline{\psi} \in E(\bar{\rho})$，我们有$R^*\overline{\psi}\rho(A) = R^*\bar{\rho}\,\rho(A)\overline{\psi} = AR^*\overline{\psi}$，因此$(R^*\overline{\psi})^* \in E(\rho)$。把这用到$\overline{\psi} = \mathcal{J}E(\overline{R}^*)\psi \in E(\bar{\rho})$，我们看到$\psi^* = R^*(\mathcal{J}E(\overline{R}^*))\psi$成立，当且仅当，对于所有$\psi' \in E(\rho)$都有$\psi^*\psi' = R^*(\mathcal{J}E(\overline{R}^*)\psi)\psi'$。

根据附录的命题344，$(E(\rho), E(R), E(\overline{R}))$是在希尔伯特空间范畴里的$E(\rho)$的共轭。(或者是在超希尔伯特空间里的，这并非关键，因为我们不使用对称。)因此存在$E(\rho)$和$E(\bar{\rho})$各自的基$\{e_i\}$，$\{f_i\}$，根据$\widehat{E(\rho)}$，$\widehat{E(\bar{\rho})}$的对偶基$\{\hat{e}_i\}$，$\{\hat{f}_i\}$，使得：

$$\cdots\cdots\cdots\cdots E(R) = \sum_i f_i \otimes e_i,\ E(\overline{R}) \sum_i e_i \otimes f_i$$

$$E(R)^* = \sum_i \hat{f}_i \otimes \hat{e}_i,\ E(\overline{R})^* = \sum_i \hat{e}_i \otimes \hat{f}_i$$

因此，对于$\psi \in E(\rho)$，$\overline{\psi} \in E(\bar{\rho})$，我们有：

$$\langle \mathcal{J}E(\overline{R}^*)\psi, \overline{\psi}\rangle = (\sum_i \hat{e}_i \otimes \hat{f}_i)(\psi \otimes \overline{\psi}) = \sum_i \hat{e}_i(\psi)\hat{f}_i(\overline{\psi})$$

从而$\mathcal{J}E(\overline{R}^*)\psi = \sum_i \overline{\hat{e}_i(\psi)}f_i$，因此：

$$\begin{aligned}E(R)^*((\mathcal{J}E(\overline{R}^*)\psi) \otimes \psi') &= (\sum_i \hat{f}_i \otimes \hat{e}_i)(\sum_i \overline{\hat{e}_j(\psi)}f_j \otimes \psi') \\ &= (\sum_i \overline{\hat{e}_i(\psi)}\,\hat{e}_i)(\psi') = \langle \psi, \psi'\rangle\end{aligned}$$

既然在$\mathfrak{F}$里，左边等于$R^*(\mathcal{J}E(\overline{R}^*)\psi)\psi'$，而右边等于$\psi^*\psi'$，提供了所希望的

恒等式$\psi^* = R^*(\mathcal{JE}(\overline{R}^*))\psi$。

对于$(A, \rho, \psi) \in \mathcal{F}^R_{E_\partial,0}$，我们所清楚的是$\gamma((A, \rho, \psi))$被包含在$\mathfrak{F}$的有限维 G 稳定子空间里，由此是处于$\mathfrak{F}_0$ 里的。任何 $F \in \mathfrak{F}_0$ 是具有 $\sigma \in \hat{G}$ 的形式 $\varepsilon_\sigma(F)$的有限多项的和。挑出一个$\mathfrak{F}_0$ 里的等距之不可约化的子空间 H_σ，它是按照类 σ 变换的，存在一个由子空间 H_σ 约化的自同态 $\rho \in \Delta_f$。由于任何 $F \in \varepsilon_\sigma(\mathfrak{F})$是一个线性组合$\sum_i A_i\psi_i$，具有 $A_i \in \mathfrak{A}$，$\psi_i \in H_\sigma$，我们具有 $F = \gamma(\sum_i(A, \rho, \psi_i))$，证明 $\gamma(\mathfrak{F}^R_{E_\partial,0}) = \mathfrak{F}_0$。

设$(A, \rho, \psi) \in \mathfrak{F}^R_0$，根据$\mathfrak{F}_0$ 的构造，我们具有一个有限的总表示$(A, \rho, \psi) = \sum_i(A_i, \rho_i, \psi_i)$，其中 ρ_i 是不可约化的，并且相互非等距的。由于，$\gamma((A, \rho, \psi)) = \sum_i A_i\psi_i$，其中空间 $E(\rho_i) \subset \mathfrak{F}$是在 G 的相互不等价不可归约表示之下变换的。因此，$\gamma((A, \rho, \psi)) = 0$，当且仅当 $A_i\psi_i = 0$，对于所有 i 等距，它按照一个在类 σ 里的表示进行变换。由于通过谐函数分析，任何 $F \in \mathfrak{F}$ 具有形式 $F = \sum_\sigma A_{\sigma,i}\psi_i^\sigma$ 的一个唯一表征，这意味着对于每一个 i，我们具有$(A_i, \rho_i, \psi_i) = 0$。因此 γ 是单射的。

因此我们证明了 γ：$\mathfrak{F}^R_{E_\partial,0} \to \mathfrak{F}_0$ 是一个同构。由于$\mathfrak{F}$的真空态 $\omega_0^{\mathfrak{F}} = (\Omega_0, \Omega)$是假定为规范不变的，$\mathfrak{F}^R_{E_\partial,0}$上的态 $\omega_0^{\mathfrak{F}} \circ \gamma$ 和 $\omega_0^{\mathfrak{A}} \circ m$ 是一致的，意味着完全网在其 GNS 表征里是幺正地等价。

推论 282。任何完备正常场网$\mathfrak{F}$是幺正地等价于一个罗伯茨场网$\mathfrak{F}^R_E$，其中我们用哪一个纤维函子 E 并不重要。

注释 283。正如之前所保证的，我们回到了函子 F：$\mathcal{C} \to \mathcal{SH}_f$的严格性问题，这是在场网的构造里假设的，但是在附录里没有证明。在后者里，我们构造了一个非严格纤维函子，即一个函子 E：$\mathcal{C} \to \mathcal{SH}_f$结合满足方程(56)、方程(57)的自然同构 $d^E_{\rho,\rho'}$：$E(\rho) \otimes E(\rho') \to E(\rho, \rho')$和 e^E：$\mathbf{1}_{\mathcal{SH}} \to E(\iota_\Delta)$。在 10.2 小节里的(代数)场代数$\mathfrak{F}_0$ 的构造，能够容易推广到如下情形，即场的积被定义为：

$$(A_1, \rho_1, \psi_1)(A_2, \rho_2, \psi_2) := (A_1\rho_1(A_2), \rho_1 \otimes \rho_2, d^E_{\rho_1,\rho_2}(\psi_1 \otimes \psi_2))$$

并且单位是$(\mathbf{1}, \iota, e^E 1_{\mathbb{C}})$。现在结合性和单位性质是方程(56)、方程(57)的明显结果。构造和证明的其余部分像之前一样完成，就像实施幺正 d^E，e^E 那样。和命题 280 一起这样的一个有趣结果，是我们能够证明严格纤维函子 E'：$\Delta_f \to \mathcal{SH}'$的存在，其中$\mathcal{SH}_f'$是有限维超希尔伯特空间的范畴的严格化。这跟范畴理

论里的严格结果是一致的。张量范畴的严格化很好地在[Kassel(卡瑟尔), 1995, Chapter XI]中得到处理，但是对张量函子的严格化最好的参考还是[Joyal and Street(乔亚尔和斯特里特), 1993a, 1]。

10.6 在$\mathfrak{A}$和$\mathfrak{F}$之间的进一步关系以及一个伽罗瓦(Galois)解释

在第9节我们花大量篇幅讨论可观测量网$\mathfrak{A}$的超选择截面的建构，跟在场网$\mathfrak{F}$的一个(整体)规范群作用的谐函数分析的关系。注意到我们并没有断言，所有的不动点网$\mathfrak{A}=\mathfrak{F}^G$的DHR表征是跟真空表征用$\mathfrak{F}$里的场连接起来的。为了看清这一点通常是假的，考虑一个具有非平凡DHR范畴的理论$\mathfrak{A}$，并且取$\mathfrak{F}:=\mathfrak{A}$为“场网”，通过平凡群$G=\{e\}$作用上去。显然，$\mathfrak{A}$的全部DHR表征并不是由$H_0$上$\mathfrak{F}$的作用所产生的。在特殊情况下，其中$\mathfrak{F}$是玻色子并且自身满足一个可观测量网上的全部要求，它可能具有非平凡DHR截面。限制具有$d(\pi)<\infty$的$\mathfrak{F}$的DHR表示π到$\mathfrak{A}$上，人们获得相同维度的$\mathfrak{A}$的DHR表征，因此分解成不可约化量的有限直和。如果π是不可约化的，并且等价于$\mathfrak{F}$的真空表征π_0，那么所有的用此方式获得的$\mathfrak{A}$的不可约化表征，跟包含在$\pi_0|_{\mathfrak{A}}$里的那些是不相交的。我们避免对这个问题进一步分析。不过，我们想指出人们能够具体化在网$\mathfrak{F}$上的条件，那就意味着$\mathfrak{A}$的全部DHR表征都包含在$\pi_0|_{\mathfrak{A}}$里。这牵涉到网上同调或者局域1上同调，它们是J. E. 罗伯茨发展起来的，并且在[Roberts, 1990, §3.4]里有评论。我们不再进一步给出准确说法，仅仅叙述如下：如果$\mathfrak{F}$具有“准平凡1上同调”，并且是通过整体规范对称的紧致群G来作用的，那么等价性(由命题206)范畴$DHR_f(\mathfrak{A})\simeq\Delta_f(\mathfrak{A})$是等价的，就像对称张量范畴到$\mathrm{Rep}_f G$一样。在[Buchholz *et al.*, 1992]中证明，例如，一个无质量标量场的理论具有准平凡1上同调。因此，如果人们取$\mathfrak{F}$为这些(同样质量)场的N复制的直积，那么$SO(N)$作用在$\mathfrak{F}$上。因此，只要$G\subset SO(N)$是封闭子群并且$\mathfrak{A}=\mathfrak{F}^G$，就有$\Delta_f(\mathfrak{A})\simeq\mathrm{Rep}_f G$。在[Doplicher and Piacitelli(多普利克尔和皮亚奇泰利), 2002]里，这个观测是跟极限建构结合起来，证明任何(第二可数的)紧致群作为DHR规范群出现的。按照类似的形式，人们证明如果$\mathfrak{F}$是具有其正则$\mathbb{Z}/2$对称的有质量费米子的理论，那么$\Delta_f(\mathfrak{F}^{\mathbb{Z}/2})\simeq\mathrm{Rep}_f\,\mathbb{Z}/2$。

有些结果处在相反方向，即从$\mathfrak{A}$的超选择结构到$\mathfrak{F}$的结构。根据在第9节

没有完全包括的[Doplicher and Roberts，1990，Theorem 3.6]，在[Doplicher and Roberts，1990]和上述第10节中重构的场网，满足“扭曲的哈格对偶性”。特别是，如果$\mathfrak{A}$没有费米表征而$\mathfrak{F}$满足哈格对偶性，在此情况下，人们可能研究范畴$DHR(\mathfrak{F})$或者$\Delta(\mathfrak{F})$。在[Conti *et al.*（孔蒂等），2001]中，下面定理得到证明。

定理284。设$\mathfrak{A}$是可观测量的一个网，使得存在至多可数的多个有限维的DHR表征，其中全部是玻色子的，那么完备场网$\mathfrak{F}$没有有限维的非平凡DHR表征。

并不想评论很多其他跟前面几节处理的那些相关已知结果，我们这一节主要评论DHR/DR理论的一个非常满意的数学解释。我们指的是这个理论跟代数场拓展的伽罗瓦理论之间的直接类似。(在后者的背景中这很清楚，我们说的“场”指的是像$\mathbb{Q}$，$\mathbb{R}$，$\mathbb{C}$的代数结构，而不是经典或者量子场的理论。)一个在后一种意义上的场$\mathbb{F}$被称为代数上封闭的，如果每个有$\mathbb{F}$中系数的多项式$P(x)$在$\mathbb{F}$里有个零(那么P是线性因子$x-a$的积)。每一个场$\mathbb{F}$是一个本质上唯一代数上封闭场$\overline{\mathbb{F}}$的子场，$\overline{\mathbb{F}}$是$\mathbb{F}$的一个代数拓展，意思是$\overline{\mathbb{F}}$是通过对$\mathbb{F}$的多项式方程的伴随的、通常超限的解获得的。群$G_{\mathbb{F}}=\mathrm{Aut}_{\mathbb{F}}(\overline{\mathbb{F}})$是紧致的，并且人们在中间场$\mathbb{F}'\subset\overline{\mathbb{F}}$、$\mathbb{F}'\supset\mathbb{F}$跟封闭子群$H\subset G$之间具有一个双射对应，这个对应是由$H\mapsto\overline{\mathbb{F}}^{H}$，$\mathbb{F}'\mapsto\mathrm{Aut}_{\mathbb{F}'}(\overline{\mathbb{F}})$给定的。一个类似的伽罗瓦对应在AQFT里成立，参看比如[Conti *et al.*，2001；Carpi and Conti，2001]。在定理284看来，完备DR场网的建构完全类似于代数封闭的建构，并且可以看作是向简化或者表现更优的理论的过渡。相反，就像把代数上封闭的场$\mathbb{F}$放在封闭子群$G\subset\mathrm{Aut}\ \mathbb{F}$的作用之下，就会导致一个代数上非封闭的场$\mathbb{F}^{G}$，取具有平凡范畴$\Delta_f(\mathfrak{F})$的一个网$\mathfrak{F}$的$G$子网(更明确的准平凡1上同调)，会得到一个具有非平凡范畴$\Delta_f(\mathfrak{A})$的网。因此通过一个非平凡DHR范畴$\Delta_f(\mathfrak{A})$显现的“错乱”，揭示了理论$\mathfrak{A}$“真的”只是一个简单理论的子理论。

当然，在物理上还不完全清楚，是具有其非平凡表征范畴$\Delta_f(\mathfrak{A})$的“可观测量”网$\mathfrak{A}=\mathfrak{F}^{G}$，还是具有平凡$\Delta_f(\mathfrak{F})$而非平凡整体对称群$G$的“场网”$\mathfrak{F}$更为基础些——至少在$\mathfrak{F}$是玻色子的时候。在[Haag，1996]里论证过物理情形的

"正确"描述应该是通过没有任何整体对称网的方式。另一方面，在[Haag，1996，Section III. 4. 2]人们发现一个意味着 $\Delta_f(\mathfrak{A})$ 的平凡性的可观测量的"好"网上的试探性假设。就像上述讨论表明的，很难找到一个理论既是一个平凡 DHR 范畴 Δ_f 又是平凡整体对称性群！单个自由有质量玻色场的理论就是这种罕见例子之一。人们是否描述了这些看法，从数学的观点看，网 $\mathfrak{A}$ 和 $\mathfrak{F}$ 两者包含同样多的信息，这个等价事实上是个很有用的工具，因为它允许从不同角度看很多问题。比如，当一个自旋统计定理能够在"场"框架下被证明，其物理解释在"可观测量"设定里就会清楚些。

10.7　自发对称性破缺

迄今为止，我们的整个分析都以理论 $\mathfrak{A}$ 的哈格对偶性公理为前提。虽然哈格对偶性在我们的范畴 $\Delta(\mathfrak{A})$ 分析里起到重要作用，但是也需要建立 $\Delta(\mathfrak{A})$ 跟表征之间的等价性，来满足先天的物理动机 DHR 标准(定义 204)。因此，似乎 DHR 表征在物理上也有非哈格对偶网的动机，同时我们的数学分析也会被卡住。所以我们将简单评论一下解决这个问题的方法，它最终具有意义深远的物理解释。

定义 285。 设 $O \mapsto \mathfrak{R}(O)$ 是一个在真空希尔伯特空间 H_0 上的冯·诺伊曼代数网，$\mathfrak{R}$ 的对偶网 $\mathfrak{R}^d$ 是赋值 $O \mapsto \mathfrak{R}(O')'$。

如果我们具有 $O_1 \subset O_2$ 那么 $O_2' \subset O_1'$，因此 $\mathfrak{R}(O_2') \subset \mathfrak{R}(O_1')$，由此 $\mathfrak{R}^d(O_1) \subset \mathfrak{R}^d(O_2)$。因此对偶网确实满足保序性。$\mathfrak{R}$ 的微观因果性等价于 $\mathfrak{R}(O) \subset \mathfrak{R}(O')' = \mathfrak{R}^d(O)$，或者简略地 $\mathfrak{R} \subset \mathfrak{R}^d$，并且 $\mathfrak{R}$ 的哈格对偶性等价于 $\mathfrak{R} = \mathfrak{R}^d$。如果 $\mathfrak{A}_1 \subset \mathfrak{A}_2$(任何 O 的包含的意义上)，那么 $\mathfrak{A}_2^d \subset \mathfrak{A}_1^d$，因此 $\mathfrak{A} \subset \mathfrak{A}^{dd}$，以及一个标准论证说明 $\mathfrak{A}^d = \mathfrak{A}^{ddd}$。不过，注意 $\mathfrak{R}$ 的微观因果性并不意味着 $\mathfrak{R}^d$ 的微观因果性！这激发了下面定义。

定义 286。 一个网 $O \mapsto \mathfrak{R}(O) \subset B(H_0)$ 满足本质对偶性，如果 $\mathfrak{R}$ 和对偶网 $\mathfrak{R}^d$(两个都用双面锥标记)满足微观因果性。

引理 287。 如果 $\mathfrak{R}$ 满足本质对偶性那么 $\mathfrak{R}^d = \mathfrak{R}^{dd}$，即 $\mathfrak{R}^d$ 满足哈格对偶性。

注释 288。 本质对偶性可能难以证明；幸运的是，本质对偶性源自于楔形对偶性，也即是对所有楔形区域(这个区域由庞加莱变换从标准楔形 $W_0 = \{x \in \mathbb{R}^{1+s} \mid x_0 \geq |x_1|\}$ 得到)都有 $\mathfrak{R}(W')' = \mathfrak{R}(W)$。除了比本质对偶性更容易证明

之外，楔形对偶性是一个非常基本的性质，可能为任何“合理的”量子场论所要求。

假设$\mathfrak{R}$满足本质对偶性，$\mathfrak{R}^d$满足哈格对偶性，并且D(H)R分析适用于它。因此我们得到一个具有共轭$\Delta_f^d(\mathfrak{R}) := \Delta_f(\mathfrak{R}^d) \simeq DHR_f(\mathfrak{R}^d)$的对称张量$*$范畴，并且我们能够建构跟$(\mathfrak{R}^d, \Delta_f(\mathfrak{R}^d))$相关的完备DR场网$\mathfrak{F}$。因此人们具有一个网的包含$\mathfrak{R} \subset \mathfrak{R}^d \subset \mathfrak{F}$。DR规范群作用在$\mathfrak{F}$上，并且我们具有$\mathfrak{F}^G = \mathfrak{R}^d$，以及$G = \mathrm{Aut}_{\mathfrak{R}^d}(\mathfrak{F})$。由于群$G$是通过使得真空矢量固定的幺正性来实现的，$G$由“未破缺的对称”组成。人们现在能够定义一个巨大的群

$$\hat{G} = \mathrm{Aut}_{\mathfrak{R}}(\mathfrak{F})$$

并且对其适当拓扑化。现在$G \subset \hat{G}$确实由幺正实现的$\hat{G}$的元素组成。关键在于，网$\mathfrak{R}$不可约化地作用在H_0上，因此一个其伴随作用使得所有代数$\mathfrak{R}(O)$逐点固定的那些幺正，必定也是一个作用在$\mathfrak{R}^d$的单位元的乘积。

涉及跟$\mathfrak{R}$结合在一起的范畴，有关范畴$\Delta(\mathfrak{R})$就没有什么可说的，但是罗伯茨证明了一个扩展函子K: $DHR(\mathfrak{R}) \to DHR(\mathfrak{R}^d)$的存在性，使得对于任何$\pi \in DHR(\mathfrak{R})$都有$K(\pi) \restriction \mathfrak{R} = \pi$，参看[Roberts，1990，§3.4]。（再者，局域1上同调理论起到关键作用，而且，这个结果在不少于三个空间维度里会由于孤立子表示的现象被破坏。）这个函子实际上是一个等价，因此自发对称性破缺在超选择结构里并没有显现。

对于我们提到的[Roberts，1974；Roberts，1990]的详细分析和评论性文章[Buchholz *et al.*，1992]，其戈德斯通(Goldstone)现象是在代数QFT的情境中被分析的。

注释：DHR超选择理论源自于一组四篇的论文系列：[Doplicher *et al.*，1969a]从一个场代数和规范群开始，然后导出超截面的性质。[Doplicher *et al.*，1969b]重构了场代数和规范群，是源自于对象都是全部一维的特殊情况里的表示的范畴重构(即Δ_f的对象的等价类构成一个阿贝尔群，具有单项积和作为逆的共轭)。[Doplicher *et al.*，1971]定义了对称$\varepsilon_{\rho_1,\rho_2}$，并且用它给出了$\Delta$对象的统计分类。

DHR理论的一般概述，参见[Roberts，1970；Roberts，1990；Roberts， 2004；Fredenhagen，1992；Fredenhagen，1994]、[Araki，1999]以及[Haag，1996]。

DR重构定理的完全证明分布于[Doplicher *et al.*，1974；Doplicher and Roberts，1972；Doplicher and Roberts，1989]以及[Doplicher and Roberts，1990]。我们在此重构定理所选择的方法是基于[Roberts，ND]和[Deligne，1990]，后者的简化主要依赖[Bichon，1998]和我们自己的文章。

对于DR重构定理的不正规阐述，参看[Doplicher and Roberts，1987；Doplicher，1991；Doplicher，1992；Doplicher，1993；Doplicher，1995]。对于重构场和规范群的目标的有趣描述，在得到解之前就写出了，参看[Roberts，1975]。

11. 重构定理的基本含义

我们现在回到驱使我们考察的基本问题(第7节)。我们也会指出几个其他情况，其中哲学界的讨论(比如有关置换对称和粒子的全同性)就可能受益于超选择理论的研究。

11.1 代数信奉者和希尔伯特空间保守主义

DHR超选择理论阐明了哲学家所问的某些问题，即有关可观测量代数的不等价表征作用的问题，但是它并不会回答我们的全部问题。我们先空置DHR理论提供帮助的那些问题，着手处理DHR理论跟基本问题相关的情况。

DHR分析要求我们固定一个真空态ω_0及其一个基本表征($\mathcal{H}_0$，π_0)。不等价DHR表征并不对应不同的真空，而是对应于一个或者相同真空态的不同局域激发。因此，DHR理论有效忽略如何选择一个真空表征的问题。(但是要注意DHR分析的威力，强烈建议——跟代数信奉者相反——表征是理论的物理内容的本质部分。)

其次，在某些最熟悉的情况下——比如自由玻色子场——DHR范畴是平凡的。那就是说$DHR(\mathfrak{A})=\{\pi_0\}$，由此有$\mathfrak{F}=\mathfrak{A}$。在此情况下，真空表征是唯一的DHR表征(跟自身有关)。因此，在这样的情况下，DHR超选择理论的复杂

工具似乎提供不了什么关于不等价表征物理重要性的见解。(不过，如果我们能够找到选择一个特殊真空表征的物理原因，那么 DHR 分析认为不再有跟解释这个现象相关的表征了。)

最后，即便在 $DHR(\mathfrak{A})$ 是非平凡的情况下，场代数本身具有不等价的表征。(总之，它仅仅是另一种大的 C^* 代数。)并且人们可能担心相同的保守派跟信奉者之间的争论也会在场代数里发生。

但是 DHR 理论还有一些关于不等价 DHR 表征以及关于场代数表征的事情要说。首先，场代数$\mathfrak{F}$作为希尔伯特空间$\mathcal{H}$上的算符得到具体建构，即$\mathfrak{F}$伴随一个选定的表征。(回顾选定的$\mathfrak{F}$表征是在希尔伯特空间$\mathcal{H}$上，在那分解成$\mathfrak{A}$超选择截面的直和。)当然，我们可能会考虑$\mathfrak{F}$的其他表征。但是在$\mathfrak{F}$的其他表征里，我们不再拥有作为$\mathfrak{A}$的 DHR 截面之间扭结子的$\mathfrak{F}$元素的直观解释。如果物理上有意义的量不得不与可观测现象有关联，那么通过$\mathfrak{A}$的$\mathfrak{F}$元素解释可能被看作解释的必要条件；由此给定的$\mathfrak{F}$表征可能因这些原因是首选。

因此，DHR 理论认为不等价表征的问题，并不因场代数而出现。考虑到可观测量代数的不等价表征的问题，我们能够把问题分成下面几个方面：

1. 存在物理上特有的真空态/表征吗？它具有什么样的特性？

2. 全部都是某些真空态的 DHR 范畴里的物理表征吗？这里我们没有管长程力理论的时候，参见[Buchholz and Fredenhagen, 1982]。在更一般的条件下，我们期望问题的形式保持不变：我们能够通过适当的场集的作用从固定的真空态达到物理态吗？

3. 如果对第二个问题的回答是否定的，那么我们将如何比较场跟真空表征没有关联的表征和有关联的表征呢？

假设第一个问题能够得到证实，并且真空表征$(\mathcal{H}_0, \pi_0)$是固定的，也假定 $DHR(\mathfrak{A})$是非平凡的。那么我们应该如何思考 $DHR(\mathfrak{A})$里的不等价表征？希尔伯特空间保守主义对目前情境的一个简单改写就告诉我们下面的结果：$DHR(\mathfrak{A})$里的表征类似于相竞争的理论，即一个正确，而其他的是剩余结构。代数信奉者对目前情境的简单改写可能会说：$DHR(\mathfrak{A})$里的表征是剩余结构；理论的物理内容是在$\mathfrak{A}$这个观测量的抽象代数里。

保守主义和信奉者都是建立在形式体系的过度简化观点基础上的：假定了

实在要素对应于抽象代数的算符，或者某些希尔伯特空间表象里的算符；并且假定可能的状态要么是在抽象代数上的全部状态，要么是状态的某些特定叶形线。但是 DHR 理论的基本看法是表征的集合自身也有结构，并且正是这个结构解释现象。所以，一个更好的立场应该是更严肃地对待所有的表象。因此，我们提出，按照表象实在论，理论内容通过如下几点给出：(i)网 $O\mapsto\mathfrak{A}(O)$；(ii)准局域代数上的动力学(即在 Aut $\mathfrak{A}$里变换群的表象)；(iii)DHR 表象的对称张量 * 范畴 $DHR(\mathfrak{A})$。

回顾保守派的说法在解释力上的优势：更多的结构(通过选择一个表象提供)所提供的实在要素也更多，从而满意的解释也更多。但是 DHR 超选择理论证明如此断言的优点是误导了：集中在一个表象就忽视了最有趣的结构，即表象之间的关系。的确，如果我们致力于一个表象，而忽视其他的，我们可能就没有场算符，没有规范群，没有玻色场和费米场的定义，也没有反粒子的定义，等等。

并且还存在看上去对表象实在论较强的异议：由于哈密顿量总是一个可观测量，不可能存在动力学演化能够把我们从一个表象的态带到一个不等价表象的态上。因此，不等价表象在动力学上是彼此孤立的，并且它们之间的抽象关系也不能解释那些最好地描述我们宇宙表象里的态的特征。

哈密顿量是可观测量的事实——因此不能从一个截面的态映射到另一种截面的态——产生我们关于场算符的进一步问题。回顾我们说的在截面$\mathcal{H}_0$里的真空“创造”状态，通过的是作用在从场代数来的元素。那就是，我们能够选择 $F\in H_\rho\subseteq\mathfrak{F}$使得 $F\Omega\in\mathcal{H}_\rho$，其中($\mathcal{H}_\rho$，$\pi_\rho$)是跟真空表象($\mathcal{H}_0$，$\pi_0$)不相交的。这里“创造”的说法指的是我们正在谈论某种动力学过程。一方面，$F\in\mathfrak{F}$能够被选择成幺正的，因此结构上映射 $\Omega\mapsto F\Omega$ 看上去像是动力学过程。但是由于哈密顿量是一个可观测量，变换 $\Omega\mapsto F\Omega$ 不是动力学上允许的。因此，在什么意义上$\mathcal{H}_0$里的态从真空是可进入的？超选择规则背后的关键看法是存在两个动力学可进入性的概念？如果如此，那么我们是如何理解这两个概念之间的差异的呢？

11.2 表象之间的解释关系

如果我们考虑一个 C^* 代数$\mathfrak{A}$，那么数学上可定义的表象(及由此在这些表象的叶形线里的态)之间的关系可由下表完全列出：

表 1. $\mathfrak{A}$的表象之间的关系

π_1 和 π_2 是等价的
π_1 和 π_2 是准等价的
π_1 和 π_2 是不相交的
π_1 和 π_2 是弱等价的

在第四个关系(特指态空间的拓扑学)之外，这些关系严格讲是那些能够定义在一个具有子对象的任意 * 范畴$\mathcal{C}$里的。在$\mathcal{C}$里的两个对象 X，Y 是等价的，如果存在一个幺正 $u \in \mathrm{Hom}(X, Y)$；是准等价的，如果存在一个等距 $v \in \mathrm{Hom}(X, Y)$；以及是不相交的，当且仅当它们不是准等价的。

现在考虑 C^* 代数$\mathfrak{A}$的正常状态空间 K。GNS 定理提供一个从 K 到$\mathfrak{A}$的表象范畴的映射 $\omega \mapsto (\mathcal{H}_\omega, \pi_\omega)$，然后我们使用这个映射诱导对应于 K 上表 1 里的那些关系：我们谈论等价性、准等价性以及不相交状态。而且，个别叶形线(其 GNS 表象是准等价的状态集合)具有丰富的几何结构，精确对应于某些希尔伯特空间$\mathcal{H}$的$\mathfrak{B}(\mathcal{H})$的正常状态空间。因此，在一条叶形线里我们具有一个在纯态(对应于$\mathcal{H}$里的射线)之间的“跃迁概率”的概念，以及一个“ω 是 ω_1 和 ω_2 之间叠加”的三方关系。当然，如果两个态处于不相交的叶形线里，那么这些关系平凡化。不相交态之间的跃迁概率是零，并且没有态可能是从不同叶形线来的态的叠加。似乎是只有那些物理上可能的事情——我们关于从不同叶形线态所能讨论的——是它们是“正交的”的。

正是前面的考虑导致哲学家担心不等价表象。这个担心是基于不相交表象似乎成为竞争者的事实，仅此而已。为了消除关于表象之间的担心，有些哲学家[Clifton and Halvorson, 2001a; Halvorson, 2004]干脆断言这些表象是现象的“互补”描述(在玻尔意义上)。当然“互补”这个词是想建议表象不仅仅是竞争者，并且选择其中一个并不意味着完全排除跟另一个的相关性。

我们想用某些关于不等价表象之间关系的事实，来代替暗示性的——而可能是误导——术语。为了阐述我们想据此说明的意图，考虑群表象的情况：设 $\mathrm{Rep}_f G$ 是紧致群 G 的幺正表象的范畴。$\mathrm{Rep}_f G$ 不仅仅是 * 范畴，而且具有单项积和共轭。那就是，对于 $\mathrm{Rep}_f G$ 里的 X，Y，存在一个积对象 $X \otimes Y$ 以及一个共

轭对象$\overline{X}$。对于我们的目的，这是在群表象跟任意 C^* 代数$\mathfrak{F}$的表象之间关键区别，对于一个任意的 C^* 代数，不存在表象的积或者表象的共轭。

在紧致群表象的情况下，$X\in \mathrm{Rep}_f G$ 通常跟 $X\otimes Y$ 和 $\overline{X}$ 都不会相交。但是在此情况下，我们不想看到 X 仅仅是作为 $X\otimes Y$ 或者 $\overline{X}$ 的竞争者；在这些表象之间存在某些有趣的关系。粗略讲，关于 X 的信息给了我们有关 $X\otimes Y$ 和 $\overline{X}$ 的信息。因此，虽然这些表象在技术上是“不相交的”，但是它们彼此并非完全没有关系。①

DHR 分析和 DR 重构定理有一个主要成就，就是证明物理表象的范畴 Δ_f 是一个具有共轭的张量 * 范畴；确实，嵌入定理（参见附录）证明了 Δ_f 等价于某些紧致群 G 的范畴 $\mathrm{Rep}_f G$。问题在于，表象范畴上的这些附加关系是否能够帮助我们找回不相交表象仅仅是相互矛盾的描述的想法。

一个类似的陈述可能有用，考虑希尔伯特空间的一个 H_1，H_2 对，并且设 $\psi_i\in H_i$ 是单位矢量。现在考虑下面两个“实在的描述”：

1. 态是 ψ_1；

2. 态是 $\psi_1\otimes\psi_2$。

我们这里讲的是：是这些相互矛盾的描述吗？在一种意义上，描述 1 和 2 是相互矛盾的，因为它们两个不能同时都是实在的完全恰当的描述。不过，描述 1 和 2 在同样的意义上又不是相互矛盾的，即在单独一个希尔伯特空间里的两个正交矢量才是相互矛盾的。这两个态描述不仅仅是相互矛盾的，而且存在一个有趣的含义，其中 ψ_1 是 $\psi_1\otimes\psi_2$ 的“部分”。确实，有关 ψ_1 的信息（比如它赋予可观测量的值）事实上给了我们有关 $\psi_1\otimes\psi_2$ 的信息，因为在 H_1 和 H_2 之间存在正则映射。

现在设 π_1，π_2 是在 DHR 范畴 Δ_f 里的对象，并且设（通常也是这样的情况）表象 H_1 和 H_2 是不相交的。这些是相互矛盾的描述吗？再者，是 π_1 和 $\pi_1\otimes\pi_2$ 是在下述意义上的相互竞争者吗？即如果一个对象（或者宇宙的对象）的态是在$\mathcal{H}_{(\pi_1\otimes\pi_2)}$，那么它就不是在$\mathcal{H}_{\pi_1}$。虽然如此，$\pi_1$ 和 π_2 不仅仅是相互竞争者，

① 但也要注意：物理学哲学家至今并不担心作为对实在的描述相竞争的不等价群表象。并且有好的理由，因为群元素并不是可观测量，而群也没有态。DHR 理论的另一个洞见是，证明了物理学家关于群表象的直觉并非全部没有基础，因为事实上可观测量代数的有趣（DHR）表象对应于紧致群的表象。

因为在某个含义上π_1是$\pi_1\otimes\pi_2$的“部分”。

但是有两点警告要提醒一下，首先，我们必须当心“部分”这个说法的使用，比如，Δ_f能够具有非平凡表象π使得$\pi\otimes\pi$等价于真空表象，然而仍然不清楚我们应不应该说“π是$\pi\otimes\pi$的部分”。第二，在ψ_1和$\psi_1\otimes\psi_2$的情况跟π和$\pi\otimes\pi$的情况之间的重要区别仍然不清楚：这两个表象是在单个C^*代数$\mathfrak{A}$上态的GNS表象。因此我们能够直接比较这些态赋予$\mathfrak{A}$里可观测量的期望值，并且它们也有重大分歧(确实，对于任何$\varepsilon>0$，存在一个可观测量$A\in\mathfrak{A}$，使得$\|A\|\leqslant 1$和$\|\omega_1(A)-\omega_2(A)\|>2-\varepsilon$)。因此，存在一个明显且经验上可以证实的含义，即在π_1里的态跟在$\pi_1\otimes\pi_2$里的态是相互竞争者。

最后，在DHR表象π跟它的共轭$\overline{\pi}$之间存在一个有趣的物理关系，即便π和$\overline{\pi}$通常是不相交的。总之，$\overline{\pi}$就像是π的逆：如果π是不可约化的，那么$\overline{\pi}$是唯一不可约化的表象，使得$\pi\otimes\overline{\pi}$包含真空表象的复制。事实上，当$\pi=\pi_0\circ\rho$，其中ρ是Δ_f的一维1元素，$d(\rho)=1$，那么这是一个精确关系：ρ是自同构并且$\overline{\rho}=\rho^{-1}$。按照场算符的方式，如果$F$产生一个荷$\xi$，那么$\overline{F}$湮灭荷$\xi$。而且，一旦$\pi$承认一种粒子解释，那么在$\overline{\pi}$里的叶形线里的态是$\pi$的叶形线里态的反粒子态[Doplicher *et al.*, 1996b]。

11.3 作为理论实体和作为剩余结构的场

从超选择理论的观点看，在可观测的和不可观测的场之间存在一个明显的区分，即一个场算符是一个可观测量，当且仅当它在所有的规范变换下是不变的。在什么程度上，场和可观测量之间的区分，跟科学哲学家在一个理论中对理论部分和观测部分的区分是相当的呢？即便这两个概念不是完全一样，这样的关联也有的。特别是，问一下在什么程度上场加上量子场论的规范部分被可观测量代数所固定似乎是件有趣的事情。

首先，场算符的等价系统概念似乎跟哲学家的“理论等价性”范畴类似。

定义289。设$\mathfrak{F}_1=(\mathfrak{F}_1,\mathcal{H}_1,G_1)$和$\mathfrak{F}_2=(\mathfrak{F}_2,\mathcal{H}_2,G_2)$是具有对$(\mathfrak{A},\omega)$规范对称的局域场系统(参见定义247)。然后$\mathfrak{A}_1$和$\mathfrak{F}_2$是理论上等价的，当且仅当它们作为局域场算符系统是幺正等价的(参见定义247)。

评论290。(i)这个定义并非完全适当，因为它并没有指向动力学，比如，

这个定义意味着对不同正质量的自由玻色场网在理论上是等价的。因为一个完全适当的定义，应该有必要要求幺正映射 W: $\mathcal{H}_1 \to \mathcal{H}_2$ 也扭结两个希尔伯特空间上的动力学群。(ii)如果 $\mathfrak{F}_1$ 和 $\mathfrak{F}_2$ 是理论上等价的，那么它们在所有物理相关意义(模动力学)上都是等价的：它们具有同样类型的对易关系(要么都是正常对易关系要么是非正常对易关系)，它们具有等距规范群，等等。

进行类比也建议我们，通过它们的可观测量代数网之间的某些等价性，来定义两个理论之间的“可观测量的等价性”。存在若干方法可以阐明如此设定的可观测性等价性的概念，哲学家们都是得心应手的。下面两个定义给出可观测等价性的最弱概念，这样的等价性并没有考虑可观测量代数的表象。

定义 291。设 $\mathfrak{F}_1$ 和 $\mathfrak{F}_2$ 是两个拥有规范对称的局域场系统，并且设 $\mathfrak{A}_1$ 和 $\mathfrak{A}_2$ 是不动点代数；即：

$$\mathfrak{A}_i = \{A \in \mathfrak{F}_i: \alpha_g(A) = A，对全部的 g \in G_i\}$$

那么我们说 $\mathfrak{F}_1$ 和 $\mathfrak{F}_2$ 是弱可观测地等价的，当且仅当存在一个从代数 $\mathfrak{A}_1$ 到代数 $\mathfrak{A}_2$ 的 $*$ 同构。

定义 292。设 $\mathfrak{F}_1$ 和 $\mathfrak{F}_2$ 是两个具有规范对称的局域场系统，设 $\mathfrak{A}_1$ 和 $\mathfrak{A}_2$ 是其不动点网，即对于双面锥 O，有：

$\mathfrak{A}_i(O) = \{A \in \mathfrak{F}_i(O): \alpha_g(A) = A，对全部的 g \in G_i\}$

那么我们说 $\mathfrak{F}_1$ 和 $\mathfrak{F}_2$ 是可观测地等价的，当且仅当存在一个同构 α: $\mathfrak{A}_1 \to \mathfrak{A}_2$(参见定义 118)。

评论 293。第一个定义较弱，因为它并不要求网结构被 $*$ 同构 α 保留。

再者，定义未指向动力学，而它可能是可观测性等价的完全适当处理的一个重要部分。虽然如此，即便有这些定义，我们能够理解有关可观测量对场的不充分决定论评论，或者关于不同理论的物理等价性评论。

1. (可观测性等价理论的建构)DR 重建定理，提供了一个一般性的、非平凡的方案，针对的是重建作为可观测性等价的非等价理论：如果($\mathfrak{A}_1$，ω_0)具有非平凡 DHR 超选择截面，那么它就能够嵌入两个非等价的场代数 $\mathfrak{F}_1$ 和 $\mathfrak{F}_2$。确实，$\mathfrak{A}$ 总是自身之上的场代数(但是不完备)，但是从 DR 重构定理来的场代数 $\mathfrak{F}$ 是完备的。

2. (超场的排除)长期以来认为超场是个理论上的人为制品。更准确地说，

已经断言任何超场理论是“物理上等价”于具有正常对易关系的理论，参见[Araki，1961]。DR 重构定理部分认同了这个断言，通过说明每一个超场理论可观测地等价于具有正常对易关系的理论。确实，假设$\mathfrak{F}_1$是一个超场理论，那么我们能够获得在$\mathfrak{F}_1$里的可观测代数$\mathfrak{A}$，并且应用 DR 重构定理建构具有正常对易关系的场代数$\mathfrak{F}_2$。由于$\mathfrak{F}_1$和$\mathfrak{F}_2$具有局域可观测量代数相同的网，它们是观测性上等价的。

3. 有人断言量子场（场代数$\mathfrak{F}$）跟可观测量（可观测量代数$\mathfrak{A}$）之间的关系，类似于坐标系跟微分几何里的流形之间的关系。然而，DR 重建定理证明（遵守正常对易关系），存在唯一的场网$\mathfrak{F}$以及跟可观测量代数（$\mathfrak{A}$，ω_0）兼容的规范群 G。因此，在两个情况之间存在一个很强的类似，因为下面的说法似乎没有意义，即认为流形的坐标系统就实在表象方面比另一个坐标系更好。

最后，我们得到有关不等价表象问题跟规范不变性问题之间的内在关联的清晰图像。

按照正常解释立场，如果物理系统的两个态能够在规范对称下等同，那么那两个态是关于同样基本实在的不同描述。因此，出于计算状态的目的，我们应该查看规范轨道下状态空间的商。类似地，也认为理论的“真实”量是规范不变量，参见[Earman，2004]。

在 DHR 超选择理论的设定里，可观测量$\mathfrak{A}$的代数确实是场代数$\mathfrak{F}$的规范不变量部分，即：

$$\mathfrak{A}=\{A\in\mathfrak{F}:\ \alpha_g(A)=A,\ 对全部的\ g\in G\}$$

其中 G 是规范群，当然这意味着，对于任何可观测量 $A\in\mathfrak{A}$，在态 ψ 跟规范变换态 $U(g)\psi$ 之间不存在区别。当然，如果 ψ 是在真空表示了的态矢量，那么 $U(g)\psi=\psi$，因为那里的规范群表象是平凡的。因此，如果正常解释立场是对的，理论的物理内容是在可观测代数$\mathfrak{A}$里；场是“描述性的东西。”

因此假设我们不管场代数$\mathfrak{F}$，并且只是希望可观测量代数$\mathfrak{A}$提供理论的物理内容。然而我们应该对$\mathfrak{A}$的表象说些什么？它们仅仅是些描述性的无价值的东西吗？如果不是，那么有没有一个$\mathfrak{A}$的正确表象，或者我们怎么就为了解释全部物理事实需要不等价的表象？

DR 重建定理表明上述两个问题集——一方面是关于规范不变量，而另一

方面是表象——是紧密结合在一起的。理论的全部结构，场代数$\mathfrak{F}$和规范群G，是唯一可以从表象$DHR(\mathfrak{A})$的范畴结构重新获得的(模完备性以及正常对易关系)。规范不变量场的本体论意义跟不等价表象的本体论意义紧密相连(我们在下一节讨论序列对称时会回到这个问题)。

当然，在DHR超选择理论里的整体规范对称跟电磁理论或者广义相对论的局域规范对称之间具有重要区别。但是不清楚的是，这个不同让DR重构定理对理解规范对称跟超选择规则之间的关系几乎没有什么趣味。

11.4 统计学、置换对称以及全同粒子

哲学家们热衷于经典物理学的麦克斯韦—玻耳兹曼(Maxwell-Boltzmann)统计跟量子物理的玻色—费米统计之间的不同。的确，挑战性的说法是玻色—费米统计为置换不变性所解释——即在拥有置换过的粒子标记状态之间没有物理上的差异——并且这也意味着量子粒子并不是经典粒子意义的“个体”。参见[French(弗伦奇)，2000；French and Rickles(弗伦奇和里克斯)，2003]对此论证的解释。

但是这样的讨论可能为过于简化的形式体系所妨碍。特别是，态常常等同于单个希尔伯特空间里的单位矢量(或者最好等同于射线)，并且没有阐明非置换不变量算符的地位。从更加灵活的形式体系看问题可能更有利，它容许在场和可观测量之间的区分，并且其中的态能够用可观测量代数的不等价表象里的矢量表示。

为什么置换不变性问题应该在量子场论(QFT)背景中重新考察，这有另一个原因。一些资料提出关于粒子个体性的形而上学问题在转到QFT时无效，因为：(i)QFT是关于场的而不是粒子的；(ii)QFT的福克空间形式已经等同置换了的态，从而排除了作为个体的粒子概念。我们已经注意到对如何理解QFT关乎场而非粒子的思想仍不清楚。再者，DR重构定理明确证明了如何理解理论里的非置换不变量和态的意义，它们都是明显的置换不变量。

不必惊讶于DHR理论跟置换不变性和统计学的问题相关：DHR的最初目标是澄清QFT里统计学的作用。正当数学成效的巅峰时刻，罗伯茨作出对DHR统计分析的下面大胆断言：

> 通过对涉及被称为粒子的"统计学"起源的超选择截面的研究，提供了一个见解……就像一个理论应该决定它的粒子态，因此也应该决定这些粒子的统计学，通常量子力学忽视了这个挑战，等于是说粒子的统计学是决定理论的参数之一，即等于告诉你一个 n 粒子态具有什么对称。QFT 做得并不好，因为它认为理论里粒子的统计是由类空分离的场的对易关系所决定的……为了适应局域可观测量决定理论的哲学观念，我们被迫完全面对这个挑战[Roberts，1976，203]。

在所引这篇文章的剩余部分，罗伯茨证明了玻色—费米粒子统计自然出现于可观测量代数物理表象的 DHR 分析。

罗伯茨的断言跟统计学和全同粒子之间的哲学争论密切相关。哲学家们问过"解释玻色—费米统计的是什么?"罗伯茨的回答是这个解释来自于可观测量代数表象的范畴结构。

让我们回顾一下玻色—费米区分是如何出现在 DHR 分析里的。在 8.3 节，范畴 Δ 的一个对象 ρ 被证明具有一个内在维度 $d(\rho)$，这个维度是有限的当且仅当 ρ 具有共轭；在此情况下我们定义一个称为 ρ 扭度(twist)的幺正算符 $\Theta_\rho \in \mathrm{End}(\rho)$，如果 ρ 是不可约化的，那么 $\Theta_\rho = \omega_\rho \mathrm{id}_\rho$，其中 $\omega_\rho = \pm 1$。那么我们规定一个"玻色子"的对象具有 $\omega_\rho = 1$，而一个"费米子"的对象具有 $\omega_\rho = -1$。

当然，ρ 并不是哲学家所称为的玻色子或者费米子的东西种类——它不是波函数。为了连接这两个体系，回顾 Δ_f 的一个对象(可观测量代数的自同态)对应于可观测量代数的表象 $\pi_0 \circ \rho$。因此，如果 ρ 是玻色子的，我们称表象 $\pi_0 \circ \rho$ 是玻色子的，而如果 ρ 是费米子的，就称表象是费米子的。最后，如果它是在玻色子表象的扭结子里的，我们就叫一个态("波函数")玻色子的，或者如果它是在费米子表象的扭结子里的，就叫费米子的。这个断言被评价为：(i)这个规定式定义恰当重现了这样的区分，即在基本非相对论 QM 里所做的玻色子和费米子波函数之间的区分；(ii)这又告诉我们有关置换不变性的什么呢?

非相对论量子力学里的玻色—费米区分

在非相对论量子力学(QM)里，n 个全同粒子的状态空间是希尔伯特空间

的n次复制的张量积$H\otimes\cdots\otimes H$，希尔伯特空间$H\otimes\cdots\otimes H$的生成靠的是积态，即具有ψ_1，…，$\psi_n\in H$的形式如下的态：

$$\psi_1\otimes\cdots\otimes\psi_n$$

定义294。我们定义置换群S_n在$H\otimes\cdots\otimes H$上的自然作用如下，设$\{\psi_1, \cdots, \psi_m\}$是$H$的正交基，并且对每一个置换$\sigma$，有：

$$U(\sigma)(\psi_{i_1}\otimes\cdots\otimes\psi_{i_n})=\psi_{\sigma(i_1)}\otimes\cdots\otimes\psi_{\sigma(i_n)}$$

并且线性扩展$U(\sigma)$到全部的H。

如果$H>1$，那么S_n的表象U是可约化的。它包含了两个S_n的一维表象，即全同表象$S_n\rightarrow 1$，以及交替表象。按照全同表象变换的$H\otimes\cdots\otimes H$矢量的子空间被叫作对称子空间，而按照交替表象变换的矢量的子空间被叫作反对称子空间。在对称子空间里的矢量叫作玻色子的，而在反对称子空间里的矢量叫作费米子的，这些传统定义在哲学界里作为置换不变性讨论的基础。

在QM里，只是置换标记不同的n粒子态——比如，$\psi_1\otimes\psi_2$和$\psi_2\otimes\psi_1$——不应该算作是可分离的。就统计学权重的目的来说，这两个符号代表一个态。这被说成是置换不变性的原理。

置换不变性(PI)：设$\mathfrak{A}$是n粒子系统的可观测量，那么对每一个态ψ以及每一个置换$\sigma\in S_n$来说，我们有：

$$\langle U(\sigma)\psi, AU(\sigma)\psi\rangle=\langle\psi, A\psi\rangle$$

置换不变性有时也叫作不可区分性假设：置换不同的两个态是不可区分的(即没有测量能够在两者间作出区分)。这被说成是PI意味着态具有玻色统计或者费米统计，不允许存在具有“仲统计”(parastatistics)。

二分法：对于态矢量ψ和置换σ，我们具有$U(\sigma)\psi=\pm\psi$。

[van Fraassen, 1991, pp. 389ff]试图从PI证明二分法的说法，详情参看[Butterfield, 1993]。换言之，这些不在对称子空间或者反对称子空间里的态是多余的结构。

这给我们留下大量困惑。首先我们到底对既不在对称子空间也不在反对称子空间的$H\otimes\cdots\otimes H$里的矢量说了些什么？它们是多余的结构吗？它们具有并非偶然在自然界中实现的可能性吗？更一般地，不是所有在$H\otimes\cdots\otimes H$里的矢量都是确定对称类型；即便在那些确定对称类型，也不是全部对称的或者全部

反对称的。对于S_n的任何不可约化表象ξ，我们说在$H\otimes\cdots\otimes H$里的波函数ψ是对称类型ξ，当且仅当ψ被包含在对应于表象ξ的子空间里。那么$H\otimes\cdots\otimes H$是确定对称类型矢量的子空间的直和。而现在大量原则表明应该存在任何对称类型的粒子，为什么我们观测不到它们呢?

对称和反对称子空间的固有特性

我们从全n粒子希尔伯特空间$H\otimes\cdots\otimes H$开始，然后我们约化到对称子空间和反对称子空间。然后我们想知道$H\otimes\cdots\otimes H$的剩余元素跟什么有关。

对称子空间和反对称子空间的内在描述在于它们都是群S_n的表象。(事实上，它们跟S_n的一维不可约化表象是准等价的。)因此我们也能够反过来，也就是说，如果我们给出S_n的一个表象(H, π)，我们就能够进一步问这个表象的内部特性。回顾S_n的不可约化表象是一一对应于具有n个空格的杨表(Young tableaux)，参见[Simon，1996]。有一种把S_n的表象自然群化成超玻色和超费米的范畴：我们通过数对$(d(\pi), \omega_\pi)$具体化表象(H, π)，其中$d(\pi)\in\{1, 2, \cdots, n\}$和$\omega_\pi=\pm 1$。

1. 对于$(d, +1)$，所有杨表的列长小于或者等于d。(在此情况下，我们说π具有秩d超玻色统计。)

2. 对于$(d, -1)$，所有杨表的行长小于或者等于d。(在此情况下，我们说π具有秩d超费米统计。)

显然表象分成超玻色子和超费米子是相互排斥的，但是没有穷尽。(比如，存在S_n的表象包含既是1表象又是交错表象的复制。)

现在假设我们处于下面的情形(罗伯茨在开头引述所生动描述的)：我们被给予了可观测量$\mathfrak{A}$代数的一个纯态ω，并且问是否其态的"内部"统计是玻色子或者费米子的。我们能够做什么? 首先我们建构了GNS表象的由ω约化的$(\mathcal{H},\pi)$，至少这使事情更具体。但是希尔伯特空间$\mathcal{H}$自身不是一个张量积，由此不存在$\mathcal{H}$上的S_n自然表象。它也不会有助于以正常方式建构$\mathcal{H}$元素的张量积，因为$\psi\otimes\cdots\otimes\psi$平凡地是玻色子。因此，显而易见的方法似乎并不告知我们$\mathcal{H}$元素的内在对称类型的任何信息。

关键的见解再次来自于DHR分析：表象$(\mathcal{H}, \pi)$自然等距于对称张量*-范畴的对象ρ，即局域化可输运自同态的范畴Δ_f。由于Δ_f有积，我们能够构建

$\rho\otimes\rho$，而对称 $\varepsilon_{\rho,\rho}$ 给予我们置换 $\rho\otimes\rho$ 的概念，这里可回顾 $\varepsilon_{\rho,\rho}\in\mathrm{Hom}(\rho\otimes\rho)$。就像我们在下节将看到的，这给我们一个 S_n 在 $\mathrm{End}(\rho\otimes\rho)$ 里的自然表象 u。而且，$(d(\rho),\ \omega_\rho)$ 对，其中 $d(\rho)$ 是 ρ 的维度，而 ω_ρ 是 ρ 的统计维度，跟作为 S_n 的超玻色或者超费米表象的 u 分类是一致的。我们也会看到 S_n 的这个自然表象 u 对应于场代数 $\mathfrak{F}$ 的"较大"希尔伯特空间里波函数的置换。

在对称张量 * 范畴里 S_n 的表象

置换群 S_n 的幺正表象很自然出现在具有幺正对称的张量 * 范畴里。设 $(\mathcal{C},\ \otimes,\ 1)$ 是一个具有幺正对称 $(c_{X,Y})$ 的张量 * 范畴。固定一个对象 $X\in\mathrm{Obj}(\mathcal{C})$，并且定义一个映射 u：$S_2\to\mathrm{End}(X\otimes Y)$，即设定：

$$u((1))=\mathrm{id}_{X\otimes X},\qquad u((1,\ 2))=c_{X,X}$$

由于 $(c_{X,X})^2=\mathrm{id}_{X\otimes X}$，$u$ 是 S_2 在 $\mathrm{End}(X\otimes X)$ 的幺正表象。这个建构能够被重述为：定义映射 u：$S_2\to\mathrm{End}(X\otimes\cdots\otimes X)$，即设定：

$$u((i,\ i+1))=\mathrm{id}_X\otimes\cdots\otimes c_{X,X}\otimes\cdots\otimes\mathrm{id}_X$$

容易证明 u 唯一地扩展到 S_n 在 $\mathrm{End}(X\otimes\cdots\otimes X)$ 里的幺正表象。

事实 295。设 $\mathcal{C}$ 是具有幺正对称和共轭的张量 * 范畴，那么对于每一个不可约化对象 $X\in\mathrm{Obj}(\mathcal{C})$，$S_n$ 在 $\mathrm{End}(X\otimes\cdots\otimes X)$ 里的约化幺正表象 u 是秩 $d(X)$ 在 $\omega_X=+1$ 的超玻色子，而如果 $\omega_X=-1$ 是秩 $d(X)$ 的超费米子。而且，统计相位 ω_X 是 $u((1,\ 2))=c_{X,X}$ 的迹（详情参见附录）。

物理解释在场算符面前变得更清楚。（通常，重构定理的关键在于证明这样的算符总是有用的。）设 $(\mathcal{H},\ \mathfrak{F},\ (G,\ k))$ 是一个具有对可观测代数 $\mathfrak{A}$ 和真空态 ω 规范对称的场系统。设 $O_1,\ \cdots,\ O_n$ 是相互类空分离的区域，设 ρ 是在 Δ_f 里不可约化对象，然后使用 ρ 的可输运性，我们可能选择 $F_i\in\mathfrak{F}(O_i)$ 使得 $F_i\Omega$ 是在截面 $\mathcal{H}_{\hat{\rho}}$ 里（回顾截面是在 Δ_f 里对象的幺正等价类 $\hat{\rho}$ 标记的）。换言之，F_i 产生了区域 O_i 里的荷 $\hat{\rho}$。设 σ 是 $\{1,\ \cdots,\ n\}$ 的置换并且考虑下面两个 $\mathcal{H}$ 里的态矢量：

$$\psi_1\times\psi_2\times\cdots\times\psi_n\equiv F_1F_2\cdots F_n\Omega \tag{52}$$

$$\psi_{\sigma(1)}\times\psi_{\sigma(2)}\times\cdots\times\psi_{\sigma(n)}\equiv F_{\sigma(1)}F_{\sigma(2)}\cdots F_{\sigma(n)}\Omega \tag{53}$$

这两个矢量通常是可区分的。事实上，如果场网具有正常对易关系，那么我们能够计算两个矢量之间的差异。假定 σ 只是置换了两个数，如果 ρ 是玻色

子，这两个矢量就会相同，而如果ρ是费米子，这两个矢量差一个符号。但是这两个矢量总是约化到可观测量$\pi(\mathfrak{A})$代数上同样的态上。的确，如果$\rho_i \in \Delta_f(O_i)$是相应态射，那么这两个矢量约化的态分别是：

$$\omega\circ(\rho_1\otimes\rho_2\otimes\cdots\otimes\rho_n)=\omega\circ(\rho_1\rho_2\cdots\rho_n) \tag{54}$$

$$\omega\circ(\rho_{\sigma(1)}\otimes\rho_{\sigma(2)}\otimes\cdots\otimes\rho_{\sigma(n)})=\omega\circ(\rho_{\sigma(1)}\rho_{\sigma(2)}\cdots\rho_{\sigma(n)}) \tag{55}$$

由于在类空分离区域局域化的自同态对易，这两个态是相等的。因此，置换不变性为可观测量而不是为场所保留。

统计学 DHR 处理的解释性结果如下：置换不变性是在下述意义上的规范对称，它保留所有可观测量值的不变性，但是改变了赋予场算符的值。通过置换联系起来的两个态是同样的还是不同的？当然，对数学问题的回答是清楚的：可观测量代数的态是同样的，场代数的态是不同的。因此，是否我们取置换对应于“真实”改变，取决于我们对场算符地位的考虑。因此置换不变性的问题仅仅是规范不变性问题的特殊版本，并相应地跟不等价表象的地位问题紧密关联。

超统计学和非阿贝尔规范群

抽象田中定理(附录 B)证明了每一个对称张量＊范畴(STC*)$\mathcal{C}$是等价于紧致子群(G,k)的表象范畴$\mathrm{Rep}_f(G,k)$。就我们的主题而言，定理证明局域化可输运态射的范畴Δ_f等价于规范群的表象范畴。而且，B.9 节证明 STC*的每一个对象X，很自然地产生对称群S_n在$\mathrm{End}(X\otimes\cdots\otimes X)$里的幺正表象，并且这个表象对应于$X$的内在统计特征。由于我们知道$(G,k)$的$(H,\pi)$表象的范畴维度对应于基本的希尔伯特空间$H$的维度，因此：

引理 296。范畴$\mathrm{Rep}_f(G,k)$具有维度大于 1 的不可约化的对象，当且仅当G是非阿贝尔群。

证明思路。G的不可约化表象集合分离G的元素，因此对于$gh\neq hg$，存在一个不可约化表象(H,π)，使得$\pi(g)\pi(h)\neq\pi(h)\pi(g)$，因此$\dim H\geqslant 2$。

联合嵌入函子保留维度的事实，直接就可以从事实 295 得到，即：

命题 297。存在一个具有超统计学的$\mathcal{C}\simeq\mathrm{Rep}_f(G,k)$的一个不可约化的对象$X$，当且仅当对应的群$G$是非阿贝尔群。

就我们目前的情况而言，这意味着存在一个具有超统计学的表象和态，当

且仅当规范群G是非阿贝尔群。① 但是我们有好的理由认为非阿贝尔规范群是物理上相关的。因此，DHR 方法毫不担心假定不存在超粒子态，以及破坏二分法证明的基本断言。

辫子群统计学

回顾 8.1 节时空是维度为 2，那么 $\varepsilon_{\rho_1,\rho_2}$不必是一个 Δ_f 上的对称，而仅仅是一个辫子化。在此情况下，在 Δ_f 里的对象并不按照对称群 S_n 的表象分类，而是 Δ_f 里的对象通过辫子群 B_n 的表象方式分类。用物理术语来说，态可能不是置换不变性，而是满足更一般的辫子群统计。

定义 298。n 束上辫子群 B_n 是有满足下面方程的大集合$\{\sigma_1, \cdots, \sigma_{n-1}\}$产生的群：

$$(1)\quad \sigma_i\sigma_j = \sigma_j\sigma_i, \qquad |i-j| \geqslant 2$$

$$(2)\quad \sigma_{i+1}\sigma_i\sigma_{i+1} = \sigma_i\sigma_{i+1}\sigma_i$$

n 束上辫子群 B_n 是能够被给出下面启发性的几何描述：一个辫子由基点的两个集合$\{a_1, \cdots, a_n\}$和$\{b_1, \cdots, b_n\}$组成，以及纱的 n 束，每一个纱束具有连在 a_i 的一个端点，而其他端点连在 b_j，而且每一个基点只接触一个纱束。我们等同两个位形，如果一个能够转换成另一个而不需要从基点分离束。在此图像下，同样的辫子具有连接到 b_i 的 a_i，并且没有缠绕。产生元素 σ_i 能够被看成是简单辫子，其中：a_i 连接到 b_{i+1}，a_{i+1}连接到 b_i，并且这两个束仅仅缠绕一次（否则 σ_i 辫子就像同样的辫子）。在这个解释下，两个辫子的积 gh 是通过连接 g 终点和 h 的起点形成的辫子。

命题 299。对于每一个 $n \in \mathbb{N}$，映射：

$$\varepsilon_\rho^{(n)}(\sigma_i) = \rho^{i-1}(\varepsilon_\rho) = I_\rho \times \cdots \times I_\rho \times \varepsilon_\rho \times I_\rho \times \cdots \times I_\rho$$

定义了一个在 End$(\rho\otimes\cdots\otimes\rho)$里辫子群 B_n 的幺正表象，对于每一个 i，$j \in \mathbb{N}$，$i \leqslant j$，我们有：

$$\varepsilon_\rho^{(i)}(g) = \varepsilon_\rho^{(j)}(\varphi_{ij}(g)), \qquad \forall_g \in S_i$$

这个定理的证明能够在[Rehren *et al.*（雷尔恩等），1990]找到，还可以从

① 但是用“超统计学”具有模糊性，我们指的是超玻色统计和超费米统计，而非玻色统计和费米统计的混合。

附录 B 的考虑明确给出。换言之，积对象$\rho\otimes\cdots\otimes\rho$ 能达到辫子群的幺正表象，它是由下面形式的幺正算子约化的：

$$I_\rho\times\cdots\times\varepsilon_{\rho,\rho}\times\cdots\times I_\rho$$

这表示了 ρ 的第 i 个复制和第$(i+1)$个复制之间的一个基本置换。

存在一个辫子群到对称群的自然自同态置换。有一点是明显的，即只要我们回顾一下对称群的定义，是精确地跟辫子群的定义相同的，后者具有每个产生子是其自身的逆附加条件。因此戴克(Dyck)的定理[Hungerford(亨格福德)，1980，78]意味着，生成元上的明显映射，唯一地扩展到群同态$\bar{f}$：$B_n\rightarrow S_n$。因此，S_n 的每一个表象 π 产生 B_n 的一个表象 $\pi\circ\bar{f}$。其口号是：服从置换统计的系统也服从辫子统计。

现在回想到令人担心的超粒子存在论证：应该存在对应于S_n 的全部不可约化表象的粒子。对于 $n\geqslant3$，不存在 S_n 的玻色表象或者费米表象，因此也将存在超粒子。

现在我们能够明白，要么这个论证的某个地方出了问题，要么问题比我们想象的更严重。由于任何具有作为对称群 S_n 的系统，也具有作为对称群 B_n，这个论证向我们保证预测对应于 B_n 的所有不可约化表象的粒子存在，但是 B_n 具有无穷多不可约化表象。(的确，它的表象具有至今反对的分类。)而且，我们现在可能重复任何群 K 的论证，这样的论证能够被同态地映射到 B_n，并且存在具有这些属性的无穷链大群。因此，将充足的原则应用到这样群表象，预言了比我们所有能够描述的粒子更多的粒子。

注释：对 DHR 表象的统计学讨论，参看[Roberts，1976；Doplicher，1975]。

致谢

汉斯·霍尔沃森想感谢的是：迈克·缪格(Michael Müger)教会他多普利克尔—罗伯茨定理；感谢编辑有用的反馈和耐心；以及大卫·贝克(David Baker)、特雷西·卢浮(Tracy Lupher)和大卫·马拉蒙特(David Malament)的纠错。迈克·缪格特别致谢 NWO 基金支持(兰兹曼的授权书号是 no. 616. 062. 384)，还要感谢朱丽恩·比雄(Julien Bichen)对附录的批判性阅读和有益意见。

参考文献

[Araki, 1961] Huzihiro Araki. On the connection of spin and commutation relations between different fields. *Journal of Mathematical Physics*, 2: 267-270, 1961.

[Araki, 1963] Huzihiro Araki. A lattice of von Neumann algebras associated with the quantum theory of a free Bose field. *Journal of Mathematical Physics*, 4, 1963.

[Araki, 1999] Huzihiro Araki. *Mathematical theory of quantum fields*. Oxford University Press, New York, 1999.

[Arntzenius, 2003] Frank Arntzenius. Is quantum mechanics pointless? *Philosophy of Science*, 70(5): 1447-1457, 2003.

[Bacciagaluppi, 1994] G. Bacciagaluppi. Separation theorems and Bell inequalities in algebraic quantum mechanics. In *Symposium on the Foundations of Modern Physics* 1993: *Quantum Measurement, Irreversibility and the Physics of Information*, pages 29-37. World Scientific, Singapore, 1994.

[Baez *et al.*, 1992] John Baez, Irving Segal, and Zheng-Fang Zhou. *Introduction to algebraic and constructive quantum field theory*. Princeton University Press, Princeton, NJ, 1992.

[Baez, 1987] John Baez. Bell's inequality for C^*-algebras. *Letters in Mathematical Physics*, 13(2): 135-136, 1987.

[Baumgärtel and Wollenberg, 1992] Hellmut Baumgärtel and Manfred Wollenberg. *Causal nets of operator algebras*. Akademie-Verlag, Berlin, 1992.

[Bell, 1987] J. S. Bell. Beables for quantum field theory. In *Quantum implications*, pages 227-234. Routledge & Kegan Paul, London, 1987.

[Beltrametti and Cassinelli, 1981] Enrico G. Beltrametti and Gianni Cassinelli. *The logic of quantum mechanics*. Addison-Wesley Publishing Co., Reading, Mass., 1981.

[Bichon, 1998] Julien Bichon. Trivialisations dans les catégories tannakiennes. *Cahiers de Topologie et Géométrie Différentielle Catégoriques*, 39(4): 243-270, 1998.

[Bichon, 1999] Julien Bichon. Galois extension for a compact quantum group. math. QA/9902031, 1999.

[Bichon, ND] Julien Bichon. Trivialisations dans les catégories tannakiennes. Preprint version of [Bichon, 1998], ND.

[Bohm and Hiley, 1995] D. Bohm and B. J. Hiley. *The undivided universe*. Routledge, London, 1995.

[Bohr and Rosenfeld, 1950] Niels Bohr and Leon Rosenfeld. Field and charge measurements in quantum electrodynamics. *Physical Review*, 78: 794-798, 1950.

[Borchers, 1967] H. -J. Borchers. A remark on a theorem of B. Misra. *Communications in Mathematical Physics*, 4: 315-323, 1967.

[Borchers, 2000] H. J. Borchers. On revolutionizing quantum field theory with Tomita's modular theory. *Journal of Mathematical Physics*, 41: 3604-3673, 2000.

[Bostelmann, 2000] Henning Bostelmann. *Lokale Algebren und Operatorprodukte am Punkt*. PhD thesis, Universität Göttingen, 2000.

[Bostelmann, 2004] Henning Bostelmann. Phase space properties and the short distance structure in quantum field theory. math-ph/0409070, 2004.

[Bratteli and Robinson, 1997] Ola Bratteli and Derek W. Robinson. *Operator algebras and quantum statistical mechanics*, volume 2. Springer, NY, 1997.

[Buchholz and Fredenhagen, 1982] Detlev Buchholz and Klaus Fredenhagen. Locality and the structure of particle states. *Communications in Mathematical Physics*, 84: 1-54, 1982.

[Buchholz and Summers, ND] Detlev Buchholz and Stephen Summers. Scattering theory in relativistic quantum field theory. In J. P. Françoise, G. Naber, and T. S. Tsun, editors, *Encyclopedia of Mathematical Physics*. Elsevier, ND.

[Buchholz and Verch, 1995] Detlev Buchholz and Rainer Verch. Scaling algebras and renormalization group in algebraic quantum field theory. *Reviews in Mathematical Physics*, 7(8): 1195-1239, 1995.

[Buchholz and Wichmann, 1986] Detlev Buchholz and Eyvind Wichmann. Causal independence and the energy-level density of states in local quantum field theory. *Communications in mathematical physics*, 106: 321-344, 1986.

[Buchholz *et al.*, 1987] D. Buchholz, C. D'Antoni, and K. Fredenhagen. The universal structure of local algebras. *Communications in Mathematical Physics*, 111 (1): 123-135, 1987.

[Buchholz *et al.*, 1991] Detlev Buchholz, Martin Porrmann, and Ulrich Stein. Dirac versus Wigner. Towards a universal particle concept in local quantum field theory. *Physics Letters. B*, 267(3): 377-381, 1991.

[Buchholz *et al.*, 1992] Detlev Buchholz, Sergio Doplicher, Roberto Longo, and John E. Roberts. A new look at Goldstone's theorem. *Reviews in Mathematical Physics*, Special Issue: 49-83, 1992.

[Buchholz, 1974] Detlev Buchholz. Product states for local algebras. *Communications in Mathematical Physics*, 36: 287-304, 1974.

[Buchholz, 1998] Detlev Buchholz. Scaling algebras in local relativistic quantum physics. In *Proceedings of the International Congress of Mathematicians, Vol. III (Berlin, 1998)*, volume Extra Vol. III, pages 109-112 (electronic), 1998.

[Butterfield, 1993] Jeremy Butterfield. Interpretation and identity in quantum theory. *Studies in History and Philosophy of Modern Physics*, 24: 443-476, 1993.

[Carpi and Conti, 2001] Sebastiano Carpi and Roberto Conti. Classification of subsystems for local nets with trivial superselection structure. *Communications in Mathematical Physics*, 217: 89-106, 2001.

[Chaiken, 1967] Jan M. Chaiken. Finite-particle representations and states of the canonical commutation relations. *Annals of Physics*, 42: 23-80, 1967.

[Chaiken, 1968] Jan M. Chaiken. Number operators for representations of the canonical commutation relations. *Communications in Mathematical Physics*, 8: 164-184, 1968.

[Clifton and Halvorson, 2001a] Rob Clifton and Hans Halvorson. Are Rindler quanta real? Inequivalent particle concepts in quantum field theory. *British Journal for the Philosophy of Science*, 52(3): 417-470, 2001.

[Clifton and Halvorson, 2001b] Rob Clifton and Hans Halvorson. Entanglement and open systems in algebraic quantum field theory. *Studies in History and Philosophy of Modern Physics*, 32(1): 1-31, 2001.

[Clifton, 1995] Rob Clifton. Independently motivating the Kochen-Dieks modal interpretation of quantum mechanics. *British Journal for the Philosophy of Science*, 46(1): 33-57, 1995.

[Clifton, 2000] Rob Clifton. The modal interpretation of algebraic quantum field theory. *Physics Letters A*, 271(3): 167-177, 2000.

[Connes and Stømer, 1978] Alain Connes and Erling Stømer. Homogeneity of the state space of factors of type III_1. *Journal of Functional Analysis*, 28(2): 187-196, 1978.

[Connes, 1994] Alain Connes. *Noncommutative geometry*. Academic Press Inc., San Diego, CA, 1994.

[Conti *et al.* , 2001] Roberto Conti, Sergio Doplicher, and John E. Roberts. Superselection theory for subsystems. *Communications in Mathematical Physics*, 218: 263-281, 2001.

[Cushing, 1994] James Cushing. Locality/separability: Is this necessarily a useful distinction? In D. Hull, M. Forbes, and R. Burian, editors, *Proceedings of PSA* 1994, volume I, pages 107-116. Philosophy of Science Association, East Lansing, Michigan, 1994.

[D'Antoni, 1990] C. D'Antoni. Technical properties of the quasi-local algebra. In *The algebraic theory of superselection sectors (Palermo*, 1989), pages 248-258. World Sci. Publishing, River Edge, NJ, 1990.

[Deligne and Milne, 1982] P. Deligne and J. S. Milne. Tannakian categories. In *Lecture Notes in Mathematics*, volume 900, pages 101-228. Springer Verlag, 1982.

[Deligne and Morgan, 1999] Pierre Deligne and John W. Morgan. Notes on supersymmetry (following Joseph Bernstein). In *Quantum fields and strings: a course for mathematicians, Vol.* 1, 2 (*Princeton*, *NJ*, 1996/1997), pages 41-97. Amer. Math. Soc. , Providence, RI, 1999.

[Deligne, 1990] Pierre Deligne. Catégories tannakiennes. In *The Grothendieck Festschrift, Vol. II*, volume 87, pages 111-195. Birkhäuser Boston, Boston, MA, 1990.

[Deligne, 2002] P. Deligne. Catégories tensorielles. *Moscow Mathematical Journal*, 2(2): 227-248, 2002.

[Dickson and Clifton, 1998] Michael Dickson and Rob Clifton. Lorentz invariance in the modal interpretation. In Dennis Dieks and Pieter Vermaas, editors, *The modal interpretation of quantum mechanics*, pages 9-47. Kluwer, 1998.

[Dieks and Vermaas, 1998] Dennis Dieks and Pieter E. Vermaas, editors. *The modal interpretation of quantum mechanics.* Kluwer Academic Publishers, Dordrecht, 1998.

[Dieks, 2000] Dennis Dieks. Consistent histories and relativistic invariance in the modal interpretation of quantum mechanics. *Physics Letters. A*, 265(5-6): 317-325, 2000.

[Doplicher and Longo, 1984] S. Doplicher and R. Longo. Standard and split inclusions of von Neumann algebras. *Invent. Math.* , 75(3): 493-536, 1984.

[Doplicher and Piacitelli, 2002] Sergio Doplicher and Gherardo Piacitelli. Any compact group is a gauge group. *Reviews in Mathematical Physics*, 14(7-8): 873-885, 2002.

[Doplicher and Roberts, 1972] Sergio Doplicher and John E. Roberts. Fields, statistics and nonabelian gauge groups. *Communications in Mathematical Physics*, 28: 331-348, 1972.

[Doplicher and Roberts, 1987] Sergio Doplicher and John E. Roberts. C^*-algebras and duality for compact groups: why there is a compact group of internal gauge symmetries in particle physics. In *VIIIth international congress on mathematical physics* (*Marseille*, 1986), pages 489-498. World Scientific Publishing, Singapore, 1987.

[Doplicher and Roberts, 1989] Sergio Doplicher and John E. Roberts. A new duality theory for compact groups. *Inventiones Mathematicae*, 98(1): 157-218, 1989.

[Doplicher and Roberts, 1990] Sergio Doplicher and John E. Roberts. Why there is a field algebra with a compact gauge group describing the superselection structure in particle physics. *Communications in Mathematical Physics*, 131(1): 51-107, 1990.

[Doplicher *et al.*, 1969a] Sergio Doplicher, Rudolf Haag, and John E. Roberts. Fields, observables and gauge transformations. I. *Communications in Mathematical Physics*, 13: 1-23, 1969.

[Doplicher *et al.*, 1969b] Sergio Doplicher, Rudolf Haag, and John E. Roberts. Fields, observables and gauge transformations. II. *Communications in Mathematical Physics*, 15: 173-200, 1969.

[Doplicher *et al.*, 1971] Sergio Doplicher, Rudolf Haag, and John E. Roberts. Local observables and particle statistics. I. *Communications in Mathematical Physics*, 23: 199-230, 1971.

[Doplicher *et al.*, 1974] Sergio Doplicher, Rudolf Haag, and John E. Roberts. Local observables and particle statistics. II. *Communications in Mathematical Physics*, 35: 49-85, 1974.

[Doplicher, 1975] Sergio Doplicher. The statistics of particles in local quantum theories. In *International Symposium on Mathematical Problems in Theoretical Physics* (*Kyoto Univ.*, *Kyoto*, 1975), pages 264-273. Springer, Berlin, 1975.

[Doplicher, 1991] Sergio Doplicher. Abstract compact group duals, operator algebras and quantum field theory. In *Proceedings of the International Congress of Mathematicians*, *Vol. I*, *II* (*Kyoto*, 1990), pages 1319-1333, Tokyo, 1991. Math. Soc. Japan.

[Doplicher, 1992] Sergio Doplicher. Progress and problems in algebraic quantum field theory. In *Ideas and methods in quantum and statistical physics* (*Oslo*, 1988), pages 390-404. Cambridge Univ. Press, Cambridge, 1992.

[Doplicher, 1993] Sergio Doplicher. Operator algebras and abstract duals: progress and problems. In *Quantum and non-commutative analysis* (*Kyoto*, 1992), pages 405-418. Kluwer

Acad. Publ. , Dordrecht, 1993.

[Doplicher, 1995] Sergio Doplicher. Quantum field theory, categories and duality. In *Advances in dynamical systems and quantum physics* (*Capri*, 1993), pages 106-116. World Sci. Publishing, River Edge, NJ, 1995.

[Driessler, 1975] W. Driessler. Comments on lightlike translations and applications in relativistic quantum field theory. *Communications in Mathematical Physics*, 44 (2): 133-141, 1975.

[Earman, 2004] John Earman. Thoroughly modern McTaggart. *Philosopher's Imprint*, 2: 1-28, 2004.

[Emch, 1972] Gérard G. Emch. *Algebraic methods in statistical mechanics and quantum field theory*. Wiley Interscience, NY, 1972.

[Fleming, 2000] Gordon N. Fleming. Reeh-Schlieder meets Newton-Wigner. *Philosophy of Science*, 67(3, suppl.): S495-S515, 2000. PSA 1998, Part II (Kansas City, MO).

[Florig and Summers, 1997] Martin Florig and Stephen J. Summers. On the statistical independenc of algebras of observables. *Journal of Mathematical Physics*, 38 (3): 1318-1328, 1997.

[Folland, 1995] Gerald B. Folland. *A course in abstract harmonic analysis*. CRC Press, Boca Raton, FL, 1995.

[Fredenhagen and Hertel, 1981] Klaus Fredenhagen and Joachim Hertel. Local algebras of observables and pointlike localized fields. *Comm. Math. Phys.* , 80(4): 555-561, 1981.

[Fredenhagen, 1981] Klaus Fredenhagen. On the existence of anti particles. *Communications in Mathematical Physics*, 79: 141-151, 1981.

[Fredenhagen, 1992] Klaus Fredenhagen. Observables, superselection sectors and gauge groups. In *New symmetry principles in quantum field theory* (*Cargese*, 1991), volume 295, pages 177-194. Plenum, New York, 1992.

[Fredenhagen, 1993] Klaus Fredenhagen. Global observables in local quantum physics. In *Quantum and non-commutative analysis*. Kluwer, Amsterdam, 1993.

[Fredenhagen, 1994] Klaus Fredenhagen. Superselection sectors. http: //unith. desy. de/research/aqft/lecturenotes/, 1994.

[French and Rickles, 2003] Stephen French and Dean Rickles. Understanding permutation symmetry. In Katherine Brading and Elena Castellani, editors, *Symmetries in Physics*, pages

212-238. Cambridge University Press, 2003.

[French, 2000] Steven French. Identity and individuality in quantum theory. 2000. http://plato.stanford.edu/entries/qt-idind/.

[Gabriel, 1962] Pierre Gabriel. Des catégories abéliennes. *Bulletin de la Société Mathématique de France*, 90: 323-448, 1962.

[Giulini, 2003] D. Giulini. Superselection rules and symmetries. In *Decoherence and the appearance of a classical world in quantum theory*. Springer, NY, 2nd edition, 2003.

[Guido and Longo, 1992] Daniele Guido and Roberto Longo. Relativistic invariance and charge conjugation in quantum field theory. *Communications in Mathematical Physics*, 148: 521-551, 1992.

[Haag and Swieca, 1965] R. Haag and J. A. Swieca. When does a quantum field theory describe particles? *Communications in Mathematical Physics*, 1: 308-320, 1965.

[Haag, 1996] Rudolf Haag. *Local quantum physics*. Springer-Verlag, Berlin, second edition, 1996.

[Hái, 2002] Phúng Hô Hái. On a theorem of Deligne on characterization of Tannakian categories. In *Arithmetic fundamental groups and noncommutative algebra (Berkeley, CA, 1999)*, volume 70 of *Proc. Sympos. Pure Math.*, pages 517-531. Amer. Math. Soc., Providence, RI, 2002.

[Halvorson and Clifton, 2000] Hans Halvorson and Rob Clifton. Generic Bell correlation between arbitrary local algebras in quantum field theory. *Journal of Mathematical Physics*, 41 (4): 1711-1717, 2000.

[Halvorson and Clifton, 2002] Hans Halvorson and Rob Clifton. No place for particles in relativistic quantum theories? *Philosophy of Science*, 69(1): 1-28, 2002.

[Halvorson, 2001] Hans Halvorson. Reeh-Schlieder defeats Newton-Wigner: on alternative localization schemes in relativistic quantum field theory. *Philosophy of Science*, 68(1): 111-133, 2001.

[Halvorson, 2004] Hans Halvorson. Complementarity of representations in quantum mechanics. *Studies in History and Philosophy of Modern Physics*, 35(1): 45-56, 2004.

[Holland, 1995] Peter R. Holland. *The quantum theory of motion*. Cambridge University Press, Cambridge, 1995.

[Horuzhy, 1990] S. S. Horuzhy. *Introduction to algebraic quantum field theory*. Kluwer

Academic Publishers, Dordrecht, 1990.

[Huggett, 1999] Nick Huggett. On the significance of permutation symmetry. *British Journal for the Philosophy of Science*, 50(3): 325-347, 1999.

[Huggett, 2000] Nick Huggett. Philosophical foundations of quantum field theory. *British Journal for the Philosophy of Science*, 51: 617-637, 2000.

[Hungerford, 1980] Thomas W. Hungerford. *Algebra*. Springer-Verlag, New York, 1980.

[Joyal and Street, 1991] André Joyal and Ross Street. An introduction to Tannaka duality and quantum groups. In *Category theory* (*Como*, 1990), volume 1488 of *Lecture Notes in Math.*, pages 413-492. Springer, Berlin, 1991.

[Joyal and Street, 1993a] André Joyal and Ross Street. Braided tensor categories. *Advances in Mathematics*, 102(1): 20-78, 1993.

[Kadison and Ringrose, 1997] Richard V. Kadison and John R. Ringrose. *Fundamentals of the theory of operator algebras. Vol. I.* American Mathematical Society, Providence, RI, 1997.

[Kassel, 1995] Christian Kassel. *Quantum groups*. Springer-Verlag, New York, 1995.

[Kitajima, 2004] Yuichiro Kitajima. A remark on the modal interpretation of algebraic quantum field theory. *Physics Letters A*, 331(3-4): 181-186, 2004.

[Kraus, 1983] Karl Kraus. *States, effects, and operations*. Springer-Verlag, Berlin, 1983.

[Landau, 1987] Lawrence J. Landau. On the violation of Bell's inequality in quantum theory. *Physics Letters. A*, 120(2): 54-56, 1987.

[Landsman, 1995] N. P. Landsman. Observation and superselection in quantum mechanics. *Studies in History and Philosophy of Modern Physics*, 26(1): 45-73, 1995.

[Longo and Roberts, 1997] Roberto Longo and John E. Roberts. A theory of dimension. *KTheory*, 11(2): 103-159, 1997.

[Longo, 1979] Roberto Longo. Notes on algebraic invariants for noncommutative dynamical systems. *Communications in Mathematical Physics*, 69(3): 195-207, 1979.

[Longo, 1989] Roberto Longo. Index of subfactors and statistics of quantum fields, I. *Communications in Mathematical Physics*, 126: 217-247, 1989.

[Mac Lane, 1998] Saunders Mac Lane. *Categories for the working mathematician*. Springer-Verlag, New York, second edition, 1998.

[McIntosh, 1970] Alan McIntosh. Hermitian bilinear forms which are not semibounded. *Bulletin of the American Mathematical Society*, 76: 732-737, 1970.

[Müger and Tuset, 2006] Michael Müger and Lars Tuset. Monoids, embedding functors and quantum groups. math. QA/0604065, 2006.

[Müger *et al.*, 2004] Michael Müger, John E. Roberts, and Lars Tuset. Representations of algebraic quantum groups and reconstruction theorems for tensor categories. *Algebras and Representation Theory*, 7(5): 517-573, 2004.

[Müger, 2000] Michael Müger. Galois theory for braided tensor categories and the modular closure. *Advances in Mathematics*, 150(2): 151-201, 2000.

[Myrvold, 2002] Wayne C. Myrvold. Modal interpretations and relativity. *Foundations of Physics*, 32(11): 1773-1784, 2002.

[Napiórkowski, 1972] Kazimierz Napiórkowski. On the independence of local algebras. *Reports on Mathematical Physics*, 3(1): 33-35, 1972.

[Pedersen, 1989] Gert K. Pedersen. *Analysis now.* Springer, New York, 1989.

[Rédei, 1998] Miklós Rédei. *Quantum logic in algebraic approach.* Kluwer Academic Publishers Group, Dordrecht, 1998.

[Redhead, 1995a] Michael L. G. Redhead. More ado about nothing. *Foundations of Physics*, 25(1): 123-137, 1995.

[Redhead, 1995b] Michael L. G. Redhead. The vacuum in relativistic quantum field theory. In David Hull, Micky Forbes, and Richard M. Burian, editors, *PSA; proceedings of the biennial meeting of the Philosophy of Science Association*, volume 2, pages 77-87, East Lansing, MI, 1995. Philosophy of Science Association.

[Reeh and Schlieder, 1961] H. Reeh and S. Schlieder. Bemerkungen zur Unitäräquivalenz von Lorentzinvarianten Felden. *Nuovo cimento* (10), 22: 1051-1068, 1961.

[Rehberg and Wollenberg, 1986] Joachim Rehberg and Manfred Wollenberg. Quantum fields as pointlike localized objects. *Mathematische Nachrichten*, 125: 259-274, 1986.

[Rehren *et al.*, 1990] Karl-Henning Rehren, Daniel Kastler, and M. Mebkhout. Introduction to the algebraic theory of superselection sectors. In Daniel Kastler, editor, *Algebraic Theory of Superselection Sectors*, pages 113-214. World Scientific, Singapore, 1990.

[Requardt, 1986] Manfred Requardt. Reeh-Schlieder-type density results in one-and *n*-body Schrödinger theory and the "unique continuation problem". *Journal of Mathematical Physics*, 27(6): 1571-1577, 1986.

[Roberts and Roepstorff, 1968] J. E. Roberts and G. Roepstorff. Some basic concepts of algebraic quantum theory. *Communications in Mathematical Physics*, 11: 321-338, 1968.

[Roberts, 1970] John E. Roberts. The structure of sectors reached by a field algebra. In *Cargese lectures in physics*, *Vol.* 4, pages 61-78. Gordon and Breach, New York, 1970.

[Roberts, 1974] John E. Roberts. Spontaneously broken symmetries and superselection rules. 1974. unpublished preprint.

[Roberts, 1975] John E. Roberts. Must there be a gauge group? *Trudy Mat. Inst. Steklov*, 135, 1975.

[Roberts, 1976] John E. Roberts. Statistics and the intertwiner calculus. In *C^*-algebras and their applications to statistical mechanics and quantum field theory* (*Proc. Internat. School of Physics "Enrico Fermi"*, *Course LX*, *Varenna*, 1973), pages 203-225. North-Holland, Amsterdam, 1976.

[Roberts, 1990] John E. Roberts. Lectures on algebraic quantum field theory. In Daniel Kastler, editor, *The algebraic theory of superselection sectors* (*Palermo*, 1989), pages 1-112. World Scientific Publishing, River Edge, NJ, 1990.

[Roberts, 2004] John E. Roberts. More lectures on algebraic quantum field theory. In Sergio Doplicher and Roberto Longo, editors, *Noncommutative geometry*, pages 263-342. Springer, Berlin, 2004.

[Roberts, ND] John E. Roberts. The reconstruction problem. Unpublished Manuscript, ND.

[Rosenberg, 2000] Alexander L. Rosenberg. The existence of fiber functors. In *The Gelfand Mathematical Seminars*, 1996-1999, Gelfand Math. Sem., pages 145-154. Birkhäuser Boston, Boston, MA, 2000.

[Rudin, 1991] Walter Rudin. *Functional analysis.* McGraw-Hill Inc., New York, second edition, 1991.

[Ruetsche and Earman, 2005] Laura Ruetsche and John Earman. Relativistic invariance and modal interpretations. Manuscript, University of Pittsburgh, 2005.

[Ruetsche, 2002] Laura Ruetsche. Interpreting quantum field theory. *Philosophy of Science*, 69(2): 348-378, 2002.

[Ruetsche, 2004] Laura Ruetsche. Intrinsically mixed states: an appreciation. *Studies in the History and Philosophy of Modern Physics*, 35: 221-239, 2004.

[Ryder, 1996] Lewis H. Ryder. *Quantum field theory.* Cambridge University Press, Cambridge second edition, 1996.

[Saavedra Rivano, 1972] Neantro Saavedra Rivano. *Catégories Tannakiennes.* Springer-Verlag, Berlin, 1972.

[Saunders, 1992] Simon Saunders. Locality, complex numbers, and relativistic quantum theory. In David Hull, Micky Forbes, and Kathleen Okruhlik, editors, *PSA; proceedings of the biennial meeting of the Philosophy of Science Association*, volume 1, pages 365-380, East Lansing, MI, 1992. Philosophy of Science Association.

[Schlieder, 1969] S. Schlieder. Einige Bemerkungen über Projektionsoperatoren (Konsequenzen eines Theorems von Borchers). *Communications in Mathematical Physics*, 13: 216-225, 1969.

[SGA, 1972] *Théorie des topos et cohomologie étale des schémas. Tome* 2. Springer-Verlag, Berlin, 1972. Séminaire de Géométrie Algébrique du Bois-Marie 1963-1964 (SGA 4), Dirigépar M. Artin, A. Grothendieck et J. L. Verdier. Avec la collaboration de N. Bourbaki, P. Deligne et B. Saint-Donat, Lecture Notes in Mathematics, Vol. 270.

[Simon, 1996] Barry Simon. *Representations of finite and compact groups.* American Mathematical Society, Providence, RI, 1996.

[Srinivas, 1980] M. D. Srinivas. Collapse postulate for observables with continuous spectra. *Communications in Mathematical Physics*, 71(2): 131-158, 1980.

[Stinespring, 1955] W. Forrest Stinespring. Positive functions on C^*-algebras. *Proceedings of the American Mathematical Society*, 6: 211-216, 1955.

[Størmer, 1970] Erling Størmer. Asymptotically abelian systems. In *Cargèse lectures in physics*, *Vol.* 4, pages 195-213. Gordon and Breach, New York, 1970.

[Streater and Wightman, 1964] R. F. Streater and A. S. Wightman. *PCT, spin and statistics and all that.* W. A. Benjamin, Inc., 1964.

[Summers and Buchholz, 2005] Stephen J. Summers and Detlev Buchholz. Quantum statistics and locality. *Physics Letters A*, 337: 17-21, 2005.

[Summers and Werner, 1987] Stephen J. Summers and Reinhard Werner. Bell's inequalities and quantum field theory. I. General setting. *Journal of Mathematical Physics*, 28(10): 2440-2447, 1987.

[Summers and Werner, 1988] Stephen J. Summers and Reinhard Werner. Maximal violation of Bell's inequalities for algebras of observables in tangent spacetime regions. *Annales de l'Institut Henri Poincaré. Physique Théorique*, 49(2): 215-243, 1988.

[Summers and Werner, 1995] Stephen J. Summers and Reinhard F. Werner. On Bell's inequalities and algebraic invariants. *Letters in Mathematical Physics*, 33(4): 321-334, 1995.

[Summers, 1990] Stephen J. Summers. On the independence of local algebras in quantum field theory. *Reviews in Mathematical Physics*, 2(2): 201-247, 1990.

[Summers, 1997] Stephen J. Summers. Bell's inequalities and algebraic structure. In *Operator algebras and quantum field theory* (*Rome*, 1996), pages 633-646. Internat. Press, Cambridge, MA, 1997.

[Summers, ND] Stephen Summers. Tomita-Takesaki modular theory. In J. P. Françoise, G. Naber, and T. S. Tsun, editors, *Encyclopedia of Mathematical Physics.* Elsevier, ND.

[Sunder, 1987] V. S. Sunder. *An invitation to von Neumann algebras.* Springer-Verlag, New York, 1987.

[Takesaki, 2002] M. Takesaki. *Theory of operator algebras. I.* Springer-Verlag, Berlin, 2002.

[Takesaki, 2003] M. Takesaki. *Theory of operator algebras. II.* Springer-Verlag, Berlin, 2003.

[Tannaka, 1939] T. Tannaka. über den Dualitätssatz der nichtkommutativen topologischen Gruppen. *Tôohoku Mathematics Journal*, 45: 1-12, 1939.

[Teller, 1995] Paul Teller. *An interpretive introduction to quantum field theory.* Princeton University Press, Princeton, NJ, 1995.

[van Fraassen, 1991] Bas C. van Fraassen. *Quantum mechanics: an empiricist view.* Oxford University Press, New York, 1991.

[Varadarajan, 2004] V. S. Varadarajan. *Supersymmetry for mathematicians: an introduction.* American Mathematical Society, Providence, RI, 2004.

[Vermaas, 1999] Pieter E. Vermaas. *A philosopher's understanding of quantum mechanics.* Cambridge University Press, Cambridge, 1999.

[Wald, 1994] Robert M. Wald. *Quantum field theory in curved spacetime and black hole thermodynamics.* University of Chicago Press, Chicago, IL, 1994.

[Wallace, forthcoming] David Wallace. In defense of naievete: The conceptual status of Lagrangian QFT. *Synthese*, forthcoming.

[Wick *et al.*, 1952] G. C. Wick, A. S. Wightman, and E. P. Wigner. The intrinsic parity of elementary particles. *Physical Review*, 88: 101-105, 1952.

[Wightman, 1964] A. S. Wightman. La théorie quantique locale et la théorie quantique des champs. *Ann. Inst. H. Poincaré Sect. A (N. S.)*, 1: 403-420, 1964.

[Wigner, 1939] E. Wigner. On unitary representations of the inhomogeneous Lorentz group. *Ann. of Math.* (2), 40(1): 149-204, 1939.

[Wizimirski, 1966] Z. Wizimirski. On the existence of the field of operators in the axiomatic quantum field theory. *Bulletin de l'Académie Polonaise des Sciences. Série des Sciences Mathématiques, Astronomiques et Physiques*, 14: 91-93, 1966.

[Wollenberg, 1986] Manfred Wollenberg. Quantum fields as pointlike localized objects. II. *Mathematische Nachrichten*, 128: 287-298, 1986.

[Yngvason, 2005] Jakob Yngvason. The role of type III factors in quantum field theory. *Reports on Mathematical Physics*, 55(1): 135-147, 2005.

附录　对称张量 * 范畴的抽象对偶性理论

迈克·缪格

这个附录的目的是给出定理 389 的一个证明，该定理在 1989 年首次被多普利克尔和罗伯茨证明，其根据是每一个具有共轭、直和、子对象以及End $\mathbf{1} = \mathbb{C}$ 的对称张量 * 范畴，等价于唯一确定紧致子群的有限维幺正表象的范畴。这些材料有部分不是新的，但是定理 403 可能算新的，也可参看评论 434。当然，这可能是第一个对称张量范畴重建定理的阐述，是既完备又简化而且效率高的证明，也是包含了田中的经典定理的一个简短明快的证明。在第一部分我们提供范畴论的必要概念和结果，其中不涉及纤维函子的概念，而第二部分就真正涉及田中理论。范畴论我们主要参考了[Mac Lane，1998]，尤其是第二版。已经对范畴论有所了解的读者可以直接跳到 B 部分，需要时再查看 A 部分。

A　范畴论预备知识

A1　基础

定义 300。一个范畴$\mathcal{C}$的组成如下。

- 对象的类 Obj $\mathcal{C}$，我们用大写字母 X，Y，…代表对象。

• 对任何两个对象 X，Y，有箭头（或者态射）的集合 $\mathrm{Hom}_{\mathcal{C}}(X, Y)$，我们写为 f：$X \to Y$ 表示 $f \in \mathrm{Hom}_{\mathcal{C}}(X, Y)$，在不引起混淆的情况下略去下标$\mathcal{C}$。

• 对任何对象 X 有一个识别箭头 $\mathrm{id}_X \in \mathrm{End}(X) = \mathrm{Hom}(X, X)$。

• 对每一个 X，Y，$Z \in \mathrm{Obj}\ \mathcal{C}$，有一个函数$\circ$：$\mathrm{Hom}(Y, Z) \times \mathrm{Hom}(X, Y) \to \mathrm{Hom}(X, Z)$，使得：

$$h \circ (g \circ f) = (h \circ g) \circ f \quad 和 \quad \mathrm{id}_Y \circ f = f,\ g \circ \mathrm{id}_Y = g$$

只要有 $f \in \mathrm{Hom}(X, Y)$，$g \in \mathrm{Hom}(Y, Z)$ 以及 $h \in \mathrm{Hom}(Z, W)$。

定义 301。一个态射 $f \in \mathrm{Hom}(X, Y)$ 是一个同构，当且仅当它是不可逆的，即存在一个 $g \in \mathrm{Hom}(Y, X)$，使得 $g \circ f = \mathrm{id}_X$ 和 $f \circ g = \mathrm{id}_Y$，如果一个同构 $X \to Y$ 存在，我们写成 $X \cong Y$。

定义 302。如果$\mathcal{C}$是个范畴，那么子范畴$\mathcal{D} \subset \mathcal{C}$被定义为一个子类 $\mathrm{Obj}\ \mathcal{D} \subset \mathrm{Obj}\ \mathcal{C}$，并且对于任何 X，$Y \in \mathrm{Obj}\ \mathcal{D}$，有一个子集合 $\mathrm{Hom}_{\mathcal{D}}(X, Y) \subset \mathrm{Hom}_{\mathcal{C}}(X, Y)$，使得对于所有 $X \in \mathrm{Obj}\ \mathcal{D}$，存在 $\mathrm{id}_X \in \mathrm{Hom}_{\mathcal{D}}(X, X)$，并且$\mathcal{D}$里的态射在$\mathcal{C}$的复合$\circ$下是封闭的。如果对所有 X，$Y \in \mathrm{Obj}\ \mathcal{D}$都存在 $\mathrm{Hom}_{\mathcal{D}}(X, Y) = \mathrm{Hom}_{\mathcal{C}}(X, Y)$，则一个子范畴$\mathcal{D} \subset \mathcal{C}$是满的。

定义 303。一个（协变）函子 F 从范畴$\mathcal{C}$到范畴$\mathcal{D}$，映射$\mathcal{C}$的对象到$\mathcal{D}$的对象，以及$\mathcal{C}$的箭头到$\mathcal{D}$的箭头，使得 $F(g \circ f) = F(g) \circ F(f)$，并且 $F(\mathrm{id}_X) = \mathrm{id}_{F(X)}$。一个逆变函子跟协变函子一样，除非它逆转了箭头的顺序。

定义 304。函子 F：$\mathcal{C} \to \mathcal{D}$是忠实的、各自满射的，如果映射

$$F_{X,Y}:\ \mathrm{Hom}_{\mathcal{C}}(X, Y) \to \mathrm{Hom}_{\mathcal{D}}(F(X), F(Y))$$

对所有 X，$Y \in \mathrm{Obj}\ \mathcal{C}$来说是内射的、各自满射的。

定义 305。一个函子 F：$\mathcal{C} \to \mathcal{D}$是本质上满射的，如果对任何 $Y \in \mathrm{Obj}\ \mathcal{D}$都存在 $X \in \mathrm{Obj}\ \mathcal{C}$，使得 $F(X) \cong Y$。

定义 306。如果 F：$\mathcal{C} \to \mathcal{D}$和 G：$\mathcal{C} \to \mathcal{D}$是函子，那么一个从 F 到 G 的自然变换 η 结合任何 $X \in \mathrm{Obj}\ \mathcal{C}$，有一个态射 $\eta_X \in \mathrm{Hom}_{\mathcal{D}}(F(X), G(X))$，使得：

$$\begin{tikzcd} F(X) \arrow[r, "F(s)"] \arrow[d, "\eta_X"'] & F(Y) \arrow[d, "\eta_Y"] \\ G(X) \arrow[r, "G(s)"'] & G(Y) \end{tikzcd}$$

对于任何箭头 $s\in \mathrm{Hom}_{\mathcal{C}}(X, Y)$交换。如果 η_X 是一个对任何 $X\in \mathrm{Obj}\ \mathcal{C}$定义的同构，那么 η 被说成是一个自然同构。

定义 307。一个函子 $F: \mathcal{C}\to\mathcal{D}$是范畴的等价，如果存在一个函子 $G: \mathcal{D}\to\mathcal{C}$ 和自然同构 $\eta: FG\to \mathrm{id}_{\mathcal{D}}$，以及 $\varepsilon: \mathrm{id}_{\mathcal{C}}\to GF$。如果存在一个等价 $F: \mathcal{C}\to\mathcal{D}$，那么两个范畴是等价的，表示为 $F\simeq G$。

定义 308。如果 Obj $\mathcal{C}$是个集合（而不正好是一类），那么一个范畴是小的。一个范畴是本质上小的，如果它等价于一个小的范畴，即 Obj $\mathcal{C}/\cong$是个集合。

评论 309。没有考虑进入基本技术层面，我们指出群的“全部表示”的范畴是个巨大的对象。然而，考虑到模等价性，这些表象是合理的基数，即是个集合。

A2 张量范畴和辫子

定义 310。给定两个范畴$\mathcal{C}$，$\mathcal{D}$，乘积范畴$\mathcal{C}\times\mathcal{D}$的定义是：

$\mathrm{Obj}(\mathcal{C}\times\mathcal{D}) = \mathrm{Obj}\ \mathcal{C}\times \mathrm{Obj}\ \mathcal{D}$

$\mathrm{Hom}_{\mathcal{C}\times\mathcal{D}}(X\times Y, Z\times W) = \mathrm{Hom}_{\mathcal{C}}(X, Z)\times \mathrm{Hom}_{\mathcal{D}}(Y, W)$

$\mathrm{id}_{X\times Y} = \mathrm{id}_X\times \mathrm{id}_Y$

具有明显的复合$(a\times b)\circ(c\times d) := (a\circ c)\times(b\circ d)$。

定义 311。一个严格的张量范畴（或者严格的幺半范畴）是一个具有特定对象 1、张量单位以及函子$\otimes$: $\mathcal{C}\times\mathcal{C}\to\mathcal{C}$的范畴$\mathcal{C}$，使得：

1. $\otimes$是在对象和态射上满足结合律，即对所有的 $X, Y, Z, X', Y', Z'\in$ Obj $\mathcal{C}$以及全部 $s: X\to X'$，$t: Y\to Y'$，$u: Z\to Z'$，都有$(X\otimes Y)\otimes Z = X\otimes(Y\otimes Z)$和$(s\otimes t)\otimes u = s\otimes(t\otimes u)$。

2. 单位对象情况自然是，对所有 $s: X\to Y$，存在 $X\otimes \mathbf{1} = X = \mathbf{1}\otimes X$ 和 $s\otimes \mathrm{id}_1 = s = \mathrm{id}_1\otimes s$。

3. 交换律：$(a\otimes b)\circ(c\otimes d) = (a\circ c)\otimes(b\circ d)$，当 $a\circ c$ 和 $b\circ d$ 被定义时成立。

评论 312。许多具有张量积的范畴并不受上述含义的限制，张量范畴是具有下面条件的范畴，包括函子$\otimes$: $\mathcal{C}\times\mathcal{C}\to\mathcal{C}$、单位 $\mathbf{1}$ 和自然同构 $\alpha_{X,Y,Z}$: $(X\otimes Y)\otimes Z\to X\otimes(Y\otimes Z)$，$\lambda_X$: $\mathbf{1}\otimes X\to X$，满足某些恒等式的 ρ_X: $X\otimes\mathbf{1}\to X$。辫子的概

念、幺半函子和幺半自然变换都推广到这些范畴。考虑非严格范畴所达到的普遍性是显而易见的：由于连贯定理，任何(辫子的/对称的)张量范畴幺半地自然等价于一个严格范畴。所有这些参看[Mac Lane，1998；Joyal and Street，1993a]。

严格说来(一语双关地说)矢量空间和希尔伯特空间的范畴并不严格，然而连贯定理让我们假定它们是严格的，因而大大简化了公式。读者如果觉得对此不放心，请将同构 α，λ，ρ 嵌入到其应该出现的地方。

定义 313。一个张量范畴$\mathcal{C}$的(完全)张量子范畴，是一个(完全)子范畴$\mathcal{D}\subset\mathcal{C}$，使得 Obj $\mathcal{D}$包含单位对象$\mathbf{1}$，并且在张量积$\otimes$下是封闭的。

定义 314。设$\mathcal{C}$，$\mathcal{D}$是严格张量范畴，一个张量函子(或者幺半函子)是一个F：$\mathcal{C}\to\mathcal{D}$函子，加上对于所有的 X，$Y\in\mathcal{C}$来说的同构 $d^F_{X,Y}$：$F(X)\otimes F(Y)\to F(X\otimes Y)$和态射 e^F：$\mathbf{1}_{\mathcal{D}}\to F(\mathbf{1}_{\mathcal{C}})$，使得：

1. 态射 $d^F_{X,Y}$就两个论证来说是自然的。
2. 对于所有的 X，Y，$Z\in\mathcal{C}$，有下面的交换图：

$$
\begin{tikzcd}
F(X)\otimes F(Y)\otimes F(Z) \arrow[r, "d^F_{X,Y}\otimes \mathrm{id}_{F(Z)}"] \arrow[d, "\mathrm{id}_{F(X)}\otimes d^F_{Y,Z}"'] & F(X\otimes Y)\otimes F(Z) \arrow[d, "d^F_{X\otimes Y,Z}"] \\
F(X)\otimes F(Y\otimes Z) \arrow[r, "d^F_{X,Y\otimes Z}"'] & F(X\otimes Y\otimes Z)
\end{tikzcd} \tag{56}
$$

3. 下面的结构是对所有的 $X\in\mathcal{C}$，$F(X)$的恒等态射：

$$
\begin{aligned}
&F(X)\equiv F(X)\otimes\mathbf{1}_{\mathcal{D}}\xrightarrow{\mathrm{id}_{F(X)}\otimes e^F}F(X)\otimes F(\mathbf{1}_{\mathcal{C}})\xrightarrow{d_{X,\mathbf{1}}}F(X\otimes\mathbf{1}_{\mathcal{C}})\equiv F(X)r\\
&F(X)\equiv\mathbf{1}_{\mathcal{D}}\otimes F(X)\xrightarrow{e^F\otimes\mathrm{id}_{F(X)}}F(\mathbf{1}_{\mathcal{C}})\otimes F(X)\xrightarrow{d_{\mathbf{1},X}}F(\mathbf{1}_{\mathcal{C}}\otimes X)\equiv F(X)
\end{aligned} \tag{57}
$$

如果$\mathcal{C}$，$\mathcal{D}$是张量 * 范畴而 F 是 * 保持的，那么同构 e，$d_{X,Y}$都要求是幺正的。

定义 315。设$\mathcal{C}$，$\mathcal{D}$是严格张量范畴，而 F，G：$\mathcal{C}\to\mathcal{D}$是张量函子，那么一个自然变换 α：$\mathcal{C}\to\mathcal{D}$是幺半的，如果：

$$\begin{tikzcd}
F(X)\otimes F(Y) \arrow[r, "d^F_{X,Y}"] \arrow[d, "\alpha_X\otimes\alpha_Y"'] & F(X\otimes Y) \arrow[d, "\alpha_{X\otimes Y}"] \\
G(X)\otimes G(Y) \arrow[r, "d^G_{X,Y}"'] & G(X\otimes Y)
\end{tikzcd}$$

对于所有 X，$Y\in\mathcal{C}$是对易的，并且合成$\mathbf{1}_{\mathcal{D}}\xrightarrow{e^F}F(\mathbf{1})\xrightarrow{\alpha_1}G(\mathbf{1})$跟 e^G 一致。

评论 316。如果所有的同构 $d_{X,Y}$和 e 是单位元，严格张量范畴之间的张量函子被称为是严格的。然而，不能说任何张量函子等价于一个严格的函子！

定义 317。如果存在一个张量函子 G：$\mathcal{D}\to\mathcal{C}$以及幺半自然同构 $GF\to\mathrm{id}_{\mathcal{C}}$ 和 $FG\to\mathrm{id}_{\mathcal{C}}$，那么一个张量函子 F：$\mathcal{C}\to\mathcal{D}$是个(张量范畴的)等价。

命题 318。一个函子 F：$\mathcal{C}\to\mathcal{D}$是一个等价，当且仅当 F 是忠实的、完全的以及本质上满射的。一个(严格)张量范畴的张量函子 F：$\mathcal{C}\to\mathcal{D}$，是一个张量范畴的等价，当且仅当 F 是忠实的、完全的以及本质上满射的。

证明。对于第一条说法参见[Mac Lane，1998，Theorem 1，p. 91]，而第二条参见[Saavedra Rivano(萨维德拉·雷瓦诺)，1972]。

定义 319。一个严格张量范畴$\mathcal{C}$的辫子，是对于所有 X，$Y\in\mathrm{Obj}\,\mathcal{C}$的一族同构 $c_{X,Y}$：$X\otimes Y\to Y\otimes X$，满足：

1. 自然性：对任何 s：$X\to X'$，t：$Y\to Y'$，下图对易。

$$\begin{tikzcd}
X\otimes Y \arrow[r, "c_{X,Y}"] \arrow[d, "s\otimes t"'] & Y\otimes X \arrow[d, "t\otimes s"] \\
X'\otimes Y' \arrow[r, "c_{X',Y'}"] & Y'\otimes X'
\end{tikzcd}$$

2. “辫子方程”保留，即下图对任何 X，Y，$Z\in\mathrm{Obj}\,\mathcal{C}$对易。

$$\begin{tikzcd}
X\otimes Y\otimes Z \arrow[r, "c_{X,Y}\otimes\mathrm{id}_Z"] \arrow[dr, "c_{X,Y\otimes Z}"'] & Y\otimes X\otimes Z \arrow[d, "\mathrm{id}_Y\otimes c_{X,Z}"] \\
& Y\otimes Z\otimes X
\end{tikzcd}
\qquad
\begin{tikzcd}
X\otimes Y\otimes Z \arrow[r, "\mathrm{id}_X\otimes c_{Y,Z}"] \arrow[dr, "c_{X\otimes Y,Z}"'] & X\otimes Z\otimes Y \arrow[d, "c_{X,Z}\otimes\mathrm{id}_Y"] \\
& Z\otimes X\otimes Y
\end{tikzcd}$$

另外，如果 $c_{Y,X}\circ c_{X,Y}=\mathrm{id}_{X\otimes Y}$对所有 X，Y 成立，辫子被说成是对称的。

一个严格辫子(对称)张量范畴，是具有一个辫子(对称)的严格张量范畴。

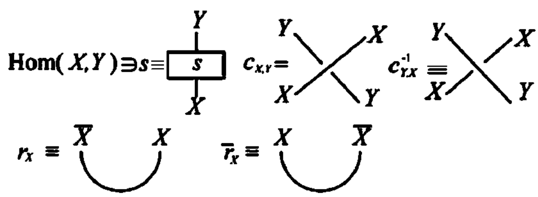
定义 320。如果$\mathcal{C}$，$\mathcal{D}$是严格辫子(对称)张量范畴，一个张量函子 F：$\mathcal{C}\to\mathcal{D}$ 是辫子(对称的)，如果：

$$F(c_{X,Y})=c_{F(X),F(Y)},\quad \forall X,\ Y\in \mathrm{Obj}\ \mathcal{C}$$

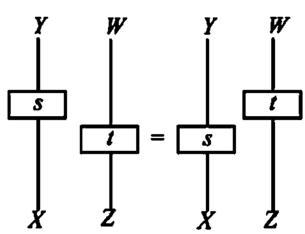
注意在左边和对应的右边，c 是$\mathcal{C}$和对应的$\mathcal{D}$的辫子。这里不存在幺半自然变换成为辫子/对称的附加条件。

A3　张量范畴的图像表示法

为了严格(辫子)张量范畴的计算，我们在有些情况下会用所谓的“纠缠图”，希望读者从前已经见过。要解释的话，更多详情参看[Kassel，1995]，我们只能说恒等态射(及相应对象)用垂线表示，一个态射 s：$X\to Y$ 的表示是一个具有分别从下面和上面进入的相应于 X 和 Y 线的方框，其组成和态射的张量积分别由垂直和水平并列表示。辫子态射用十字交叉表示，而对偶态射 r，$\bar{r}$ 用弧线表示：



如果 c 是对称的，辫子里的两条线不能画断。使用这个图线表示的原因是，只要牵涉具有“不同种类的输入和输出”，比如像 s：$A\to B\otimes C\otimes D$，那么张量范畴里即便相对简单的公式都会变得十分难懂，这种情况辫子和对偶态射都涉及时更糟糕。下面所画的是一个完备公式的例子。关于 s：$X\to Y$，t：$Z\to W$ 的交换律 $s\otimes\mathrm{id}_W\circ\mathrm{id}_X\otimes t=\mathrm{id}_Y\otimes t\circ s\otimes\mathrm{id}_Z$ 被画为：



这个图(正确地!)指出我们可以把态射彼此拽动。

A4 加法、$\mathbb{C}$ 线性和 * 范畴

定义 321。如果所有的幺半集合是阿贝尔群并且复合 $\circ$ 是双加的，那么一个范畴是一个 Ab 范畴。

定义 322。设 X，Y，Z 是 Ab 范畴里的对象，那么 Z 是 X 和 Y 的直和，表示为 $Z\cong X\oplus Y$，如果存在态射 u：$X\to Z$，u'：$Z\to X$，v：$Y\to Z$，v'：$Z\to Y$，使得 $u'\circ u=\mathrm{id}_X$，$v'\circ v=\mathrm{id}_Y$ 以及 $u\circ u'+v\circ v'=\mathrm{id}_Z$。(注意任何 $Z'\cong Z$ 也是 X 和 Y 的直和，因此直和只是取决于态射而定义，这就是为什么我们不写成 $Z=X\oplus Y$ 的原因。)只要对两个对象 X 和 Y 都存在一个直和 $Z\cong X\oplus Y$，我们就说$\mathcal{C}$具有直和。

定义 323。如果对于任何 $X\in\mathcal{C}$，集合 $\mathrm{Hom}(X,\ \mathbf{0})$ 和 $\mathrm{Hom}(\mathbf{0},\ X)$ 两者都明确包含一个元素，那么一个范畴$\mathcal{C}$里的对象 $\mathbf{0}$ 被称为一个零对象。一个到或者来自于一个零对象的态射被称为零态射。

它直接源自于任何两个零对象是同构的定义，如果一个范畴并不具有零对象，那么可以直接添加一个。如果 z 是零态射的并且 f 是任何态射，那么$z\circ f$，$f\circ z$，$z\otimes f$，$f\otimes z$ 都是零态射(只要它们有意义)。

定义 324。一个加法范畴是一个具有零对象和直和的 Ab 范畴。

例子 325。阿贝尔群的范畴(具有作为零的平凡群)。

定义 326。如果对于所有 X，$Y\in\mathrm{Obj}\ \mathcal{C}$，$\mathrm{Hom}(X,\ Y)$ 是一个$\mathbb{C}$ 矢量空间，并且复合映射$\circ$：$(f,\ g)\mapsto g\circ f$ 是双线性的，那么范畴$\mathcal{C}$被称为$\mathbb{C}$ 线性。如果$\mathcal{C}$是个张量范畴，我们要求$\otimes$：$(f,\ g)\mapsto g\otimes f$ 也是双线性的。$\mathbb{C}$ 线性范畴之间的函子总是假定为$\mathbb{C}$ 线性的，即 $\mathrm{Hom}_{\mathcal{C}}(X,\ Y)\to\mathrm{Hom}_{\mathcal{D}}(F(X),\ F(Y))$一定是$\mathbb{C}$ 线性的。

定义 327。一个在$\mathbb{C}$ 线性范畴上的正 * 运算是映射族，即把任何态射 $s\in\mathrm{Hom}(X,\ Y)$跟态射 $s^*\in\mathrm{Hom}(Y,\ X)$结合的映射。这个映射在 $s^*\circ s=0$ 意味着 $s=0$ 的意义上，一定是反线性的、对合的($(s^*)^*=s$)以及正的。一个 * 范畴是一个具有正 * 运算的$\mathbb{C}$ 线性范畴。一个张量 * 范畴，是一个具有正 * 运算，使得对所有 s，t 有$(s\otimes t)^*=s^*\otimes t^*$ 的张量范畴。我们只考虑张量 * 范畴的幺正辫子(对称性)。

定义 328。如果 $v^* \circ v = \mathrm{id}_X$，那么一个在 * 范畴里的态射 v：$X \to Y$ 被称为等距。如果它满足 $v \circ v^* = \mathrm{id}_Y$，等距 v 被称为幺正的。如果 $p = p \circ p = p^*$，态射 $p \in \mathrm{End}\ X$ 被称为一个投影。如果对于任何投影 $p \in \mathrm{End}\ X$ 在一个等距 v：$Y \to X$，使得 $v \circ v^* = p$，我们说 $\mathcal{C}$ 具有子对象。在一个 * 范畴里，我们加强定义 322 通过要求 $u' = u^*$，$v' = v^*$，即 u，v 必须是等距的。

定义 329。如果对于任何态射 s 有 $F(s^*) = F(s)^*$，* 范畴之间的函子 F 是一个 * 保持。结合张量 * 范畴的函子，又被张量 * 范畴之间的函子结合起来，它们的同构 $d_{X,Y}$，e 被要求是幺正的。

定义 330。设 $\mathcal{C}$ 是一个张量 * 范畴，并且 $X \in \mathrm{Obj}\ \mathcal{C}$，一个对象 $\overline{X} \in \mathrm{Obj}\ \mathcal{C}$ 被称为是 X 的共轭对象，如果存在同构 r：$\mathbf{1} \to \overline{X} \otimes X$，以及 r：$\mathbf{1} \to X \otimes \overline{X}$ 满足"共轭方程"：

$$\mathrm{id}_X \otimes r^* \circ \bar{r} \otimes \mathrm{id}_X = \mathrm{id}_X$$

$$\mathrm{id}_{\overline{X}} \otimes \bar{r}^* \circ r \otimes \mathrm{id}_{\overline{X}} = \mathrm{id}_{\overline{X}}$$

我们就说 $(\overline{X}, r, \bar{r})$ 是 X 的共轭。如果 $\mathcal{C}$ 的任何非零对象具有共轭，那么我们说 $\mathcal{C}$ 具有共轭。

注意一个零对象不可能有共轭。如果 $(\overline{X}, r, \bar{r})$，$(\overline{X}', r', \bar{r}')$ 两者都是 X 的共轭，那么很容易证明 $\mathrm{id}_{\overline{X}'} \otimes \bar{r}^* \circ r' \otimes \mathrm{id}_{\overline{X}}$：$\overline{X} \to \overline{X}'$ 是幺正的，因此共轭对于幺正等价是唯一的。

定义 331。如果 $\mathrm{End}\ X = \mathbb{C}\ \mathrm{id}_X$，一个在 $\mathbb{C}$ 线性范畴里的对象 X 是不可约化的。

定义 332。一个 TC^* 是具有有限维 Hom 集合、共轭、直和、子空间以及不可约化单位 **1** 的张量 * 范畴。一个 BTC^* 是一个具有幺正辫子的 TC^*，一个 STC^* 是一个具有幺正对称的 TC^*。

例子 333。有限维希尔伯特空间 $\mathcal{H}$ 的 * 范畴是一个 STC^*，对称 $c_{H,H'}$：$H \otimes H' \to H' \otimes H$ 是由翻转同构 $\sum$：$x \otimes y \mapsto y \otimes x$ 给定的，一个对象 H 的共轭是希尔伯特空间对偶 $\overline{H}$，挑选出一个具有对偶基 $\{f_i\}$ 的 H 一个基 $\{e_i\}$，那么共轭态射可通过下面得到：

$$r = \sum_i f_i \otimes e_i,\quad \bar{r} = \sum_i e_i \otimes f_i$$

按照同样的方法，人们看到紧致群 G 的有限维幺正表象的范畴 $\mathrm{Rep}_f G$ 是一个 STC^*。

引理 334。一个 TC^* 是半单的，即每一个对象是不可约化对象的有限直和。

证明。对于任何 $X \in \mathcal{C}$，End X 是一个具有正对合的有限维 $\mathbb{C}$ 代数，这样一个代数是半单的，即多矩阵代数。因此 id_X 是一个投影 p_i 的直和，这种投影在 $p_i \mathrm{End}\, X p_i \cong \mathbb{C}$ 的意义上是最小的。由于 $\mathcal{C}$ 具有子对象，存在对应于 p_i 对象 X_i，因为 p_i 的最小性是不可约化的，显然，$X \cong \bigoplus_i X_i$。

定义 335。一个共轭方程的解 $(X, r, \bar{r})$ 叫作标准的，如果对于所有的 $s \in$ End X 都有：

$$r^* \circ \mathrm{id}_{\overline{X}} \otimes s \circ r = \bar{r}^* \circ s \otimes \mathrm{id}_{\overline{X}} \circ \bar{r}$$

在此情况下，$(\overline{X}, r, \bar{r})$ 被叫作一个标准共轭。

引理 336。设 $\mathcal{C}$ 是一个 TC^*，以及 $(\overline{X}, r, \bar{r})$ 是一个 $X \in \mathcal{C}$ 的共轭，设 v_i：$X_i \to X$，w_i：$\overline{X}_i \to \overline{X}$ 是把 X，$\overline{X}$ 的直和分解作用成不可约化的两个等距，那么 $(\overline{X}, r, \bar{r})$ 是一个标准共轭，当且仅当 $(\overline{X}_i, w_i^* \otimes v_i^* \circ r, v_i^* \otimes w_i^* \circ \bar{r})$ 是 X_i 对所有 i 的标准共轭。任何对象承认一个标准共轭。

证明。对于等价性断言，参看[Longo and Roberts, 1997]，特别是引理 3.9。注意在[Longo and Roberts, 1997]里，标准为上述说法里的性质所定义。我们采用这一点来证明任何对象都承认一个标准共轭。如果 X 是不可约化的，我们就有 End $X = \mathbb{C}\, \mathrm{id}_X$。因此标准条件约化到 $r^* \circ r = \bar{r}^* \circ \bar{r}$，故共轭 $(\overline{X}, r, \bar{r})$ 能够通过重新标度 r，$\bar{r}$ 作为标准。在一般情况下，我们使用半单性来寻找把 X 变成不可约化的 X_i 的直和分解。设 $(\overline{X}_i, r_i, \bar{r}_i)$ 是 X_i 的标准共轭，并且 $\overline{X} = \bigoplus \overline{X}_i$。设 v_i：$X_i \to X$，ω_i：$\overline{X}_i \to \overline{X}$ 是作用在直和上的等距。定义 $r = \sum_i w_i \otimes v_i \circ r_i$ 和 $\bar{r} = \sum_i v_i \otimes w_i \circ \bar{r}_i$，在引理第一部分里的规则适用并提供了 $(\overline{X}, r, \bar{r})$ 的标准。

引理 337。设 $(\overline{X}, r, \bar{r})$ 是 X 的一个(标准)共轭，设 $p \in$ End X 是个投影，并且定义 $\bar{p} = r^* \otimes \mathrm{id}_{\overline{X}} \circ \mathrm{id}_{\overline{X}} \otimes p \otimes \mathrm{id}_{\overline{X}} \circ \mathrm{id}_{\overline{X}} \otimes \bar{r} \in \mathrm{End}\, \overline{X}$。如果 v：$Y \to X$，w：$\overline{Y} \to \overline{X}$ 是等距，使得 $v \circ v^* = p$，$w \circ w^* = \bar{p}$，那么 $(\overline{Y}, w^* \otimes v^* \circ r, v^* \otimes w^* \circ \bar{r})$ 就是一个 Y 的(标准)共轭。

证明。略。简单证明参看[Longo and Roberts，1997]或者[Müger，2000]。

引理 338。 如果$(\overline{X}, r, \overline{r})$，$(\overline{Y}, r', \overline{r}')$是$X$，$Y$相应的(标准)共轭，那么$(\overline{Y}\otimes\overline{X}, r'', \overline{r}'')$是一个$X\otimes Y$的(标准)共轭(其中$r''=\mathrm{id}_{\overline{Y}}\otimes r\otimes \mathrm{id}_Y\circ r'$和$\overline{r}''=\mathrm{id}_X\otimes\overline{r}'\otimes\mathrm{id}_{\overline{X}}\circ\overline{r}$)。

证明。$(\overline{Y}\otimes\overline{X}, r'', \overline{r}'')$是个共轭，这一点只是一个简单的计算，标准不够明显，原因在于映射 $\mathrm{End}\,X\otimes\mathrm{End}Y\to\mathrm{End}X\otimes Y$ 没必要是满射。当然，它使用了引理 336 给出的标准的其他特征。

命题 339。 设$\mathcal{C}$是一个TC^*，设$X\in\mathcal{C}$并且设$(\overline{X}, r, \overline{r})$是一个标准共轭，那么映射

$$\mathrm{Tr}_X:\ \mathrm{End}X\to\mathbb{C}\ ,\ s\mapsto r^*\circ\mathrm{id}_{\overline{X}}\otimes s\circ r$$

是明确定义了的，即$(\overline{X}, r, \overline{r})$的选择具有独立性，它被称为迹，满足：

$$\mathrm{Tr}_X(s\circ t)=\mathrm{Tr}_Y(t\circ s)\ \forall s:\ Y\to X,\ t:\ X\to Y,$$

$$\mathrm{Tr}_{X\otimes Y}(s\otimes t)=\mathrm{Tr}_X(s)\mathrm{Tr}_Y(t)\quad \forall s\in\mathrm{End}\,X,\ t\in\mathrm{End}\,Y$$

证明。作为读者的简单练习。

定义 340。 设$\mathcal{C}$是一个TC^*并且$X\in\mathcal{C}$，X的维度定义为$d(X)=\mathrm{Tr}_X(\mathrm{id}_X)$，即对于任何标准共轭$(\overline{X}, r, \overline{r})$有$d(X)=r^*\circ r$。

引理 341。 维度满足加法律($d(X\oplus Y)=d(X)+d(Y)$)和乘法律($d(X\otimes Y)=d(X)d(Y)$)。而且对任何对象有$d(\overline{X})=d(X)\geqslant 1$，其中$d(X)=1$意味着$X\otimes\overline{X}\cong 1$，即$X$是可逆的。

证明。可加性直接由标准共轭的讨论得到，维度的可乘性来自引理 338。

如果$(\overline{X}, r, \overline{r})$是$X$的标准共轭，那么$(\overline{X}, \overline{r}, r)$是$\overline{X}$的标准共轭，意味着$d(\overline{X})=d(X)$。$*$运算的正定性意味着$d(X)=r^*\circ r>0$。由于$X\otimes\overline{X}$包含$\mathbf{1}$作为直和项，我们有$d(X)^2\geqslant 1$，故$d(X)\geqslant 1$。最后，如果$d(X)=1$，那么$\mathbf{1}$是$X\otimes\overline{X}$的唯一直和项，即$X\otimes\overline{X}\cong 1$。类似地，$\overline{X}\otimes X\cong 1$。

定义 342。 设$\mathcal{C}$是一个BTC^*并且$X\in\mathcal{C}$，扭$\Theta(X)\in\mathrm{End}X$被定义为：

$$\Theta(X)=r^*\otimes\mathrm{id}_{\overline{X}}\circ\mathrm{id}_{\overline{X}}\otimes c_{X,X}\circ r\otimes\mathrm{id}_X$$

其中$(\overline{X}, r, \overline{r})$是共轭方程的标准解。

引理 343。 设$\mathcal{C}$是一个BTC^*，那么：

(i) $\Theta(X)$是明确定义的，即完全不依赖于$(\overline{X}, r, \overline{r})$的选择；

(ii) 对于任何态射 $s: X \to Y$，我们有 $\Theta(Y) \circ s = s \circ \Theta(X)$（即 Θ 是$\mathcal{C}$的单位函子的自然变换）；

(iii) $\Theta(X)$是幺正的；

(iv) 对所有的 X 和 Y 都有：

$$\Theta(X \otimes Y) = \Theta(X) \otimes \Theta(Y) \circ c_{Y,X} \circ c_{X,Y}$$

(v) 如果$\mathcal{C}$是一个 STC^*，这可以简化成 $\Theta(X)^2 = \mathrm{id}_X$，并且对所有 $X, Y \in \mathcal{C}$ 有 $\Theta(X \otimes Y) = \Theta(X) \otimes \Theta(Y)$（即 Θ 是$\mathcal{C}$的单位函子的一个幺半自然变换）。如果 X，Y 是不可约化的，我们具有 $\omega(X) = \pm 1$，并且对于所有不可约化的直和项 $Z < X \otimes Y$ 都有 $\omega_Z = \omega_X \omega_Y$。

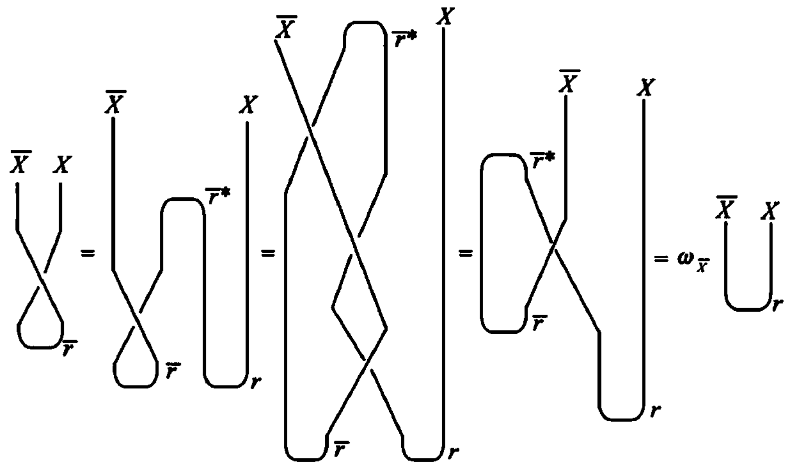
证明。(i)为命题 339 所证。其他证明都是不太难的计算，我们可以参看[Longo and Roberts，1997]或者[Müger，2000]。我们来看看(v)：在一个 STC^* 里，我们具有 $c^*_{X,X} = c^{-1}_{X,X} = c_{X,X}$，意味着 $\Theta(X)^* = \Theta(X)$，结合幺正性可以导出 $\Theta(X)^2 = \mathrm{id}_X$。$\Theta$ 在 STC^* 里的可乘性，源自于 $c_{Y,X} \circ c_{X,Y} = \mathrm{id}$。如果 X，Y 是不可约化的，我们有 $\Theta(X) = \omega_X \mathrm{id}_X$，$\Theta(Y) = \omega_Y \mathrm{id}_Y$，由此而有 $\Theta(X \otimes Y) = \omega_X \omega_Y \mathrm{id}_{X \otimes Y}$。由于对于不可约化的 $Z < X \otimes Y$ 有 $\omega(Z) = \omega_X \omega_Y$，这可以由 Θ 的自然性得到。

下面主要重复[Longo and Roberts，1997]里命题4.4 和4.5 的工作。

命题 344。 设$\mathcal{C}$，$\mathcal{D}$都是 BTC^*，并且 $E: \mathcal{C} \to \mathcal{D}$是个 * 保持的辫子张量函子。如果$(\overline{X}, r, \overline{r})$是 $X \in \mathcal{C}$标准共轭，那么$(E(\overline{X}), (d^E_{\overline{X},X})^{-1} \circ E(r) \circ e^E, d^E_{X,\overline{X}} \circ E(\overline{r}) \circ e^E)$是 $E(X)$的标准共轭，特别是：

$$d(E(X)) = d(X), \ \Theta(E(X)) = E(\Theta(X)) \quad \forall X \in \mathcal{C}$$

证明。我们暂且假设 E 是严格的，而 X 是不可约化的。设$(\overline{X}, r, \overline{r})$是标准共轭，由于 E 保留了共轭方程，$(E(\overline{X}), E(r), E(\overline{r}))$对 $E(X)$是共轭的，但是如果 E 不是满的，那么是否标准就需要证明。我们从下图开始：



因此有 $c^*_{\overline{X},X}\circ\overline{r}=\omega_{\overline{X}}\cdot r$，这等价于 $c_{\overline{X},X}\circ r=\overline{\omega}_{\overline{X}}\overline{r}$。现在我们设 $s\in\mathrm{End}E(X)$，并且计算：

$$
\begin{aligned}
&E(r^*)\circ \mathrm{id}_{E(\overline{X})}\otimes s\circ E(r)\\
&=E(r^*)\circ c^*_{E(\overline{X}),E(X)}\circ c_{E(\overline{X}),E(X)}\circ \mathrm{id}_{E(\overline{X})}\otimes s\circ E(r)\\
&=(c_{E(\overline{X}),E(X)}\circ E(r))^*\circ c_{E(\overline{X}),E(X)}\circ \mathrm{id}_{E(\overline{X})}\otimes s\circ E(r)\\
&=(c_{E(\overline{X}),E(X)}\circ E(r))^*\circ s\otimes\mathrm{id}_{E(\overline{X})}\circ c_{E(\overline{X}),E(X)}\circ E(r)\\
&=E(c_{\overline{X},X}\circ r)^*\circ s\otimes\mathrm{id}_{E(\overline{X})}\circ E(c_{\overline{X},X}\circ r)\\
&=E(\overline{\omega}_{\overline{X}}\overline{r})^*\circ s\otimes\mathrm{id}_{E(\overline{X})}\circ E(\overline{\omega}_{\overline{X}}\overline{r})\\
&=E(\overline{r})^*\circ s\otimes\mathrm{id}_{E(\overline{X})}\circ E(\overline{r})
\end{aligned}
$$

这意味着$(E(\overline{X}),E(r),E(\overline{r}))$是$E(X)$的一个标准共轭。(我们应用了辫子的幺正性，$E$ 是 * 保持的并且是辫子的这个事实，$c_{\overline{X},X}\circ r=\overline{\omega}_{\overline{X}}\overline{r}$ 和 $|\omega_{\overline{X}}|=1$。)

现在设 X 是可约化的，$(\overline{X},r,\overline{r})$是标准共轭，并且设 $v_i: X_i\to X$，$w_i: \overline{X}_i\to\overline{X}$，在作用是把成分分解成不可约化量时，都是等距的。定义 $r_i=w_i^*\otimes v_i^*\circ r$ 和 $\overline{r}_i=v_i^*\otimes w_i^*\circ\overline{r}$，那么$(\overline{X},r_i,\overline{r}_i)$根据引理 336 就是标准的。因此$(E(\overline{X}_i),E(r_i),E(\overline{r}_i))$根据该证明的第一部分是标准的。回顾 $E(r)=E(\sum_i w_i\otimes v_i\circ r_i)=\sum_i E(w_i)\otimes E(v_i)\circ E(r_i)$ 及其对 $E(\overline{r})$ 类似的情况，因此有$(E(\overline{X}),E(r),E(\overline{r}))$是标准的，因为它是标准共轭的直和。

如果 E 并不严格，我们不得不嵌入幺正 $d^E_{X,Y}$：$E(X)\otimes E(Y)\to E(X\otimes Y)$，$e^E$：$\mathbf{1}\to E(\mathbf{1})$在上述计算中是明显的，但是无济于事。$E$ 保持维度不变，原因在于维度由标准共轭的方式确定。最后$(E(\overline{X}),\ E(r),\ E(\overline{r}))$的标准结合 $E(c_{X,Y})=c_{E(X),E(Y)}$推出 $\Theta(E(X))=E(\Theta(X))$。

我们通过评论 * 范畴跟[Doplicher and Roberts, 1989; Longo and Roberts, 1997] 中 C^*张量范畴的更一般概念之间的关系来结束这一节。

定义 345。一个 C^* 范畴是一个具有正 * 运算$\mathbb{C}$ 线性范畴，其中 Hom(X, Y)是所有 X，Y 的一个巴拿赫空间，并且对于所有 s：$X\to Y$，t：$Y\to Z$ 都有 $\|s\circ t\|_{\mathrm{Hom}(X,Z)}\leqslant\|s\|_{\mathrm{Hom}(X,Y)}\cdot\|t\|_{\mathrm{Hom}(Y,Z)}$，以及对于 s：$X\to Y$ 都有 $\|s^*\circ s\|_{\mathrm{End}\,X}=\|s\|^2_{\mathrm{Hom}(X,Y)}$。(因此每一个 End X 是 C^* 代数。)一个 C^* 张量范畴是一个 C^* 范畴，而一个张量范畴使得对所有 s，t 都有 $\|s\otimes t\|\leqslant\|s\|\cdot\|t\|$。

命题 346。[Longo and Roberts, 1997]设$\mathcal{C}$是一个具有直和和不可约化单位的C^*张量范畴，只要 X，$Y\in\mathcal{C}$容许共轭，那么就有 Hom$(X, Y)<\infty$。因此一个具有直和、子对象、共轭以及不可约化单位的 C^* 张量范畴就是一个 TC^*。反过来，给定一个 TC^*，就存在唯一的一个范数，在以$\mathcal{C}$为背景形成 C^*张量范畴的 Hom(X, Y)空间上。

证明。假设 $X\in\mathcal{C}$具有共轭$(X,\ r,\ \overline{r})$，那么映射 End $X\to$Hom$(\mathbf{1},\ \overline{X}\otimes X)$，$s\mapsto \mathrm{id}_{\overline{X}}\otimes s\circ r$ 是矢量空间的一个同构，因为 $t\mapsto\overline{r}^*\otimes\mathrm{id}_X\circ\mathrm{id}_X\otimes t$ 是它的逆，就像使用共轭范畴所证明的那样。由于 Hom$(\mathbf{1},\ \overline{X}\otimes X)$是关于内积$\langle a,\ b\rangle\mathrm{id}_1=a^*\circ b$ 的前希尔伯特空间，并且它是完备的，因为$\mathcal{C}$是 C^* 张量范畴。选择 Hom$(\mathbf{1},\ \overline{X}\otimes X)$里的一个正交基$(e_i)_{i\in I}$，那么每一个 e_i：$\mathbf{1}\to\overline{X}\otimes X$ 是一个等距，并且 $e_i^*\circ e_j=0$ 其中 $i\neq j$，意味着$\overline{X}\otimes X$ 包含作为直和项的 1 的$\#I$ 复制。由于 X 具有一个共轭，同样 $\overline{X}\otimes X$ 也具有，但是如果$\#I$ 是无限的就不可能了。因此，Hom$(\mathbf{1},\ \overline{X}\otimes X)$和 End X 是有限维的。

假定任意 X，Y 具有共轭，选择一个拥有 u：$X\to Z$，v：$Y\to Z$ 的直和 $Z\cong X\otimes Y$，那么 Z 也具有共轭，参看引理 336，因此 dim End $Z<\infty$。由于，由 $s\mapsto v\circ s\circ u^*$ 给定的映射 Hom$(X, Y)\to$End Z 是单射的，因为它具有可逆的 $t\mapsto v^*\circ t\circ u$。这意味着 dim Hom$(X, Y)<\infty$。

我们略去说明 $TC^* \Rightarrow C^*$ 张量范畴的证明，由于它在后面用不着，它可以在[Müger，2000]找到。

这个结果证明了附录 B 里所作的假设等价于[Doplicher and Roberts，1989]里面的，它是按照 C^* 张量范畴形式化的。

A5　阿贝尔范畴

在附录 B 的第二部分，那是一个纯粹代数性质的，我们需要某些源自于阿贝尔范畴理论的基本事实，好的参考书，比如有[Gabriel(加布里埃)，1962]和[Mac Lane，1998，Chapter Ⅷ]。

定义 347。一个态射 s：$X\to Y$ 被称为首一的(monic)，如果 $s\circ t_1 = s\circ t_2$ 隐含着 $t_1 = t_2$，只要 t_1，t_2 是具有目标 X 和相同源的态射。一个态射 s：$X\to Y$ 被称为 epi，如果 $t_1\circ s = t_2\circ s$ 隐含着 $t_1 = t_2$，只要 t_1，t_2 是具有源 X 和相同目标的态射。

定义 348。设$\mathcal{C}$是一个加法范畴，给定一个态射 f：$X\to Y$，态射 k：$Z\to X$ 是 f 的一个核，如果 $f\circ k = 0$，并且给定任意态射 k'：$Z'\to X$，使得 $f\circ k' = 0$，则存在一个唯一的态射 l：$Z'\to Z$ 使得 $k' = k\circ l$。

f：$X\to Y$ 的上核是一个态射 c：$Y\to Z$，如果 $c\circ f = 0$ 并且给定任意态射 c'：$Y\to Z'$使得 $c'\circ f = 0$，存在唯一 d：$Z\to Z'$使得 $c' = d\circ c$。

这个定义的一个简单结果是，任何核都是首一的，而任何上核都是 epic 的。

定义 349。一个加法范畴$\mathcal{C}$是阿贝尔的，如果：

1. 任何态射具有一个核和一个上核；
2. 任何首一的态射是某些态射的核；
3. 任何首一的态射是某些态射的上核。

命题 350。设$\mathcal{C}$是一个阿贝尔范畴，那么：

(i) 任何首一的是其上核的核，而任何 epi 是其核的上核；

(ii) 一个态射是一个同构，当且仅当它是首一的和 epic；

(iii) 在阿贝尔范畴里的任何态射 f：$X\to Y$ 承认一个因子化 $f = m\circ e$，其中 e：$X\to Z$ 是 epi，而 m：$Z\to Y$ 是首一的。给定另一个 epi e'：$X\to Z'$和首一的 m'：

$Z' \to Y$，使得 $f = m' \circ e'$。存在一个同构 u：$Z \to Z'$，使得 $e' = u \circ e$ 和 $m = m' \circ u$。

证明。详细证明参看[Mac Lane，1998，Chapter Ⅷ]。第(ii)条里的“仅当”是常规的。关于(iii)：定义 $m = \ker(\operatorname{coker}(f))$，$m$ 是首一的。就 $(\operatorname{coker} f) \circ f = 0$ 的观点来看，f 因子当对唯一 e 的 $f = m \circ e$。下一步人们证明 e 是 epi，并且 $e = \operatorname{coker}(\ker(f))$。

定义 351。在阿贝尔范畴里态射 f：$X \to Y$ 的像，是在 f 的首一的 epic 因子化 $X \xrightarrow{e} Z \xrightarrow{m} Y$ 里的首一的 m：$Z \to Y$(对同构唯一)。

在一个具体的阿贝尔范畴，对象 Z 是同构于 f 的普通像，是 Y 的子集，是专业术语。

定义 352。如果对给定的任意满态射 p：$A \to B$ 以及任意态射 b：$P \to B$，存在一个态射 a：$P \to A$ 使得 $b = p \circ a$，则在一个阿贝尔范畴里的一个对象 P 是投影的。

引理 353。任意 TC^*，具有零对象的$\mathcal{C}$是阿贝尔的。

证明。显然$\mathcal{C}$是可加的，定义 349 的其他要求来自于半单性的一点作用，参看引理 334。

A6 在阿贝尔对称张量范畴里的交换代数

熟知的交换环代数的绝大部分，它们的理想和模(处于阿贝尔群的范畴 Ab 里)能够推广到其他阿贝尔对称甚或辫子张量范畴。我们做的工作只是陈述后面需要的那些事实，其中有些可能是新的。

定义 354。设$\mathcal{D}$是一个严格张量范畴，那么$\mathcal{D}$中的一个幺半群是一个三元组 (Q, m, η)，其中 $Q \in \mathcal{D}$，m：$Q \otimes Q \to Q$ 和 η：$\mathbf{1} \to Q$ 都是态射，满足：

$$m \circ (m \otimes \mathrm{id}_Q) = m \circ (\mathrm{id}_Q \otimes m),\ m \circ \eta \otimes \mathrm{id}_Q = \mathrm{id}_Q = m \circ \mathrm{id}_Q \otimes \eta$$

如果$\mathcal{D}$是辫子的，那么幺半群被认为是对易的，只要 $m \circ c_{Q,Q} = m$。

定义 355。设 (Q, m, η) 是一个在严格张量范畴$\mathcal{D}$里的幺半群，那么一个 Q 模(在$\mathcal{D}$里)是一个 (M, μ) 对，其中 $M \in \mathcal{D}$和 μ：$Q \otimes M \to M$ 满足：

$$\mu \circ \mathrm{id}_Q \otimes \mu = \mu \circ m \otimes \mathrm{id}_M,\ \mu \circ \eta \otimes \mathrm{id}_M = \mathrm{id}_M$$

一个 Q 模的态射 s：$(M, \mu) \to (R, \rho)$，是一个满足 $s \circ \mu = \rho \circ \mathrm{id}_Q \otimes s$ 的态射 $s \in \mathrm{Hom}_{\mathcal{D}}(M, R)$。在$\mathcal{D}$里的 Q 模及其态射形成一个范畴 $Q-\mathrm{Mod}_{\mathcal{D}}$，如果$\mathcal{D}$

是k线性的，那么$Q-\mathrm{Mod}_{\mathcal{D}}$也是线性$k$的。在范畴里的 Hom 集合$Q-\mathrm{Mod}$都用$\mathrm{Hom}_Q(\cdot, \cdot)$表示。

评论 356。

1. 前面的定义，是 Vect 里相应概念的明显推广，以直接的方式推广到非严格张量范畴。

2. 如果(M, μ)是Q模并且$X\in\mathcal{D}$，那么$(Q\otimes X, \mu\otimes\mathrm{id}_X)$是$Q$模的。

3. 如果$\mathcal{D}$具有直和，我们就能够定义两个Q模(M_1, μ_1)，(M_2, μ_2)直和(R, ρ)，具体而言，如果v_i：$M_i\to R$，$i=1$，2 是对应于$R\cong M_1\oplus M_2$等距的，那么$\rho=v_1\circ\mu_1\circ\mathrm{id}_Q\otimes v_1^*+v_2\circ\mu_2\circ\mathrm{id}_Q\otimes v_2^*$提供一个$Q$模结构。

4. 给定$\mathcal{D}$里的一个幺半群(Q, m, η)，我们具有一个明显的Q模(Q, m)，并且对于任意$n\in\mathbb{N}$，我们能够考虑$n\cdot(Q, m)$，Q模(Q, m)的n复制的直和。

定义 357。设$\mathcal{D}$是一个具有单位$\mathbf{1}$的严格张量范畴，并且设(Q, m, η)是一个$\mathcal{D}$里的幺半群，我们定义一个集范畴里的幺半群Γ_Q为$\Gamma_Q=\mathrm{Hom}(\mathbf{1}, Q)$，乘法通过$s\bullet t=m\circ t\otimes s$给定，并且单位为$\eta$。如果$\mathcal{D}$是辫子的，并且$(Q, m, \eta)$是交换的，那么$\Gamma_Q$是交换的。

引理 358。设$\mathcal{D}$是一个严格张量范畴，并且(Q, m, η)是$\mathcal{D}$里的一个幺半群，那么存在一个幺半群的同构γ：$\mathrm{End}_Q((Q, m))\to(\Gamma_Q, \bullet, \eta)$，其如下给出：

$$\gamma: \mathrm{End}_Q((Q, m))\to\mathrm{Hom}(\mathbf{1}, Q), \ u\mapsto u\circ\eta$$

$$\gamma^{-1}: \mathrm{Hom}(\mathbf{1}, Q)\to\mathrm{End}_Q((Q, m)), \ s\mapsto m\circ\mathrm{id}_Q\otimes s$$

如果$\mathcal{D}$(由此包括$Q-\mathrm{Mod}_{\mathcal{D}}$)是$k$线性的，那么$\gamma$是$k$代数的一个同构。如果$\mathcal{D}$是辫子的并且幺半群$(Q, m, \eta)$是交换的，那么幺半群($k$代数)$(\Gamma_Q, \bullet, \eta)$，从而包括$\mathrm{End}_Q((Q, m))$都是交换的。

证明。$(\Gamma_Q, \bullet, \eta)$是幺半群(分别结合$k$代数)，这一点直接得到，因为$(Q, m, \eta)$是幺半群，对于$s\in\mathrm{Hom}(\mathbf{1}, Q)$，我们靠幺半群公理有$\gamma(\gamma^{-1}(s))=m\circ\mathrm{id}_Q\otimes s\circ\eta=s$，另一方面，对于$u\in\mathrm{End}_Q((Q, m))$我们有：

$$\gamma^{-1}(\gamma(u)) = m\circ \mathrm{id}_Q\otimes(u\circ\eta) = m\circ \mathrm{id}_Q\otimes u\circ \mathrm{id}_Q\otimes\eta = u\circ m\circ \mathrm{id}_Q\otimes\eta = u$$

其中第三个等式是由于 s 是一个 Q 模映射(参看定义 355)，显然 $\gamma(\mathrm{id}_Q)=\eta$，而且：

$$\gamma^{-1}(s)\circ\gamma^{-1}(t) = (m\circ \mathrm{id}_Q\otimes s)\circ(m\circ \mathrm{id}_Q\otimes t) = m\circ m\otimes \mathrm{id}_Q\circ \mathrm{id}_Q\otimes t\otimes s$$
$$= m\circ \mathrm{id}_Q\otimes m\circ \mathrm{id}_Q\otimes t\otimes s = \gamma^{-1}(s\bullet t)$$

如果$\mathcal{D}$是辫子的，并且幺半群(Q, m, η)是对易的，那么：

$$s\bullet t = m\circ t\otimes s = m\circ c_{Q,Q}\circ s\otimes t = m\circ s\otimes t = t\cdot s$$

其中我们使用了辫子的自然性和幺半群的交换性。

评论 359。

1. 我们已经看到，在任意抽象张量范畴里幺半群(Q, m, η)产生具体的幺半群$(\Gamma_Q, \bullet, \eta)$，即处于范畴集合(Sets)里。后者具有笛卡尔积作为张量积，并且任意一个元素集合是具有单位 **1** 的张量。因此对于任意 $X\in$ Sets，Hom$(\mathbf{1}, X)$是处在对应于 X 元素的双射里。因此，如果$\mathcal{D}=$ Sets，那么幺半群(Q, m, η)和$(\Gamma_Q, \bullet, \eta)$是同构的。因为这个原因，我们称 Γ_Q 为 Q 的元素的幺半群，甚至当$\mathcal{D}$是一个抽象范畴。

2. 在辫子张量范畴$\mathcal{D}$里的交换幺半群(Q, m, η)的情况下，$\mathrm{End}_Q((Q, m))$的对易性具有十分自然的解释：如果$\mathcal{D}$具有余等子，这保持在任意阿贝尔范畴里，那么范畴 $Q-\mathrm{Mod}_{\mathcal{D}}$ 又重新是一个张量范畴，并且 Q 模(Q, m)是它的单位对象。在具有单位 **1** 的任意张量里，End **1** 是可交换的幺半群(交换 k 代数如果$\mathcal{D}$是 k 线性的)。这是为什么 $\mathrm{End}_Q((Q, m))$是可对易的真正原因。更多的是：如果$\mathcal{D}$是对称的，并且 Q 是阿贝尔的，那么张量范畴 $Q-\mathrm{Mod}_{\mathcal{D}}$ 重新是对称的。(在辫子的情况下这不必是正确的，但是 $Q-\mathrm{Mod}_{\mathcal{D}}$ 总具有一个辫子化的特定完全子范畴。)

现在我们特意看一下阿贝尔范畴。

命题 360。设(Q, m, η)是一个在阿贝尔严格张量范畴$\mathcal{D}$里的幺半群，那么范畴 $Q-\mathrm{Mod}_{\mathcal{D}}$ 是阿贝尔的。

证明。略。(这是一个关于阿贝尔范畴的好练习。)

定义 361。设$\mathcal{D}$是一个阿贝尔严格对称张量范畴。一个在对易幺半群(Q, m, η)里的理想，是一个在范畴 $Q-\mathrm{Mod}$ 里的首一的 j：$(J, \mu_J)\to(Q, m)$。

一个理想j：$(J, \mu_J)\to(Q, m)$被称为真正的，如果j不是一个同构(即不是epi)。如果j：$(J, \mu_J)\to(Q, m)$和j'：$(J', \mu_{J'})\to(Q, m)$都是理想，那么j：$(J, \mu_J)\to(Q, m)$被包含在j'：$(J', \mu_{J'})\to(Q, m)$，表示为$j<j'$，如果存在一个首一的$i\in \mathrm{Hom}_Q((J, \mu_J), (J', \mu_{J'}))$使得$j'\circ i=j$。一个在$(Q, m, \eta)$里的真正理想$j$：$(J, \mu_J)\to(Q, m)$被称为极大的，如果任何包含$j$：$(J, \mu_J)\to(Q, m)$的真正理想$j'$：$(J', \mu_{J'})\to(Q, m)$，是同构于$j$：$(J, \mu_J)\to(Q, m)$。

引理362。设$\mathcal{D}$是一个本质上小的阿贝尔严格对称张量范畴，(Q, m, η)是在$\mathcal{D}$里的交换幺半群，那么任何在(Q, m, η)里真正的理想j：$(J, \mu_J)\to(Q, m)$，被包含在一个极大理想$\tilde{j}$：$(\tilde{J}, \tilde{\mu})\to(Q, m)$里。

证明。在(Q, m, η)里的理想并不必然形成一个集合，但是同构类却形成一个集合，因为$\mathcal{D}$假定在本质上小。在(Q, m, η)里的理想上关系$<$产生理想的同构类集合的部分序，关于这个部分序的极大元素肯定是极大理想的同构类。现在我们把佐恩(Zorn)引理用来完成在交换代数里的证明。

就像在范畴R-Mod里一样，我们能够用一个理想来商除一个交换幺半群：

引理363。设$\mathcal{D}$是一个阿贝尔严格对称张量范畴，(Q, m, η)是交换幺半群，并且j：$(J, \mu_J)\to(Q, m)$是一个理想。设$p=\mathrm{coker}\, j$：$(Q, m)\to(B, \mu_B)$，那么存在唯一态射m_B：$B\otimes B\to B$和η_B：$\mathbf{1}\to B$，使得：

1. (B, m_B, η_B)是一个对易的幺半群；

2. $p\circ m=m_B\circ p\otimes p$；

3. $p\circ\eta=\eta_B$。

幺半群(B, m_B, η_B)被叫作(Q, m, η)除理想j：$(J, \mu_J)\to(Q, m)$得到的商，它是非平凡的(B不是一个零对象)，当且仅当该理想是符合规则的。

而且，由$s\mapsto p\circ s$给定的映射p_Γ：Γ：$\Gamma_Q\to\Gamma_B$，是一个交换代数的同态，如果单位$\mathbf{1}\in\mathcal{D}$是一个投射对象，这就是一个满射。

证明。m_B，η_B的建构本质上是作为交换代数进行的，尽管存在没有元素使其更加抽象的这个事实。由于p：$(Q, m)\to(B, \mu_B)$是j的上核，B是非零的，当且仅当j不是epi，即如果理想是符合规则的。方程$p\circ m=m_B\circ p\otimes p$和$p\otimes\eta=\eta_B$意味着$p_\Gamma$是单位的同态。如果$\mathbf{1}$是投射的，那么正是定义352意味着

对于任何 s: $\mathbf{1} \to B$ 存在 t: $\mathbf{1} \to Q$，使得 $s = p \circ t$，因此 p_{Γ} 是满射。

引理 364。 设$\mathcal{D}$是一个本质上小的阿贝尔严格对称张量范畴，假设(Q, m, η)是一个$\mathcal{D}$里的交换幺半群，并且j: $(J, \mu) \to (Q, m)$是一个理想，设(B, m_B, η_B)是商幺半群。那么在(B, m_B, η_B)里理想的等价类，跟包含j: $(J, \mu) \to (Q, m)$的(Q, μ, η)里的理想j': $(J', \mu') \to (Q, m)$的等价类之间的双射对应。

特别是，如果j是一个极大理想，那么在(B, m_B, η_B)里的所有理想要么是零，要么是同构于(B, m_B)。

证明。就像在通常对易代数里的一样。

引理 365。 设 k 是一个场，并且(Q, m, η)是一个在严格对称阿贝尔 k 线性范畴$\mathcal{D}$里的交换幺半群，如果在(Q, m, η)里任何非零理想是同构于(Q, m)，那么对易单位 k 代数 $\mathrm{End}_Q((Q, m))$是一个场。

证明。设 $s \in \mathrm{End}_Q((Q, m))$是非零的，那么 $\mathrm{im}\, s \neq 0$ 是在(Q, m)里的非零理想，因此必定同构于(Q, m)。因此 $\mathrm{im}\, s$ 进而 s 是 epi。由于 $s \neq 0$，核 $\ker s$ 不同构于(Q, m)，从而它一定是零，因此 s 是首一的。根据命题 350，s 是一个同构，因此对易代数 k 代数 $\mathrm{End}_Q((Q, m))$是一个扩展 k 的场。

下述引理是借助于[Bichon，1998]。

引理 366。 设$\mathcal{D}$是一个阿贝尔代数严格对称张量范畴，并且(Q, m, η)是一个在它里面的交换幺半群，那么任何 $\mathrm{End}_Q((Q, m))$里的满态射是一个同构。

证明。设 $g \in \mathrm{End}_Q((Q, m))$是一个满射，并且设$j$: $(J, \mu_J) \to (Q, m)$是一个在(Q, m, η)里的理想。由于 Q − Mod 是一个其单位为(Q, m)的张量范畴，因此存在一个同构 $s \in \mathrm{Hom}_Q((J, \mu_J), (Q \otimes_Q J, \mu_{Q \otimes_Q} J))$。设 $h \in \mathrm{End}_Q((J, \mu_J))$是合成：

$$(J, \mu_J) \xrightarrow{s} (Q \otimes_Q J, \mu_{Q\otimes_Q} J) \xrightarrow{g \otimes \mathrm{id}_J} (Q \otimes_Q J, \mu_{Q\otimes_Q} J) \xrightarrow{s^{-1}} (J, \mu_J)$$

由于 Q − Mod 的张量积$\otimes_Q$ 是右正合的，$g \otimes \mathrm{id}_J$ 是 epi。由于$j \circ h = g \circ j$，并且如果我们取$(j$: $(J, \mu_J) \to (Q, m)) = \ker g$，我们具有 $j \circ h = 0$，从而有 $j = 0$，因为 h 是 epi。因此，g 是首一的，从而 g 是一个同构。

A7　归纳极限 Ind 范畴

我们需要归纳极限概念的范畴方案。为了我们的目的，ℕ 上的归纳极限就会如此，但是为了用在目前理论上，我们需要一些定义。

定义 367。 如果$\mathcal{I}$，$\mathcal{C}$ 是范畴并且 F：$\mathcal{I}\to\mathcal{C}$ 是函子，那么 F 的上极限（或者归纳极限）由一个对象 $Z\in\mathcal{C}$，以及对于在$\mathcal{C}$任何 $X\in\mathcal{I}$里的态射 i_X：$F(X)\to Z$ 构成，使得：

1. 对于$\mathcal{I}$里的任何态射s：$X\to Y$都有 $i_Y\circ F(s)=i_X$。

2. 给定 $Z'\in\mathcal{C}$和一个$\mathcal{C}$里的态射 j_X：$F(X)\to Z'$族，使得对$\mathcal{I}$里的任何态射 s：$X\to Y$都有 $j_Y\circ F(s)=j_X$，存在一个唯一态射 ι：$Z\to Z'$，使得对所有 $X\in\mathcal{I}$都有 $j_X=\iota\circ i_X$。

上面要求的第二个性质是一个普遍属性，它意味着 F 的任意两个上极限是同构的，因此同构本质上是唯一的，只要它存在。

定义 368。 一个范畴$\mathcal{I}$是滤子化的，如果它不为空，并且：

1. 对任意两个对象 X，$Y\in\mathcal{I}$，存在一个 $Z\in\mathcal{Z}$和态射 i：$X\to Z$，j：$Y\to Z$；

2. 对于任何两个$\mathcal{I}$里的态射 u，v：$X\to Y$，存在一个态射 w：$Y\to Z$，使得 $w\circ u=w\circ v$。

注意任何定向的偏序集（I，$\leqslant$）是滤子化范畴，如果我们取对象为 I 的元素，并且箭头是有序对$\{(i,j):i\leqslant j\}$。

定义 369。 设$\mathcal{C}$是一个范畴，那么范畴 Ind $\mathcal{C}$被定义为函子范畴，其对象都是函子 F：$\mathcal{I}\to\mathcal{C}$，其中$\mathcal{I}$是一个小滤子化范畴。对于 F：$\mathcal{I}\to\mathcal{C}$，$F'$：$\mathcal{I}'\to\mathcal{C}$，Hom 集合被定义为：

$$\mathrm{Hom}_{\mathrm{Ind}\,\mathcal{C}}(F,F')=\varprojlim_{X}\varinjlim_{Y}\mathrm{Hom}_{\mathcal{C}}(F(X),F'(Y))$$

等式右边的元素组成一个族$(f_{X,Y}: F(X)\to F'(Y))_{X\in\mathcal{I},Y\in\mathcal{I}'}$满足$\mathcal{I}'$里任何 s：$Y\to Y'$的 $F'(s)\circ f_{X,Y}=f_{X,Y'}$，以及$\mathcal{I}$里任何 t：$X\to X'$的$f_{X',Y}\circ F(t)=f_{X,Y}$。我们把它留作练习得出态射的复合。

有些 Ind $\mathcal{C}$的性质很明显，它包括作为子范畴的$\mathcal{C}$：对于任何 $X\in\mathcal{C}$，我们赋予函子 F：$\mathcal{I}\to\mathcal{C}$，其中只具有一个对象 $*$，并且 $F(*)=X$。这个嵌入显然是完全的和忠实的。如果$\mathcal{C}$是一个 Ab 范畴/可加的/ℂ 线性，那么 Ind $\mathcal{C}$也如此。

如果$\mathcal{C}$是一个严格(对称)张量范畴，那么 Ind $\mathcal{C}$也如此。F：$\mathcal{I}\to\mathcal{C}$和 F：$\mathcal{I}'\to\mathcal{C}$的张量积被定义为$\mathcal{I}''=\mathcal{I}\times\mathcal{I}'$(这是滤子化范畴)，并且 $F\otimes F'$：$\mathcal{I}''\ni X\times Y\mapsto F(X)\otimes F'(Y)$。对于我们需要的其他结果，我们仅仅引述[SGA，1972]，在那我们也提到这个证明。

定理 370。Ind $\mathcal{C}$具有对全部小滤子化指数范畴$\mathcal{I}$的上极限，如果$\mathcal{C}$是一个阿贝尔范畴$\mathcal{C}$，那么 Ind $\mathcal{C}$是阿贝尔的。

因此任何阿贝尔(对称单项的)范畴是一个阿贝尔(对称单项的)范畴的完全子范畴，她在滤子化上极限之下是完备的。对于我们来说，这意味着在 Ind $\mathcal{C}$ 里我们能够取由$\mathbb{N}$标记的无限直和，定义$\bigoplus_{i\in\mathbb{N}}X_i$为函子 F：$\mathcal{I}\to\mathcal{C}$的上极限，其中$\mathcal{I}$是解释为滤子化范畴的偏序集$\mathbb{N}$，并且 $F(n)=\bigoplus_{i=1}^{n}X_i$ 跟显态射 $F(n)\to F(m)$在 $n\leqslant m$ 时结合在一起。

引理 371。如果$\mathcal{C}$是一个 TC^*，那么任何对象 $X\in\mathcal{C}$，作为 Ind $\mathcal{C}$的对象，都是投影的。

证明。首先假设 X 是不可约化的，并且考虑到 s：$X\to B$。给定一个在 Ind $\mathcal{C}$ 里的 epi p：$A\to B$，我们具有 $A=\lim_{\to}A_i$，其中有 $A_i\in\mathcal{C}$，并且对 B 有类似情况。而且，$\mathrm{Hom}(A,B)=\lim_{\leftarrow}\lim_{\to}\mathrm{Hom}_{\mathcal{C}}(A_i,B_j)$和 $\mathrm{Hom}(X,B)=\lim_{\to}\mathrm{Hom}_{\mathcal{C}}(X,B_j)$。由于 X 不可约化并且$\mathcal{C}$是半单的，只要 s_j：$X\to B_j$ 是非零的，则 X 是 B_j的直和。由于 p：$A\to B$ 是 epi，就 i 充分大来说，复合 $A_i\to B_j$ 是 epi。通过$\mathcal{C}$的半单性，那么 s_j 上升到态射 $X\to A_i$。结合所有条件给出一个态射 $\hat{s}$：$X\to A$，使得 $p\circ\hat{s}=s$。

现在假设 X 是一个 X_i 有限直和，具有等距 v_i：$X_i\to X$ 和 s：$X\to B$ 的不可约化。定义 $s_i=s\circ v_i$：$X_i\to B$，该证明的前半部分假设 $\hat{s}_i$：$X_i\to A$，使得$p\circ\hat{s}_i=s_i$，现在定义 $\hat{s}=\sum_i\hat{s}_i\circ v_i^*$：$X\to A$。我们有：

$$p\circ\hat{s}=\sum_i p\circ\hat{s}_i\circ v_i^*=\sum_i s_i\circ v_i^*=\sum_i s\circ v_i\circ v_i^*=s$$

这证明 X 的投影性质。

B 对称张量 * 范畴的抽象对偶性理论

前面我们给出在 AQFT 建构中所需要结果的自洽陈述。证明的某些部分推迟到附录的剩余部分，读者可以快速(或随意)浏览或者直接跳过。

B1 纤维函子和具体的田中定理，第一部分

设 $\mathrm{Vect}_{\mathbb{C}}$ 代表有限维$\mathbb{C}$ 矢量空间的$\mathbb{C}$ 线性对称张量范畴，而$\mathcal{H}$代表有限维希尔伯特空间的 STC^*。我们假定两个张量范畴都是严格的，这等于是删除结合性以及源自标记的单位同构 α，λ，ρ。两个范畴都有正则对称 $\sum$ 、翻转同构 $\sum_{V,V'}$：$V\otimes V'\to V'\otimes V$。

定义 372。设$\mathcal{C}$是一个 STC^*，一个$\mathcal{C}$的纤维函子，是一个忠实的$\mathbb{C}$ 线性张量函子 E：$\mathcal{C}\to\mathrm{Vect}_{\mathbb{C}}$。一个$\mathcal{C}$的 * 保持纤维函子，是一个张量 * 范畴的忠实函子 E：$\mathcal{C}\to\mathcal{H}$。$E$ 是对称的，如果 $E(c_{X,Y})=\sum_{E(X),E(Y)}$，即$\mathcal{C}$的对称被映射成 $\mathrm{Vect}_{\mathbb{C}}$ 或者$\mathcal{H}$各自的正则对称性。

一个装配了对称 * 保持纤维函子的 STC^*，被称为具体的，因为它等价于希尔伯特空间的范畴$\mathcal{H}$的一个(非完全!)张量子范畴。我们在此附录关注的主要是：(1)纤维函子存在的结果；(2)纤维函子的唯一性；(3)纤维函子的存在性。关于(2)我们将证明：

定理 373。设$\mathcal{C}$是一个 STC^*，并且设 E_1，E_2：$\mathcal{C}\to\mathcal{H}$是 * 保持对称纤维函子。那么 $E_1\cong E_2$，即存在一个幺正单项自然同构 α：$E_1\to E_2$。

现在我们假设一个对称 * 保持纤维函子，对 STC^* 来说$\mathcal{C}$是给定的。设 $G_E\subset\mathrm{Nat}_{\otimes}E$ 代表 E(对自己)的幺正单项自然包含集合，这显然是一个具有恒等自然变换为单位的群，G_E 能够等同于 $\Pi_{X\in\mathcal{C}}\mathcal{U}(E(X))$的子集，其中$\mathcal{U}(E(X))$是在有限维希尔伯特空间 $E(X)$上的幺正紧致群。这些群的积是靠吉洪诺夫(Tychonov)定理紧致的，比如参看[Pedersen(佩德森)，1989，Theorem 1.6.10]，并且由于 G_E 是封闭子集，它自己是紧致的。这个积和逆映射是连续的，因此 G_E 是紧致拓扑群。正是靠其定义，群 G_E 作用在希尔伯特空间 $E(X)$，$X\in\mathcal{C}$，

通过的是幺正表示π_X，即$\pi_X(g)=g_X$，其中g_X是自然变换$g\in G_E$在X的成分。

命题 374。存在一个忠实的对称张量$*$函子F：$\mathcal{C}\to \mathrm{Rep}_f G_E$，使得$K\circ F=E$，其中$K$：$\mathrm{Rep}_f G_E\to\mathcal{H}$是可遗函子$(H,\pi)\mapsto H$。

证明。我们定义对于所有的$X\in\mathcal{C}$有$F(X)=(E(X),\pi_X)\in \mathrm{Rep}_f G_E$，对于所有的$s\in \mathrm{Hom}(X,Y)$有$F(s)=E(s)$。对于$s$：$X\to Y$我们有：

$$F(s)\pi_X(g)=F(s)g_X=g_Y F(s)=\pi_Y(g)F(s)$$

由于g：$E\to E$是一个自然变换，因此F是个函子，它是明显$*$保持的并且忠实的。在对任何$g\in G_E$有$g_1=\mathrm{id}_{E(1)}$看来，我们具有$F(\mathbf{1}_{\mathcal{C}})=(\mathbb{C},\pi_0)=\mathbf{1}_{\mathrm{Rep}_f G_E}$，其中$\pi_0$是个平凡表示。为了理解$F$是个张量$*$范畴函子，我们必须生成分别满足式(56)和式(57)的$d^F_{X,Y}$：$F(X)\otimes F(Y)\to F(X\otimes Y)$，$X$，$Y\in\mathcal{C}$和$e$：$\mathbf{1}_{\mathrm{Rep}_f G_E}\to F(\mathbf{1}_{\mathcal{C}})$。我们说选择$e^F=e^E$，$d^F_{X,Y}=d^E_{X,Y}$完成了这个任务，其中$e^E$和$d^E_{X,Y}$是伴随张量函子$E$：$\mathcal{C}\to\mathcal{H}$的单位元。显然$e^E$和$d^E_{X,Y}$满足式(56)和式(57)，但是我们必须证明它们是$\mathrm{Rep}_f G_E$里的态射。对于$d^E_{X,Y}$这可以由下面式子推出：

$$d^F_{X,Y}\circ(\pi_X(g)\otimes\pi_Y(g))=d^E_{X,Y}\circ g_X\otimes g_Y=g_{X\otimes Y}\circ d^E_{X,Y}=\pi_{X\otimes Y}(g)\circ d^F_{X,Y}$$

其中我们已经知道g是单项自然变换。由于，通过自然单项变换的定义，我们对于所有$g\in G_E$有$g_1=\mathrm{id}_{E(1)}$，即$F(\mathbf{1})=(E(\mathbf{1}),\pi_1)$是平凡表示。如果严格的单位$\mathbf{1}_{\mathcal{H}}=\mathbb{C}$是在$E$的像中，那么，由于自然属性它也传递了平凡表象，因此e^F事实上是表示的一个态射。在$\mathbf{1}_{\mathcal{H}}\notin E(\mathcal{C})$情况下，我们人为地用平凡表象装备$\mathbf{1}_{\mathcal{H}}$。由于$\mathrm{Rep}_f G_E$的对称性，是靠由$c((H,\pi),(H',\pi'))=c(H,H')$所给式子定义的，其中式子右边指的是范畴$\mathcal{H}$，并且由于$E$遵守这样的对称性，因此$F$也如此。所以$K\circ F=E$是明显的。

下述命题的证明被推迟，因为它要求进一步的准备。

命题 375。设$\mathcal{C}$是一个STC^*，并且E：$\mathcal{C}\to\mathcal{H}$是一个对称$*$保持纤维函子。设$G_E$和$F$：$\mathcal{C}\to \mathrm{Rep}_f G_E$如上所定义的，那么下面说法成立：

(i) 如果$X\in\mathcal{C}$是不可约化的，那么$\mathrm{span}_{\mathbb{C}}\{\pi_X(g),g\in G_E\}$在$\mathrm{End}\,E(X)$是稠的；

(ii) 如果X，$Y\in\mathcal{C}$是不可约化的，并且$X\not\cong Y$，那么$\mathrm{span}_{\mathbb{C}}\{\pi_X(g)\oplus\pi_Y(g),g\in G_E\}$在$\mathrm{End}\,E(X)\oplus\mathrm{End}\,E(Y)$里是稠的。

定理 376。设$\mathcal{C}$是一个 STC^*，并且 E：$\mathcal{C}\to\mathcal{H}$是一个对称 * 保持纤维函子。设 G_E 和 F：$\mathcal{C}\to \mathrm{Rep}_f G_E$ 是如上所定义的，那么 F 是一个对称张量 * 范畴的一个等价。

证明。我们已经知道 F 是一个忠实对称张量函子，在命题 318 看来，还需要证明的是，F 是完全的并且本质上是满射的。

由于范畴$\mathcal{C}$和 $\mathrm{Rep}_f G_E$ 是半单的，为了证明 F 是完全的，必须证明：(a)如果 $X\in\mathcal{C}$是不可约化的，$F(X)\in \mathrm{Rep}_f G_E$ 是不可约化的；并且(b)只要 X，$Y\in\mathcal{C}$是不可约化的和不等价的，那么 $\mathrm{Hom}(F(X), F(Y))=\{0\}$。由于，命题 375 的(i)明显蕴含 $\mathrm{End}(F(X))=\mathbb{C}\,\mathrm{id}$，这是 $F(X)$要求的不可约化性。现在设 X，$Y\in\mathcal{C}$是不可约化的并且是非同构的，并且设 $s\in \mathrm{Hom}(F(X), F(Y))$，即 $s\in \mathrm{Hom}(E(X), E(Y))$和对于所有 $g\in G_E$ 都有 $s\pi_X(g)=\pi_Y(g)s$。那么命题 375 的(ii)意味着对于所有 $u\in \mathrm{End}\,E(X)$和 $v\in \mathrm{End}\,E(Y)$都有 $su=vs$。就 $u=0$ 和 $v=1$ 而言，这意味着 $s=0$，因此不可约化表示 $F(X)=(E(X), \pi_X)$和 $F(Y)=(E(X), \pi_Y)$是非同构的。这就证明了 F 是完全的。

因此，F 是具有 $\mathrm{Rep}_f G_E$ 的完全张量子范畴$\mathcal{C}$的一个等价。如果 $g\in G_E$ 是非平凡的，靠 G_E 的定义直接可得，存在一个 $X\in\mathcal{C}$使得 $g_X\neq \mathrm{id}_{E(X)}$，但是这意味着 $\pi_X(g)\neq 1$。换言之，表示$\{F(X), X\in\mathcal{C}\}$分离了 G_E 的点。但这是彼得—外尔(Peter-Weyl)定理的一个著名结果，一个完全单项子范畴 $\mathrm{Rep}_f G_E$ 分离了 G_E 的点，当且仅当它实际上等价于 $\mathrm{Rep}_f G_E$。因此函子 F 本质上是满射的，而我们做到了。

由于它们是如此重要，我们以幺正自洽的方式重述定理 373 和 376。

定理 377。设$\mathcal{C}$是一个 STC^*，并且 E：$\mathcal{C}\to\mathcal{H}$是一个 * 保持对称纤维函子。设 G_E 是具有源自于 $\Pi_{X\in\mathcal{C}}\mathcal{U}(E(X))$拓扑学 E 的幺正单项自然变换群，那么 G_E 是紧致的，并且函子 F：$\mathcal{C}\to \mathrm{Rep}_f G_E$，$X\mapsto(E(X), \pi_X)$，其中 $\pi_X(g)=g_X$，是一个 STC^* 的等价。如果 E_1，E_2：$\mathcal{C}\to\mathcal{H}$是 * 保持对称纤维函子，那么 $E_1\cong E_2$ 从而 $G_{E_1}\cong G_{E_2}$。

评论 378。上述定理本质上是田中经典结果[Tannaka，1939]用现代语言进行的重述。它可能被推广到(尽管没有唯一性部分)设置$\mathcal{C}$是仅有的辫子甚或没有辫子的情况，这导致量子群的一个(具体)田中理论，有兴趣的读者可以参看

[Joyal and Street, 1991]和[Müger *et al.*, 2004]的评论。

在我们回过头来证明定理373(B4小节)和命题375(B5小节)之前，我们确定一下纤维函子存在的必要条件，这将引导我们对定理377的推广。

B2 紧致子群和抽象田中定理

按照定理377，一个STC^*承认一个对称$*$保持纤维函子，就像对称张量$*$范畴那样，是等价于紧致群G的有限维幺正表象，其中G对于同构是唯一决定性的。关于纤维函子的存在，其结果是扭Θ(定义342)提供了一个阻碍，正好是仅有的一个。

定义379。一个STC^*被如此称谓，即便对所有$X \in \mathcal{C}$有$\Theta(X) = \mathrm{id}_X$。

例子380。一个简单计算，使用例子333里给出的r，$\bar{r}$，$c_{X,Y}$精确公式，证明有限维希尔伯特空间的STC^* $\mathcal{H}$是偶数。同样的观点也适合紧致群G的有限维幺正表象的范畴$\mathrm{Rep}_f G$。

这就是认为一个STC^*为了承认一个纤维函子必定是偶数，事实上：

命题381。如果一个STC^* $\mathcal{C}$承认一个$a*$保持对称纤维函子E，那么它是偶数的。

证明。根据命题344，我们具有$E(\Theta(X)) = \Theta(E(X))$，由于$\mathcal{H}$是偶数，这等于是说$\mathrm{id}_{E(X)} = E(\mathrm{id}_X)$，由于$E$是忠实的，这意味着$\Theta(X) = \mathrm{id}_X$。

幸运的是，这是目前仅有的阻碍，从下一小节开始，我们将证明：

定理382。任何偶数STC^*承认一个$a*$保持对称纤维函子E：$\mathcal{C} \to \mathcal{H}$。

结合定理377我们得到：

定理383。设$\mathcal{C}$是一个偶数STC^*，那么存在一个紧致群G，独一无二同构，使得存在STC^*s的一个等价F：$\mathcal{C} \to \mathrm{Rep}_f G$。

定理383对于应用到量子场论还不够普遍，这正是本章主题。结合DHR理论，我们看到不可归约DHR截面的扭是± 1，正负取决于截面是玻色子还是费米子。一般来说由于我们无法事先排除费米截面，我们不能把我们限制在偶数STC^*内。因此我们真正需要的是所有STC^*的刻画，这就要求对紧致群概念的推广。

定义384。一个(紧致)超群是一个(G, k)对，其中G是一个紧的豪斯道夫

(Hausdoff)群，而k是在G中心的二阶群元，一个(紧致)超群的同构α：$(G, k)\xrightarrow{\cong}(G', k')$，是一个(拓扑学的)群的一个同构$\alpha$：$G\to G'$，使得$\alpha(k)=k'$。

定义385。一个紧致超群(G, k)的(有限维、幺正、连续)表征，正好是G的一个(有限维、幺正、连续)表征(H, π)。扭结子和表征的张量积都是为群定义的，因此$\mathrm{Rep}_{(f)}(G, k)\cong \mathrm{Rep}_{(f)}G$作为$C^*$张量的张量范畴。由于$k$是在$G$中心，在$\mathrm{Rep}_{(f)}(G, k)$里的同构自动保持由$\pi(k)$约化的$\mathbb{Z}_2$阶。$\mathrm{Rep}_{(f)}(G, k)$是用对称$\sum_k$装配如下：对任何$(H, \pi)\in \mathrm{Rep}(G, k)$，设$P_{\pm}^{\pi}=(\mathrm{id}+\pi(k))/2$分别是在表示空间$H$的偶数和奇数子空间上的射影算子。那么：

$$\sum\nolimits_k((H, \pi), (H', \pi'))=\sum(H, H')(1-2P_-^{\pi}\otimes P_-^{\pi'})$$

其中$\sum(H, H')$：$H\otimes H'\to H'\otimes H$是普通翻转同构$x\otimes y\mapsto y\otimes x$。因此，对于同类的$x\in H$，$y\in H'$，我们具有$\sum_k((H, \pi), (H', \pi'))$：$x\otimes y\mapsto \pm y\otimes x$，其中负号出现在当且仅当$x\in H_-$和$y\in H'_-$时。在$(G, k)=(\{e, k\}, k)$情况下，我们称$\mathrm{Rep}_f(G, k)$超希尔伯特空间的范畴$\mathcal{SH}$。

评论386。注意k的作用约化了H上$\mathbb{Z}_2$阶，它在G作用下稳定的。由于上述定义的对称$\sum_k$是明确处于有限维超希尔伯特空间的范畴上$\mathcal{SH}$的那一个，我们看到存在一个可遗对称张量函子$\mathrm{Rep}_f(G, k)\to\mathcal{SH}$。

引理387。如上所定义的$\sum_k$是一个在范畴$\mathrm{Rep}(G, k)$上的对称，因此$\mathrm{Rep}_f(G, k)$是一个STC^*。对于任何对象$X=(H, \pi)\in \mathrm{Rep}_f(G, k)$，扭$\Theta(X)$是由$\pi(k)$给定的。

证明。所断言的大多数性质是直接源自于Rep_fG的那些性质，显然$\sum_k((H,\pi),(H',\pi'))\circ\sum_k((H', \pi'), (H, \pi))$是$H'\otimes H$的恒等式。我们仅仅需要证明张量积的自然性和兼容性，这是一个简单练习。同样的说法符合等式$\Theta((H, \pi))=\pi(k)$。

我们需要一个定理383(的证明)的推论：

推论388。对于任意紧致群G，Rep_fG上的单位函子的幺正单项自然变换，形成一个同构于中心$Z(G)$的阿贝尔群。

证明。如果 $k\in Z(G)$ 和 $(H, \pi)\in \mathrm{Rep}_f G$ 是不可约化的，那么 $\pi(k)=\omega_{(H,\pi)}\mathrm{id}_H$。其中 $\omega_{(H,\pi)}$ 是一个标量，定义 $\Theta((H, \pi))=\omega_{(H,\pi)}\mathrm{id}_{(H,\pi)}$，并且推广到不可约化的对象，从而定义一个 $\mathrm{Rep}_f G$ 的幺正单项自然同构。相反，设 $\{\Theta((H, \pi))\}$ 是一个 $\mathrm{Rep}_f G$ 的单位函子的幺正单项同构，并且 $K: \mathrm{Rep}_f G\to\mathcal{H}$ 是一个可遗函子，那么族 $(\alpha_{(H,\pi)}=K(\Theta((H, \pi))))$ 是一个 K 的幺正单项自然同构，根据定理 377，存在一个 $g\in G$，使得对于所有 $(H, \pi)\in \mathrm{Rep}_f G$ 都有 $\alpha_{(H,\pi)}=\pi(g)$。由于 $\pi(g)$ 是任何可约化 (H, π) 单位的积，g 根据舒尔定理处于 $Z(G)$ 之中。显然上述对应是阿贝尔群的一个同构。

以定理 382 为模板我们能够证明这个附录的主要结果。

定理 389。设 $\mathcal{C}$ 是一个 STC^*，那么存在一个紧致超群 (G, k)，独一无二的同构，并且是一个对称张量 * 范畴的等价 $F: \mathcal{C}\to\mathrm{Rep}_f(G, k)$。特别是如果 $K: \mathrm{Rep}_f(G, k)\to\mathcal{SH}$ 是可遗函子，合成 $E=K\circ F: \mathcal{C}\to\mathcal{SH}$ 是一个“超纤维函子”，即忠实的对称 * 保持张量函子成为超希尔伯特空间的 STC^*。

证明。我们定义一个新 $STC^*\ \tilde{\mathcal{C}}$（$\mathcal{C}$ 的“玻色子化”）如下。作为一个张量 * 范畴，$\tilde{\mathcal{C}}$ 跟 $\mathcal{C}$ 一致，对称 $\tilde{c}$ 被定义为：

$$\tilde{c}_{X,Y}=(-1)^{(1-\Theta(X))(1-\Theta(Y))/4}c_{X,Y}$$

对于不可约化的 $X, Y\in \mathrm{Obj}\ \mathcal{C}=\mathrm{Obj}\ \tilde{\mathcal{C}}$，并且由于自然性被扩展到所有对象。容易证明 $(\tilde{\mathcal{C}}, \tilde{c})$ 是一个重新得到的对称张量范畴，事实上是一个偶数的范畴。因此根据定理 383，存在一个紧致群 G，使得有作为 STC^* 的 $\tilde{\mathcal{C}}\simeq\mathrm{Rep}_f G$。把推论 388 应用到范畴 $\tilde{\mathcal{C}}\simeq\mathrm{Rep}_f G$ 和族 $(\Theta(X))_{X\in\mathcal{C}}$，按照源范畴 $\mathcal{C}$ 被定义，证明一个元素 $k\in Z(G)$，$k^2=e$ 的存在，使得对于 $(H, \pi)\in\tilde{\mathcal{C}}\simeq\mathrm{Rep}_f G$ 都有 $\Theta((H, \pi))=\pi(k)$。显然 (G, k) 是一个超群。我们断言作为 STC^*s 的 $\mathcal{C}\simeq\mathrm{Rep}_f(G, k)$。忽略对称性，这显然是正确的，因为作为张量 * 范畴的 $\mathrm{Rep}_f(G, k)\simeq\mathrm{Rep}_f G$。$\mathcal{C}$ 和 $\mathrm{Rep}_f(G, k)$ 等价，作为 STC^*s 的，即考虑到对称性，源自于 $\mathcal{C}$ 跟 $\tilde{\mathcal{C}}$ 相关联这一事实，就像 $\mathrm{Rep}_f(G, k)$ 是明确跟 $\mathrm{Rep}_f G$ 关联的一样，即靠的是由族 $(\Theta(H, \pi)=\pi(k))$ 作用下对称的一个扭。最后，我们观察到 (G, k) 的唯一结果，来自于定理 383 里的 G 的唯一性和推论 388 里 k 的情况。

评论 390。 如上所述，定理 389 由多普利克尔和罗伯茨在[Doplicher and Roberts，1989，Section 7]明确证明，没有用到超群术语，这仅仅是表面差异(注意到我们的超群并不是由这个名称通常所指的那样)。如上所述，这个证明约化成偶范畴和紧致群。独立地并且几乎同时地，德利涅[Deligne，1990]证明了一个类似于(前)代数群定理 382 的结果，说明定理 383 的代数类似，由[Saavedra Rivano，1972；Deligne and Milne，1982]完成。最近，德利涅也讨论了特殊情况，参看[Deligne，2002]。

这就完成了本附录主要结果的讨论，现在我们回过来证明定理 373、命题 375 和定理 382。

B3　源自于纤维函子的某些代数

设$\mathcal{C}$是一个 TC^* 以及 E_1，E_2：$\mathcal{C}\to\mathrm{Vect}_{\mathbb{C}}$ 纤维函子。回顾它们伴随自然同构 $d^i_{X,Y}$：$E_i(X)\otimes E_i(Y)\to E_i(X\otimes Y)$ 和 e^i：$\mathbf{1}_{\mathrm{Vect}}=\mathbb{C}\to E_i(\mathbf{1}_{\mathcal{C}})$。考虑到 c_矢量空间：

$$A_0(E_1,\ E_2)=\bigoplus_{X\in\mathcal{C}}\mathrm{Hom}(E_2(X),\ E_1(X))$$

对于 $X\in\mathcal{C}$和 $s\in\mathrm{Hom}(E_2(X),\ E_1(X))$，我们为 $A_0(E_1,\ E_2)$的元素写$[X,\ s]$，其在 X 取值 s，而在其他地方为零。显然，A_0 肯定由这些元的有限线性组合构成。我们把 $A_0(E_1,\ E_2)$转换成一个$\mathbb{C}$ 代数，通过定义$[X,\ s]\cdot[Y,\ t]=[X\otimes Y,\ u]$，其中 u 是合成：

$$E_2(X\otimes Y)\xrightarrow{(d^2_{X,Y})^{-1}}E_2(X)\otimes E_2(Y)\xrightarrow{s\otimes t}E_1(X)\otimes E_1(Y)\xrightarrow{d^1_{X,Y}}E_1(X\otimes Y)$$

由于$\mathcal{C}$是严格的，我们有$(X\otimes Y)\otimes Z=X\otimes(Y\otimes Z)$和 $\mathbf{1}\otimes X=X=X\otimes\mathbf{1}$。跟满足同构 $d^i_{X,Y}$的 2－上循环类型方程(56)结合，这意味着 $A_0(E_1,\ E_2)$是可结合的。对 $i=1$，2 具有 e^i 的 $d^i_{X,Y}$的兼容性(57)，意味着$[\mathbf{1},\ e^1\circ(e^2)^{-1}]$是代数 $A_0(E_1,\ E_2)$的一个单元。

引理 391。 设$\mathcal{C}$是一个 TC^* 和 E_1，E_2：$\mathcal{C}\to\mathrm{Vect}_{\mathbb{C}}$ 纤维范畴，子空间 $I(E_1,\ E_2)=\mathrm{span}_{\mathbb{C}}\{[X,\ a\circ E_2(s)]-[Y,\ E_1(s)]\circ a]\mid s:\ X\to Y,\ a:\ E_2(Y)\to E_1(X)\}$是一个双侧理想。

证明。为了证明 $I(E_1,\ E_2)\subset A_0(E_1,\ E_2)$是一个理想，设 s：$X\to Y$，$a\in$

$\mathrm{Hom}(E_2(Y), E_1(X))$，因此$[X, a\circ E_2(s)]-[Y, E_1(s)\circ a]\in I(E_1, E_2)$，并设$[Z, t]\in A_0(E_1, E_2)$，那么：

$$
\begin{aligned}
&([X, a\circ E_2(s)]-[Y, E_1(s)\circ a])\cdot[Z, t]\\
&=[X\otimes Z, d^1_{X,Z}\circ(a\circ E_2(s))\otimes t\circ(d^2_{X,Z})^{-1}]\\
&-[Y\otimes Z, d^1_{Y,Z}\circ(E_1(s)\circ a)\otimes t\circ(d^2_{Y,Z})^{-1}]\\
&=[X\otimes Z, d^1_{X,Z}\circ a\otimes t\circ(d^2_{Y,Z})\circ E_2(s\otimes id_Z)]-\\
&\quad[Y\otimes Z, E_1(s\otimes id_Z)\circ d^1_{X,Z}\circ a\otimes t\circ(d^2_{Y,Z})^{-1}]\\
&=[X', a'\circ E_2(s')]-[Y', E_1(s')\circ a']\in I(E_1, E_2)
\end{aligned}
$$

其中在第二个等式里我们使用了d^i的自然性，而在最后一行我们写成$X'=X\otimes Z$，$Y'=Y\otimes Z$，$s'=s\otimes\mathrm{id}_Z$：$X'\to Y'$以及$a'=d^1_{X,Z}\circ a\otimes t\circ(d^2_{Y,Z})^{-1}\in\mathrm{Hom}(E_2(Y'), E_1(X'))$，是为了明确结果是在$I(E_1, E_2)$里的。这证明了$I(E_1, E_2)$是在$A_0(E_1, E_2)$里的左理想。类似地，可以证明它是右理想。

我们用$A(E_1, E_2)$代表商代数$A_0(E_1, E_2)/I(E_1, E_2)$，它也能够理解成通过符号$[X, s]$推广的代数，其中$X\in\mathcal{C}$，$s\in\mathrm{Hom}(E_2(X), E_1(X))$，满足关系$[X, s]+[X, t]=[X, s+t]$和$[X, a\circ E_2(s)]=[Y, E_1(s)\circ a]$，只要满足$s$：$X\to Y$，$a\in\mathrm{Hom}(E_2(Y), E_1(X))$。因此，我们再次用$[X, s]$代表在$A(E_1, E_2)$里$[X, s]\in A_0(E_1, E_2)$的像不会引起混淆。

命题392。设$\mathcal{C}$是一个STC^*和E_1，E_2：$\mathcal{C}\to\mathrm{Vect}_{\mathbb{C}}$纤维函子。如果$E_1$，$E_2$是对称的，那么$A(E_1, E_2)$是对易的。

证明。设$\mathcal{C}$是对称的，并且纤维函子满足$E_i(c_{X,Y})=\sum_{E_i(X),E_i(Y)}$，设$[A, u]$，$[B, v]\in A_0(E_1, E_2)$，因此：

$$A, B\in\mathcal{C}\text{和}u: E_2(A)\to E_1(A),\ v: E_2(B)\to E_1(B)$$

那么：

$$[A, u]\cdot[B, v]=[A\otimes B, d^1_{A,B}\circ u\otimes v\circ(d^2_{A,B})^{-1}]$$

和：

$$
\begin{aligned}
&[B, v]\cdot[A, u]=[B\otimes A, d^1_{B,A}\circ v\otimes u\circ(d^2_{B,A})^{-1}]\\
&=[B\otimes A, d^1_{B,A}\circ\textstyle\sum_{E_1(A),E_2(B)}\circ u\otimes v\circ\sum_{E_2(B),E_1(A)}\circ(d^2_{B,A})^{-1}]\\
&=[B\otimes A, d^1_{B,A}\circ E_1(c_{B,A})\circ u\otimes v\circ E_2(c_{B,A})\circ(d^2_{B,A})^{-1}]
\end{aligned}
$$

$$=[B\otimes A,\ E_1(c_{A,B})\circ d^1_{A,B}\circ u\otimes v\circ(d^2_{A,B})^{-1}\circ E_2(c_{B,A})]$$

根据 $X=A\otimes B$，$Y=B\otimes A$，$s=c_{A,B}$和 $a=d^1_{A,B}\circ u\otimes v\circ(d^2_{A,B})^{-1}\circ E_2(c_{B,A})$，我们得到：

$$[A,\ u]\cdot[B,\ v]=[X,\ a\circ E_2(s)]$$

$$[B,\ v]\cdot[A,\ u]=[Y,\ E_1(s)\circ a]$$

因此：

$$[A,\ u]\cdot[B,\ v]-[B,\ v]\cdot[A,\ u]$$
$$=[X,\ a\circ E_2(s)]-[Y,\ E_1(s)\circ a]\in I(E_1,\ E_2)$$

说明$[A_0(E_1,\ E_2),\ A_0(E_1,\ E_2)]\subset I(E_1,\ E_2)$。这表明 $A(E_1,\ E_2)=A_0(E_1,\ E_2)/I(E_1,\ E_2)$的对易性。

命题393。设$\mathcal{C}$是一个 TC^*，并且设 E_1，E_2：$\mathcal{C}\to\mathcal{H}$是一个 * 保持纤维函子，那么$A(E_1,\ E_2)$具有一个正 * 运算，即一个反线性和反乘法对合，使得 $a^*a=0$ 隐含 $a=0$。

证明。我们在 $A_0(E_1,\ E_2)$上定义一个 * 运算★，设$[X,\ s]\in A_0(E_1,\ E_2)$，选择一个标准共轭$(\overline{X}_i,\ r_i,\ \overline{r}_i)$，定义 $t\in\mathrm{Hom}_{\mathcal{H}}(E_2(\overline{X}),\ E_1(\overline{X}))$为：

$$t=\mathrm{id}_{E_1(\overline{X})}\otimes E_2(\overline{r}^*)\circ\mathrm{id}_{E_1(\overline{X})}\otimes s^*\otimes\mathrm{id}_{E_2(\overline{X})}\circ E_1(r)\otimes\mathrm{id}_{E_2(\overline{X})}$$

并且取$[X,\ s]^*:=[\overline{X},\ t]$。当然，$s^*$是通过使用在希尔伯特空间 $E_1(X)$，$E_2(X)$上的内积来定义。如果我们选择另一个 X 的标准共轭$(\overline{X}',\ r',\ \overline{r}')$，我们知道存在一个幺正 u：$\overline{X}\to\overline{X}'$，使得 $r'=u\otimes\mathrm{id}_X\circ r$ 和 $\overline{r}'=\mathrm{id}_X\otimes u\circ\overline{r}$。使用$(\overline{X}',\ r',\ \overline{r}')$我们得到$([X,\ s]^*)':=[\overline{X}',\ t']$，其中 t'通过用 r'，$\overline{r}'$替代 r，$\overline{r}$来定义。现在，$[\overline{X},\ t]-[\overline{X}',\ t']$等于：

$$[\overline{X},\ \mathrm{id}_{E_1(\overline{X})}\otimes E_2(\overline{r}^*)\circ\mathrm{id}_{E_1(\overline{X})}\otimes s^*\otimes\mathrm{id}_{E_2(\overline{X})}\circ E_1(r)\otimes\mathrm{id}_{E_2(\overline{X})}]$$
$$-[\overline{X}',\ \mathrm{id}_{E_1(\overline{X}')}\otimes E_2(\overline{r}'^*)\circ\mathrm{id}_{E_1(\overline{X}')}\otimes s^*\otimes\mathrm{id}_{E_2(\overline{X}')}\circ E_1(r')\otimes\mathrm{id}_{E_2(\overline{X}')}]$$
$$=[\overline{X},\ (\mathrm{id}_{E_1(\overline{X})}\otimes E_2(\overline{r}'^*)\circ\mathrm{id}_{E_1(\overline{X})}\otimes s^*\otimes\mathrm{id}_{E_2(\overline{X}')}\circ E_1(r)\otimes\mathrm{id}_{E_2(\overline{X}')})\circ E_2(u)]$$
$$-[\overline{X}',\ E_1(u)\circ(\mathrm{id}_{E_1(\overline{X})}\otimes E_2(\overline{r}'^*)\circ\mathrm{id}_{E_1(\overline{X})}\otimes s^*\otimes\mathrm{id}_{E_2(\overline{X}')}\circ E_1(r)\otimes\mathrm{id}_{E_2(\overline{X}')})]$$

这是处于命题398 里定义的理想 $I(E_1,\ E_2)$里的。因此，如果$[X,\ s]^*$依赖于选择 X 的共轭$(\overline{X},\ r,\ \overline{r})$，它的像 $\gamma([X,\ s]^*)\in A(E_1,\ E_2)$并不如此。

为了能够通过 $x^*:=\gamma\circ\bigstar\circ\gamma^{-1}$定义一个 $A(E_1,\ E_2)$上的 * 运算，我们必须

证明合成映射 $\gamma\circ\star$：$A_0(E_1, E_2)\to A(E_1, E_2)$，把 $I(E_1, E_2)$ 映射到零。为了达到这个目的，设 $X, Y\in\mathcal{C}$，s：$X\to Y$，$a\in \mathrm{Hom}(E_2(Y), E_1(X))$，并且选择共轭 $(\overline{X}, r_X, \overline{r}_X)$，$(\overline{Y}, r_Y, \overline{r}_Y)$，那么：

$$\begin{aligned}
&[X, a\circ E_2(s)]^{\star}-[Y, E_1(s)\circ a]^{\star}\\
&=[\overline{X}, \mathrm{id}_{E_1(\overline{X})}\otimes E_2(\overline{r_X}^*)\circ \mathrm{id}_{E_1(\overline{X})}\otimes(a\circ E_2(s))^*\otimes \mathrm{id}_{E_2(\overline{X})}\circ E_1(r_X)\otimes \mathrm{id}_{E_2(\overline{X})}]\\
&-[\overline{Y}, \mathrm{id}_{E_1(\overline{X})}\otimes E_2(\overline{r_Y}^*)\circ \mathrm{id}_{E_1(\overline{X})}\otimes(E_1(s)\circ a)^*\otimes \mathrm{id}_{E_2(\overline{X})}\circ E_1(r_Y)\otimes \mathrm{id}_{E_2(\overline{X})}]\\
&=[\overline{X}, \tilde{a}\circ E_2(\tilde{s})]-[\overline{Y}, E_1(\tilde{s})\circ \tilde{a}]
\end{aligned}$$

其中 $\tilde{a}\in \mathrm{Hom}_{\mathcal{H}}(E_2(\overline{Y}), E_1(\overline{X}))$ 和 $\tilde{s}\in \mathrm{Hom}(\overline{X}, \overline{Y})$ 定义为：

$$\tilde{a}=\mathrm{id}_{E_1(\overline{X})}\otimes E_2(\overline{r}_X^*)\circ \mathrm{id}_{E_1(\overline{X})}\otimes a^*\otimes \mathrm{id}_{E_2(\overline{Y})}\circ E_1(r_X)\otimes \mathrm{id}_{E_2(\overline{Y})}$$

$$\tilde{s}=\mathrm{id}_{\overline{Y}}\otimes\overline{r}_X^*\circ \mathrm{id}_{\overline{Y}}\otimes s^*\otimes \mathrm{id}_{\overline{X}}\circ r_Y\otimes \mathrm{id}_{\overline{X}}$$

这显然是在 $I(E_1, E_2)$ 里，因此 $x^*:=\gamma\circ\star\circ\gamma^{-1}(x)$ 定义 $A(E_1, E_2)$ 上的一个 * 运算。

现在很明显，在 $A(E_1, E_2)$ 上的最终 * 映射是可加的和反线性的，它也是合取的和反乘法的，正如人们通过共轭范畴的适当使用所证明的。我们省去烦琐的直接计算，接下来证明 * 运算的正性。考虑 $[\overline{X}, s]\in A_0(E_1, E_2)$，选择共轭 $(\overline{X}, r, \overline{r})$ 并且计算 $[X, s]^*\cdot[X, s]=[\overline{X}\otimes X, t]$，其中 t 等于：

$$d^1_{\overline{X},X}\circ(\mathrm{id}_{E_1(\overline{X})}\otimes E_2(\overline{r}^*)\circ \mathrm{id}_{E_1(\overline{X})}\otimes s^*\otimes \mathrm{id}_{E_2(\overline{X})}\circ E_1(r)\otimes \mathrm{id}_{E_2(\overline{X})})\otimes s\circ(d^2_{\overline{X},X})^*$$

现在有：

$$\begin{aligned}
&[\overline{X}\otimes X, t]=[\overline{X}\otimes X, E_1(r^*)\circ E_1(r)\circ t]=[\mathbf{1}, E_1(r)\circ t\circ E_2(r^*)]\\
&=[\mathbf{1}, E_1(r^*)\circ(\mathrm{id}_{E_1(\overline{X})}\otimes E_2(\overline{r}^*)\circ \mathrm{id}_{E_1(\overline{X})}\otimes s^*\otimes \mathrm{id}_{E_2(\overline{X})}\circ E_1(r)\otimes \mathrm{id}_{E_2(\overline{X})})\otimes s\circ E_2(r)]\\
&=[\mathbf{1}, E_1(r^*)\circ \mathrm{id}\otimes(s\circ s^*)\circ E_1(r)]=[\mathbf{1}, u^*u]
\end{aligned}$$

其中我们使用了共轭范畴，并且取 $u=\mathrm{id}\otimes s^*\circ E_1(r)$，因此，$[X, s]^*\cdot[X, s]=[\mathbf{1}, u^*u]$ 是零，当且仅当 u^*u 是零，根据 $\mathcal{H}$ 里 * 运算的正性，这得以保持当且仅当 $u=0$。再次使用共轭范畴，我们看到这等价于 $s=0$。因此对于形式 $[X, s]$ 的元 $a\in A(E_1, E_2)$，说明 $a^*a=0\Rightarrow a=0$ 成立。对于一般的 $a=\sum_i(X_i, s_i)$，我们选择等距 v_i：$X_i\to X$，使得 $\sum_i v_i\circ v_i^*=\mathrm{id}_X$（即 $X\cong\bigoplus_i X_i$）。那么 $[X_i, s_i]=[X, E_1(v_i)\circ s_i\circ E_2(v_i^*)]$，因此：

$$\sum_i [X_i, s_i] = [X, \sum_i E_1(v_i) \circ s_i \circ E_2(v_i^*)]$$

这表明 $A(E_1, E_2)$ 的任何元都能写成 $[X, s]$，而我们正是这样做的。

命题394。设 $\mathcal{C}$ 是一个 TC^*，并且设 $E_1, E_2: \mathcal{C} \to \mathcal{H}$ 是一个 * 保持纤维函子，那么 $\|a\| = \inf_b{}' \sup_{X \in \mathcal{C}} \|b_X\|_{\mathrm{End}E(X)}$。其中下确界是遍布在 $a \in A(E_1, E_2)$ 的 $b \in A_0(E_1, E_2)$ 的所有表征，定义一个在 $A(E_1, E_2)$ 上的 C^* 范数。

证明。设 $[X, s], [Y, t] \in A_0(E_1, E_2)$，那么 $[X, s] \cdot [Y, t] = [X \otimes Y, u]$，其中 $u = d^1_{X,Y} \circ s \otimes t \circ (d^2_{X,Y})^{-1}$。由于 $d^1_{X,Y}$，$d^2_{X,Y}$ 是幺正的，我们具有 $\|[X \otimes Y, u]\| = \|u\| \leqslant \|s\| \cdot \|t\|$，因此 $\|b\| = \sup_{X \in \mathcal{C}} \|b_X\|_{\mathrm{End}E(X)}$ 定义 $A_0(E_1, E)$ 上的次可乘性范数，而上述 $\|a\|$ 公式是一个商代数 $A_0(E_1, E_2)/I(E_1, E_2)$ 上的范数的常规定义。这个范数满足 $\|[X, s]\| = \|s\|$，因为每一个 $a \in A(E_1, E_2)$ 都能够写成 $[X, s]$，我们具有 $\|a\| = 0 \Rightarrow a = 0$。最后，在命题393的证明里的计算意味着：

$$\|[X, s]^*[X, s]\| = \|[1, u^*u]\| = \|u^*u\| = \|u\|^2 = \|s\|^2 = \|[X, s]\|^2$$

这是 C^* 条件。

定义395。设 $\mathcal{C}$ 是一个 TC^*，并且设 $E_1, E_2: \mathcal{C} \to \mathcal{H}$ 是 * 保持纤维函子。那么 $\mathcal{A}(E_1, E_2)$ 代表 $A(E_1, E_2)$ 的 $\|\cdot\|$ 完成。(这是一个幺正 C^* 代数，如果 $\mathcal{C}$，E_1，E_2 都是对称的，那么它就是对易的。)

B4 纤维函子的唯一性

引理396。参见[Joyal and Street, 1993a]。设 $\mathcal{C}$ 是一个 TC^*，$\mathcal{D}$ 是一个严格张量范畴，而 $E_1, E_2: \mathcal{C} \to \mathcal{D}$ 是一个严格张量函子。那么任何单项自然变换 $\alpha: E_1 \to E_2$ 是一个自然同构。

证明。完全可以证明任何组元 $\alpha_X: E_1(X) \to E_2(X)$ 具有一个双侧逆元 $\beta_X: E_2(X) \to E_1(X)$。族 $\{\beta_X, X \in \mathcal{C}\}$ 就会自动成为一个自然变换。如果 $(\overline{X}, r, \bar{r})$ 是 X 的共轭，α 的单项性意味着：

$$E_2(r^*) \circ \alpha_{\overline{X}} \otimes \alpha_X = E_2(r^*) \circ \alpha_{\overline{X} \otimes X} = \alpha_1 \circ E_1(r^*) = E_1(r^*) \tag{58}$$

如果我们现在定义：

$$\beta_X = \mathrm{id}_{E_1(X)} \otimes E_2(r^*) \circ \mathrm{id}_{E_1(X)} \otimes \alpha_{\overline{X}} \otimes \mathrm{id}_{E_2(X)} \circ E_1(\bar{r}) \otimes \mathrm{id}_{E_2(X)}$$

我们就有：

$\beta_X\circ\alpha_X=(\mathrm{id}_{E_1(X)}\otimes E_2(r^*)\circ\mathrm{id}_{E_1(X)}\otimes\alpha_{\bar{X}}\otimes\mathrm{id}_{E_2(X)}\circ E_1(\bar{r})\otimes\mathrm{id}_{E_2(X)})\circ\alpha_X$

$=\mathrm{id}_{E_1(X)}\otimes E_2(r^*)\circ\mathrm{id}_{E_1(X)}\otimes\alpha_{\bar{X}}\otimes\alpha_X\circ E_1(\bar{r})\otimes\mathrm{id}_{E_1(X)}$

$=\mathrm{id}_{E_1(X)}\otimes E_1(r^*)\circ E_1(\bar{r})\otimes\mathrm{id}_{E_1(X)}=\mathrm{id}_{E_1(X)}$

对 $\alpha_X\circ\beta_X=\mathrm{id}_{E_2(X)}$ 的论证也类似。

评论 397。这个引理仍然正确，如果人们允许 E_1，E_2（甚或$\mathcal{C}$，$\mathcal{D}$）是非严格的。为了改写这个证明，人们用$(d^{E_1}_{\bar{X},X})^{-1}\circ E_1(r)\circ e^{E_1}$（这是一个态射 $\mathbf{1}_{\mathrm{Vect}}\to E_1(\bar{X})\otimes E_1(X)$）替代 $E_1(r)$（一个态射 $E_1(\mathbf{1})\to E_1(\bar{X}\otimes X)$）。类似情况也适合 $E_2(\bar{r})$。

命题 398。设$\mathcal{C}$是一个 TC^*，并且 E_1，E_2：$\mathcal{C}\to\mathrm{Vect}_{\mathbb{C}}$ 是纤维函子。给定的自然变换 $E_1\to E_2$ 在 $A_0(E_1,E_2)$跟下面矢量空间之间的对：

$$\begin{aligned}\mathrm{Nat}(E_1,E_2)\ =\ \{(\alpha_X)_{X\in\mathcal{C}}\in\prod_{X\in\mathcal{C}}\mathrm{Hom}(E_1(X),E_2(X))\mid\\ E_2(s)\circ\alpha_X\ =\ \alpha_Y\circ E_1(s)\ \forall s:X\to Y\}\end{aligned}$$

对于$(\alpha_X)\in\mathrm{Nat}(E_1,E_2)$和 $a\in A_0(E_1,E_2)$，通过

$$\langle\alpha,\ a\rangle=\sum_{X\in\mathcal{C}}Tr_{E_1(X)}(a_X\alpha_X)\tag{59}$$

从而下降到一个 $\mathrm{Nat}(E_1,E_2)$跟商代数 $A(E_1,E_2)=A_0(E_1,E_2)/I(E_1,E_2)$之间的对，使得 $\mathrm{Nat}(E_1,E_2)\cong A(E_1,E_2)^*$。

在此同构下，一个元 $a\in A(E_1,E_2)^*$对应于 $\mathrm{Nat}_{\otimes}(E_1,E_2)$的一个元，即一个单项自然变换（从而通过引理396同构），当且仅当它是一个特征，即可乘性。

证明。直和 $A_0(E_1,E_2)$的对偶矢量空间是直积 $\prod_{X\in\mathcal{C}}\mathrm{Hom}(E_2(X),E_1(x))^*$，并且由于 $\mathrm{Hom}(E_2(X),E_1(X))\times\mathrm{Hom}(E_1(X),E_2(X))$，$s\times t\mapsto Tr(s\circ t)$之间的配对是非简并的，我们有关于式(59)所给的配对的：

$$A_0(E_1,E_2)^*\cong\prod_{X\in\mathcal{C}}\mathrm{Hom}(E_1(X),E_2(X))$$

现在，$A(E_1,E_2)$是 $A_0(E_1,E_2)$跟子空间 $I(E_1,E_2)$的商，因此对偶空间 $A(E_1,E_2)^*$肯定由 $A_0(E_1,E_2)^*$的那些元素构成，它们在 $I(E_1,E_2)$上等同于零。假定$(a_X)_{X\in\mathcal{C}}$对所有 $a\in I(E_1,E_2)$满足$\langle\alpha,a\rangle=0$，等价地对于所有 s：$X\to Y$ 和 a：$E_2(Y)\to E_1(X)$有$\langle\alpha,[X,a\circ E_2(s)]-[Y,E_1(s)\circ a]\rangle=0$。通过配对的定义(59)，这等价于说对于所有 s：$X\to Y$ 和 $a\in\mathrm{Hom}(E_2(Y),E_1(X))$都有：

$$Tr_{E_1X}(a \circ E_2(s) \circ \alpha_X) - Tr_{E_1(Y)}(E_1(s) \circ a \circ \alpha_Y) = 0$$

迹的非简并性意味着，$\alpha = (\alpha_X)_{X \in \mathcal{C}}$一定对于所有 s：$X \to Y$ 都满足$E_2(s) \circ \alpha_X = \alpha_Y \circ E_1(s)$，因此 $\alpha \in \mathrm{Nat}(E_1, E_2)$，这意味着：

$$A(E_1, E_2)^* \cong \mathrm{Nat}(E_1, E_2)$$

现在我们考虑这个问题，当对应于 $\alpha \in \mathrm{Nat}(E_1, E_2)$ 的泛函 $\phi \in A(E_1, E_2)^*$ 是个特性(即可乘性)的时候。这个情况对应于下述时候：

$\langle \alpha, [X, s] \cdot [Y, t] \rangle = \langle \alpha, [X, s] \rangle \langle \alpha, [Y, t] \rangle \quad \forall [X, s], [Y, t] \in A(E_1, E_2)$

严格地说，$[X, s]$，$[Y, t]$是在 $A_0(E_1, E_2)$里对一些在 $A(E_1, E_2)$里的元素的表征。按照(59)和 $A(E_1, E_2)$里积的定义，这等同于：

$$Tr_{E_1(X \otimes Y)}(d^1_{X,Y} \circ s \otimes t \circ (d^2_{X,Y})^{-1} \circ \alpha_{X \otimes Y}) = Tr_{E_1(X)}(s \circ \alpha_X) Tr_{E_1(Y)}(t \circ \alpha_Y)$$

$$= Tr_{E_1(X) \otimes E_2(X)}((s \circ \alpha_X) \otimes (t \circ \alpha_Y))$$

$$= Tr_{E_1(X) \otimes E_2(X)}(s \otimes t \circ \alpha_X \otimes \alpha_Y)$$

按照循环不变性和迹的非简并性，这对于所有 s：$E_2(X) \to E_1(X)$ 和 t：$E_2(Y) \to E_1(Y)$是成立的，当且仅当：

$$\alpha_{X \otimes Y} = d^2_{X,Y} \circ \alpha_X \otimes \alpha_Y \circ (d^1_{X,Y})^{-1} \quad \forall X, Y \in \mathcal{C}$$

这确实是 $\alpha \in \mathrm{Nat}(E_1, E_2)$成为单项的条件，即 $\alpha \in \mathrm{Nat}_{\otimes}(E_1, E_2)$。

命题 399。设$\mathcal{C}$是 TC^* 并且设 E_1，E_2：$\mathcal{C} \to \mathcal{H}$是 * 保持纤维函子，那么一个单项自然变换 $\alpha \in \mathrm{Nat}_{\otimes}(E_1, E_2)$是幺正的(即每一个 α_X 是幺正的)，当且仅当对应的特征 $\phi \in A(E_1, E_2)$是一个 * 同构(即 $\phi(a^*) = \overline{\phi(a)}$)。

证明。设 $\alpha \in \mathrm{Nat}_{\otimes}(E_1, E_2)$和$[X, s] \in A(E_1, E_2)$，通过 $A(E_1, E_2)$和 $\mathrm{Nat}(E_1, E_2)$对的定义：

$$\phi([X, s]) = \langle \alpha, [X, s] \rangle = Tr_{E_1(X)}(s \circ \alpha_X)$$

使用$\overline{Tr(AB)} = Tr(A^* B^*)$，因此：

$$\overline{\phi([X, s])} = Tr_{E_1(X)}(s^* \circ \alpha^*_X)$$

另一方面，有：

$\phi([X, s]^*) = \langle \alpha, [\overline{X}, \mathrm{id}_{E_1(\overline{X})} \otimes E_2(\overline{r}^*) \circ \mathrm{id}_{E_1(\overline{X})} \otimes s^* \otimes \mathrm{id}_{E_2(\overline{X})} \circ E_1(r) \otimes \mathrm{id}_{E_2(\overline{X})}] \rangle$

$$= Tr_{E_1(\overline{X})}(\mathrm{id}_{E_1(\overline{X})} \otimes E_2(\overline{r}^*) \circ \mathrm{id}_{E_1(X)} \otimes s^* \otimes \mathrm{id}_{E_2(\overline{X})} \circ E_1(r) \otimes \mathrm{id}_{E_2(\overline{X})} \circ \alpha_{\overline{X}})$$

$$= E_2(\overline{r}^*) \circ s^* \otimes \alpha_{\overline{X}} \circ E_1(\overline{r})$$

$$= E_2(\overline{r}^*) \circ (\alpha_X \circ \alpha_X^{-1} \circ s^*) \otimes \alpha_{\overline{X}} \circ E_1(\overline{r})$$

$$= E_1(\overline{r}^*) \circ (\alpha_X^{-1} \circ s^*) \otimes \mathrm{id}_{E_1(\overline{X})} \circ E_1(\overline{r})$$

$$= Tr_{E_1(X)}(\alpha_X^{-1} \circ s^*)$$

在第四步我们使用了 α 的可逆性(引理 396)，而在第五个等式我们用了式(58)中 X 和 $\overline{X}$ 的相互交换以及用 $\overline{r}$ 代替 r。现在迹的非简并性，意味着对所有 $[X, s] \in [E_1, E_2]$ 都有 $\overline{\phi([X, s])} = \phi([X, s]^*)$，当且仅当对于所有 $X \in C$ 都有 $\alpha_X^* = \alpha_X^{-1}$，就像断言的那样。

现在我们可以证明我们杰出断言的第一个。

定理 373 的证明。通过上述建构，$A(E_1, E_2)$ 的 $\|\cdot\|$ 封闭 $\mathcal{A}(E_1, E_2)$，是一个对易的幺正 C^* 代数。因此，它具有(很多)特征，即幺正 * 同构到 $\mathbb{C}$ (参看下述定理 401)。这样一个特征仅限于 $A(E_1, E_2)$，并且根据命题 398 和 399 对应于一个幺正单项自然变换 $\alpha \in \mathrm{Nat}(E_1, E_2)$。

评论 400。

1. 代数 $A(E_1, E_2)$ 的讨论受启发于那个预印本[Bichon, ND]，而在正式版本[Bichon, 1998]里没有放进去。定理 373 的上述证明首次出现在[Bichon, 1999]；

2. 引理 396 意味着由纤维函子和单项自然变换构成的范畴是一个广群，即任何态射是不可逆的。那么定理 373 就暗含着由对称 * 保持纤维函子和幺正单项自然变换构成的范畴是一个可迁群，即所有对象都是同构的。这个广群是非平凡的，这是定理 382 的说法，其证明形成这一节的主要内容，从 B6 小节开始。

B5 具体的田中定理 第二部分

为了证明命题 375，我们需要前几节的形式化体系。我们把前面定义的对易的幺正 C^* 代数 $\mathcal{A}(E, E)$ 写成 $\mathcal{A}(E)$。为了研究这个代数，我们需要一些有关对易单位 C^* 代数的结果，它们可能集中在比如[Pedersen, 1989]。

定理 401。 设 $\mathcal{A}$ 是一个对易的幺正 C^* 代数，设 $\mathcal{A}^*$ 是其巴拿赫空间对偶，

并且设：

$P(\mathcal{A})=\{\phi\in\mathcal{A}^* \mid \phi(1)=1,\ \|\phi\|\leqslant 1\}$

$X(\mathcal{A})=\{\phi\in\mathcal{A}^* \mid \phi(1)=1,\ \phi(ab)=\phi(a)\phi(b),\ \phi(a^*)=\overline{\phi(a)}$ $\forall a,\ b\in\mathcal{A}\}$

$P(\mathcal{A})$和$X(\mathcal{A})$都具有在$\mathcal{A}$上的ω^*拓扑，对应于$\phi_\iota\to\phi$当且仅当对于所有$a\in\mathcal{A}$都有$\phi_\iota(a)\to\phi(a)$。那么：

(i) $X(\mathcal{A})\subset P(\mathcal{A})$（即 * 特征具有范数$\leqslant 1$）；

(ii) $X(\mathcal{A})$是在$P(\mathcal{A})$上w^*拓扑紧致的；

(iii) 由$a\mapsto(\phi\mapsto\phi(a))$给定的映射$\mathcal{A}\to C(X(\mathcal{A}))$，是一个$C^*$代数的同构；

(iv) $X(\mathcal{A})$的凸包

$$\left\{\sum_{i=1}^{N} c_i\phi_i,\ N\in\mathbb{N},\ c_i\in\mathbb{R}_+,\ \sum_i c_i=1,\ \phi_i\in X(\mathcal{A})\right\}$$

在$P(\mathcal{A})$里是w^*稠的。

证明。

(i) 巴拿赫代数的任何幺正 * 同态α满足$\|\alpha(a)\|\leqslant\|a\|$；

(ii) 根据阿劳格鲁（Alaoglu）的定理[Pedersen，1989，Theorem 2.5.2]，$\mathcal{A}^*$的单位球是关于w^*拓扑紧致的，并且封闭子集$X(\mathcal{A})\subset P(\mathcal{A})\subset\mathcal{A}^*$也如此；

(iii) 这是盖尔范德的定理，参看[Pedersen，1989，Theorem 4.3.13]；

(iv) 这是克莱因—米尔曼（Krein-Milman）定理，参看定理 2.5.4 结合[Pedersen，1989]里的命题 2.5.7。

定理 401。(ii) 意味着$\mathcal{A}(E)$的 * 特征的集合$X\equiv X(\mathcal{A}(E))$，是关于$w^*$拓扑的紧致豪斯道夫空间。根据(iii)和命题 399，X的元素是跟函子E的幺正单项变换的集合G_E处于双射对应之中。

引理 402。双射$X\cong G_E$关于上述定义的拓扑是一个同胚。

证明。根据$\prod_{X\in\mathcal{C}}\mathcal{U}(E(X))$上积拓扑的定义，一个在$G_E$里的网$(g_\iota)$是收敛的，当且仅当在$\mathcal{U}(E(X))$里的网$(g_\iota,\ X)$对于任何$X\in\mathcal{C}$收敛。另一方面，在$X$里的$(\phi_\iota)$网收敛，当且仅当$(\phi_\iota(a))$对于任何$a\in\mathcal{A}(E)$在$\mathbb{C}$中收敛。从建立在命题 398 里的对应$\phi\leftrightarrow g$形式看来，这两个收敛概念是一致的。

同胚 $X \cong G_E$ 允许把 G_E 自动具有的拓扑群结构移到紧致空间 X，现在我们来完成我们第二个杰出断言的证明。

命题 375 的证明。由于$\mathcal{C}$是半单的并且本质上小，存在不可约化对象的一个集合 I 和族$\{X_i,\ i \in I\}$，使得任何对象(同构于)来自这个集合的对象的一个有限直和。如果 $\mathrm{Nat}(E) \equiv \mathrm{Nat}(E,\ E)$是从 E 到自身的自然变换的空间，具有我们能够跟族$(\alpha_i = \alpha_{X_i})_{i \in I}$结合的任何 $\alpha \in \mathrm{Nat}(E)$，这是 $\prod_{i \in I} \mathrm{End}E(X_i)$的一个元素。$\mathcal{C}$的半单性和 α 的自然性意味着，任何这样一个元素都是恰好产生于 E 的自然变换。假若还不明确，可以参看[Müger *et al.*，2004，Prop. 5.4]里的证明。按照这种方法我们得到矢量空间的一个同构：

$$\gamma:\ \mathrm{Nat}(E) \to \prod_{i \in I} \mathrm{End}E(X_i),\ \alpha \mapsto (\alpha_{X_i})_{i \in I}$$

现在看看线性映射：

$$\delta:\ \bigoplus_{i \in I} \mathrm{End}E(X_i) \to A(E),\ (a_i) \mapsto \sum_i\ [X_i,\ a_i]$$

由于任何 $a \in A(E)$都能写成$[X,\ s]$(命题 393 的证明)，并且任何$[X,\ s]$都是具有不可约化 X_i 的元素$[X_i,\ s_i]$的直和，所以 δ 是满射的，一旦理解为一个到$A_0(E)$的映射，δ 就是单射的。作为 $\mathrm{Hom}(X_i,\ X_j) = \{0\}$ 在 $i \neq j$ 时的一个结果，δ 的在 $A_0(E)$里的像有具备理想 $I(E)$的平凡截面，这是商映射 $A_0(E) \to A(E)$的核，因此 δ 是单射从而是一个(矢量空间的，而不是代数的)同构。如果在 $A(E)$上的 C^* 范数通过 δ 被拖回，我们就得到$\bigoplus_{i \in I}\mathrm{End}E(X_i)$上的范数：

$$\|(a_i)_{i \in I}\| = \sup_{i \in I} \|a_i\|_{\mathrm{End}E(X_i)}$$

因此我们具有范数闭包的同构 $\bar{\delta}$：$\overline{\bigoplus_{i \in I}\mathrm{End}E(X_i)}^{\|\cdot\|} \to \mathcal{A}(E)$。关于同构 γ，δ，命题 398 的对$\langle \cdot,\ \cdot \rangle$：$\mathrm{Nat}(E) \times A(E) \to \mathbb{C}$ 变成：

$$\langle \cdot,\ \cdot \rangle^{\sim}:\ \prod_{i \in I} \mathrm{End}E(X_i) \times \bigoplus_{i \in I} \mathrm{End}E(X_i) \to \mathbb{C},\ (\alpha_{X_i}) \times (a_i) \mapsto \sum_{i \in I} Tr_{E(X_i)}(\alpha_i a_i)$$

更准确地是，$\langle \cdot,\ \delta(\cdot) \rangle = \langle \gamma(\cdot),\ \cdot \rangle^{\sim}$ 作为映射 $\mathrm{Nat}(E) \times \bigoplus_{i \in I}\mathrm{End}E(X_i) \to \mathbb{C}$。因此如果 $\alpha \in \mathrm{Nat}(E)$是如此，使得 $\gamma(\alpha) \in \prod_{i \in I} \mathrm{End}E(X_i)$仅仅具有有限多个非零组分(即 $\gamma(\alpha) \in \bigoplus_{i \in I}\mathrm{End}E(X_i)$)，那么$\langle \alpha,\ \cdot \rangle \in A(E)^*$扩展到 $\mathcal{A}(E)^*$的元素。

现在定理 401 的(iv)意味着，任何 $\phi \in \mathcal{A}(E)^*$ 是在$\mathcal{A}(E)$的 * - 特征

$X(\mathcal{A}(E))$的$\mathbb{C}$生成里一个网(ϕ_ι)的w^*极限。因此对于任何$(\alpha_i) \in \bigoplus_{i\in I}\mathrm{End}E(X_i)$存在这样一个网$(\phi_\iota)$，对

$$w^* - \lim\phi_\iota = \langle \gamma^{-1}((\alpha_i)), \ \cdot \ \rangle \in \mathcal{A}(E)^*$$

限制ϕ_ι到$A(E)$并且使用同构$\mathrm{Nat}E \cong A(E)^*$，我们就得到在收敛于$\gamma^{-1}((\alpha_i))$的$\mathrm{Nat}E$里的一个网。根据命题398，399，同构$A(E)^* \to \mathrm{Nat}E$映射$X(\mathcal{A}(E))$的元素到$E$的幺正自然单项变换，即$G_E$的元素。因此，特别是对于任何有限$S \subset I$，我们具有：

$$\overline{\mathrm{span}}_{\mathbb{C}}\{\underbrace{\pi_{s_1}(g)\oplus\cdots\oplus\pi_{s|S|}(g)}_{所有s\in S} \in G_E\} = \bigoplus_{s\in S} g\mathrm{End}E(X_s)$$

这显然比命题375里断言的要多得多。

这就完成进入定理377证明的全部组成部分的证明。从这个证明来看，显然对易的C^*代数$\mathcal{A}(E)$仅仅是在紧致群G_E上的连续函数的代数，然而$\mathcal{A}(E)$是G_E有限维表示的矩阵元的线性开张。

B6 形成对称纤维函子*保持

这一小节的目标是证明下面的结果，这看上去有些新。

定理403。一个承认对称纤维函子$\mathcal{C}\to\mathrm{Vect}_{\mathbb{C}}$的偶$STC^*\ \mathcal{C}$，也承认对称*保持纤维函子$\mathcal{C}\to\mathcal{H}$。

引理404。设$\mathcal{C}$是一个STC^*，并且E: $\mathcal{C}\to\mathrm{Vect}_{\mathcal{C}}$是一个对称纤维函子。选择任意正定内积$\langle\cdot, \ \cdot\rangle_X^0$(即希尔伯特空间结构)在整个空间$E(X)$，$X\in\mathcal{C}$上。那么映射$X\mapsto E(X)$和$s\mapsto E(s^*)^\dagger$，其中$E(s^*)^\dagger$是关于内积$\langle\cdot, \ \cdot\rangle_X^0$的$E(s^*)$的伴随，就定义了一个忠实的函子$\tilde{E}$: $\mathbb{C}\to\mathrm{Vect}_{\mathbb{C}}$，由于$d_{X,Y}^{\tilde{E}} = ((d_{X,Y}^E)^\dagger)^{-1}$和$e^{\tilde{E}} = ((e^E)^\dagger)^{-1}$，这是一个对称纤维函子。

证明。首先注意到$s\mapsto\tilde{E}(s)$是一个$\mathbb{C}$线性，并且真正定义了一个函子，由于$\tilde{E}(\mathrm{id}_X) = \mathrm{id}_{\tilde{E}(X)}$且：

$\tilde{E}(s\circ t) = E((s\circ t)^*)^\dagger = E(t^*\circ s^*)^\dagger = (E(t^*)\circ E(s^*))^\dagger = E(s^*)^\dagger\circ E(t^*)^\dagger = \tilde{E}(s)\circ\tilde{E}(t)$

E的忠实性明显意味着$\tilde{E}$的忠实性。由于$d_{X,Y}^{\tilde{E}} = ((d_{X,Y}^E)^\dagger)^{-1}$和$e^{\tilde{E}} =$

$((e^{E})^{\dagger})^{-1}$，(56)和(57)的对易性是显然的。由于 E 是一个张量函子，我们对于所有 s：$X \to X'$ 和 t：$Y \to Y'$ 具有：

$$E(s \otimes t) \circ d^{E}_{X,Y} = d^{E}_{X',Y'} \circ E(s) \otimes E(t)$$

这等价于：

$$(E(s \otimes t))^{\dagger} \circ ((d^{E}_{X',Y'})^{-1})^{\dagger} = ((d^{E}_{X,Y})^{-1})^{\dagger} \circ (E(s) \otimes E(t))^{\dagger}$$

由于这对所有 s，t 都成立，我们证明族 $(d^{\tilde{E}}_{X,Y})$ 的自然性，因此 $\tilde{E}$ 是一个张量函子。计算 $\tilde{E}(c_{X,Y}) = E(c^{*}_{X,Y})^{\dagger} = E(c_{Y,X})^{\dagger} = \sum^{\dagger}_{E(Y),E(X)} = \sum_{E(X),E(Y)}$，其中我们使用了 $\sum^{\dagger}_{H,H'} = \sum_{H',H}$，证明 $\tilde{E}$ 也是对称的，因此 $\tilde{E}$ 是一个对称纤维函子。

现在 B3 小节的讨论为我们运用和提供了一个对易的幺正ℂ代数 $A(E, \tilde{E})$。不过，我们不能由命题 393 得出结论说 $A(E, \tilde{E})$ 是一个 * 代数，因为 E，$\tilde{E}$ 不是 * 保持。事实上，对于任意对称纤维从函子 E_1，E_2，没有 $A(E_1, E_2)$ 上正 * 运算存在的理由，但是在目前的情况下，两个函子都由 $E_2(s) = E_1(s^{*})^{\dagger}$ 关联在一起，确实有下面命题：

命题 405。 设 $\mathcal{C}$ 是一个 STC^{*}，E：$\mathcal{C} \to \mathrm{Vect}_{\mathbb{C}}$ 是一个对称纤维函子，而 $\tilde{E}$ 定义如上，那么：

$$[X, s]^{\star} = [X, s^{\dagger}]$$

这是明确定义了的，并且是一个在 $A(E, \tilde{E})$ 上的正 * 运算，关于这个 * 运算，来自命题 394 的范数 $\|\cdot\|$ 是一个 C^{*} 范数，即对于所有 $\|a^{\star}a\| = \|a\|^{2}$ 都有 $a \in A(E, \tilde{E})$。

证明。对于 $[X, s] \in A_0(E, \tilde{E})$，我们定义 $[X, s]^{\star} = [X, s^{\dagger}]$，其中 $s^{\dagger}$ 是 $s \in \mathrm{End}E(X)$ 关于 $E(X)$ 上内积的伴随，显然，★是对合和反线性的。现在，如果 s：$X \to Y$，$a \in \mathrm{Hom}(E_2(Y), E_1(X))$，那么：

$$([X, a \circ E_2(s)]) - [Y, E_1(s) \circ a])^{\star} = [X, a \circ E(s^{*})^{\dagger}]^{\star} - [Y, E(s) \circ a]^{\star}$$
$$= [X, E(s^{*}) \circ a^{\dagger}] - [Y, a^{\dagger} \circ E(s)^{\dagger}] = [X, E_1(s^{*}) \circ a^{\dagger}] - [Y, a^{\dagger} \circ E_2(s^{*})]$$

由于 $s^{*} \in \mathrm{Hom}(Y, X)$ 和 $a^{\dagger} \in \mathrm{Hom}(E(X), E(Y))$，这个表达式的右手边再次处于 $I(E, \tilde{E})$ 中。因此 $I(E, \tilde{E})$ 是在★下稳定的，而★下降到 $A(E, \tilde{E})$ 上

的一个反线性对合。在 $A_0(E,\tilde{E})$ 里我们具有：

$$
\begin{aligned}
&([X,s]\cdot[Y,t])^{\star}=[X\otimes Y,\ d^{\tilde{E}}_{X,Y}\circ s\otimes t\circ(d^{E}_{X,Y})^{-1}]^{\star}\\
&=[X\otimes Y,\ (d^{E}_{X,Y}{}^{\dagger})^{-1}\circ s\otimes t\circ(d^{E}_{X,Y})^{-1}]^{\star}\\
&=[X\otimes Y,\ (d^{E}_{X,Y}{}^{\dagger})^{-1}\circ s^{\dagger}\otimes t^{\dagger}\circ(d^{E}_{X,Y})^{-1}]\\
&=[X\otimes Y,\ (d^{\tilde{E}}_{X,Y})\circ s^{\dagger}\otimes t^{\dagger}\circ(d^{E}_{X,Y})^{-1}]\\
&=[X,s]^{\star}\cdot[Y,t]^{\star}
\end{aligned}
$$

跟 $A(E,\tilde{E})$ 的对易性结合，这意味着★是反可乘性的。回顾存在一个同构 δ：$\bigoplus_{i\in I}\mathrm{End}E(X_i)\to A(E,\tilde{E})$，使得 $\|\delta((a_i)_{i\in I})\|=\sup_i\|a_i\|$，其中 $\|\cdot\|$ 是在 B3 小节定义的范数。根据★的定义我们具有 $\delta((a_i))^{\star}=\delta((a_i^{\dagger}))$，意味着 $\|a^{\star}a\|=\|a\|^2$。因此 $(A(E,\tilde{E}),\ \star,\ \|\cdot\|)$ 是一个前 C^* 代数。

注意对合★跟 B3 小节定义的那个没有任何关系！

命题 406。设 $\mathcal{C}$ 是个 STC^*，并且 E：$\mathcal{C}\to\mathrm{Vect}_{\mathbb{C}}$ 是个对称纤维函子。由于如上定义的 $\tilde{E}$，存在一个自然单项同构 α：$E\to\tilde{E}$，其组分 α_X 是正的，即对所有非零 $u\in E(X)$ 都有 $\langle u,\ \alpha_X u\rangle^0_X>0$。

证明。跟 B4 小节一样，$A(E,\tilde{E})$ 的范数完备化 $\mathcal{A}(E,\tilde{E})$，是一个对易的幺正 C^* 代数，从而承认一个 * 特征 ϕ：$\mathcal{A}(E,\tilde{E})$。限制 $A(E,\tilde{E})$，命题 398 提供一个单项自然同构 α：$E\to\tilde{E}$。但是我们知道的更多，特征 ϕ 是正的，即对于所有 $a\neq0$ 有 $\phi(a^{\star}a)>0$。由于 $a=[X,s]$ 以及考虑到式(59)，我们有：

$$
\begin{aligned}
\phi(a^{\star}a)&=\phi([X,\ s^{\dagger}s])=Tr_{E(X)}(s^{\dagger}s\,\alpha_X)=Tr_{E(X)}(s\,\alpha_X s^{\dagger})\\
&=\sum_i\langle e_i,\ s\,\alpha_X s^{\dagger}e_i\rangle^0_X=\sum_i\langle s^{\dagger}e_i,\ \alpha_X s^{\dagger}e_i\rangle^0_X
\end{aligned}
$$

其中 $\{e_i\}$ 是关于 $\langle\cdot,\ \cdot\rangle^0_X$ 正交的任何 $E(X)$ 的基。这对所有 $a=[X,s]\in\mathcal{A}(E,\tilde{E})$ 都是正的，当且仅当对于所有 $u\in E(X)\langle u,\ \alpha_X u\rangle^0_X>0$。

现在我们开始证明这一节的主要结果，这是更加具体的定理 403 方案。

定理 407。设 $\mathcal{C}$ 是一个偶 STC^*，并且 E：$\mathcal{C}\to\mathrm{Vect}_{\mathbb{C}}$ 是一个对称纤维函子。那么存在空间 $E(X)$，$X\in\mathcal{C}$ 上的希尔伯特空间结构，即正定内积 $\langle\cdot,\ \cdot\rangle_X$，

使得 X $\mapsto$(E(X)$\langle\cdot,\ \cdot\rangle_X$)是一个 * 保持对称纤维函子$\mathcal{C}\to\mathcal{H}$。

选择在空间 $E(X)$，$X\in\mathcal{C}$上的非简并内积$\langle\cdot,\ \cdot\rangle_X^0$，由于 $E(\mathbf{1})$是一维的，并且通过 e^E1 扩张，其中 $1\in\mathbb{C}=\mathbf{1}_{\mathrm{Vect}_{\mathbb{C}}}$，我们就能够通过$\langle ae^E1,\ be^E1\rangle_1^0=\bar{a}b$ 定义$\langle\cdot,\ \cdot\rangle_1^0$，跟接下来会假定的一样。设 $\bar{E}$ 和 $\alpha\in\mathrm{Nat}_{\otimes}(E,\ \bar{E})$如上所述。定义在空间$E(X)$上的新内积$\langle\cdot,\ \cdot\rangle_X$为：

$$\langle v,\ u\rangle_X=\langle v,\ \alpha_X u\rangle_X^0$$

(α_X)的自然性：

$$\alpha_Y\circ E(s)=\bar{E}(s)\circ\alpha_X=E(s^*)^\dagger\circ\alpha_X\quad\forall s:\ X\to Y$$

这意味着对所有 s：$X\to Y$，$u\in E(X)$，$v\in E(Y)$都有：

$$\langle v,\ E(s)u\rangle_Y=\langle v,\ \alpha_Y E(s)u\rangle_Y^0=\langle v,\ E(s^*)^\dagger\alpha_X u\rangle_Y^0$$

$$=\langle E(s^*)v,\ \alpha_X u\rangle_X^0=\langle E(s^*)v,\ u\rangle_X$$

这跟 $E(s^*)=E(s)^*$一样，其中现在 $E(s)^*$代表 $E(s)$关于内积$\langle\cdot,\ \cdot\rangle$的伴随。

因此函子 $X\mapsto(E(X),\ \langle\cdot,\ \cdot\rangle_X)$是 * 保持的。新内积$\langle\cdot,\ \cdot\rangle_X$是非简并的，因为 α^X 是可逆的，并且正性质$\langle u,\ \alpha_X u\rangle_X^0>0$ 对于 $u\neq0$，意味着$(E(X),\ \langle\cdot,\ \cdot\rangle_X)$是希尔伯特空间。自然同构 α：$E\to\bar{E}$ 的单项

$$\alpha_{X\otimes Y}\circ d_{X,Y}^E=d_{X,Y}^{\bar{E}}\circ\alpha_X\otimes\alpha_Y=((d_{X,Y}^E)^\dagger)^{-1}\circ\alpha_X\otimes\alpha_Y\quad\forall X,\ Y$$

等价于

$$\alpha_X\otimes\alpha_Y=(d_{X,Y}^E)^\dagger\circ\alpha_{X\otimes Y}\circ d_{X,Y}^E \tag{60}$$

借助于这一点我们有：

$$\langle d_{X,Y}^E(u'\otimes v'),\ d_{X,Y}^E(u\otimes v)\rangle_{X\otimes Y}=\langle d_{X,Y}^E(u'\otimes v'),\ \alpha_{X\otimes Y}\circ d_{X,Y}^E(u\otimes v)\rangle_{X\otimes Y}^0$$

$$=\langle(u'\otimes v'),\ (d_{X,Y}^E)^\dagger\circ\alpha_{X\otimes Y}\circ d_{X,Y}^E(u\otimes v)\rangle_{X\otimes Y}^0$$

$$=\langle(u'\otimes v'),\ (\alpha_X\otimes\alpha_Y)(u\otimes v)\rangle_{X\otimes Y}^0$$

$$=\langle u',\ \alpha_X u\rangle_X^0\langle v',\ \alpha_Y v\rangle_Y^0=\langle u',\ u\rangle_X\langle v',\ v\rangle_Y$$

因此同构 $d_{X,Y}^E$：$E(X)\otimes E(Y)\to E(X\otimes Y)$关于内积$\langle\cdot,\ \cdot\rangle$是幺正的。

现在，d^E 和 e^E 的兼容性(57)意味着 $d_{1,1}^E\circ e^E1\otimes e^E1=e^E1$，从而通过我们对内积$\langle\cdot,\ \cdot\rangle_1^0$ 的使用可以得到：

$$\langle d_{1,1}^{E}(ae^{E}1\otimes be^{E}1),\ d_{1,1}^{E}(ce^{E}1\otimes de^{E}1)\rangle_{1\otimes 1}^{0}=\langle abe^{E}1,\ cde^{E}1\rangle_{1}^{0}$$
$$=\overline{ab}cd=\langle ae^{E}1,\ ce^{E}1\rangle_{1}^{0}\langle be^{E}1,\ de^{E}1\rangle_{1}^{0}$$

这意味着 $d_{1,1}^{E}$：$E(1)\otimes E(1)-E(1)$是关于内积$\langle\cdot,\ \cdot\rangle_{1}^{0}$幺正的，取式(60)里的 $X=Y=1$ 并且使用 $\alpha_1=\lambda\mathrm{id}_{E(1)}$，我们得到 $\lambda^2=\lambda$。由于 α_1 是可逆的，我们具有 $\lambda=1$，因此 $\alpha_1=\mathrm{id}_{E(1)}$从而$\langle\cdot,\ \cdot\rangle_1=\langle\cdot,\ \cdot\rangle_1^0$。现在：

$$\langle e^{E}1,\ e^{E}1\rangle_{1}=\langle e^{E}1,\ \alpha_1 e^{E}u\rangle_{1}^{0}=\langle e^{E}1,\ e^{E}1\rangle_{1}^{0}=1=\langle 1,\ 1\rangle_{\mathbb{C}}$$

因此$(e^{E})^{*}e^{E}=\mathrm{id}_{\mathbb{C}}$。根据所涉及空间的一维性，我们也具有 $e^{E}(e^{E})^{*}=\mathrm{id}_{E(1)}$，因此 e^{E}：$1\to E(1)$关于新内积$\langle\cdot,\ \cdot\rangle$是幺正的。

B7 约化到有限生成范畴

定义 408。一个可加性张量范畴$\mathcal{C}$是一个有限生成的，如果存在一个对象 $Z\in\mathcal{C}$,使得任何对象 $X\in\mathcal{C}$是 Z 的下面某些张量幂的直和，$Z^{\otimes n}=\underbrace{Z\otimes\cdots\otimes Z}_{n\text{个因子}}$，$n\in\mathbb{N}$。

引理 409。设$\mathcal{C}$是一个 TC^{*}，那么$\mathcal{C}$的有限生成张量子范畴形成一个直接系统，并且$\mathcal{C}$是后者的归纳极限：

$$\mathcal{C}\cong\varinjlim_{\iota\in I}\mathcal{C}_{\iota}$$

证明。考虑$\mathcal{C}$的全部完全张量子范畴，由于$\mathcal{C}$是本质上小，这样的子范畴的等价类形成一个集合，嵌入偏序的集合。如果$\mathcal{C}_1$，$\mathcal{C}_2\subset\mathcal{C}$是有限生成的，通过对象 X_1，X_2 表示，那么包含$\mathcal{C}_1$和$\mathcal{C}_2$的最小张量子范畴是由 $X_1\oplus X_2$ 生成的，因此我们具有一个直接的系统。显然存在一个完全而忠实的张量算子$\varinjlim_{\iota\in I}\mathcal{C}_{\iota}\to\mathcal{C}$。由于任何对象 X 是包含在有限生成元子范畴(比如由生成的那个)，这个函子本质上是满射的，从而有一个等价的范畴，参看[Mac Lane，1998]，张量范畴的事实参看[Saavedra Rivano，1972]。

评论 410。

1. 考虑有限生成范畴的原因是，这些范畴的纤维函子的存在问题能够通过使用强大的纯代数方法解决。而一般情况能够被约化成用引理 409 有限生成的那个。

2. 注意到我们并不要求生成元 Z 成为不可约化的，因此如果我们仅仅事先

知道$\mathcal{C}$是由对象的有限集Z_1，…，Z_r生成，直和$Z=\bigoplus_i Z_i$就会是$\mathcal{C}$的一个(可约化的)产生元，这是为什么只有一个单产生对象出现在定义里的原因。

3. 如果G是一个紧致群，范畴$\mathrm{Rep}_f G$是有限生成的当且仅当G是一个李群。证明：⇐是紧致李群的著名表示理论；⇒是众所周知G的有限维表示分离了G的元。因此，如果(H,π)是$\mathrm{Rep}_f G$的产生元，显然π必定是忠实的。因此G同构于紧致李群$\mathcal{U}(H)$的一个闭合子群，因此它是一个李群。

4. 引理409里的下标集合I能够作为可数的，当且仅当$\mathcal{C}$具有不可约化对象的可数的很多同构类。范畴$\mathrm{Rep}_f G$，其中G是一个紧致群，具有这个性质，当且仅当G是第二可数且等价可度量的。

在B8—B11小节我们将证明下面的结果。

定理411。一个有限产生的偶STC^*承认一个对称纤维函子E：$\mathcal{C}\to\mathrm{Vect}_{\mathbb{C}}$。

定理382的证明。根据引理409，我们能够表示$\mathcal{C}$为有限产生范畴的一个归纳极限$\lim\limits_{\overrightarrow{\iota\in I}}\mathcal{C}_i$，现在定理411提供我们对称纤维函子$E_i$：$\mathcal{C}_i\to\mathrm{Vect}_{\mathbb{C}}$，$i\in I$，并且定理407把后者变成*保持对称纤维函子$E_i$：$\mathcal{C}_i\to\mathcal{H}$。根据定理377，我们获得具有在$E_i(X)$，$X\in\mathcal{C}_i$上的$\pi_i$，$X$紧致群$G_i=\mathrm{Nat}_{\otimes}E_i$(事实上是评论410.3里的紧致李群)，使得函子$F_i$：$\mathcal{C}_i\to\mathrm{Rep}_f G_i$，$X\mapsto E_i(X)$，$\pi_{i,X}$都是等价的。现在设$i\leqslant j$，蕴含着$\mathcal{C}_i$是$\mathcal{C}_j$的一个完全子范畴，那么$E_j\upharpoonright\mathcal{C}_i$是一个$\mathcal{C}_i$的纤维函子，从而定理373意味着幺正自然同构$\alpha^{i,j}$：$F_1\to F_2\upharpoonright\mathcal{C}_i$的存在。(注意$\alpha^{i,j}$不是唯一的!)现在，通过定义任何$g\in G_2$为$E_2$的单项自然自同构的幺正($g_X\in\mathcal{U}(E_2(X)))_{X\in\mathcal{C}_1}$的族。对于任何$X\in\mathcal{C}_1$，定义$h_X:=\alpha_X^{i,j}\circ g_X\circ(\alpha_X^{i,j})^*$，我们看到族$(h_X\in\mathcal{U}(E_1(X)))_{X\in\mathcal{C}_1}$是$E_1$的一个幺正单项自然子同构，即$G_1$的一个元素。按这种方式，我们得到一个映射$\beta^{i,j}$：$G_j\to G_i$，它显然是群同态和连续的。根据舒尔(Schur)的引理，单位元的$\alpha_X^{i,j}$对不可约化X的一个相位是唯一的。因此对于这样的X，$\beta_X^{i,j}$是独立于已选择的$\alpha^{i,j}$，从而$\beta^{i,j}$是唯一决定的。按照$\mathrm{Rep}_f G$的完全张量子范畴跟商G/N之间的伽罗瓦对应来看，它也是满射的，其中$N\subset G$是一个封闭的正规子群。现在逆极限

$$G=\lim_{\overleftarrow{i\in I}}G_i=\{(g_i\in G_i)_{i\in I}\mid\beta^{i,j}(g_j)=g_i\text{只要 }i\leqslant j\}$$

是一个紧致群，具有对所有$i\in I$的明显满射同态γ_i：$G\to G_i$。现在我们定义一

个函子E：$\mathcal{C}\to \mathrm{Rep}_f G$如下：对于任何$X\in \mathcal{C}$挑选一个$i\in I$，使得$X\in \mathcal{C}_i$并且定义$F(X)=(E_i(X),\ \pi_i(X)\circ \gamma_i)$。显然这是一个$\mathrm{Rep}_f G$里的对象，并且它的等距类是独立于$i\in I$的选择。按照这种方式我们得到一个从$\mathcal{C}=\lim_{\to}\mathcal{C}_i$到$\mathrm{Rep}_f G\cong \lim_{\to}\mathrm{Rep}_f G_i$的函子，限制到等价$\mathcal{C}_i\to \mathrm{Rep}_f G_i$。因此$E$是完全的和忠实的。最后，$E$本质上是满射的，因为$G=\lim_{\leftarrow}G_i$的任何有限维表征是通过群$G_i$之一来因子化的。

评论 412。按照评论 410.3 的观点，上述证明也表明每一个紧致群是一个紧致李群的逆极限。

B8 来自单项的纤维函子

我们证明定理 411 的策略本质上是德利涅方法之一[Deligne，1990]，不过用相当基本的对易范畴代数替代了对称代数范畴里的代数几何。已经存在几个这一证明的描述[Bichon，1998；Rosenberg(罗森堡)，2000；Hái(海)，2002]，我们发现其中[Bichon，1998]最为有用，也可参看[Bichon，ND]，当然我们会给出比这些文献更为详细的证明，并且我们某些方面更加简化。

下面的结果显然证明在 A.6 小节里引入的概念到我们证明定理 411 的目标之间的关联。

命题 413。设$\mathcal{C}$是一个TC^*，并且$\hat{\mathcal{C}}$是一个包含作为完全张量子范畴$\mathcal{C}$的$\mathbb{C}$线性严格张量范畴。设$(Q,\ m,\ \eta)$是一个$\hat{\mathcal{C}}$里的单项并且满足：

(i) $\dim \mathrm{Hom}_{\hat{\mathcal{C}}}(\mathbf{1},\ Q)=1$(即$\mathrm{Hom}_{\hat{\mathcal{C}}}(\mathbf{1},\ Q)=\mathbb{C}_\eta$)；

(ii)对于任何$X\in \mathcal{C}$，存在$n(X)\in \mathbb{Z}_+$使得任何$X\neq 0$的时候$n(X)\neq 0$，并且有Q模的一个同构α_X：$(Q\otimes X,\ m\otimes \mathrm{id}_X)\to n(X)\cdot(Q,\ m)$。

那么函子E：$\mathcal{C}\to \mathrm{Vect}_{\mathbb{C}}$定义为：

$$E:\ \mathcal{C}\to \mathcal{H},\ X\mapsto \mathrm{Hom}_{\hat{\mathcal{C}}}(\mathbf{1},\ Q\otimes X)$$

结合：

$$E(s)\phi=\mathrm{id}_Q\otimes s\circ \phi,\ s:\ X\to Y,\ \phi\in \mathrm{Hom}(\mathbf{1},\ Q\otimes X) \tag{61}$$

这是一个忠实(强)张量函子，并且满足$\dim_{\mathbb{C}} E(X)=n(X)$。

如果$\hat{\mathcal{C}}$有一个关于$(Q,\ m,\ \eta)$是对易的对称c，那么E是关于$\mathrm{Vect}\mathbb{C}$的对

称 $\sum$ 的对称单项，即 $E(c_{X,Y}) = \sum_{E(X),E(Y)}$。

证明。我们具有 $E(X) = \mathrm{Hom}(\mathbf{1},\ Q\otimes X) \cong \mathrm{Hom}(\mathbf{1},\ n(X)Q) \cong d(X)\mathrm{Hom}(\mathbf{1},\ Q) \cong \mathbb{C}^{n(X)}$，因此 $E(X)$ 是维度 $n(X)$ 的一个矢量空间。由于对任何非零 $X\in\mathcal{C}$ 的 $E(X)\neq 0$，因此函子 E 是忠实的。

为了理解 E 是单项的，首先来看根据(ii)我们有 $E(\mathbf{1}) = \mathrm{Hom}(\mathbf{1},\ Q) = \mathbb{C}_{\eta}$。因此，存在一个根据 $c\mapsto c\eta$ 定义的正则同构 e：$\mathbb{C} = \mathbf{1}_{\mathrm{Vect}_{\mathbb{C}}}\to E(\mathbf{1}) = \mathrm{Hom}(\mathbf{1},\ Q)$。下一步我们来定义态射为：

$[d^{E}_{X,Y}:\ E(X)\otimes E(Y)\to E(X\otimes Y),\ \phi\otimes\psi\mapsto m\otimes\mathrm{id}_{X\otimes Y}\circ \mathrm{id}_{Q}\otimes\phi\otimes\mathrm{id}_{Y}\circ\psi]$

根据映射 $E(s)$：$E(X)\to E(Y)$ 的定义(61)，显然族($d^{E}_{X,Y}$)关于两个论证都是自然的。来自张量函子要求的方程

$$d^{E}_{X_1\otimes X_2,X_3}\circ d^{E}_{X_1,X_2}\otimes\mathrm{id}_{E(X_3)} = d^{E}_{X_1,X_2\otimes X_3}\circ \mathrm{id}_{E_1}\otimes d^{E}_{X_2,X_3}\quad \forall X_1,\ X_2,\ X_3\in\mathcal{C}$$

是 m 的可结合性的直接结果。该证明留作练习。

$(E,\ (d_{X,Y}),\ e)$ 满足统一公理几乎是显然的，第一个条件根据的是：

$$d_{X,\mathbf{1}}(\mathrm{id}_{E(X)}\otimes e)\phi = d_{X,\mathbf{1}}(\phi\otimes\eta) = m\otimes\mathrm{id}_{X}\circ\mathrm{id}_{Q}\otimes\phi\circ\eta = \phi$$

并且第二条类似得证。

到此为止，我们已经证明 E 是一个弱张量函子，为的是 e：$\mathbf{1}_{\mathcal{H}}\to E(\mathbf{1}_{\mathcal{C}})$ 是个同构。为了得到 E 是一个(强)张量函子，接下来要证明态射 $d^{E}_{X,Y}$ 都是同构的。设 $X,\ Y\in\mathcal{C}$，我们考虑双线性映射：

$$\gamma_{X,Y}:\ \mathrm{Hom}_{Q}(Q,\ Q\otimes X)\boxtimes\mathrm{Hom}_{Q}(Q,\ Q\otimes Y)\to\mathrm{Hom}_{Q}(Q,\ Q\otimes X\otimes Y)$$
$$s\boxtimes t\mapsto s\otimes\mathrm{id}_{Y}\circ t$$

我们对 $\mathrm{Vect}_{\mathbb{C}}$ 的张量积写成⊠而不是 $\otimes_{\mathbb{C}}$，为的是避免跟 $Q-\mathrm{Mod}$ 里的张量积混淆。根据 2，我们具有 Q 模态射 s_i：$Q\to Q\otimes X$，s'_i：$Q\otimes X\to Q$，对 $i=1,\ \cdots,\ n(X)$，满足 $s'_i\circ s_j = \delta_{ij}\mathrm{id}_{Q}$，和 $\sum_i s_i\circ s'_i = \mathrm{id}_{Q\otimes X}$，以及对于 Y 代替 X 时类似的态射 $t_i,\ t'_i,\ i=1,\ \cdots,\ n(Y)$。那么 $\gamma_{ij} = \gamma_{X,Y}(s_i\otimes t_j)$ 是线性无关的，因为它们满足具有 $\gamma'_{i'j'} = t'_j\circ s'_i\otimes\mathrm{id}_{Y}$ 的 $\gamma'_{i'j'}\circ\gamma_{ij} = \delta_{i'i}\delta_{j'j}\mathrm{id}_{Q}$。$\gamma_{X,Y}$ 的双射性现在源自于 $\gamma_{X,Y}$ 的定义域和值域都含有 $n(X)n(Y)$ 的事实。求助于同构 δ_X：$\mathrm{Hom}_{Q}(Q,\ Q\otimes X)\mapsto\mathrm{Hom}(\mathbf{1},\ Q\otimes X)$，人们容易证明：

$$d^{E}_{X,Y} = \delta_{X\otimes Y}\circ\gamma_{X,Y}\circ\delta_X^{-1}\boxtimes\delta_Y^{-1}$$

这意味着 $d^E_{X,Y}$是对于任何 X，$Y\in\mathcal{C}$的一个同构。

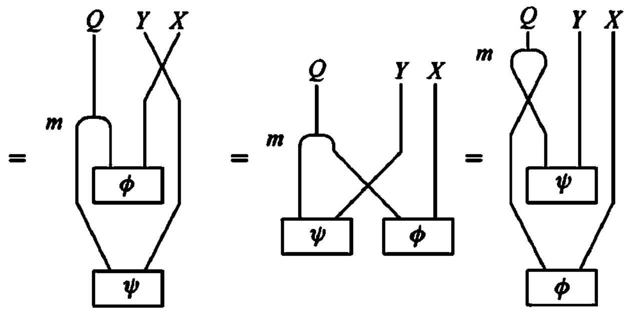
我们现在假设$\hat{\mathcal{C}}$具有一个对称 c 并且$(Q,\ m,\ \eta)$是对易的，为了证明 E 是一个对称张量函子，我们必须证明对于所有 X，$Y\in\mathcal{C}$都有：

$$E(c_{X,Y})\circ d^E_{X,Y} = \sum{}_{E(X),E(Y)}\circ d^E_{Y,X}$$

设 $\phi\in E(X)$，$\psi\in E(Y)$。

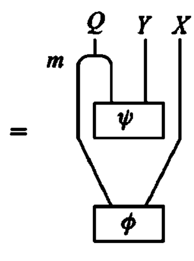
根据 E 的定义我们得到：

$$(E(c_{X,Y})\circ d^E_{X,Y})(\phi\otimes\psi) = \mathrm{id}_Q\otimes c_{X,Y}\circ m\otimes\mathrm{id}_{X\otimes Y}\circ\mathrm{id}_Q\otimes\phi\otimes\mathrm{id}_Y\circ\psi$$



另一方面，有：

$$(d^E_{Y,X}\circ c_{E(X),E(Y)})(\phi\otimes\psi) = (d^E_{Y,X}\circ\sum{}_{E(X),E(Y)})(\phi\otimes\psi) = d^E_{Y,X}(\psi\otimes\phi)$$



如果 m 是可对易的，即 $m=m\circ c_{Q,Q}$，那么这两个表述是一致的，证毕。

评论 414。1. 命题里的性质(ii)被称为“吸收性质”。

2. 命题 413 里的条件实际上对于一个纤维函子的存在是必要的！假设一个张量 * 范畴$\mathcal{C}$承认一个 * 保持纤维函子 E：$\mathcal{C}\to\mathcal{H}$。根据[Müger *et al.*，2004]，其中评论和扩展了瓦伦纳维奇(Woronowicz)、山上(Yamagami)及其他人的工作，存在一个离散代数量子群$(A,\ \Delta)$，使得$\mathcal{C}\simeq\mathrm{Rep}_f(A,\ \Delta)$。在[Müger and Tuset(缪格和图塞特)，2006]里证明了取$\hat{\mathcal{C}}\simeq\mathrm{Rep}(A,\ \Delta)$(即任意维度的表示)，

以及$Q=\pi_l$，存在一个单项(Q, m, η)满足命题 413 的条件。即人们能够取$Q=\pi_l$，剩下的规则表象。在[Müger and Tuset, 2006]里它已经证明：(i) $\dim \mathrm{Hom}(\pi_0, \pi_l)=1$，即存在一个非零态射$\eta$：$\pi_0 \to \pi_l$，是唯一可正规化的；(ii) π_l具有所要求的吸收性质；(iii)存在一个态射m：$\pi_l \to \pi_l$，使得$(Q=\pi_l, m, \eta)$是一个单项。

3. 在上述情况下，剩下的规则表象π_l处于$\mathrm{Rep}_f(A, \Delta)$里，当且仅当$A$是有限维的。这就指出范畴$\mathcal{C}$通常太小不足以包含所希望性质的单项。事实上，假设我们能够取$\hat{\mathcal{C}}=\mathcal{C}$，那么对于任何不可约化的$X \in \mathcal{C}$，我们具有$\dim \mathrm{Hom}(X, Q)=\dim \mathrm{Hom}(\mathbf{1}, Q \otimes \overline{X})=n(\overline{X})>0$。因此$Q$包含所有作为直和项的不可约化对象。由于$\mathcal{C}$里的任何对象是简单对象的有限直和，$\hat{\mathcal{C}}=\mathcal{C}$是可能的，仅当$\mathcal{C}$是只具有简单对象的有限多同构类。事实上，即便在这样的情况下，我们对(Q, m, η)的建构也要求更大范畴$\hat{\mathcal{C}}$的使用。正是这里 A7 小节的范畴 Ind $\mathcal{C}$起作用了。

由于我们已经把建构一个纤维函子的困难，约化成有限生成张量范畴的情况，我们要一个适合那个情况的上述结果的方案。

推论 415。设$\mathcal{C}$是一个具有单项产生元$Z \in \mathcal{C}$的TC^*，并且设$\hat{\mathcal{C}}$是一个包含作为一个完整张量子范畴$\mathcal{C}$的$\mathbb{C}$线性严格张量范畴。如果(Q, m, η)是一个$\hat{\mathcal{C}}$里的单项，满足：

(i) $\dim \mathrm{Hom}_{\hat{\mathcal{C}}}(\mathbf{1}, Q)=1$；

(ii)存在$d \in \mathbb{N}$和一个Q模的同构α_Z：$(Q \otimes Z, m \otimes \mathrm{id}_Z) \to d \cdot (Q, m)$。

那么得到命题 413 里的假设(iii)。因此E：$X \to \mathrm{Hom}_{\hat{\mathcal{C}}}(\mathbf{1}, Q \otimes X)$是一个纤维函子。

证明。如果$X \in \mathcal{C}$，存在$n \in \mathbb{N}$使得$X < Z^{\otimes n}$。具体而言，存在态射u：$X \to Z^{\otimes n}$和v：$Z^{\otimes n} \to X$，使得$v \circ u=\mathrm{id}_X$。那么态射$\tilde{u}=\mathrm{id}_Q \otimes u$：$Q \otimes X \to Q \otimes Z^{\otimes n}$和$\tilde{v}=\mathrm{id}_Q \otimes v$：$Q \otimes Z^{\otimes n} \to Q \otimes X$是$Q$模的态射。因此$Q$模$(Q \otimes X, m \otimes \mathrm{id}_X)$是$(Q \otimes Z^{\otimes n}, m \otimes \mathrm{id}_{Z^{\otimes n}})$的直和项。根据假设，后者是同构于$(Q, m)$复制$d^n$的直和。根据引理 358 和假设(i)，$\mathrm{End}_Q((Q, m)) \cong \mathbb{C}$，因此$(Q, m) \in Q-\mathrm{Mod}$是不

可约化的。因此$d^n \cdot (Q, m)$的直和$(Q\otimes X, m\otimes \mathrm{id}_X)$是一个$(Q, m)$的 r 复制的直和，其中 $r\leqslant d^m$ 并且任何 $X\neq 0$ 时候 $r\neq 0$。因此命题 413 里的假设(ii)得证。

按照推论 415，证明定理 411 等于寻找一个包含作为完整子范畴$\mathcal{C}$的对称张量范畴$\hat{\mathcal{C}}$，以及一个$\hat{\mathcal{C}}$里的对易单项(Q, m, η)，使得 $\dim \mathrm{Hom}(\mathbf{1}, Q)=1$ 和 $Q\otimes Z\cong d\otimes Q$，作为$\mathcal{C}$的一个适当单项产生之 Z 的 Q 模。这将会在 B11 小节里完成，建立在通过范畴$\mathcal{C}$的置换对称的分析基础上。

B9 对称群作用、行列式和维度的完整性

我们现在开始讨论对称群 P_n，$n\in\mathbb{N}$ 的某些表示，它们主要存在于具有幺正对称的张量 * 范畴里。众所周知 n 标记的对称群 P_n 具有表示：

$$P_n=(\sigma_1, \cdots, \sigma_{n-1} \mid |i-j| \geqslant 2\Rightarrow \sigma_i\sigma_j=\sigma_j\sigma_i,$$
$$\sigma_i\sigma_{i+1}\sigma_i=\sigma_{i+1}\sigma_i\sigma_{i+1}\ \forall_i\in\{1, \cdots, n-1\}, \sigma_i^2=1\ \forall_i)$$

由于$\mathcal{C}$是严格的，我们可以定义张量幂 $X^{\otimes n}$，$n\in\mathbb{N}$，按照任何 $X\in\mathcal{C}$的明显方法。我们对任何 $X\in\mathcal{C}$设定 $X^{\otimes 0}=\mathbf{1}$。

引理 416。设$\mathcal{C}$是一个 STC^*，设 $X\in\mathcal{C}$和 $n\in\mathbb{N}$，那么：

$$\prod\nolimits_n^X: \sigma_i\mapsto \mathrm{id}_{X^{\otimes i-1}}\otimes c_{X,X}\otimes \mathrm{id}_{X^{\otimes n-i-1}}$$

唯一决定一个同态 $\prod_n^X$，从群 P_n 到 $\mathrm{End}X^{\otimes n}$的幺正群。

证明。显然 $\prod_n^X(\sigma_i)$和 $\prod_n^X(\sigma_j)$有交换性，如果 $|i-j|\geqslant 2$。$\prod_n^X(\sigma_i)^2=\mathrm{id}_{X^{\otimes n}}$是同样明显的。最后：

$$\prod\nolimits_n^X(\sigma_i)\circ\prod\nolimits_n^X(\sigma_{i+1})\circ\prod\nolimits_n^X(\sigma_i)=\prod\nolimits_n^X(\sigma_{i+1})\circ\prod\nolimits_n^X(\sigma_i)\circ\prod\nolimits_n^X(\sigma_{i+1})$$

源自于满足对称 c 的杨—巴斯克(Yang-Baxter)方程。

评论 417。降低关系 $\sigma_i^2=1$，跟上面一样的公式把阿廷(Artin)辫子群 B_n 的同态定义成 $\mathrm{End}X^{\otimes n}$。不过，下面的考虑没有一个是辫子情况里的已知类似物。

回顾存在一个同态符号：$P_n\to\{1, -1\}$，符号映射。

引理 418。设$\mathcal{C}$是一个 STC^*，对于任何 $X\in\mathcal{C}$，我们都通过 $S_0^X=A_0^X=\mathrm{id}_1$ 在 $\mathrm{End}X^{\otimes 0}=\mathrm{End}\mathbf{1}$ 里定义正交投影，对于任何 $n\in\mathbb{N}$，态射：

$$S_n^X = \frac{1}{n!} \sum_{\sigma \in P_n} \prod_n^X (\sigma)$$

$$A_n^X = \frac{1}{n!} \sum_{\sigma \in P_n} \mathrm{sgn}(\sigma) \prod_n^X (\sigma)$$

满足：

$$\prod_n^X (\sigma) \circ S_n^X = S_n^X \circ \prod_n^X (\sigma) = S_n^X$$

$$\prod_n^X (\sigma) \circ A_n^X = A_n^X \circ \prod_n^X (\sigma) = \mathrm{sgn}(\sigma) A_n^X$$

对于所有的 $\sigma \in P_n$，从而是在 * 代数 $\mathrm{End}X^{\otimes n}$ 里的正交投影。

证明。直接计算可得。

定义 419。对应于幂等元 S_n^X 和 A_n^X 的 $X^{\otimes n}$（对等距定义的）子对象，是通过 $S_n(X)$ 和 $A_n(X)$ 分别表示的。

下面的命题在[Doplicher and Roberts, 1989]和[Deligne, 1990]都有证明。

命题 420。设 $\mathcal{C}$ 是一个偶 STC^*，对于任何 $X \in \mathcal{C}$ 我们都有：

$$Tr_{X^{\otimes n}} A_n^X = \frac{d(X)(d(X)-1)(d(X)-2)\cdots(d(X)-n+1)}{n!} \quad \forall n \in \mathbb{N} \qquad (62)$$

证明。（概略）$\mathcal{C}$ 是偶的这个事实的关键应用，即对所有 $X \in \mathcal{C}$ 有 $\Theta(X) = \mathrm{id}_X$，人们就能证明：

$$Tr_{X^{\otimes n}} \prod_n^X (\sigma) = d(X)^{\#\sigma} \quad \forall X \in \mathcal{C}, \ \sigma \in P_n$$

其中 $\#\sigma$ 是置换 σ 分解成的环数。熟悉三角图的读者会发现这个公式几乎是显然的：扭 $\Theta(X)$ 的平凡性意味着在第一赖德迈斯特（Reidemeister）移动下有不变性，因此置换 σ 的闭包等价于 $\#\sigma$ 环，每一个环给出一个因子 $d(X)$。现在这个结果直接源自于 A_n^X 和下述公式的定义

$$\sum_{\sigma \in P_n} \mathrm{sgn}(\sigma) z^{\#\sigma} = z(z-1)(z-2)\cdots(z-n+1)$$

适用于所有 $n \in \mathbb{N}$ 和 $z \in \mathbb{C}$，正如人们通过对 n 的归纳所能证明的。

推论 421。在一个 STC^* 里，我们对于任何非零 $X \in \mathcal{C}$ 都有 $d(X) \in \mathbb{N}$。

证明。首先假设 $\mathcal{C}$ 是偶的，并且设 $X \in \mathcal{C}$。由于 $\mathcal{C}$ 有子对象而存在一个对象 $A_n(X) \in \mathcal{C}$ 和态射 $s: A_n(X) \to X^{\otimes n}$，使得 $s^* \circ s = \mathrm{id}_{A_n(X)}$ 和 $s \circ s^* = A_n^X$。那么根据命题 339 里的第一部分和第二部分，我们得到：

$$Tr_{X^{\otimes n}}A_n^X = Tr_{X^{\otimes n}}(s \circ s^*) = Tr_{A_n(X)}(s^* \circ s) = Tr_{A_n(X)}\mathrm{id}_{A_n(X)} = d(A_n(X))$$

由于 * 范畴里的任何对象的维度是非负的，因此我们得出对于所有 $n \in \mathbb{N}$ 都有 $Tr_{X^{\otimes n}}A_n^X \geqslant 0$。从公式(62)的右边，对于 $Tr_{X^{\otimes n}}A_n^X$ 我们看到，$Tr_{X^{\otimes n}}A_n^X$ 将会变成对某些 $n \in \mathbb{N}$ 是负的，除了 $d(X) \in \mathbb{N}$ 之外。

如果$\mathcal{C}$是奇的，上述论证给出在玻色子化的范畴$\tilde{\mathcal{C}}$里的维度的整体性。由于范畴维度是独立于辫子的，我们具有 $d_{\mathcal{C}}(X) = d_{\tilde{\mathcal{C}}}(X)$，证毕。

设$\mathcal{C}$是一个 STC^* 和 $X \in \mathcal{C}$是非零的，并且设 $d = d(X) \in \mathbb{N}$。考虑 $X^{\otimes d}$的子对象 $A_d(X)$，这是在推论 421 的证明里引入的，对应于在引理 418 里定义的正交投影 $A_d^X \in \mathrm{End}X^{\otimes d}$，那么：

$$d(A_d(X)) = Tr_{X^{\otimes d}}A_d^X = \frac{d!}{d!} = 1$$

我们看到 $A_d(X)$是$\mathcal{C}$的一个不可约化的并且可逆的对象（具有逆$\overline{A_d(X)}$）。

定义 422。 $A^{d(X)}(X)$的同构类被称为是 X 的行列式 $\det X$。

引理 423。 设$\mathcal{C}$是一个 STC^* 并且 X，$Y \in \mathcal{C}$，那么：

(i) $\det \overline{X} \cong \overline{\det(X)}$；

(ii) $\det(X \oplus Y) \cong \det X \otimes \det Y$；

(iii) $\det(X \oplus \overline{X}) \cong \mathbf{1}$。

证明。(i)设$(\overline{X}, r, \overline{r})$是 X 的一个标准左逆，通过引理 338 的归纳运用，人们获得对于任意 $n \in \mathbb{N}$ 的 $X^{\otimes n}$的左逆$(\overline{X}^{\otimes n}, r_n, \overline{r}_n)$。现在如果 $\sigma = \sigma_{i_1} \cdots \sigma_{i_t} \in P_n$，人们就能够证明：

$$\prod{}_n^{\overline{X}}(\sigma') = r_n^* \otimes \mathrm{id}_{\overline{X}^{\otimes n}} \circ \mathrm{id}_{\overline{X}^{\otimes n}} \otimes \prod{}_n^X(\sigma) \otimes \mathrm{id}_{\overline{X}^{\otimes n}} \circ \mathrm{id}_{\overline{X}^{\otimes n}} \otimes \overline{r}_n$$

其中 $\sigma' = \sigma_{n-i_t}^{-1} \cdots \sigma_{n-i_1}^{-1}$，特别是 $\mathrm{sgn}\sigma' = \mathrm{sgn}\sigma$，这意味着对于所有 $n \in \mathbb{N}$ 都有：

$$A_n^{\overline{X}} = r_n^* \otimes \mathrm{id}_{\overline{X}^{\otimes n}} \circ \mathrm{id}_{\overline{X}^{\otimes n}} \otimes A_n^X \otimes \mathrm{id}_{\overline{X}^{\otimes n}} \circ \mathrm{id}_{\overline{X}^{\otimes n}} \otimes \overline{r}_n$$

现在该断言源自于引理 337。

(ii)对于任意 $X \in \mathcal{C}$，我们简称为 $d_X = d(X)$ 和 $A^X = A_{d_X}^X \in \mathrm{End}X^{\otimes d_X}$，设 u：$X \to Z$，v：$Y \to Z$ 是实现 $Z \cong X \otimes Y$ 的等距。那么 $X^{\otimes d_X}$是 $Z^{\otimes d_X}$的一个子对象，并且 $Y^{\otimes d_Y}$也类似。根据定义，$\det Z$ 是对应于投影子 $A^Z \in \mathrm{End}Z^{\otimes d_Z}$的 $Z^{\otimes d_Z}$的子对象。另一方面，$\det X \otimes \det Y$ 是对应于 $A^X \otimes A^Y$ 的 $X^{\otimes d_X} \otimes Y^{\otimes d_Y}$的子对象，因此它同构于

$Z^{\otimes d_z}$的子对象，对应的投影子是：

$$u\otimes\cdots\otimes u\otimes v\otimes\cdots\otimes v\circ A^X\otimes A^Y\circ u^*\otimes\cdots\otimes u^*\otimes v^*\otimes\cdots\otimes v^*\in \mathrm{End}Z^{\otimes d_z}$$

其中存在 d_X 的因子 u 和 u^* 以及 d_Y 的因子 v 和 v^*。这等于：

$$\frac{1}{\mathrm{d}_X!\ \mathrm{d}_Y!}\sum_{\substack{\sigma\in P_{d_x}\\ \sigma'\in P_{d_y}}}\mathrm{sgn}(\sigma)\mathrm{sgn}(\sigma')u\otimes\cdots\otimes u\otimes v\otimes\cdots\otimes v$$

$$\circ\prod\nolimits_{d_x}^{X}(\sigma)\otimes\prod\nolimits_{d_y}^{Y}(\sigma')\circ u^*\otimes\cdots\otimes u^*\otimes v^*\otimes\cdots\otimes v^*$$

根据辫子的自然性，这等于：

$$\frac{1}{d_X!\ d_Y!}\sum_{\substack{\sigma\in P_{d_x}\\ \sigma'\in P_{d_y}}}\mathrm{sgn}(\sigma)\mathrm{sgn}(\sigma')\prod\nolimits_{d_x}^{Z}(\sigma)\otimes\prod\nolimits_{d_y}^{Z}(\sigma')\circ p_X\otimes\cdots\otimes p_X\otimes p_Y\otimes\cdots\otimes p_Y$$

其中 $p_X=u\circ u^*$，$p_Y=v\circ v^*$。靠 σ 和 σ'的并列 $\sigma\times\sigma'\in P_{d_x+d_y}=P_{d_z}$，这就变成：

$$\frac{1}{d_X!\ d_Y!}\sum_{\substack{\sigma\in P_{d_x}\\ \sigma'\in P_{d_y}}}\mathrm{sgn}(\sigma\times\sigma')\prod\nolimits_{d_z}^{Z}(\sigma\times\sigma')\circ p_X\otimes\cdots\otimes p_X\otimes p_Y\otimes\cdots\otimes p_Y \quad (63)$$

另一方面，属于 2^{d_z}的：

$$\cdots\cdots\cdots\cdots\cdots A^Z=\frac{1}{d_Z!}\sum_{\sigma\in P_{d_z}}\mathrm{sgn}(\sigma)\prod\nolimits_{d_z}^{Z}(\sigma)$$

$$=\Big(\sum_{\sigma\in P_{d_z}}\mathrm{sgn}(\sigma)\prod\nolimits_{d_z}^{Z}(\sigma)\Big)\circ(p_X+p_Y)\otimes\cdots\otimes(p_X+p_Y)$$

改成这种可以被分解的，只有具有 d_X 的因子 p_X 和 d_Y 的因子 p_Y 的那些是非零的，因为对于 $n>d_X$ 有 $A_n^X=0$，以及对于 $n>d_Y$ 有 $A_n^Y=0$。我们由此剩下 $d_Z!/d_X!\ d_Y!$ 项之和，并且得出我们看到它们全部等于 $d_X!\ d_Y!/d_Z!$ 倍于(63)，因此和等于式(63)。这证明同构 $\det Z\cong\det X\otimes\det Y$。

最后，(iii)来自于

$$\det(X\oplus\overline{X})\cong\det X\otimes\det\overline{X}\cong\det X\otimes\overline{\det X}\cong\det X\otimes(\det X)^{-1}\cong\mathbf{1}$$

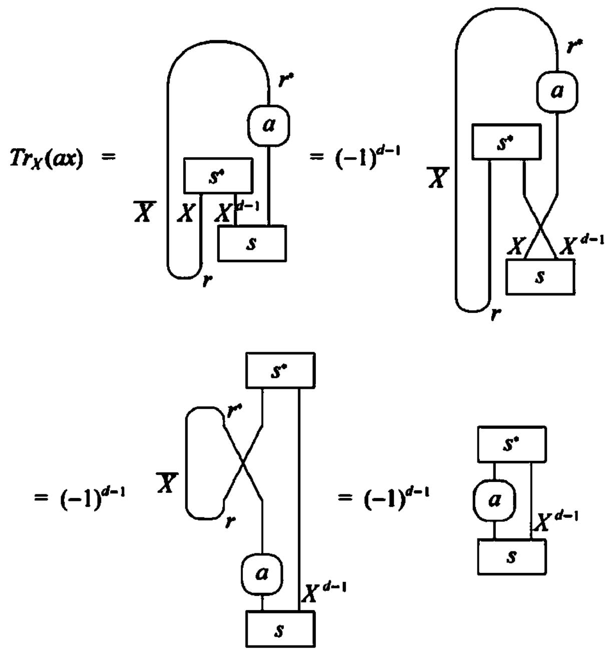
其中我们使用了这个引理的(i)和(ii)、$d(\det X)=\mathbf{1}$ 以及引理 341 的(iii)。

对于后者的运用我们举一个可计算结果为例。

引理 424。设 X 满足 $X\cong\mathbf{1}$，并且写成 $d=d(X)$。如果 s：$\mathbf{1}\to X^{\otimes d}$是一个对于 $s\circ s^*=A_d^X$ 的等距，那么：

$$s^*\otimes\mathrm{id}_X\circ\mathrm{id}_X\otimes s=(-1)^{d-1}d^{-1}\mathrm{id}_X \quad (64)$$

证明。我们缩写 $x = s^* \otimes \mathrm{id}_X \circ \mathrm{id}_X \otimes s$，并且注意到根据迹的非简并性，它足以证明，对于所有 $a \in \mathrm{End}X$ 都有 $Tr_X(ax) = (-1)^{d-1}d^{-1}Tr_X(a)$。为了证明这一点，设$(\overline{X}, r, \overline{r})$是共轭方程的标准解，并且计算：



我们依次使用 s 的整个反对称性(引理418)、扭 Θ_X 的平凡性和辫子的自然性质，现在：

$$s^* \circ a \otimes \mathrm{id}_{X^{\otimes d-1}} \circ s = Tr_1(s^* \circ a \otimes \mathrm{id}_{X^{\otimes d-1}} \circ s)$$

$$= Tr_{X^{\otimes d}}(a \otimes \mathrm{id}_{X^{\otimes d-1}} \circ s \circ s^*) = Tr_{X^{\otimes d}}(a \otimes \mathrm{id}_{X^{\otimes d-1}} \circ A_d^X)$$

为了完成这个证明，我们需要证明这个等式 $d^{-1}Tr_X a$，它是由命题420的证明的适当作修改而成。根据那里所给的相同论证，它足以证明 $Tr_{X^{\otimes d}}(a \otimes \mathrm{id}_{X^{\otimes d-1}} \circ \prod_d^X(\sigma)) = d^{\#\sigma-1}Tr_X a$。再者，计算 σ 分解成环计算的一个集合，目前明确地涉及上标1。因此，对于所有 n 上标的任何环排列 σ，它足以开始证明 $Tr_{X^{\otimes n}}(a \otimes \mathrm{id}_{X^{\otimes n-1}} \circ \prod_n^X(\sigma)) = Tr_X a$。在适当地方添加 a，计算在本质上跟以前一样。唯一不同之处是留下 $Tr_X a$ 替代 $Tr_X \mathrm{id}_X = d(X)$，从而产生所希望的结果。

评论425。具有行列式1的对象被称为在[Doplicher and Roberts，1989]里的特殊情况，在那也能找到这一小节的全部结果。

这就结束了反对称化和行列式的讨论，接下来我们转向对称化和对称代数。正是在这里我们需要我们在 A7 小节引入的 Ind 范畴。

B10 对称代数

在“通常的”代数里人们是在矢量空间 V 上定义对称代数 $S(V)$ 的，除非 $V=\{0\}$，这是非平凡矢量空间的无穷直和。我们将需要对对称张量范畴的这个建构进行推广而不是其他 Vect。在对象的无穷直和在 C^* 范畴的设定（定义 345）这个意义下，在下面的考虑下一个更加方便的设定由阿贝尔范畴理论给出。

引理 426。设$\mathcal{C}$是一个 STC^* 并且 $X\in\mathcal{C}$，对于每一个 $n\in\mathbb{N}$，选择一个对象 $S_n(X)$ 和一个等距 u_n：$S_n(X)\to X^{\otimes n}$，使得 $u_n\circ u_n^*=S_n^X$。另外，设 $u_0=\mathrm{id}_1$，被解释成一个从 $S_0(X)=1$ 到 $X^0=1$ 的态射。态射 $m_{i,j}$：$S_i(X)\otimes S_j(X)\to S_{i+j}(X)$ 被定义为：

$$m_{i,j}:\ S_i(X)\otimes S_j(X)\xrightarrow{u_i\otimes u_j}X^{\otimes i}\otimes X^{\otimes j}\equiv X^{\otimes(i+j)}\xrightarrow{u_{i+j}^*}S_{i+j}(X)$$

对于所有的 i，j，$k\in\mathbb{Z}_+$满足：

$$m_{i+j,k}\circ m_{i,j}\otimes\mathrm{id}_{S_k(X)}=m_{i,j+k}\circ\mathrm{id}_{S_i(X)}\otimes m_{j,k}$$

而且，

$$m_{i,j}=m_{j,i}\circ c_{S_i(X),S_j(X)}\quad\forall_{i,j}$$

以及

$$m_{i,0}=m_{0,i}=\mathrm{id}_{S_i(X)}$$

证明。作为对所有的 $\sigma\in P_n$ 都有的一个结果 $S_n^X\circ\prod_n^X(\sigma)=S_n^X(\sigma)$，参看引理 418，我们有：

$$s_{i+j+k}^X\circ S_{i+j}^X\otimes\mathrm{id}_{X^{\otimes k}}\circ S_i^X\otimes S_j^X\otimes\mathrm{id}_{X^{\otimes k}}=S_{i+j+k}^X\circ S_{i+j}^X\otimes\mathrm{id}_{X^{\otimes k}}=S_{i+j+k}^X$$

$$S_{i+j+k}^X\circ\mathrm{id}_{X^{\otimes i}}\otimes S_{j+k}^X\circ\mathrm{id}_{X^{\otimes i}}\otimes S_j^X\otimes S_k^X=S_{i+j+k}^X\circ\mathrm{id}_{X^{\otimes i}}\otimes S_{j+k}^X=S_{i+j+k}^X$$

全部用 u_{i+j+k}^*在左边相乘，而用 $u_i\otimes u_j\otimes u_k$ 右边相乘，并使用 $u_i^*\circ S_i^X=u_n^*$ 和 $S_i^X\circ u_i=u_i$，这就有：

$$u_{i+j+k}^*\circ S_{i+j}^X\otimes\mathrm{id}_{X^{\otimes k}}\circ u_i\otimes u_j\otimes u_k=u_{i+j+k}^*\circ u_i\otimes u_j\otimes u_k$$

$$=u_{i+j+k}^*\circ\mathrm{id}_{X^{\otimes i}}\otimes S_{j+k}^X\circ u_i\otimes u_j\otimes u_k$$

再次使用 $S_{i+j}^{X}=u_{i+j}\circ u_{i+j}^{*}$，我们就有我们想证明的第一个等式，进一步可以得到：

$$m_{j,i}\circ c_{S_i(X),S_j(X)}=u_{i+j}^{*}\circ u_j\otimes u_i\circ c_{S_i(X),S_j(X)}=u_{i+j}^{*}\circ c_{X^{\otimes i},X^{\otimes j}}\circ u_i\otimes u_j$$

$$=u_{i+j}^{*}\circ\prod{}_{i+j}^{X}(\sigma)\circ u_i\otimes u_j=u_{i+j}^{*}\circ S_{i+j}^{X}\circ\prod{}_{i+j}^{X}(\sigma)u_i\otimes u_j$$

$$=u_{i+j}^{*}\circ S_{i+j}^{X}\circ u_i\otimes u_j=u_{i+j}^{*}\circ u_i\otimes u_j=m_{i,j}$$

其中 $\sigma\in P_{i+j}$是把第一个 i 跟剩下的 j 组进行交换的转置，最后一个说法在 $S_0(X)=1$ 里是明显的。

按照引理 353 的观点，$\mathcal{C}$(有放进去的零对象)是一个阿贝尔范畴，因此存在一个阿贝尔$\mathbb{C}$ 线性严格对称张量范畴 Ind $\mathcal{C}$，它包含作为一个完全子空间并且关于可滤归纳极限完备的$\mathcal{C}$。因此，对于任何在 STC^{*} $\mathcal{C}$里的对象 X，存在一个对象

$$S(X)=\varinjlim_{n\to\infty}\bigoplus_{i=0}^{n}S_n(X)$$

跟单同态 v_n：$S_n(X)\to S(X)$结合在一起。

命题 427。设$\mathcal{C}$是一个 STC^{*} 并且 $X\in\mathcal{C}$，那么存在一个态射

$$m_{S(X)}:\ S(X)\otimes S(X)\to S(X)$$

使得：

$$m_{S(X)}\circ v_i\otimes v_j=v_{i+j}\circ m_{i,j}:\ S_i(X)\otimes S_j(X)\to S(X)$$

并且$(S(X),\ m_{S(X)},\ \eta_{S(X)}\equiv v_0)$是在 Ind $\mathcal{C}$里的对易的。

证明。这等于是使用

$$\mathrm{Hom}_{\mathrm{Ind}\mathcal{C}}(S(X)\otimes S(X),\ S(X))$$

$$=\varprojlim_{m}\varinjlim_{n}\mathrm{Hom}_{\mathcal{C}}\Big(\bigoplus_{i,j=0}^{m}S_i(X)\otimes S_j(X),\ \bigoplus_{k=0}^{n}S_k(X)\Big)$$

来把 $m_{i,j}$：$S_i(X)\otimes S_j(X)\to S_{i+j}(X)$集结成一个大的态射 $S(X)\otimes S(X)\to S(X)$。我们省略了除明确的细节以外的冗长琐碎，结合性($m_{S(X)}\circ m_{S(X)}\otimes\mathrm{id}_{S(X)}=m_{S(X)}\circ\mathrm{id}_{S(X)}\otimes m_{S(X)}$)和交换性($m_{S(X)}=m_{S(X)}\circ c_{S(X),S(X)}$)就源自于引理 426 里建立起来的 $m_{i,j}$的各自属性。统一性 $m_{S(X)}\circ\mathrm{id}_{S(X)}\otimes v_0=\mathrm{id}_{S(X)}\otimes v_0=\mathrm{id}_{S(X)}$ 遵从于 $m_{i,0}=m_{0,i}=\mathrm{id}_{S_i(X)}$。

我们现在研究对称化运算跟反对称化之间的相互作用，即在行列式代数跟对称代数之间，它们是嵌入定理的核心。我们先来看看两个在严格对称张量范

畴里的对易幺半群（Q_i，m_i，η_i），三元组（$Q_1 \otimes Q_2$，$m_{Q_1 \otimes Q_2}$，$\eta_{Q_1 \otimes Q_2}$），其中 $\eta_{Q_1 \otimes Q_2} = \eta_1 \otimes \eta_2$ 和 $m_{Q_1 \otimes Q_2} = m_1 \otimes m_2 \circ \mathrm{id}_{Q_1} \otimes c_{Q_2, Q_1} \otimes \mathrm{id}_{Q_2}$，定义了一个对易幺半群，直积$(Q_1, m_1, \eta_1) \times (Q_2, m_2, \eta_2)$。直积 × 严格结合的，因此乘法直积由归纳明确定义。

引理 428。设C是一个 STC，并且假设 $Z \in C$满足 $\det Z \cong 1$。我们写为 $d = d(Z)$，并且挑选 s：$\mathbf{1} \to Z^{\otimes d}$，$s'$：$Z^{\otimes d} \to \mathbf{1}$，使得 $s' \circ s = \mathrm{id}_1$ 和 $s \circ s' = A_d^Z$。设 $S(Z)$是 Z 上具有正则嵌入 v_0：$\mathbf{1} \to S(Z)$，v_1：$Z \to S(Z)$。考虑 $Q = S(Z)^{\otimes d}$上的对易幺半群结构由下式给出：

$$(Q, m_Q, \eta_Q) = (S(Z), m_{s(Z)}, \eta_{S(Z)})^{\times d}$$

定义态射f：$1 \to Q$ 和 u_i：$Z \to Q$，t_i：$Z^{\otimes(d-1)} \to Q$，$i = 1, \cdots, d$ 为：

$$f = \underbrace{v_1 \otimes \cdots \otimes v_1}_{d\text{个因子}} \circ s$$

$$u_i = \underbrace{v_0 \otimes \cdots \otimes v_0}_{i-1\text{个因子}} \otimes v_1 \otimes \underbrace{v_0 \otimes \cdots \otimes v_0}_{d-i\text{个因子}}$$

$$t_i(-1)^{d-i} \underbrace{v_1 \otimes \cdots \otimes v_1}_{i-1\text{个因子}} \otimes v_0 \otimes \underbrace{v_1 \otimes \cdots \otimes v_1}_{d-i\text{个因子}}$$

那么 s，f，u_i，t_j 满足：

$$m_Q \circ t_j \otimes u_i \circ s = \delta_{ij} f \quad \forall_{i,j} \in \{1, \cdots, d\} \tag{65}$$

证明。首先注意到 s：$\mathbf{1} \to Z^{\otimes d}$就像要求的那样存在，因为 $\det Z \cong \mathbf{1}$ 和 f 是一个首一的合成，因此是非零的。我们计算：

$$\begin{aligned}
m_Q \circ t_i \otimes u_i \circ s &= (-1)^{d-i} \mathrm{id}_{S(Z)^{(i-1)}} \otimes c_{S(Z)^{\otimes(d-i)}, S(Z)} \circ v_1 \otimes v_1 \otimes \cdots \otimes v_1 \circ s \\
&= (-1)^{d-i} v_1 \otimes v_1 \otimes \cdots \otimes v_1 \circ \mathrm{id}_{Z^{\otimes(i-1)}} \otimes c_{Z^{\otimes(d-i)}, Z} \circ s \\
&= v_1 \otimes v_1 \otimes \cdots v_1 \circ s \\
&= f
\end{aligned}$$

在第一个等式里我们用了作为$(S(Z), m_{S(Z)}, \eta_{S(Z)})$的 d 重直积的(Q, m_Q, η_Q)的定义，以及 $v_0 = \eta_{S(Z)}$是统一的事实，在第二个式子里辫子的自然性，以及在第三个里的引理 418。为了看到 $m_Q \circ t_j \otimes u_i \circ s = 0$，如果 $i \neq j$，考虑到 $j = d-1$，$i = d$。那么 $m_Q \circ t_j \otimes u_i \circ s$：$\mathbf{1} \to S(Z)^{\otimes d} \equiv Q$ 是合成：

$$1 \xrightarrow{s} Z^{\otimes d} \xrightarrow{\overbrace{v_1 \otimes \cdots \otimes v_1}^{d-2\text{个因数}} \otimes v_0 \otimes v_1 \otimes v_1} S(Z)^{\otimes(d+1)} \xrightarrow{\mathrm{id}_{S(Z)^{\otimes(d-1)}} \otimes m_{S(Z)}} S(Z)^{\otimes d}$$

现在，已知：

$$
\begin{aligned}
&\mathrm{id}_{S(Z)^{\otimes(d-1)}}\otimes m_{S(Z)}\circ v_1\otimes\cdots\otimes v_1\otimes v_0\otimes v_1\otimes v_1\circ s\\
&=\mathrm{id}_{S(Z)^{\otimes(d-1)}}\otimes(M_{S(Z)}\circ c_{S(Z),S(Z)})\circ v_1\otimes\cdots\otimes v_1\otimes v_0\otimes v_1\otimes v_1\circ s\\
&=\mathrm{id}_{S(Z)^{\otimes(d-1)}}\otimes m_{S(Z)}\circ \mathrm{id}_{S(Z)^{\otimes(d-1)}}\otimes c_{S(Z),S(Z)}\circ v_1\otimes\cdots\otimes v_1\otimes v_0\otimes v_1\otimes v_1\circ s\\
&=\mathrm{id}_{S(Z)^{\otimes(d-1)}}\otimes m_{S(Z)}\circ v_1\otimes\cdots\otimes v_1\otimes v_0\otimes v_1\otimes v_1\circ \mathrm{id}_{Z^{\otimes(d-2)}}\otimes c_{Z,Z}\circ s\\
&=-\mathrm{id}_{S(Z)^{\otimes(d-1)}}\otimes m_{S(Z)}\circ v_1\otimes\cdots\otimes v_1\otimes v_0\otimes v_1\otimes v_1\circ s
\end{aligned}
$$

其中我们在第一步用到 $m_{S(Z)}$ 的对易性，而在最后一步用了 s 的反对称性。因此 $m_Q\circ u_d\otimes t_{d-1}\circ s=-m_Q\circ u_d\otimes t_{d-1}\circ s=0$。对于 $i\neq j$ 时这个论证完全一样，只是在详细写出来时则变得十分烦琐。

评论 429。引理 428 和下面的命题 430，两个都来自[Bichon，1998]，且在通向我们的重构定理的道路上是关键步骤。

B11 吸收交换性单项的建构

贯穿这一节，设$\mathcal{C}$是一个具有幺半生成元 Z 的偶 STC^*，考虑到在 Ind $\mathcal{C}$里的可交换幺半群$(Q, m, \eta)=(S(Z), m_{S(Z)}, \eta_{S(Z)})^{\times d(Z)}$，以及定义在引理428 里的态射 s，s'，f，u_i，t_j。那么 $m_0\in \mathrm{End}Q$ 被定义为：

$$m_0=m_Q\circ \mathrm{id}_Q\otimes(f-\eta_Q)=m_Q\circ \mathrm{id}_Q\otimes f-\mathrm{id}_Q$$

是一个 Q 模映射，因此 $m_0\in \mathrm{End}_Q((Q, m_Q))$。那么它的像 $j=\mathrm{im}\ m_0$：$(J, \mu_J)\to(Q, m_Q)$（在阿贝尔范畴 $Q-\mathrm{Mod}$ 里）定义了一个在(Q, m, η)里的理想 j：$(J, \mu_J)\to(Q, m)$，这是一个真理想，当且仅当 j 不是一个同构，当且仅当 m_0 不是一个同构。这个问题先放一放，我们有：

命题 430。设$\mathcal{C}$是一个偶对称 STC^*，并且设 $Z\in\mathcal{C}$是如此，使得 $\det Z\cong\mathbf{1}$。设(Q, m, η)和 s，s'，f，u_i，t_j 被定义为就像在引理 428 里那样，并且 m_0 也跟上面一样。设 j'：$(J', \mu')\to(Q, m)$是在(Q, m, η)里的任何真理想，包含理想 j：$(J, \mu)\to(Q, m)$，其中 $j=\mathrm{im}\ m_0$。设(B, m_B, η_B)是商幺半群单项。那么存在一个 B 模的同构：

$$(B\otimes Z, m\otimes\mathrm{id}_Z)\cong d(Z)\cdot(B, m_B)$$

证明。由于理想是真正的，商(B, m_B, η_B)是非平凡的，并且我们具有一

个 epi，即 p：$Q \to B$ 满足：

$$p \circ m_Q = m_B \circ p \otimes p \tag{66}$$

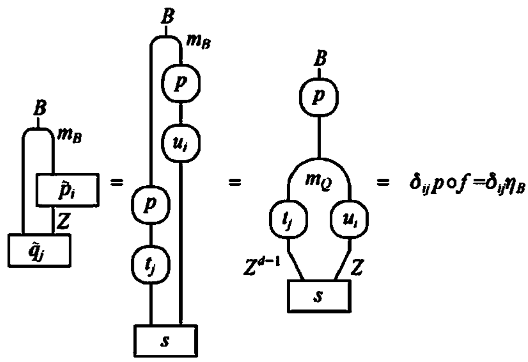
$$p \circ f = p \circ \eta_Q = \eta_B \tag{67}$$

为了证明所断言的 B 模的同构 $B \otimes Z \cong dB$，我们定义态射 $\tilde{q}_i \in \mathrm{Hom}(1, B \otimes Z)$，$\tilde{p}_i \in \mathrm{Hom}(Z, B)$，$i = 1, \cdots, d$ 为下述合成：

$$\tilde{q}_i: 1 \xrightarrow{s} Z^{\otimes d} \equiv Z^{\otimes(d-1)} \otimes Z \xrightarrow{t_i \otimes \mathrm{id}_Z} Q \otimes Z \xrightarrow{p \otimes \mathrm{id}_Z} B \otimes Z$$

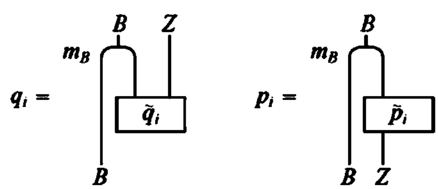
$$\tilde{p}_i: Z \xrightarrow{u_i} Q \xrightarrow{p} B$$

依次使用式(66)、式(65)和式(67)我们推断：

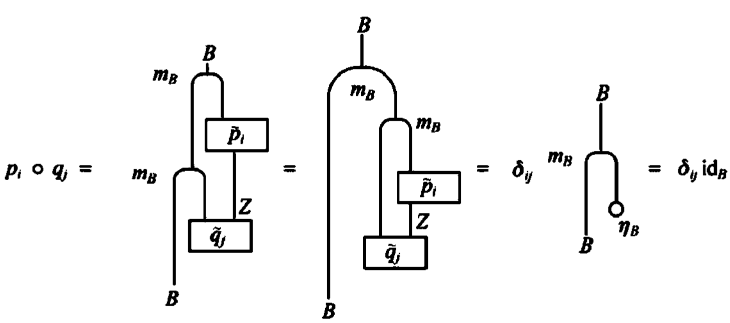


(68)

对于 $i = 1, \cdots, d$，定义：



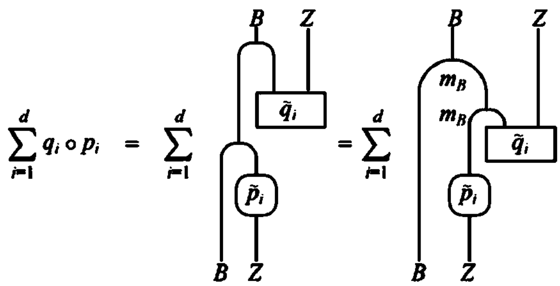
我们发现：

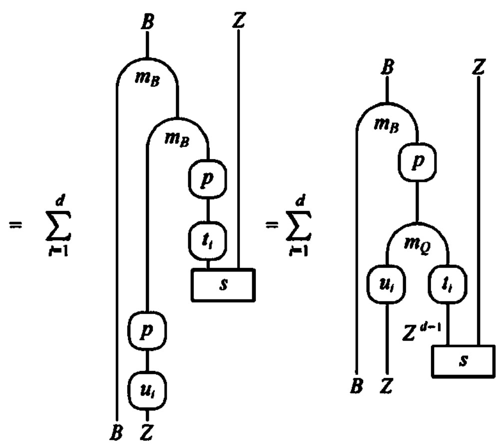


其中接下来的最后一步我们用了式(68)。从定义来看 p_i，q_i 明显是 B 模的态射。我们从而证明了 B 模 $(B\otimes Z,\ m_B\otimes \mathrm{id}_Z)$ 具有 d 个直和项 $(B,\ m_B)$，由此：

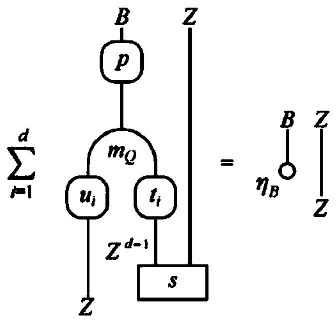
$$(B\otimes Z,\ m_B\otimes \mathrm{id}_Z)\cong\underbrace{(B,\ m_B)\oplus\cdots\oplus(B,\ m_B)}_{d\text{ 个直和}}\oplus(N,\ \mu_N)$$

接下来要证明的是 $N=0$，或者等价的说法，$\sum_{i=1}^{d}q_i\circ p_i=\mathrm{id}_{B\otimes Z}$。对这个结果的一个简单论证是在[Deligne，1990；Bichon，1998]里给出的，但是由于它有些抽象，我们给出一个按部就班的计算式证明，我们计算





跟 $\eta_B\otimes \mathrm{id}_Z$ 结合，证明这等于 $\mathrm{id}_{B\otimes Z}$，当且仅当



(69)

在(Q, m_Q, η_Q)的定义看来，式(69)的左边等于：

$$\sum_{i=1}^{d}(-1)^{d-i}(p\circ c_{S(Z),S(Z)^{\otimes(i-1)}}\otimes \mathrm{id}_{S(Z)^{\otimes(d-i)}}\circ v_1\otimes\cdots\otimes v_1)\otimes \mathrm{id}_Z\circ \mathrm{id}_Z\otimes s=$$

$$(p\circ v_1\otimes\cdots\otimes v_1)\otimes \mathrm{id}_Z\circ(\sum_{i=1}^{d}(-1)^{d-i}c_{Z,Z^{\otimes(i-1)}}\otimes \mathrm{id}_{Z^{\otimes(d-i)}}\otimes \mathrm{id}_Z\circ\ \mathrm{id}_Z\otimes s)$$

(70)

写 $K_i=c_{Z,Z^{\otimes(i-1)}}\otimes \mathrm{id}_{Z^{\otimes(d-i+1)}}\circ \mathrm{id}_Z\otimes s$，其中 $i\in\{1,\cdots,d\}$，人们容易证明对于所有 $j\in\{1,\cdots,i-1\}$ 都有：

$$\prod\nolimits_{d+1}^{Z}(\sigma_j)\circ K_i=\begin{cases}K_{i-1}: & j=i-1\\ K_{i+1}: & j=i\\ -K_i: & \text{其他情况}\end{cases}$$

这就意味着在(70)的大括号里的态射 $Z\to Z^{\otimes(d+1)}$，是关于第 d 分支总体上反对称的，即用 $\prod_{d+1}^{Z}(\sigma_j), j=1,\cdots,d-1$ 从左边乘而改变其符号。我们由此能够插入 $A_d^Z=s\circ s'$在适当位置并且看到式(70)等于：

$$=(p\circ v_1\otimes\cdots\otimes v_1)\otimes \mathrm{id}_Z\circ(s\circ s')\otimes \mathrm{id}_Z\circ(\sum_{i=1}^{d}(-1)^{d-i}c_{Z,Z^{\otimes(i-1)}}\otimes \mathrm{id}_{Z^{\otimes(d-i)}}\otimes \mathrm{id}_Z\circ \mathrm{id}_Z\otimes s)$$

$$=(p\circ v_1\otimes\cdots\otimes v_1\circ s)\otimes \mathrm{id}_Z\circ(\sum_{i=1}^{d}(-1)^{d-i}s'\otimes \mathrm{id}_Z\circ c_{Z,Z^{\otimes(i-1)}}\otimes \mathrm{id}_{Z^{\otimes(d-i)}}\otimes \mathrm{id}_Z\circ \mathrm{id}_Z\otimes s)$$

现在，$p\circ v_1\otimes\cdots\otimes v_1\circ s=p\circ f=\eta_B$。另一方面，通过 s 的总的反对称性，我们有 $s'\circ c_{Z,Z^{\otimes(i-1)}}\otimes \mathrm{id}_{Z^{\otimes(d-i)}}=(-1)^{i-1}s'$，由此：

$$\sum_{i=1}^{d}(-1)^{d-i}\quad s'\otimes \mathrm{id}_Z\circ c_{Z,Z^{\otimes(i-1)}}\otimes \mathrm{id}_{Z^{\otimes(d-i)}}\otimes \mathrm{id}_Z\circ \mathrm{id}_Z\otimes s$$

$$=\sum_{i=1}^{d}(-1)^{d-i}(-1)^{i-1}s'\otimes \mathrm{id}_Z\circ \mathrm{id}_Z\otimes s$$

$$=d(-1)^{d-1}s'\otimes \mathrm{id}_Z\circ \mathrm{id}_Z\otimes s=\mathrm{id}_Z$$

其中最后一个等式是由引理424所提供的。因此式(69)是正确的，意味着$\sum_{i=1}^{d}q_i\circ p_i=\mathrm{id}_{B\otimes Z}$以及因而被断言的$B$模的同构$B\otimes Z\cong d(Z)B$。

引理431。设$\mathcal{C}$，Z和单项(Q,m,η)就像在引理428里那样，那么交换代数$\Gamma_Q=\mathrm{Hom}(\mathbf{1},Q)$是$\mathbb{Z}_+$级的，并且具有最大可数的维度。

证明。根据Q的建构，我们具有：

$$\Gamma_Q=\mathrm{Hom}(\mathbf{1},Q)=\varinjlim_{n}\bigoplus_{i=0}^{n}\mathrm{Hom}(\mathbf{1},S_i(Z))=\bigoplus_{i\geqslant 0}\mathrm{Hom}(\mathbf{1},S_i(Z))$$

在右手边的每一个直和项都处在$\mathcal{C}$里，因此具有有限维度。因此Γ_Q具有最大可数的维度。Γ_Q是$\mathbb{Z}_+$级代数这一点是明确来自于m_Q的定义，按照态引理426里的射$m_{i,j}$：$S_i(X)\otimes S_j(X)\to S_{i+j}(X)$。

定理432。设$Z\in\mathcal{C}$使得有$\det Z\cong 1$，那么存在一个在Ind $\mathcal{C}$里的可交换幺半群(B,m_B,η_B)，使得$\dim \mathrm{Hom}_{\mathrm{Ind}\mathcal{C}}(\mathbf{1},B)=1$，并且存在一个$B$模的同构$B\otimes Z\cong d(Z)B$。

证明。设(Q,m,η)和理想$j=\mathrm{im}\ m_0$：$(J,\mu)\to(Q,m)$跟前面一样。假设j是一个同构，因此是个epi。那么m_0是epi，由此根据引理366是一个同构。特别是，由$s\mapsto s\bullet(f-\eta)$给定的映射$\Gamma_Q\to\Gamma_Q$是一个同构，因此$f-\eta\in\Gamma_Q$是可逆的。不过，由于$\Gamma_Q$是$\mathbb{Z}_+$级的，并且$F-\eta\in\Gamma_Q$并不处于零级部分，因此这是不可能的。因此理想$j$是真理想。根据引理362存在一个包含$j$：$(J,\mu)\to(Q,m)$的最大理想$\tilde{j}$：$(\tilde{J},\tilde{\mu})\to(Q,m)$。如果幺半群$(B,m_B,\eta_B)$用$j$：$(\tilde{J},\tilde{\mu})\to(Q,m)$除$(Q,m,\eta_Q)$的商，命题430意味着$B$模的同构$B\otimes Z\cong d(Z)\cdot B$。由引理364，商模$(B,m_B,\eta_B)$没有真的非零理想，因此，根据引理365，可交换$\mathbb{C}$代数$\mathrm{End}_B((B,m_B))$是扩展$k$的一个场。根据引理385，作

为一个$\mathbb{C}$代数有$\mathrm{End}_B((B,\ m))\cong \mathrm{Hom}(\mathbf{1},\ B)=:\ \Gamma_B$。根据引理371，单位$\mathbf{1}\in \mathrm{Ind}\ \mathcal{C}$是投影的，因此引理363意味着$\Gamma_B$是一个$\Gamma_Q$的商，以及根据引理431它具有最大的可数的维度。现在下面的引理433的使用，并给出$\Gamma_B=\mathbb{C}$，还由此得到期望的结果：

$$\dim \mathrm{Hom}(\mathbf{1},\ B)=1$$

引理433。设$K\supset\mathbb{C}$是一个$\mathbb{C}$的场扩展，如果$[K:\ \mathbb{C}]\equiv \dim_{\mathbb{C}} K$是最大可数的，则$K=\mathbb{C}$。

证明。假设$x\in K$是超越于$\mathbb{C}$之上的，我们断言集合$\{\frac{1}{x+a}\mid a\in\mathbb{C}\}\subset K$是线性独立于$\mathbb{C}$：假设$\sum_{i=1}^{N}\frac{b_i}{x+a_i}=0$,其中$a_i$是成对不同的,并且$b_i\in\mathbb{C}$。用$\prod_i(x+a_i)$(在$K$里是非零的）去乘,我们得到一个关于$x$的多项式方程$\sum_{i=1}^{N}b_i\prod_{j\neq i}(x+a_j)=0=\sum_{k=0}^{N-1}c_kx^k$。由于$x$是超越的,我们得到对于所有的$k=0,\cdots,N-1$都有的$c_k=0$。这给我们$N$个线性方程$\sum_{i=1}^{N}M_{ki}b_i=0,k=1,\cdots,N$,其中$M_{ki}=\sum_{\substack{s\subset\{1,\cdots,N\}-\{i\}\\ \#S=k-1}}\prod_{s\in S}a_s$。这个矩阵能够通过基本元素行变换变成矩阵($V_{ki}=a_i^{k-1}$)。由范得蒙(Vandermonde)的公式，$\det V=\prod_{i<j}(a_j-a_i)\neq0$，因此$M\mathbf{b}=0$的唯一解是$b_1=\cdots=b_N=0$，证明是线性无关的。由于$\mathbb{C}$是不可数的，这跟$k$具有$\mathbb{C}$上的可数的维度假设相矛盾。因此，$K/\mathbb{C}$是代数的，从而有$K=\mathbb{C}$，因为$\mathbb{C}$是代数上闭合的。

最后我们得到：

定理411的证明。如果$\mathcal{C}$是一个偶STC^*，具有幺半产生元Z，引理423允许我们假设$\det Z\cong\mathbf{1}$(用$Z\oplus\overline{Z}$替代Z)。现在定理432提供了一个在$\mathrm{Ind}\ \mathcal{C}$里的单项$(B,\ m,\ \eta)$，满足推论415的假设，这就产生了一个对称性纤维函子E：$\mathcal{C}\to\mathrm{Vect}_{\mathbb{C}}$。

评论 434。指出我们的定理 411 证明跟[Deligne，1990；Bichon，1998]方法的差异似乎是有教育意义的。在[Deligne，1990]里，一个可交换幺半群(Q，m，η)，从而有一个 Q 模的同构 $Q\otimes Z\cong d(Z)Q$，是通过一些复杂的归纳过程得到的。我们所给的幺半群的明确建构归功于[Bichon，1998]。德利涅进行的是通过注意到，对于任何 $X\in\mathcal{C}$，k 矢量空间 $\mathrm{Hom}(\mathbf{1},\ Q\otimes X)$，是一个可交换环 $\Gamma_Q:=\mathrm{End}_Q((Q,\ m))\cong\mathrm{Hom}(\mathbf{1},\ Q)$上的模，以及函子是一个关于 Γ_Q - Mod(而不是 $\mathrm{Vect}_{\mathbb{C}}$ 的)张量积的单项 $\widetilde{E}$：$X\mapsto\mathrm{Hom}(\mathbf{1},\ Q\otimes X)$。现在，一个关于在 Γ_Q 里的最大理想 J 的商的过程，被用来获得张量函子 E：$\mathcal{C}\to K$ - Vect，其中 $K=\Gamma_Q/J$ 是基础场 k 的场扩展。如果 $\mathrm{Hom}(\mathbf{1},\ Q)$是属于最大可数维的，那么$[K:k]\leqslant\aleph_0$，并且如果 k 是不可数的以及代数上闭合的，因此就有 $K=k$。

我们的方法在两个方面不一样。不太重要的方面是，我们强调 $\det Z\cong\mathbf{1}$ 使幺半群(Q，m，η)的建构稍微比[Bichon，1998]里更清楚些，重要的方面是，我们用 Ind $\mathcal{C}$里的 Q 模的范畴里面，而不是用 Γ_Q 模的范畴里的最大理想，产生一个具有 $\Gamma_{Q'}=\mathbb{C}$ 的 Ind $\mathcal{C}$里的单项(Q'，m'，η')。除了以一种更加直接的形式产生一个对称相位函子 E：$\mathcal{C}\to\mathrm{Vect}_{\mathbb{C}}$，这具有额外的好处，正如我们在最后一个小节里将会证明的，可以容许恢复群 $\mathrm{Nat}_{\otimes}E$，而不管纤维函子及其自然变换！其最终原因是，由于嵌入函子的唯一性，Ind $\mathcal{C}$里的幺半群(Q'，m'，η')不外乎是在 Rep$G$ 里的幺半群(π_l，$\widetilde{m}$，$\widetilde{\eta}$)，它源自于 G 的左正则表示，参看[Müger and Tuset，2006]。

B12 补遗

在前一个小节里我们得到纤维函子的存在性证明，以及通过具体的田中定理证明等价性$\mathcal{C}\simeq\mathrm{Rep}_f(G,\ k)$，其中($G$，$k$)是一个紧致子群。然而，我们将要证明群 $\mathrm{Nat}_{\otimes}E$ 以及在某些情况下 G 是如何能够直接从幺半群(Q，m，η)删除的，而不管纤维函子、自然变换等。

定义 435。在严格张量范畴$\mathcal{C}$里幺半群(Q, m, η)的自同构群是：

$$\mathrm{Aut}(Q, m, \eta) = \{g \in \mathrm{Aut}Q \mid g \circ m = m \circ g \otimes g, \ g \circ \eta = \eta\}$$

命题 436。设$\mathcal{C}$是一个 STC^*，并且(Q, m, η)是一个满足下面条件的在 $\mathrm{Ind}\,\mathcal{C}$里的幺半群：

(i) $\dim \mathrm{Hom}_{\mathrm{Ind}\,\mathcal{C}}(\mathbf{1}, Q) = 1$；

(ii) 对于每个 $X \in \mathcal{C}$，存在 $n(X) \in \mathbb{Z}_+$，使得任何 $X \neq 0$ 时候都有 $n(X) \neq 0$，以及一个 Q 模的同构 α_X：$(Q \otimes X, m \otimes \mathrm{id}_X) \rightarrow n(X) \cdot (Q, m)$。

那么在命题 413 里构造的函子的幺半群自然子同构的群 $\mathrm{Nat}_{\otimes} E$，是正则地同构于群 $\mathrm{Aut}(Q, m, \eta)$。

证明。设 $g \in \mathrm{Aut}(Q, m, \eta)$，对于任何 $X \in \mathcal{C}$我们把 $g_X \in \mathrm{End}E(X)$ 定义为：

$$g_X \psi = g \otimes \mathrm{id}_X \circ \psi \quad \forall \psi \in E(X) = \mathrm{Hom}(\mathbf{1}, Q \otimes X)$$

从$(g_X)_{X \in \mathcal{C}}$的定义和函子 E 的定义，直接得到$(g_X)_{X \in \mathcal{C}}$是一个从 E 到自身的自然变换。我们必须证明这个自然变换是幺半的，即

$$\begin{array}{ccc} E(X) \otimes E(Y) & \xrightarrow{d_{X,Y}} & E(X \otimes Y) \\ \Big\downarrow {\scriptstyle g_X \otimes g_Y} & & \Big\downarrow {\scriptstyle g_{X \otimes Y}} \\ E(X) \otimes E(Y) & \xrightarrow{d_{X,Y}} & E(X \otimes Y) \end{array}$$

对易。为此，考虑 $\phi \in E(X) = \mathrm{Hom}(\mathbf{1}, Q \otimes X)$，$\psi \in E(X) = \mathrm{Hom}(\mathbf{1}, Q \otimes Y)$，以及具有刚定义的$(g_X)_{X \in \mathcal{C}}$的 $g \in \mathrm{Aut}(Q, m, \eta)$，那么在 $g_{X \otimes Y} \circ d_{X,Y}$之下的 $\phi \boxtimes \psi \in E(X) \otimes E(Y)$的像是：

$$g \otimes \mathrm{id}_{X \otimes Y} \circ m \otimes \mathrm{id}_{X \otimes Y} \circ \mathrm{id}_Q \otimes \phi \otimes \mathrm{id}_Y \circ \psi$$

而在 $d_{X,Y} \circ g_X \otimes g_Y$ 之下其像为：

$$m \otimes \mathrm{id}_{X \otimes Y} \circ g \otimes g \otimes \mathrm{id}_{X \otimes Y} \circ \mathrm{id}_Q \otimes \phi \otimes \mathrm{id}_Y \circ \psi$$

按照 $g \circ m = m \circ g \otimes g$ 来看，这两个表示是一致的，因此$(g_X) \in \mathrm{Nat}_{\otimes} E$。非

常容易得到映射 σ：$\mathrm{Aut}(Q, m, \eta) \to \mathrm{Nat}_{\otimes}E$，因此得到的是一个群同态。

我们断言 σ 是一个同构，重要的是我们在 Ind $\mathcal{C}$里而不是$\hat{\mathcal{C}}$里进行的，因为这意味着 Q 是在$\mathcal{C}$里的对象的归纳极限。假设(i)和(ii)，然后给出对于所有的 $X \in \mathcal{C}$都有 $\mathrm{Hom}(X, Q) \cong \mathrm{Hom}(1, Q \otimes \overline{X} \cong \mathbb{C}^{n(\overline{X})}$，因此(用 $n(X) = n(\overline{X}) = \dim E(X)$)有：

$$Q \cong \varinjlim_{S \subset I} \bigoplus_{i \in S} n(X_i) X_i \quad \text{和} \quad \mathrm{End}Q \cong \prod_{i \in I} \mathrm{End}E(X_i) \tag{71}$$

其中 S 贯穿 I 的有限集。现在假设 $\sigma(g)$ 是一个全同自然变换，即对于所有 $X \in \mathcal{C}$都有 $g \otimes \mathrm{id}_X \circ \phi = \phi$，并且 $\phi \in \mathrm{Hom}(1, Q \otimes X)$。由于$\mathcal{C}$里共轭的存在，这就等价于对所有 $Y \in \mathcal{C}$都有 $g \circ s = s$，并且 $s \in \mathrm{Hom}(Y, Q)$。由于 Q 是$\mathcal{C}$里的一个归纳极限，这意味着 $g = \mathrm{id}_Q$。

如果现在 $\alpha \in \mathrm{Nat}_{\otimes}E$，我们首先注意到，根据 396 可知 α 是一个自然同构。由同构 $\mathrm{Nat}E \cong \prod_{i \in I} \mathrm{End}E(X_i)$（参看命题 375 的证明）与式(71)，我们得到一个映射 $\mathrm{Nat}_{\otimes}E \to \mathrm{Aut}Q$。返回前面的计算表明，任何 $\alpha \in \mathrm{Nat}_{\otimes}E$ 产生 $\mathrm{Aut}(Q, m, \eta)$ 的一个元。

评论 437。这个结果表明，群 $\mathrm{Nat}_{\otimes}E$ 能够直接从 Ind $\mathcal{C}$里的吸收幺半群(Q, m, η)恢复。一般来说就像 B1 小节里定义的那样，紧致群 G 是 $\mathrm{Nat}_{\otimes}E$ 的一个真子群，后者作为 G 的前代数包络线。在 $G = U(1)$，$SU(2)$，$U(2)$ 的情况下，比如可能分别是$\mathbb{C}^{\times}$，$SL(2, \mathbb{C})$，$GL(2, \mathbb{C})$。但是，如果$\mathcal{C}$是有限的(即具有简单对象的有限多个同构类)，那么 $\mathrm{Nat}_{\otimes}E$ 是有限的，并且 $G = \mathrm{Nat}_{\otimes}E$。有趣的是，即便在有限$\mathcal{C}$的情况下，那里的幺半群$(Q, m, \eta)$实际上处于$\mathcal{C}$中，似乎没有办法在不用 Ind $\mathcal{C}$作为中间步骤时就可以恢复 G。

第九章　经典统计物理基础概论

乔斯·尤菲克

1. 引言

一般认为，建立起一个成熟的、系统化的理论版本的好处在于，人们可以忘掉它以前的发展历史。当我们以研究一个物理学理论的概念结构为目的时，这种认识当然是正确的。例如，在讨论经典力学的基础时，我们就不需要考虑巴黎经院哲学家们的工作。在讨论量子力学的基础时，我们只需要从冯·诺伊曼(Von Neumann)公理开始，可以忽略旧量子论。然而，由于统计物理还没有发展出一套能被普遍接受的形式化公理，因此我们没有别的选择，只能详述它的历史。

这并不是因为试图建立统计物理基础的尝试没有出现或者说过于不足，参见[Ehrenfest and Ehrenfest-Afanassjewa(埃伦费斯特和埃伦费斯特—阿凡纳斯捷瓦)，1912；ter Haar(特·海尔)，1955；Penrose(彭罗斯)，1979；Sklar(斯克拉)，1993；Emch and Liu(埃姆什和刘)，2001]等。相反，上述文献的研究表明，统计物理学已经发展出大量的学派，它们各自具有自己的纲领和技术手段。与量子力学和相对论不同的是，这个领域缺乏一套可以被大多数参与者接受的普遍性的假定，当然，

尽管很多假定都是重叠的。但是，一个普遍的情况是，几乎所有的学派都承认统计力学的发现之父，麦克斯韦（Maxwell）、玻耳兹曼（Boltzmann）和吉布斯（Gibbs）是他们的先驱。

从大体上理解，统计物理可以被描述为物理学中的这样一个分支，它想要描述热的行为以及大量散乱物质的性质，也就是说，与微观粒子成分及其动力学相关的宏观层面的性质。① 在这篇综述文章中，我们只涉及由经典物理学支配的包含有限数量微观成分的研究内容。关于量子统计物理的讨论请参见[Emch，2006]，其中也涉及了对无限系统的研究。

上面的描述是故意有些模糊的，以上文字尚未详细描述热的行为是什么，并且也没有说明通过哪些可能的方式能对其进行描述。我们现在稍微将它展开来叙述。统计物理学可以分为两个基本部分。第一部分是关于宏观物质系统的力学模型。例如，气体可以被模型化为质点粒子组成的系统，或者硬球组成的系统，或者混合物组成的系统等。类似地，可以用格子结构来使固体模型化等。总的来说，力学模型的特点及其动力学性质，取决于我们所关注的系统。

该理论的第二个部分，也是所有的研究方式都一致的地方，就是引入概率和统计因素。有时，教科书在解释为什么需要引入这部分内容时，会指出对于拥有大量自由度的机械模型来说，是无法得到其运动方程的精确解的。但是这种为了弥补不足而引入的动机确实低估了概率在统计物理学中所具有的建构性与解释性作用。另外稍好一些的原因，可以在很多教科书上找到，就是即使能够得到动力学方程的具体细节，但这些细节最终与描述热行为也是毫不相干的。这种说法是部分正确的，但是也很难像其所希望的那样让人满意。的确，并不是所有的细节都与微观动力学相关，比如说相位变换，但是人天生就希望得到更多的明确信息，以确切区分哪些细节是相关的，哪些是无关的。

因此，统计物理中一个最著名的基础问题就是指出并阐明概率假设在理论中的地位。正如我们将看到的，这个任务已经导致了对概率的两种不同认识的大致区别，其中一种是作为一个在力学（动力学）中被明确定义的概念，另一种

① 这里“大量散乱”的定义和“微观/宏观”的区别应该从相对的意义上理解。因此，统计物理学可以描述星系或者星云，那里的组成恒星可以被理解为“微观成分”。

是(统计力学中)一个在概念上非独立的因素。

接下来就是并不具有一致认识的部分。以下是一个(不完全的)清单。

——关于多大数量微观组分为大量的假设(典型的量级为 10^{23} 或更多)。

——关于动力学不规律性质(如遍历)的假设。

——特殊初始条件的选择。

——外界对系统的影响所扮演的角色,也即系统是否为开放的,是否能够与环境进行能量/动量交换的假设,以及在这种外界影响下,动力学敏感性的预设。

——基于微观元素排列的宏观量的对称性。

——分辨率的限制或宏观观测者的实验精确性。

——对由因果律引起的时间不对称性的吁求。

在统计物理学的方法中,以上这些部分的角色都是有争议的。很多"大师"认为是必需的、不可或缺的部分,但在另外一些人看来是不充分的或多余的。因此统计物理学基础的一个主要目标应该是将上述想法分类并整理出一个子集,能够用其以精确和合乎逻辑的方式,为统计物理学构建一个统一的且足够全面的理论框架。

另外一个先前曾模糊讨论过的问题是对热行为和宏观物理性质的一般认识。我们有两种资源来描述这个论题。其中一个就是与理论物理中的其他(旧的)分支进行比较,它们应该与统计物理学有同样的目标,但不依赖于以上的两个主要组成部分即力学模型和概率论证。两个主要的例子分别是:热力学和流体力学。另一个资源,当然就是观察,这为理论提供了丰富的现象,其中有一个具有足够的理论解释作用(如湍流)。

显然,我们要检测统计物理学有多么成功,就是为了回答这个问题:通过这样的研究方式,我们能在多大程度上重现之前非统计理论在其胜任的经验科学中得到的结果,以及在其不能胜任的领域对其进行改进。因此,统计物理的基本原理还为有关交互理论间的关系(如约化)这样的哲学观点,提供了一个测

试平台。比较[Brush（布鲁斯），1977；Sklar，1993；Batterman（巴特曼），2002]。不过我将不会探讨这个问题。在下面的部分我将主要介绍上面提到的四个方面的理论：热力学、流体力学、动力学理论及统计物理学，对它们进行一个概述。

1.1　热力学

正统热力学的研究道路是与这几个人物联系在一起的，他们是克劳修斯（Clausius）、开尔文（Kelvin）和普朗克（Planck）。在这里，人们旨在描述宏观物体的热性质，而刻意回避有关微观实体的一切假设，正是这些微观实体有可能构成了所研究的物体。相反，该方法旨在从一系列经验性原理的限制出发，获得某些对所有这类物体成立的一般性规律。

在这种方式下，宏观物体（或热力学系统）被认为是一类黑箱，可通过做功和热交换的手段与环境产生相互作用。最基本的经验性原则是，这类宏观系统，当只有自身存在时，即从环境中分离后，最终会在一个平衡态上静止下来，在该平衡态上不会再有进一步的可观测到的变化发生。此外，为简单起见，我们说均匀物体的这种平衡态可以由少量的宏观变量的值来完全描述。

其他经验性原理表明了哪些类型的过程被认为是不可能的。通过巧妙的论证，我们就可以通过这些原理证明某些量（尤其是绝对温度、能量和熵）是作为“态函数”而存在的，即对于所有这样的系统来说，在热力学平衡态空间下定义的函数。

虽然该理论关注的是过程，但是它对这类过程的描述也是极其有限的。在一般情况下，这个过程需要系统经过一系列的非平衡态，这期间热力学状态函数将不能被定义，因此也不能由该理论所提供的工具来对这个过程进行详细地描述。所以我们被限定为只能仅仅考虑某些特殊类型的过程，即那些以平衡态开始和结束的过程。甚至更特别的是这样的一些过程，它们进行得如此细微和缓慢，以至于我们能够以任意小的错误率假设在整个过程中系统一直处于平衡

态过程。后面的这种过程被称为准静态过程，或有时被称为是可逆过程。①

当然，由于平衡态在定义上是假设如果不受干扰的话就会保持平衡，因此只要有外部干预如推动活塞或去除隔板等，所有这类过程就会被触发。对于第一类过程来说，正统的热力学只能涉及初始态和最终态。第二类过程则可以被（近似地）描述为平衡态空间中的一条曲线。

该研究方法的优点是它的一般性。虽然该方法最初在19世纪末期是被用于对气体和液体进行研究，但是它也能被推广到研究磁体和其他系统的行为。事实上，该假设独立于其所研究的微观成分，这就意味着正统热力学的方法也可以且已经被应用于本质上属于量子力学的系统（如光子气体）以及更多的（比如黑洞）这样的外来客体，参见[Rovelli（罗韦利），2006]。

考虑到统计物理的基础，热力学的两个方面具有突出的重要性。首先，它面临的任务是提供一个与平衡态相对应的概念，并提供一个与热力学定律“所有处于不平衡状态的孤立系统都要向平衡态演化”相对应的定律。其次，统计物理应该能够描述热力学第二定律，即孤立的绝热系统的熵不能减少。显然，这些对应的定律应该是属于统计学的；即平均来说它们会成立，或它们会以高概率成立，但不能和热力学通常采用的一般性陈述一致。

1.2 流体力学

如果我们认为统计物理的目标仅仅是重塑热力学定律的话，那我们就错了。在理论物理中，还有许多其他的理论为刻画热行为提供了更详细但不是那么普遍的描述。一个具体的例子是流体力学或者说流体动力学。与热力学相比，流体力学确实依赖于对微观成分的假设。它将流体模型化为连续介质或连续实体。按照现在的说法，那就是场论。此外，它旨在描述某些宏观量在时间过程中，即在非平衡过程中的演化。它本身就是一个比热力学能提供更多信息且描述更详细的理论，当然，代价就是其所依据的经验被限制在流体范围内。

这里不详细地深入下去，更全面的论述请参见[Landau and Lifshitz（朗道和

① 但是，读者应该注意，热力学中的“可逆”一词有许多不同的含义。具体的讨论请参见[Uffink（尤菲克），2001]。

利弗席兹)，1987；de Groot and Mazur(德·格鲁特和马祖尔)，1961]。流体力学假设存在三种基本的场：质量密度 $\rho(\vec{x}, t)$、流速场 $\vec{v}(\vec{x}, t)$ 和温度场 $T(\vec{x}, t)$。因此也有三个基本的场方程，它们以微分形式表达了质量守恒、动量守恒和能量守恒。可惜的是，这些方程引入了更多的量：压力 $P(\vec{x}, t)$、应力张量 $\pi(\vec{x}, t)$、能量密度 $u(\vec{x}, t)$、剪切黏度 η 和体积黏度 ζ 以及导热系数 κ，它们都会由于不同的本质关系或状态方程而与基本场相互关联(依赖于具体的流体)，以使得场方程能够闭合，即使它们对于求解是有意义的。

由此产生的方程在时间反演下显而易见是不对称的。但流体力学的另一个显著特点是方程完全是闭合的。那就是说，只需要知道极少数的宏观量的信息就可以预测那些量的演变。它们的行为换个词来说就是自治的(autonomous)。这种自治性，也适用于其他理论或公式，它们用来描述远离平衡态的系统的过程：例如扩散理论、金属中的电传导、傅里叶热传导方程等。尽管具有很大的微观自由度，但是少数宏观量的演变看起来一般只依赖于这些宏观量的瞬时值。除了这类理论呈现出的时间反演不对称性以外，统计物理也应该完美地解释这些演化方程所表现出的非凡的自主性。

1.3　分子运动论

现在把注意力转向我们需要考虑的第二类理论，即依赖于对内部微观成分的假设或模型化，或者与所考虑系统的动力学相关的那些理论。如前所述，可粗略地将它们分为两个子类：分子运动论和统计力学。

分子运动论，也称为气体运动学理论、气体动力学理论、热的分子动力学理论等，其主要的出发点就是假设系统(特别是气体)由分子组成。因此热性质和热行为，是与这些分子的运动特别相关的。

分子运动论最早的现代版本是丹尼尔·伯努利(Daniel Bernoulli)(1731 年)的版本。之后近一个世纪，伯努利的工作都没有得到进一步发展。但在 19 世纪中叶它又重获关注。经克劳修斯、麦克斯韦和玻耳兹曼之手，该理论发展成为一个更一般且复杂的理论结构。克劳修斯通过考虑粒子之间的碰撞扩展了伯努利的模型，用以阐释令人惊叹的分子速度(其量级为 10^3 米/秒)与相对缓慢的

扩散率是能够相容的。然而，他没有建立一个关于碰撞及其影响的系统性理论。是麦克斯韦首先认识到碰撞往往会产生以各种不同速度运动的粒子，而不是以统一速度运动的粒子，他进而追问在一个平衡态下速度的值怎么可能是不同的。因而，麦克斯韦将概率和统计因素的概念引入到了分子运动学理论中。

从 1868 年起，玻耳兹曼进一步拓展了麦克斯韦的研究。在他 1872 年著名的回忆录中，他提到自己得到了一个分布函数的演化方程，即玻耳兹曼方程，并声称孤立气体的每个非稳定分布函数将会逐步向麦克斯韦形式，即平衡态演化。然而，沿着这个方向，玻耳兹曼提出了各种假设和理想化情况，例如忽略多粒子碰撞效应，这就将其推论的有效性限制在稀薄气体范围内，此外麦克斯韦于 1867 年提出了分子混沌假设（或磁数假设，Stoßzahlansatz，后来被称为“分子无序性假设”）。

玻耳兹曼方程或该方程的变体，是物理学家关于气体的主要理论。流体力学方程，以及其他输送方程都可以由其推出。然而，众所周知，它仅仅是一个近似方程，通常被看作是得到更精确方程的第一步。但是最重要的概念性问题在于其时间的不对称性，这凸显了一个事实，就是玻耳兹曼方程本身不可能仅仅从力学中得到。在玻耳兹曼的一生中，这导致了两个著名的反对意见，洛施密特（Loschmidt）提出的可逆性反对（Umkehreinwand）和策梅罗（Zermelo）提出的重复性反对（Wiederkehreinwand）。第三个重要的挑战，只是近期才由[Lanford（兰福德），1975]提出，它涉及的是玻耳兹曼方程与“气体系统是由哈密顿（Hamiltonian）动力学支配的力学系统”这一假设之间的一致性问题。

1.4 统计力学

分子运动论和统计力学之间只有一个模糊的界线。主要的区分标准是由埃伦费斯特夫妇（1912 年）提出的。分子运动论被埃伦费斯特夫妇称为“统计力学研究的早期公式化”或“分子的运动统计学”。这里，分子的状态，尤其是其速度，被看作是随机变量，概率被认为适用于这种分子状态的运动。这些概率本身是由整个气体系统的状态决定的。它们或者被设想为处于某个特定状态下的分子的相对数量，或者被看作分子处于该状态上经过的相对时间（麦克斯韦选择的是第一种观点，玻耳兹曼则在两者之间摇摆不定）。重要的是我们要强调

在这两个观点下，被谈论到的“概率”都是由气体的力学性质决定的。因此，在这种方法中，力学和统计概念之间确实并没有一个明确的分界。

逐渐地，理论出现了一些变化，埃伦费斯特夫妇称之为“统计力学研究的现代公式化”或“气体模型的运动统计学”，或者就是现在我们所认为的统计力学。在后一种方法中，概率不再被应用到单个分子的状态上，而是被应用于整个气体系统的状态。因此，气体的状态不再决定概率分布的情况，而是自身成为了一个随机变量。

后一种方法的优点是，分子间的相互作用也会被纳入考虑的范围。事实上，这种方法的研究对象不一定局限为气体，在原则上也可以适用于液体或固体。(这是“气体理论”这个名称之所以被遗弃的原因。)然而，付出的代价是概率自身变得更加抽象。由于概率被用于描述整个系统的力学状态，因此它不再由该力学状态决定。相反，在统计力学中，概率通常被设想为是以“系综”的方式来确定的，即由一个虚拟出来的以我们所谈论的系统为模型的复制系统的集合决定的。但是，不管我们想把这种结构指定为什么角色，最主要的一点是目前概率是一个独立概念，不能再被退化为系统的力学性质。

在历史进程中指出这一转变是不容易的，这可不像说麦克斯韦在 19 世纪 60 年代的工作肯定属于第一类，吉布斯 1902 年的作品属于第二类那样简单。玻耳兹曼自己的工作是介于二者之间的。他早期的贡献显然属于气体运动学理论(虽然他在 1868 年的论文中已经将概率应用于整个气体系统)，他在 1877 年之后的工作，通常被视为是统计力学理论中的一部分。然而，玻耳兹曼自己从来没有明确区分这两个不同的理论，任何想将他的工作划出一个确切分界的尝试看起来都有点太武断了。

以一种概念性的观点看，从气体分子运动论到统计力学的转变提出了两个主要的基本性问题。第一个问题是，我们选择一个特定系综的根据是什么，或概率分布是这个系综的特征的理由是什么？吉布斯没有系统地讨论过这个问题，只是讨论了平衡系综(即正则系综或微正则系综等)这个特例，对该特例来说，概率分布由一些特殊简单的形式所规定。第二个问题是如何将基于系综的概率与早期在单个气体模型的运动学方法下得到的概率联系起来。

埃伦费斯特夫妇在其论文[1912]中首先认识到这两个问题，并提供了部分答案。即采用玻耳兹曼的一个特定假设，他们称之为遍历假设。他们指出，对于一个孤立系统，微正则分布是唯一稳定的概率分布。因此，如果我们要求描述热平衡的孤立系统的一个系综必须由一个稳定的分布来表示，那么符合目的的唯一选择就是微正则分布。同样地，他们指出，根据遍历假设，无限时间平均和系综平均是相同的。那么，这就为早期气体分子运动学方法下的概率与基于系综的概率提供了一种关联，至少在平衡态下以及在无限时间极限下如此。然而，埃伦费斯特夫妇同时对遍历假设的正确性表示强烈的质疑。这种质疑很快就被证实了，1913年，罗森塔尔(Rosenthal)和普兰彻瑞(Plancherel)证明了在真实气体模型中这个假设是站不住脚的。

因此，埃伦费斯特夫妇对玻耳兹曼工作的重建突出了遍历假设的作用，表明其在玻耳兹曼的思想中，其扮演了一个基本且稳定的角色。尽管这一观点确实为他多方面的工作提供了一个一致性的认识，但从历史的角度来看它肯定是不正确的。玻耳兹曼自已也对这一假设保持严重的疑虑，并尽可能地避免使用它，特别是在他1872年和1877b的两篇重要论文中反映出这一点。从埃伦费斯特夫妇开始，许多作者在文章中讨论了玻耳兹曼的工作，尤其重要的是克莱因(Klein)[1973]和布鲁斯[1976]。

尽管如此，埃伦费斯特夫妇的分析确实导致了一个关于统计物理学基础的且多少有些清晰的描述方案或者说观点的产生，其中遍历是一个重要的特征。原先遍历假设的失败并没有使这个方案停下来，这个假设被替换为另一个(较弱)的假设，即该系统是“度量可递的”(现在，“遍历”经常被用作其同义词)。更重要的是，伯克霍夫和冯·诺伊曼的特定数学成果(遍历定理)表明，对于这个新意义下的遍历系统，以一些数学上的限制性条件为模，所期望的结果确实可以被证明，而这些限制性条件起初并没有引起足够的重视。

因此，出现了一种关于统计力学基础的遍历观点或“标准”观点，参见[Khinchin(辛钦)，1949，44]。在这种观点下，统计力学的形式体系出现了：一个具体的系统，比如说一个装有气体的容器，被描述成一个具有非常大自由度的力学系统。所有物理量都是系统动力学变量的函数，或者换个说法是其相空间的函数。然而，对这类物理量的实验或观测却不记录这些物理量的瞬时

值。相反，每一次观测都需要持续一段时间，在人类的标准下这段时间可能非常短，但在微观层面上是相当长的，即在其间微观态经历了许多变化，例如持续不断的分子碰撞。因此，我们可以记录的只是物理量在很长一段时间内的平均值。因此，这些平均数在经验上是有意义的，不幸的是在理论上和分析中它们是难以控制的。时间平均值依赖于轨迹，且只能通过对运动方程的积分来计算。在一个给定的系综上，相位函数的期望值，即相位平均则具有相反的特征，也就是说，它很容易计算，但不会立即由经验得到。然而，遍历保证了这两个平均值(几乎处处)是相等的。因此，可以将二者的优点结合，将理论上的便利性与经验上的有意义性等同起来。

统计力学显然是一个比分子运动论更强大的理论，同热力学一样，它在解释气体和其他处于平衡态的系统以及为其建构模型方面是尤其成功的。非平衡态统计力学却仍是一个问题重重的领域。

1.5　简介

本章的结构如下。第 2 节中，我将简单阐述一下正统热力学，并在第 2.2 节中简单回顾一些非正统热力学的研究方法。第 3 节关注的是气体的分子运动理论，重点是麦克斯韦 1860 年和 1867 年的奠基性论文，考察了麦克斯韦概率论证的意义和地位。

第 4 节(有选择性地)专注于玻耳兹曼的工作，如前面提到的，可视为分子运动论和统计力学之间的那部分研究。重点将是他 1868 年的论文以及他最著名的 1872 年和 1877 年的论文。此外，来自洛施密特［1877］和策梅罗［1897］的反对意见也会和玻耳兹曼的回应放在一起讨论。我们的讨论强调的是玻耳兹曼这些年来使用的假设和方法的多样性，如分子混沌假设、遍历假设、1877 年的组合论证以及他的研究结果的开放性，并对他在 19 世纪 90 年代提出的 H 定理进行统计性解读。

接下来，第 5 节介绍了吉布斯版本［1902］的统计力学，并强调他和玻耳兹曼研究之间的本质区别。第 6 和第 7 节概述了统计力学近期的一些发展，特别是，我们回顾了现代遍历理论的一些成果，以及旨在发展一个更加系统化的非平衡理论的研究方法，如 BBGKY 方法和兰福德方法。BBGKY 由博戈柳博夫、

波恩、格林、柯克伍德和依云（Bogolyubov，Born，Green，Kirkwood and Yvon）的首字母命名。第7节通过这些方法的结合扩展了讨论的内容，这里统一在随机动力学的名义下，其中包括那些被称为“粗粒化”和“干涉主义”或“开放系统”的方法。在所有的情况下，我们应当注意的问题是这些方法在论及非平衡态时是否令人满意，以及它们是如何成功的。

正如本前言所解释的，讨论的主题是精心挑选过的。还有许多重要的论题以及统计物理学基础的发展过程我将不会涉及。在此我为那些希望了解更多的读者们列出其中最引人注目的一些观点及参考文献。

——麦克斯韦妖和朗道原理[Klein，1970；Earman and Norton（厄尔曼和诺顿），1998；1999；Leff and Rex（利夫和瑞克斯），2003；Bennett（本内特），2003；Norton，2005；Maroney（马罗尼），2005；Ladyman *et al.*（雷德曼等），2006]。

——玻耳兹曼在18世纪80年代的工作（例如单环系统）[Klein，1972；1974；Bierhalter（比尔海特），1992；Gallavotti（格拉瓦蒂），1999；Uffink，2005]。

——吉布斯悖论[van Kampen（范·坎朋），1984；Jaynes（杰恩斯），1992；Huggett（哈格特），1999；Saunders（桑德斯），2006]。

——分支系统[Schrödinger（薛定谔），1950；Reichenbach（莱辛巴赫），1956；Kroes（克罗斯），1985；Winsberg（文斯伯格），2004]。

——统计力学中概率的主观解释[Tolman（托尔曼），1938；Jaynes，1983；von Plato（冯·柏拉图），1991；van Lith（范·利兹），2001a；Balian（巴里安），2005]。

——普里高津和布鲁塞尔—奥斯丁学派[Obcemea and Brändas（奥博塞米和巴尔达斯），1983；Batterman，1991；Karakostas（卡拉特斯特），1996；Edens（伊登斯），2001；Bishop（毕肖），2004]。

2. 正统热力学

2.1 克劳修斯—开尔文—普拉克方法

热力学是一个旨在通过宏观可观测量(通常是温度、压强、体积等)来描述宏观物理对象特征，以及在特定类型的相互作用(通常是与环境发生热交换或做功)下描述其变化的理论。

该理论的经典版本，出现在1850年左右，被视为是一种方法论的起点，其观点是该理论的基本法则应该独立于与所研究物体的微观成分有关的任何特殊假设。它们是基于经验性原理，即在实验事实基础上做出的大胆的一般性陈述，而不是基于假设性的因而是不可检测的假定，如原子假设。

这种方法论形成的原因是双重的。首先，科学目标的主流观点是实证主义者的经验哲学，它极度推崇超越理论假设的可直接经验验证的陈述。但实证主义观点的影响从来没有如此彻底，以至于物理学家全部都避免进行猜测。事实上，热力学的许多主要创始人，热切地沉迷于自己做出的关于物质的微观物理成分的独特假说。

第二个原因更加务实。众多的微观物理假说和猜测在19世纪中叶就已经那样出色，而在它们之间通过抉择而胜出的前景是如此暗淡，因此，不依赖于这样的假设来得到并呈现结果的方法具有很明显的优势。克劳修斯在1857年表示，他坚决相信关于气体性质的运动学观点，他还提到，此前他没有透露过这种观点，是为了不至于将这种信念与他在热力学性质研究方面开展的工作相混淆[Clausius, 1857, 353]。①

如果不按历史进程②，人们可能会说，热力学的第一个核心概念是平衡态。这样的认识被视为一个经验事实，即一个有限体积内的宏观客体，当只有其自身存在时，即从环境中被孤立出来后，最终会在一个固定的状态上静止下来，

① 如果我们比较他与朗肯的声望，就会清晰地体会到做出这种选择的明智性。朗肯其实在克劳修斯之前就在寻找熵函数(他称之为“热力学势能”)。然而，这一结果在很大程度上被忽视了，由于它植根于朗肯关于原子旋涡的相当复杂的理论中。

② 有关详细资料我参阅了[Uffink, 2001]。

在该态上再不会有进一步的可观测到的变化发生。（“负第一定律”，参考第2.1节末尾的备注中提到的“第零定律”。）这一静止状态，被称为（热的）平衡态。此外，为简单起见，对于均匀物质，这个态可以由少数几个宏观变量的值完全描述。尤其是对于流体（气体或液体）来说，两个独立的变量就足以确定平衡状态。

因此对于流体有三个变量：压强 p、温度 θ 和体积 V，由一个所谓的状态方程联系起来，其中，按照欧拉(Euler)的观点，习惯性地将压强表达为剩下两个变量的函数：

$$p=p(\theta, V) \tag{1}$$

这个函数的形式随流体的变化而不同，对于 n 摩尔理想气体来说，其为：

$$p(\theta, V)=nR\theta/V \tag{2}$$

这里 R 为气体常数，θ 为热力学温标下的温度值。

热力学内容是在三方面的研究基础上发展起来的。首先是量热学，在18世纪其理论就已经发展到完善的程度，主要是由约瑟夫·布莱克(Joseph Black)完成的[Fox（福克斯），1971；Truesdell（特鲁斯德尔），1980；Chang（张），2003；2004]。它主要研究物体在对系统吸热或放热过程中的温度变化。当然，这里（默认）的前提是，这种热交换的过程是如此地细微和缓慢，以至于系统总是可以被视为处于平衡态。用现代术语来说，它的过程是“准静态”的。因此，状态方程在这个过程中是成立的。

量热学的工具是微积分。对于流体吸收的无限小的热量 $đQ$ 来说，我们有：

$$đQ = c_V d\theta + \Lambda_\theta dV \tag{3}$$

其中 c_V 被称为定容积下的热容量，Λ_θ 为恒定温度下的潜热。c_V 和 Λ_θ 都被假设是 θ 和 V 的函数。表示热增量 $đQ$ 的符号 $đ$ 不必是一个正合微分，即并没有将 Q 视为是一个状态的函数。

因此，在某个过程中流体吸收的总热量 Q，可以被表示为一个在平面 (θ, V) 上沿路径 P 的线积分：

$$Q(\mathcal{P}) = \int_{\mathcal{P}} đQ = \int_{\mathcal{P}} (c_V d\theta + \Lambda_\theta dV) \tag{4}$$

与上述的讨论类似，我们可以研究比流体更一般的热物体中的准静态热交换。

事实上，量热学已经足够一般化，可以用来描述相变过程，比如说从水到冰，只要假设 Λ_θ 是不连续的。

所有这一切都与热本身是否是一个实体无关。事实上，在这个问题上布莱克自己是希望保持中立的。即便如此，量热学中的术语多少还是内含了热是一种实体这样的假定，通常被称为热质，许多18世纪和19世纪初的研究者都采纳了这一观点[Fox，1971]。在这种观点下，谈及物体中包含的热量是有意义的，这就要求 $đQ$ 是一个正合微分，或者换句话说，对于相同的初始点和最终点来说，所有路径 $\mathcal{P}$下的积分 $Q(\mathcal{P})$ 都是相同的。但这在经验上是错误的，如果我们也考虑做功的效果的话。

19世纪40年代焦耳(Joule)和迈尔(Mayer)等人的研究，导致了一种观点，即热和做功是"等价的"，或更确切一点说，在每一个循环过程 $\mathcal{C}$中，系统吸收的热量 $Q(\mathcal{C})$ 与系统所做的功是成正比的。或者我们可以将对系统做的功$W(\mathcal{C})$看作是正功：

$$JQ(\mathcal{C})+W(\mathcal{C})=0 \tag{5}$$

这里 $J\approx4.2$ 牛·米/卡为焦耳常数，现在习惯于取该常数为1。这就是所谓的热力学第一定律。

对可逆过程来说，它还可以被表示为热力学平衡态的一个态空间 Ω_{eq} 上的线积分：

$$\oint_{C}(đQ+đW)=0 \tag{6}$$

其中：

$$đW=-pdV \tag{7}$$

假设式(6)对于平衡态空间上的所有闭合路径都成立，那么必然存在 Ω_{eq} 上的一个函数 U 使得：

$$dU=đQ+đW \tag{8}$$

热力学的第三个部分产生于对热和做功间关系的研究，具体说就是热机效率的研究。1824年，卡诺得到下面的定理。

考虑一个执行以下循环 $\mathcal{C}$的任意系统：(a)从一个温度为 θ_+ 的储

热器中吸收热量 $Q_+(C)$；(b)将热量 $Q_-(C)$ 释放到温度为 θ_- 的储热器中，其中 $\theta_- < \theta_+$；(c)在循环的其他阶段没有热交换；(d)对第三个物体做功 $W(C)$。令 $\eta(C) := \frac{W(C)}{Q_+(C)}$ 为热机的循环效率。那么：

(1)所有可逆循环的效率相同。这个效率是两个温度的一个普适函数，即：

$$\eta(C) = \eta(\theta_+, \theta_-) \tag{9}$$

(2)其他所有循环的效率小于或等于可逆循环的效率。

卡诺假设热是一个守恒的物质，从而得到这个结果，因此，对所有的 C 来说，$Q_+(C) = Q_-(C)$，它也被看作这样一个原理，即(第一类)永动机不能实现。

事实上，卡诺并没有使用可逆/不可逆这种二分法来描述他的定理的两个部分。①

他实际上用了两种不同的方式来表述循环，其中会产生一个最大的效率值。(a)在他证明卡诺循环属于第(1)类时，他的关键性假设是，该循环“可以在相反方向上以相反的顺序进行”[Carnot，1824，11]。但在后文中，他提出了产生最大效率循环的充要条件。(b)在所有涉及热交换的阶段，只有温度相同的物体之间存在热接触，或者更确切地说，它们之间的温度差微乎其微。

卡诺定理是引人瞩目的，因为它并不需要假设该循环运行所处的热系统的任何性质。因此，当他的工作首次被物理学界的先驱(汤姆逊，后来被称为开尔文勋爵，1848 年)知道的时候，就被认为是通向研究热交换和做功的一般性理论，即开尔文命名为“热力学”理论的一条重要线索。事实上，开尔文在他关于该主题的第一篇论文(1848 年)中就已经表明卡诺的普适函数 η 可以用来给出一个温度的绝对量度，即一个不依赖于任何特定物质属性的量度。

不幸的是，同时期的研究表明卡诺关于热守恒的假设是违反第一定律的。克劳修斯和开尔文发表了一系列的论文，在不同的基础上重建了卡诺定理，即在式

① 事实上，[Truesdell，1980]认为他的定理的表述是不正确的。进一步的讨论请参见[Uffink，2001]。

(5)表示的第一定律基础上，或在 $Q_+(C) = Q_-(C) + W(C)$ 且排除第二类永动机的原则下，并将他的成果转化为研究一般热力学系统及其在热和做功的影响下如何变化的一般性论题。在大多数情况下，这些研究仅涉及卡诺定理的第一部分。他们的研究揭示了现在被称为第二定律的第一部分的内容，其内容如下。

首先，开尔文重新将他在 1848 年提出的绝对温标形式化为 $T(\theta)$，其中普适效率可以被明确表示为：

$$\eta(T_+, T_-) = 1 - \frac{T_-}{T_+} \tag{10}$$

这里 $T_i = T(\theta_i)$。由于效率 η 还能被表示为 $W/Q_+ = 1 - (Q_- / Q_+)$，这等价于：

$$\frac{Q_-}{T_-} = \frac{Q_+}{T_+} \tag{11}$$

接下来，改变 Q 的符号约定，当吸收热量时 Q 为正，释放热量时为负，并推广为包含任意多储热器的一个循环，这样就得到：

$$\sum_i \frac{Q_i}{T_i} = 0 \tag{12}$$

在这种情况下，系统被看作是一个可逆循环，其中在循环过程中储热器的温度连续变化，这就可推广为：

$$\oint_C \frac{đQ}{T} = 0 \tag{13}$$

这里，T 仍然表示与系统发生相互作用的储热器的温度，而不是系统自身的温度。然而，卡诺对于可逆性判据的充要条件要求，在涉及热交换过程的所有阶段，储热器和系统的温度都应是相同的。因此，在这种情况下，我们可以将 T 等同于系统自身的温度。

这个结果的好处是积分式(13)现在可以完全由系统的量表示。克劳修斯在 1865 年通过应用这个著名的定理，指出存在一个定义在系统热平衡态上的被称为熵 S 的函数，满足：

$$S(s_1) - S(s_2) = \int_{s_1}^{s_1} \frac{đQ}{T} \tag{14}$$

或者，以它更著名的形式：

$$\frac{đQ}{T} = dS \tag{15}$$

这个结果经常被表述为：$đQ$ 有一个积分因子（即 T），被 T 除后，非正合（不完备的，不可积的）微分 $đQ$ 变成了正合（完备的，可积的）微分。对 1 摩尔理想气体，即 c_V 为常数、Λ_θ 为零且满足式（2）理想气体定律的流体来说，我们会发现，例如：

$$S(T,V) = c_V \ln T + R\ln V + 常数 \tag{16}$$

这个熵函数的存在，也使得式（8）可逆过程的第一定律可以方便地改写为：

$$dU = TdS - pdV \tag{17}$$

这里又是用我们所讨论的系统的属性表示的。

然而，第二定律前半部分本身的重要性，从来就没有引起太多争议。相比之下，上述结果的扩展覆盖了卡诺定理的第二部分，却引起了相当多的思考，而且还与对"（不）可逆过程"的理解密切相关。

卡诺定理的第二部分最初并没有引起很多关注。克劳修斯[1854]只用了一小段内容说明它，得到了一个有关"不可逆"循环的结果：

$$\oint \frac{dQ}{T} \leqslant 0 \tag{18}$$

但这个结果非常不容易运用，因为这里 T 指的是与系统相关的储热器的温度，不（一定）是系统本身的温度。

在 1865 年的论文中，克劳修斯将不可逆过程放在了更加突出位置。如果一个不可逆循环过程是一个一般过程，既包括一个可能是非准静态（记为 non-qs）的阶段，从 s_i 到 s_f，也包括一个准静态（记为 qs）的阶段，从 s_f 回到 s_i，那么我们可以把式（18）写为：

$$\int_{s_i}^{s_f}{}_{\text{non-qs}} \frac{đQ}{T} + \int_{s_f}^{s_i}{}_{\text{qs}} \frac{đQ}{T} \leqslant 0 \tag{19}$$

将式（14）应用到左边第二项，我们得到：

$$\int_{s_i}^{s_f}{}_{\text{non-qs}} \frac{đQ}{T} \leqslant S(s_f) - S(s_i) \tag{20}$$

此外，如果我们假设一般的非准静态过程是绝热的，即 $đQ = 0$，结果就是：

$$S(s_i) \leqslant S(s_f) \tag{21}$$

换句话说，在任意绝热过程中，末态的熵不能小于初态的熵。

备注：1. 循环积分$\oint$和非正合微分$đ$都是现代的符号。克劳修斯以及他之后的玻耳兹曼，通常会把式(13)和式(18)的左边写为$\int \frac{dQ}{T}$。

2. 需要注意的一个重点是，严格来说，第二定律的克劳修斯形式化并不要求在时间进程中的绝热孤立系统的熵是普遍地单调增加的。事实上，在正统热力学中，熵的定义仅适用于平衡态。因此，在该理论下，考虑在非准静态过程中系统的熵如何变化是没有意义的。在一般意义下，我们能说的是当系统从一个平衡态开始，经过一个绝热过程后，再次在一个平衡态结束，后面那个状态的熵不低于前者。

尽管如此，第二定律往往被理解为，随着时间进程熵必须是连续地单调增加的；对于绝热孤立系统，经常被表示为更严格的形式：

$$\frac{dS}{dt} \geqslant 0 \tag{22}$$

然而，在正统的热力学中这种要求并没有根据。

3. 对第二定律的另一种常见的误解是，它只要求孤立系统的熵不会减少。应当指出，这仅仅是克劳修斯得出的部分结果：第二定律可以推广到一般绝热过程，即从始至终系统一直保持绝热的过程。换句话说，除了热交换，系统可以与环境之间发生其他任意的相互作用。（例如搅拌保温瓶中的液体，就像焦耳的“桨轮”实验那样。）

4. 另一点需要注意的是，克劳修斯的结果，即绝热孤立系统的熵永不减少来自这样一个假设：我们可以找到一个可逆过程，它能连接终态和初态，以完成一个循环。事实上，如果不存在这样一个过程，这两个状态下熵的差异则不能被定义。这样的准静态过程的存在对许多预想中的应用是不成问题的（如假设s_f和s_i是流体的平衡态），但在更一般的情况下，可能并不是很明显（如假设我们考虑一个复杂系统如在一个活细胞中远离平衡态的过程）。因此，[Kirchoff（基尔霍夫），1894，69]曾提出一个警告，即熵是否增加取决于准静态过程是否存在。

5. 除了众所周知的热力学第一和第二定律以外，后来的先驱者们还发现了

一些更基本的假设或经验原则，它们在理论中往往是作为传统而被默认的假设——或有时是明确的但没有被命名——这类假设可能也需要被放到类似的基础位置上。

其中我们最熟悉的是所谓的第零定律，它由福勒和古根海姆(Fowler and Guggenheim)(1939 年)命名。为了将其引入，我们考虑热平衡的关系。这指的是两个系统的平衡态之间的关系，如果我们发现两个系统之间存在直接热接触——即只发生热交换的相互作用，那么无论何时，由这两个系统组成的复合系统将一定会处于一个平衡态。第零定律现在是一个关于传递关系的假设，即如果对于物体 A 和 B 的状态该定律成立，且对于物体 B 和 C 的状态也成立的话，那么它也同样适用于物体 A 和 C。①

2.2 热力学的次正统版本

即使在正统的热力学框架内，也有不同于克劳修斯—开尔文—普朗克方法的研究方法。其中最重要的无疑是吉布斯在 1873 年至 1878 年间所提出的方法[Gibbs，1906]。吉布斯的方法在思想上与他的欧洲同行们并不相同。他并没有将热力学状态变量 U，T 或 S 的存在性或唯一性与经验性原理联系起来，而只是简单地假设它们存在。此外，吉布斯的重点在于对平衡态而非对过程的描述。

前人通常认为变量的选取是为了更好地使热力学量的表示符合日常习惯，如热力学(平衡)态空间上的坐标系统的选取。对于流体，人们可以等价地选择

---

① 事实上，仅仅通过传递性是不足以对其进行描述的。实际需要的假设是，热平衡是一个等价关系，即它是可传递的、相互的和对称的，参见[Boyling(博伊林)，1972，45]。可以将这一认识提升为一个基本“定律”的思想是，这一基于温度概念的假设依据的只能是经验层面的现象。

另一个这样的假设也经常会被提到但很少会以此命名，那就是任何存在与有限的体积中的系统，如果使其孤立，就会趋向于演化为一个平衡态，这有时也被称为“第零定律”，参见[Uhlenbeck and Ford(乌伦贝克和福特)，1963，5；Lebowitz(莱博维兹)，1994，135]，从而与福勒和古根海姆命名的对象产生冲突。因此[Brown and Uffink，2001]也提出将其命名为负第一定律。请注意，这个假设已经引入了一个明确的时间不对称性，这是在比第二定律更深的层次上说的——而不是根据第二定律得到的。然而，大多数 19 世纪(和许多 20 世纪)的学者们并不重视这种区别，因此就正如下面我们将看到的，这个负第一定律往往被归入第二定律。

变量(p, V),(V, T)等,因为它们是独立的且都能表示唯一的热力学平衡态。① 因此我们能够以任意此类变量等价地表示量 U, S 等。不过,对于哪些选择比其他选择"更好",吉布斯有深刻的见解。例如,将熵作为能量和体积的函数$S(U, V)$,或将能量作为熵和体积的函数 $U(S, V)$,那么其他所有热力学量可以由它决定,但是在其他选择下这一般不成立。例如,如果我们只知道对于1摩尔气体来说 $S(T, V)$ 可以由式(2)给出,那么我们就推不出状态方程 $p = RT/V$ 和 $U = c_V T$。相反,如果给出函数 $S(U, V) = c_V \ln U + R \ln V$ + 常数,那么这些方程就能通过求其偏导数得到:$\frac{p}{T} = \left(\frac{\partial S}{\partial V}\right)U$ 和$\frac{1}{T} = \left(\frac{\partial S}{\partial U}\right)V$。

正因为如此,吉布斯将下式

$$dU = TdS - pdV \quad 或 \quad dS = \frac{1}{T}dU + \frac{p}{T}dV \tag{23}$$

称为基本方程。② 当然,这并不意味着选择其他的变量是不好的。相反,我们也能为这样的变量对找到等价的基本方程,通过对 U 进行勒让德变换。即(T, V)对应亥姆霍兹(Helmholtz)自由能 $F = U - TS$,(p, S)对应焓 $U + pV$,(p, T)对应吉布斯自由能 $U + pV - TS$。此外,吉布斯将这些因素由均质流体扩展到由多种化学成分和物理相位构成的非均匀体。

另一个主要的创新之处在于,吉布斯提出了变分原理来将稳定平衡态与不确定和不稳定平衡态区分开来。粗略地说,这一原理要求对于稳定平衡态,函数$S(U, V)$ 应该是凹的。这个原理在描述热力学系统的相变时有很大的价值,例如:范·德·瓦尔斯(Van der Waals)气体,被麦克斯韦用来得到其著名的"麦克斯韦结构"或等价的面积律[Klein, 1978]。吉布斯的工作被证明在化学热力学和物理化学的发展过程中也起到了重要的作用。

---

① 后面的情况也可能是错误的:类似水的液体就存在不同的平衡态,它们有相同的 p 和 V,但有不同的 T,见[Thomsen and Hartka(汤姆森和哈塔卡), 1962]。

② 注意吉布斯的观点和克劳修斯—开尔文—普朗克的观点的不同之处在于这些学者将式(23)看作热力学第一定律的表达式,将微分解释为可逆过程中的无穷小增量,可与式(17)比较。另一方面,在吉布斯看来,式(23)并不表示一个过程,而是热力学平衡态空间上的一个微分方程,它的解 $U(S, V)$ 或 $S(V, U)$ 包含了有关该系统平衡属性的所有信息,包括状态方程、比热和潜热、压缩性等——它不仅仅是第一定律。

正统热力学的另一类方法是特别专注于为理论建构一个更严格的形式框架。这通常被称为公理热力学。当然，选择在一个公理框架内研究物理理论本身并不能反映出任何在物理假设或哲学观点上的偏好。然而，克劳修斯—开尔文—普朗克方法依赖于经验性和直观概念，在经验上看起来似乎足够清楚——但令人惊讶的是它们往往很难定义。因此，公理化方法企图将这些经验性原理替换为概念上更加精确、但也更抽象从而也更脱离经验的陈述。第一个例子是卡拉特欧多(Carathéodory)于1909年所做的工作。后来的公理化方法由其他很多人发展，如[Giles，1964]、[Boyling，1972]、[Jauch，1972；1975]和[Lieb and Yngvason，1999]。所有这些方法都因他们选择的原始概念和他们制定的公理而不同，因此也因他们所取得的成果和实现的目标而不同。然而，粗略地看，我们可以说，他们都特别注重于论证在什么条件下我们能保证在一个适当的结构下熵和其他状态函数的数学存在性和唯一性。

20世纪40年代以来，在我们称之为“非平衡热力学”或“不可逆过程热力学”的物理学方面，很多人做了大量的工作，如[de Groot，1951；Prigogine，1955；de Groot and Mazur，1961；Yourgrau(尤格诺)，1966；Truesdell，1969；Müller，2003]。这类工作的目的是将正统热力学朝描述非平衡态系统的方向延伸。通常，我们假设，热力学量被表示为空间和时间中的一个连续的变量场，而平衡条件在热力学系统内的每个无限小区域内成立。此外，我们还应注意，这个领域的研究者看起来能被划分为几个不同的学派(他们使用的名称如“广义热力学”“一般热力学”“理性热力学”等)，且相互之间并不认同，参见[Hutter and Wang（胡特尔和王），2003]。

这类工作已经产生了许多成功的应用。但是公正地说，迄今为止，几乎所有的注意力都集中于实际应用。例如，公理化热力学试图去回答这类问题(例如在什么情况下我们能表明在形式上应用的非平衡量是存在的和唯一的)，这些问题在很大程度上是没有解决的，而且也确实引起一些质疑，可比较[Meixner(迈克斯纳)，1969；Meixner，1970]。该理论的另一个固有的限制是，按照非平衡态至少在一个无限小的局部区域内可以近似为平衡态的假设，这种方法对于远离平衡态的环状系统(如湍流或活细胞)来说是无能为力的。

应该提到的是最后一种类型的方法，即统计热力学的方法。这里的基本思

路是，在仍然不引入关于热力学系统微观成分的假说的同时，我们拒绝接受正统热力学的关键假设，即一个平衡态就是所有量达到恒定值的态，以适用于如布朗运动或热噪声等的涨落现象。因此这种想法使得至少有一些热力学量被描述为随机量，它们在时间进程中，以不同的概率得到不同的值。在这个方向上的工作主要由西拉德(Szilard)[1925]，曼德尔布罗(Mandelbrot)[1956；1962；1964]，蒂萨和奎伊(Tisza and Quay)[1963]完成。

当然，关键的问题还是如何选择合适的概率分布。方法之一主要由蒂萨[1966]提出，他是从爱因斯坦[1910]的论文中获得灵感的，该方法依赖于玻耳兹曼原理的倒置：玻耳兹曼认为，在统计力学中，热力学熵的概念应该与概率的对数有关；而爱因斯坦认为热力学已经给出了熵的概念，我们可以通过两个平衡态的熵差的指数来定义这两个平衡态的相对概率。其他方法从数理统计中借用了更加复杂的结果。例如，曼德尔布罗借用了皮特曼—考夫曼—戴蒙斯定理，该定理指出足够的估计量只对于概率分布的“指数族”而存在，他用该定理从能量是系统温度的充分估计量这个假设出发推出了典型的概率分布，另见[Uffink and Van Lith，1999]。

3. 分子运动论——从伯努利到麦克斯韦

3.1　19 世纪中期的概率

概率论的历史可以追溯到统计物理学出现前的至少两个世纪。通常，我们认为这一理论诞生于 1650 年左右帕斯卡和费尔玛的书信中。经过雅克布·伯努利(Jacob Bernoulli)[1713]，亚伯拉罕·德·美弗(Abraham de Moivre)[1738]和皮埃尔—西蒙·德·拉普拉斯(Pierre-Simon de Laplace)[1813]等人的努力，它被提炼成一个成熟的数学学科，参见[Hacking(汉克)，1975]。

在这一传统中，概率通常被称为“经典概率”，其概念被看作我们对信仰的确定性程度。这里要注意非常重要的两点。首先，在这个特定的认识下，概率是驻留在头脑中的。自然界中是不存在不确定性或机会性事件的。事实上，经典传统下的所有作者，都强调他们坚持严格决定论，无论是诉诸于神的无所不知(伯努利，德·美弗)还是借助于力学定律和初始条件(拉普拉斯)。因此概

率表示对某些事件状态的判断，而不是这些事件状态本身。因此，古典学者们从来不认为概率在描述自然或物理过程的诸如此类的事件时会发挥任何作用。①第二，虽然伯努利自己使用“主观”一词强调了概率涉及的是我们自己，以及我们所拥有的知识，但是对于概率论主观解释的追随者们来说，经典解释是不能成立的，他们把概率看作一个主观的(尽管是理性的)人的信仰的程度，而这个人的信仰可能基于他一时的兴致、他的偏见以及他的个人观点。

当然，如果不是伴随着一些在具体案例中为概率赋值的规则的话，概率的经典概念将一点都不会流传下来。唯一可用到的规则是所谓的“不充分理由原则”：当我们没有理由相信某个事件会实现而另外一个不会时，我们应该给它们分配相同的概率，参见[Uffink，1995]。与之密切相关的另一个版本的规则是，两个或两个以上的变量应该是独立的，当我们没有理由相信它们会相互影响时。

从1650年至1813年这整个时期内，经典的观点是占主导地位的，确切地说是唯一存在的对概率的认识，然而大约从1830年开始它受到了侵蚀。对此有那么几个原因，但最重要的也许是——看起来很矛盾——伴随着巨大的成功该理论被应用到了多个学科。在拉普拉斯有影响力的《概率论哲学》(*Essai philosophique sûr les Probabilités*)的唤醒下，科学家们发现了概率论在司法裁决、人口统计、社会科学和遗传研究等方面的应用价值。事实上，有人可能会说：除了物理以外，概率几乎无处不在，参见[Hacking，1990]。在大规模现象的频繁事件中，人们发现了惊人的规律性，观测数据(比如说每秒滴到瓦片上的雨滴的数量与普鲁士军队中每年被马踢死的士兵人数相同)导致了另一种观点的出现，即概率不再是对主观(不)确定性的表征，反而是自然界中一种独特的规律性的表现(泊松，凯特莱)。在这段时期之后，我们发现了概率定律的思想，即概率论的定理反映出了自然所坚持的如定律那样的行为。在这种关于概率的替代性的频率观点中，不充分理由原则再也没有立足之地。相反，确定概

① 可以将丹尼尔·伯努利作为一个例子。他非常熟悉他的叔叔雅各布在概率方面的工作，而且他自己事实上也是18世纪最著名的概率学家之一。然而，在他关于气体分子运动论的工作(将在3.2节讨论)中，他没有找到任何机会将他自己所研究的这两个领域的专业知识联系起来。

率值的显而易见的方法是收集关于事件发生频率的经验数据。然而，在这个世纪结束之前——(可以说)在1919年之前都没出现一个能明确替代概率的经典概念的方案，然后在这之后的几年内出现了不少于3种的替代方案：凯恩斯的逻辑解释、冯·米塞斯(von Mises)的频率解释以及拉姆齐(Ramsey)和德·费耐迪(De Finetti)的主观解释。更详细的论述请参见[Fine，1973]、[Gallavotti，2004]或[Emch，2005]。

大致可以总结如下，我们可以说在1850年左右概率领域处于一种不稳定和混乱的状态。只存在两种相互竞争的观点，经典解释与频率解释，并且还经常会令人困惑地混在一起。庞加莱(Poincaré)[1896]的一句名言准确地描述了当时的结果，所有的数学家似乎都相信概率性规律来自对经验的陈述，而所有的自然科学家似乎都认为它是由纯数学推导出的定理。

麦克斯韦和玻耳兹曼在19世纪60年代的工作出现于这个混乱时代的中期。他们的工作自然应该反映出在19世纪上半叶概率概念的这种歧义。尽管如此，他们似乎主要是从频率出发将概率看作一个描述多粒子系统的客观量，而且可以由该系统力学状态来明确定义。然而，这一点麦克斯韦并不比玻耳兹曼更清楚。

逐渐地，概率从这个机械背景下解放了出来。玻耳兹曼[1871b]和麦克斯韦[1879]分别发现概率描述的是由很多的多粒子系统构成的系综而非单一的系统。吉布斯在1902年的书中将此看作一个统一连贯的观点。然而，这个系综解释仍然是相当模糊的以至于可以得出不同的解读。关于系综的主观认识，与伯努利和拉普拉斯的经典解释密切相关，在20世纪50年代杰恩斯的工作就已经出现。本文将不对这种方法进行进一步讨论。更多的细节，请参见[Jaynes，1983；Uffink，1995；1996；Balian，2005]。

3.2　从伯努利到麦克斯韦(1860年)

气体运动学(有时也被称为气体动力学)通常可以追溯到丹尼尔·伯努利1738年写的《流体力学》(*Hydrodynamica*)中的一段。当然，以前的作者相当熟悉气体是由大量但有限的微观粒子组成的这个观点。然而，他们通常用静态模型来解释气体压强现象，假设这些粒子之间存在斥力。

伯努利是第一个将压强解释为与微粒运动有关的人。他认为气体由大量的

粒子组成，它们穿越空间从这里移动到那里，并不断碰撞容器壁产生压力。有了这个模型，伯努利就能够得到理想气体定律，即在恒定的温度下 $pV=$ 常数，预言了稠密状态下该定律的修正，并指出温度应该与粒子速度的平方成正比。尽管这一努力在最初是成功的，但在接下来的一个世纪中都没有被关于气体运动理论的研究进一步跟进。相比之下，将气体看作一个连续统一体模型的观点却是相当受欢迎的，因为它允许使用微积分这个强大的工具。事实上，在 19 世纪初期，例如沃特斯顿和赫拉帕斯所做的关于分子运动学方面的那仅有的一点研究，几乎完全被他们同时代的学者忽略了，参见[Brush，1976]。

尽管如此，19 世纪 50 年代分子运动的观点还是在克罗尼格(Kroning)和克劳修斯的工作中得以复苏。这种复苏的主要动力就是关于热和做功等价的焦耳—迈尔原理，它导出了热力学第一定律，并使得热本身就是气体粒子运动的一种形式这一认识看起来更合理。(克劳修斯 1857 年的题为“我们称之为热的那种运动”的论文就很好地表达了这种观点，随后斯蒂芬·布鲁斯在 1976 年研究这一时期的历史时也采纳了这种观点。)

克劳修斯也认识到气体粒子之间相互碰撞的重要性，它可以用来解释为什么对比于粒子那相当快的运动速度(在正常室温下估计为 400 米/秒或更快)，扩散的速度却相对缓慢。事实上，他认为，尽管它们的速度很快，但是平均自由程，即一个粒子在两次碰撞之间经过的典型距离，可能是相当小的(微米级)，以至于粒子每秒的平均位移也相应地要小得多。

不过，克劳修斯没有过多地关注碰撞也会改变速度的大小这个事实。的确，尽管他的论文有时会提到“平均速度”或“概率定律”等短语，但他并没有详细给出一个精确的平均过程或概率假设，而且他的计算经常是在粗略的简化基础上进行的(如假设除某个粒子运动外其他粒子都静止)。

麦克斯韦(1860)

麦克斯韦在 1860 年完成的论文，真正标志着分子运动学的重生。麦克斯韦意识到，如果气体由很大数量的 N 个运动粒子组成，那么由于相互碰撞，它们的速度将会不断变化，因此，处于静态的气体就应该由慢粒子和快粒子的混合物构成。更重要的是，对于麦克斯韦来说，这不是一个需要通过简化的假设来取代的令人烦恼的难题，而是一个值得进一步研究的重要特征。

于是他提出下面这个问题。

提议四：计算在大量相同粒子发生多次碰撞之后，速度值在给定界限之间的粒子的平均数，参见［Maxwell，1860，380］。

他将这个想要得到的平均数表示为 $Nf(\vec{v})\,d^3\vec{v}$，且通过两个让人印象深刻的假设找到了这个问题的解。(i) 分布函数 $f(\vec{v})$ 应该分解为速度的正交分量的函数，即存在一些函数 g，使得：

$$f(\vec{v}) = g(v_x)g(v_y)g(v_z) \tag{24}$$

(ii) 分布函数应该是球对称的，即：

$$f(\vec{v})\text{ 只取决于 } v = \|\vec{v}\| \tag{25}$$

他发现要想满足这样的函数方程，只有：

$$f(\vec{v}) = Ae^{-v^2/B} \tag{26}$$

这里常数 A 由归一化公式 $A = (B\pi)^{-3/2}$ 给定，常数 B 由绝对温度下的速度平方的平均值确定——即采取现代的写法 $\frac{3}{2}kT = \frac{m}{2}\langle v^2\rangle$—— 这样就得到：

$$f(\vec{v}) = \left(\frac{m}{2\pi kT}\right)^{3/2} e^{-mv^2/2kT} \tag{27}$$

麦克斯韦的结果导出了一些新的且令人意想不到的预言，最引人注目的是气体的黏度不依赖于它的密度，后来它就被实验验证了。麦克斯韦另一个著名的预言是，在这个模型中具体的热容比 $\gamma = \frac{c_V}{c_p}$ 的值一定是 $\frac{4}{3}$。这不同于实验上得到的值 $\gamma = 1.408$。①

麦克斯韦最先用分布函数 f 来表示气体的状态。他也是第一个将 $f(\vec{v})\,d^3\vec{v}$ 称为概率的人。显然，麦克斯韦采用了概率的频率解释。速度在范围 $d^3\vec{v}$ 内的概率正是速度在这个范围内的气体粒子的相对数量。它指的是气体系统的一个客观的力学性质，而不是我们的主观观点。

---

① 更为普遍的是，$c_V/c_p = (f+2)/f$，其中 f 是分子的自由度数目。这种所谓的 c_V/c_p 异常困扰了气体理论半个世纪。实验值为 1.4 左右，部分原因是，大多数普通气体为二原子分子，对其来说典型的 $f = 6$。需要应用量子理论来解释，在室温下这些自由度中的一个被"冻结"了。与麦克斯韦预言相一致的实验最早是由孔特和沃伯格于 1875 年利用汞蒸气做的。相关的详细信息，请参阅［Brush，1976，pp. 353—356］。

现在，这种观点面对的一个明显问题是，如果气体由有限数目的粒子组成，那么速度分布必定是离散的，即用狄拉克(Dirac)δ符号表示为：

$$f(\vec{v}) = \frac{1}{N}\sum_{i=1}^{N}\delta(\vec{v} - \vec{v}_i) \tag{28}$$

且气体的能量如果是有限的和确定的，那么分布应该有一个边界。式(26)的函数则没有这些性质。

目前尚不清楚麦克斯韦是如何回应这些问题的。有理由认为在某种意义上，他会把式(26)视为仅是对实际气体状态的一个足够好的近似表示，① 而不是一个确切的表示，就像在偶然性实验中的真实频率永远不会完全与其预期值相同。麦克斯韦通过插图建议将连续的分布函数看作点的离散云，其中每个点代表一个速度矢量的端点，参见图1。见[Maxwell，1875]。这表明他认为实际的分布更符合式(28)而非式(26)。但是在什么意义下麦克斯韦分布能近似看作实际的速度分布，仍然是一个未解决的问题。

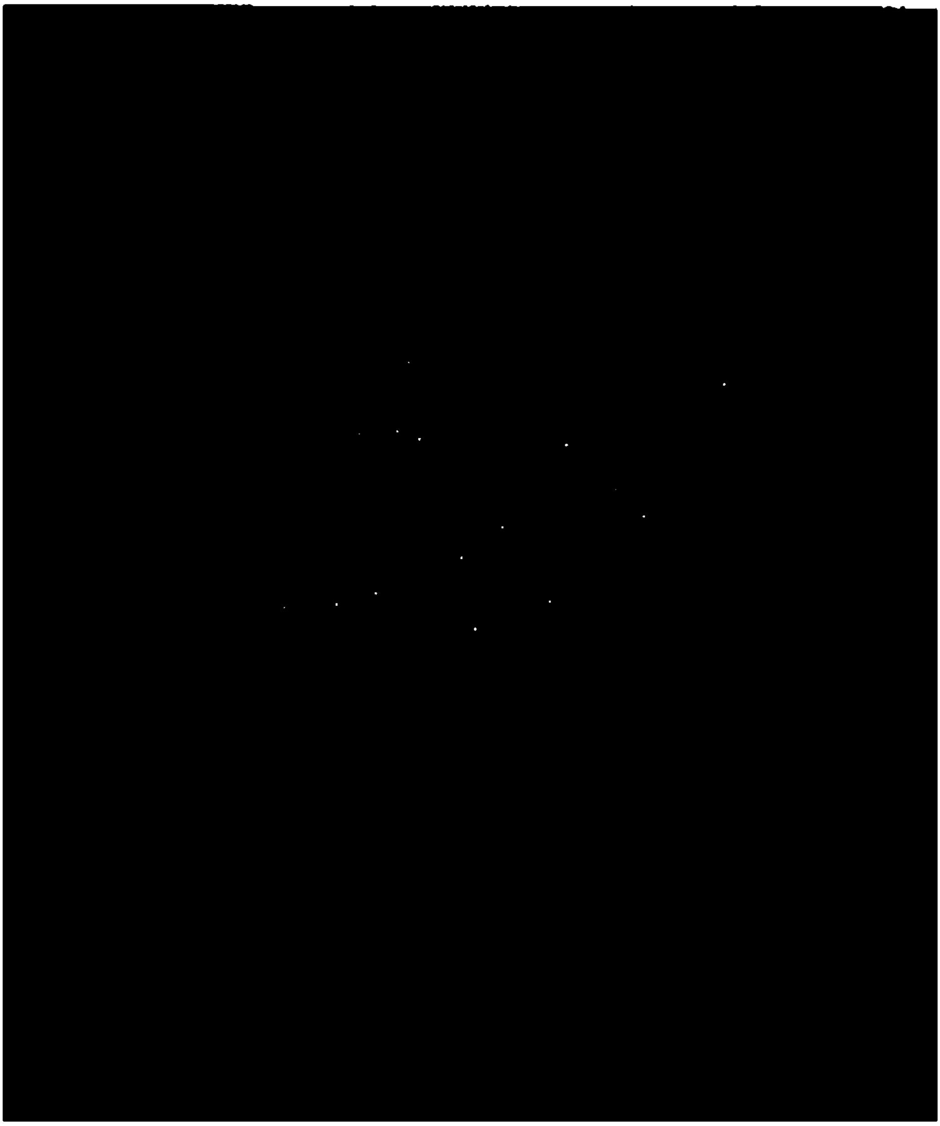
这里一个可选择的方法是更加强调上面引述于麦克斯韦的话中的“平均”二字。也就是说，f或许并不代表实际的速度分布，而是平均分布。接下来的问题是，是什么样的平均？由于已经使用了粒子的平均，因而唯一合理的选择可能就是时间平均或相似气体系统构成的系综平均。但我找不到证据能表明麦克斯韦在他的论文中设想了这种方法。或许有些人认为式(26)的分布是一种预期，即它代表对具有一定速度的粒子数的一个合理的主观猜测。但在这种情况下，麦克斯韦对概率的解释，最终将会成为经典。

虽然这是可能的，但是值得注意的是，通过引入用连续函数表示状态的数学方法(泛函方程)，分子运动学才取得了超越伯努利的工作进展，其代价是使得与物理现实相关的态的概念变得更抽象。

一个更为紧迫的问题是麦克斯韦用来得到他的分配形式的假设式(24)和式(25)与其意指的频率解释并不相协调。这些假设似乎反映了关于对称性的先验要求，并且也许是基于对某种形式的不充分理由原则的需求。在这个意义上，

---

① [Boltzmann，1896b]也表达过这一观点。例如，他写道：“对于有限数量的分子来说，麦克斯韦分布不能被准确地实现，只能被视为一个很好的近似。”见[Boltzmann，1909，Ⅲ，p. 569]。



速率的图解

图 1　麦克斯韦分布的图示[Maxwell，1875]，

图中每个点代表一个速度矢量的端点

根据我们的知识，如果没有理由表明速度的各正交分量之间存在联系的话，我们就有权力假定它们是独立的。

这种对麦克斯韦动机的解读得到了来自文献的支持，他在 1867 年描述

1860 年给出的式(24)的假设时写道："对分子在 x 轴方向的速度大小处于某个范围内的概率所做的假设，是不会受到我们已知的其在 y 轴方向的速度大小的影响的。"参见[Maxwell，1867，emphasis added]。

有人指出，参见[Brush，1976，Nol. Ⅱ，pp. 183—188]，麦克斯韦在 1860 年的论断似乎是受到了赫歇尔(Herschel)[1850]对克托莱在概率方面研究所做的评论的启发。这个综述文章中有非常重要的相似性论断，且该论断被应用于确定一个神枪手射击靶子时，子弹会落于该靶子一定距离范围内的概率。此外，赫歇尔的文章坚定地支持概率的经典解释，并给予了不充分理由原则一定的地位。事实上，他将(类似的)式(25)解释为："该式只是表明了我们对错误(即偏离靶子)出现的原因及其作用方式是完全无知的。"见[Herschel，1850，398]。如果麦克斯韦确实借鉴了这么多赫歇尔的观点的话，那么有可能他也同意，或者至少是受到了这种由式(25)条件给出的动机的启发。[①]

无论麦克斯韦提出这些假设的最初原因是什么，这些假设存在的问题仍是很清楚的，因为尽管他将问题形式化了(即确定"大量碰撞后"函数 f 的形式)，但是这些假设根本就没有涉及碰撞。事实上，任何表明其合理性的原因，看起来也适用于由非碰撞(如能被完全穿透的)粒子组成的气体模型。正如麦克斯韦后来评论其他人的论文中的某些推论时指出的，我们也可以说，应用"大量碰撞后"这个条件是"为了使读者能够对材料系统形成一个主观印象，而非证明的条件"，参见[Garber *et al.*，1995，359]和[Maxwell，1879]。

3.3 麦克斯韦(1867 年)

无论麦克斯韦第一篇论文的价值和问题是什么，他在接下来的 1867 年关于气体理论的论文中舍弃了他以前试图从式(24)和式(25)的假设出发得到分布函数的尝试，认为那是"不可靠"的，并提出了一个完全不同的论断。这一次，他考虑

---

① 有趣的是，我们注意到赫歇尔的评论最早且尖刻地批评了莱斯利·埃利斯(Leslie Ellis)将不充分理由原则应用于事件的频率，包含对其著名言论的批评："仅仅无知是不能作为任何推理的基础的，那会是无中生有。如果我们对事物一无所知的话，情况不可能是这样的，其实我们还是知道一些的。"见[Ellis，1850]。不过，不能确定麦克斯韦是否知道这些批评。

了一个由等质量点粒子构成的模型，这些点粒子间存在中心斥力，该斥力与它们彼此间距离的五分之一次幂成正比。更重要的是，这个论断用到了碰撞。

麦克斯韦考虑了一对粒子之间的弹性碰撞，其初速度分别为 $\vec{v}_1$，$\vec{v}_2$，末速度分别为$\vec{v}_1'$，$\vec{v}_2'$。① 这些量由动量和能量守恒定律相联系起来，由此产生了四个方程，其中的两个参数取决于碰撞过程的几何因素。

我们可以很方便地考虑这样一个坐标系，在其中粒子 1 最初是静止的，相对速度 $\vec{v}_2-\vec{v}_1$ 沿 z 轴负方向，使用柱坐标。如果(b, ϕ, z)表示粒子 2 中心的轨迹坐标，那么，在碰撞之前，我们知道 b =常数，ϕ=常数，$z(t)=z_0-\|\vec{v}_2-\vec{v}_1\|t$。在这些粒子是弹性硬球的情况下，只有在碰撞参数 b 小于球体直径 d 时，碰撞才会发生。然后碰撞后的速度由 $\|\vec{v}_1-\vec{v}_2\|$，b 和 ϕ 确定。转换回实验室坐标系，末速度$\vec{v}_1'$，$\vec{v}_2'$可以被表示为 $\vec{v}_1$、$\vec{v}_2$，b 和 ϕ 的函数。

麦克斯韦在这里提出了后来被埃伦费斯特称为分子混沌假设的假设。在时间 dt 内发生碰撞的次数 $N(\vec{v}_1, \vec{v}_2)$，即在三维微分空间 $dV=bdbd\phi dz=\|\vec{v}_1-\vec{v}_2\|bdbd\phi dt$ 中，由速度为 $\vec{v}_1$，$\vec{v}_2$ 的微分量 $d^3\vec{v}_1d^3\vec{v}_2$ 变为速度为$\vec{v}_1'$，$\vec{v}_2'$的微分量 $d^3\vec{v}_1'd^3\vec{v}_2'$ 的粒子数，正比于速度为 $\vec{v}_1$ 的微分量为 $d^3\vec{v}_1$ 的粒子数，即 $Nf(\vec{v}_1)d^3\vec{v}_1$，与速度为 $\vec{v}_2$ 的微分量为 $d^3\vec{v}_2$ 的粒子数，即 $Nf(\vec{v}_2)d^3\vec{v}_2$，以及三维微分空间三者的乘积。因此：

$$N(\vec{v}_1, \vec{v}_2)=N^2f(\vec{v}_1)f(\vec{v}_2)\|\vec{v}_2-\vec{v}_1\|d^3\vec{v}_1d^3\vec{v}_2bdbd\phi dt \tag{29}$$

由于碰撞定律的时间反转不变性，对于所谓的反碰撞也可以应用类似的考虑，其中初速度$\vec{v}_1'$，$\vec{v}_2'$和末速度 $\vec{v}_1$，$\vec{v}_2$ 互换。碰撞次数为：

$$N(\vec{v}_1', \vec{v}_2')=N^2f(\vec{v}_1')f(\vec{v}_2')\|\vec{v}_2'-\vec{v}_1'\|d^3\vec{v}_1'd^3\vec{v}_2'bdbd\phi dt \tag{30}$$

麦克斯韦认为，速度分布将保持稳定，即在时间进程中是不变的，如果这两类碰撞的数量相等的话，即如果：

① 在无限相互作用的观点下，“初”和“末”是在渐近的意义上理解的，即在极限 $t\rightarrow\pm\infty$ 下。本文后面的另一种方法是，用麦克斯韦在 1860 年考虑的硬球模型取代 1867 年的这个模型。

$$N(\vec{v}_1{}', \vec{v}_2{}') = N(\vec{v}_1, \vec{v}_2) \tag{31}$$

此外，碰撞定律要求 $\| \vec{v}_2{}' - \vec{v}_1{}' \| = \| \vec{v}_2 - \vec{v}_1 \|$ 且 $d^3 \vec{v}_1{}' d^3 \vec{v}_2{}' = d^3\vec{v}_1 d^3\vec{v}_2$。因此，条件(31)可以简化为：

$$\text{对所有的 } \vec{v}_1, \vec{v}_2 \text{ 来说}, f(\vec{v}_1)f(\vec{v}_2) = f(\vec{v}_1{}')f(\vec{v}_2{}') \tag{32}$$

这正是式(26)麦克斯韦分布的情况。因此，麦克斯韦认为，式(26)的分布是一个“可能”的形式。

他又断言，对于稳定分布来说，这也是唯一可能的形式。该断言认为，稳定分布 f 只能由式(32)给出，现在也被称为细致平衡原理，参见[Tolman, 1938, 165]。① 虽然他的论证相当简短，但看起来其想法是对于违反式(32)的分配，一定存在(基于分子混乱假设的)两个速度对②，$\vec{v}_1, \vec{v}_2$ 和 $\vec{u}_1, \vec{u}_2$，它们满足 $\vec{v}_1 + \vec{v}_2 = \vec{u}_1 + \vec{u}_2$ 且 $v_1^2 + v_2^2 = u_1^2 + u_2^2$，以至于碰撞将主要使 $(\vec{v}_1, \vec{v}_2) \longrightarrow (\vec{u}_1, \vec{u}_2)$ 而非 $(\vec{u}_1, \vec{u}_2) \longrightarrow (\vec{v}_1, \vec{v}_2)$ 的转变发生。但是，由于分布是稳定的，必然存在第三个速度对 $(\vec{w}_1, \vec{w}_2)$ 满足类似的关系，对此碰撞将主要使 $(\vec{u}_1, \vec{u}_2) \longrightarrow (\vec{w}_1, \vec{w}_2)$ 发生，等等。现在，只有全部此类序列形成一个循环，分布才能保持稳定。因此就会存在一个速度对的循环 $(\vec{v}_1, \vec{v}_2) \longrightarrow (\vec{u}_1, \vec{u}_2) \longrightarrow \cdots \longrightarrow (\vec{v}_1, \vec{v}_2)$，碰撞粒子经过这个循环，最终回到自己最初的速度。

麦克斯韦则认为：“对于分子的速度为什么要满足这样的一个循环而不是可逆的，我们不可能给出一个理由。”见[Maxwell, 1867, 45]。他认为这两种循环出现的可能性应该是均等的，因此，$(\vec{v}_1, \vec{v}_2) \longrightarrow (\vec{v}_1{}', \vec{v}_2{}')$ 这样的碰撞循环同 $(\vec{v}_1{}', \vec{v}_2{}') \longrightarrow (\vec{v}_1, \vec{v}_2)$ 这样的碰撞循环是具有同等概率的，即条件式(32)成立。

注解：首先，麦克斯韦在 1867 年关于分布函数式(26)的推导的一个明显的优势是碰撞在其中起到至关重要的作用。如果分子之间没有碰撞的话，这个论断就是不适当的。第二点要注意的是分布式(26)之所以被挑选出来，是因为

① 读者需要注意，“细致平衡”这个名称还具有不同于此处用法的含义[Tolman, 1938, 521]。

② 事实上，麦克斯韦讨论的只是单个粒子的速度情况，为清晰起见，我将他的论证放在粒子对上进行。

它的稳定性，而不是因为其球对称和因子分解性。这也是对他以前论文的一个重大改进，因为稳定性对热平衡来说是至关重要的。

论断中的一个关键因素仍是一个对独立性的假设。但现在，在分子混沌假设中碰撞粒子的初速度是独立假定的，而不是单粒子的正交速度分量。麦克斯韦并没有进一步阐释为什么我们应该这样假设。他清楚地认为这是显而易见的。然而，似乎有理由认为，在他的思想背后，他一定相信某种版本的不充分理由原理，即我们将两个碰撞粒子的初速度看作独立的，是因为我们没有理由做出其他假设。虽然仍是从不充分理由出发的论断，但这至少比 1860 年的论文中的那个更加合理。

该论文的一个主要不足在于他粗略地指出麦克斯韦分布式(26)将是唯一的稳定分布。这种断言可能会在两方面被否定：(a)上面谈论过的有关循环的论断，它使得麦克斯韦论证了细致平衡；(b)声称麦克斯韦分布是唯一与此条件相匹配的分布。

麦克斯韦根本没有做出针对(b)方面的论证；但很快玻耳兹曼(1868 年)就给出了证明——麦克斯韦对玻耳兹曼的证明也给予了应有的赞扬。但(a)部分更有趣。我们已经看到，麦克斯韦在这里明确依赖于不充分理由原理。在这一点上他受到了玻耳兹曼[1872]和格思里(Guthrie)[1874]的批评。

玻耳兹曼认为，麦克斯韦回避了问题的实质。如果我们假设这两种循环发生的概率不是相等的，那么这个假设本身就提供了一个为两种类型的碰撞分配不相等概率的理由。① 这种说法由玻耳兹曼提出，但是至少在我看来，其准

---

① 更确切地说，玻耳兹曼的论证如下：为了证明(一对)分子的速度从 $(\vec{v}_1,\vec{v}_2)$ 变化到 $(\vec{v}_1',\vec{v}_2')$ 的可能性大于相反变化的可能性(这个假说)是不可能的。麦克斯韦认为，应该存在一个有关速度的封闭循环，分子的速度变化会以某种顺序而不是其他顺序完成这个循环。然而，他认为这是不可能的，因为我们不能找到一个理由，说明为什么分子会按照这种顺序而非其他顺序。但在我看来，这最后的声明已经假设了其想要证明的论断已经得到了证明。事实上，如果我们假设已经证明，速度从 $(\vec{v}_1,\vec{v}_2)$ 变化到 $(\vec{v}_1',\vec{v}_2')$ 的可能性与相反变化的可能性一样，那么我们当然没有理由相信为什么循环会以某种顺序而非其他顺序完成。但是，如果我们假设该声明尚未被证明，那么分子速度更可能从 $(\vec{v}_1,\vec{v}_2)$ 变化到 $(\vec{u}_1,\vec{u}_2)$，而不是相反；更可能从 $(\vec{u}_1,\vec{u}_2)$ 变化到 $(\vec{w}_1,\vec{w}_2)$，而不是相反，等等，正是这个事实说明为什么循环会以某种顺序而非其他顺序完成。见[Boltzmann，1909，Ⅰ，p. 319]。

备远没有麦克斯韦就不充分理由给出的论断准备充分。事实上，在第 4 节中我们将看到，相比于麦克斯韦，他对概率的认识更像是一个彻底的频率论者。

事实上玻耳兹曼后来一再提到一个气体的反例，其中所有粒子都一字排开，使得它们只发生中心碰撞，且在平行壁间垂直运动 [Boltzmann，1872；Boltzmann，1909，Ⅰ，p. 358；Boltzmann，1878；Boltzmann，1909，Ⅱ，p. 285]。在这种情况下，速度分布

$$\frac{1}{2}(\delta(v-v_0)+\delta(v+v_0)) \tag{33}$$

也是不变的。

对麦克斯韦工作的最终评价：正如我们所看到的，精确描述概率的麦克斯韦解释并不容易。在他 1860 年的论文中，他认为一个特定分子状态的概率就是处于该状态的粒子的相对数量。① 然而，我们也看到，他将概率与知识的状态联系了起来。因此，他的认识可能被定位为介于经典认识和频率论者的观点之间。

请注意，麦克斯韦从来没有试图重新定义第二定律。在一定程度上，他似乎满足于气体热平衡的统计描述。②他在 1867 年后所有的著作都表明，他确信从力学原理推导出第二定律是不可能的。事实上，他对第二定律的一般性评论表明，他认为第二定律“只具有统计确定性”，可参见给泰特的信，日期不详，另见[Garber *et al.*，1995，180]，而统计因素是不适用于力学原理的。事实上，麦克斯韦很乐意看到玻耳兹曼和克劳修斯争论到底是谁最先将热力学第二定律概括到了力学中：

> 这真是一个笑柄，我看到那些博学的德国人争夺关于 $\theta\Delta cs$ 的第二定律是“哈密顿定理(Hamiltonsche Prinzip)”的发现优先权，……同时，哈密顿定理存在于不受统计因素干扰的领域，而德国人伊卡利

① 很明显，这个术语在他 1867 年的论文中完全没有出现。

② 除了布鲁斯[1976，344]分析的麦克斯韦[1860]中的一个相当差劲的论证。

> (Icari) 则在 nephelococcygia① 拍打着他透明的翅膀，在那里对人文科学的无知和限制已经赋予了无形的希腊神后不能传达的价值。见给泰特的信，1873 及[Garber *et al.*，1995，225]。

显然，麦克斯韦认为第二定律源于力学“这个不受统计因素干扰的领域”是一种幻想。这一点更为生动地体现在他的思想实验“麦克斯韦妖”中，在这个思想实验中，他表明如何利用力学定律来违反第二定律。在关于麦克斯韦妖的大量文献中，我参阅了[Earman and Norton，1998；1999；Leff and Rex，2003；Bennett，2003；Norton，2005]。

但是麦克斯韦并没有努力将第二定律重建于统一的统计/力学基础之上。事实上，他就该主题所给出的很少的评论似乎指向了另一个方向，例如[Maxwell，1873；Maxwell，1878b]，他区分了他所谓的“统计方法”和“历史”或“力学”(或有时称为“运动”)方法。这是对相同系统的两种描述方式。但是，麦克斯韦认为不是要将它们统一，而是要让它们竞争，甚至它们就是不相容的——有人说是“互补的”——方法，而且使用哪个是合适的，取决于我们自己的知识、能力和兴趣。例如：

> 涉及物体的质量时，由于我们看不到单个的分子，因此我们不得不采取我所说的统计方法，而放弃严格的力学方法，因为在力学方法中每次运动都由微积分得到[Maxwell，1872，309]。

在这方面，他的立场与玻耳兹曼形成了鲜明的对比，后者在一生的工作中都力图发现这种统一的基础。

① “克拉乌考克诺兰(Cloudcuckooland)”，阿里斯托芬作品《鸟》(*The Bird*)中的一个虚构世界。

4. 玻耳兹曼①

4.1 早期工作：分子混沌假设和遍历假设

玻耳兹曼在1866年的一篇论文中就曾考虑过从力学推导出第二定律的问题。当时，他并不知道麦克斯韦的工作。但在1868年，他读了麦克斯韦在1860年和1867年所写的论文。和麦克斯韦一样，他侧重于研究处于热平衡态的气体，而不是第二定律。他还采用麦克斯韦的想法，用概率分布表征热平衡态，将分子混沌假设看作重要的力学假设。但是在这篇宏大的论文中，玻耳兹曼介绍了一种完全不同的研究方法，这依赖于我们现在所谓的遍历假设。

正如我们在3.3节中所看到的，麦克斯韦从两个特殊气体模型(即1860年的硬球气体模型和1867年的相互之间存在大小正比于中心距离r^5的排斥力的点粒子模型)中得出了他的平衡分布。他注意到这种分布一旦达到，将会在时间轴上保持稳定(当气体保持孤立的情况下)，也认为(但不是非常有说服力的)这是唯一的稳定分布。

在[1868a]论文的第一节中，玻耳兹曼旨在将系统看作由平面上无限数量的处于运动状态的硬质圆盘构成，然后重现和改善这些结果。他认为，平衡分布很明显应该与圆盘的位置无关，且它们的速度沿每个方向的可能性都是相等的。因此，考虑速度$v = \|\vec{v}\|$的各个值的概率分布就可以了。然而，玻耳兹曼在思想上关于概率的解释一开始就与麦克斯韦有所不同。他通过下面的方式引进概率分布：

> 令$\phi(v)dv$为很长一段时间内，一个圆盘的速度介于v和$v+dv$之间的所有短暂瞬时间的总和，令N为单位面积内圆盘的平均数量，那么
>
> $$N\phi(v)dv \tag{34}$$
>
> 就是单位面积内速度介于v和$v+dv$之间的圆盘的数量。② 见[Boltz-

① 这节的部分出版于[Uffink, 2004]。

② 这里和下文中的"Abh"指的是玻耳兹曼的三卷科学论文集[Boltzmann, 1909]。

mann, 1909, I, p. 50]。

因此，$\phi(v)dv$ 是作为相对时间的量而被引入的，它表示在这段时间内一个(给定的)圆盘处于特定的速度。但是，同时它又被理解为是处于这个速度的圆盘的相对数量。

这个值得注意的引用表明了他是如何为一个函数赋予两种不同含义的。我们将会看到，这段模棱两可的话，在玻耳兹曼的著作中以不同的形式出现了一遍又一遍。①事实上，我相信，这正是遍历问题的中心，埃伦费斯特夫妇将其放在相当显著的位置(参见6.1)。当然，无论是对时间还是粒子的平均，这里概率都是在严格的力学意义上定义的，因此，它是气体的客观性质。

接着，他对双圆盘碰撞过程进行了细致的力学描述。如果指定碰撞前两个圆盘初始速度的变量处于一个给定的无穷小范围内，那么玻耳兹曼就能确定碰撞如何将变量 $(\vec{v}_i, \vec{v}_j)$ 的初值转变为另一个范围内的末值$(\vec{v}'_i, \vec{v}'_j)$。在这一点上他引入了分子混沌假设的二维类比，以得到单位时间内碰撞的次数。就像麦克斯韦那样，这相当于假设这种碰撞的次数与乘积 $\phi(v_1)\phi(v_2)$ 成正比。事实上：

$$N(\vec{v}_1, \vec{v}_2) \propto N^2 \frac{\phi(v_1)}{v_1} \frac{\phi(v_2)}{v_2} \| \vec{v}_2 - \vec{v}_1 \| dv_1 dv_2 dt \tag{35}$$

其中比例常数取决于碰撞的几何结构。

他指出，如果对于所有的速度 v_i, v_j 和所有的圆盘数 i, j 来说，使得处于第一个范围 $dv_i dv_j$ 内的速度值(v_i, v_j) 转变为处于范围 $dv'_i dv'_j$内的速度值 v'_i, v'_j 的碰撞发生的可能性与逆碰撞，即使得处于第一个范围 $dv'_i dv'_j$ 内的速度值 v'_i, v'_j 转变为处于范围 $dv_i dv_j$ 内的速度值(v_i, v_j) 的碰撞的可能性相同，那么分布 ϕ 将保持稳定。他指出：“因此，这种分布是我们希望得到的。”见[Boltzmann, 1909, I, p. 55]。其实，这是第一次在论文中提及希望得到稳定的概率分布。

① 这并不是说他总是将这两种概率的解释混为一谈。一些论文只采用一个明确和一贯的解释。但在不同的论文中，甚至在一篇论文的不同部分中选择也不同。事实上，在[Boltzmann, 1871c]中，他甚至将不同解释下的概率应用到了同一个方程中，从而得到了联合概率。随后在1872年，他再次将它们合并。甚至在他的最后一篇论文[Boltzmann and Nabl, 1904]中，玻耳兹曼认为关于概率的这两个含义只是一种单纯的思想上的论证。

在分子混沌假设的二维版本下，上述希望使得：

$$\frac{\phi(v_i)}{v_i}\frac{\phi(v_j)}{v_j}=\frac{\phi(v'_i)}{v'_i}\frac{\phi(v'_j)}{v'_j} \tag{36}$$

他表明[Boltzmann,1909,p. 57]对 v_1,v_2,v'_1,v'_2 的所有选择来说,满足条件式(36)且兼容能量方程 $v_1^2+v_2^2=v_1'^2+v_2'^2$ 的唯一形式为：

$$\phi(v)=2hve^{-hv^2} \tag{37}$$

其中 h 为常数。定义 $f(v)\coloneqq v\phi(v)$,从而我们得到了麦克斯韦分布式(26)的二维版本。玻耳兹曼在该论文中没有说明是否条件式(36)对 ϕ 的稳定性是必要的。

在 1868a 的下一小节中，玻耳兹曼重复了这个推导，只是在设定上略有不同。首先，他回到问题的三维版本，假设系统由硬球组成，并且假设一个特定的球是由外部势 $V(\vec{x})$ 加速的。他表示,如果其他所有球体的速度是按照麦克斯韦分布式(26)分布的话,发现该特定球体坐标为 $\vec{x}$、速度为 $\vec{v}$ 的概率分布为 $f(\vec{v},\vec{x})\propto e^{-h(\frac{1}{2}mv^2+V(\vec{x}))}$[Boltzmann，1909，I，p. 63]。在随后的小节中，他将硬球替换为存在短程相互作用势的物质点，并得到了类似的结果。

在这个时候论证突然转变了方向，参见[1868a]论文的第一节末尾。玻耳兹曼不再以这种方式继续，而是宣布在[Boltzmann，1909，80]文中所有已经讨论过的情况和尚未涉及的情况，全都服从一个更一般的定理。这个定理，我们将看到它依赖于遍历假设，就是本论文的第二和第三部分讨论的主题。我将在第三部分对此进行讨论，并涉及部分麦克斯韦[1879]的阐释，因为它多少比玻耳兹曼自己的解释简单和清晰一点。

遍历假设

考虑由 N 个质点组成的一般力学系统，其中每一个质点的质量为 m，且受一个随机的时间无关势①支配。在现代符号中，用 $x=(\vec{q}_1,\vec{p}_1;\cdots;\vec{q}_N,\vec{p}_N)$

① 尽管玻耳兹曼在这里没有提到它，但他在前一节中附加了这样的规定：粒子被封闭在一个有限的空间内，由完全弹性墙壁包围。

表示系统的正则位置坐标和动量。那么其哈密顿量①为：

$$H(x) = \frac{1}{2m}\sum_{i}^{N}\vec{p}_i^2 + U(\vec{q}_1,\cdots,\vec{q}_N) \tag{38}$$

态 x 可被描述为机械相空间中的相点 Γ。根据哈密顿运动方程，相点随时间变化，因此可以得到一个轨迹 $x_t(t\in\mathbb{R})$。这个轨迹被限定只能位于一个给定的能量曲面 $\Gamma_E=\{x\in\Gamma: H(x)=E\}$ 上。玻耳兹曼想得到相点位于范围 $dx=d^3\vec{q}_1\cdots d^3\vec{p}_N$ 内的概率（即在很长一段时期内的一个时间片断），对于一些函数 χ，我们可以将其记作：

$$\rho(x)dx = \chi(x)\delta(H(x)-E)dx \tag{39}$$

看起来玻耳兹曼隐含地假设了这种分布是稳定的。如果"很长一段时期"被理解为是无限时间的话，那么这个性质当然是一定的。他认为，根据刘维尔（Liouville）定理，对"可能的"能量曲面，即实际经过的轨迹上的所有点来说，χ 是常量。对于其他所有点来说，χ 为零。如果我们忽视后者，那么函数 χ 就一定是整个能量曲面上的常量，而概率密度 ρ 取微正则分布的形式：

$$\rho_{\mathrm{mc}}(x) = \frac{1}{\omega(E)}\delta(H(x)-E) \tag{40}$$

其中

$$\omega(E) = \int\delta H(x) = Edx \tag{41}$$

为所谓的结构函数。

特别地，我们可以结合动量得到位置 $\vec{q}_1$，…，$\vec{q}_N$ 的边缘概率密度：

$$\rho_{\mathrm{mc}}(\vec{q}_1,\cdots,\vec{q}_N) := \int\rho_{\mathrm{mc}}(x)d^3\vec{p}_1\cdots d^3\vec{p}_N =$$

$$\frac{2m}{\omega(E)}\int\delta\left(\sum\vec{p}_i^2 - 2m(E-U(q))\right)d\vec{p}_1\cdots d\vec{p}_N \tag{42}$$

可以准确地得到动量积分（它是 $n=3N$ 维空间中半径为 $R=\sqrt{2m(E-U)}$ 的

---

① 事实上，玻耳兹曼允许 N 的质量是不同的，但限制势只能是来自外部，或是由两粒子相互作用产生的，即 $H(x)=\sum_i\frac{\vec{p}_i^2}{2m_i}+\sum_{i\leqslant j}U_{ij}(\|\vec{q}_i-\vec{q}_j\|)+\sum_i U_i(\vec{q}_i)$。

超球面面积的$2R^{-1}$倍)，因此可得：

$$\rho_{mc}(\vec{q}_1,\cdots,\vec{q}_N) = \frac{2m(\pi)^{n/2}}{\omega(E)\Gamma\left(\frac{n}{2}\right)}(2m(E-U(q)))^{(n-2)/2} \tag{43}$$

其中Γ表示欧拉的伽玛函数：$\Gamma(x) := \int_0^\infty t^{x-1}e^{-t}dt$。

类似的，发现第一个粒子的动量分量为p_{1x}，且发现所有粒子的位置为$\vec{q}_1$，…，$\vec{q}_N$的位置概率密度为：

$$\begin{aligned}\rho_{mc}(p_{1x},\vec{q}_1,\cdots,\vec{q}_N) &= \int\rho_{mc}(x)dp_{1y}dp_{1z}d^3\vec{p}_2\cdots d^3\vec{p}_N \\ &= \frac{2m\pi^{(n-1)/2}}{\omega(E)\Gamma\left(\frac{n-1}{2}\right)}(2m(E-U(q))-p_{1x}^2)^{(n-3)/2}\end{aligned} \tag{44}$$

这两个结果可能会适时地出现在条件概率的表达式中，给定第一个粒子的位置的值为$\vec{q}_1$，…，$\vec{q}_N$，那么其动量的x分量的取值介于p和$p+dp$之间的概率，可以通过求式(44)和式(43)的比值得到：

$$\rho_{mc}(p\mid\vec{q}_1,\cdots,\vec{q}_N)dp = \frac{1}{\sqrt{2m\pi}}\frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)}\frac{\left(E-U-\frac{p^2}{2m}\right)^{(n-2)/2}}{(E-U)^{(n-3)/2}}dp \tag{45}$$

在本质上，这就是玻耳兹曼曾指出的一般性定理。此外，他表明在极限$n\to\infty$下，每个自由度的自由动能$\kappa := (E-U)/n$为常数，表达式(45)变为：

$$\frac{1}{\sqrt{4\pi m\kappa}}\exp\left(-\frac{p^2}{4m\kappa}\right)dp \tag{46}$$

因此，这个概率与麦克斯韦分布式(26)具有相同的形式，如果$\kappa = \frac{1}{2}kT$等于1的话。据推测，当玻耳兹曼声称在他的论文的第一部分中讨论的所有特殊情况都遵循一般性定理时，在他的想法中考虑的应该就是这个结果。不过，我们应该注意到，既然U(并因此κ)取决于坐标，那么κ = 常数这个条件会因($\vec{q}_1$，…,$\vec{q}_n$)的值的不同而不同。

下面是对这个结果的一些评论。

1. 这种方法与依赖于分子混沌假设的那种方法之间的差异是相当明显的。

相反，不同于关注由除了偶然碰撞外全部自由运动的粒子构成的气体模型，这里玻耳兹曼假设了一个更一般的具有任意相互作用势 $U(\vec{q}_1, \cdots, \vec{q}_N)$ 的哈密顿模型。此外，概率密度 ρ 是定义在相空间的，而不是分子速度空间。这是第一次概率因素被应用于力学系统的整体状态，而不是单个粒子的状态。如果说从气体分子运动学过渡到统计力学之间存在一个间隙的话（如埃伦费斯特夫妇在 1912 年和克莱因在 1973 年认为的那样），那么转变似乎就是在这里发生的。

但当然，对于玻耳兹曼来说，这种转变并没有涉及主要概念的变化，这是因为他将概率看作一个如同相对时间的概念。因此，整个系统处于一个特定状态的概率仍然被视为系统处于该状态上的时间，在此时间段内状态由系统占据。换句话说，他不需要系综或非力学概率假设。

然而，我们应该注意到混淆相对时间和相对粒子数量的那种对概率的认识——在 1868 年的论文的第一节中——相对来说是没有太大问题的，但是在对 ρ 的解释中就不行了。因此，条件概率 $\rho(p \mid \vec{q}_1, \cdots, \vec{q}_N)$ 为我们提供的是总系统处于某个状态上的相对时间，该状态为对于所有给定的位置的值粒子 1 的动量的 x 分量介于 p 和 $p+dp$ 之间。我们不能直接得出这和动量为 p 的粒子的相对数有关的结论。事实上，我们不能保证针对粒子 1 的概率式(45)对于其他粒子也一样，除非我们假设 U 在粒子变换时是不变的。因此，尽管形式相同，但是分布式(46)和(26)的含义完全不同。

2. 令粒子的数目是无限的，那么式(45)就转变成了式(46)，这似乎也是热力学极限的第一个例子。由于只有在此极限下才能重新得到麦克斯韦分布，因此玻耳兹曼的过程解决了麦克斯韦分布提出的一些问题。对于有限数量的粒子，分布式(45)也提供了支持，即对 $p_i^2 \geqslant 2m(E-U)$ 的那些值来说，$\rho_{mc}=0$。因此，在处理有限能量气体时，我们并不会遇到麻烦。

3. 最重要的是，式(45)和式(46)揭示了一个更一般性的观点。这表明对于一个处于定态的孤立系统，如果其分子数量趋向于无穷，那么分子速度的概率就是麦克斯韦形式的。值得注意的是，这个证明意味着除了哈密尔顿量的假设外，还需要对碰撞做出一些特定假设，并且需要涉及有关力学模型的其他细节。但是，它的确不要求研究对象是气体。

4. 目前这个结果的主要缺点是其假设的轨迹实际上经过了能量曲面上的所有点。这就是现在我们说的遍历假设(Ergodic hypothesis)。①

玻耳兹曼在文章的最后一页又回到了这个问题上[Boltzmann, 1909, 96]。他指出，他的定理也可能有例外，例如，在周期性轨迹的情况下。不过，玻耳兹曼认识到这种情况很容易受到来自外界的微小干扰的影响。它们会被破坏，例如，与一个刚好经过的单个自由原子发生相互作用。因此，他认为这些例外只是提供了一些不稳定平衡态的案例。

尽管如此，玻耳兹曼也一定会对他自己的说法感到不满。根据他论文集中的编辑脚注[Boltzmann, 1909, I, p. 96]，玻耳兹曼论文的私人复印稿中在页边空白处有一个手写的评论，说这一点仍然是有疑问的，而且至今也没有证明即使存在一个单一的外部原子发生的相互作用，系统也会遍历所有满足能量方程的可能值。

---

① 关于该假设如何得名，文献资料中的解释也较为混乱。玻耳兹曼在1884年引入"遍历(Ergode)"这个词，并在他关于气体理论的讲座[Boltzmann, 1898]中使用了这个词，埃伦费斯特借用了玻耳兹曼的"遍历"这个名称，但实际上玻耳兹曼是怎么理解遍历的呢?布鲁斯在他翻译的[Boltzmann, 1898, 297]中，以及在[Brush, 1976, 364]中指出，玻耳兹曼用这个词来命名由相空间中微正则分布描述的稳定系综。换句话说，在玻耳兹曼在1898年的背景下，遍历的只是一个微正则系综，似乎和我们今天所谓的遍历假设没有任何关系。布鲁斯指出是埃伦费斯特混淆了这个术语从而引出了歧义。然而，在他最早于1884年引入这个短语的时候，遍历这个名称指的是具有一个运动整体，即其总能量的一重积分的稳定系综。作为结果，该系综的确是微正则的，但更重要的是，系综中的每个元素都满足经过每一个具有给定总能量的相点这个假设。事实上，在这种情境下，成为遍历的元素就意味着满足这一假设。因此，埃伦费斯特将这个假设命名为"遍历的"实际上是有道理的。另一个关于这个词的争议涉及词源学。至少在埃伦费斯特时期，人们的共同看法是这个词来自ergos(工作)和hodos(路径)。然而格拉瓦蒂[1994]认为，它"毫无疑问"是来自ergos的eidos(相似)。现在我们必须承认格拉瓦蒂的认识，因为ergode在词源上的后缀的"—ode"和玻耳兹曼在这篇论文中造的很多词，如Holode、Monode、Orthode和Planode等都是相同的，而在这后四个词中，路径这种解读是不太自然的。不过，我们不认为"Eidos"这种说法会更自然。此外，如果玻耳兹曼熟悉词源学的话，那他将会类似于小行星(planetoid)、椭圆体(ellipsoid)等词那样将其写为"Ergoide"。事实上，通过他在[Boltzmann, 1892]中造了"Momentoide"这个术语来描述类动量的自由度(即那些在哈密顿量中为一个二次方项的量)，我们可以证明他熟悉这种习惯性用法。[Cercignani(瑟斯格纳尼), 1998, 141]中提到的论据(即格拉瓦蒂的父亲是一位古典主义者)在这个问题上并不能说服我们。

对遍历假设的质疑

玻耳兹曼的下一篇论文[1868b]主要通过一个相对简单易解的力学模型来检验遍历假设的有效性。这篇论文也给出了一个很好的关于遍历假设的隐喻表述：如果相点是一个光源，且它运动得非常迅速，以至于在我们看来整个势能面是被均匀照亮的[Boltzmann，1909，I，p. 103]。然而，他的疑虑仍然没有得到消除。他的下一篇关于气体理论的论文[1871a]又回到了对一个详细的气体力学模型的研究，这次是由多原子分子组成的，并避免对遍历假设的依赖。当他在[1871b]不考虑遍历假设时，这次他反而更为谨慎。事实上，也正是在这里，他才第一次将他所疑虑的假设描述为一个假说，具体叙述如下：

> 物体中大量不规则的热运动和大量的作用力，使得其原子由于我们所谓的热运动，能够遍历所有满足能量(守恒)定理的位置和速度对应的状态[Boltzmann，1909，I，p. 284]。①

注意，玻耳兹曼的这一假说针对的是任意物体，即没有限制必须是气体。在文末他还备注道："这里并没有证明这一假说对热物体来说是实现了的，甚至没有证明它是可实现的。"见[Boltzmann，I，p. 287]。

关于遍历假设在玻耳兹曼思想中的作用和地位，在当今的评论中有很大的分歧。事实上，人们经常会怀疑玻耳兹曼怎么可能曾经相信轨迹会遍历能量曲面上所有的点，因为，埃伦费斯特夫妇于1911年猜测，且普兰彻瑞和罗森塔尔在1913年已证明，当能量曲面的维度大于1时，这在数学上是不可能的(见第6.1节)。

事实上，他在[1868a，Abh. I，p. 96]和[1871b，Abh. I，p. 284]中就将外界干涉视为遍历假设的部分动机。这可能被视为"干涉观点"的证据，即认为这种外界影响对于解释热现象来说是至关重要的，参见[Blatt(布拉特)，1959；Ridderbos and Redhead(瑞德鲍斯和雷德黑德)，1998]。然而，即使玻耳兹曼清楚地表

① 遍历假设的另一个等价表述为：哈密顿量只是哈密顿运动方程的独立积分。这种表述也出现在同一篇论文中[Boltzmann，1909，pp. 281—282]。

达了这些干扰可能有助于提出遍历假设，他也从来没有很认真考虑过这个想法。上面提到的论文[1868a]中的旁注表明，即使系统被干扰，也不容易证明遍历假设，而且他所有对这一假设的进一步研究考虑的要么是完全与环境相隔离的系统，要么是最多只受恒等外力影响的系统。因此，在他的思想中，“干涉”并没有发挥显著作用。①

也有学者认为，玻耳兹曼后来习惯于将连续变量离散化，在这种观点下，他会以某种方式将能量曲面看作一个离散簇，其中只包含有限多的离散单元[Gallavotti，1994]。显然，在这种认识下，罗森塔尔和普兰彻瑞证明的在数学上不可行的理论不再适用。现在看来玻耳兹曼偏爱将连续变量离散化的观点绝对是正确的，而且后来他越来越多地应用了这个方法(虽然他常常会认为这是虚构的，纯粹为了达到说明的目的，且这样更容易理解，参照4.2节中的内容)。然而，没有任何证据表明在论文[1868]和[1871b]中，玻耳兹曼隐含地假设了力学相空间或能量曲面的离散结构。

相反，他的论文[1871b]的背景介绍清楚地表明了他是如何设计这个假设的，正如[Brush，1976]所论证的那样。就是在引入该假说的前一段中，玻耳兹曼讨论轨迹时用了一个简单的例子：势能为 $U(x, y) = ax^2 + by^2$ 的二维谐振子。在这个系统中，位形点(x, y)沿矩形表面移动。参见图2。另见[Cercignani，1998，148]。

然后他指出，如果 a/b 为有理数(实际上，$\sqrt{a/b}$是有理数)，那么运动是周期性的。但是，如果这个值是无理数，那么在时间进程上，轨迹将“遍历这个矩形的所有点”[Boltzmann，1909，271]。他指出，在这种情况下，x 和 y 是相互独立的，因为对于每个 x 值，在矩形范围内的任意间隔内都有无限个 y 的值。玻耳兹曼考虑的是任意小尺寸上 x 和 y 值的间隔，并强调了比率 a/b 为有理数和无理数时的区别，这都表示他没有隐含地假设相空间在本质上是离散的，因为这会让这种区别变得毫无意义。

---

① 确实，他后来也在极少数的情况下提到过外部干扰，不过只是认为那“不是必需的”。可参见[Boltzmann，1895b]，另见[Boltzmann，1896，§91]。

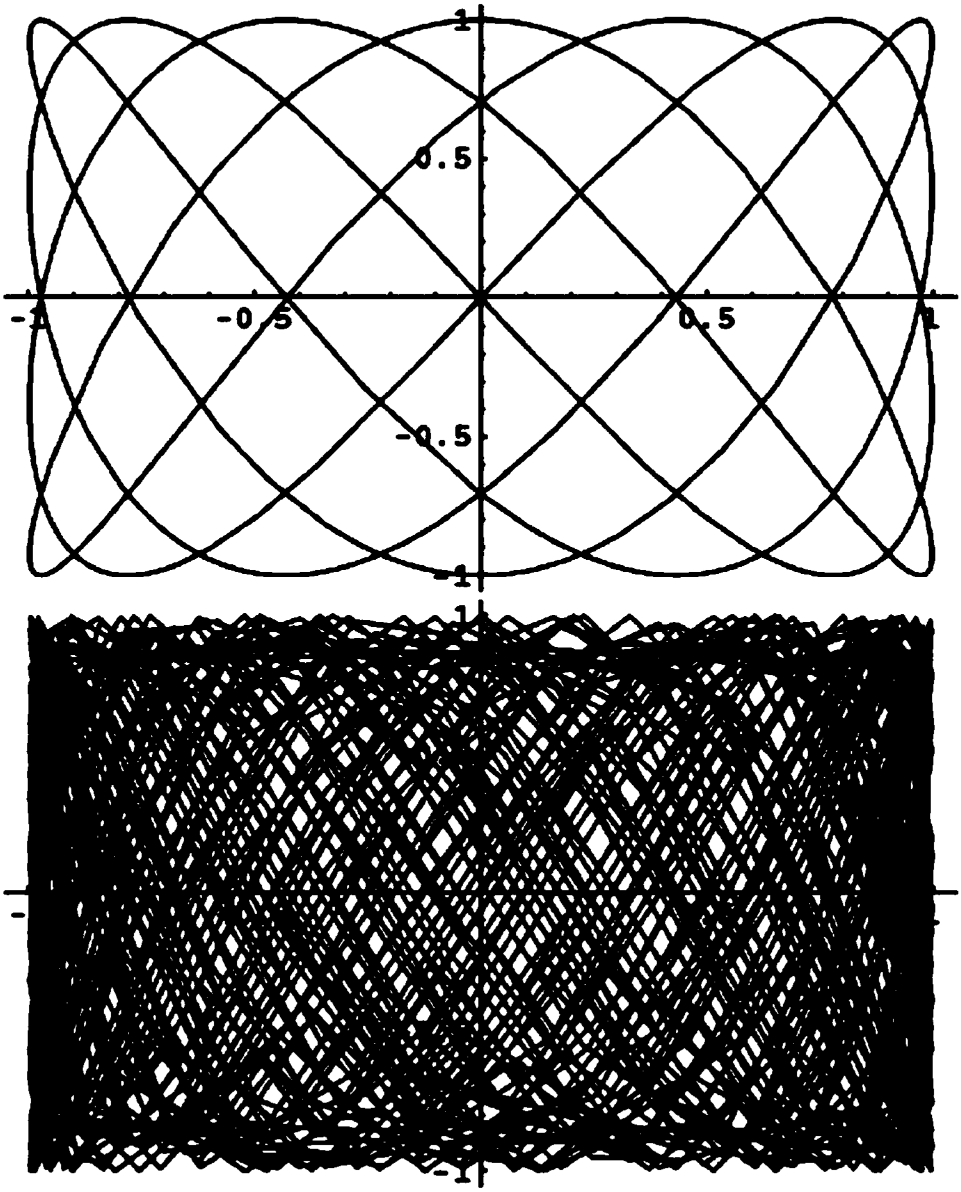


图2　一个势能为 $U(x, y)=ax^2+by^2$ 的二维谐振子构型空间中的轨迹。插图表明了两种情况下的区别：(i) $\sqrt{a/b}$为有理数（这里为4/7）的情况；(ii)为无理数（$1/e$）的情况。后者的轨迹这里只画出了一部分

现在很明显，如果用现代的方式表述，人们会说，如果 a/b 为无理数，那么轨迹在矩形上是稠密的，但并不代表它会遍历所有的点。玻耳兹曼不知道这种表述方式。事实上，他可能不清楚康托(Cantor)的见解：连续统比可数无穷多个点包含更多的点。因此，正确的表述是，在 $\sqrt{a/b}$ 为无理数的情况下，对于每个 x 的值，轨迹将在任何小的区间内经过无限多个 y 的值，这种表述可以很容易地使他(错误地)相信，在时间进程中，轨迹会遍历所有的 x 和 y 的值。

因此，看起来非常合理的是，鉴于上述讨论出现在遍历假设之前，因而玻耳兹曼对遍历假设的理解确实就是埃伦费斯特夫妇所说的准遍历假设(quasiergodic hypothesis)：假设轨迹在能量曲面上是稠密的(即任意接近于经过了每一个点)。①高维相空间上的准遍历假设在数学上是可能的。然而，准遍历假设得不出我们想要的结论：只有能量表面上稳定概率分布是微正则的。

然而，玻耳兹曼对自己的假设的有效性仍然表示怀疑，并试图从别的途径实现用力学表征热平衡的目的。事实上，无论是在他之前的论文[1871a]还是他的下一篇论文[1871c]中，都出现了替代性的观点，并明确说明它们回避了这一假设。事实上，直到 19 世纪 80 年代(受到麦克斯韦 1879 年对玻耳兹曼 1868 年的论文最后一节的评论的启发)之前，他都没有重新考虑遍历假设。在那个时候，也许是受麦克斯韦的激励，他表现得对遍历假设更有信心了。然而，在 1885 年之后，这种信心再度消失，尽管在以后的论文中他偶尔会提到该假设，但是他从来没有强调其有效性。最值得注意的是，他在 1896 年和 1898 年关于气体理论的演讲中甚至没有提到遍历假设。

总而言之，对玻耳兹曼来说，遍历假设究竟发挥了什么样的作用呢？似乎玻耳兹曼认为遍历假设是一个特殊的力学假设，它可能正确或可能不正确，这取决于系统的性质，也许还取决于它的初始状态及环境的影响。它的作用仅仅是它有助于得出一个更一般性的结果：对于令该假设为真的任何系统，其平衡态由式(45)给出，通过它我们可以重新得到在 $N\rightarrow\infty$ 极限下类似于麦克斯韦分布的分布，而不需要知道内部粒子间相互作用的任何细节，也不需要知道表征

---

① 或其他与准遍历假设相一致的假设。碰巧，玻耳兹曼的例子与“度量可递性”的测量理论假说一致，参见 6.1 节。

的系统究竟是一种气体、液体、固体还是任何其他的热物体。

正如我们在 1.4 中的讨论，埃伦费斯特夫妇在 1912 年曾表明，遍历假设起到了更为基本的作用。特别是，如果假设是正确的，（无限）长时间上的平均就等同于具有微正则分布的相平均。因此，他们认为玻耳兹曼依赖于遍历假设，是为了将时间平均等同于相平均，或换句话说，将概率的这两个含义（相对时间和相空间中的相对容积）画等号，但没有证据表明，无论是在 19 世纪 70 年代，还是更晚的时候，玻耳兹曼是沿这条路线推理的。他根本没有针对模糊地使用时间平均和粒子平均或相平均给出任何理由。可能他并没有针对这个问题考虑太多，认为这只是一种尝试。

4.2 玻耳兹曼方程和 H 定理（1872 年）

1872 年，玻耳兹曼发表了他最重要的论文之一，它包含两个著名的结果，现在分别被称为玻耳兹曼方程和 H 定理其中后者是玻耳兹曼重新建立相当于第二定律的一般性定理的依据。该论文受到众多学者的研究和评论，文献［Brush，1966］是对整个文本的翻译。因此，就当前的目的来说，简单地总结一些要点可能已经足够了。但是，就其实际内容来说，现在还存在争议。

争议主要涉及的问题是，论文得到的结果是否是运动力学方程的必然结果？或者玻耳兹曼是否明确承认允许有例外的情况存在。克莱因写道：

> 在他 1872 年的回忆录中我找不到任何迹象表明玻耳兹曼设想了后来我们称之为 H 定理的可能的例外情况［Klein，1973，73］。

克莱因认为，玻耳兹曼是由于 1877 年洛施密特对他进行了批判才承认存在这样的例外。文献［von Plato，1994］中对此表达了相反的意见。柏拉图将克莱因的观点称为“流行的观念”，他认为玻耳兹曼在 1872 年就深知他的 H 定理存在例外，因而“已经对他未来会受到的批评留有一手”。事实上，冯·柏拉图指出：

> 与被广泛认可的观点相反，玻耳兹曼在 1872 年并没有声称第二

定律和麦克斯韦分布是分子运动论的必然结果[von Plato, 1994, 81]。

因此，讨论并解决这一争议是很有意义的。

玻耳兹曼[1872]一开始就评价了概率论在气体理论背景下的作用。气体中的粒子数量是如此巨大，且它们的运动如此迅速，以至于我们能观测到的只有平均值。对平均数的确定，属于概率计算的范畴。因此，“关于热的力学理论的问题，确实就是概率计算的问题”[Boltzmann, 1909, I, p. 317]。但是，玻耳兹曼说，因此就认为热的理论包含不确定性是一个错误的观点。

他强调，我们不应该混淆没有被完全证明的断言和通过严格推导得到的概率论定理。后者是由它们的前提得到的必然结果，就像其他任何理论那样。只要我们能够观测足够大数量的案例，那它们就能被经验所证实。不过，这个条件在热的理论中不应该是个大问题，因为宏观物体中有极大数量的分子。然而，在这种背景下，我们必须加倍保证以最严格的程序执行。

因此，本论文一开始所表达的信息似乎非常清楚：玻耳兹曼即将得到的结果号称经得起双重检查且绝对严谨，尽管如此，它们仍是理论上的结果。它们与经验的关系可能不太牢固，因为任何概率陈述只可能由对足够大数量的独立数据的观测重现。因此，玻耳兹曼允许在理论和观测的关系方面存在例外，而不是在前提和结论的关系方面。

接着他指出他认为概率意味着什么，并重申了他在1868年的论文中就给出的关于相对时间段和粒子的相对数的含糊其辞的认识：

> 如果我们想……建立一个精确的理论……在所有需求之前，我们要确定各种状态的概率，即假设在很长一段时间的过程中，一个分子和与其相同的分子的不同状态的概率，以及对不同的分子来说，同时出现不同状态的概率。也就是说，我们必须计算状态介于某些限定值范围内的分子数与分子总数的关系[Boltzmann, 1909, I, p. 317]。

然而，这种模棱两可的话不是错误的。这篇论文大部分时候所预期的概率

的意义，始终是处于特定状态的分子的相对数量。只有在论文的最后[Boltzmann，1909，I，p.400]（突然）回到了概率的时间平均解释。

玻耳兹曼说，麦克斯韦和他都尝试确定气体系统的这些概率，但都没有得到一个通解。然而，仔细检验，发现"只根据运动方程得到这些概率似乎也不是那么不可能……"[Boltzmann，1909，I，p.317]。事实上，他宣布，对于由任意数量分子组成的气体来说，他已经解决了这个问题。他的目的是要证明，在这样一个气体分子系统中，无论状态的初次分布如何，它必然会演化为由麦克斯韦结构描述的分布（同上述文献的320页）。

在下一节中，他专门研究了最简单的单原子气体的情况，并且更完整地说明了他想要解决的问题。气体分子被限定在一个有完全弹性壁的容器中。只有当它们接近对方直到非常小的距离时，它们之间才存在相互作用。这些相互作用可以由弹性体之间的碰撞模拟。事实上，这些物体被模型化为硬球[Boltzmann，1909，I，p.320]。玻耳兹曼用一个依赖于时间的分布函数$f_t(\vec{v})$表示气体的状态，我们称之为"状态分布"，它提供给我们的是在每一个时刻t速度介于$\vec{v}$和$\vec{v}+d^3\vec{v}$之间的相对分子数量。①

他还陈述了两个更特殊的假设。

1. 在气体的初始状态下，每个方向上的速度是等概率的。即：

$$f_0(\vec{v}) = f_0(v) \tag{47}$$

对后面的任意时刻，这个假设也是成立的。

2. 容器内的气体在空间上是均匀的。也就是说，当速度在任意给定范围内时，位置在某个特定空间区域R内的分子的相对数量，不依赖于R在可用空间中的具体位置。

接下来玻耳兹曼用来计算在单位时间内速度为$\vec{v}_1$的粒子数量的变化的关键假设是分子混沌假设，即式(29)和式(30)。

① 事实上玻耳兹曼是用动能分布函数而非速度分布函数来表述的。我将其转成后者，因为这是现在更常见的表述。

对于现代读者来说，这个方程的结构中还存在一些不成文的假设。首先，分子的数量必须足够大，以使它们速度的(离散)分布可以很好地近似为一个连续可微函数f。其次，只有发生两体碰撞时，f才会变化。这意味着，气体的密度要较低(以便可以忽略三个粒子的碰撞)，但不能太低(这将使得碰撞太少，以至于f根本不会改变)。这两方面的要求已经强到足够放入数学上的精确形式中。现在的确切说明应该是在所谓的玻耳兹曼—格兰德极限下(参见6.4节)。最后(未说明)的假设是，所有上述假设在时间进程中一直有效。

他的目标是通过对式(29)和式(30)求微分，再对除了$\vec{v}$和t外的所有变量求积分，来构造一个微分—积分(differentio-integral)演化方程。其结果(在现代符号标记下)就是玻耳兹曼方程：

$$\frac{\partial f_t(\vec{v}_1)}{\partial t} = N\int_0^d b\,db\int_0^{2\pi}d\phi\int_{\mathbb{R}^3}d^3\vec{v}_2\,\|\vec{v}_2-\vec{v}_1\|\,(f_t(\vec{v}_1{}')f_t(\vec{v}_2{}')-f_t(\vec{v}_1)f_t(\vec{v}_2)) \quad (48)$$

它描述了在时间进程上f的变化，在初始时刻这个函数是给定的。回顾3.3节，带撇的速度被认为是不带撇的速度和碰撞几何参数的函数：$\vec{v}'_i=\vec{v}'_i(\vec{v}_1,\vec{v}_2,b,\phi)$，$d$表示硬球的直径。

H 定理

假设玻耳兹曼方程式(48)在所有时刻都成立，那么我们可以证明，经过几个著名的操作后，下面的量：

$$H[f_t] := \int f_t(\vec{v})\ln f_t(\vec{v})\,d^3\vec{v} \quad (49)$$

是关于时间的单调减函数，即：

$$\frac{dH[f_t]}{dt} \leqslant 0 \quad (50)$$

并且对于麦克斯韦分布来说，它是不变的。即：

$$\frac{dH[f_t]}{dt} = 0 \qquad (\forall t)\ \text{当且仅当}\quad f_t(v) = Ae^{-Bv^2} \quad (51)$$

玻耳兹曼在论文的第一节得到如下结论：

> 因此，我们已经严格证明，无论动能的初始分布怎样，随着时间的推移，它必然逼近由麦克斯韦所发现的形式。……这个(证明)实际

> 上得到了更多的意义，因为它适用于多原子气体分子的理论。在这里，我们也可以证明，对于某个量[H]，由于分子运动，这个量只能减小或在极限情况下保持不变。因此，我们可以证明，在由任意多物质点构成的系统中，由于原子运动，总是存在一个量，不会增大，最大就是成为一个常数因子，这个量精确地等于我在[Boltzmann, 1871c]中发现的著名积分$\int dQ/T$的值。
>
> 这里提供了一种与之前的方式完全不同的对第二定律的解析证明。至此，有人已经尝试证明对于一个可逆(umkehrbaren)循环[①]过程，$\int dQ/T=0$；但对于唯一发生在自然界的不可逆循环过程来说，却并不成立，它始终是负值，可逆过程仅仅是理想化的，我们可以或多或少地接近它，但永远不能完全实现。然而，在这里，我们马上得到的结果是，只有在限定情况下，$\int dQ/T$才会一般为负值或零……[Boltzmann, 1909, I, p. 345]。

因此，正如在他在1866年的论文中那样，玻耳兹曼声称对第二定律做出了严格的、分析的和一般性的证明。到现在为止，从我们对论文(第1节)的讨论中，我们发现克莱因的解释似乎比冯·柏拉图的更合理。我将会在4.2节中对这个争议做进一步讨论，不过我们要先关注一下这篇论文其他部分的内容。

玻耳兹曼[1872]的其他部分

第二节题为“用求和替换积分”，它主要重复了一个较早的论点，在这里假设分子的动能只能在离散集合{0, ϵ, 2ϵ, …, $p\epsilon$}中取值。玻耳兹曼表明在极限$\epsilon\to0, p\epsilon\to\infty$下，与前面的结果相同。

许多读者可能会对这种操作感到惊讶，无论从教学还是逻辑的角度看，这似乎都显得相当多余。然而，一些人已经感觉到它预示了量子理论的问世。玻

---

① “循环”这个术语在布鲁斯的翻译中被丢掉了，尽管在原文中确实谈到了“Kreisprozeß”。循环积分的符号$\oint$是后来才引进的。

耳兹曼对这种兜圈子方式的解释是这种离散的方法比前一种显得更清晰。他认为，积分只不过是对无穷多个无限小元素求和的象征性方式，而离散计算会提供更多的理解。然而，他并不认为，这更接近于物理现实。即使如此，该节最终还是取了极限，并得到了和以前相同的结果。

第三部分考虑的情况是，气体是非均匀的，即放弃上述条件 2。对于这种情况，玻耳兹曼引入了一个广义分布函数 $f_t(\vec{r},\vec{v})$，这样 $f_t d^3\vec{r}d^3\vec{v}$ 代表位置限制在 $\vec{r}$ 附近的体积 $d^3\vec{r}$ 中、速度限定在 $\vec{v}$ 附近的速度 $d^3\vec{v}$ 下的粒子的相对数量。

他得到了相应的广义玻耳兹曼方程：

$$\frac{\partial f_t(\vec{r},\vec{v})}{\partial t} + \vec{v}\cdot\nabla_x f_t + \frac{\vec{F}}{m}\cdot\nabla_v f_t =$$

$$N\int bdbd\phi d^3\vec{v}_2 \parallel \vec{v}_2 - \vec{v}_1 \parallel \Big(f_t(\vec{r},\vec{v}'_1)f_t(\vec{r},\vec{v}'_2) - f_t(\vec{r},\vec{v}_1)f_t(\vec{r},\vec{v}_2)\Big) \quad (52)$$

这里 $\vec{F}$ 表示气体存在于其中的外力场。这时 H 的形式为 $H[f_t] := \int f_t(\vec{r},\vec{v})d^3\vec{r}d^3\vec{v}$；并且得到了一个广义 H 定理 $dH/dt \leqslant 0$。

最后三节专门讨论多原子分子，旨在得到这种情况下的一般结果。当然，这样做的主要原因是，适当推广分子混沌假设。这个假设的表述在本质上与他在论文[1871a]中关于多原子分子所做的表述是相同的，后来被证明是错误的，并由洛伦兹（Lorentz）做出了纠正。我不会在这里讨论这个问题，参见[Lorentz，1887；Boltzmann，1887b；Tolman，1938]。

在论文的最后有一个有趣的段落，在那里他扩展了 H 和熵的关系。他考虑了处于平衡态的单原子气体。这个态的稳定分布如下：

$$f^*(\vec{r},\vec{v}) = V^{-1}\left(\frac{3m}{4\pi T}\right)^{3/2}\exp\left(\frac{-3mv^2}{4T}\right) \quad (53)$$

其中 V 是容器的体积。请注意，与式(27)比较，玻耳兹曼采用了温度的单位，使得 $k=2/3$。他表明：

$$H[f^*] := \int f^*\log f^*\,dxdv = -N\log V\left(\frac{4\pi T}{3m}\right)^{3/2} - \frac{3}{2}N \quad (54)$$

假设 $S = -kNH[f^*]$ 时上式与式(16)中理想气体的热力学表示一致，最多只在

积分常数上有区别。针对多原子气体，也会得出类似的结果。

备注和问题

1. 概率的作用。正如我们所看到的，H 定理形成了玻耳兹曼重建性断言的基础，该断言表明玻耳兹曼得到了一个至少对于气体而言相当于完整的第二定律(即两部分都包括在内)的新定理。与他在 1866 年的论文中的断言的主要区别在于，在推导过程中他开始极力强调概率计算的作用。很显然，这里阐述的概率概念是建立在彻底的频率主义之上的，并且他将“概率定律”看作经验性陈述。此外，概率可以用力学术语完全表述：概率分布 f 不过是分子状态处于一定限度内的粒子的相对数量。因此，他两方面的陈述之间没有冲突，一方面是，“热的力学理论中的问题其实就是概率计算中的问题”，另一方面，概率自身就是从运动方程中推导出来的。事实上，在我看来，玻耳兹曼在这篇论文中对概率作用的强调只是为了说明，概率论为讨论气体理论中的力学问题提供了一个特别有用且适当的语言。本论文中没有任何迹象表明，概率论能够起到为非力学性质(即运动方程的独立性)提供假设的作用，参见[Boltzmann and Nabl，1904，520]。

2. 分子混沌假设的作用。注意，玻耳兹曼强调他的证明具有一般性、严密性和“分析性”。他并没有强调论证中用到的特别假设。事实上，分子混沌假设后来被认为是导致 H 定理中时间不对称性的关键假设。对其介绍如下：

> 我们只能通过考虑两个粒子之间的相对速度，以一种确实单调的方式确定(碰撞的次数)。但撇开这种单调不说，由于这种考虑没有丝毫的困难，也没有任何特殊的意义，且因为结果是如此简单，我们几乎可以说它是不证自明的，我只是陈述了这个结果。见[Boltzmann，1909，I，p. 32]。

因此，很自然地，与玻耳兹曼同时代的学者一定理解他的断言，即 H 定理

必然是从力学气体模型的动力学中推导出来的。[①] 我在该论文中找不到任何证据表明他希望该断言被有所保留地解读，正如[von plato，1991，81]所认为的那样。

那么就完全没有证据表明冯·柏拉图读过这篇论文吗？冯·柏拉图引用了第2节中的一段话，其中玻耳兹曼已重复了以前的分析，他假设能量只能取离散值，并用求和替换了所有积分。当然，他得到了相同的结论，但在这里他增加了一个边注，简单地提及了非均匀气体的情况：

> 无论初始分布是怎么样的，随着时间进程，将有且仅有一种分布。……该陈述已在状态分布初始为均匀的情况下被证实。在其他情况下，它一定也是有效的，即当分子最初分布不均匀时，随着时间的推移，它们相互之间混合得越来越厉害，以至于很长一段时间后状态分布变为均匀的了。不过有一种非常特殊的情况，例如当所有分子最初沿一条直线运动，被容器壁反射后又回到这条直线上[Boltzmann，1909，I，p. 358]。

显然，最后这句话，导致了一种观点，即玻耳兹曼确实已设想了一种关于他的断言的例外情况。不过，我想说这段话并没有说服我。的确，在上述引用中玻耳兹曼已说明存在例外。但他提到这种例外只是为了推广气体最初不是均匀的情况下的结果，即丢弃了上述条件(2)。但毫无疑问的是，玻耳兹曼是在条件(1)和(2)的假设下，声称 H 定理是严格有效的。奇怪的是，当他更系统地论及非均匀气体时，见[1872]中的第3节，并没有提到关于“H 只会减少”这个断言的任何例外情况[Boltzmann，1909，I，p. 362]。

事实上，当洛施密特提出反对意见时，碰巧也是用的一个非均匀气体的例子(虽然没什么必要依赖这一点)。因此，如果真如冯柏拉图所说的那样，玻耳

---

① 确实这就是玻耳兹曼的断言被理解的方式。例如，在1888年推荐其加入普鲁士科学院的推荐信中提到其主要功绩时就认为，他证明了无论气体的初态如何，它必然会走向麦克斯韦分布[Kirsten and Körber，1975，109]。

兹曼在 1872 年的论文中就为“未来可能会出现的批评留有一手”的话，那么我们可以预料他对洛施密特的反对意见的回复应该是指出洛施密特是正确的，因为他已经预见到了这种反对意见。然而正相反，他说洛施密特是错的(见 4.3 节)。

但是，抛开玻耳兹曼是否设想了关于他的 H 定理的例外情况这个历史问题，似乎更重要的是追问玻耳兹曼为论证分子混沌假设给出了怎样的理由。要回答这个问题，必须多少有点猜测才行，因为正如我们已经看到的，玻耳兹曼指出这个假设“几乎是不证自明的”和“没有特殊的意义”，因此大概不需要作进一步的解释。但我仍然有信心作出下面的评论。

首先，我们已经看到，麦克斯韦在早期使用这个假设时，从来没有远离不充分理由原则的观点。因此，按他的方法，我们可能认为分子混沌假设表达的观点是，没有理由猜测任何将要碰撞的粒子之间存在影响或相互关系。那么这个假设是作为一个概率论假设出现的，它反映出“合理的判断”，而与力学无关。

相比之下，玻耳兹曼对麦克斯韦方法的批评表明，他不认可不充分理由这种说法。但由于分子混沌假设显然不能被看作一个力学假设——像遍历假设那样——那么唯一的选择就是，它必须取决于有关气体力学状态的特殊假设。事实上，后来在 1895—1896 年间，玻耳兹曼更加明确地承认在其 H 定理的证明过程中需要一个拟设——将它称为“假设 A”[Boltzmann，1895]或“分子无序假说”[Boltzmann，1896]——他将其表述为一个关于气体状态的假设。

然而，即使在那些年里，他也认为该假说表达的是“自由支配随意事件”[Boltzmann，1895，Abh，Ⅲ，p. 546]“即概率定律适用于计算碰撞的次数”[Boltzmann，1895b]。同样，他认为令该假设失败的态是不自然的，“是专门为违反概率定律而设想的”[Boltzmann，1896，§3]。然而，我认为不应该将这些引述的话视为是在宣称分子混沌假设是概率论自身的后果。相反，从玻耳兹曼对“概率定律”的经验性理解中，我们可以看出，玻耳兹曼认为，作为一个有关经验内容的事实，该假设“几乎总是”成立的，即使气体最初是非常远离平衡态的。

3. H 定理和第二定律。需要注意的是玻耳兹曼误解了或者故意低估了其研

究结果的意义。玻耳兹曼方程和H定理涉及的都是存在于一个固定的容器中的大量气体，且气体物质隔离于它的环境进行演化。毫无疑问在这个过程中气体存在热交换，更不用说在一个不可逆的循环过程中。因此在前文的引述中将他的观点与克劳修斯的积分$\int dQ/T$（即前面式(18)中的$\oint đQ/T$）做对比，实际上是完全不恰当的。

玻耳兹曼结果的真正意义在于：(i)将熵的概念推广到了非平衡状态①；(ii)指出随着孤立的气体从非平衡态走向平衡态，这种非平衡态下的熵kH会单调增加。因此，它与第二定律的关系是间接的：一方面，玻耳兹曼证明的结果比需要证明的更多，因为第二定律不涉及非平衡态的熵，也没说它会单调增加；另一方面他证明的结果还不够，因为玻耳兹曼没有考虑在一般绝热过程中熵是否增加。

4.3 玻耳兹曼(1877a)：关于可逆性的反对意见

按照[Klein，1973]的观点，玻耳兹曼似乎已经不满足于他在1871年和1872年论文中的讨论，而在接下来的几年中把注意力转向了其他方面。1875年，他转回气体理论的研究，讨论了在外力作用影响下气体的玻耳兹曼方程的扩展。但该论文不存在任何根本性的思想变化。但是，它确实包含了一些进一步的阐释，例如，它第一次提到推导玻耳兹曼方程要求气体是稀薄的，以保证三个或更多粒子之间的碰撞可以忽略。

然而，1875年论文中的一个结果，导致了两年后他与洛施密特之间的辩论。玻耳兹曼表明，式(52)意味着处于外力场(如地球引力)中的平衡气体在所有高度上应具有相同的平均动能，因此，应该保持一个统一的温度，而其压力和密度则自然会随高度而变化。这个结论与直观上认为的分子向上运动时，它们必须克服重力场做功，并因此而在更高的高度上具有较低的动能的观点一致。

---

① 玻耳兹曼强调他所说的熵应该被理解为是从热力学熵扩展到非平衡态下的熵[Boltzmann，1877；1909，Ⅱ，p.218；1896，§5]。当然并不能保证这种扩展对非平衡态熵来说是独一无二的。

这里，玻耳兹曼(1875 年)不是第一次得出相反的结论，洛施密特也不是第一个挑战他观点的人。麦克斯韦和格思里在 1873 年已就同一主题进行了辩论。但实际上他们的主要争论点我们并不十分关心。洛施密特和玻耳兹曼之间的讨论的另一个问题却相当重要，其中洛施密特的讨论内容只出现在旁注中。洛施密特讨论的情况是，通过考虑处于均匀引力场中的气体容器，最初除一个原子外，其他所有原子都静止在容器的底部。这个唯一运动的原子，会通过碰撞激发其他原子，使它们开始运动，直到出现由麦克斯韦分布描述的“稳定态”。他继续阐述道：

顺便说一下，我们应该谨慎地对待这种说法，即一个系统从任意初始状态开始，达到所谓的稳定态，那么就能一直完好地保持这种稳定态。我有充分的理由相信，我们可以预测这种稳定态只能保持很短一段时间。

事实上，如果在上述情况下，经过一段足以达到稳定态的时间 τ，我们突然假定所有原子的速度方向反转了，那么我们将得到一个初始状态，它看起来与稳定态具有相同的特征。在相当长的一段时间内，这会是正确的，但是渐渐地稳定态会退化，经过时间 τ 后，我们将不可避免地回到我们最初的状态：即只有一个原子吸收了系统的所有动能……，而所有其他分子都静止在容器的底部。

很显然，在每一个任意的系统中，事件的过程必将会退化，如果它的所有元素的速度都会反转的话[Loschmidt，1876，139]。

玻耳兹曼的回应[1877a]

玻耳兹曼对洛施密特的回应多少有点混乱。一方面，他承认，洛施密特的反对是“相当巧妙的，对正确认识第二定律具有十分重要的意义”；另一方面，

他指责其为“谬论”和“诡辩”。[①]但是，在后来的两页中，这种说法又变为“最重要的，因为它说明了如何将第二定律和概率论紧密联系起来”。

回应的要点是这样的。首先，玻耳兹曼以极为清晰的方式抓住了问题的核心：

> 从物体的性质和它们彼此之间力的相互作用规律出发，而没有假设初始条件，所有企图证明
>
> $$\int \frac{dQ}{T} \leqslant 0 \qquad (55)$$
>
> 的尝试必然是徒劳无功的[Boltzmann，1909，Ⅱ，p. 119—121]。

这里提出的观点通常被称为可逆性反对，并且自从 H 定理(在 19 世纪 90 年代才出现这种叫法)于 1872 年作为对 $\int \frac{dQ}{T} \leqslant 0$ 的一般性证明出现以来，就意味着它是会失效的。然而，玻耳兹曼旨在说明，这种反对是一种谬论。他的论点可以被分为五个要点。

1. 承认给不出证明。玻耳兹曼说，无法证明每个分布都会绝对必然地发展演化为均匀分布，他声称这个事实“已经由概率论表明”。事实上，他认为，即使一个相当非均匀的状态分布，很大程度上虽然是不可能的，但也并非完全不可能。因此，他承认，存在 H 不断增加的初始状态，同时也存在 H 不断减小的初始状态。承认这个，当然会很难与他宣称要达到的目的相一致，即他想表明需要对初始状态做出假设这个结论是荒谬的。

请注意，这段话预示了一个重要的观念转变。鉴于 1872 年的论文认为状态分布 f_t 定义了概率(即分子速度)，而这里状态分布本身就是某个能够表示较高或较低程度“可能性”的量。这就是说：概率归因于状态分布，即状态分布本

---

① 事实上，玻耳兹曼称这个结论为——通过正确的各种手段和标准得到的(在我看来，他预料不到反对的理由)一个谬误。事实上，当我们比较这个判断与他在《气体理论讲义》(*Lectures on Gas Theory*)中的话[1898，442]时，即“[H 定理]的单边性唯一地有赖于初始条件”，就会很容易发现，玻耳兹曼有多么不理解洛施密特的反对意见。

身被视为一个随机变量。在他(1877b)的论文中，这种观点的转变显得更加明确，我们将在下面的第4.4节中进行讨论。

2. 反思"概率"的意义。玻耳兹曼认为，每个状态分布，无论是均匀的还是非均匀的，同样是不可能的。但与非均匀状态分布相比，均匀状态分布要多"无限多"个。在这里，我们见证了另一种概念上的转变。在1872年术语"状态分布"指的是函数$f(\vec{v})$或$f(\vec{r},\vec{v})$代表的状态与处于各种分子状态的分子相对数。在这个意义上，当然只能有一种均匀状态分布，即麦克斯韦分布函数(53)。但是，由于玻耳兹曼在这里声称有很多种，显然他认为术语"状态分布"表示更复杂的类型，这包括单个分子的速度和位置，以至于分子的不同排列就会产生不同的状态分布。这就是说，他在我们现在所谓的微观状态的意义上使用这个术语，在几个月之后，他在论文[1877b]中称之为"Komplexion"，在该文中他用"状态分布"这个术语表示宏观状态。

注意到玻耳兹曼假设每个Komplexion都是同等可能的(或不可能的)，以便一个特殊的状态分布的概率由相对数来确定。事实上，他注意到，通过确定它们数量的比率来计算状态分布的概率很可能是有意义的，随后他在论文[1877b]中还采纳了这个建议。

确实，这标志着另一个概念变化。概率归因于状态分布，而不是由其来定义，它们由等概率假设来确定。玻耳兹曼并没有明确地提出这个假设。在第3.1节的讨论中，我们能够推测出在他的思想中一定有类似于拉普拉斯的不充分理由原则那样的理论，这使得根据我们信息所认为的任何两种同等可能的情况，同样也是等概率的。但是，这将意味着一次更大的概念转变，并不仅仅是因为总的来说玻耳兹曼是一个关心概率论的频率论者。而且，不充分理由原则或任何类似的假设，只有在概率是一个非机械概念这样的观点下才有意义，它反映了我们关于一个系统的信念或信息。我无法找到任何证据表明，他接受了这样的观点。当然也可以猜想，他默默地回到了遍历假设。但是，这个猜想似乎也是不可能的，因为自1871年之后他就一直避免提及这个假设。

3. 关于演化的断言。玻耳兹曼说："只有从均匀(微观)状态分布比非均匀状态分布多得多这个事实出发才能得到结论：随着时间推移，分布有较大概率

会变得均匀。"更明确地，他说道：

> 我们能证明，经过很长一段时间的演化后，无限多的初始状态，会实现更高均匀度的状态分布，而非较低均匀度的状态分布，即使是在后一种情况下，这些状态经过更长时间的演化，也将会变得更加均匀[Boltzmann，1909，Ⅱ，p. 120][①]。

请注意，这是关于微观状态演化的断言。事实上，这是后来埃伦费斯特夫妇称之为统计学 H 定理中的第一种情况，不过也许我们称其为 H 定理的统计学解读会更好些，因为尽管玻耳兹曼做出了断言，但他并没有提供证明。

4. 洛施密特初态的(不)可能性。 玻耳兹曼坚持认为洛施密特考虑的初始条件，仅有极小的可能性。这是因为它是通过一个非均匀微观状态的时间演化和速度反向而获得的。由于时间演化和速度反向都是一对一的映射(或更重要的是它们保持刘维尔测量不变性)，这些操作不会影响状态的数量或概率。因此，洛施密特状态的概率就等于是被构造出来的特殊的非均匀状态的概率。但是，由上述第 2 点可知，均匀状态比非均匀状态多无限多个，所以洛施密特状态的概率是非常小的。

5. 从(不)可能性(probability)到更(不)可能(possibility)。 玻耳兹曼回应的最后一个部分宣称概率是非常小的，小到实际上是不可能的。

玻耳兹曼基于以上五点得到的结论是，洛施密特选择的态被认为实际上几乎是不可能的。请注意这完全是一个静态的说法，也就是说，它的逻辑仅仅依赖于第 1，2，4 和 5 点，并不需要对演化作任何假设，与动力学演化中态(或测量)守恒的一般性质无关。事实上，论证中并没有用到上面的第 3 点，即 H 定理的统计解读。

作为结果，该论证虽然是完全一致的，但它表达了比玻耳兹曼可能希望的更多的内涵。同忽略洛施密特初始状态的原因一样，我们也应该排除其他的非均匀态。特别地，我们应该分配相同的概率给洛施密特的初始状态而不考虑速

① 在译文[Brush，2003，366]中丢掉了"即使是在后一种情况"后面的内容。

度的反向。但是，那些状态能出现在实验室中，并且据推测，不应该被认为实际上是不可能的。事实上，如果我们采用的规则是，考虑到所有非均匀态的低概率，它们就能被忽略的话，那么我们最终将只考虑均匀态，即理论将会简化为对平衡态的描述，H 定理会简化为 $dH/dt=0$，从而失去了所有的时间不对称性。

这种针对洛施密特反驳意见的回应的确是显得太低劣了。我们从玻耳兹曼的论证中想要看到的是关于时间演化的假设发挥了更大的作用，以此来证实他对 H 定理的统计解读。

总结：从这个角度我们应当看到，玻耳兹曼甚至更强烈地强调第二定律和概率论之间的密切关系。即便如此，这个关系究竟是什么也并不总是明确的。此外，我们会质疑他关于初始状态的概率性的思考是否命中要害。就像力学那样，概率论同样不要求时间的方向。

可逆性问题的明确出现是玻耳兹曼在牛津演讲之后巴宝利(Burbury)[1894a]和布莱恩(Bryan)[1894]的论文中，他们在《自然》(*Nature*)的专栏上展开了激烈的争论。他们指出，分子混沌假设就已经包含了一个关于时间不对称性的假设。

事实上，这个假设要求满足$(\vec{v}_1,\vec{v}_2)\longrightarrow(\vec{v}_1{}',\vec{v}_2{}')$那种碰撞的次数是与乘积$f(\vec{v}_1)f(\vec{v}_2)$成比例的，其中，$\vec{v}_1,\vec{v}_2$是碰撞前的速度。如果我们想将该条件替换为碰撞的次数与碰撞后的速度$\vec{v}_1{}',\vec{v}_2{}'$的函数的乘积成比例的话，那么根据类似的推理，我们将会得到，$dH/dt\geqslant 0$。当然，现在的问题是，为什么我们应该更倾向于一个假设而非另一个，如果不会因此陷入某种双重标准的话。我参阅了[Price(普瑞斯)，1996]关于这一危险的详细讨论。有一件事是肯定的，那就是，任何对假设的这种倾向都不能单独从力学或概率论中得到。

4.4 玻耳兹曼[1877b]的组合论证

玻耳兹曼接下来的文章[1877b]通常被认为与他之前采用的概念基础产生了重大分歧。事实上，概念的转移已经隐含在他对洛施密特的回复中，而在这篇文章中则变得更加明确。事实上，根据[ter Haar, 1955, 296]和[Klein, 1973, 83]，正是这篇文章标志着运动学理论向统计力学的转变。此外，这篇

论文建立了熵和“概率”之间的联系“$S = k\log W$”，后来被称为“玻耳兹曼定理”，并刻在了他的墓碑上。

玻耳兹曼在他的论文一开始就指出，他的目标是要澄清第二定律和概率计算之间的关系。他指出他曾反复强调第二定律是与概率计算相关的。特别地，他指出在他1872年的一篇论文中通过说明一定的量（即H）只能减小来证实了这一关系，因此，它必然在热平衡态下取得其最小值。然而，概率论与第二定律的这种关系在他之后的论文[1877a]中变得更加明显。玻耳兹曼指出，他现在将解决那篇论文中提到的问题，即通过确定状态数量的比率来计算状态的概率分布。

他还宣称，当系统从不太可能的状态开始演化的话，那么它总是朝着更可能的状态的方向进行，直到它达到最可能的状态，即热平衡态。当应用于第二定律时，他说：“我们能够将通常被称为熵的量，与所讨论的状态的概率等同起来。”还说：“根据目前的解释，（第二定律）只是陈述了一个观点，即一个复合系统的总状态的概率总是增大的。”他指出[Abh. II，pp. 165—6]，究竟所有这一切意味着什么，在本文的后面部分中将进行说明。

组合论证

为简单起见，我们采用埃伦费斯特夫妇的术语，将该论证重新表述如下。除了Γ，即包含总的气体系统的可能状态x的力学相空间之外，我们考虑所谓的μ空间，即单个分子的状态空间。对于单原子气体，这只是一个将$(\vec{r},\vec{v})$作为坐标的六维欧氏空间。每个机械状态x，我们可以将μ空间中由N个点组成的集合与每个力学状态x联系起来；其中每个点代表一个分子。

现在，将μ空间分割成m个不连续的单元格：$\mu = \omega_1 \cup \cdots \cup \omega_m$。这些单元格是根据坐标划分的，且为大小相等的矩形。另外，假设单元格ω_i中每个分子的能量的值为ϵ_i，且只取决于i。对于每个x，此后也称之为微观态（玻耳兹曼使用的术语是Komplexion），我们定义宏观态或“态分布”为$Z := (n_1, \cdots, n_m)$，其中n_i为分子状态位于单元格ω_i中的粒子的数量。宏观态和微观态之间的关系显然不是唯一的，因为许多不同的微观态（例如通过对分子序列进行重新排列得到的微观态）会导致相同的宏观态。我们可将每个特定的宏观态Z_0与相应

的微观态的集联系起来：

$$\Gamma_{Z_0} := \{x \in \Gamma: Z(x) = Z_0\} \tag{56}$$

该集的相空间体积 $|\Gamma_{Z_0}|$ 是与保持宏观态 Z_0 不变的粒子排列方式的数量成比例的。事实上，当单元格 ω_i 的六维量为 $\delta\omega$，即每个单元格都相同时，集合 Γ_Z 的相空间体积为：

$$|\Gamma_Z| = \frac{N!}{n_1! \cdots n_m!}(\delta\omega)^N \tag{57}$$

此外，假设 $n_i \gg 1$ 对于所有 i 都成立，且对阶乘应用斯特林(Stirling)近似，就会发现：

$$\ln\Gamma_Z \approx N\ln N - \sum_i n_i \ln n_i + N\ln\delta\omega \tag{58}$$

这种表达式事实上是与 H 函数的离散近似成比例的。事实上，令：

$$n_i = Nf(\vec{r}_i, \vec{v}_i)\delta\omega \tag{59}$$

这里 $(\vec{r}_i, \vec{v}_i)$ 在 ω_i 中表示点的坐标，我们可知：

$$\begin{aligned}\sum_i n_i \ln n_i &= \sum_i Nf(\vec{r}_i, \vec{v}_i)\ln(Nf(\vec{r}_i, \vec{v}_i)\delta\omega)\delta\omega \\ &\approx N\int f(\vec{r}, \vec{v})(\ln f(\vec{r}, \vec{v}) + \ln N + \ln\delta\omega)d^3\vec{r}d^3\vec{v} \\ &= NH + N\ln N + N\ln\delta\omega\end{aligned} \tag{60}$$

因此，以式(58)的观点看：

$$-NH \approx \ln|\Gamma_Z| \tag{61}$$

此外，由于玻耳兹曼已经将 $-kNH$ 等同于一个宏观态的熵，因此我们也能将熵视作与相空间中相应区域的体积的对数是成比例的。现在，$\ln|\Gamma_Z|$ 经常被称为玻耳兹曼熵。

玻耳兹曼接下来考虑的问题是，对于给定的总粒子数 N 和总能量 E：

$$N = \sum_{i=1}^{m} n_i, \qquad E = \sum_{i=1}^{m} n_i \epsilon_i \tag{62}$$

如何选择 Z 才能使区域 Γ_Z 最大化。可以很容易地通过拉格朗日(Lagrange)乘子法解决该问题。在斯特林近似式(58)下我们发现：

$$n_i = \mu e^{\lambda \epsilon_i} \tag{63}$$

这是麦克斯韦分布的离散形式。这里，μ 和 λ 由约束式(62)中的 N 和 E 来确定。

玻耳兹曼提出用最大体积的宏观态来表示平衡态。更一般的是，他还将体积作为宏观态的“概率”或“可置换性”。因此，他在这里将第二定律表达为系统的一种发展趋势，认为系统朝着更可能的宏观态演化，直到达到平衡态这一它所能达到的最可能的状态。

备注和问题

1. 力学的作用。在当前的论证中，并没有做出力学假设。特别是，遍历假设是否成立以及粒子如何碰撞与该论证并不相关。乍一看，这似乎使得当前的论证比之前的更普遍。事实上，玻耳兹曼在论文[Boltzmann，1909，II，p. 223]的结尾处表明，同样的论证也适用于稠密气体，甚至固体。

然而，应该注意到，下面的这个假设是很强的，即总能量可表示为 $E=\sum_i n_i\epsilon_i$ 的形式，其中每个粒子的能量只取决于它所在的单元格，而与其他粒子的状态无关。只有当所有粒子之间都不存在相互作用时，该假设才是不依赖于数目 N 而成立的。因此该论证的有效性实际上仅限于理想气体，参见[Uhlenbeck and Ford，1963]。

2. 单元格的选取。乍看之下，人们也许希望，将 μ 空间分割成单元格的程序仅仅是一个技巧或教学方式，最终可以通过取极限 $\delta\omega\to 0$ 的方式将其消除，类似于玻耳兹曼 1872 年的论文中的那个步骤。但这个希望破灭了，因为表达式(58)是发散的。事实上，如果我们没有采用有限分区的方式，那么使用组合论证的整个前景将会消失。并且，按照位置和速度变量赋予所有单元格相同的体积这个特殊选取方式并不是不证自明的，正如玻耳兹曼自己所表明的那样。事实上，在他着手这里给出的论证之前，他就曾讨论过用能量而不是用位置和速度来描述粒子的特征。这导致他按照能量将 μ 空间划分成具有大小相等的 $\delta\epsilon$ 的单元格。然后，他表明，组合论证对于重现三维空间中运动粒子的麦克

斯韦分布来说是失败的。① 在论文[Boltzmann, 1909, II, p. 190]中，他又通过转为按位置和速度选择具有同样大小 $\delta\omega$ 的单元格的方式对这种失败进行补救。后一种选择显然是“正确的”，如果在它能够导出所需结果这个意义上说的话。然而，既然选择显然不能被归为一个习惯问题，因此如何证明该问题仍然是没有得到解答的。

现代的评论者们在寻找能够导致对这些单元格的大小做出选择的动机的方向时，彻底分为了两派。有些人认为应该根据测量仪器的实际分辨率或人的观测能力做出选择。这样做实际上是否有利于划分具有相同相空间体积的单元格的这个问题几乎并没有被提及。其他人[Popper(波普尔), 1982; Redhead, 1995]则拒绝诉诸于观测能力，理由是这会将一个“主观的”或“人类中心主义的”元素引入到不可逆解释中，也可参见[Jaynes, 1965, Grünbaum(格拉堡), 1973; Denbigh and Denbigh(登比夫妇), 1985; Ridderbos, 2002]。

3. 微观与宏观。在该论证中最本质的步骤是区别微观态和宏观态。这确实是一个决定性的新要素，它允许玻耳兹曼对概率的概念和作用做出完整的重新解释。

1872 年及这之前，状态的分布 f 被认为等同于概率，即分子状态，参照第 4.2 节中的注释。另一方面，在当前的工作中，分布或其离散类比 Z，则是对气体宏观态的描述，而概率被分配给了宏观态。从本质上讲，状态分布的角色已经从定义一个概率测量变成了一个随机变量。它以前的角色被一个新的想法所取代：概率并不是分配给粒子的，而是将气体作为一个整体而分配给其宏观态的，并以其在相空间中所对应的体积作为测量对象。

另一个创新点是玻耳兹曼已经改变了他对平衡态的认识。与先前认为平衡态是稳态的不同，在玻耳兹曼的新观点中，它被看作是占据最大体积的宏观态(即相空间中的一个区域)。其结果是，处于平衡态的系统并不需要保持在平衡态：在时间进程中，该系统的微观态可能在这个平衡区域附近出现涨落。玻耳

① 问题在于对理想气体来说，所有能量都是机械能，$\delta\epsilon \propto v\delta v$。另一方面，对于三维粒子来说，$\delta\omega \propto v^2\delta v$。因此式(59)和式(63)中的函数 f 与 v 的关系并不一致。正如玻耳兹曼注意到的，在二维情况下(即在平面上运动的圆盘)，这两种选择是相容的。

兹曼在其论文[Boltzmann, 1878]中简单地研究了这种涨落的概率。在30年后，由爱因斯坦和斯莫卢霍夫斯基(Smoluchowski)给出的关于涨落现象的实验预言为统计力学提供了引人注目的经验性支持。

4. 但是如何演化? 或许这是最重要的问题。那么玻耳兹曼针对洛施密特的反驳的论文[1877b]和他的首次回应[1877a]之间究竟有什么关系? 首次回应(参见第4.3小节)可以被解读为宣告了其进一步研究的两个主题:

> 从不同状态分布的相对数出发，我们甚至能计算其概率。这可能会产生一个确定热平衡态的有趣方法[Boltzmann, 1909, II, p. 121]。

这是一个关于平衡态的问题。宣告的第二个主题是玻耳兹曼所说的:“这种情况完全与第二定律类似。”[Boltzmann, 1909, II, p. 121]。因为均匀分布比非均匀分布多得多，因此系统应该是极不可能从均匀状态分布演化为非均匀状态分布的。这是一个关于演化的问题(参见4.3节中的第3点)。换句话说，我们想要看到诸如统计性H定理那样的理论实际上是成立的。

玻耳兹曼[1877b]被广泛地理解为沿着所宣告的这两个主题而进行的后续研究。事实上，玻耳兹曼在论文[Boltzmann, 1909, II, p. 165]的引言中重复了上面的引文，就表明他会考虑这个问题。他确实详尽地研究了该问题。他曾指出:

> 我们的主要目的不是停留在讨论热平衡上，而是研究热力学第二定律与概率的关系[Boltzmann, 1909, II, p. 166]。

因此，论文[1877b]的主要目的显然是要解决有关演化的问题，以及说明演化如何与第二定律相关。事实上，这时人们很自然会想到，因为对可逆性的反驳，就是关于演化的问题。即便如此，一个显著的事实是，论文[1877b]几乎没有涉及其所说的“主要目标”。事实上，我只在该论文的剩余部分找到一段话，提及了它与第二定律的关系。

这段话出现在论文的第五部分[Boltzmann, 1909, II, pp. 216—7]。文中

说明处于平衡态下对单原子气体的“可置换测量”$\ln|\Gamma_Z|$（在玻耳兹曼标记法下为 Ω）是与热力学熵成正比的，且其系数为任意积分常数，之后，文中也指出通过选择适当的常数可以得出如下结论：①

$$\int \frac{dQ}{T} = \frac{2}{3}\Omega\left[= \frac{2}{3}\ln|\Gamma_Z|\right] \tag{64}$$

并且：

> 众所周知，当物体构成的系统经过可逆变化时，所有这些物体的熵的总和保持不变，但在此过程中包括不可逆（nichtumkehrbar）变化时，总熵必然会增加。这可以由我们所熟知的在不可逆转循环过程中 $\int dQ/T$ 为负这个事实得到。根据式（64）的观点，对所有物体所做的全部可置换性测量的总和 $\sum\Omega$，或它们的总的可置换性测量也必须会增加。因此，可置换性是一个与熵相同的量，它取决于相乘性常数或相加性常数，但熵在一个不可逆物体，原文如此，应理解为“过程”中仍有意义，在这一进程中，它不断地增加［Boltzmann，1909，II，pp. 216—7］。

如何解决有关演化的问题呢，它是否针对可逆性异议给出了一个令人满意的反驳呢？在过去的文献中，关于玻耳兹曼的回应在实际上涉及的和完成的内容来说，至少存在四方面的看法。

4α. **依赖于微观和宏观的区分。**最近出现了一种观点，例如戈德斯坦（Goldstein）［2001］认为玻耳兹曼按照他自己的论证，就已经充分且直接地解释了为什么熵应该趋于增加。特别是，这种观点认为，与非平衡相位点的集合相比，所有的平衡相位点的集合 Γ_{eq} 占据了有压倒性优势的大的相位空间这个事实，就已经提供了足够的论据。

① 事实上，式（64）最接近于他墓碑上所刻的那个著名的公式，因为 $\Omega = \ln W$，且玻耳兹曼采取的温标使得 $k = 2/3$。

> 对于能量 E 的一个非平衡相点 x，支配由 x 到 x_t 的运动的哈密顿动力学，如果能够避免以相当快的速度将 x_t 引入到 Γ_{eq} 中，并避免它在其中保持非常长的时间的话，那该力学肯定是令人难以置信的特殊——当然除非 x 本身是人难以置信的特殊[Goldstein，2001，6]。

实际上，这种观点可能会被认为是忠于史实的。正如我们所看到的，[Boltzmann，1877a]确实声称，演化趋向平衡态的大概率性确实是从状态数量的巨大差异中得出的。

从现代视角来看，这种观点的主要困难在于我们很难坚持认为它是适当的。某些状态会演化为其他的状态，并不仅仅是因为后者具有更多的数量，或者因为它们构成了一个更大程度的集合。系统的演化只依赖于其初始状态和哈密顿量。有关演化的问题只能通过动力学的方法得到答案，而不是仅仅通过对集合的测量。举一个极端的例子，x_t 的轨迹，即集合 $\{x_t: t\in\mathbb{R}\}$ 无论如何都是一个测度为零的集合，因此是非常特殊的。与此相反，它的补集，即给定轨迹以外的状态集则是巨大的：它的测度为 1。当然，我们不能断言系统会不可避免地进入到轨迹以外的状态集上。另一个例子是非相互作用粒子构成的系统，例如理想气体。在这种情况下，所有单个粒子的能量都是守恒的，因为这些守恒量，相点只能存在于非常有限的相空间区域内。①

当然，从中得到的经验就是，为了得到关于系统为什么会倾向于从非平衡态演化到平衡态这个问题的令人满意的论证，我们就应该对其动力学做出一些假设。在任何情况下，诸如“合理的”或“荒谬的”这样的判断在一定程度上仍然是一种尝试。可逆性反驳要求的是数学证明（正如谚语所说，这是要说服不讲道理的人）。

4β. 依赖于遍历假设。看待这个问题的第二种观点，也许是最广为人知的，是由埃伦费斯特夫妇提出的。从本质上来说，他们认为，在玻耳兹曼的论证中他以某种方式依赖于遍历假设。

---

① 我们应该注意，多少有些讽刺意味的是，在前面的评论 1 中，这是仅有的一个与玻耳兹曼的论证相容的例子。这导致了辛钦的“方法论悖论”。

如果遍历假设成立，那么某个状态在相空间中能量曲面上的不同区域所存在的时间与其体积成正比。这就是说，系统在沿其轨迹演化的过程中，体积很小的区域对应于状态高度非均匀分布的情况，轨迹只是偶尔才会经过；而体积较大的区域对应于状态的均匀分布，轨迹则会经常出现于此。

这就会使下面的情况看上去是合理的，即如果系统从一个非常小的区域(不大可能的状态)开始演化，它就会表现出一种朝着具有压倒性优势的平衡态演化的趋向。当然，这种"趋向"在一定意义下会被解释为：相同的遍历假设意味着，系统不可能永远处于平衡态，而是必然会在平衡态附近涨落。事实上，我们将不得不说从不可能态向可能态演化的这种趋势，其自身是一个概率性事件：如同对于大部分初始状态，或在大部分时间内成立的那种理论一样，或者就像有关平均行为的某些理论或其他形式那样。总之，我们将希望得到 H 定理的某种统计版本。统计性的 H 定理究竟应该表达什么，这在埃伦费斯特夫妇的观点下仍然是一个悬而未决的问题。事实上，他们区别了几种不同的解释，所谓的"浓度曲线"和"H 曲线束"[Ehrenfest and Ehrentest-Afanassjewa，1912，pp. 31—35]。

现在，不可否认的是，埃伦费斯特夫妇对玻耳兹曼意图的解读有一些明显的优势。特别是，即使尚未有人在遍历假设或在度规传递性假设的基础上成功地证明统计性 H 定理(参见第6.1)，但是我们仍希望 H 定理的一些统计版本为真。

这里存在的一个问题是，玻耳兹曼在他的论文中使用的假设限制只能应用于非相互作用分子的情况，而对它们来说遍历假设被证明为不成立。但更重要的是，玻耳兹曼在[1877b]中很显然并不是沿着这样的路线来论证的。的确，他从没提到过遍历假设。事实上，他后来在评论论文[1877b]和[1868]中分别提出的遍历假设之间的关系时说道：

> 在那时候，即在[1877b]中……我并不希望涉及系统是否能够遍历与能量方程相容的所有可能态这个问题[Boltzmann，1881a，Abh. Ⅱ，p. 572]。

4γ. **依赖于 H 定理**。第三种观点，也是笔者一直到最近都坚持的一种观

点，就是在[1877b]中玻耳兹曼仅仅依赖于1872年提出的 H 定理的有效性。毕竟，正是1872年的这篇论文将 $-NH$ 解释为熵(以相乘性和相加性常数为模)，基于所谓的熵永不减少定理。[1877b]论文中提出了一个新的方案，即将宏观态的熵与 $\ln|\Gamma_Z|$联系起来。但是这个方案，如果说不是从式(61)，$\ln|\Gamma_Z|$(近似地)等于 $-NH$ 得到的话，至少也是由其促成的，因为在论文[1872]中 $-NH$ 被解释为熵。因此，看起来似乎我们可以合理地推测玻耳兹曼的思想依赖于该论文的结果，以及依赖于不大可能的态会向可能的态演化这个断言，即$\ln|\Gamma_Z|$表现出随时间增加的趋势，如同他已经证明的依赖于 H 定理那样。① 这种解读的缺点是它使得玻耳兹曼对可逆性反驳的回应显得相当不堪一击。由于在[1877]中阐述的该反驳质疑 H 定理的有效性，因此以这个定理的有效性为前提的对此所做的回应是无力的。

4δ. **绕过 H 定理。**[Janssen(扬森)，2002]给出了一种不同的解读。他指出："在玻耳兹曼1877年的论文中，系统永远不会从高可能的态演化为低可能的态，这种说法只是对第二定律的一种新表述，而不是作为 H 定理的结果。"见第13页。事实上，论文[1877b]从没有明确地提及 H 定理。然而，我们是否认识到这点却并不十分肯定。在前一篇论文[1877a]中也没有提到该定理，而只是讨论了"证明 $\int\frac{dQ}{T}\leqslant 0$ 的那些尝试"。不过，这通常被视为是隐含地涉及了现在被称为 H 定理的那些内容，只是在当时还没有一个专门的名字。事实上，H 定理本身是在1872年才被描述为是对 $\int\frac{dQ}{T}\leqslant 0$ 的一种新的证明。因此，[1877b]中并没有明确提到 H 定理，这个事实本身并不能确切地说明他不打算提及它。

即便如此，他在介绍熵增加原理时将其看作是众所周知的，而且也根本没有提到1872年的论文，这确实使扬森的解读显得好像是有道理的。因此，或许玻耳兹曼仅仅是利用第二定律的经验合理性作为该陈述的依据，而与气体分

① 玻耳兹曼后来在1896年的阐释正是沿着这一方向进行的，从而支持了这种推测。

子运动论中的任何命题都无关。[①] 这当然会更严重地侵蚀以下观点的基础，即玻耳兹曼提出的 H 定理的一个统计版本，或者说任何关于时间演化概率性的某个定理。

不可逆反驳关注的并不是现象上的第二定律与 H 定理之间的关系，而是 H 定理和运动力学方程之间的关系。因此，即使扬森的解读与玻耳兹曼的观点一致，它也不会让[1877b]论文针对洛施密特的反驳变得令人信服。

4ϵ. 瓮的类比——定义上的成功？ 冒着过度探讨该问题的风险，我还想给出第五种解读。玻耳兹曼在[1877b]中详细地讨论了从一个瓮中重复抽签的过程。从现代术语上讲，他考虑了一个伯努利过程，即针对一个具有有限数量的可能结果的实验，进行的一系列独立重复的分布式尝试。形象地讲，就是考虑一个其中放置了 m 个不同标签的瓮，我们对其进行 N 次抽签为一个系列，那么在该系列中第 i 个标签被抽到 n_i 次（$\sum_{i=1}^{m} n_i = N$）。他通过一个"状态分布"$Z=(n_1, \cdots, n_m)$来重复此序列。在该讨论中，这些状态分布的概率第一次被认为等同于（规范化）排列的数量，而 Z 可由排列确定。换句话说：

$$\mathrm{Prob}(Z) \propto \frac{N!}{n_1! \cdots n_m!} \tag{65}$$

但在该讨论中[Boltzmann，1909，II，p. 171]，他认为，我们能以另一种方式重新定义概率，即在一个 N 次抽签序列下，在后面的抽签中某个事件发生的相对频率。因此，即使在一次特定的实验中，某个不大可能的状态 Z 出现了，我们仍然可以认为，在后面的抽签中，更可能的状态会出现。玻耳兹曼提到在连续性重复试验中 Z 的变化就是演化。他接着说：

---

① 我们可以在后面的段落中找到对这种解读的支持。例如，在文献[Boltzmann，1897b]中写道："经验表明相互作用的物体构成的系统'最初'总是会被发现处于一个不大可能的态，然后将会很快达到最可能的态（平衡态）。"见[Boltzmann，1909，III，p. 607]。这里，玻耳兹曼也是认为从不太可能的态演化为较可能的态的趋势是一个经验事实，而不是任何定理的结果。

> 因此，最有可能的状态分布必须被定义为那个大多数(状态)都会演化成的状态[Boltzmann, 1909, II, p. 172]。

虽然他没有非常明确地讲出关键性的问题，但是他对在瓮中抽签的探讨无疑是对气体状态分布演变的类比。因此，在后来的例子中，玻耳兹曼也认为最有可能的状态分布就是大多数状态都将会演化成的那个，这个事实就不是那么令人难以置信的。反过来，这就意味着他认为有关演化的问题是某种需要证明的理论，也不依赖于诸如遍历假设或分子混沌假设那类假设的有效性，而是从一开始就已经解决了的事情。这就解释了为什么玻耳兹曼没有进一步费心解决这个问题。

即便如此，这种解读也有严重的缺点。先不必说在论证过程中对概念进行重新定义并不是一个明智的做法，仅仅将孤立气体的演化类比为一个伯努利过程的做法就是不可靠的。对前者来说，演化是由确定性的运动规律支配的，而在后者的情况下，我们会完全避免涉及任何基础的动力学理论，而只是遵循重复独立实验的概率性规律。请参见6.2节。

将上述关于玻耳兹曼针对可逆性反驳进行的回应的讨论总结如下：在上面所有对他1877年的两篇论文的解读中，似乎都指出了在令人信服地回应洛施密特的质疑这件事情上，玻耳兹曼已经做到的和他需要做的事情之间还存在着差距，即解决状态分布演化的问题并证明非均匀分布在某种统计学意义上趋向于演化为均匀分布，或证明 H 定理的任何重新表述，这些差距还是很明显的。

4.5 回归异议

庞加莱

1890年，在他关于天体力学中三体问题的著名论著中，庞加莱提出了现在所称之为的回归定理。大致上说，该定理表明，对于每个具有有界相空间的力学系统，无论它的初始状态是什么，在有限的时间内，几乎都会回到一个与该初始状态任意接近的状态，并且这种过程会重复无限多次。

用现代术语来说，可以归结为如下定理。

回归定理：考虑动力学系统①$\langle \Gamma, \mathcal{A}, \mu, T \rangle$，其中$\mu(\Gamma) < \infty$．令$A \in \mathcal{A}$为 Γ 的任意可测量子集，然后对一个给定的时间 τ，定义集合：

$$B = \{x:\ x \in A\ \&\ \forall t \geqslant \tau:\ T_t x \notin A\} \tag{66}$$

那么：

$$\mu(B) = 0 \tag{67}$$

特别地，对于一个哈密顿系统，如果我们选择 Γ 为能量曲面 Γ_E，取 A 为 Γ_E 上的一个“微小”区域，例如为正则坐标中一个直径为 ϵ 的开放球体，那么该定理指出，在这个区域中点的集合，其演化过程是这样的，即经过一段时间 τ 后，它们将不会回到区域 A 的测度为零。换句话说，几乎每一个从 A 内开始的轨迹，在我们选择的任意长的有限时间之后，将会返回到 A。

庞加莱在更早的时候就已经表达过他反对热力学关于不可逆现象所做的力学解释，例如[Poincaré，1889]。但只有应用了他的新定理，他才能够更准确地说明这一点。在其 1893 年的论文中，他指出，热的力学概念与我们关于不可逆过程的经验是矛盾的。根据当时英文版的运动学理论，庞加莱认为：

> 世界最初会趋向于一个能保持很长时间而不会出现明显变化的状态，并且这是符合经验的，但它不会永远这样保持下去，它并不违背上面引用的定理，它只是在该状态保持了一段非常长的时间，分子的数量越多，则这段时间越长。这个状态并不是宇宙最终的死寂状态，而是一种休眠状态，在亿万个世纪之后宇宙会从这种状态中苏醒过来。
>
> 根据这一理论，想要观察到热量从冷的物体传递到热的物体的话，就不需要麦克斯韦那样敏锐的视觉、智力和灵活性了，只要有一点点耐心就够了[Brush，2003，380]。

① 动力学系统的定义请参见 6.1 节。简单地讲，Γ 为相空间，$\mathcal{A}$是 Γ 的可测量子集的一个族，T 是时间演化 T_t：$\Gamma \times \mathbb{R} \longrightarrow \Gamma$ 的一个单参数连续群。

他的结论是，这些结果与经验相抵触，并导致一种“对力学的明确质疑”[Brush，2003，381]。

当然，庞加莱的“一点点耐心”，甚至“亿万个世纪”，仍是一种乐观的轻描淡写的说法。玻耳兹曼估计在1立方厘米的空气中出现一次重复所需的时间是$10^{10^{19}}$秒(见下文)，这完全超出了经验的范围。因此，庞加莱指出运动学理论的结果与经验是相矛盾的，这个论断还是过于草率了。

庞加莱的文章，似乎并没有受到同时代德语物理学界的关注——也许是因为他仅仅是批判英文版的理论而已。然而，当策梅罗于1896年提出了一个略微有所区别的观点时，玻耳兹曼注意到了这个问题。在策梅罗的观点中，一个最主要的区别就是他认为经验并不重要。

策梅罗的说法

策梅罗[1896a]指出，对于一个具有有界相空间的哈密顿力学系统来说，庞加莱定理意味着，撇开单个状态构成的集合不说，每个态都必然会几乎完全准确地回归到其初始状态，而且确实会经常重复这种回归。因此，对于相空间上任意连续函数F，$F(x_t)$不可能随时间而单调增加(除非初始状态是单个状态)；当初始状态重现时，只要存在一个有限的增加，就必须存在一个相应的减少，这种说法的最新证明请参见[Olsen(奥尔森)，1993]。因此，我们不可能得到“不可逆的”过程。沿着这条路径，策梅罗指出有多种选择可以避免这个问题。

1. 我们假设气体系统没有有界的相空间。这可以通过让粒子达到无限远距离或无限大速度来实现。然而，第一种选择会被气体存在于一个有限体积的容器中这个假设所否定。如果气体由相互之间在短距离上存在吸引力的点粒子(如粒子间存在$F\propto r^{-2}$的吸引力，就可以把粒子加速到任意快的速度)构成的话，那么第二个选择就可以实现。然而，在物理学基础上，我们还应该假设粒子之间在非常小的距离上也存在斥力。

2. 另一种可能性是假设粒子之间存在速度相关力的作用。然而，这一点要么违背能量守恒定律，要么违背作用力与反作用力定律，这两者都是基础的原子理论。

3. H 定理只对那些特殊的初始状态成立，这些态是回归定理的例外，我们假设只有在本质上这些态能实现。策梅罗说，这种选择是不可驳倒的。事实上，可逆性异议已经表明，并非所有的初始状态都符合第二定律。然而，这里我们不得不排除所有能想象到的初态中的绝大部分，因为回归定理的例外只是构成了一个总范围(即在现代语言中为测度)为零的集合。此外，状态变量的最小变化也会把一个单态转变为一个回归态，从而足以毁掉这个假设。因此，这个假设“在物理学中是相当独特的，我不相信任何人会长时间对它感到满意”。

然后就只留下了两个主要选择。

4. 必须修改卡诺—克劳修斯原理(Carnot-Clausius principle)。①

5. 运动学理论必须以一种本质上不同的方式重新建构，甚至应该被完全放弃。

在最后的这两个选择之间，策梅罗并没有表示出偏向于哪一个。他的结论是，他的目的是尽可能清楚地解释哪些能被严格证明，并希望这将有助于形成新的讨论并最终解决这个问题。

我想强调的是，在我看来，策梅罗的论证是完全正确的。如果他还有任何问题的话，只能说他并没有注意到，在玻耳兹曼当时最新发表的论文中，已经使 H 定理重新焕发了光彩。

玻耳兹曼的回应

在[Boltzmann，1896b]的回应中，他一开始就重复指出，气体定理是统计性的。特别是，他说他经常尽可能清楚地强调麦克斯韦分布定律不是一个从普通力学中得出的定理，也不能由力学假设来证明。② 同样地，从分子的观点看，第二定律仅仅是一种概率陈述。他用嘲讽的方式继续评论：

---

① 根据这个术语，策梅罗显然提及了第二定律，按推测应该也包括第零定律。

② 正如我们所见到的，玻耳兹曼在1877年就已经指出了这一点。然而，我们可能会注意到，仅仅在几年之后[Boltzmann，1892]，在给出麦克斯韦分布的又一推论之后(这次推广到适用于由硬质粒子构成的气体，它具有任意数量的自由度，为哈密顿量提供了二次方项)，他得出结论：“我因此相信，作为一个分析力学的定理，它的正确性(即麦克斯韦分布律)不容置疑”[Boltzmann，1909，III，p. 432]。但是，正如我们在其他地方看到的，玻耳兹曼认为，一些结果本质上取决于概率论这个陈述，与它们可以由力学定理推导出来这个陈述并不是相互排斥的。

策梅罗的论文表明，我的论著被人误解了；不过我仍然感到很高兴，因为这是第一次表明这些工作在德国引起了关注。①

玻耳兹曼认为庞加莱的回归定理“显然是正确的”，但策梅罗将该定理应用到气体理论中是不正确的。他的反驳论证与他在《自然》(*Nature*)(1895年)上发表的报告非常相似，策梅罗很显然没看过这篇论文。

更具体地说，该论证过程是这样的。考虑一种气体，它存在于由完全光滑和完全弹性的壁构成的容器中，最初处于任意的初始状态，随着时间进程而演化。每隔一段时间 t，我们可以计算一次 $H(t)$。进一步，我们考虑此函数的曲线，玻耳兹曼称之为 H 曲线。在他对策梅罗的第二次回复[Boltzmann，1897a]中给出了这个图。这个 H 曲线的一个粗略的现代版本，如图3。

除了运动是“规律性”的那些情况，例如所有分子都在一个共同的平面运动的情况之外，玻耳兹曼给出了曲线的下列属性：

(ⅰ)在大多数时间内，$H(t)$ 非常接近于其最小值，记作 $H_{\min}$。此外，当 $H(t)$ 的值非常接近于 $H_{\min}$ 时，分子速度的分布与麦克斯韦分布相比只有很小的偏离；

(ⅱ)曲线偶尔会，但很少达到一个峰值，它可能远大于 $H_{\min}$；

(ⅲ)峰值出现的概率随其高度而锐减。

现在假设，在一些初始时刻 $t=0$ 时，函数需要一个非常大的值 H_0，它远高于最小值。那么，玻耳兹曼认为，这个态很有可能随着时间进程而接近于麦克斯韦分布，即 $H(t)$ 将会朝着 $H_{\min}$ 的方向减少；随后在那个值上保持相当长的一段时间，以至于这个态在短到微乎其微的持续时间之内几乎不会偏离麦克斯韦分布。尽管如此，如果等待足够长时间的话，我们总会发现一个新的峰值，而事实上，初始状态最终也会重新出现。因此，在数学意义上，这些演化是周期性的，完全符合庞加莱的回归定理。

---

① 八年前，玻耳兹曼在柏林已经被认为是基尔霍夫的继任者，并且是普鲁士科学院的成员。因此他抱怨说自己的工作在德国没有受到关注，这一点很难让人相信。

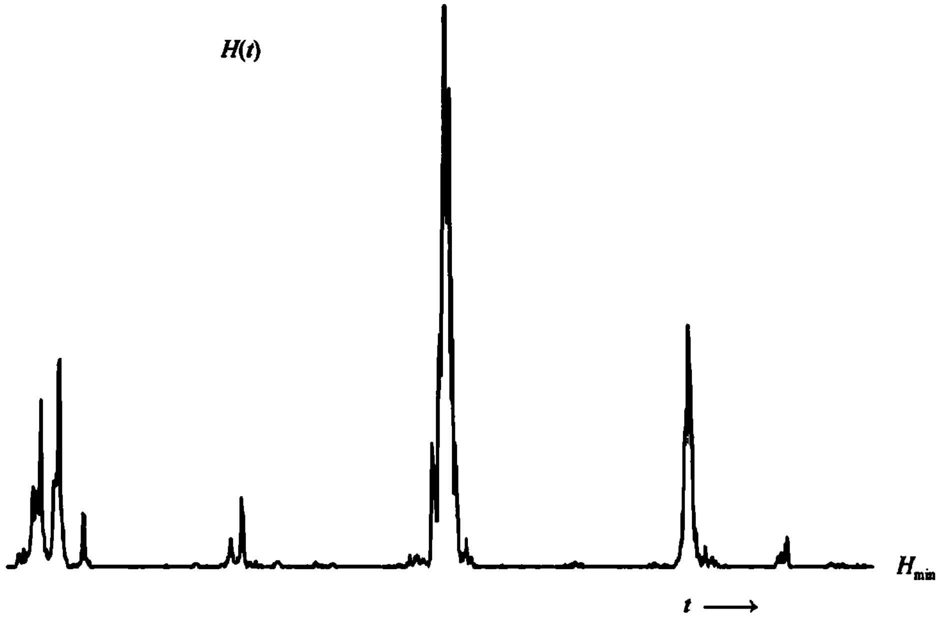


图3　H 曲线的形式化例子

那么，策梅罗的论证失败在什么地方呢？策梅罗曾声称只有非常特别的态才具有不断接近麦克斯韦分布的性质，与所有可能的状态相比，这些特殊的状态加起来也是无限少的。玻耳兹曼认为这是不正确的。对于绝大多数状态来说，H 曲线具有上面所说的定性特征。

玻耳兹曼也同意这个(他自称是策梅罗的)结论，即力学观点必须以某种方式进行修改或被放弃。他认为，只有当这一观点导致一些与经验相悖的结果时，这个结论才会成立。但是，玻耳兹曼指出，回归所需的时间是如此之长，以至于没有人能活到观测到它们。

为了证实有关回归时间长的这种说法，他还在一个在附录中估算了在室温和常压下 1 立方厘米气体的回归时间。假设在一个样品中有10^9 个分子，① 而在相应的 μ 空间中所选的单元格为六维立方体，在(物理)空间上其宽度为10^{-9}

① 事实上，现在估算时认为 1 立方厘米气体中包含 10^{19} 个分子，这就使得玻耳兹曼估算的回归时间会变得更长，即 $10^{10^{19}}$。

米，在速度空间中其速度为 1 米/秒，玻耳兹曼计算了不同宏观态的数量，即分子分布于这些单元格中的不同的方式的数量(大约)为10^{10^9}。然后，他假定，在先前的宏观态回归之前，该系统必须遍历所有其他的宏观态。即使分子之间经常发生碰撞，从而使系统的宏观态每秒变化10^{27}次，而它遍历如此大量宏观态也需要$10^{10^9-27}\approx10^{10^9}$秒。事实上，这个时间是如此之大，以至于无论我们用秒、年、千年或是其他什么单位来表示它，对它的数量级都不会有多少影响。

其结果就是，根据玻耳兹曼的观点，如果我们认为热是分子运动的一种形式，服从力学的一般规律，并假设系统的初始状态是不太可能的，那么对所有能观测到的现象来说，我们就会得到一个相当于第二定律的定理。他最后又嘲讽道：

> 因此，所有对关于自然的机械论观点的反对都是空洞的，是基于错误的认识之上的。不过，不管是谁，只要能克服困难，提出一个对气体理论的清醒认识，那么他就确实应该遵从策梅罗先生的建议，下决心彻底抛弃这个理论［Boltzmann，1909，III，p. 576］。

策梅罗的答复

策梅罗[1896b]指出，玻耳兹曼的回应证实了他的观点，承认庞加莱定理是正确的，且适用于一个由气体分子构成的封闭系统。因此，在这样的系统中，“在严格意义上来说，所有(原文如此)运动都是周期性的且不是不可逆的”。这样的话，气体运动学理论就不能像第二定律所要求的那样断言，熵是严格地单调增加的。他补充道：“我觉得这种一般性的澄清完全不是多余的。”见[Brush，2003，404]。

因此，策梅罗认为，他的主要观点得到了承认：热力学和运动学理论之间确实存在冲突，我们确实需要尝试抛弃其中之一。策梅罗承认，观察庞加莱回归可能确实超出了人类经验的范围。他(正确地)指出，玻耳兹曼对回归时间进行估算的前提是系统在回归到初始状态之前会访问相空间中的所有单元格。这种估算是不准确的，因为这个假设多少有些特别。在一般情况下，回归时间不会“是如此‘令人欣慰的’大”［Brush，2003，405］。但是，正如我强调过的，在策梅罗的反对中，与经验的关系根本不是问题。

策梅罗的回应主要是分析玻耳兹曼假设的合理性及其导致的结果，这个假设认为回归初态的可能性非常小，即 H_0 是非常大的。策梅罗认为，即使为了得到如玻耳兹曼所设想的那么一个与第二定律相近似的或符合经验的理论，即能够使得一个能持续很久但不是永久的平衡态出现的这么一种方法，但对于一个特定的初态来说，它也不足以表明这个结果。相反，我们不得不表明演化总是在同样的意义下进行的，至少在可观测的时间之内。

正如策梅罗理解的那样，玻耳兹曼不仅仅假设对于 H 来说，初始状态有一个非常高的值，而且还假设通常初态为一个最大值，或者刚刚经过了一个最大值。如果这个假设是理所当然的，那么很显然人们只能观测到 H 曲线减少的那一侧。然而，策梅罗反驳说，人们可以选择任何时间作为起始时间。为了得到令人满意的一般结果，因此这个额外的假设将不得不适用于任何时刻。但这样的话 H 曲线将不得不完全由最大值构成。但是，策梅罗认为，这是荒谬的，因为曲线不能是常数。策梅罗的结论是：玻耳兹曼需要对关于初态的假设做进一步的物理解释。

此外，策梅罗指出，概率论本身与时间的方向是无关的，以至于我们不能根据其推导出在某种特定意义上演化会更加特殊。他还指出，玻耳兹曼显然将状态的持续时间和其扩展（即它保持在某个区域内的相对时间和该区域在相空间中所占的相对体积）等同了起来。“我发现他实际上并没有证明这一点”[Brush，2003，406]。

玻耳兹曼的第二次回应

在他的第二次回应[1897a]中，玻耳兹曼驳斥了策梅罗的要求，即他认为玻耳兹曼需要针对相关系统初态的假设给出一个物理解释的要求，玻耳兹曼指出问题并不是一个随意选择的初态会怎样，而是处于当前状态的宇宙中的系统会怎样。

他认为，我们应该抛弃（诚然是不可证明的）这样的假设，即宇宙（或至少是我们周围的一个非常大的那部分宇宙）开始于一个非常不可能的状态，而且现在仍然处于一个不可能的状态。这样如果我们考虑一个突然从宇宙的其余部分中隔离出来的小系统（例如一个气体系统），那么就会有以下几种可能。(1)该系统可能已经处于平衡态，即 H 接近于其最小值。玻耳兹曼说，这是最

有可能的情况。但是在该系统不处于平衡态的那些极少数情况中。(2)最有可能的情况是 H 是 H 曲线上的一个最大值，因此，不论时间的方向如何，它都将会减少。(3)更罕见的情况是初值 H 处于 H 曲线单调减的那一侧。但是这种情况与第(4)种情况即 H 处于 H 曲线单调增的那一侧的概率相当。①

因此，玻耳兹曼对于 H 最初处于一个最大值这种说法的解释是，这对于一个不处于平衡态的系统来说是最有可能的一种情况，如果这个系统孤立于当前宇宙中的其他部分的话。

这或许是玻耳兹曼第一次为他的断言给出解释，作为对系统初态的一种假设，最终他将其与关于宇宙初始条件的假设联系了起来。今天，这通常被称为前假设(post-hypothesis)，参见[Albert(阿尔伯特)，2000；Winsberg，2004；Callender(卡伦德)，2004；Earman，2006]。

在论文的最后他评论道，如果对于所有的可观测现象来说，气体理论的力学概念与克劳修斯—卡诺概念(即热力学)一致的话，那么这种力学观点的优势就在于它最终可以预言新的现象，特别是预言悬浮在液体中的微粒的运动。八年后爱因斯坦对布朗运动的研究证实了这些预言。

不过，他并没有回应策梅罗的要求，即更明确地证明论断(1)—(3)，或证明相空间体积与时间平均的等同性。他直言不讳地指出，在利用相空间体积的方法测量概率方面，他领先了时代 30 年(这是真的)，他又补充说，一直以来他都是这样(这是假的)。即便如此，我们也不能将这种说法解释为玻耳兹曼拒绝接受关于概率的时间平均概念。接下来他又指出，最有可能的状态出现的频率也最大，除了那些数量非常少的初态以外。他没有对此做出证明。这再次为埃伦费斯特夫妇的推测提供了一个实例，他们推测玻耳兹曼可能在他的思想深处就相信遍历假设。

备注

玻耳兹曼对策梅罗的回应被认为是“非常清晰且完全无误的”[Lebowitz，

① 后来埃伦菲斯特于 1912 年增加了最后一种可能，即(5)：H 可能处于 H 曲线上一个局域的最小值，因此无论时间朝哪个方向，它都会增大。但是基于同样的原因，这种情况甚至比玻耳兹曼提到的那些情况更加罕见。

1999，S347]。然而，正如上面的内容及接下来的备注中所言的，我并不同意这个看法。关于策梅罗—玻耳兹曼争论的其他评论，请参见[Klein，1973；Curd(科德)，1982；Batterman，1990；Cercrgnani(瑟斯格纳尼)，1998；Brush，1999；Earman，2006]。

1. 迫在眉睫的问题。很明显至少在一个主要的争议点上，玻耳兹曼和策梅罗彼此之间谈论的并不是一回事。策梅罗认为在气体动力学理论中，并不存在一个持续的朝向最终稳定态的“趋向”，显然他是在 $t\to\infty$ 这个极限的意义下谈论这一点的。但玻耳兹曼的回复表明，他认为这个“趋向”并不确定，而只是可能的，它会持续很长时间，但这段时间是有限的。他的 H 曲线图清楚地表明，$\lim_{t\to\infty} H(t)$并不存在。

他对 H 定理的统计解读，正如在上述断言(1)—(3)中表明的那样，在[Boltzmann，1895]中就已经说得很清楚了，因此玻耳兹曼辩解道，他的工作被忽视了。但是，公平地说，我们必须注意的是，即使在此期间，玻耳兹曼也向他的读者传递了混淆的信息。事实上，在玻耳兹曼发表于1896年的关于气体理论的报告的第一卷中，他就强调了 H 定理的必要性和无例外的一般性，就像他在最初的论文[Boltzmann，1872]中所说的那样，只是增加一种认识，即该定理依赖于分子无序性假设(他后来称之为分子混沌假设)：①

> 被命名为 H 的这个量只能减少，最多它只能保持不变……这里我们所做的唯一假设是，速度分布最初“在分子层面上是无序的”并且这种无序会保持下去。因此在这种条件下，我们证明被称为 H 的量只能减小，而速度分布必定会更接近麦克斯韦分布[Boltzmann，1896，§5，p. 38]。

可能并不只是策梅罗认为玻耳兹曼曾真正打算做出最后的断言，策梅罗至

① 在他对策梅罗的回应中，玻耳兹曼声称，他在关于气体理论的讲座中对 H 定理进行讨论时，已经明确地强调了他假设分子的数量是无限的，所以，回归定理并不适用。但是，我在该文中找不到任何提及该假设的地方。与此相反，这个假设第一次出现在第6章，在那里他说“我们稍后将给出这个假设”，这表明前面的讨论并不依赖于它。

少同样合理地指出，玻耳兹曼的澄清“完全不是多余的”。

另一方面，玻耳兹曼将策梅罗的说法错误地表述为力学观点应该被抛弃这样的结论。正如我们已经看到的那样，策梅罗只是主张运动学理论的严格有效性和热力学的严格有效性之间存在一种两难推理。策梅罗的分析并不与实证问题相关。不过，当玻耳兹曼提及反对意见并没有指出与经验相冲突的地方时，他显然是正确的。因此，作为对庞加莱而非策梅罗的反击来说，他的回应显得更加成功。

2. *H* 定理的统计解读。这一点也与玻耳兹曼针对 *H* 曲线的特性提出的论断(1)—(3)有关。它们共同构成了也许对“*H* 定理统计解读”来说是最清楚的说明和最明确的形式。然而，他们与最初的定理的联系并不紧密。目前还不清楚的是，例如，这些论断是否仍然依赖于分子混沌假设以及稀薄气体假设，等等。因此，我们如何论证其有效性这个问题仍然存在。玻耳兹曼无法提供证明。在 1895 年《自然》杂志上发表的论文中，他认为他在以前的论文已经对此做出了证明，并直接地说道：“我不会在这里重复我已经给出过的证明。”见[Boltzmann，1909，Ⅲ，p. 541]。但可以肯定，玻耳兹曼从来没有给出过任何关于 *H* 随时间演化的概率性的证明，这一点对他的理论来说仍然是一个缺陷。当然，我们可以考虑如何弥合这一缺陷，例如，正如埃伦费斯特夫妇表明的那样，玻耳兹曼隐含地且默默地，或者以其他的方式依赖于遍历假设，但这里我不会对此做进一步讨论。迄今为止，兰福德提供了一种建构和证明统计性 *H* 定理的最成功的现代尝试，见下文中的第 6.4 节。

5. 吉布斯统计力学

在严格意义上，即作为一个前后一致且很有系统性的理论来说，统计力学的诞生是以吉布斯的书的出版(1902 年)为标志的。这本书的标题是《统计力学的基本原理；对热力学合理基础的特殊扩展》(*Elementary Principles in Statistical Mechanics; developed with especial reference to the rational foundation of thermodynamics*)。他的出发点是一个受哈密顿运动方程支配的一般力学系统，其(微观)状态由力学相空间 Γ 中的点表示。

吉布斯避免对这一系统的微观结构做出特定的假设。他提到了由双原子分子构成的气体，其比热存在异常值这个众所周知的问题，他的评论如下：

> 这类困难对作者解释自然奥秘的尝试造成了干扰，并迫使他在推导一些有关力学的统计分支的较为明显的命题时，不得不满足于更为有限的目标［Gibbs，1902，viii］。

这句话很清楚地表明吉布斯主要关心的是逻辑上的一致性，而非分子结构（事实上，该书只有最后一章专门研究由分子构成的系统）。这就导致了他的研究方法不同于麦克斯韦和玻耳兹曼。①

吉布斯的逻辑方案只包含两个部分，即力学和概率。这里概率不是可退化为独立系统的力学状态的，而是以我们现在熟悉的“系综”的方式来理解的：

> 我们可以设想存在着大量具有相同性质的系统，但在一个给定的瞬间，它们的结构和速度是不同的，它们之间不仅仅只有微小的不同，而且它们可能包括了每一种可能的结构和速度组合。在这里我们并不关注某个特定系统的一系列连续的结构，而是想确定在任意时刻，全部这些系统的各种可被设想的结构和速度是怎样分布的，如果在某一时刻的分布已经给定的话［Gibbs，1902，v］。

以及：

> 一般来说，如果我们想对一个物体做最准确且最简单的描述的话，那么我们所知道的一个方法就是，说它是从一个我们已经完全描述了大量物体（系综）中随机挑选出来的。

① 这也与爱因斯坦的研究方法不同，爱因斯坦在一系列论文［1902；1903；1904］中独立地得到了与吉布斯的结果非常接近的形式，不过他将其作为在实证上验证分子/原子假说的一种工具，参见［Gearhart，1990；Navarro，1998；Uffink，2006］。

请注意，吉布斯多少有点不承认关于概率的一切特殊解释。（当然，对概率做出权威解释是20世纪20年代以后的事了，我们不能假设吉布斯会预先想到了这些权威诠释的诞生。）现代的频率论者（对他们来说，一个事件的概率就是在很多类似的案例中，该事件发生的频率）在理解吉布斯提出的系综时不会有任何困难，并且他们大概会将这一概念与冯·米塞斯的 Kollektiv 概念等同起来。另一方面，像杰恩斯那样赞成对概率做主观解释（在这类解释中，一个事件的概率被理解为对有关该事件的知识或信念的一种陈述）的学者都强调，在吉布斯研究方法中，系综只不过是“想象出来”的，是表达我们知识的一种工具。

系综通常被表示为 Γ 上的一个概率密度函数 ρ，因此 $\int_A \rho(x)dx$ 就是系综中微观状态 $x=(\vec{q}_1,\ \vec{p}_1;\ \cdots;\ \vec{q}_N,\ \vec{p}_N)$ 位于区域 A 中的系统的相对数。在 $t=0$ 的时刻，系综密度 ρ_0 的演化是由哈密顿运动方程所决定的。对于（形式化的）时间演化算符 T_t 来说，我们有：

$$\rho_t(x)=\rho_0(T_{-t}x) \tag{68}$$

或者，写为另一种形式：

$$\frac{\partial\rho_t(x)}{\partial t}=\{H,\ \rho\} \tag{69}$$

这里$\{\cdot,\ \cdot\}$表示泊松括号：

$$\{H,\rho\} = \sum_{i=1}^{N} \frac{\partial H}{\partial \vec{q}_i}\frac{\partial \rho}{\partial \vec{p}_i} - \frac{\partial H}{\partial \vec{p}_i}\frac{\partial \rho}{\partial \vec{q}_i} \tag{70}$$

一种特殊情况是密度函数为常数的情况，即：

$$\forall t:\ \frac{\partial\rho_t(x)}{\partial t}=0 \tag{71}$$

这就是吉布斯所说的统计平衡条件。吉布斯指出，任何一个可被记作哈密顿函数的密度都是常数，他还区分了几种不同的特殊情况，其中最重要的就是：

$$\rho_E(x)=\frac{1}{\omega(E)}\delta(H(x)-E)\text{（微正则）} \tag{72}$$

$$\rho_\theta(x)=\frac{1}{Z(\theta)}\exp(-H(x)/\theta)\text{（正则）} \tag{73}$$

$$\rho_{\theta,\alpha}(x,\ N)=\frac{1}{N!\ Z(\theta,\ \alpha)}\exp(-H(x)/\theta+\alpha N)\text{（巨正则）} \tag{74}$$

其中 $\omega(E)$，$Z(\theta)$和 $Z(\theta,\ \alpha)$是归一化因子。接下来我将主要讨论正则和微正则系综。

5.1　统计平衡的热力学类比

正如书的副标题所说的，吉布斯的主要目的是为热力学提供一个"合理的基础"。他在处理这一问题时相当谨慎，他指出正则和微正则系综的关系与热力学结果之间存在某些相似之处。但吉布斯并没有提出将热力学归纳为统计力学。

由吉布斯指出的第一个相似性是针对存在热接触的 A 和 B 两个系统来说的。取相空间的积 $\Gamma_{AB}=\Gamma_A\times\Gamma_B$ 和哈密顿量 $H_{AB}=H_A+H_B+H_{\text{int}}$，他得到一个统计力学模型。如果 A 和 B 都由正则系综来描述，且 H_{int}与系统哈密顿量相比是"无限小"的话，那么当 $\theta_A=\theta_B$ 时复合系统就处于统计平衡态。他认为，这一点"完全类似于……热力学中相应的案例"。在热力学中，"判断两个物体温度相等的最简单方法就是要看它们存在热接触时，是否会保持在热平衡态上"。显然，吉布斯告诉我们统计平衡类似于热平衡，而 θ 类似于系统的温度。①

第二点相似之处在于重新得到热力学的"基本方程"式(23)：

$$dU = TdS + \sum_i F_i da_i \tag{75}$$

其中 a_i 为所谓的外部参数(例如体积)，F_i 为总合成力(例如负压强)。对于正则系综，吉布斯得到一个与形式上类似于上述基本方程的关系式：②

$$d\langle H\rangle = \theta d\sigma - \sum_i \langle A_i\rangle da_i \tag{76}$$

这里$\langle H\rangle$是在正则系综中哈密顿量的期望值，θ 是系综的模数，σ 是所谓的正则分布的吉布斯熵：

$$\sigma[\rho_\theta] = -\int\rho_\theta(x)\ln\rho_\theta(x)dx \tag{77}$$

---

① 关于参数 θ 的性质以及其性质与温度的类比的更为详细的讨论，可见于爱因斯坦(1902 年)的论文中。其中还讨论了热平衡的传递性，即热力学第零定律(参见第 2 段)。

② 更多细节参见[Uhlenbeck and Ford，1963；van Lith，2001b]。

a_i 是哈密顿量形式上的参数，$\langle A_i \rangle = \langle \frac{\partial H}{\partial a_i} \rangle$ 表示"总合成力"①。该方程表明，正则系综平均值类似于相应的热力学量，而 θ 和 σ 分别类似于温度和熵。②

注意式(76)中 θ 和 σ 特有的作用，它们不是相空间函数的期望值，而是系综密度 ρ_θ 的参数和函数。这里有一个重要的概念蕴含。前面的量可能被认为是在系综中的每个独立系统所拥有的某些属性的系综的平均数。但是，对于温度 θ 和熵 σ 来说，事实并非如此。在 θ 的情况下，我们可以减少这种对比——至少当 H 是一个运动学项和势能项的总和的情况下，其中运动学项是以二次方形式即动量出现的，即 $H = \sum_i \alpha_i p_i^2 + U(q_1, \cdots, q_n)$ ——根据著名的能量均分定理。该定理说明对于每个自由度来说，θ 等于预期动能的2倍：

$$\frac{\theta}{2} = \alpha_i \langle p_i^2 \rangle \theta \tag{78}$$

因此，在这种情况下，我们能找到正则期望值等于 θ 的相函数，并把这样的函数的值视为一个独立系统的温度。③ 但相空间上没有如下的函数 χ 存在：

$$\text{对所有的 } \theta\text{，有 } \sigma[\rho_\theta] = \langle \chi \rangle_\theta \tag{79}$$

因此，吉布斯熵不能被解释为系综中独立个体的某些属性的平均值。

接下来的问题是，对于微正则系综，是否也能得到一个类似于式(76)的微分方程。在这种情况下，我们很自然会考虑同上面一样的表达式 $\langle A_i \rangle$ 和 $\langle H \rangle$，在这里是否能作为微正则系综的期望值，以使得 $\langle H \rangle_{mc} = E$ 显然成立。那么问题就是如何找到 T 和 S 的微正则类比。[Gibbs，1902，pp. 124—128，169—171]中提出如下建议：

$$T \leftrightarrow \left(\frac{d \ln \Omega(E)}{dE} \right)^{-1} \tag{80}$$

① 如果我们想要证实 $-\langle \frac{\partial H}{\partial V} \rangle$ 真正等于压强——即每单位面积容器壁的平均力的话——还需要更加详细的论证，这类论证请参见[Martin-Löf(马丁—洛夫)，1979，pp. 21－25]。

② 这个推导中的一个关键假设是，微分式表示可逆过程中的无穷小元素，在此可逆过程中概率密度始终保持其正则形态。这个假设与动力学演化是相冲突的[van Lith，2001b，141]。

③ 要了解其他能等价于温度的更一般定义上的相函数，参见[Rugh(鲁)，1997；Jepps *et al.*，2000]。

$$S \leftrightarrow \ln\Omega(E) \tag{81}$$

其中：

$$\Omega(E) := \int_{H(x)\leqslant E} dp_1 \cdots dq_n \tag{82}$$

是积分的结构函数。

值得注意的是，在后面的文章中[Gibbs，1902，pp. 172—178]还提供了第二类温度和熵的等价值，即：

$$T \leftrightarrow \left(\frac{d\ln\omega(E)}{dE}\right)^{-1} \tag{83}$$

$$S \leftrightarrow \ln\omega(E) \tag{84}$$

其中，ω 为结构函数：

$$\omega(E) = \frac{d\Omega(E)}{dE} = \int_{H(x)=E} dx$$

对于这种选择，关系式(75)被再次重建。因此，对于统计力学量来说，似乎存在多种选择可作为热力学的对应量。虽然吉布斯讨论了对这两套方案的各种支持和反对意见——取决于我们是否把能量或温度看作独立变量，以及我们是否偏爱最有可能的值的预期值——但他没有明确地指出他更偏爱哪套方案，正如他所说的，系统(80)(81)更自然，而系统(83)(84)更简单。不过，吉布斯认为[Gibbs，1902，183]两套方案对于一个非常大的自由度来说是近似于一致的。然而，这意味着在他的方法中，仍然存在一种不完全决定性，只在热力学极限下人们才有希望避免它。

表达式(81)和(84)也被称为“体积熵”和“表面熵”。在现代教科书中，后一种选择要流行得多，或许是因为它与微正则系综中的吉布斯熵一致：$\sigma[\rho_E] = \ln\omega(E)$。然而，已经有人指出也存在一般性的理论原因去选择体积熵(81)，特别是，因为它与表面熵不同，是一个绝热不变量，参见［Hertz(赫兹)，1910；Rugh，2001；Campisi(坎皮西)，2005］。

当然，所有这一切都严格地受限于分析(统计)平衡态。在非平衡态的情况下，人们显然希望进一步与热力学进行类比，以获得回到平衡态的方式(“负第一定律”)以及知道平衡态开始和结束时的绝热过程中熵增加的情况，甚至完全在统计力学的基础上重新得到动力学方程。吉布斯对这些问题的探讨是第 5.3

和5.4节的主题。

但吉布斯还指出，这种将温度和熵与它们在统计力学中的对应量进行的比较“是不完整的，如果不考虑它们在单位和零点方面之间的区别，以及用作数值规范的数字的话”[Gibbs，1902，183]。这将会在下面的5.2节中继续讨论。

5.2 单位，零点和因子 N!

吉布斯在提出类似于熵的多种表达式，即式(77)、式(81)和式(84)的时候，都没有讨论过它们的“单位和零点”，即没有探讨过它们的物理度量，以及那些可能会被添加到这些表达式中的常量。这是很自然的，因为吉布斯在选择那些表达式时，是基于它们在重构基本方程时表现出的形式上的优点，而在基本方程中只出现了 TdS 的组合。后来他指出当我们将温度的对应量——即正则分布中的参数 θ，或微正则分布中的函数(80)或(83)——乘某个常数 K，并且将熵的对应量——(77)、(81)和(84)——乘 $1/K$，则基本方程保持不变，从而探讨了熵的物理度量的问题。将其应用到由 N 个分子构成的单原子理想气体的简单案例中，他的结论是，为了使温度的对应量等于理想气体温度，则$1/K$的值应为：

$$\frac{1}{K}=\frac{2}{3}\frac{c_V}{N} \tag{85}$$

其中，c_V 为定容比热。他指出：“物理学家已经确认这个值是不依赖于所探讨的单原子气体种类的一个常数”[Gibbs，1902，185]。事实上，在现代的符号中，$1/K=k$，即玻耳兹曼常数。

关于“零点”的问题，吉布斯指出，所有被提出可以作为熵的对应量的表达式都具有相空间体积的对数的维度，因此它们会受到我们对长度、质量以及时间的单位所做的选择的影响，并以附加一些常数的方式表现出来，参见[Gibbs，1902，19，183]。但是，即使某些选择对于这些单位来说是固定的，也还是需要为熵的对应量，即可能取决于基本方程中的不变量的那些任意表达式进一步添加常数。然而，当对熵的差进行比较时，它们的值会消失。而且，由于只有熵的差具有物理意义，因此确定这些常数的问题就显得无关紧要了。不过，吉布斯接着指出“当比较同一物质的不同量时，任意物体的熵都有任意

的积分常数这个原则，是随极限的不同而变化的"[Gibbs，1902，206]。在其论著的最后一章中，他构建了关于积分常数是如何取决于粒子数 N 的进一步条件。

吉布斯通过提出下面的问题来开始研究。考虑一个 N 粒子系统的相位（微观状态），即（$\vec{q}_1$，$\vec{p}_1$；…；$\vec{q}_N$，$\vec{p}_N$），这个系统中的粒子被认为是"不可区分的""完全相同的"或"完美地相似的"。[1]现在，如果我们对这样一个系统中的粒子执行一种排列，那么是否我们应将该结果视为一个特殊的相位？吉布斯首先指出，将这些相位看作是相同的，这种认识"看起来与统计学方法的精神一致"。他说，这可能会使得对于这样的粒子来说，除了数量上的区别以外，它们是全同的，即当比较重新排列后的系统和未排列的系统时，"我们在第一个系统中的任意特定粒子与第二个系统中的任意特定粒子之间找不到任何区别"[Gibbs，1902，187]。

不过，他很快就放弃了这种说法，指出对于"同时客观存在"的系统来说，所有这些都是正确的，但很难适用于"想象的创造物"。相反，吉布斯认为：

> "一个系统中几个粒子的完美相似性丝毫不会影响我们将某种情况下的一个特定粒子与在另一种情况下的一个特定粒子区分开来。问题是在对我们所涉及的问题进行讨论时，按照实际的便利性需求，我们应该选择哪种情况。"见[Gibbs，1902，188]。

因此，他研究了两种不同的选择，将其区分为序列变化后的相位是相同的一般相（generic phase）和不同的特殊相（specific phase）这两种观点。用现代的术语来说，特殊相空间是通过辨识一次排序中的所有相位点而得到的，一般相空间可被看作特殊相空间的商数空间，参见[Leinaas and Myrheim（雷纳斯和米海

---

① 据推测，这些术语（至少）意味着哈密顿量与这些粒子的排列无关，即它们是同等质量且是以完全同样的方式相互作用的。

姆)，1977]。一般情况下，在 N 个粒子构成的系统的相位上有 $N!$ 种不同的排列，①因此存在 $N!$ 个不同的特殊相位对应于一个一般相位。这使得一般相空间的测度相比于特殊相空间来说是按一个总体因数 $\frac{1}{N!}$ 缩减的。由于熵的对应量的规模都等于相空间测度的对数，因此相比于由特殊相位来计算的熵来说，这个因数可以被看作熵的一个新的积分常数，即 $-\ln N!$。吉布斯得出结论，当 N 是常数时，不论我们使用一般熵还是特殊熵都是一样的，因为这只会影响加到熵上面的任意积分常数[Gibbs，1902，206]。②

然而，吉布斯指出，如果我们比较具有不同粒子数的系统的熵，会发现情况不是这样的。例如，考虑两个相同的气体系统，它们具有相同的能量 U、体积 V 和粒子数 N，且处于连通的容器中，我们将每个气体系统的熵记为 $S(U, V, N)$。吉布斯认为总系统的熵等于两个气体系统的熵的总和：

$$S_{\text{tot}} = 2S(U, V, N) \tag{86}$$

现在假设阀被打开了，使得两个容器连通起来。吉布斯说："我们不认为这会使熵产生任何变化，虽然气体扩散到另一个容器中，而且如果气体是不同的话，这个过程会使熵增加。"见[Gibbs，1902，pp. 206—207]。因此，这种新情况下的熵为：

$$S'_{\text{tot}} = S_{\text{tot}} \tag{87}$$

但这是一个能量为 $2U$、体积为 $2V$ 和粒子数为 $2N$ 的气体构成的新系统。因此，我们得到：

$$S'_{\text{tot}} = S(2U, 2V, 2N) = 2S(U, V, N) \tag{88}$$

方程右边表示的是熵的延展性。该条件对于(至少对于大的 N)一般熵来说是满足的，但对于特殊熵来说并不满足。吉布斯总结说："因此显而易见的是，涉及一般相位时，系统是平衡态，涉及特殊相位时，系统不是平衡态，这是在对熵进行估算时，我们不得不处理的问题……除非考虑的是各种分子的数目都恒

① 这假定粒子的分子状态($\vec{p}_i$，$\vec{q}_i$)不一致。然而，其中一个或多个分子的状态确实一致的特殊相位空间中的点构成了一个勒贝格(Lebesgue)测度为零的集合。

② 同样的结论对玻耳兹曼熵式(61)也成立[Huggett，1999]。

定的物体的热力学。”见[Gibbs，1902，207]。

在吉布斯书中最后几页所探讨的问题，也许是此书最具争议性的内容，至少它已经形成了许多进一步的讨论。后来的许多学者都认为，在相空间的测量中加入一个因子 $1/N!$ 对于获得“正确”的结果来说是必需的，并最终归结于所考虑的完全相似的粒子在形而上的唯一性或“存在的个体性”方面的缺失。有些人甚至认为，量子力学解释需要对此做出解释。例如[Huang，1987，154]中写道：“为什么我们必须把……除以 $N!$ 才能得到对状态的正确计数，这在经典力学中是不可能被理解的。究其原因是这本质上就是属于量子力学范畴的……”然而，许多人否认这样的认识[Becker(贝克尔)，1967；van Kampen，1984；Ray(雷)，1984]。对该问题的各种意见和普遍困惑就讨论到这里，否则我就离题太远了。

需要注意的是吉布斯驳斥了那些从本体的形而上学出发对想象的创造物所做的论证。(我想这可能说明 N 粒子系统的相位是理论概念，而非物质客体。)此外，吉布斯并没有宣称，一般的观点是正确的，而特殊的观点不正确的，他喜欢从“实际的便利性”出发来解决这一问题。的确，其观点的几个方面是依赖于那些一般被认为是传统的假设。例如“可加性”要求式(86)可以被更详细地扩展为：

$$S_{tot}(U_1,\ V_1,\ N_1;\ U_2,\ V_2,\ N_2)+K_{tot}=S_1(U_1,\ V_1,\ N_1)+K_1+S_2(U_2,\ V_2,\ N_2)+K_2 \tag{89}$$

被应用于 S_1 和 S_2 为相同的函数且它们的自变量取相同值的这种特殊情况。这里要注意的一点是，只有我们同时假设 $K_{tot}=K_1+K_2$ 和 $K_1=K_2$ 时，这个关系式才能得到式(86)。此外，涉及式(87)时，他则谨慎地说“我们不认为这是做出了任何改变”——这表明他并不想说明这个公式表示的是一个事实还是约定俗成的选择。但是，总的来讲，我们似乎可以公正地说，吉布斯关于实际便利性的这个标准仅仅是重建了我们通常认为是热力学熵所具有的那些属性。

在评论的最后，我们应该注意到这里提到的吉布斯所做的对比，即热力学中相同气体的混合，即允许两个气体系统相互扩散，不会改变熵的大小，而对于不同气体的混合来说，此过程会导致熵增加，毫无疑问地涉及在他 1875 年的论文中针对该问题做出的一些讨论[Gibbs，1906，pp. 165—167]。他针对相

同流体混合时的熵与不同流体混合时的熵所做的对比，现在普遍被称为吉布斯悖论。更确切地说，这个“悖论”是这样的，只要物质不同，那么这些物质构成的不同流体相互混合时熵就会是一个常数(在上述情况下为 $kT\ \ln 2$)，当它们完全相同时，熵就突然变为零，从而它否定了我们可能希望的一个直观预期，即混合熵应随着被混合物质相似程度的逐渐增大而逐渐减小。现在请注意，在特殊的观点下，无论不同物质混合还是相同物质混合都会导致熵增加：在该观点下不存在吉布斯悖论，因为当物质变得越来越相似时，不会出现突然的变化。另一方面，吉布斯采用了一般的观点，即将相空间测度除以 $N!$，则重建了包括吉布斯悖论——不同气体混合和相同的气体混合之间存在的不连续性——在内的热力学熵的常规属性。

尽管如此，许多作者似乎仍然相信，除以 $N!$ 是解决吉布斯悖论的一个方案。但是，这显然是不对的，相反，是特殊的观点避免了这个悖论，而一般的观点则恢复了有关熵的统计力学对应量的吉布斯悖论。具有讽刺意味的是，在统计力学中，“吉布斯悖论”有时被用来表明或暗示在特殊的观点中，吉布斯悖论就是存在的，因此只有回到原来的悖论上，才能解决吉布斯悖论。

5.3 吉布斯与熵增加

正如我们看到的那样，吉布斯熵可以被定义为相空间 Γ 上一个关于任意概率密度函数 ρ 的泛函：①

$$\sigma[\rho] = -\int \rho(x)\ln\rho(x)\,dx \tag{90}$$

这个表达式具有很多著名的且有用的性质。例如，在能量超曲面 $H(x)=E$ 上的所有概率密度中，微正则密度式(72)具有最高的熵值。同样，我们可以看到，在具有一个给定期望值 $\langle H\rangle_\rho$ 的所有分布 ρ 中，正则分布式(73)具有最高的熵值，在所有 $\langle H\rangle$ 和 $\langle N\rangle$ 都给定的分布中，巨正则系综具有最高的熵值。

① 实际上吉布斯并没有使用熵这个术语来指称该表达式。他称函数 $\ln\rho$ 为“概率的指数”，称 $-\sigma$ 为“概率的平均指数”。正如我们已经看到的那样，在微正则系综中，吉布斯指出了多个而不是一个关于熵的对应量，并清楚地认识到：“当涉及有限数量自由度的系统时，可能还有……并且一定有其他的表达式，可以被视为……熵”[Gibbs，1902，169]。

但是，假设 ρ 不是固定不变的。它会随时间进程而演变，即 $\rho_t(x)=\rho(T_{-t}x)$。我们可能会问，在时间进程中熵是否会增加。然而，刘维尔定理直接表明：

$$\sigma[\rho_t]=\sigma[\rho_0] \tag{91}$$

尽管表面上与玻耳兹曼的 H 类似，但吉布斯熵是不随时间变化的。对第二定律的解释，或走向平衡态的过程，不可能是这么简单的。

不过，吉布斯警告我们要非常谨慎地进行研究。刘维尔定理可以被理解为这样的陈述，即 ρ_t 的运动可以被比作（多维）不可压缩流体的运动。因此，他将 ρ 的演化比拟为在不可压缩的介质中搅拌液体［Gibbs，1902，pp. 143—151］。在这种情况下，液体的平均密度，以及其密度的任意函数的平均值，都不会改变。不过，一个熟悉的经验事实是，搅拌往往会使得混合物更加均匀，或达到一个均匀密度的状态，对此，表达式 $-\int\rho\ln\rho dx$ 的值会增大，直到实现其最大值。

吉布斯通过对密度概念进行定义解决了这一矛盾。当然，当后者变为零时，这通常被看作是空间体积元中液体量的极限。如果我们采纳这个定义，即首先采取该限制，然后再考虑搅拌运动的话，我们得出的结论是 $-\int\rho\ln\rho dx$ 会保持不变。但是，如果我们考虑定义在一个确定的有限（非零的）体积元之上的密度，在搅拌无限长时间后，密度可能“显著地”变得一致，如果我们随后令体积元变得微乎其微地小，结果也不会受到影响。因此，正如吉布斯所看到的，这是有关我们在取这两个极限时如何安排顺序的问题。

吉布斯意识到，并非所有相空间中的运动都能产生朝向统计平衡发展的这种趋向，正如不是每一次对不可压缩流体进行的搅动都能导致混合明显变得更加均匀。尽管如此，他还是初步得出结论：“我们也许能从已经给出的这类原因中推断出，朝统计平衡的极限条件方向发展的趋向是一条普遍规则，如果初始条件不具有那样的特质的话。”参见［Gibbs，1902，148］。

5.4　粗粒化

目前关于吉布斯思想最常见的现代表述是通过将相空间分割成单元格的方

法来给出的，通常我们称之为“粗粒化”。我们用每个单元格上的分布函数的相平均$\rho(x)dx$取代原来的分布函数$\rho(x)$来进行研究，我们用到以下映射：

$$\mathcal{CG}:\rho(x)\mapsto\mathcal{CG}\rho(x)=\sum_i\hat{\rho}(i)\,\mathbb{1}_{\omega_i}(x)\tag{92}$$

其中：

$$\hat{\rho}(i):=\frac{\int_{\omega_i}\rho(x)dx}{\int_{\omega_i}dx}\tag{93}$$

且$\mathbb{1}$代表特征函数：

$$1_A(x)\begin{cases}1 & \text{如果}\quad x\in A\\ 0 & \text{其他}\end{cases}\tag{94}$$

通常的认识是，这种分割是与我们观测能力的有限精度相一致的，这种“粗粒化”的分布可能会被认为足以对我们所观测到的现象进行充分描述。显然，Γ上在单元格内相差不太大的任意函数的平均值是大致相同的，无论我们是使用精细化分布还是粗粒化分布对其进行描述。

对于任意ρ，我们可以将其粗粒化熵$\sum[\rho]$定义为式(92)和式(90)的结合：

$$\sum[\rho]:=\sigma[\mathcal{CG}\rho]\tag{95}$$

粗粒化熵在时间进程上不需要是守恒的。事实上，参见[Tolman，1938，172]我们很容易得到：

$$\sum[\rho]\geqslant\sigma[\rho]\tag{96}$$

因此，如果我们假设在某些初始时刻$\rho_0=\mathcal{CG}\rho_0$，例如，如果对于某些单元格$i$来说，$\rho_0\propto\frac{1}{V_i}1\omega_i$，那么对于所有$t$，有：

$$\sum[\rho_t]\geqslant\sigma[\rho_t]=\sigma[\rho_0]=\sum[\rho_0]\tag{97}$$

然而，这并不意味着随着$t\to\infty$，$\sum[\rho_t]$不会减小，或者它会达到一个极值。

如果一个类似于搅拌液体那样的特性在ρ_t的动力学演化过程中也存在的话，那么我们就有：

$$\lim_{t\to\infty}\sum[\rho_t] = \sum[\rho_{mc}] \tag{98}$$

因此，在粗粒化水平上朝向平衡态的趋向就可能会出现。当然，这种收敛要求对力学做出一个非平凡的假设。在现代的研究中，该假设就是，此系统具有混合特性(见6.1节)。

5.5　注解

吉布斯统计力学仅仅利用哈密顿力学和概率论就构建了一套具有明确表述的概念和方法的形式系统。它通过引入一个特定形式的哈密顿量，就能够并且经常用于计算气体和其他系统的平衡性质。吉布斯留下的主要问题是：第一，选择某个特定平衡系综的动机；第二，热力学的对应量不是被唯一定义的。然而，很多细致的研究已经表明，假设分子间相互作用是温和的，具有它们所需性质的唯一的热力学态函数就会在“热力学极限下”得到(参见第6.3节)。

1. 选择系综的动机。虽然除了简单性以外，吉布斯对于为什么选择这三个系综作为表示平衡态的候选对象，并没有提供更多的建议。从现在的观点出发我们会找到其他原因，首先，微正则系综被单独挑选出来用于描述具有特定能量E的隔热系统组成的系综。

针对这个目的出现了各种不同的讨论。由玻耳兹曼(1868年)提出，后来由爱因斯坦(1902年)更清楚地表明的一种观点是，微正则系综对一个由具有固定能量的孤立系统构成的系综来说，是唯一的稳定密度，如果我们承认遍历假设的话。不幸的是，在这些讨论中，遍历假设对于任何相空间为二维或更高维的系统来说是不成立的(参见6.1节)。

与此相关但更有希望的一种观点依赖于一个定理，即在那些关于P_{mc}来说是绝对连续的所有测量中，按照$P_{mc}(A)=\int_A\rho_{mc}(x)dx$这个表达式与微正则系综相关的测量$P_{mc}$是唯一的稳定测量，如果假设系统是度量可递的(同样参见6.1节)。

这种观点适用于更一般的系统，但它的结论较弱。特别是，我们现在将不得不论证，物理上令人感兴趣的系统确实是度量可递的，以及为什么在涉及微正则分布时不是绝对连续的测量在某种程度上应该被忽略。第一个问题仍然是

一个悬而未决的问题，即使对于硬球模型来说也是如此（我们将在6.1节进行讨论）。第二个问题可以通过各种不同的方式来回答。

例如，在[Penrose，1979，1941]中采取了一个原则，即每一个系综应该被描述为一个（分段）连续的密度函数，以排除“物理上不合理的情况”。此假设意味着根据拉东—尼克丁（Radon-Nikodym）定理，涉及微正则系综的系综测量是绝对连续的。[Kurth，1960，78]也提出了类似的假设。[Malament and Zabell（马拉蒙特和萨贝尔），1980]则提出另一种论证，文中假设与在物理上有意义的系综相关的测量 P 应该具有一种称为“平移连续性”的性质。粗略地说，这个概念意味着分配给任意测量集的概率在处于能量曲面上的集合的小位移范围内，应该是一个连续函数。马拉蒙特和萨贝尔指出，涉及 μ_{mc} 时，这个性质就等价于 P 的绝对连续性，从而单独挑选出了微正则测量，如果系统是度量可递的话，关于此问题更具体的讨论，请参见[van Lith，2001b]。

第三种方法，是由托尔曼和杰恩斯提出的，他们或多或少地假设微正则密度，该方法适当地描述了我们关于具有给定能量的微观系统的知识（无论系统是否是度量可递的）。

一旦微正则系综被认为是对一个具有给定能量的孤立系统的专门描述的话，我们花费更少的努力就能够找到其他系综的相应地位。正则分布被认为描述了一个小系统 S_1，当它与一个较大的系统 S_2 之间存在较低的能量接触，即发生“热浴”的时候，见[Gibbs，1902，pp. 180—183]。在这里，他假设总系统是孤立的，且可由一个微正则分布描述，其中总系统有一个哈密顿量 $H_{tot}=H_1+H_2+H_{int}$，且 $H_2 \gg H_1 \gg H_{int}$。爱因斯坦（1902年）和马丁—洛夫（1979年）对这种观点给出了更详细的论证。同样，巨正则系综可以从一个与大系统进行能量和粒子交换的小系统得到，参见[van Kampen，1984]。

2. 系综的“等价性”。在物理教科书中经常会出现的争议是，在不同系综之间做出选择（例如正则和微正则）时，它们之间的实际关联性被丢掉了，因为有断言认为它们都是等价的。关于这种争议的可能的最早版本，请参见[Lorentz，1916，32]，近期的陈述，请参见[Thompson，1972，72；Huang，1987，pp. 161—162]。这个断言意味着如果构成系综成分的数量增加，即 $N \to \infty$ 时，且总哈密顿正比于 N，则在此极限下由它们中的每个系统得出的热力学

关系将会是一致的。

然而，这些论证不应该被误认为解决了各种系综间的经验等价性问题，即使是在此极限下。例如，我们可以证明，微正则系综能够描述某些亚稳定的热力学态(例如负热容量)，而这些态是不能被正则系综描述的，参见[Touchette(杜彻特)，2003；Touchette *et al.*，2004]以及其中所引的参考文献。

3. 粗粒化熵。粗粒化的方法容易使人想起玻耳兹曼在他的论文[1877b]中建构的单元格(参见4.4节中的讨论)。它们的主要区别是，在这里我们假设对相空间Γ进行分割，而玻耳兹曼分割的是μ空间。尽管如此，关于特定分割的起源和地位的类似论题还是可以放在一起讨论的。如果我们假设这种分割可用于描述我们对系统的了解程度的话，也就是说，如果我们认为我们所知道的就是其状态是否位于一个特定单元格ω_i中的话，那么就可以认为这种分割的地位是主观的。如果我们认为，这种分割代表了人类观测可能性时其精确度的极限，而这种精确度或许还可以通过仪器来提高的话，也就是说，除了系统的状态处于某些单元格ω_i中之外，关于系统我们不可能实现更多观测的话，我们就可以认为这种分割的选择是客观的，这种客观性指的是某种可由认识论上的共同体观测到或者观测不到的客观事实。当然，我们仍然可以坚持认为粗粒化是一种人类中心主义的做法(参见7.5节的讨论)。但是，需要注意的是，吉布斯本人并不赞同相空间中的单元格应该具有某个特定的尺寸，而是认可取极限的方法，在这种极限下它们的尺寸按不同的阶趋向于零。

4. 统计平衡。最后是有关吉布斯的平衡态概念的评论。这根本不同于在1877年玻耳兹曼提出的关于平衡态的概念，他把平衡态看作一个宏观态，该宏观态对应于相空间中占据最大体积的区域(参见4.4节)。对于吉布斯，统计平衡态只适用于系综。而且，由于任何给定的系统可以被看作属于无数个不同的系综，因此说一个独立的系统是否处于统计平衡态是无意义的。与此相反，在玻耳兹曼的案例中，平衡态可以属于一个独立的系统(即如果该系统的微观态是集合$\Gamma_{eq}\subset\Gamma$中的一个元素的话)。但不能保证任何时候它都会处于该状态。

因此，我们可以说，对比于平衡态的正统热力学概念(它是稳定的，且是独立系统的一个属性)，玻耳兹曼[1877b]和吉布斯在面对哪些特征应该保留以及哪些特征应该放弃这个问题时做出了相反的选择。进一步的讨论详见

[Uffink，1996b；Callender，1999；Lavis(拉维斯)，2005]。

6. 统计力学的现代方法

本节将把前面章节中那种或多或少的历史性说明抛在脑后，转而有选择地对统计物理学中一些有影响力的现代方法做一个概述。特别地，我们专注于分析遍历理论(6.1—6.2 节)、有关热力学极限的理论(6.3 节)、兰福德关于玻耳兹曼方程所做的工作(6.4 节)，以及 BBGKY 方法(6.5 节)。

6.1 遍历理论

当埃伦费斯特夫妇在他们 1912 年完成的著名的百科全书式的论文中批判性地回顾玻耳兹曼和吉布斯关于统计物理的研究方法时，他们指出了与遍历假设相关的三个问题。

1. 玻耳兹曼在使用相空间区域的“概率”这个概念时是有歧义的(既可以理解为该区域的相对体积也可以理解为系统的轨迹出现在该区域的相对时间)。

2. 微正则概率分布或其他仅依赖于系统哈密顿量的概率分布的优越地位。

3. 玻耳兹曼的观点：系统的微观状态，如果最初出现在对应于一个非平衡的宏观态的相空间区域中，那么它应趋向于演化为这样一种结果，即它的轨迹在绝大多数时段里会落在对应于平衡的宏观态 Γ_{eq} 的相空间区域内。

对于这三个问题，应用遍历假设能得到一些多少比较确切的解决方案。因此，埃伦费斯特夫妇表明玻耳兹曼对上述问题的回答取决于遍历假设。正如我们已经看到的，只是针对于玻耳兹曼在他 1868 年的论文中处理问题(2)这个案例时，这一判断才是正确的。从埃伦费斯特夫妇的视角看，遍历假设未被证实的这种状态，当然会使得这些问题的这种悬而未决状态更加明显。

在后来的研究中，“遍历问题”变得更加直接地与上述的第一个问题相关，即相位平均和时间平均的等价性问题。此问题可具体阐述如下。考虑一个哈密顿系统以及定义在其相空间 Γ 之上的一些函数 f。对于一个初态为 x_0 的系统来说，f 的(无穷大)时间平均可以被定义为：

$$\overline{f(x_0)} = \lim_{T\to\infty}\frac{1}{T}\int_0^T f(T_t x_0)\,dt \tag{99}$$

这里 T_t 是演化算符。另一方面，对于密度为$\rho_t(x)$的系统组成的系综，f的系综平均为：

$$\langle f\rangle_t = \int f(x)\rho_t(x)\,dx \tag{100}$$

遍历问题是一个关于是否或是在哪种情况下，平均时间和系综平均是等价的，即$\overline{f(x_0)} \stackrel{?}{=} \langle f\rangle_t$ 的问题。请注意，这两种平均之间的差异是明显的。对比于$\langle f\rangle$，$\bar{f}$依赖于初态 x_0。事实上，初始相位点的每个选择都会导致相空间上的另一条轨迹，因此，在一般情况下，需要给出另一种时间平均。其次，$\langle f\rangle$一般依赖于时间，而$\bar{f}$是与时间无关的。因此，对于该问题，我们不应该期望得到一个普遍性的肯定回答。

然而，在稳定系综(统计平衡)的情况下，最后的相异性消失了。我们选择一个更为特殊的情况，即对于微正则系综 ρ_{mc}来说，遍历问题的最简单版本是这样的一个问题：

$$\overline{f(x_0)} \stackrel{?}{=} \langle f\rangle_{mc} \tag{101}$$

现在，很明显，如果玻耳兹曼的遍历假设是真的，即如果系统的轨迹经过能量超曲面 Γ_E 上的所有点，那么我们想要得到的等价性就成立。事实上，取 Γ_E 上的两个任意点 x 和 y。遍历假设意味着，存在一个时间 τ，使得 $y = T_\tau x$。因此：

$$\begin{aligned}
\overline{f(y)} &= \lim_{T\to\infty}\frac{1}{T}\int_0^T f(T_{t+\tau}x)\,dt \\
&= \lim_{T\to\infty}\frac{1}{T}\left(\int_0^\tau f(T_t x)\,dt + \int_0^T f(T_t x)\,dt\right) \\
&= \lim_{T\to\infty}\frac{1}{T}\int_0^\tau f(T_t x)\,dt = \overline{f(x)}
\end{aligned}$$

换句话说，在 Γ_E 上$\bar{f}$必然是恒定的，因此，也等于微正则期望值。

根据后来的文献，我们注意到另一个推论：遍历假设意味着 ρ_{mc}是 Γ_E 上唯一的稳定密度(参见 4.1 节)。

埃伦费斯特夫妇怀疑遍历假设的有效性，正如玻耳兹曼自己也是对此表示

怀疑的那样，因此他们提出了一个替代的假设，他们称之为准遍历假设。该假设陈述如下：轨迹在 Γ_E 中是密集的，即 x_t 将经过 Γ_E 上的每一个开子集，从而任意接近于经过 Γ_E 上的每一点。该系统可以被称为是准遍历的，如果该假说适用于系统的所有轨迹的话。正如我们已经看到的那样，这种提法似乎实际上更接近于玻耳兹曼的想法，至少是他在 1871 年的想法，这比他自己对该假设的公式化表述更好。

在埃伦费斯特夫妇提出该观点后不久，就有人从数学上证明当 Γ_E 是高于一维的流形时，遍历假设不成立[Rosenthal，1913；Plancherel，1913]。另一方面，准遍历假设却不能立即被驳倒。事实上，它可以很好地满足哈密顿系统的统计力学。不幸的是，我们一直不清楚它如何能解决遍历问题。乍看之下，人们可能希望准遍历系统的时间平均和微正则平均对连续函数是一致的，微正则密度 ρ_{mc}是唯一的连续稳定密度。但是，甚至这也是未知的。已知的是准遍历系统可能没有一个唯一的稳定测量[Nemytskii and Stepanov（尼梅特斯基和斯捷潘诺夫），1960，392]。这并不是说，准遍历是一个完全无创造性的概念。在拓扑遍历理论中，这个条件是以“极小性”的名义出现的，并且它暗示了一些令人关注的定理，见[Petersen（皮特森），1983，152ff]。

虽然罗森塔尔—普兰彻瑞的结果在 1913 年似乎为遍历理论敲响了丧钟，但是在 20 世纪 30 年代初一个意想不到的复兴又出现了。这些新的成果是由数学的急速发展以及考夫曼的结果而导致的，它们表明了哈密顿力学是如何被嵌入到希尔伯特（Hilbert）空间的形式化体系中的，在其中演变算符 T_t 被表示为一个幺正群。这为处理这个问题提供了一系列数学技巧（如频谱分析）。

冯·诺伊曼在 1932 年的一篇论文中得到了第一个结果，该论文有一个有前途的（但其实是误导性的）标题：“准遍历假设的证明”。伯克霍夫（G. D. Birkhoff）1931 年的一篇题为“遍历定理的证明”的论文使他的定理得以巩固，而且甚至比冯·诺伊曼的论文发表的时间更早。

由于他们的工作，以及后来有关遍历理论的所有工作，涉及了更精确的数学概念，因此我们可能需要先介绍一个该问题的更抽象的背景。一个抽象的动力学系统被定义为一个数组$\langle\Gamma,\ \mathcal{A},\ \mu,\ T\rangle$，其中 Γ 是一个任意集合，$\mathcal{A}$是 Γ 的子集的 σ 代数，被称为 Γ 中的“可测量”集，μ 是一个关于 Γ 的概率测度，T

表示 Γ 中的一对一变换 T_t 构成的一个单参数群(其中 $t\in\mathbb{R}$ 或 $t\in\mathbb{Z}$),T_t 表示演化算符。变换 T_t 被假设为测量不变的,即对所有 $A\in\mathcal{A}$,存在 $\mu(T_tA)=\mu(A)$。在统计力学的一个更具体的背景中,人们可以将 Γ 看作势能超曲面,将 $\mathcal{A}$ 看作其波莱尔(Borel)子集的集合,将 μ 看作微正则概率测度,将 T 看作由哈密顿方程决定的演化。

冯·诺伊曼—伯克霍夫遍历定理可以形式化地表述如下。

> 遍历定理:令 $\langle\Gamma, \mathcal{A}, \mu, T\rangle$ 为任意动力学系统,f 为 Γ 上的可积函数。那么:
>
> (i) $\overline{f(x)}=\lim_{T\to\infty}\frac{1}{T}\int_0^T f(T_tx)\,dt$ 对几乎所有 x 都存在,即令 $\overline{f(x)}$ 不存在的态 $x\in\Gamma$ 的集的 μ 测度为零;
>
> (ii) $\overline{f(x)}=\langle f\rangle_\mu$ 对几乎所有 x 都成立,当且仅当系统是度量可递的。

这里,度量可递的意思是,我们不可能将 Γ 分为两个实际可测的区域,使得从任何一个区域开始的轨迹永远不会经过另一个区域。更确切地说,度量可递性如下。

> 一个系统被称为是度量可递的①,当且仅当如下说法成立:对于所有将 Γ 划分为不相交的集 A_1 和 A_2,且使得 $T_tA_1=A_1$ 和 $T_tA_2=A_2$ 的任意分割法来说,$\mu(A_1)=0$ 或 $\mu(A_2)=0$ 成立。

不难理解为什么这个定理被认为是通过一个简单的重新诠释,就成功地解决了原来的遍历问题。首先,度量可递性在测量理论的意义上抓住了一种思想,即轨迹广泛地出现在整个能量超曲面上,只是对测度为零的集合例外。其

① 这个名字用得多少有点不当,因为这个条件与距离上的测量是完全无关的,而是纯粹关于测量理论的。度量可递的系统,也被称为是"测量不可分解的",或者特别在后来论文中是"遍历的"。我会坚持使用旧名称,以避免与遍历假说相混淆。

次，该定理保证了时间平均和微正则系综平均的等价性，虽然仅对可积函数成立，并且，同样对测度为零的集合例外。但是，这似乎对于大多数物理学家来说已经足够好了。

因此，遍历定理被视为一次巨大的成功。用莱辛巴赫的话来讲就是：

> 玻耳兹曼引入的……名为遍历假设的……假设，认为相空间点会经过能量超曲面上的每一个点。这种提法很容易被证明是站不住脚的。它被P. 埃伦菲斯特和T. 埃伦菲斯特的形式化结构所取代，即这条路径以任意小的距离 ϵ 接近于每一个点，ϵ 是由我们选择的，且它大于0。
>
> 仍然存在的问题是，是否遍历假设必须被视为一个独立的假设，或者是否它能从正则方程中推导出来，就像刘维尔定理那样。
>
> 这个问题……最终由约翰·冯·诺伊曼和乔治·伯克霍夫通过独创性的研究所解决，他们表明，第二种选择是正确的。……有了冯·诺伊曼和伯克霍夫定理，确定性物理学达到了其最高的完善程度：关于基本过程的严格决定论，导致了宏观现象的统计规律[Reichenbach, 1956, 78]。

不幸的是，该引文所陈述的一切内容几乎都是不正确的。

问题

1. 度量可递的系统是否存在？ 当然，最直接的问题就是度量可递的系统是否存在。对于数学意义上的“存在”来说，答案是肯定的。更有趣的问题是，是否可以证明任何足够实在地与统计力学相关的模型是度量可递的。

有一些力学系统已被明确证明是度量可递的。例如，在一个具有凸面散射性的容器中的一个硬球，或只限于在“体育场”（由两个半圆连接两条平行线段形成的平面）内运动的圆盘，或其三维等价体在由圆柱体与两个半球围成的空间中运动的硬球。但在统计力学范围内我们只对由许多粒子构成的系统感兴趣。

在 1963 年的论文中，西奈(Sinai)宣布，他找到一种证明由 N 个硬球构成的气体的度量可递性的办法。因此遍历定理，最终似乎是与物理上感兴趣的气体模型相关联。当然，硬球模型也是一种理想化模型，但物理学家普遍预期，过渡到更复杂的气体系统模型只会使度量可递性更可能和更合理，虽然这很难证明。

这个问题的证明是非常烦琐的，而且西奈的证明也是相当复杂的，实际上，也从来没有被完整地发表过。只是有不少部分结果发表过。事实上，完成这些证明所需的思想和技术的发展对如今被称为“遍历理论”的严谨数学理论的出现作出了巨大的贡献。而且，由于西奈的断言看起来是那么合理，因此很多书籍和文章将其作为严格证明了的事实，例如[Lebowitz and Penrose(勒波维茨和彭罗斯)，1973；Sklar，1993]。

但直到 20 世纪 80 年代，完整证明发表的一再延迟使得人们怀疑该陈述的有效性。最后，[Sinai and Chernov(西奈和切尔诺夫)，1987，185]中写道：“对于普遍性案例来说，[Sinai，1963]中给出的论断，必须被认为是不成熟的。”已经被严格证明的是，3 个硬球构成的系统是度量可递的。最近，萨斯(Szász)[1996]、斯莫伊和萨斯(Simányi and Szász)[1999]对这个问题做了进一步的研究。他们已经确定，对于一个 N 硬球模型，遍历要素，即度量可递性在其上成立的那个能量超曲面上的子集，具有正测度。无论如何，整个问题仍有待人们解决。

2. 无限时间。在时间平均式(99)的定义中，我们取极限 $T \to \infty$。这带来了一些问题：

(i)时间平均是有意义的，因为它是在实验上可以得到的。我们所希望的是，它能表示 f 的平衡值。但极限 $T \to \infty$ 没有告诉我们在有限的时间内会发生什么。最好的情况中，在经验上可得到的是在 T 很大但是有限时 $\frac{1}{T}\int_0^T f(T_t x_0)\,dt$ 的值。这个表达式可能仍然与极限值相差任意远。

(ii)当该系统不处于平衡态时，该极限甚至可能是存在的。一个时间平均值不需要是一个平衡值，因为一般来说有：

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T f(T_t x)\,dt \neq \lim_{t\to\infty} f(T_t x) \qquad (102)$$

例如，对于周期性运动来说，等式左边存在而右边不存在。

(iii) 在经验上来说，平衡态的到来一般相当迅速。但使 $\frac{1}{T}\int_0^T f(T_t x_0)\,dt$ 非常接近于 $\langle f\rangle_{mc}$ 所需的时间 T 可能是非常长的，即达到玻耳兹曼估计的庞加莱回归时间的量级！参见[Jaynes, 1967, 94]。

3. 测度为零的问题。遍历定理提供的结果是，对于度量可递的系统，$\overline{f(x)} = \langle f\rangle_{mc}$ 成立，除了由测度为零的微观态构成的集合以外。所以这里的建议是，这个例外的集合在一定意义上是可以忽略不计的。而且，从概率测度 μ_{mc} 的角度来判断，这显然是真的。但在任何其他意义上测度为零的集合都是不可忽略的。众所周知，如果我们将"测度大小"与其他判断集合的"大小"的自然准则，例如它们的基数，尺寸或贝尔(Baire)类等进行比较的话，则会发现这种比较并不合适。以许多其他的标准看的话，测度为零的集可能会出奇地大[Sklar, 1993, pp. 181—188]。

更重要的是，人们可能会选择另一项测量 μ'，这样，μ 测度为零的集合可能其 μ' 测度并不为零，反之亦然。当然这说明对测量的选择决定了哪些集合的测度为零。因此，如果我们决定根据微正则测度为零而无视或忽视某些集合，那么我们其实已经预设了微正则测量的优越地位。但是，这意味着遍历定理作为一种导致微正则测量具有优越地位的方法，会被贬低到只不过是一种自视过高的方法而已。

6.2 混合性质、K 系统和伯努利系统

遍历理论，这一由伯克霍夫和冯·诺伊曼的定理构成的数学领域，可以被认为是专门研究下面这个问题的，即在何种程度上一个确定的、时间反演不变的动力学系统可能会产生宏观尺度上的随机行为，通过假设其动力学的各种特殊属性的方式。

在该理论的现代形式中，它按照度量可递性的强弱程度将这些属性分成了不同层次。或许最重要的是混合属性，"K 系统"和伯努利系统特有的一个属

性。在这方面层次越高，“随机”行为被表现得越明显。微观层面上的演化在所有情况下都是遵循决定性演化规律的。在此主题的(扩展的)文献中，这种层次结构又被区分成不同的层次(如“弱混合”“弱伯努利”“非常弱的伯努利”等)，而且一些属性也不能适应于严格的线性层次结构(像“阿诺索夫”属性，它依赖于拓扑的概念，而不是依赖于动力学系统的纯粹测量理论表征)。这超出了本文的讨论范围。

混合

混合的概念通常被认为是由吉布斯提出的，在他将系综的演化对比于在不可压缩的液体中搅拌染料的情况时(参见5.4节)，即使最初的液体和染料粒子分别占据各自的空间，但搅拌最终将会使染料粒子均匀地分布在液体中。其正式定义为：

混合：一个动力学系统$\langle\Gamma, \mathcal{A}, \mu, T\rangle$被认为是混合的，当且仅当$\forall A, B\in\mathcal{A}$，

$$\lim_{t\to\infty}\mu(T_tA\cap B) = \mu(A)\mu(B) \tag{103}$$

在直观意义下，混合属性表达了这样的一种观点：动力学演化将用这样一种方式彻底搅拌相位点，该方式使得最初存在于A中的点最终均匀地分布在Γ上的所有可测量子集B中。人们可以很容易地证明，通过应用该条件到恒定集A并选择$B=A$，那么混合的确是一个比度量可递性更强的属性，相反的陈述则不成立(例如，一维谐振子是度量可递的，但不是混合的)。

同样，从概率测量或密度角度讲，也存在一个有趣的推论。考虑一个混合系统，以及一个随时间变化的概率密度ρ_t，相对于微正则测度μ，这样的ρ_t绝对是连续的。这意味着所有$A\in\mathcal{A}$的集，$\mu(A)=0$，且$\int_A\rho_t(x)dx = 0$，或等价地，ρ_t是一个关于μ可积的适当密度函数。在该情况下，随着$t\to\infty$，与ρ_t相关联的概率测度收敛为一个微正则测量。因此，一个由具有绝对连续密度的混合系统构成的系综，会渐进地趋向于统计平衡。请注意，同样的结果对于$T\to-\infty$也成立，因而就不会与时间反演不变性产生矛盾。那么它会与庞加莱回归定

理相冲突吗？也不会，因为回归定理关注的是微观相位点，而不是概率密度。即使几乎所有的轨迹最终都会返回它们原来的出发点附近，每个相位点的回归时间也不同，因此这类点构成的系综，其演化也会表现出确定地趋向于统计平衡态。

还应该注意的是，如果该结果被用于说明微正则测量的特殊地位（即作为对所有绝对连续概率分布演化趋向的唯一测量），该策略将会再次因为上述的那一点原因而遭到质疑，即绝对连续性的条件已经将微正则测量看作一个优先的选择。

尽管混合特性的描述是简洁的，但我们或多或少应该重复一下在遍历定理背景下提出的批判性言论。首先，该条件只考虑了极限 $t\to\infty$，这并未说明收敛速度如何。其次，强加的条件是平凡为真的，如果我们选择 A 或 B 测度为零的集合的话。因此，混合属性并没有说明在时间演化过程中这类集合会怎样。第三，我们仍然面临混合属性关于与统计力学相关的物理系统是否成立的问题。由于该属性是比度量可递性更强的论断，因此这个问题至少是同样难以回答的。

K 系统

接下来的一个重要概念是 K 系统的概念，“K”代表柯尔莫哥洛夫（Kolmogorov）。为简单起见，我们假定时间是离散的，因此对 $t\in\mathbb{Z}$ 来说，$T_t=T^t$。存在一个完全类似于连续时间的定义，我们称之为 K 流，参见［Emch，Definiton 10.3.2］。

> K 系统：① 一个动力学系统 $\langle\Gamma, \mathcal{A}, \mu, T\rangle$ 被称为一个 K 系统，当存在一个子代数 $\mathcal{A}_0\subset\mathcal{A}$ 时，使得：
>
> 1. $T^n\mathcal{A}_0\subset T^m\mathcal{A}_0$ 在 $m<n$ 时成立，$\subset$ 表示包含于；
>
> 2. 包含 $\cup_{n=1}^{\infty}T^{-n}\mathcal{A}_0$ 的最小的 σ 代数是 $\mathcal{A}$；
>
> 3. $\cap_{n=1}^{\infty}T^n\mathcal{A}_0=\mathcal{N}$ 成立，其中 $\mathcal{N}$ 是仅包含测度为零或 1 的集合构成的 σ 代数。

---

① 在该定义形式化的过程中存在一个相当大的变化［Cornfeld *et al.*（科恩菲尔德等），1982；Batterman，1991；Berkovitz *et al.*，2006］。这里的公式化也与前面的有一点不同。如果我们分别将 T 的指数位置上的 n 和 m 替换为 $-n$ 和 $-m$，则较常见的定义都是相同的。

乍看之下，这个定义是非常抽象的。通过下面的例子，我们可能会获得一些直观的认识。考虑一个将 Γ 划分为不相交单元格的有限分割 $\alpha=\{A_1, \cdots, A_m\}$ 以及与该分割相关的系统状态的所谓粗粒化历史。换句话说，我们记录的不是详细轨迹 x_t，而只是每个瞬间时刻状态所处的单元格 A_i 的标签 i，直到 $t=0$ 的时刻：

$$\cdots, i_{-k}, \cdots, i_{-3}, i_{-2}, i_{-1}, i_0 \qquad i_{-k}\in\{1, \cdots, m\}, k\in\mathbb{N} \tag{104}$$

这一个序列完全由 $t=0$ 时刻的微观态 x 确定：

$$i_{-k}(x) = \sum_{j=1}^{m} j\,\mathbf{1}_{A_j}(T^{-k}x) \tag{105}$$

这里1 代表特征函数(94)。是的，正如我们将看到的，对于一个 K 系统，这一序列在某些方面常常地表现出随机序列的特点。注意到：

$$i_{-k}(x) = j \Leftrightarrow T^{-k}x \in A_j \Leftrightarrow x \in T^k A_j \tag{106}$$

因此我们也可以通过将演化应用于分割好的单元格的方式来表示粗粒化的过程。如果 $T\alpha \coloneqq \{TA_1, \cdots, TA_m\}$，令 $\alpha \vee T\alpha \coloneqq \{A_i \cup TA_j : i, j=1, \cdots, m\}$ 表示对 α 和 $T\alpha$ 的一般精细化。当然，说 x 属于 $A_i \cup TA_j$，就等价于提供了序列(104)的后两项。以这样的方式继续下去，我们就能建立精细化：

$$\bigvee_{k=0}^{\infty} T^k\alpha = \alpha \vee T\alpha \vee T^2\alpha \cdots \vee T^k\alpha \vee \cdots \tag{107}$$

其中每个元素对应于一个特殊粗粒化历史（104），直到 $t=0$。集合(107)不再是有限的，但仍然是 Γ 的一个可数分区。

现在取 $\mathcal{A}_0$ 为由分区 $\vee_{k=0}^{\infty} T^k\alpha$ 产生的 σ 代数。显然，由该代数表示的事件仅仅是完全确定会发生的事件，如果那些粗粒化历史已知的话。换言之，对于所有 $A\in\mathcal{A}_0$，$\mu(A \mid C)$ 为 0 或 1，如果 C 是式(107)中的一项的话。很容易看到，$T^{-m}\mathcal{A}_0$ 仅仅是由 $T^{-m}\vee_{k=0}^{\infty} T^k\alpha = \vee_{k=-m}^{\infty} T^k\alpha$ 产生的 σ 代数，即由 $t=m$ 时刻之前的粗粒化历史描述的分区所形成的 σ 代数。由于该分区包含 $t=n$ 时刻的所有过程，如果 $n<m$ 的话，因此我们有：

$$T^{-m}\mathcal{A}_0 \subseteq T^{-n}\mathcal{A}_0, \text{对于全部的 } n<m \text{ 成立} \tag{108}$$

这等价于条件 1，但要将“$\subset$”替换为“$\subseteq$”。

此外，为了解释条件 2，我们要注意，包括 $\cup_{n=1}^{N} T^{-n}\mathcal{A}_0$ 的最小 σ 代数是由所有分区 $\vee_{k=-n}^{\infty} T^k\alpha$（其中 $n\leqslant N$）的并集产生的，以式(108)的观点看就是 $T^{-N}\mathcal{A}_0$。

因此，条件2只是说，如果我们将记录粗粒化历史的时间延长到 $t=N>0$，并令 $N\to\infty$，则该分区最终将变得足够精细，能够产生$\mathcal{A}$中的所有可测量集合。这是动力学的一个很强的属性。这意味着，整个粗粒化记录，从 $-\infty$ 到∞，提供将$\mathcal{A}$中全部可测量集合分离出来所需的一切信息(除非，有可能的它们的测度为零的集合不同)。

同样地，为了解释条件3，请注意式(108)意味着$\cap_{n=1}^{N}T^{n}\mathcal{A}_0=T^{N}\mathcal{A}_0$，这是由$\vee_{k=0}^{\infty}T^{k}\alpha$产生的，即，时间为 $-N$ 之前的粗粒化历史。因此，条件3表达的要求是，当我们令 $N\to\infty$ 时，由时间 $t=-N$ 之前的粗粒化历史决定的那类事件退化为表示概率为1或0的那些集合的"无意义的"代数。换句话说，对于$A\in\mathcal{A}$的每个事件，其中$0<\mu(A)<1$，在粗粒化历史的一些早期阶段，A 的发生是不确定的。

然而，K 系统真正显著的特点在于条件1所要求的严格内容：在任意时刻 n，由时间上在 n 之前的粗粒化历史决定的事件集，要严格小于时刻 $n+1$ 决定的事件集。由于后者是从前者的分区$\vee T^{-k}\alpha$加上分区 $T^{-(n+1)}\alpha$ 产生的，因此这意味着，在每个时刻 n，关于在下一时刻 $n+1$，分区上的哪个单元格会被占用这个问题的答案是不能从我们关于之前粗粒化历史的认识中得到的。对于由决定性运动定律产生的序列来说，这是一个相当引人注目的特性。当然，对于随机序列，例如扔硬币或轮盘赌来说，这是很常见的。

在尝试澄清的过程中，我们已经预设了一个特殊的有限分区 α。有人可能会问，是否每个柯尔莫哥洛夫系统都存在这样的分区。答案是肯定的，只要系统遵循一些温和的条件，即$\langle\Gamma,\ \mathcal{A},\ \mu\rangle$是一个勒贝格空间①。另一个问题是对于这个特殊分区，关于粗粒化历史的陈述是否是特别的。答案是否定的。我们可以证明，如果它们对于一些分区 α 成立，则它们也对 Γ 上的任意一个有限分区成立(粗略地说就是因为分区$\vee_{n}T^{n}\alpha$产生了所有事件的 σ 代数，因此由另一个有限分区构建的粗粒化历史，可以以 α 项的粗粒化历史重构)。

---

① 粗略地说，这个条件意味着，在测量理论的意义上，$\langle\Gamma,\ \mathcal{A},\ \mu\rangle$与时间间隔[0, 1]是同构的，都具有勒贝格测度。参见[Cornfeld *et al.*, 1982, 449]的精确定义。

伯努利系统(Bernoulli systems)

在遍历系统中上述属性最强的是伯努利系统。要介绍这类动力学系统的定义，首先了解通常所说的"伯努利"表是有意义的。考虑一个基本的概率性设备，可得到的结果为$\{A_1, \cdots, A_m\}$，概率为p_j。伯努利表被定义为一个概率空间，它是由在该设备上进行独立同分布重复试验产生的双无限序列而得到的。从形式上看，概率为$\{p_j\}$的集合(或"字母表")，$\alpha=\{1, \cdots, m\}$的伯努利表为概率空间$\langle \Gamma, \mathcal{A}, \mu\rangle$，其中$\Gamma$是由所有双无限序列构成的集合：

$$\eta=(\cdots, i_{-2}, i_{-1}, i_0, i_1, i_2, \cdots,) \quad i_k \in \{1, \cdots, m\}; \ k \in \mathbb{Z} \tag{109}$$

$\mathcal{A}$被定义为Γ上的最小σ代数，它包含集合：

$$A_k^j := \{\eta \in \Gamma: \ i_k = j\} \tag{110}$$

$\mathcal{A}$也被认为是柱代数。我们进一步要求伯努利表满足：

$$\mu(A_k^j) = p_j, \text{ 对全部的 } k \in \mathbb{Z} \text{ 成立} \tag{111}$$

我们可以把这种概率空间转换为一个动力学系统，通过引入变换T^m的离散群的方法，其中$m \in \mathbb{Z}$，T代表变换，即将Γ上的序列η中的每个元素向左移的变换：

$$\text{对全部的 } k \in \mathbb{Z}, \ T(i_k) = i_{k-1} \text{成立} \tag{112}$$

> 因此，我们定义伯努利系统：
>
> 具有离散时间演化T的动力学系统$\langle \Gamma, \mathcal{A}, \mu, T\rangle$是一个伯努利系统，当且仅当$\Gamma$上存在一个有限分区$\alpha=\{A_1, \cdots, A_m\}$，使得对于具有分布
>
> $$p_i = \mu(A_i) \quad i \in \{1, \cdots, m\} \tag{113}$$
>
> 的α来说，双无限的粗粒化历史是(同构于)伯努利表的。

因此，对于伯努利系统，α的粗粒化历史表现得就像相互独立的抽签过程那样是随机的。这些历史表现得完全无关，即使我们知道从负无穷到时刻n这段时间内的全部粗粒化历史，但是对于在时刻$n+1$的状态位置，我们所能做的最好预测，也不会比我们不知道任何事情时所做的预测更好。可以证明，每一个伯努利系统是一个K系统，但是，反过来则不成立。

讨论

遍历理论已经发展成为一个成熟的数学分支学科，出现了许多有意义的结果，也存在许多未解决的问题，要了解该领域目前的发展，请参阅[Cornfeld *et al.*，1982；Petersen，1983；Mañé(曼尼)，1987]。然而，这方面的研究与统计力学基础之间的关系往往会让人产生疑问。因此，[Earman and Rédei，1996]认为，这方面的研究与解释在平衡态统计力学中“为什么相平均有意义”是不相关的，[Albert，2000，70]甚至指出花费大量努力来严格证明遍历“不过就是在浪费时间——如果从统计力学基础的出发点来看的话”。进一步的讨论，请参见[Farquhar(法夸尔)，1964；Sklar，1973；Friedman(弗里德曼)，1976；Malament and Zabell，1980；Leeds(利兹)，1989；van Lith，2001；Frigg(弗丽嘉)，2004；Berkovitz *et al.*(伯克维兹等)，2006]。

这种认识通常是基于上面已经探讨过的问题，即对于在统计力学中我们感兴趣的物理模型来说，即使是遍历层次中最低的属性也难以确定，并且还有在经验上无限时间平均无法达到和测度为零的问题。此外，我们还经常用到柯尔莫哥洛夫—阿诺德—莫泽(KAM)结果，① 以缓和遍历是哈密顿系统的一般属性这个期望。这些困难是很严重的，但是，在我看来，它们不足以说明我们确实应该抛弃遍历理论。

相反，[Khinchin，1949；Malament and Zabell，1980；Pitowsky(皮托斯基)，2001]已经指出，该理论将会在更大程度上取得进展，如果我们按下面的条件发展理论的话：(1)系综平均和时间平均的等价性不需要对所有的可积函数成立，只需要对一个物理上有意义的子类成立就行；(2)在无限时间极限式(99)和式(103)下，确定收敛速度的重要条件；(3)放宽判定平衡态的条件。事实上涉及条件(1)的重要进展已经在“有关热力学极限的理论”中得到阐述，

① 很粗略地说，KAM定理表明，轨迹被限制在具有微小正测度的相空间中的不变集合内——因此不是度量可递的——那些哈密顿系统，当它们的哈密顿量增加一个足够小的扰动时，还会仍然保持该属性，更详细的介绍，请参阅[Tabor，1989]。这个结论破坏了曾经非常普遍的一种希望，即非度量可递系统在哈密顿系统中是罕见的、理想化的例外，并且通过确认其环境中的微扰，它们也总是可以变成一个度量可递系统。正如我们所知道的，玻耳兹曼(1868年)曾经对遍历假说就表达过这样的希望。

如第6.3节所述。很明显，进一步的变化在数学上可能是难以控制的，而且可能得到的那些结果将不会像现有的遍历理论的结果那样具有简单性和普遍性。但是，并没有理由表明在这些方向上的进展是不可能的。参见[Vranas(维拉纳斯)，1998；van Lith，2001b]。

测度为零的问题，我认为，“仅仅”通过我们上面讨论过的那种对测量的理论设定是无法解决的。关键的一点是，关于动力学系统的任何测量理论探讨，只是对测度为零的集合来说是不同的，但是在测量理论上来讲，它们是同构的，或者通常是等同的。测量理论没有办法区分测度为零的集合和空集。任何试图回答测度为零问题的方法，都需要使用其他的数学概念。我们只有赋予相空间更多的物理相关结构，例如拓扑结构或辛形式，才能够预期进一步的发展，参见[Butterfield(巴特菲尔德)，2006；Belot，2006]。

此外，即使遍历理论不能解释在平衡态统计力学中“为什么相位平均有效”的问题，也并不意味着它是在浪费时间。回想一下，相位平均和时间平均的等价性只是埃伦费斯特夫妇指出的几点原因中的一个，由于这些原因，埃伦费斯特夫妇认为玻耳兹曼的断言需要利用遍历假设来证实。另一点原因是他在1877年的断言：最初处于非平衡宏观态的系统应趋向于朝着平衡宏观态演化。

具有讽刺意味的是，针对遍历理论的一些批评放弃了尝试说明在何种意义上，以及在什么条件下微观态会呈现出在整个能量超曲面上散播的趋势，而依赖于一个字面上的且不可实现的希望，即不需要任何动力学假设在背后支持，这就“往往”会发生。显然，在此遍历层次被证明可能仍然是有意义的。

然而，不可否认的是，我们可以提供许多具体的例子，说明即使系统在上述意义上不是遍历的，平衡态统计力学仍然成立。在固体中，比如说在一个冰块中，分子被紧紧锁定在它们的晶格位置，相位点只能访问能量超曲面上的微小区域。类似地，对于在重力场中的一个∩形容器内的气体/液体混合物，分子可能会在极大比例的时间内被限制处于该容器某一半部分的底部，即使对应于另一半部分的区域在动力学上也是可访问的。不过我们仍然想用统计力学来解释它们的热性能。

总的来说，即使我们承认遍历理论不能提供有关所有想要案例的全部描述，但这并不意味着它是无意义的。我认为，在定性和概念层面，遍历理论最

重要的成就之一是，它明确表明，即使在很强的伯努利系统的意义上讲，微观层面上的严格决定论与宏观层面上的随机行为并不是相悖的。这意味着一个随机演化模型，例如玻耳兹曼在1877年用到的瓮中抽签模型，或埃伦费斯特夫妇给出的狗虱模型(参见7.2节)，并不一定与一个潜在的确定性动力学相冲突。

6.3 辛钦的方法和热力学极限

在遍历理论的“硬核”版本中，如同在前两节中描述的那样，我们关注于抽象的动力学系统，即我们使用的唯一假设与动力学演化过程所在的测量空间有关。这个动力学过程没必要产生于一个哈密顿量。另外，在这种方法中系统是否具有很大自由度也是无关紧要的。事实上，“贝克(Beker)变换”是遍历理论学家们所偏爱的一个例子，因为它提供了一个包含遍历层次全部重要属性的动力学系统，它使用单位面积作为相空间，因此它只有两个自由度。另一方面，拥有很大自由度的哈密顿系统，甚至不能满足遍历层次中最低的那一级，即度量可递性。

遍历理论在这方面经常会受到批评，因为我们可以证明，需要统计力学来解释的宏观系统的热现象，只有当系统的自由度很大时才会表现出来。正如辛钦所说的：

> 伯克霍夫和他的追随者们得到的结果……适合描述最普遍的动力学系统……。做这些研究的学者们并不关心我们这本书主要感兴趣的统计力学的基础问题。其目的是，得到那些最一般形式下的结果，尤其是这些结果既适用于只有少量自由度的系统，也适用于具有非常大自由度的系统。
>
> 从我们的角度来看，我们必须背离这一趋势。我们不必限制自己忽略统计力学所研究的系统(首先它们的基本属性就是它们具有非常大的自由度)的特殊属性。此外，我们并不要求用所有函数的时间平均来替代相位平均，事实上如果这种替代是令人满意的话，那么这些函数就具有许多特殊属性，使得这种替代在这些情况下是显而易

见的[Khinchin，1949，62]。

因此，部分原因是为了给遍历理论提供补充，部分原因是为了与遍历理论相竞争，辛钦研究了一类特殊的遍历问题，它们将大自由度作为一个重要因素，但只适用于一类特殊函数，即所谓的求和函数。

特别是，考虑一个由 N 个点粒子构成的哈密顿动力学系统 $\langle \Gamma, \mathcal{A}, T, \mu \rangle$。也就是说，我们假设：$x = (\vec{q}_1, \vec{p}_1; \cdots; \vec{q}_N, \vec{p}_N) \in \Gamma \subset \mathbb{R}^{6N}$。$\Gamma$ 上的一个函数 f 是一个求和函数，如果：

$$f(x) = \sum_{i=1}^{N} \phi_i(x_i) \tag{114}$$

其中 $x_i = (\vec{p}_i, \vec{q}_i)$ 是粒子 i 的分子态。① 在进一步假设哈密顿量本身就是一个求和函数后，辛钦证明了如下定理。

辛钦遍历定理：对于所有求和函数 f 来说，存在两个正常数 κ_1，κ_2，对所有 N：

$$\mu\left(\left\{x \in \Gamma : \left|\frac{\overline{f(x)} - \langle f \rangle_\mu}{\langle f \rangle_\mu}\right| \geqslant \kappa_1 N^{-1/4}\right\}\right) \leqslant \kappa_2 N^{-1/4} \tag{115}$$

简单来说就是，随着 N 越来越大，$\bar{f}$ 和 $\langle f \rangle$ 在很大程度上不同的这个集合的测度会趋向于零。

那么，这个定理为解决遍历问题提供了一种替代策略，它说明求和函数的时间平均和微正则相位平均是大致相等的，至少在能量超曲面的一个非常大的子集上是相等的，如果提供的粒子数量足够多的话。当然，这个“大致相等”是远远弱于冯·诺伊曼和伯克霍夫的遍历定理要求的“几乎无处不在的”严格相等。此外，它仅对于求和函数(114)成立。然而，这里并不需要度量可递性的假设，也不需要有关遍历层次的任何更严格的属性。

① 请注意，辛钦并不要求累加函数对于粒子的排列来说是对称的。

这种针对遍历问题的方法，其优点是显而易见的。第一，我们可以避免遍历对物理上感兴趣的系统来说很难获得的问题。第二，大自由度具有重要的作用，如上所述，在对热行为所做的任意一种解释中，它似乎都是必要的，或者至少是不受限制的一个因素。[①]第三，从物理动机出发对特殊函数做出了选择。

然而，这种方法也存在一些问题和缺点。首先，关于“无限次”的问题，辛钦的方法并不比原来的遍历方法更好或更坏。其次，由于粗略的等价性并不是“几乎在所有地方”都成立，而是将测度随 N 增大而变小的子集排除在外，遍历理论的测度为零的问题，就转变成了一个所谓的“测量 ϵ 问题”：如果我们希望能得到在实践中时间平均和相位平均(粗略地)相等这个结论，我们就应该承认，该等价性不成立的集合，即式(115)左侧的集合是可以忽略不计的。这个问题甚至比测度为零的问题更糟。例如，我们不能认为其密度函数支持这类集合的系综由于绝对连续性或平移连续性而被排除在外。此外，如果我们想将结果应用到确实不是度量可递的系统时，有可能存在一类运动方程的积分，使系统的轨迹在所有时刻都被锁定在 Γ 上的一个很小的子集内，在这种情况下，出于实践目的，这样的集合不能被忽略，参见[Farquhar，1964]。

辛钦认为，我们在统计力学中遇到的大部分从物理学上看是重要的相位函数，都是求和函数，参见[Khinchin，1949，63，97]。然而，从物理学角度来看，这种观点显然过于狭隘。这意味着所有取决于粒子间的关联或相互作用的量都被排除掉了。

最后，存在一个“方法论悖论”[Khinchin，1949，pp. 41—43]。这指的是辛钦不得不承认哈密顿量本身也是一个求和函数这个事实。我想强调的是，做出这种假设不只是为了让哈密顿量成为一个能够使该定理适用的函数，而且该假设对于推导该定理来说也是至关重要的。正如辛钦清楚地指出的，这是矛盾的，因为对于最终一定会出现的平衡态来说，它的一个本质特征就是粒子之间

① 当然这一点还有待讨论。有些作者认为，小系统也可以表现出应该用统计力学来解释的热行为。然而，热力学量(如温度等)的定义，对于这样的小系统来说是有争议的[Hill(希尔)，1987；Feshbach(费什巴赫)，1987；Rugh，2001；Gross and Votyakov(格罗斯和沃提亚科夫)，2000]。

存在相互作用(例如碰撞)，而如果哈密顿量是一个求和函数的话，这种可能性就会被否定。

因此辛钦的观点是，这个假设不能从字面上理解。相反，我们应该假设，在哈密顿量中确实存在相互作用项，但是只有当粒子之间的距离非常短时，它们才会表现出来，这使得它们可以被忽略不计，除非它们存在于相空间很小的一部分上时。尽管如此，他的工作仍然表现出一个奇怪的特点，即他的定理是为了适用于与推出该定理的假设不一致的那些情况中，参见文献[Morrison(莫里森)，2000，pp. 46—47]。在接下来的章节中，我们将会看到，后续的工作解决了这个矛盾，并弥补了辛钦方法的诸多缺点。

关于热力学极限的理论

辛钦的研究方法在范德林德和马祖尔(1963)那里得到了进一步的发展，而到了20世纪60年代末到70年代初，范·霍夫(van Hove)、杨、李、费舍尔、格里菲斯(Griffiths)、迷尼洛斯(Minlos)、吕埃勒(Ruelle)、兰福德等人则分别独立地在此方面开展了工作，从而将该方法发展为一个所谓的可得到“严格结果”的方法或“有关热力学极限的理论”。这方面最有用的文献为[Ruelle，1969；Lanford，1973；Martin-Löf，1979]。下面的论述主要是基于文献[Lanford，1973]，该文献是最易懂的，同时也是与我们的目的最具相关性的，因为它明确地解决了遍历问题，请参见[van Lith，2001b]。

正如在辛钦的工作中，该方法的目的是为处于平衡态的宏观物体的热行为提供一种解释方案，主要是依赖于以下几个关键点来实现的：

- 我们在相空间上采用了微正则测量；
- 可观测量为一类特殊的相位函数 F(见下文)；
- 粒子 N 的数量是非常大的。

它表明，在一定条件下，即下面指定的“热力学极限”下，对于 F/N 的微正则概率分布会集中于一些固定值附近的狭小区域内。这个结果类似于辛钦的遍历定理。然而，正如我们将要看到的，目前的结果越有力，所需的假设则越弱。

一开始，我们首先假设 N 粒子的哈密顿量，其形式为：

$$H(x) = \sum_{i}^{N} \frac{\vec{p}_i^{\,2}}{2m} + U(\vec{q}_1, \cdots, \vec{q}_N) \tag{116}$$

被定义在相空间Γ上。由于技术上的原因，在位形空间中进行研究，并忽略动量要更方便也更简单一些。考虑函数 $F(\vec{q}_1, \cdots, \vec{q}_n)$ 的一个序列，其中 $n=1$，2，…，其参数的数量不定，换种说法就是单值函数 F 定义在如下空间上。

$$\cup_{n=1}^{\infty}(\mathbb{R}^3)^n \tag{117}$$

这样的一个函数被称为是“可观测的”，如果它具有下列性质的话。

(a)连续性：对于每个 n，$F(\vec{q}_1, \cdots, \vec{q}_n)$ 是 $\mathbb{R}^{3n}$ 上的一个连续函数。

(b)对称性：对于每个 n，$F(\vec{q}_1, \cdots, \vec{q}_n)$ 对其参数的不同排列来说是不变的。

(c)平移不变性：对于每个 n，以及每个 $\vec{a}\in\mathbb{R}^3$，$F(\vec{q}_1+\vec{a}, \cdots, \vec{q}_n+\vec{a}) = F(\vec{q}_1, \cdots, \vec{q}_n)$。

(d)归一化：$F(\vec{q}_1)=0$。

(e)有限程：存在一个实数 $R\in\mathbb{R}$，使得对于每个 n，下面的陈述成立，即假设我们将 n 个粒子分成两组，分别表示为 $i=1, \cdots, m$，和 $i'=1, \cdots, m'$，这里 $m+m'=n$。如果对于所有的 i，i'，$|\vec{q}_i-\vec{q}_{i'}|>R$ 成立，那么 $F(\vec{q}_1, \cdots, \vec{q}_m; \vec{q}_1, \cdots, \vec{q}_{m'}) = F(\vec{q}_1, \cdots, \vec{q}_m) + F(\vec{q}_1, \cdots, \vec{q}_{m'})$。

在大多数情况下，这些条件是自然的且是不证自明的。需要注意的是对称条件(b)是非常强的。这可以对应于玻耳兹曼[1877b]的组合方法，在这种方法下人们认为，由于在粒子排列变换下宏观态保持不变，因此宏观态占据了相空间中的绝大部分(见第4.4节)。我们还应进一步注意，条件(e)表明如果所有粒子彼此之间相距足够远，那么 F 就可以被简化为一个求和函数。这也意味着，由兰福德描述的可观测量只能被认为对应于外延量。(回想一下，一个热力学量如果它随系统的大小成比例地变化的话，能被称为外延量；如果它与系统的大小无关的话它被称为强度量。)因此在这种方法下，强度量(如温度和压力)不能被表述为可观测量，而是被认为等同于其他量的适当导出量，当我们已达到热力学极限之后。

另外，假设式(116)中的势能函数 U 也满足上述条件。此外，假定势能是

稳定的[①]，即：

(f)稳定性：存在一个数 $B \in \mathbb{R}$，使得对于所有 n 以及所有 $\vec{q}_1$，…，$\vec{q}_n$，

$$U(\vec{q}_1, \cdots, \vec{q}_n) \geqslant -nB \tag{118}$$

这个条件——这是违反比如牛顿引力定律的——避免了当 n 变大时，每粒子的势能会趋于负无穷大，即它避免了系统的塌缩。

对于这些结果，施加一个更强的条件是有用的。

(f′)超稳定性：势能 U 被认为是超稳定的，如果对于$\mathbb{R}^3$ 上紧致支集中每个连续函数 Φ 来说，

$$U(\vec{q}_1,\cdots,\vec{q}_N) + \lambda \sum_{i \neq j} \Phi(\vec{q}_i - \vec{q}_j) \tag{119}$$

对于一个足够小的 $\lambda > 0$ 是稳定的。换句话说，一个稳定势是超稳的，如果它受到有限范围内一个持续的两体相互作用势的微扰时，仍能保持稳定的话。

在辛钦的方法中，假设(f)或(f′)不仅仅是因为我们想把势能算作一类可观测量才是必要的，而且对于证明热力学极限的存在性，它们也是非常重要的。当然，假设相互作用势是连续的且是在有限范围内的，仍然是限制太过了，以至于不能模拟真实的分子间作用力。正如兰福德指出的，我们可以将条件(e)弱化为一个有关“弱缓和”势的条件，[②]该势随着距离增大而快速减小[Fisher, 1964, 386; Ruelle, 1969, 32]，虽然这使得求证的技术过程更复杂。此外，很明显的是，就算仅仅是为了避免另一场灾难，即随着 n 的增大，每个粒子的势能会趋向于 $+\infty$，从而使系统可能会趋向于爆炸(这可能会发生，例如，一个只存在纯粹相斥的库仑力的系统)，给出弱化长程相互作用的这类条件也是

---

① 严格地讲，结构的微正则测量的热力学极限的存在性，是不需要条件(f)成立的。然而，当这些结果扩展到相空间时(或使用正则测量时)，条件(f)成立是必要的。还要注意的是术语“稳定”，在这里指的是哈密顿量的扩展下界。这应区分于热力学中稳定性，在其中它所表达的是熵函数的凹性。

② 为简单起见，如果势 U 是粒子对相互作用势 $U = \sum_{i \neq j} \phi(\vec{q}_i - \vec{q}_j)$ 的总和，那么它就是弱缓和的，当且仅当存在实常数 $R, D, \epsilon > 0$ 时，使得当 $\|\vec{r}\| \geqslant R$ 时，存在 $\phi(\vec{r}) \leqslant D\|\vec{r}\|^{3+\epsilon}$。

必要的。

现在，如果这些假设是适当的，那么我们介绍该观点如下。选择满足上述条件的给定的势 U 和一个可观测量 F，再选取两个数 u 和 ρ，分别表示每个粒子的(势)能和粒子密度(随着 N 变大会处在一个极限下)，一个有界的开放区域 $\Lambda\subset\mathbb{R}^3$，以及一个大整数 N，使得 $\frac{N}{V(\Lambda)}\approx\rho$，在这里，$V(\Lambda)$ 表示 Λ 的体积。此外，应选择一个很小的数 $\delta u>0$，构造位形空间上的(超密)能量超曲面，即壳层：

$$\Omega_{\Lambda,N,u,\delta u}=\left\{(\vec{q}_1,\ \cdots,\ \vec{q}_N)\in\Lambda^N:\ \frac{U(\vec{q}_1,\ \cdots,\ \vec{q}_N)}{N}\in(u-\delta u,\ u+\delta u)\right\} \tag{120}$$

令 μ 表示 Λ^N 上的勒贝格测度，那么对上面的集合的(归一化)限制可能会被称为“超密位形微正则测度”。需要注意的是：

$$\omega^{\mathrm{cf}}(E):=\int_{\Lambda^N}d\vec{q}_1\cdots\vec{q}_N\delta(U(\vec{q}_1\cdots\vec{q}_N)-E) \tag{121}$$

可被视为结构函数(41)的位形等价式。从而

$$\mu(\Omega_{\Lambda,N,u,\delta u})=\int_{\Lambda^N}d\vec{q}_1\cdots\vec{q}_N\mathbb{1}_{(u-\delta u,u+\delta u)}(U/N)=\int_{N(u-\delta u)}^{N(u+\delta u)}dE\omega^{\mathrm{cf}}(E) \tag{122}$$

因此 $\frac{1}{2N\delta u}\mu(\Omega_{\Lambda,N,u,\delta u})$ 提供了一个位形结构函数的超密的或平滑的版本。这里用超密的壳层替代稀疏的超曲面的原因当然是为了规避后者中可能会出现的奇点。在任何情况下，我们都可以预期的是，当 δu 很小时，这个表达式将表示微正则熵式(84)的位形部分。通过增加另一个因数 $1/N!$，这个熵有机会变成是外延的。① 可请参阅5.2节。

涉及位形空间上的这个超密微正则测量时，我们对 F/N 的概率分布感兴趣。为了这个目的，我们选择一个任意开放区间 J，并定义：

---

① 例如，如果这个系统是一个理想气体，即如果 $U(\vec{q}_1,\ \cdots,\ \vec{q}_N)\equiv0$，我们将会有 $\omega^{\mathrm{cf}}(E)=V^N=\left(\frac{N}{\rho}\right)^N$，使得 $\ln\frac{1}{N!}\omega^{\mathrm{cf}}(E)$ 随着 N 的变化而按比例变化，但是 $\ln\omega^{\mathrm{cf}}(E)$ 却不会。

$$\mathcal{V}(\Lambda,N,u,\delta u,F,J) := \frac{1}{N!}\mu\left(\left\{(\vec{q}_1,\cdots,\vec{q}_N) \in \Omega_{u,\delta u}:\frac{F(\vec{q}_1,\cdots,\vec{q}_N)}{N} \in J\right\}\right) \tag{123}$$

因此：

$$\frac{1}{\mu(\Omega_{\Lambda,N,u,\delta u})}\mathcal{V}(\Lambda,N,u,\delta u,F,J) = \frac{\mathcal{V}(\Lambda,N,u,\delta u,F,J)}{\mathcal{V}(\Lambda,N,u,\delta u,F,\mathbb{R})} \tag{124}$$

给出 F/N 位于区间 J 的概率，当涉及上述微正则测量时。

我们希望研究热力学极限下该概率的变化方式，即随着 N 变大，以及 $V(\Lambda)$ 正比于 N 增大，会有 $N/V(\Lambda)=\rho$。这一变化方式将取决于对极限过程的精确描述，特别是取决于 Λ 的形状。兰福德选择取范·霍夫意义上的极限，对于所有 $r>0$，在距离 Λ 的边界为 r 范围内的所有点构成的集合的体积除以 Λ 的体积的商，如果随着 N 趋于无穷大而趋向于零的话，那么$\mathbb{R}^3$ 上的有界开区域 Λ 构成的序列在范·霍夫意义上就被称为是变为无穷大的。换句话说，与内部的体积相比，靠近表面的点的体积就变得可以忽略不计了。这避免了表面效应会在极限情况下起作用的可能——也不用担心粒子与容器壁之间的相互作用会被考虑在内。

现在，最主要的结果是：

> 热力学极限的存在性。随着 $N\to\infty$，以及 Λ 在范·霍夫意义下变得无限大，以这种方式，$N/V(\Lambda)=\rho$，则下面两种情况中有一种会实现：
>
> (α) $\mathcal{V}(\Lambda,N,u,\delta u,F,J)$ 随 N 的增大而趋向于零的速度快于指数级；
>
> (β) $\mathcal{V}(\Lambda,N,u,\delta u,F,J) \approx e^{Ns(\rho,u,\delta u,F,J)}$，其中 $s(\rho,F,J)$ 不依赖于 Λ 或 N，除非通过表达式$\dfrac{N}{V(\Lambda)}=\rho$ 使它们产生关联。

换句话说，这个结果断言下式是存在的：

$$s(\rho,u,\delta u,F,J) := \lim_{N\to\infty}\frac{1}{N}\ln \mathcal{V}(\Lambda,N,u,\delta u,F,J) \tag{125}$$

这里，s 或者是有限的或者为 $-\infty$（进一步来说，对于 s 的参数的所有值来说，

$s=-\infty$ 的概率可以忽略)。这就为解释为什么概率式(123)是 J 的函数提供了一些线索。如果 J_1 和 J_2 为两个开区间，N 很大，并且为了标记的方便性，我们不考虑其他变量，那么我们认为：

$$\frac{\mu\left(\frac{F}{N}\in J_1\right)}{\mu\left(\frac{F}{N}\in J_2\right)}=\frac{\mathcal{V}(J_1)}{\mathcal{V}(J_2)}\approx e^{N(s(J_1)-s(J_2))} \tag{126}$$

如果 $s(J_2)>s(J_1)$，那么随着 N 的增大上式的值会指数性地趋向于零。因此，对于大型系统，概率 $\mu\left(\frac{F}{N}\in J\right)$ 只有对那些 $s(J)$ 足够大的开区间 J 来说才是可计算的。

我们可以得到下面这个更强的论断。通过集函数 $s(J)$ 我们可以定义一个点函数 s：

$$s(x):=\inf_{\substack{J\ni x\\ J\text{ open}}} s(J) \tag{127}$$

然后，我们可以表明，相反地，对于所有开区间 J：

$$s(J)=\sup_{x\in J} s(x) \tag{128}$$

此外，这是第二重要的结果，我们可以表明：

$$s(x)\text{是凹的} \tag{129}$$

进一步讲，$s(x)$ 在其域上的一个凸的开子集内是有限的[Lanford, 1973, 26]。

现在，很明显一个凹函数 $s(x)$ 可以有三种一般形态，关于它是否会达到其最大值，有三种情况：(1)从不；(2)恰有一次，例如在某个点 x_0 处；(3)在某些区间内。在情况(1)中，F/N 在热力学极限下会“逃逸到无穷大”，由于超稳定性条件(f′)需要被满足，这种情况可以被排除。针对我们的目的来说，情况(2)是我们最感兴趣的一种情况。在这种情况下，我们会考虑区间 $J_2=(x_0-\epsilon, x_0+\epsilon)$，其中 $\epsilon>0$ 为一个任意小量，J_1 是不包含 x_0 的任意开区间。从式(127)、式(128)中我们可以推断出 $s(J_2)>s(J_1)$，并能够从式(126)得出结论，F/N 从 J_2 而非 J_1 中取值的相对概率，随系统尺度的增大而指数性地趋于零。

因此，我们得到了所需的结果：当 N 逐渐变大时，F/N 的概率分布接近于一个 δ 函数。或者换句话说，在位形能量超壳层的绝大部分上，函数 F/N

变得大致为常量：

$$\lim_{N\to\infty}\mu\left(\left\{(\bar{q}_1,\cdots,\bar{q}_N)\in\Omega_{\Lambda,N,u,\delta u}:\left|\frac{F(\bar{q}_1,\cdots,\bar{q}_N)}{N}-x_0\right|>\epsilon\right\}\right)=0 \quad (130)$$

最后在情况(3)中，我们只能得出这样一个结论：概率分布趋向于集中在某些区域内，但是在这些区域内，具体的分布情况仍未确定。我们可以证明，如果这些区域是有界的，那么这种情况就与相变是相关的。①

备注

1. 相变。首先，很显然热力学极限理论的一个显著优点是，与遍历理论相反，在原则上它能够通过一个微观相互作用的模型来解释和预言相变的发生，在后来的工作中通常是与再归一化技术相结合在一起的。事实上，这种解释和预言的能力是其声名鹊起的主要原因，而不是关于遍历问题它必须说些什么。更重要的是，人们往往认为，相变在任何有限系统中是绝对不可能发生的，因此热力学极限是绝对必需的条件[Styer(斯泰尔)，2004；Kadanoff(卡达诺夫)，2000]。

这个论点提出一个问题，就是我们的经验，包括我们对真正物理客体的相变的认识，始终不能脱离有限系统。这样一个指出相变只存在于热力学极限下的理论，必然会被认为是理想化的。这个结论不会令众多物理学家们感到惊讶，因为理想化在理论物理学中是无处不在的。然而，令人感到奇怪的是，这种特殊的理想化似乎是“无法控制”的。进一步讨论参见[Sklar，2002]和[Liu(刘)，1999；Callender，2001；Batterman，2005]。我也注意到，最近有人提出了另一种观点。当然，在这种观点中，相变是与微正则超曲面$\{x: H(x)=$

① (不严密地说)要看到这种相关性，应该注意，如果我们从定义式(123)中去掉条件 $F/N\in J$——或等价地选择 $J=\mathbb{R}$——那么式(125)中的 s 可以被解释为每粒子的(超密、位形)微正则熵。现在将 s 看作开区间$(u-\delta u,\ u+\delta u)$上的一个函数，这样 s 就具有了为 $s(J)$ 而建构的那些属性，因为 U 本身就属于可观测量的类。因此，这里也存在一个类似于式(127)的点函数 $s(u)$，而且这个函数是凹的。事实上，如果我们在符号标记中再增加回来一个变量，记作 $s(\rho,\ u)$ 的话，该函数相对于两个变量来说是凹的。因此，在情况(3)中，在某些区间，比如说$[u_0',\ u_0'']$上，该函数相对于 u 是不变的。这意味着对于一系列 u 和 ρ 的值来说，存在一系列具有相同温度 $T=\left(\frac{\partial s}{\partial u}\right)_\rho^{-1}$ 的热力学态，这正是范·德·瓦尔斯气体中紧缩相变发生时的情况。

$E\}$上的拓扑结构变化相关的，其中 E 是可变的。当然，该理论与热力学极限理论的本质区别在于，这样的拓扑结构变化，可能会发生在有限的——事实上，即使是在很小的——系统中，参见[Gross, 1997; Gross and Votyakov, 2000; Casetti *et al.*（卡塞蒂等），2003]。然而，这可能正是热力学极限在处理遍历问题方面的长处，下面我会专注于讨论这个问题。

2. 重新探讨遍历问题。当对比于遍历理论或辛钦的方法时，热力学极限的理论有很大的优势。在辛钦的工作中，在物理学上有意义的系统中如何保证度量可递性的问题并没有出现，因为该方法并不需要做这种假设。另外，在辛钦的工作中，该方法仅适用于特殊函数。但由上述假设(a－f)或(a－f′)确定的这类函数并不是严格意义上的(对称)求和函数，而且它们允许粒子之间短距离相互作用的存在。因此，不同于辛钦所认为的那样，这里不存在方法论悖论。

然而，人们可能会质疑是否这些假设对物理上有意义的系统来说约束力不够。一方面，很显然，我们需要一些关于缓和性与稳定性的条件，以保证在热力学极限下，类似于系统内爆或爆炸的这类灾难性后果不会发生。另一方面，这些假设是否又有些太强了，以至于不能模拟现实的热系统。库仑相互作用(Coulomb interaction)，根据[Lieb and Lebowitz（利布和勒波维茨），1973, 138]的观点来看，是“与实际物质相关的”真正的力，但是它既没有被缓和也不是稳定的。莱纳德(Lenard)、戴森(Dyson)、勒波维茨和利布对这个力所做的研究，已经将热力学极限理论所得的主要成果推广到那些相互之间只存在纯粹库仑力的系统上(如果系统的净电荷为零或很小的话)，无论是在经典力学还是量子力学中(对于费米子而言)都成立，参见[Lieb, 1976]及其中引用的文献。那么，这个结果应覆盖大多数关于普通物质的微观模型，只要相对论效应和电磁力可以忽略不计。但要注意，这种扩展使用的是正则测量，而不是微正则测量，并且在与这些系综不等价的例子中(参见5.5节)，人们可能会担心这个结果是否适用于处于亚稳态的普通物质(类似于过饱和蒸汽和过热或过冷的液体)。

另外的一个显著要点是，不同于辛钦的结果式(115)，式(130)的结果根本不涉及时间平均。相反，在位形能量超曲面的一个很大的子集上，F/N 的瞬时值被认为几乎是恒定的。因此，对于无限时间极限来说，该理论也不存在问

题。事实上，动力学或时间演化对于目前的研究结果来说都不发挥任何作用，而且它与遍历理论的程序之间的差异比其与辛钦的方法之间的差异更加明显。

3. 遗留的问题。对比于解决遍历问题的另外两种方法，该理论还存在两个问题。首先，仍然存在的问题是为何要选择位形微正则测量（即为什么归一化勒贝格测度仅限于能源超壳层上）。兰福德直接指出，热力学极限的理论对于解决这个问题没有提供任何帮助：

> 一个更深刻的问题是，如何理解为什么当勒贝格测度是非常不可能的时候，相关事件就不会在自然界中发生。不幸的是，关于这个问题我无话可说。见[Lanford，1973，2]。

因此为了回答这个问题，我们不得不借助于其他方法来寻找原因。

其次，是测量 $-\varepsilon$ 的问题。根据式(130)，如果 N 很大，等式 $F/N \approx x$ 成立，小测度的集除外。那么我们能得出这样的结论：这个集合是可以被忽略不计，或者说，它的状态在自然界里不存在吗？事实上，结果式(130)的瞬时值是如此强，以至于我们应该要注意而不是给出太多的断言。例如，我们可能会错误地宣称，对于宏观系统（即 $N \approx 10^{27}$ 的系统），式(130)左边的集合在自然界中是不存在的。相反，一个关于经验的残酷事实是，宏观系统也会处于非平衡态。在这种状态下，可观测量的瞬时值随时间而略有不同，因此它们的微正则平均也是不同的。所以，它们的微观态必然会位于我们想要忽略的那个具有微小测度的集合内。当然，有人可能会质疑是否还需要更大的 N，比如 $N = 10^{100}$，但这只能说明在此极限下理想化是“不可控”的，即我们仍然无法控制 N 必须有多大，才能保证合理地用热力学极限来替代一个有限系统。

更多观点。还存在一些其他观点，在前面讨论的方法中没有涉及，它们对应于下述方法。这种方法的关键之处在于非常精巧地取极限的顺序。我们首先必须取一个超密能量壳层，然后取 N，Λ 为范·霍夫意义上的无穷大，最后取 δu 为零。不过，人们可能会问，这种取极限方式是否很清楚且很显然地是正确的，因为毕竟还有其他的取极限方式，以及其他非等价的极限存在（比如费舍尔意义上的极限），这些取极限的顺序显然是不对易的（能量超壳层的密度与

$N\delta u$ 成正比)，也有人[Compagner(卡姆佩格纳)，1989]提出过其他一些方法，如"连续极限"。

最后，为了能充分地与经典统计力学联系起来，我们还不得不将这些约束提升到位形空间，使它们在相空间中起作用。[Lanford，1973，2]将这作为一个"简单的习题"留给读者。让我们来看看是否我们能够提供这方面的细节。

假设我们从相空间上的超密微正则测量开始，它具有相同密度 $2N\delta u$，总能量值在 $E_0 = Ne_0$附近。其概率密度由下式给出：

$$\rho Ne_0,\ N\delta u(\vec{p}_1,\ \cdots,\ \vec{p}_N;\ \vec{q}_1,\ \cdots,\ \vec{q}_N) = \frac{1}{2N\delta u}\int_{E_0 - N\delta u}^{E_0 + N\delta u}\frac{1}{\omega(E)}\delta(H(x) - E)dE \tag{131}$$

对于哈密顿量(116)，可以对其动量进行积分，如玻耳兹曼[1868]所示，参见式(43)，这将产生一个边缘密度：

$$\bar{p}Ne_0,\ N\delta u(\vec{q}_1,\ \cdots,\ \vec{q}_N) = \frac{1}{2N\delta u}\frac{2m\pi^{3N/2}}{\Gamma(\frac{3n}{2})} \tag{132}$$

$$\int_{E_0 - N\delta u}^{E_0 + N\delta u}\frac{1}{\omega(E)}(2m(E - U(q)))^{(3N-2)/2}dE$$

这不完全是兰福德选择的位形空间上的归一化勒贝格测度，但由于因子 $(2m(E - U(q)))^{(3N-2)/2}$是关于 U 的一个连续函数——至少当 $E_0 - N\delta u - U > 0$ 时——对于壳层上的勒贝格测度来说这是绝对连续的，并且在 $\delta u \to 0$ 的极限下将收敛到勒贝格测度。

但是，在一个完整的相空间设定中，物理量也会依赖于动量，即它们会是 $F(\vec{p}_1,\ \cdots,\ \vec{p}_N;\ \vec{q}_1,\ \cdots,\ \vec{q}_N)$这样的函数，即使我们像对前一组参数一样假设其对第二组参数的依赖性同样要满足条件(a－f)，其概率分布也不会总是由位形微正则测量确定。例如，让 F_1和 F_2为位形空间上的两个可观测量，F_1/N 和 F_2/N 将在热力学极限下分别收敛到不同的值，比如说 x_1和 x_2，并且如果令 G 为一个关于动量的任意对称函数，则它会使两个不同值的概率分别为 1/2。例如：

$$G(\vec{p}_1,\ \cdots,\ \vec{p}_N) = \begin{cases} 1 & \text{如果 } \sum_i \vec{p}_i \cdot \vec{n} \geqslant 0 \\ 0 & \text{其他情况下} \end{cases} \tag{133}$$

对于某些特定的单位向量 $\vec{n}$。现在，考虑相空间上的函数：

$$A(\vec{p}_1, \cdots, \vec{p}_N; \vec{q}_1, \cdots, \vec{q}_N) =$$
$$G(\vec{p}_1, \cdots, \vec{p}_N)F_1(\vec{q}_1, \cdots, \vec{q}_N) + G'(\vec{p}_1, \cdots, \vec{p}_N)F_2(\vec{q}_1, \cdots, \vec{q}_N) \tag{134}$$

其中 $G' = 1 - G$。如果我们先对动量进行积分，我们就会得到 $\tilde{A} = \frac{1}{2}(F_1 + F_2)$，在热力学极限下它会收敛到$\frac{1}{2}(x_1 + x_2)$。然而，我们将会错误地认为，在相空间的一个绝对大的部分上，A 几乎等于$\frac{1}{2}(x_1 + x_2)(x_1 + x_2)/2$。实际上，在(粗略为)一半的相空间上它几乎等于 x_1，在另一半上它几乎等于 x_2。

这种将式(130)扩展到相空间函数上的做法会需要对这类函数的形式做出一些额外的假设，例如，它们对动量的依赖性仅表现为它们是动能的函数，即：

$$A(\vec{p}_1, \cdots, \vec{p}_N; \vec{q}_1, \cdots, \vec{q}_N) = \psi(\sum \frac{\vec{p}_i^{\,2}}{2m}) + F(\vec{q}_1, \cdots, \vec{q}_N) \tag{135}$$

对于某些连续函数 ψ 成立。

6.4　玻耳兹曼方程的兰福德研究方法

现在我们来考虑研究非平衡态统计力学的一些现代方法。其中，兰福德和其他人［Lanford，1975；Lanford，1976；Lanford，1981；Spohn（斯伯尼），1991；Cercignani *et al.*，1994］开发出的研究方法值得我们特别关注，因为它在概念上比其他任何现代的统计物理研究方法都更接近于玻耳兹曼于 1872 年在玻耳兹曼方程和 H 定理方面所做的研究。此外，兰福德提出并试图回答的问题，是与著名的可逆性和反驳问题同样重要的。此外，到目前为止，他得到的结果最好地说明了玻耳兹曼方程或 H 定理的统计解读对于硬质球气体也成立。

兰福德提出的问题是玻耳兹曼方程与基础哈密顿动力学的一致性问题。事实上，如果我们考虑的是力学系统如稀薄气体的微观态，似乎对其时间演化我们可以提供两种相互竞争的解释。

(1)一方面，给定气体的力学微观态 x_0，我们可以构建一个态分布 $f(\vec{r},\vec{v})$，

使得$f(\vec{r},\vec{v})d^3\vec{v}d^3\vec{r}$能给出位置介于$\vec{r}$和$\vec{r}+d^3\vec{r}$之间以及速度介于$\vec{v}$和$\vec{v}+d^3\vec{v}$之间的分子的相对数量。据推测，该分布应该是由微观态x_0唯一确定的。我们采用符号$f^{[x_0]}$来明确地表示这种依赖性。那么，该函数在理想情况下被认为是玻耳兹曼方程(48)的初始条件，解这个方程——假设它有一个唯一解——将会给出分布函数在后来时刻的形态，$f_t^{[x_0]}(\vec{r},\vec{v})$。

(2)另一方面，利用哈密顿方程，我们可以得到x_0在一段时间t内的演化。其结果为$x_t=T_tx_0$。那么后来的这个态x_t也是由状态分布$f^{[x_t]}(\vec{r},\vec{v})$确定的。

一个明显的问题是从最初的微观态得出后来的状态分布的这两种方式是否是相同的，即两类时间演化是否是一致的。换句话说，该问题就是，下图是否成立：

$$\begin{tikzcd} x_0 \arrow[r, "\text{哈密顿演化}"] \arrow[d] & x_t \arrow[d] \\ f^{[x_0]} \arrow[r, "\text{玻耳兹曼演化}"] & f_t^{[x_0]} \stackrel{?}{=} f^{[x_t]} \end{tikzcd} \tag{136}$$

这里要解决的第一个问题是微观态和状态分布f之间的精确关系是怎样的。很明显，在这个函数表示气体系统物理性质的条件下，它应该是由瞬时微观态x来决定的。同样清楚的是，为了满足玻耳兹曼方程，它被认为是连续的且对时间可微的，这不论从字面上还是本质来说都是不正确的。

因此，就像玻耳兹曼所做的那样，让我们假设该气体包含N个硬球，球的直径都为d，质量都为m，它们存在于一个确定的有界空间区域Λ内，其中$|\Lambda|=V$。给定系统的一个微观态x，我们就可以形成一个“精确”的状态分布：

$$F^{[x]}(\vec{r},\vec{v}):=\frac{1}{N}\sum_i^N\delta^3(\vec{r}-\vec{q}_i)\delta^3\left(\vec{v}-\frac{\vec{p}_i}{m}\right) \tag{137}$$

当然，这个分布不是一个严格意义上的函数，它既不连续也不可微，显然不是一个可以插入到玻耳兹曼方程中的适当对象。然而，我们可以合理地假设，我们应该能以极限的方式表达玻耳兹曼的思想，即粒子的数量N趋于无穷。但是，显然在分析这个极限时必须要小心。

一方面，我们应该保持气体是稀薄的，以使得涉及三个或更多粒子碰撞的情况是罕见的，相比于两个粒子间的碰撞，我们可以放心地忽略它们。另一方

面，气体又不能太稀薄，否则碰撞就太罕见了，而不能促成 f 的改变。一个适当的极限，就像兰福德所认为的那样，就是所谓的玻耳兹曼—格兰德极限，其中 $N\to\infty$，且：①

$$\frac{Nd^2}{v} = \text{常数} > 0 \tag{138}$$

将该极限记为"$N \xrightarrow{BG} \infty$"，其中它被隐含地理解为 $d \propto N^{-1/2}$。那么，该期望就是，在玻耳兹曼—格兰德极限下，具体分布 $F^{[x^n]}$ 会趋向于一个连续函数，而该函数可作为玻耳兹曼方程的一个适当的初始条件。为了达到这个目的，我们不得不为 μ 空间 $\Lambda\times\mathbb{R}^3$ 上的分布引入一个相关的收敛概念。一个合理的选择是承认分布函数的一个任意序列 f_n（或者是适当的密度函数或者是分布意义上的函数）会收敛到一个分布 f，即 $f_n\to f$，当且仅当下列条件成立：

对于每个长方体 $\Delta\subset\Lambda\times\mathbb{R}^3$：

$$\lim_{n\to\infty}\int_\Delta f_N d^3\vec{r}\,d^3\vec{v} = \int_\Delta f d^3\vec{r}\,d^3\vec{v} \tag{139}$$

和

$$\lim_{n\to\infty}\int \vec{v}^2 f_n d^3\vec{r}\,d^3\vec{v} = \int \vec{v}^2 f d^3\vec{r}\,d^3\vec{v} \tag{140}$$

其中第二个条件保证了平均动能的收敛性。

在上述两个分布相互接近的意义上，引入一个（适当的或不适当的）分布之间的距离函数也是方便的。这就是说，我们可以在 $\Lambda\times\mathbb{R}^3$ 上定义一个密度函数之间的距离 $d(f, g)$，使得：

$$d(f_n, f)\to 0 \Rightarrow f_n\to f \tag{141}$$

存在很多距离函数能够满足这个条件，但我不会去讨论如何从中挑选一个特定函数的问题。

那么，重申一次，我们所希望的就是，当 $N \xrightarrow{BG} \infty$ 时，在上述意义上 $F^{[x^n]}\to f$，其中 f 足够光滑可以充当玻耳兹曼方程的一个初始条件，并且在此定义

---

① 该条件可以被粗略地解释为 Nd^2/V 是与"平均自由路径"，即两次碰撞之间一个粒子运动距离的典型规模成正比的，或者还可以通过指出玻耳兹曼方程中的碰撞积分与 Nd^2/V 成正比，为了保证排列不变，我们保持玻耳兹曼方程不变。

下，玻耳兹曼和哈密顿演化在图(136)可交换的意义上是一致的。但很显然这仍然是一个棘手的问题。事实上，增大 N，意味着从一个力学系统过渡到另一个有更多粒子的力学系统。然而，并没有一个明显的算法能从态 x^N 构造出态 x^{N+1}，因此没有办法在单个态的层面强制收敛。

不过，人们可能相信一种乐观的猜想，如果属实的话，那它将会以一种近似的方式解决玻耳兹曼演化和哈密顿演化之间的一致性问题，如果 N 非常大的话。

> 乐观的猜想：如果 $F^{[x_t^N]}$ 接近 f，那么对于所有的 $t>0$，$F^{[x_t^N]}$ 就接近 f_t，其中 f_t 为具有初始条件 f 的玻耳兹曼方程的解。

兰福德[1976]指出，乐观的猜想不可能是正确的。这是可逆性反驳的一个直接后果。事实上，假设对所有 $x\in\Gamma$，且 $t>0$，这是真的(在这里，我们暂时去掉 x^N 的上标 N，以简写符号)。考虑由 x 通过反转全部动量得到的相点 Rx，即 $R(\vec{q}_1, \vec{p}_1; \cdots; \vec{q}_N, \vec{p}_N)=(\vec{q}_1, -\vec{p}_1; \cdots; \vec{q}_N, -\vec{p}_N)$。如果 $F^{[x]}(\vec{r}, \vec{v})$ 接近于分布 $f(\vec{r}, \vec{v})$，那么 $F^{[Rx]}(\vec{r}, \vec{v})$ 就接近于 $f(\vec{r}, -\vec{v})$。但是，随着 x 演化为 x_t，Rx_t 就演化为 $T_tRx_t=RT_{-t}x_t=Rx$。因此，$F^{[T_tRx_t]}(\vec{r}, \vec{v})=F^{[Rx]}(\vec{r}, \vec{v})$ 就接近于 $f(\vec{r}, -\vec{v})$。但关于 Rx_t 的猜想的有效性将要求 $F^{[T_tRx_t]}(\vec{r}, \vec{v})$ 接近于 $f_t(\vec{r}, -\vec{v})$，然而这两个状态分布是绝对不会相互接近的，除了在一些极个别情况下。

但是，即使乐观的猜想一般来说是错误的，人们可能也会希望它"很有可能"是真实的，即具有非常大的概率为真，至少在一些有限的时间段内。为了使这种策略更加明确，兰福德采取在 Γ 上进行概率测量的方法，或者更精确地说是在 Γ_N 的序列上进行一系列概率测量。

因此，除了在关于该问题的纯粹动力学理论探讨中引入一个统计元素之外，此方法还有一个明确的优势。正如上面提到的，不存在一个明显的算法可用来构造一个玻耳兹曼—格兰德极限下的微观态序列。但是对于测量来说，这

是不同的。例如微正则测量，不仅仅是一个 N 粒子系统的势能超曲面上的测量，对于每个 N，它定义一个此类测量的算法序列。

根据上述讨论，我们现在可以陈述兰福德定理如下[Lanford，1975；1976]。

令 $t \mapsto f_t$ 为玻耳兹曼方程的一些解，比如说当 $t \in [0, a) \subset \mathbb{R}$ 时。对于每个 N，令 ΔN 表示 N 个粒子构成的相空间 Γ_N 中的集合，在这个集合中 $F^{[x^N]}$ 接近于 f_0（玻耳兹曼方程解的初始条件），在下述意义下，对一些选定的距离函数 d 以及公差 $\epsilon > 0$ 有：

$$\Delta_N = \{x^N \in \Gamma_N:\ d(F^{[x^N]}, f_0) < \epsilon\} \tag{142}$$

另外，对于每个 N，限制微正则测量 μ_N 在 ΔN 上：

$$\mu_{\Delta,N}(\cdot) := \mu_N(\cdot \mid \Delta_N) \tag{143}$$

换言之，$\mu_{\Delta,N}$ 是对不同 Γ_N 进行的测量序列，它分配测度 1 给接近于 f_0 的微观态 $x^N \in \Gamma_N$ 的集合，在 $d(F^{[x^N]}, f_0) < \epsilon$ 这一意义上。那么：

$\exists \tau$，$0 < \tau < a$，使得对所有的 t，有 $0 < t < \tau$：

$$\mu_{\Delta,N}(\{x^N \in \Gamma_N:\ d(F^{[x^N]}, f_t) < \epsilon\}) > 1 - \delta \tag{144}$$

其中，随着 $\epsilon \longrightarrow 0$ 且 $N \xrightarrow{BG} \infty$，$\delta \longrightarrow 0$。

换句话说，如果通过对被限制于那些状态 x^N（其精确状态分布接近于一个给定的初始函数 f_0）的 Γ_N 进行的微正则测量来判断的话，遵循哈密顿动力学的很大比例即 $(1-\delta)$ 的演化，会以这样的一种方式进行，即它们状态后来的精确分布 $F^{[x^N]}$ 必然接近于函数 f_t，就如同遵循玻耳兹曼方程从 f_0 演化一样。

备注

兰福德定理表明，玻耳兹曼方程的统计版本和近似版本，可以由哈密尔顿力学和玻耳兹曼—格兰德极限下选择的初始条件来推出。这是一个引人注目的成就，在一定意义上它清楚地表达了玻耳兹曼的直觉。根据[Lanford，1976，14]，该定理说明，玻耳兹曼方程的近似有效性，以及 H 定理可仅仅从力学加上考虑初始条件推导出来。

不过，已经得到的结果有几个明显的问题，并且兰福德也承认所有这些问题。首先，它存在一些缺点，这使得这些结果在证明玻耳兹曼方程在现实生活中的物理系统里也有效时，并不能显示出任何实际的影响。气体密度为 N/d^3，且在玻耳兹曼—格兰德极限下会变为零。因此，这个结果对于极稀薄的气体成立。此外，该结果成立的那个时间段，即 τ，取决于式(138)中的常数，这也为气体平均自由程提供了一个粗略的量级。事实证明，如果考虑的是相同量级的话，τ 大约是两次碰撞之间的平均时间间隔的五分之二。这段时间短得令人失望，在室内的温度和空气密度下，τ 是微秒量级的物理量。因此，该定理对于证明玻耳兹曼方程能够应用于宏观现象来说没有任何帮助，因为宏观现象需要长得多的时间。

然而，需要注意的是，这么短的时间并不是毫无意义的。通常的一个误解是，该定理表明玻耳兹曼方程是在如此短的时间内才有效，以至于粒子并没有机会发生碰撞，事实上在两次碰撞之间平均间隔的五分之二时间内，粒子碰撞的概率约为 40% 。

另一个问题是，与玻耳兹曼自己的推导相比，这里似乎并没有明确提及分子混沌假设。这种认识仅仅在某种程度上是对的。在对该定理的一个更详细的叙述中，该定理并不是在微正则测量的基础上提出的，而是在(相空间序列)Γ_N 上的任意测量序列 ν_N 的基础上提出的，可参见[Lanford，1975；1976]这些测量是基于各种假设的。其中之一就是，对于微正则测量 μ_N，每个 ν_N 都应该是绝对连续的，即 ν_N 应该具有一个适当的密度函数：

$$d\nu_N(x) = n_N(x_1, \cdots, x_N)dx_1\cdots dx_N \tag{145}$$

这里 $x_i = (\vec{q}_i, \vec{p}_i)$ 表示第 i 个粒子的正则坐标。另外，对于每个 N，及 $m < N$，我们定义约化密度函数为：

$$n_N^{(m)}(x_1, \cdots, x_m) := \frac{N!}{(N-m)!}\frac{1}{N^m}\int n_N(x_1, \cdots, x_N)dx_{m+1}\cdots dx_N \tag{146}$$

为前 m 个粒子的(弱再归一化)边缘概率分布。现在的重要假设是：

$$\lim_{N \xrightarrow{BG} \infty} n_N^{(m)}(x_1, \cdots, x_m) = n^{(1)}(x_1)\cdots n^{(1)}(x_m) \tag{147}$$

在 $(\Lambda \times \mathbb{R}^3)^m$ 的紧致子集上一致成立。这个假设(可以证明对微正则测量成立)

容易被认为是分子混沌假设在测量理论方面的等价假设。它要求在玻耳兹曼—格兰德极限下，任何一个粒子对或包含 m 个粒子的元组在 $t=0$ 时刻的分子的各种量是统计独立的。兰福德也明确指出，如果我们采取可逆性反驳的结构，即如果在时间 t 内，$0<t<\tau$，我们按照 Δ_N 中的态 x 反转动量，并试着将该定理应用到集 $\Delta_N'=\{Rx_t;\ x\in\Delta_N\}$ 的话，那么假设式(146)将不能成立。

但是另一方面则需要更积极。即兰福德的定理并不需要明确地重复假设分子混沌假设成立。确实，一个了不起的成就是，一旦分解条件式(146)在时刻 $t=0$ 时成立，那么它也将在时间 $0<t<\tau$ 范围内成立，虽然是以一种较弱的形式(在测量中收敛，而不是一致收敛)。这有时被称为"混沌的蔓延"[Cercignani *et al.*, 1994]。

但兰福德定理考虑的主要概念性问题，是明显的不可逆性或时间反演非不变性来自何处。在这个问题上，已经出现了一些不同的意见。在[Lanford, 1975, 110]中认为，不可逆性是超过了玻耳兹曼—格兰德极限的结果。相反，在[Lanford, 1976]中认为，这是条件式(146)与初始条件(即 $x_N\in\Delta_N$)同时被满足的结果。

不过，我的观点与此不同。该定理对于 $-\tau<t<0$ 同样成立，其条件是 f_t 现在是反玻耳兹曼方程的一个解。这意味着该定理在时间反演下其实是不变的。

6.5 BBGKY 方法

所谓的 BBGKY 结构，是以博戈柳博夫、波恩、格林、柯克伍德和依云的首字母命名的，是对吉布斯和玻耳兹曼方法的一种独特的混合描述。该方法的目标是利用约化概率密度来描述系综的演化，以此来确定是否在合适的条件下可以得到类似玻耳兹曼方程的表达式——因此它是一种关于统计平衡的方法。

首先，考虑一个任意的随时间变化的概率密度 ρ_t。则 ρ 的演化是由包含哈密顿量的刘维尔方程确定的：

$$\frac{\partial \rho_t}{\partial t}=\{H,\ \rho\} \tag{148}$$

目前这种方法的首要任务是观测对于与统计力学中相关的系统来说，这个哈密顿量是否在粒子的不同排列下是对称的。事实上，由 N 个无法区分的粒子

构成的系统的哈密顿量，其形式通常如下：

$$H(\vec{q}_1,\vec{p}_1;\cdots;\vec{q}_N,\vec{p}_N) = \sum_{=1}^{N} \frac{\vec{p}_i^{\,2}}{2m} + \sum_{i}^{N} V(\vec{q}_i) + \sum_{i<j}^{N} \phi(\| q_i - \vec{q}_j \|) \quad (149)$$

其中，V 是表示有界空间区域 Λ 的容器壁的势，即：

$$V(\vec{q}) = \begin{cases} 0 & \text{如果} \quad \vec{q} \in \Lambda \\ \infty & \text{其他情况} \end{cases} \quad (150)$$

ϕ 为粒子 i 和 j 之间的相互作用势。这不仅在粒子标签的不同排列下是对称的，甚至还具有更特殊的属性，即它是从不依赖于两个以上粒子坐标的那些函数的总和(参见 6.3 节中的讨论)。

让我们再次使用符号 $x=(\vec{q}_1, \vec{p}_1; \cdots; \vec{q}_N, \vec{p}_N)=(x_1, \cdots, x_N)$，那么 $x_i=(\vec{q}_i, \vec{p}_i)$，考虑那些定义为 ρ 的次要项的约化概率密度函数的序列：

$$\begin{aligned} \rho^{(1)}(x_1) &:= \int \rho_t(x)\, dx_2 \cdots dx_N \\ \cdots\cdots\cdots\cdots &\vdots \\ \rho^{(m)}(x_1,\cdots,x_m) &= \int \rho_t(x)\, dx_{m+1} \cdots dx_N \end{aligned} \quad (151)$$

这里，$\rho^{(m)}$ 给出那些位于指定位置 $\vec{q}_1, \cdots, \vec{q}_m$，具有确定动量 $\vec{p}_1, \cdots, \vec{p}_m$ 的点 $1, \cdots, m$ 的概率密度，而所有剩余的粒子处于任意位置并具有任意动量。

哈密顿量的对称性并不意味着 ρ 的对称性。但是，有人可能会认为，我们应该要求概率密度是对称的，如果所有可观测量都是对称的话。在这种情况下，当两个或两个以上粒子的微观态互换，或者我们用所有排列下的平均值替换 ρ 且不改变任意可观测量的期望值时，我们就观察不到任何变化。无论如何这是可能的，事实上我们现在假设 ρ 在粒子标签的不同排列下是对称的。换句话说，从现在开始 $\rho^{(m)}$ 给出我们任意选择的由具有特定位置和动量值的 m 个粒子构成的集合的概率密度。

现在的指导思想是，对于相关的宏观量我们并不需要 ρ_t 随时间演化的具体形式。相反，我们将关注点放在由分级结构式(151)得到的一些边界值上。例如，假设一个可以表示为相位函数 A 的物理量，是 Γ 上的一个对称求和函数：

$$A(x)=\sum_{i=1}^{N}A(x_i) \tag{152}$$

那么：

$$\langle A\rangle=N\int A(x_1)\rho^{(1)}(x_1)dx_1 \tag{153}$$

这是一个相当大的简化形式。但是，这并不是说，我们可以如此容易地计算出$\langle A\rangle$随时间的演化。

我们特别考虑式(151)中的$\rho_t^{(1)}$。它是单粒子分布函数，任意粒子处于单粒子态$(\vec{p},\ \vec{q})$的概率。这个分布函数在一定程度上类似于玻耳兹曼的函数f。但请注意，ρ_1是一个边缘概率分布，它表征一个系综，而(在这种情况下)f是一个随机变量，表示单一气体的某个属性：

$$f(\vec{r},\ \vec{v})=\frac{1}{N}\sum_i\delta(\vec{q}_i-\vec{r})\delta(\vec{v}-\frac{\vec{p}_i}{m}) \tag{154}$$

那么$\rho_t^{(1)}$如何演化呢？由刘维尔方程，我们得到：

$$\frac{\partial\rho^{(1)}(x_1)}{\partial t}=\int\{H,\ \rho\}d^3\vec{p}_2\cdots\vec{p}_Nd\,\vec{q}_2\cdots\vec{q}_N \tag{155}$$

这里我们可以简单地把这个泊松括号看作是关于ρ的微分算符，它通常被称为刘维尔算子$\mathcal{L}$：

$$\mathcal{L}\rho:=\sum_{i=1}^{N}\left(\frac{\partial H}{\partial\vec{q}_i}\cdot\frac{\partial}{\partial\vec{p}_i}-\frac{\partial H}{\partial\vec{p}_i}\cdot\frac{\partial}{\partial\vec{q}_i}\right)\rho \tag{156}$$

对于式(149)的哈密顿量，它可展开为：

$$\mathcal{L}=\sum_{i}^{N}\mathcal{L}_i^{(1)}+\sum_{i<j}^{N}\mathcal{L}_{ij}^{(2)} \tag{157}$$

其中

$$\mathcal{L}_i^{(1)}:=\vec{p}_i\cdot\frac{\partial}{\partial\vec{q}_i} \tag{158}$$

以及

$$\mathcal{L}_{ij}^{(2)}:=\frac{\partial\phi_{ij}}{\partial\vec{q}_i}\cdot\left(\frac{\partial}{\partial\vec{p}_i}-\frac{\partial}{\partial\vec{p}_j}\right) \tag{159}$$

$\rho^{(1)}$的演化由下式给出：

$$\frac{\partial \rho_t^{(1)}(x_1)}{\partial t}=\mathcal{L}_1^{(1)}\rho_t^{(1)}(x_1)+\int dx_2\,\mathcal{L}_{12}^{(2)}\rho^{(2)}(x_1,\ x_2) \tag{160}$$

更一般地，对于高阶约化分布函数$\rho^{(m)}$（$m\geqslant 2$时），演化由下式确定：

$$\frac{\partial \rho_t^{(m)}(x_1,\ \cdots,\ x_m)}{\partial t}=\sum_{i=1}^{m}\mathcal{L}_i^{(1)}\rho_t^{(m)}(x_1,\ \cdots,\ x_m)+\sum_{i<j=1}^{m}\mathcal{L}_{ij}^{(2)}\rho_t^{(m)}(x_1,\ \cdots,\ x_m)$$
$$+\sum_{i}^{m}\int dx_{m+1}\,\mathcal{L}_{i,m+1}^{(2)}\rho_t^{(m+1)}(x_1,\cdots,x_{m+1}) \tag{161}$$

方程式(160)和(161)构成了 BBGKY 分级结构。它严格等价于相互对称的ρ和H的哈密顿形式体系，当H不包含与三个或更多粒子之间相互作用有关的项时。正如人们所预期的，解这些方程与解原来的哈密顿方程一样难。特别是，这些方程不是封闭的：为了确定$\rho_t^{(1)}$是如何演化的，我们需要知道$\rho_t^{(2)}$的情况；为了知道$\rho_t^{(2)}$的演化，我们需要知道$\rho_t^{(3)}$；等等。

克服这个问题的常规方法是切断分级结构之间的联系。即假设对于一个有限的m来说，$\rho^{(m)}$是$\rho^{(l)}$的泛函，其中$l<m$。特别地，如果我们只考虑最简单的情况($m=2$)，以及泛函的最简单形式，我们可以取$\rho^{(2)}$来进行因式分解，在很遥远的过去($t\to-\infty$)，如果下式给定的话：

$$\rho_t^{(2)}(x_1,\ x_2)=\rho_t^{(1)}(x_1)\rho_t^{(1)}(x_2);\ \text{如果}\ t\to-\infty \tag{162}$$

即要求任意粒子对的分子态在它们发生相互作用之前都是不相关的。这是分子混沌假设式(29)的相似式，但现在，当然是在系综的约化分布函数的意义上进行公式化的。

可以证明，在均匀介质的情况下，即当$\rho^{(2)}$在位置$\vec{q}_1$和$\vec{q}_2$上一致，即当$\rho^{(2)}(x_1,\ x_2)=\rho^{(2)}(\vec{p}_1,\ \vec{p}_2)$且$\phi$是一个有限范围内的相互作用势时，$\rho^{(1)}$的演化方程在形式上与玻耳兹曼方程式(48)相同。也就是说，在式(160)中，我们可令$\mathcal{L}_i^{(1)}=0$，从而：

$$\begin{aligned}\frac{\partial \rho_t^{(1)}(\vec{p}_1)}{\partial t}&=\int \mathcal{L}_{12}^{(2)}\rho(\vec{p}_1,\vec{p}_2)d^3\vec{p}_2\\&=\frac{N}{m}\int bdbd\phi\int d\vec{p}_2\,\|\vec{p}_2-\vec{p}_1\|\\&\quad(\rho_t^{(1)}(\vec{p}_1{}')\rho_t^{(1)}(\vec{p}_2{}')-\rho_t^{(1)}(\vec{p}_1)\rho_t^{(1)}(\vec{p}_2))\end{aligned} \tag{163}$$

更多细节，请参见[Uhlenbeck and ford，1963，131]。

备注

BBGKY 方法看起来完全是吉布斯式的，也就是说，它将相空间的概率密度看作其基本概念工具。而吉布斯并没有广泛使用一种新引入的因素，该因素是它对排列对称性的依赖。这极大地拓展了吉布斯自已的工作，通过为约化(或边缘化)密度函数提供一个关于演化方程的系统性分级结构，就可以将微扰理论的方法应用于约化密度函数。玻耳兹曼方程的一个以系综为基础的等价方程可以由这种方法得到，我们将其作为稀薄气体的一阶近似，这里的稀薄气体是指在其中粒子碰撞次数远小于平均自由时间。对于非均匀气体，玻耳兹曼方程不能如此轻易就获得——正如人们所预料的，在物理层面，我们还需要额外的假设以保证其有效性。

将这种方法与兰福德的方法进行比较是有启发性的。他关于玻耳兹曼方程的等价方程是为了另一类函数而给出的，即单粒子分布函数 $F^{[x]}$，即具有分子态 $(\vec{r},\vec{v})$ 的粒子的精确相对数，而非 $\rho^{(1)}$。当然，两者之间是存在联系的。注意 $F^{[x]}$ 是一个求和函数。对比式(137)，我们可以发现：

$$\langle F^{[x]}\rangle = \int \rho^{(1)}(\vec{p}_1,\ \vec{q}_1) f\ \ \delta(\vec{r}-\vec{q}_1)\delta\left(\vec{v}-\frac{\vec{p}_1}{m}\right) dp_1 dq_1 = \rho^{(1)}(\vec{r},\ \vec{v}) \quad (164)$$

换句话说，单粒子分布函数 $\rho^{(1)}$，是确切态分布的预期值。因此，看起来兰福德描述的是态的确切分布演化的概率，而 BBGKY 的结果式(164)描述的是态的确切分布的平均值的演化。因此，兰福德的结果包含了更多的信息。

有人在这里可能会认为，通过利用类似于证明遍历理论时用到的论证方法，我们可以证明，真实的粒子分布可能会被看作等价于其系综平均。特别是，我们从辛钦的工作(参见 6.3 节)中已经看到，对于足够大的系统，一个求和函数例如 $F^{[x]}$ 明显偏离其期望值的概率是可以忽略不计的。然而，一个严重的问题是，这种对辛钦结果的解读，仅对于平衡态是成立的，即它们仅适用于微正则分布 ρ_{mc}，而不适用于这里所面对的任意与时间相关的密度 ρ_t。

这个作为结果的方程的时间不对称性不是从方程的分级结构中得到的，而是从基于系综的类似于分子混沌假设的等价式(162)中得到的。也就是说，在这种

方法中，时间不对称性是通过一个系综上的初始条件，即存在初始相关性这个条件引入的。就像对原来的玻耳兹曼方程那样，可以证明，如果边界条件变为碰撞后动量是相互无关的，即式(162)换作对 $t\to\infty$ 成立，那么我们就会得到反玻耳兹曼方程，见[Uhlenbeck and Ford，1963，127]。

7. 随机动力学

7.1 简介

近几十年来，出现了一些非平衡态统计力学的方法，虽然它们彼此在基础认识和哲学观点方面是明显不同的，但它们仍然趋向于发展出一个统一的数学框架。我将称此框架为“随机动力学”，因为这种方法的主要特征就是，它将一个力学系统的态的演化描述为按照随机映射进行的，而不是按照一个确定的且时间反演不变的哈密顿动力学进行的。①

使用这种随机动力学类型的原因源于不同的背景，我们至少可以找到三种不同的观点。

1.“粗粒化”，参见[van Kampen，1962；Penrose，1970]。在这种观点下，我们假设，在微观层面，系统可被定性为一个(哈密顿)动力学系统，它具有确定的时间反演不变的动力学。然而，在宏观层面，我们只关心宏观态的演化，即我们关心的是如何将微观相空间分割(或粗粒化)为离散的单元格。通常的观点是，这些单元格的形状和尺寸是根据我们观测能力的极限而选择的。7.5 节给出了对这一观点的详细阐述。

在宏观层面上，演化现在不再需要被描述为确定的。给定的只是系统的瞬时宏观态，一般来说下一刻的宏观态是不确定的，即使底层的微观演化是确定性的。相反，可以给出一个转移概率，用它来规定从任何给定的初始宏观态过渡到后来的宏观态的概率。就随机过程来说，虽然没有进一步的假设，我们不可能对于宏观特征态的演化进行一般性的描述，但是我们可以描述系综的演化

① 此外，苏达山(Swdarshan)及其同事已经在这个意义上使用过这个名称，参见[Sudarshan *et al.*，1961；Mehra and Sudarshan(梅拉和苏达山)，1972]。

或这些态的概率分布。

2. “干涉”“寻迹”或“开放系统”，参见[Blatt，1959；Davies(戴维斯)，1976；Lindblad(林德布拉德)，1976；Lindblad，1983；Ridderbos，2002]。在这种观点下，我们假设所要描述的系统不是孤立的，而是与环境之间存在相互作用的。我们假设，总的系统，包括我们关注的系统和环境可以被描述为一个具有时间反演不变性和确定性动力学的(哈密顿)动力学系统。如果我们用 $x \in \Gamma^{(s)}$ 表示系统的状态，$y \in \Gamma^{(e)}$ 表示环境的态，那么它们的联合演化可以由一个单参数演化变换组来给出，它是由复合系统的哈密顿运动方程：U_t：$(x, y) \mapsto U_t(x, y) \in \Gamma^{(s)} \times \Gamma^{(e)}$ 产生的。状态 x 随时间进程的演化是通过在每个 t 时刻投影到 $\Gamma^{(s)}$ 上 $U_t(x, y)$ 的坐标来表达，我们称这个投影结果为 x_t。显然，系统的这种约化时间演化，一般是不确定的，例如，$\Gamma^{(s)}$ 上由 x_t 描述的轨迹自己可能就会相交。

同样，我们可能会希望，对于由系统及其环境组成的系综来说，这种非确定性演化仍然可以被描述为一个随机过程，至少当我们做出一些更深一层的合理的假设时。

3. 第三个观点是否定本质上的确定性或时间反演不变性动力学的存在[Mackey(麦基)，1992；2001]，或者认为其不可知[Streater(斯特赖特尔)，1995]，认为应该仅仅把一个系统的演化看作由一个随机过程描述，这种随机过程有资格成为一种新的基本的动力学。

因此，虽然采用这种方法的学者们在动机方面不同，并且在他们各自的领域对此给出了不同的解释，但是正如我们看到的，非平衡态统计力学在数学形式上显然是统一的。很明显，他们希望从随机动力学角度来对系统的演化所做的这种描述能以某种方式使得演化会典型地表现为“不可逆的行为”，即玻耳兹曼演化方程所认为的那种“朝向平衡态演化的趋势”，在 H 定理中也有一个随机的类似趋势。我们想要恢复那种非平衡态热系统通常会表现出来的自主性和不可逆性。

我们将会看到，这在很大程度上可以用相对较小的努力实现，一旦我们做出关键的技术性假设的话。随机过程实际上是一个齐次的马尔可夫过程(Markov process)，或者等价地，服从所谓的主方程。更难的问题是关于这种方法的

中心假设是否仍然能与底层的确定性的具有时间反演不变性的动力学相兼容，以及在什么意义上这种方法得到的结果会体现出时间不对称性。事实上，我们将会看到，后面的这个问题中提到了相冲突的直观认识，取决于我们是采取从一个概率的观点还是从动力学的角度来理解这种数学形式。

从一种基础的观点来看，随机动力学提出了一种解释不可逆行为的新方法，它区别于正统的哈密顿或动力学系统的方法。在该方法中，任何对不可逆现象的说明，都需要涉及一些特殊的初始条件或动力学假设。此外，众所周知的是，这种系统构成的系综中(精细化)吉布斯熵是守恒的，这说明在推导熵增加结果时，不能依赖于这种形式的熵。

然而，在随机动力学中，人们可能希望发现某个不依赖于特殊初始条件的不可逆过程，即希望该过程内构于随机动力学演化过程。此外，由于刘维尔定理不适用于这种情况，因此存在这样一种可能性，即我们可以通过这类动力学得出吉布斯熵是真正增加的。

正如刚才提到的，随机动力学中的核心技术性假设是我们所描述的过程具有马尔可夫性。[①]事实上，不可逆行为的一般特征可以毫不费力地从马尔可夫特征中得到，或者从与其紧密相关的"主方程"中得到。因此，对随机动力学的关注大部分已经转向了得到这种马尔可夫性所需的假设上，或者稍微再强一些的论断是，已经转向了得到不可逆的马尔可夫过程所需的假设上 [Mackey, 1992]。关于这种假设最有名的一个案例是[van Kampen, 1962]的"重复随机性假设"。同样，对这种方法的批判 [Sklar, 1993; Redhead, 1995; Callender, 1999]也将他们的反对意见集中于这种假设为何是合理的且普遍的这个问题上了(参见7.5节)。

我相信，争论的双方都完全没有找到要点。许多学者都不加批判地假设关于(不可逆)马尔可夫过程的假设确实会导致非时间反演不变性的结果。然而，事实上，(无论对于可逆的还是不可逆马尔可夫过程来说)马尔可夫性都是时间反演不变的。因此，任何需要获得该特性的论证并不用预设时间不对称性。事

① 有些作者认为，该方法可以而且应该推广到包含非马尔可夫随机过程。不过，在这里我将只关注马尔可夫过程。

实上，我将会论证，这些由马尔可夫性引出的关于不可逆行为的讨论受直觉所累。这是由于我们已习惯于从给定的初始状态来预测未来的状态，而不习惯于反推之前的状态。正如我们将要看到的，对于随机动力学中不可逆性的一个适当描述来说，我们需要注意另外一个问题，即前向转移概率和后向转移概率之间的差异。

在接下来的一节中，我首先在 7.2 节将从随机过程理论出发来重述齐次马尔可夫过程的标准定义。在 7.3 节用动力学的语言列出这些概念，引入主方程，并讨论它与玻耳兹曼方程之间的相似性。在 7.4 节中，我们将回顾那些从表面上看揭示了齐次马尔可夫过程中的不可逆行为的研究结果。在 7.5 节，我们将转向讨论给出马尔可夫性的物理动机及其问题，7.6 节关注的问题是看似不可逆的结果是如何从时间对称性假设中获得的。最后，7.7 节指出，关于这些问题更有希望的一些讨论应该从对随机过程可逆性的新定义开始。

7.2　马尔可夫过程的定义

在开始之前，我们先考虑这样一个例子。在物理文献中关于随机过程的最早的讨论之一是 P·埃伦菲斯特和 T·埃伦菲斯特(1907 年)提出的所谓的“狗蚤模型”。

考虑 N 个跳蚤，将其编号为 1 到 N，它们分别寄生在两只狗身上。狗 1 和 2 身上的跳蚤数分别为 n_1 和 $n_2 = N - n_1$。此外，我们想象有一个装有 N 个签的瓮，其中的签上分别标有 1 至 N 的数字。我们摇动这个瓮，掉出来很多签(然后再放回)，同时具有相应编号的跳蚤按顺序跳到另一个狗身上。这个过程每秒都在重复。

不难看出，这种模式在一定程度上体现出一个“朝向平衡的趋势”，假设最初所有或近乎所有的跳蚤都寄生在狗 1 身上。然后，非常有可能的是，第一次抽签会使狗 1 身上的跳蚤移动至狗 2 身上。但是，一旦狗 2 身上的跳蚤数量增加，那么跳蚤跳回到狗 1 身上的概率也会增加。这类典型行为，即 $|n_1 - n_2|$ 作为时间的函数是与玻耳兹曼 H 曲线相类似的，如果最初 $|n_1 - n_2|$ 很大的话，那么它会趋向于减小，并在大部分时间内保持处于 $n_1 \approx n_2$ 的“均衡”值。但是请注意，对比于气体理论中的玻耳兹曼 H 曲线，这里的“演化”是完全随机的，即

是完全由偶然事件造成的，在底层并没有确定性运动方程存在。

一般说来，一个随机过程在数学上而言不过是测量空间 X 上的一个概率测量 P，其元素在这里将被记为 ξ，它包含无穷多个随机变量 Y_t，其中 $t\in\mathbb{R}$（或有时候 $t\in\mathbb{Z}$）。从物理上来说，我们将 t 解释为时间，将 Y 解释为表征宏观态的宏观变量——比如说狗身上跳蚤的数目，或分子态位于 μ 空间中单元格上的分子的数目，等等。另外，ξ 表示确定 $Y_t(\xi)$ 的值的系统的总历史。因此 Y_t 可被认为是涉及时间进程的一个单一随机变量 Y。

初看起来，“过程”这个名称对于概率测量来说似乎有点不自然。从物理学的角度来看，随机变量 Y_t 得到其值 $Y_t(\xi)=y_t$ 的这种现实化操作应该被称为一个过程。然而，在数学文献中，通常将确定所有这类现实化操作概率的测量表示为“随机过程”。

为方便起见，我们这里假设变量 Y_t 只能取有限的离散值，比如说 $y_t\in\mathcal{Y}=\{1,\cdots,m\}$。不过，对于连续变量，这个理论的形式在很大程度也类似。

在时刻 $t_1,\cdots,t_n$，当 $n=1,2,\cdots$ 时，概率测量 P 为事件提供了确定的概率，因此 Y_t 在这些时刻只能取特定值 $y_1,\cdots,y_n$：

$$\begin{aligned}&P_{(1)}(y_1,t_1)\\&P_{(2)}(y_2,t_2;y_1,t_1)\\&\quad\vdots\\&P_{(n)}(y_n,t_n;\cdots;y_1,t_1)\\&\quad\vdots\end{aligned}\tag{165}$$

这里，$P_{(n)}(y_n,t_n;\cdots;y_1,t_1)$ 表示在时刻 $t_1,\cdots,t_n$，变量 Y_t 取值为 $y_1,\cdots,y_n$ 时的联合概率，其中 $y_i\in\mathcal{Y}$。这是下式的缩写，即：

$$P_{(n)}(y_n,t_n;\cdots;y_1,t_1):=P(\{\xi\in X:Y_{t_n}(\xi)=y_n\&\cdots\&Y_{t_1}(\xi)=y_1\})\tag{166}$$

显然，概率式(165)是归一化的且是非负的，而且每个 $P_{(n)}$ 是所有高阶概率分布的一个边际化：

$$P_{(n)}(y_n,t_n;\cdots;y_1,t_1)=\sum_{y_{n+m}}\cdots\sum_{y_{n+1}}P_{(n+m)}(y_{n+m},t_{n+m};\cdots;y_1,t_1)\tag{167}$$

事实上，概率测量 P 由层次结构式(165)唯一地确定。[①]

同样，我们能以熟悉的方式定义条件概率，如：

$$P_{(1|n-1)}(y_n, t_n \mid y_{n-1}, t_{n-1}; \cdots; y_1, t_1) :=$$

$$\frac{P_{(n)}(y_n, t_n; \cdots; y_1, t_1)}{P_{(n-1)}(y_{n-1}, t_{n-1}; \cdots; y_1, t_1)} \tag{168}$$

提供了 Y_{t_n}取值为 y_n 的概率，在 $Y_{t_{n-1}}$，…，Y_{t_1}取值为 y_{n-1}，…，y_1 的条件下。

理论上，在联合概率分布式和条件概率分布式(165)和(168)中时间可以是以任意的顺序出现的。但是，从现在开始，我们按照惯例，认为它们是有序的，即 $t_1<\cdots<t_n$。

一类特殊且重要的随机过程可以通过增加这样的一个假设来获得，即假设条件概率只取决于最后时刻的条件。这就是说，对于所有的 n 以及所有 y_1，…，y_n 和 $t_1<\cdots<t_n$：

$$P_{(1|n)}(y_n, t_n \mid y_{n-1}, t_{n-1}; \cdots; y_n, t_n) = P_{(1|1)}(y_n, t_n \mid y_{n-1}, t_{n-1}) \tag{169}$$

这就是马尔可夫性以及那些称之为马尔可夫过程的随机过程。

我们经常为这个假设做出的解释是，马尔可夫过程“没有记忆”。为了更准确地解释这个词，请考虑以下情况。假设我们了解变量 Y 的一小段历史，在时刻 t_1，…，t_{n-1}，它的值分别为 y_1，…，y_{n-1}。在这个信息的基础上，我们想预测变量 Y 在下一时刻 t_n 的值 y_n。马尔可夫性式(169)表明，即使我们知道变量 Y 前期的全部历史来做这种预测的话，也并不会比我们仅知道前一时刻 t_{n-1}时 Y 的取值为 y_{n-1}就做出的预测更好或更坏。因此，关于过去取值的额外信息与如何预测未来的取值是不相关的。

对于一个马尔可夫过程，联合概率分布式(165)的层次结构服从一个严格的要求。事实上，它们可完全由下面的步骤确定。

(a)任意选择的初始时刻 $t=0$ 时，$P_{(1)}(y, 0)$的具体情况。

(b)条件概率 $P_{(1|1)}(y_2, t_2 \mid y_1, t_1)$可适用于所有的 $t_2>t_1$。事实上：

$$P_{(1)}(y, t) = \sum_{y_0} P_{(1|1)}(y, t \mid y_0, 0) P_{(1)}(y_0, 0) \tag{170}$$

---

① 至少，当我们假设 X 中可测集的 σ 代数是式(166)右边那种形式的集所产生的柱代数时。

且对于联合概率分布 $P_{(n)}$，我们发现：

$$P_{(n)}(y_n, t_n; \cdots; y_1, t_1) = P_{(1|1)}(y_n, t_n | y_{n-1}, t_{n-1})$$
$$P_{(1|1)}(y_{n-1}, t_{n-1} | y_{n-2}, t_{n-2}) \times \cdots \times P_{(1|1)}(y_2, t_2 | y_1, t_1) P_{(1)}(y_1, t_1) \tag{171}$$

从马尔可夫性可以断定，条件概率 $P_{(1|1)}$ 具有以下属性，它被称为查普曼—柯尔莫哥洛夫(Chapman-Kolmogorov)方程：

$$P_{(1|1)}(y_3, t_3 | y_1, t_1) = \sum_{y_2} P_{(1|1)}(y_3, t_3 | y_2, t_2)$$
$$P_{(1|1)}(y_2, t_2 | y_1, t_1), \text{对 } t_1 < t_2 < t_3 \tag{172}$$

因此，对于一个马尔可夫过程，层次结构式(165)是完全可以由初始时刻的概率 $P_{(1)}$ 以及具有满足查普曼—柯尔莫哥洛夫方程的条件概率 $P_{(1|1)}$ 的一个系统来描述。因此，关于马尔可夫过程的研究主要集中于这两方面。①

另一个特殊假设是齐次性。如果条件概率 $P_{(1|1)}(y_2, t_2 | y_1, t_1)$ 不依赖于具体的两个时刻 t_1 和 t_2，而是只依赖于它们的差值 $t = t_2 - t_1$，即它们在时间平移下不变的话，那么这个马尔可夫过程就被称为是齐次的。在这种情况下，我们可以将其记作：

$$P_{(1|1)}(y_2, t_2 | y_1, t_1) = T_t(y_2, y_1) \tag{173}$$

这类条件概率也被称为转移概率。

对一个马尔可夫过程的定义是时间对称的吗？在式(169)中 Y_m 的概率分布的取值以先前 Y_t 的值为条件，这当然是很特别的情况。原则上，式(165)或式(168)中并没有某种对这种排序的要求。这样的话，也许我们要求的就是，先前 Y_t 的一个值的概率，是以后来的一部分行为为已知条件的(或者更确切地说，是以在前面时刻以及后面时刻的行为为条件的)。

初看之下，马尔可夫性并不要求后一种情况的出现。因此，人们可能很容

---

① 但是，请注意，虽然每一个马尔可夫过程都可以由这两个步骤完全描述：(i)初始分布 $P_{(1)}(y, 0)$；(ii)遵循查普曼—柯尔莫哥洛夫方程和方程式(171)的转移概率 $P_{(1|1)}$ 的集合，但并不是每个满足(i)和(ii)的随机过程都是马尔可夫过程。请参阅[van Kampen, 1981, 83]的反例。然而，事实是我们确实可以按照式(171)通过这两个步骤来定义一个唯一的马尔可夫过程。

易就误以为这个定义是时间不对称的。然而，情况并不是这样。可以证明式(169)等价于：

$$P_{(1\mid n-1)}(y_1,\ t_1\mid y_2,\ t_2;\ \cdots;\ y_n,\ t_n)=P_{(1\mid 1)}(y_1,\ t_1\mid y_2,\ t_2)\tag{174}$$

这里 $t_1<t_2<\cdots<t_n$ 的约定仍然有效。因此，一个马尔可夫过程不仅是“没有记忆”的，而且也是“没有预见性”的。一些学者采用了马尔可夫过程的一个(等效)定义，参见[Kelly(凯利)，1979]，它是明确地满足时间对称性的：假设在时间序列中段的某个时刻 t_i 的值 y_1 给定。那么一个随机过程是马尔可夫过程的条件为：

$$P_{(n\mid 1)}(y_n,\ t_n;\ \cdots;\ y_1,\ t_1\mid y_i,\ t_i)=P_{(n-i\mid 1)}(y_n,\ t_n;\ \cdots;\ y_{i+1},\ t_{i+1}\mid y_i,\ t_i)$$
$$P_{(i-1\mid 1)}(y_{i-1},\ t_{i-1};\ y_1,\ t_1\mid y_i,\ t_i)\tag{175}$$

对于所有的 $n=1,\ 2,\ \cdots$以及所有的 $1\leqslant i\leqslant n$。换个说法就是：过去和将来是无关的，如果我们以现在为条件的话。

7.3 随机动力学

$t>0$ 时刻的齐次马尔可夫过程可完全由初始概率分布 $P_{(1)}(y,\ 0)$ 和式(173)定义的转移概率 $T_t(y_2\mid y_1)$ 确定。P 和 T 之间在符号上的差异还可用于改进一个特定的概念性步骤。也就是说，这个观点是将 T_t 视为一个随机演化算子。因此，我们可以把 $T_t(y_2\mid y_1)$ 作为一个矩阵的元素，这个矩阵表示一个(线性)操作 T，它可以确定初始分布 $P_{(1)}(y,\ 0)$是如何演化为 $t>0$ 时刻的分布的。后面我会修改这种符号并将 $P_{(1)}(y,\ t)$ 记作 $P_t(y)$：

$$P_t(y)=(T_tP)(y):=\sum_{y'}T_t(y\mid y')P_0(y')\tag{176}$$

那么查普曼—柯尔莫哥洛夫方程式(172)就可以被简写为：

$$T_{t+t'}=T_t\circ T_t',\ \text{对}\ t,\ t'\geqslant 0\tag{177}$$

其中∘代表矩阵相乘，并且我们现在扩展该符号，以包括单位算符：

$$\mathbb{1}(y,\ y')=T_0(y,\ y'):=\delta_{y,y'}\tag{178}$$

这里 δ 表示克罗内克符号 δ。

式(177)(几乎)可以被解释为演化算符 T 的群合成性。有意义的是我们应该注意这在多大程度上是基于马尔可夫性的。事实上，对于任意的条件概率，

比如说，如果 A_i，B_j 和 C_k 分别表示由完备的且相互排斥的事件组成的三个族（即对于 $i \neq i'$，$j \neq j'$，以及 $k \neq k'$ 来说，$\cup_i A_i = \cup_j B_j = \cup_k C_k = \mathcal{Y}$；$A_i \cap A_{i'} = B_j \cap B_{j'} = C_k \cap C_{k'} = \emptyset$）。总概率的规则由下式给出：

$$P(A_i \mid C_k) = \sum_j P(A_i \mid B_j, C_k) P(B_j \mid C_k) \tag{179}$$

一般说来，这个规则不能被视为普通的矩阵乘法或群合成！但马尔可夫性使得式(179)中的 $P(A_i \mid B_j, C_k)$ 简化为 $P(A_i \mid B_j)$，因此式(179)中的求和与我们熟悉的矩阵乘法规则相一致。

我上面说“几乎”，是因为对比于通常的群属性，这里还存在一些区别。在查普曼—柯尔莫哥洛夫方程式(177)中，所有的时间都必须是正的。因此，在一般情况中，$t>0$ 时，T_t 甚至都不能被定义，所以也不会有：

$$T_{-t} \circ T_t = \mathbb{1} \tag{180}$$

一个在式(177)的$\circ$操作下封闭，且 $T_0 = \mathbb{1}$ 的算符 $\{T_t, t \geqslant 0\}$ 的群被称为半群。它不同于一般意义上的不要求元素 T_t 是可逆的（即不需要有逆元素的）那种群。T_t 缺乏逆元素可能有多方面的原因：可能是 T_t 不具有逆元素，即它不是一个一对一映射，也可能 T_t 具有逆矩阵 T_t^{inv}，但它是非随机的（例如，它可能有负矩阵元）。在 7.4 节和 7.7 节中我们会再探讨逆矩阵的作用。

有关马尔可夫过程的理论与线性代数有一个很强的且很自然的关联。有时候，该理论完全以这种图景呈现，我们一开始就引入了一个随机矩阵半群，比如说，$m \times m$ 矩阵 T，其中 $T_{ij} \geqslant 0$ 且 $\sum_i T_{ij} = 1$。或者，更抽象地，我们设想一类态 P，模为 $\| P \|_1 = 1$ 的巴拿赫(Banach)空间的元素，和一个随机映射 T_t 半群，其中 $t \geqslant 0$ 时满足 T_t 的是线性的、正的，且保证模 $\| T_t P \|_1 = \| P \|_1$ 不变这个条件，参见[Streater，1995]。

因此当 t 是离散的时（$t \in \mathbb{N}$），概率分布 P（现在被视为一个向量或一个态）的演化就特别简单：

$$P_t = T^t P_0，\text{其中 } T^t = \underbrace{T \circ \cdots \circ T}_{t\text{个}} \tag{181}$$

离散时间下的齐次马尔可夫过程也被称为马尔可夫链。

显然，如果我们将族 $\{T_t\}$ 看作一个随机演化算符半群，或一种随机形式的动力学，那么将 $P_0(y)$ 看作一个可能的初始态就变得引人注目了，如果演化算

符T_t独立地给定的话。不过，如果从我们一开始采取的概率化形式的角度看的话，这可能是一个意想不到的想法：$P_{(1)}$和$P_{(1|1)}$分别是一个独立的、给定的概率测量P的两个方面。那么关于它们是可以分别被指定的独立概率的观点看起来就不自然了。但是，当然并没有出现对该观点的正式反对意见，因为由遵守查普曼—柯尔莫哥洛夫方程的一系列转移概率T_t和一个任意初始概率分布$P_0(y)=P_{(1)}(y, 0)$结合起来可以定义一个唯一的齐次马尔可夫过程。事实上，有时人们甚至更进了一步，仅用转移概率来完全确定齐次马尔可夫过程，而不考虑初始状态$P_0(y)$，就像通常描述确定性系统的动力学时并不需要假设任何特定的初始状态一样。

对于马尔可夫链，确定$P_t(y)$的演化这个目标现在已经由方程(181)完全解决了。在连续时间的情况下，确定这种演化的任务一般通过微分方程来完成。这样的方程可以直接通过在很小的时间段内对转移概率做泰勒展开的方式得到[van Kampen, 1981, pp. 101—103]——根据一个适当的连续性假设。

(在符号方面有轻微变化的)结果是：

$$\frac{\partial P_t(y)}{\partial t}=\sum_{y'}\left(W(y\mid y')P_t(y')-W(y'\mid y)P_t(y)\right) \tag{182}$$

这里，$W(y\mid y')$是每单位时间内从y'到y的转移概率。这个微分方程，最早是由泡利(Pauli)在1928年得到的，它被称为主方程(这个名字已经变得通行起来，因为这种类型的方程包含了种类繁多的进程)。

对这个方程的解释是具有启发性的。概率$P_t(y)$的变化是通过在得失之间做出平衡来确定的：值y相对于y'的所有可能值的概率在时间dt内增大，因为转移是从y'到y的。每单位时间内的增加值为$\sum_{y'}W(y\mid y')P_t(y')$。但是，在同一段时间$dt$内，$P_t(y)$也会减小，这是由于发生了一系列从值$y$转移到其他所有可能值$y'$的事件的结果。这提供了第二个术语。

在这种“平衡”方面，主方程类似于玻耳兹曼方程式(48)，尽管它们的推导过程完全不同，而且$P_t(y)$与玻耳兹曼的$f_t(y)$相比有不同的意义。前者是一个概率分布，而后者是粒子的分布。但它们都是关于t的一阶微分方程。主方程与玻耳兹曼方程在数学上的一个关键区别是它在P上是线性的，因此也更容易求解。

事实上，主方程的任意解都可以形式化地表示为：

$$P_t = e^{tL}P_0 \tag{183}$$

这里 L 表示算符

$$L(y \mid y') := W(y \mid y') - \sum_{y''} W(y'' \mid y')\delta_{y,y'} \tag{184}$$

一般解式(183)类似于离散时间情况下的式(181)，因此它表明主方程等价于连续时间下关于齐次马尔可夫过程的假设。

最后的评论(与后面的讨论无关)。通过考虑关于粒子对的一个马尔可夫过程，即设想粒子对以一定的转移概率从初始状态(i, j)转移到状态(k, l)这样一个过程，我们可以多做一些与玻耳兹曼方程的类比，参见[Alberti and Uhlmann(阿尔贝蒂和乌尔曼)，1982，30]。令 $W(i, j \mid k, l)$表示单位时间内的转移概率。那么主方程的形式如下：

$$\frac{\partial P_t(i, j)}{\partial t} = \sum_{k,l} \left(W(i, j \mid k, l)P_t(k, l) - W(k, l \mid i, j)P_t(i, j)\right) \tag{185}$$

现在假设转移$(i, j) \to (k, l)$和$(k, l) \to (i, j)$是等概率的，从而使得单位时间内的转移概率是对称的，$W(i, j \mid k, l) = W(k, l \mid i, j)$，并且作为分子混沌假设的类似方程，右侧的 $P(i, j)$可由其各项积代替：

$$P(i, j) \to \sum_j P(i, j) \cdot \sum_i P(i, j) = P'(i)P''(j) \tag{186}$$

将上面的式(185)对 j 求和，我们最终得到：

$$\frac{\partial P_t'(i)}{\partial t} = \sum_j \frac{\partial P_t(i, j)}{\partial t} = \sum_{j,k,l} T(i, j \mid k, l)(P_t'(k)P_t''(l) - P_t'(i)P_t''(j)) \tag{187}$$

即一个更显著的玻耳兹曼方程式(48)的类似方程。但请注意，虽然式(185)描述的是一个马尔可夫过程，但是作为式(186)的结果，最后这个方程式(187)并不是这样的：在 P 中，它不再是线性的。

7.4 趋向于平衡及熵增加的趋势?

关于一个齐次马尔可夫过程的 $P_t(y)$的演化，我们能一般性地总结些什么呢？一个直接的结果就是：相对熵是单调非递减的。这就是说，如果我们将

$$H(P, Q) := -\sum_{y \in \mathcal{Y}} P(y)\ln\frac{P(y)}{Q(y)} \tag{188}$$

定义为概率分布 P 相对于 Q 的相对熵，那么我们可以证明，参见例如[Moran，1961；Mackey，1991，30]：

$$H(P_t,\ Q_t) \geqslant H(P,\ Q) \tag{189}$$

其中，$P_t = T_t P$，$Q_t = T_t Q$，而 T_t 是半群式(181)或式(183)中的元素。

我们也可以证明，一个非零相对熵至少对于某些概率分布 P 和 Q 来说是增加的，随机矩阵 T_t 一定是不可逆的。

在某种意义上，相对熵 $H(P \mid Q)$ 被认为可以用来测量 P 和 Q 的"相似"程度。①事实上，它会达到其最大值(即 0)，当且仅当 $P = Q$ 时；它会变为 $-\infty$，如果 P 和 Q 有不相交支集，即对所有 $y \in \mathcal{Y}$，$P(y)Q(y) = 0$。因此，式(189)表明，如果随机过程是不可逆的，那么随着时间的推移，一对概率分布 P_t 和 Q_t 通常会变得越来越相似。

因此，似乎在这个框架下，我们已经得到了"不可逆行为"的一个普遍的弱的特性。当然，上面的结果并没有暗示一个概率分布的"绝对"熵 $H(P) := -\sum_y P(y)\ln P(y)$ 是非递减的。但是，现在假设这个过程中有一个定态。换句话说，存在一个概率分布 $P^*(y)$，使得：

$$T_t P^* = P^* \tag{190}$$

很明显，我们的目的是将这样一个分布看作对平衡态的一种描述。如果有这样的稳定分布 P^*，我们可以应用前面的结果，并记作：

$$H(P,\ P^*) \leqslant H(T_t P,\ T_t P^*) = H(P_t,\ P^*) \tag{191}$$

换句话说，随着时间的推移，分布 $T_t P$ 将越来越类似于稳定分布而不是 P。如果稳定分布是均匀的，即：

$$P^*(y) = \frac{1}{m} \tag{192}$$

那么不仅相对熵增加，而且绝对熵 $H(P) := -\sum_y P(y)\ln P(y)$ 也会增加，因为：

$$H(P,\ P^*) = H(P) - \ln m \tag{193}$$

为了对"朝向平衡的趋势"做出一个令人满意的描述，我们还应考虑下面的问题。

---

① 当然，这是一种非对称性意义上的"相似性"，因为 $H(P,\ Q) \neq H(Q,\ P)$。

(i)存在这样一个稳定分布吗?

(ii)如果存在的话，它是唯一的吗?

(iii)$H(P_t)$的单调性意味着$\lim_{t\to\infty}P_t=P^*$吗? 更难的问题，我们放到下一小节7.5中讨论，它们分别是:

(iv)这种方法所需的假设的出现是由什么动机导致的，或者如何判断它们与底层的时间确定性动力学是(不)相容的;

(v)这种行为又是如何与马尔可夫过程的时间对称性相容的。

附录(i)。由式(190)所定义的定态，可以被看作是T_t的一个具有本征值1的本征矢量，或者按照式(183)，L是本征值为0的本征矢量。需要注意的是T或L不一定是厄米的(或者更确切地说，不一定是对称的，因为我们研究的是实矩阵)。因此，并不能由谱定理保证本征矢量是存在的。进一步讲，即使对应于本征值的本征矢量是存在的，它也不会自然就适合作为一个概率分布，因为它的组成元可能不是正的。

然而，事实证明，根据佩龙(Perron)(1907年)和弗罗宾尼斯(Frobenius)(1909年)提出的一个定理，每一个随机矩阵确实具有一个本征矢量，且其元都是非负的，其本征值为1，见[Gantmacher(甘德曼彻)，1959; van Harn and Holewijn(范·哈恩和霍尔维基)，1991]。但是，如果集合$\mathcal{Y}$是无限的或连续的，这并不一定总是正确的。

后一种情况的一个著名例子是所谓的维纳过程，通常用于描述布朗运动。它是由跃迁概率密度来描述的:

$$T_t(y\mid y')=\frac{1}{\sqrt{2\pi t}}\exp\frac{(y-y')^2}{2t},\quad y,\ y'\in\mathbb{R} \tag{194}$$

任意的初始概率密度ρ_0的演化可以被记作一个卷积:

$$\rho_t(y)=\int T_t(y\mid y')\rho_0(y')dy' \tag{195}$$

它随着时间的推移逐渐变低、变平滑且变宽，但不会接近任何稳定概率密度。因为对于每一个ρ_0的选择，它都是成立的，因此在这种情况中，并不存在稳定分布。

但是，将其看作一个严重的缺陷则是不合理的。事实上，在热力学中我们也会发现排放到自由空间中的一丝气体同样也会扩散，变得比以往任何时候都

更稀薄，但永远不会趋向于平衡态。只有存在于一个体积有限的容器内的系统会趋向于热力学平衡。

然而，对于在一定范围内具有有限测度的连续变量，可以保证一个稳定分布是存在的，如果概率密度 ρ_y 在所有时刻都是有界的，即 $\exists M \in \mathbb{R}$，使得 $\forall t$，$\rho_t \leqslant M$，参见[Mackey，1992，36]。

附录(ii)。是否稳定解不是唯一的这个问题在某种程度上更难解决。这个问题类似于遍历问题中的度量可递性(参见6.1节)。

一般说来，完全可能的情况是 Y 的范围$\mathcal{Y}$可以被分割成两个不相交的区域，比如说 A 和 B，即$\mathcal{Y}=A\cup B$，因此就不存在从 A 到 B 的转移，反之亦然(或发生这种转移的概率为零)。也就是说，随机演化 T_t可能具有下面的性质：

$$T_t(Y\in A \mid Y\in B)=T_t(Y\in B \mid Y\in A)=0 \tag{196}$$

换句话说，(用传统标记法重新进行表示)其矩阵可用下面的形式表示：

$$\begin{pmatrix} T_A & 0 \\ 0 & T_B \end{pmatrix} \tag{197}$$

那么这个矩阵被称为是(完全)约化的。在这种情况下，稳定分布一般不会是唯一的：如果 P_A^* 是在区域 A 上有支集的一个稳定分布，P_B^* 是在区域 B 上有支集的一个稳定分布，那么每一个凸组合

$$\alpha P_A^*(y)+(1-\alpha)P_B^*(y) \quad 其中\ 0\leqslant\alpha\leqslant 1 \tag{198}$$

也是稳定的。因此，为了得到一个唯一的稳定解，我们不得不假设一个类似于度量可递性的性质。这就是说，我们应该要求使式(196)成立的每种将$\mathcal{Y}$划分为不相交集合 A 和 B 的分割法，在 $P(A)=0$ 或 $P(B)=0$ 的意义上都是"平凡的"。

因此，人们可能会问，当且仅当转移概率 T_τ 不可约化时，稳定分布 P^* 就是唯一的吗？正如我们在6.1节中所看到的，在遍历问题中，得到的答案是肯定的(至少当我们假设 P^* 在微正则测量下是绝对连续的时候)。但在当前情况下，并不是这样！

这就必须考虑所谓的"瞬态"现象，它在哈密顿动力学中并没有对应物。让我们通过一个例子来引入这个概念。考虑一个如下形式的随机矩阵：

$$\begin{pmatrix} T_A & T' \\ 0 & T_B \end{pmatrix} \tag{199}$$

这里 T'是仅具有非负项的矩阵。那么：

$$\begin{pmatrix} T_A & T' \\ 0 & T_B \end{pmatrix}\begin{pmatrix} P_A \\ 0 \end{pmatrix} = \begin{pmatrix} T_A P_A \\ 0 \end{pmatrix}, \quad \begin{pmatrix} T_A & T' \\ 0 & T_B \end{pmatrix}\begin{pmatrix} 0 \\ P_B \end{pmatrix} = \begin{pmatrix} T' P_B \\ T_B P_B \end{pmatrix} \qquad (200)$$

因此，这里 $a \to b$ 类型的转移概率为零，但 $b \to a$ 类型的转移会以一个正概率发生，在这里，a，b 分别代表子集 A 和 B 中的任意元素。很显然在这种情况下，区域 B 最终将会被"抽空"。也就是说，元素位于区域 B 中的总概率（即 $\| T'P_B \|$）会指数性地减为零。在 B 中具有支集的分布被称为"瞬态"，集合 A 被称为"吸收集"或"俘获集"。

很显然，这种瞬态在测定稳定分布时将不会起到任何作用，而且如果以此为目的的话，它们可能会被简单地忽略掉。因此，在这个例子中，仅有的稳态是那些在 A 中有支集的态。而且如果 T_A是可约化的话，则不仅仅只有一个稳态。

从式(199)（通过行和列的排列）我们可以得到具有可约化 T_A的一个矩阵 T，被称为是不完全可约化的[van Kampen，1981，108]。此外，如果一个随机矩阵既不是完全可约化的也不是不完全可约化的，那么它就被称为是不可约化的。另一种（等价的）标准是所有状态相互之间是否能够"通信"，即对于每一对 i，$j \in \mathcal{Y}$，是否存在某些时刻 t，使得 $P_t(j \mid i) > 0$。

佩龙—弗罗宾尼斯定理保证只要 T 是不可约化的，那么就存在一个唯一的稳定分布。此外，我们还可以证明一个类似于遍历定理的理论[Petersen，1983，52]。

针对马尔可夫过程的遍历定理　如果转移概率 T_t是不可约化的，P_t的时间平均收敛于一个唯一的稳定解：

$$\lim_{\tau \to \infty} \frac{1}{\tau} \int_0^{\tau} T_t P(y)\, dt = P^*(y) \qquad (201)$$

附录(iii)。如果存在一个唯一的稳定分布 P^*，那么对于 P 的每一个选择，$T_t P$ 会收敛到 P^* 吗？同样，答案不完全是肯定的，即使式(201)有效！例如，存在齐次马尔可夫链和不可约化马尔可夫链，对于它们来说，P_t可分为两部分（$P_t = Q_t + R_t$），它们分别具有以下属性[Mackey，1992，71]。

1. Q_t 是一个具有 $\| Q_t \| \to 0$ 性质的项。这是一个瞬态项。

2. 剩余部分 R_t 是周期性的，即在某段有限时间 τ 之后，演化会重复：$R_{t+\tau}=R_\tau$。

这些过程被称为是渐近周期性的。它们经常与一个唯一的稳定分布 P^* 一起出现，并表现为严格的单调递增特征，但仍不会收敛到 P^*。在这种情况下，相对熵 $H(P_t, P^*)$ 单调增加完全是因为这个瞬态项。对于周期性的部分 R_t，转移概率是置换矩阵，它在经过 τ 次重复之后，就会返回到单位算符。

此外，如果我们约定 P^* 是相同的，那么我们就可以说(即使在这个例子中)：在周期 τ 的一个循环过程中得到的各种形式的 R_t，其相对熵 $H(R_t, P^*)$ 都具有相同的值，但熵严格低于 $H(P^*, P^*)=0$。事实上，P^* 是 R_t 的平均，即对应于式(201)，$P^*=\frac{1}{\tau}\sum_{t=1}^{t=\tau}R_t$。

进一步的技术性假设可以被引入到这样的例子中，并因此而严格收敛到唯一的稳定分布，例如通过强加一个"精确性"条件的方式[Mackey，1992]。但是，它偏离我们详细讨论的这个问题太远了。

总之，这似乎是"不可逆行为"较弱的方面，即相对熵的单调非递减性是齐次马尔可夫过程(事实上是所有随机过程)的一个基本特征，而当转移概率不可逆时，则不是一般性特征。在问题(i)、(ii)和(iii)的答案是肯定的这个意义上，可以得到这种行为的更强的版本，但代价是要给出额外的技术性假设。

7.5　提出马尔可夫性的动机及反对意义

附录(iv)。我们现在讨论下面的问题。做出马尔可夫性假设的动机是什么？当然，答案将依赖于对我们关于形式体系的解释，基于"粗粒化""开放系统"或干涉方法的不同，这个答案可能也不同(参见7.1节)。我将在下一节中讨论粗粒化方法，然后再从干涉的角度来考虑类似的问题。

粗粒化与重复随机性假设

在当前的观点来看，我们假设所考虑的系统确实是一个孤立的哈密顿系统，但马尔可夫性按照推测可以从相空间的一个分区中得到。但是，那究竟是如何实现的呢？

最清晰且最毫无保留地阐明这种观点的论文之一是[van Kampen，1962]。

就如在5.4节中那样，我们假设哈密顿相空间Γ中——或能量曲面Γ_E上——存在一些特殊的将其分为不相交单元格的分割法：$\Gamma=\omega_1\cup\cdots\cup\omega_m$。考虑在此相空间上概率密度为$\rho$的任意一个系综。其演化可以由算符表示为：

$$U_t^*\rho(x):=\rho(U_{-t}x) \tag{202}$$

在这里——为了避免混淆符号——我们用U_t表示哈密顿演化算符，先前我们将其表示为T_t，例如在式(68)中以及整个第6节中。将这种分割好的单元格之间的转移概率定义为：

$$T_t(j\mid i):=P(x_t\in\omega_j\mid x\in\omega_i)=P(U_tx\in\omega_j\mid x\in\omega_j)=\frac{\int_{(U_{-t}\omega_j)\cap\omega_i}\rho(x)dx}{\int_{\omega_i}\rho(x)dx} \tag{203}$$

显然，这样的转移概率将是均匀的，因为哈密顿演化U_t具有时间平移不变性。进一步，令$\hat{p}_0(i):=P(x\in\omega_i)=\int_{\omega_i}\rho(x)dx, i\in\mathcal{Y}=\{1,\cdots,m\}$为$t=0$时刻的任意初始粗粒化概率分布。

使用式(92)定义的粗粒化映射，我们也可以将t时刻的粗粒化分布表示为：

$$\mathcal{CG}U_t^*\rho(x)=\sum_{ji}T_t(j\mid i)\hat{p}_0(i)\frac{1}{\mu(\omega_j)}\mathbb{1}\,\omega_j(x) \tag{204}$$

这里μ是Γ上的正则测量，或Γ_E上的微正则测量。这个表达式表明，既然我们只对粗粒化的历史感兴趣，那么只要知道转移概率式(203)以及初始粗粒化分布就足够了。

但为了得到在前一节中所提出的问题的答案，我们需要证明转移概率可以定义一个马尔可夫过程，也就是说，它们服从查普曼—柯尔莫哥洛夫方程式(172)：

$$T_{t'+t}(k\mid i)=T_{t'}(k\mid j)T_t(j\mid i)\text{，对全部的 }t,\ t'>0 \tag{205}$$

在时刻t，t'和$t+t'$应用式(204)，很容易表明查普曼—柯尔莫哥洛夫方程等价于：

$$\mathcal{CG}U_{t'+t}^*=\mathcal{CG}U_{t'}^*\,\mathcal{CG}U_t^*\text{，对全部的 }t,\ t'>0 \tag{206}$$

换句话说，$t+t'$时刻的粗粒化概率分布可以由以下几步得到：首次在时间

段 t 内应用哈密顿动力学演化，然后执行一次粗粒化操作，接着在时间段 t' 内应用哈密顿动力学演化，然后再次粗粒化。对比于关系式 $U^*_{t'+t}=U^*_{t'}U^*_t$，我们可以看到，查普曼—柯尔莫哥洛夫条件可以通过满足粗粒化要求的条件的方式得到，即允许在演化的任意中间阶段，可以重新排列每个单元格中的相点。当然，这种在演化过程中的粗粒化会抹掉过去演化的所有信息，除了在那个时刻态所处于的单元格的标签以外，而且这也与将马尔可夫性看作无记忆特性的观点很好地结合了起来。

更重要的是，重复应用粗粒化会导致吉布斯熵的单调非递减性。为了简单起见，如果我们将时间间隔划分成 m 段的时间 τ，那么我们有：

$$\rho_{m\tau}=\underbrace{CGU^*_\tau\ CGU^*_\tau\cdots CGU^*_\tau}_{m\text{ times}}\,\rho \tag{207}$$

并且由式(96)：

$$\sigma[\rho_{m\tau}]\geqslant\sigma[\rho_{(m-1)\tau}]\geqslant\cdots\geqslant\sigma[\rho_\tau]\geqslant\sigma[\rho_0] \tag{208}$$

但是由于 τ 的选择是任意的，因此我们可以得出如下结论：$\sigma[\rho_t]$ 是单调非递减的。

因此，范·坎朋认为，动力学演化中增加的一个新部分是：在演化的任意阶段，我们应该应用一次对分布的粗粒化。重要的是需要注意只在一个单一瞬间这样做一次是不够的。在演化的每一个阶段，我们都需要一次又一次地对分布进行粗粒化。范·坎朋[1962，193]称其为重复随机性假设。

这一假设的理由是什么？范·坎朋指出，这并“不是没道理的”(同上，第182页)，因为事实是它在现象物理学中是成功的。热力学和对宏观系统的其他现象学描述(扩散方程、输送方程、流体力学和福克尔—普朗克方程等)都是用非常少的变量描述宏观系统的。这意味着，相比于微观相空间来说，它们的状态描述是非常粗糙的。但它们的演化方程具有自主性和确定性：宏观量的变化是由那些相同变量的瞬时值给出的。很显然，这些方程的成功表明精确的微观状态相比于这种粗糙的描述并不能增加任何有用的信息。同时，范·坎朋也承认，粗粒化过程显然并不总是成功的。我们很难建构一种从相空间到单元格的分割，对此马尔可夫性完全失效。

很显然，单元格的选择必须是“恰到好处的”[van Kampen，1962，183]。

但目前还没有明确的方法指出如何做到这一点。[van Kampen，1981，80]认为，如何找到正确的选择是“物理学家的艺术”，之所以称之为艺术是因为他或她在实践中能够通过把一般性原则及其独特的创造性结合起来而得到成功，但他们并不能提供一般性的准则。给出重复随机性假设的理由是，我们可以由它导出马尔可夫性，并且它成功地提供了一种对粗粒化分布演化的自洽的和确定性的描述，由此我们可以得到主方程。

值得注意的是，范·坎朋因此给出了关于如何选择单元格的“一般性”观点(参见5.4节)，即单元格的选择应该与我们有限的观测能力相对应。宏观变量的可观测性不足以保证重复随机性假设能成功。可以想象(在实践中也会出现)的是，从可观测量角度给出的特定分割法不会导致一个马尔可夫过程。在这种情况下，可观测变量的选择是根本不够的，我们必须加入一些其他的(不可观测的)量，直到我们(希望)得到一个完备集，即一个变量集合，该集合中的变量可自行描述其演化。其中一个例子就是自旋回波实验，(可观测到的)系统的总磁化强度不会提供一个适当的粗粒化描述。关于这一主题的进一步讨论，请参见[Blatt，1959；Ridderbos and Redhead，1998；Lavis，2004；Balian，2005]。

除了重复随机性假设适用于什么样的分割方式这个未解决的问题之外，还有一些反对意见是关于重复随机性假设本身的。范·坎朋实际上只是建议我们大胆地接受重复随机性假设，不要被它的不确定地位分散注意力，只要看到它是成功的就行。对于有些学者，比如在[Sklar，1993]中，将该假设看作“再随机化假设”，这使该问题表现得非常突出。他们要求对为什么使用这个假设给出一个理由，从而可以解释为什么该方法是成功的。事实上，即使[van Kampen，1981，80]也将这种成功描述为一个“奇迹般的事实!”。当然，如果给出的理由依赖于它的成功的话，则这种要求是不能得到满足的。但是，在我看来，这并不意味着那是一个无效的理由。

令许多学者反感的另一点原因似乎是，重复粗粒化操作在偏离由U_t规定的真正的动力学演化时看起来是“人为”添加的。因此，熵增加和朝向平衡态的趋势，看上去是下面这个事实的结果：我们反复地改动概率分布，以擦去过去的所有信息，同时又拒绝对此过程做出一个动力学解释。[Redhead，1995，31]将此过程描述为“在理论物理学中我遇到过的最具欺骗性的策略之一”，类似的

反对意见请参见[Blatt，1959][Sklar，1993]以及[Callender，1999]。

有人可能会问，重复随机性假设与动力学演化之间的对比必然会像范·坎朋与其批评者所认为的那样单调吗？毕竟，正如我们在6.2节中看到的，存在这样的一些动力学系统，它们在遍历层次结构中的等级是如此之高，以至于对于相空间的某些分区来说，它们具有伯努利性(参见6.2节)。由于马尔可夫性弱于伯努利性，我们可以推断也存在一些动力学系统，其粗粒化演化定义一个齐次马尔可夫过程。[1]因此，人们可能会振振有词地指出，作为其原因的马尔可夫性，或重复随机性假设，并不需要来自外部的那只抛弃信息的“手”所做出的超自然的干预，系统微观相空间中的一个足够复杂的确定性动力学自身就能完成这种工作。然而，遍历层次结构不具有的那些性质都依赖于一个给定的保测(measure-preserving)演化。因此，虽然一些动力学系统可能具有马尔可夫性，但它们只会导致稳定的马尔可夫过程。其保测动力学仍然意味着吉布斯熵保持不变。因此，式(208)的结果只有当所有不等式都取等号时才能得到。为了得到粗粒化的非平凡形式，我们确实应该暂时放弃保测动力学。

综上所述，虽然如何选择特殊的分区方式仍然是一个悬而未决的问题，但重复随机性假设和微观层面动力学的确定性特征之间并不冲突。然而，据我所知，假设式(206)对于统计力学中探讨的哈密顿系统是否能够成立，还是一个尚待解决的问题。

干涉或“开放系统”

随机动力学的另一种研究方法会涉及开放系统。这里的观点是，该系统持续地与环境发生相互作用，并且这导致了朝向平衡态的趋势。

事实上，一种无法抗拒的观点是，对于现实系统来说，完全孤立是一种不切实际的理想化。对于微观层面的演化，系统与环境相互作用的实际影响是巨大的。让我们回到[Borel，1914]的论文中来看一个众所周知的例子，他估算了如果将天狼星上的1克物质移动1厘米，其万有引力的变化会对地球上一个气

① 严格来说，这仅适用于离散动力系统。对于连续时间，例如哈密顿动力学，马尔可夫性只能通过在重复随机性假设中增加一个时间平滑过程来得到[Emch，1965]和[Emch and Liu，2001，pp. 484—486]。

缸中气体的微观演化造成什么影响。在正常条件下，影响是如此之大，以至于对于一个典型的气体分子，它可能决定了该分子是否会大致在 50 次间接碰撞之后，碰撞那个给定的分子。这就是说，在天狼星上是否移动了那一克物质，在大约 10^{-6} 秒之后，就会相当明显地体现在与其对应的微观动力学的演化上。换句话说，这样一个系统的力学演化对于针对初始态的扰动来说是极为敏感的，以至于即使最微小的环境变化也会对微观轨迹造成很大的变化。但我们无法控制环境的状态。是否能把这种不可逆行为看作是外界不可控微扰的结果呢?①

令$(x,\ y)$为一个总系统的状态，其中，像以前一样，$x\in\Gamma^{(s)}$表示对象系统的状态而$y\in\Gamma^{(e)}$表示环境的状态。我们假设，总系统由以下形式的一个哈密顿量支配：

$$H_{\mathrm{tot}}(x,\ y)=H_{(s)}+H_{(e)}+\lambda H_{\mathrm{int}}(x,\ y) \tag{209}$$

因此对于总系统构成的系综，其概率密度演化如下：

$$\rho_t(x,\ y)=U_t^*\rho_0(x,\ y)=\rho(U_{-t}(x,\ y)) \tag{210}$$

即它是一个具有时间对称性和确定性的保测演化。

在每个时刻，我们可以分别定义系统和环境的边缘分布为：

$$\rho_t^{(s)}(x)=\int dy\,\rho_t(x,\ y) \tag{211}$$

$$\rho_t^{(e)}(x)=\int dx\,\rho_t(x,\ y) \tag{212}$$

当然，我们更感兴趣的是对象系统，即式(211)。进一步假设在时刻 $t=0$ 时，总密度因子为：

$$\rho_0(x,\ y)=\rho_0^{(s)}(x)\rho_0^{(e)}(y) \tag{213}$$

关于$\rho_t^{(s)}(x)$的演化，我们能说些什么呢？它是形成了一个马尔可夫过程呢，还是它表明了熵增加呢？

一个直接的结果是这样的，参见例如 [Penrose and Percival(彭罗斯和珀西瓦尔)，1962]。

---

① 请注意，这里的术语“开放系统”是指一个与系统发生(弱)相互作用的系统。这应该区别于其他物理学分支中“开放系统”的概念，在那里它表示一个与其环境之间存在粒子交换的系统。

$$\sigma[\rho_t^{(s)}]+\sigma[\rho_t^{(e)}]\geqslant\sigma[\rho_0^{(s)}]+\sigma[\rho_0^{(e)}] \tag{214}$$

其中，σ 表示吉布斯精细化熵式(90)。该结果基于如下事实，即 $\sigma[\rho_t]$是守恒的，且联合概率分布的熵总是小于或等于它们边际熵的总和，如果联合分布可分解时等号成立。这给出了总系统熵变的一个形式，但是不足以得出对象系统自身将朝着平衡态演化这个结论，甚至得不出其熵会单调增加这一结论。请注意，式(214)在 $t\leqslant 0$ 时也成立。

其实，这显然是不能预期的。系统与环境之间存在的某种相互作用，可能会导致系统远离平衡态。关于这种情况我们将不得不给出额外的假设。通常这一系列的假设如下。

(a)环境的范围是非常大的(甚至是无穷大的)，即 $\Gamma^{(e)}$ 的维度远大于 $\Gamma^{(s)}$，且 $H_{(s)}\ll H_{(e)}$。

(b)系统和环境之间的耦合很弱，即 λ 非常小。

(c)环境最初处于热力学平衡态，例如 $\rho^{(e)}(y)$是正则的：

$$\rho_0^{(e)}=\frac{1}{Z(\beta)}e^{-\beta H^{(e)}} \tag{215}$$

(d)人们考虑的时间跨度相对于环境的弛豫时间很长，但相对于整个系统的庞加莱回归时间很短。

即使是这样，得到系统边缘态演化的主方程式(211)，或者表明其演化是由一个半群所导致的，仍是一项重要任务，它保证了这会形成一个马尔可夫过程。学者们详细地研究了许多具体模型，参见[Spohn，1980]。有学者得到一般性的定理(虽然大多数是建立在量子力学基础上的)，参见[Davies，1974；Davies，1976a；Lindblad，1976；Gorini *et al.*，1976]。但它们与以前的方法也有一点类似之处，那就是，似乎这里也一样必须引入类似“重复随机性”的假设[Mehra and Sudarshan，1972；van Kampen，1994；Maes and Netočný(梅斯和纳托茨)，2003]。

虽然我可能冒一些风险，即可能会过于简单化或歪曲这种分析的结果，但我相信我可以将它们概括如下。它们表明在所谓的“弱耦合”限制或其他类似的一些限制性过程下，时间的演化式(211)可被建模为：

$$\rho_t^{(s)}(x)=T_t\rho^{(s)}(x),\quad t\geqslant 0 \tag{216}$$

这里算符T_t构成一个半群，而环境仍处于其稳定平衡态：

$$\rho_t^{(e)}(y)=\rho_0^{(e)}(y),\quad t\geqslant 0 \tag{217}$$

这些结果的确立将允许我们从式(214)中推断出，系统的熵是单调非递减的。

为了评估这些研究成果，我们可以方便地定义，对于 $\rho_0^{(e)}$ 的一个确定选择，总系统概率分布的线性映射为：

$$\mathcal{TR}:\ \rho(x,\ y)\mapsto\mathcal{TR}\rho(x,\ y)=\int\rho(x,\ y)dy\cdot\rho_0(y) \tag{218}$$

这个映射去掉了系统和环境之间的相关性，并将环境的边缘分布重新表示为其初始平衡形式。

现在，不难看出，对全部的 t，$t'\geqslant 0$ 查普曼—柯尔莫哥洛夫方程(等价于半群性)可以表示为：

$$\mathcal{TR}U_{t+t'}^{*}=\mathcal{TR}U_{t'}^{*}\,\mathcal{TR}U_{t}^{*} \tag{219}$$

它是式(206)的类似式。

因此，在粗粒化和开放系统方法之间有一种很强的形式相似性。事实上，环境变量发挥的作用等价于粗粒化方法中单元格里内部坐标的作用。关于过去的确切微观信息在这里被转变为系统与环境之间相互关联的形式。现在通过假设在后面时段内，态可以有效地由式(213)那种形式的积替换，如果忽略系统对环境的反作用的话，此信息就会被去除。映射$\mathcal{CG}$和$\mathcal{TR}$都是线性和幂等映射，它们可以被视为中岛和萨瓦兹格(Nakajima and Zwanzig)的投影算符方法的特殊情况，这种方法所做的是更系统且更抽象的阐述，有时也被称为子动力学。

开放系统方法的支持者们，例如在[Morrison，1966；Redhead，1995]中，认为对比于粗粒化方法，现在的步骤是“客观的”。可以推测，这意味着存在一个假设的事实，关于这种相关性是否确实“输出到了环境中”。然而，这两种方法之间的相似性会使人怀疑，粗粒化方法遇到的任何问题在开放系统方法中都会转变为类似的问题。事实上，我们在上一节讨论的如何找到一个特殊分区方法的问题，在这里就变为了如何区分“系统”和“环境”的问题。毫无疑问，在实际应用中，这种区分也是任意的。

7.6 马尔可夫性能解释不可逆行为吗?

附录(v)。最后，我想讨论的可能是最具争议且最令人惊讶的问题：马尔

可夫性或重复随机性假设，是时间反演非不变性结果的原因吗？

我们已经看到，每一个不可逆齐次马尔可夫过程所表现出的“不可逆行为”，都是表现在下面这个意义上的，即不同的初始概率分布随着时间的推移往往会变得越来越相似。在特定技术性条件下，我们还能获得更强的结果，例如：朝向某个唯一平衡态的演化趋势，绝对熵的单调非递减性，等等。所有这些结果看起来很明显是时间不对称的。然而，我们也看到了马尔可夫性是明确地具有时间对称性的。那么该如何调和这个矛盾？

一开始，我们就应该注意到，人们经常断言马尔可夫性是得到时间不对称性结果的关键。例如，彭罗斯写道：

> ……远离平衡态系统的行为在时间反演下是不对称的，例如，热量总是从较热物体流向较冷物体，从来不会从较冷物体到较热物体。如果这种行为仅仅源于动力学的对称定律的话，那么这可能确实是一个悖论，因此，我们必须承认一个事实，即在理论变得足够丰富，可以解释非平衡态行为之前，我们必须在应用动力学对称定律时，增加一些额外的假设，它们的性质是非动力学的且具有时间反演不对称性。在目前的理论中，这个额外的假设就是马尔可夫假设。参见[Penrose，1970，41]。

在前面的章节中，我们已经质疑过马尔可夫性是“非动力学的”这个论断。但是，现在我们感兴趣的问题是能否假设马尔可夫性在时间反演下是不对称的。我们可以在该方法的拥护者的论文中找到许多类似的陈述，例如重复随机性假设是“为了得到不可逆方程，而不得不补充到统计力学中的附加成分”[van Kampen，1962，182]，或者马尔可夫过程的不可逆性提供了热力学行为的起源[Mackey，1992]。

但如果给定了马尔可夫性是明确地具有时间对称性的，那么结果怎么会是这样的呢？为了探讨这个问题，需要考虑另一个问题。假如时间为负，那么给定的概率分布 $P(y, 0)$ 是如何演化的呢？因此，重新从式(170)开始，现在让我们令 $t\leqslant 0$。我们仍然有：

$$P(y,\ t)=\sum_{y'} P(y,\ t \mid y',\ 0)P(y',\ 0) \tag{220}$$

这些条件概率 $P(y,\ t,\ \mid y',\ 0)$满足“时间反演下”的马尔可夫性式(174)，就是说增加对后来的值的的详细描述并不会反过来影响前期的值。因此，我们得到的 $t \leqslant t' \leqslant t''$：

$$P(y,\ t \mid y'',\ t'')=\sum_{y'} P(y,\ t \mid y',\ t')P(y',\ t' \mid y'',\ t'') \tag{221}$$

这是查普曼—柯尔莫哥洛夫方程对应的时间反演下的类似方程。

因此，我们也可以将负时间下的这些条件概率看作反向演化算符。如果我们能够假设它们在时间平移下不变，即它们只依赖于时间差 $\tau=t-t'$，那么我们可以将其记作：

$$S_\tau(y \mid y') := P(y,t \mid y,t'),\text{其中 } \tau=t-t' \leqslant 0 \tag{222}$$

并得到算符 S_τ 的一个二次半群，它满足：

$$S_{\tau+\tau'}=S_\tau \circ S_{\tau'} \quad \tau,\tau' \leqslant 0 \tag{223}$$

它造成朝向过去的随机演化。

此外，这些反向条件概率通过贝叶斯定理与正向条件概率产生了联系：

$$P_{(\mid\mid)}(y,\ t \mid y',\ t')=\frac{P_{(\mid\mid)}(y',\ t' \mid y,\ t)P(y,\ t)}{P(y',\ t')} \tag{224}$$

如果这个过程像以前一样是齐次的，那么这就变为：

$$P_{(\mid\mid)}(y,\ t \mid y',\ t')=\frac{T_{-\tau}(y' \mid y)P_t(y)}{P_{t'}(y')};\ \tau=t-t'<0 \tag{225}$$

对于 $t<t'$，矩阵 $P_{(\mid\mid)}(y,\ t \mid y',\ t')$总是会给出 T_t的正确的“逆”。也就是说：

$$\sum_{y'} P(y,\ t \mid y',\ t')(T_{t'-t}P_t)(y')=P_t(y) \tag{226}$$

首先，请注意式(225)不是 T_t矩阵的逆！事实上，式(225)的右边既取决于 P_t 和 $P_{t'}$，还取决于 T。即使矩阵的逆 $T^{(\text{inv})}$ 不存在，或者不是一个真正的随机矩阵，朝向过去的演化仍然会受贝叶斯反演，即转移概率式(225)的支配。

还要注意的是，如果正向转移概率是齐次的，而反向转移概率却不必然如此。例如，在式(225)中，如果我们将 t 和 t'分别平移 δ，那么我们会发现：

$$P(y,\ t+\delta \mid y',\ t'+\delta)=\frac{T_{-\tau}(y' \mid y)P(y,\ t+\delta)}{P(y',\ t'+\delta)}$$

这里，上式右边一般仍然取决于 δ。在初始分布自身是稳定的特殊情况下，

反向转移概率是齐次的，无论正向转移概率是否是齐次的。如果 $P(y, t)$ 是不稳定的，我们可能仍会得出同样的结论，只要非稳定性仅限于 $\mathcal{Y}$ 的这些元素 y 或 y'，即对所有 t，$T_t(y \mid y') = 0$。否则，这两个概念在逻辑上就成为独立的了。

这就产生了一个意想不到的新问题。通常情况下，相比于时间反演不变性，均匀性假设(或时间平移不变性)可以被看作在哲学领域内是无足轻重的。但在这里我们可以看到，针对具有正向转移概率的系统假设时间平移不变性，不等价于对具有反向转移概率的系统做出同样的不变性假设。如果我们认为，这两个假设中有一个是明显的，那应该怎么解释另一个是失败的呢？并且我们应该如何解释为什么我们认为其中一个是明显的，同时还可以避免受到类似[Price，1996]所提出的关于使用了"双重标准"的那类指责？

但是，熵会增加吗？我们在此之前已经看到对于每个可逆马尔可夫过程，分布 P 相对于平衡分布的相对熵会增加，并且分布的演化会趋向平衡(得出这个结论并不需要假设过程的均匀性)。但反向演化算符也形成了一个马尔可夫过程，对这个过程同样的结论完全成立。这看起来是矛盾的。如果 $T_t P_0 = P_t$，那么我们也有 $P_t = S_{-t} P_0$。P_t 的熵不可能同时既高于又低于 P_0 的熵！我们可以举一个例子来说明如何解决这个明显的问题，即 S 的稳定解与 T 的稳定解并不相同！

看下面的例子。考虑一个 $\mathcal{Y} = \{1, 2\}$ 的马尔可夫链，令：

$$T = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix} \tag{227}$$

选择一个初始分布 $P_0 = \begin{pmatrix} \alpha \\ 1-\alpha \end{pmatrix}$。经过一个步骤，我们已经得到：

$$TP = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix} \tag{228}$$

这也是(唯一的)稳定分布 P^*。后向转移概率由贝叶斯定理给出，我们很容易

发现：

$$S=\begin{pmatrix}\alpha & \alpha\\ 1-\alpha & 1-\alpha\end{pmatrix} \tag{229}$$

这个转移概率的稳定分布是：

$$\tilde{P}^{*}=\begin{pmatrix}\alpha\\ 1-\alpha\end{pmatrix} \tag{230}$$

这就是说，对于正向演化算符，变换：

$$\begin{pmatrix}\alpha\\ 1-\alpha\end{pmatrix}\xrightarrow{T}\begin{pmatrix}\frac{1}{2}\\ \frac{1}{2}\end{pmatrix} \tag{231}$$

是一个从非稳定初始分布到稳定分布的变换。相对熵会增加：$H(P_0, P^{*})\leqslant H(TP, P^{*})$。但是，对于反向演化，类似地：

$$\begin{pmatrix}\frac{1}{2}\\ \frac{1}{2}\end{pmatrix}\xrightarrow{S}\begin{pmatrix}\alpha\\ 1-\alpha\end{pmatrix} \tag{232}$$

表示从一个非稳定初始分布到稳定分布 $\tilde{P}^{*}$ 的演化，同样地，这里相对熵也增加：$H(P_1, \tilde{P}^{*})\leqslant H(P_0, \tilde{P}^{*})$。

造成不可逆马尔可夫过程具有一个内置的时间不对称性这种错觉的原因是(至少部分地是)由于，我们习惯于将 T_τ 看作一个关于自主选择分布 P_0 的确定的演化算符。当然，在物理学的其他问题中，我们对这种观点是非常熟悉的，在其中确定性演化算符普遍形成了一个群，并且在我们内心中，它们在正负时间下都可以使用。

事实上，这些算符一般没有逆这个事实似乎反映了这样一个观点，即马尔可夫过程没有任何记忆，并且在过程中会“丢失信息”，而这是导致不可逆行为的原因，具体体现在时间不对称的主方程、相对熵或绝对熵的增加以及朝向平衡的趋向上。但实际上，每一个马尔可夫过程除了包括一个前向转移概率系统之外，也包括一个后向转移概率系统，这再次形成了一个半群(当它们是齐次

的）。如果我们认为它们是给定的，那么我们就会得到我们之前得到的所有结论，不过现在是对负时间而言的。

我的结论是不可逆性既不是内置于马尔可夫性中的，也不是内置于转移概率的不可逆性中的；或者说既不是内置于重复随机性假设中①，也不是内置于主方程或半群的属性中。相反，之所以表现出不可逆行为是由于选择依赖于正向转移概率，而不是后向转移概率。之前伊登斯[2001]已经在普里高津及其同事的方案背景下得到了类似的结论。在这里我主要的关注点是，同样的结论也适用于更“主流”的方法，比如粗粒化或开放系统的方法。

7.7　随机过程的可逆性

为了不以破坏性的注释结束这一章，让我强调一下，我并没有断言从随机动力学推导出不可逆行为是不可能的。相反，我断言的是得到正向转移概率的理想特性的动机是不够的，我们还应该表明，这些特性对于反向转移概率是缺失的。

为了用非平衡态统计力学的方法对不可逆问题做一次较系统的讨论，我们首先应该合理地定义一个（不）可逆随机过程意味着什么，这种定义应该能够在哈密顿统计力学中准确地表述隐藏于其原先背景之下的直观判断。

看起来很普遍的一个一般性定义[Kelly，1979，5]是要求一个随机过程是可逆的，当且仅当对所有的 n 和 t_1，…，t_n 以及 τ：

$$P_{(n)}(y_1,\ t_1,;\ \cdots;\ y_n,\ t_n) = P_{(n)}(y_1,\ \tau - t_n;\ \cdots;\ y_n,\ \tau - t_n) \qquad (233)$$

一个仅限于马尔可夫过程的类似定义，请参见[Grimmett and Stirzaker（格里梅特和斯丁扎克），1982，219]。此定义的一个直接推论是如果单次概率 $P_{(1)}(y,\ t)$是稳定的，即它处于统计平衡，那么这个随机过程只能是可逆的。

① 在最近的工作中，范·坎朋承认，重复随机性假设本身并不会导致不可逆性：“这个重复随机性假设……在最初的时间间隔 Δt 就明显地假设了随机性，从而打破了时间对称性。对于这个假设，除了它是唯一可以做的假设，并且它有效之外，我们找不到任何逻辑上的证据。如果我们假设随机性出现在每个 Δt 的最后，那么扩散系数，黏度系数等的正负将会出现错误；如果我们假设随机性出现在中间，则不会表现出不可逆性。”[van Kampen，2002，475]给出最初的重点。

事实上，这个定义似乎完全消解了调和可逆性和不可逆性的行为这个问题。正如[Kelly，1979，19]在讨论埃伦菲斯特模型时指出的：“可逆性和熵增加现象之间没有冲突——可逆性是一个平衡态模型的属性，而熵增加是趋向于平衡态这个过程的属性。”

但显然，这种观点把这个问题想得太简单了，因此，这不是非平衡态统计力学的恰当定义。回想一下埃伦费斯特的狗蚤模型(见7.2节)最初提出的背景，当时他是在尝试说明从初始非平衡分布(例如，这样一个概率分布，即对应于所有跳蚤都寄居在同一只狗身上这种状态的概率为1)演化到接近于平衡分布的这种倾向是可以与一种随机的具有时间对称性的动力学相协调的。

如果我们想将对初始条件的探讨与对动力学的探讨完全分开，那么我们就希望提到一个仅与随机动力学相关的(不)可逆性概念，而它与初次分配是否稳定无关。

看起来满足这种直观判断的另一种定义如下，一个随机过程是可逆的，仅当对所有 y 和 y' 及 $t'>t$：

$$P_{(1|1)}(y,\ t\mid y',\ t')=P_{(1|1)}(y,\ t'\mid y',\ t) \tag{234}$$

在这种情况下，我们不能得出这一进程必然是稳定的这个结论，而事实上，埃伦费斯特的模型是可逆随机过程的一个例子。我相信这个定义把握住了这样一种直觉，即如果在某一时刻我们得到状态 y'，那么该时刻前面那个时刻状态为 y 的条件概率等于后面那个时刻状态为 y 的条件概率。

根据这项建议，如果我们想在随机动力学中发现不可逆行为或“时间之箭”等现象的“起源”的话，就必须以找到并且激发某些条件为基础，在那些条件下前向转移概率与后向转移概率是不同的，在式(234)不成立这个意义上。否则，不可逆行为在本质上是与初始条件有关的假设的一个推论，在理论上它是一个与那些可以由哈密顿动力学得到的结论相同的结果。

参考文献

[Albert, 2000] D. Z. Albert. *Time and Chance*. Cambridge, Mass.: Harvard University Press, 2000.

[Alberti and Uhlmann, 1982] P. M. Alberti amd A. Uhlmann. *Stochasticity and partial order. Doubly stochastic maps and unitary mixing.* Dordrecht: Reidel, 1982.

[Balescu, 1997] R. Balescu. *Statistical dynamics.* London: Imperial College Press, 1997.

[Balian, 2005] R. Balian. Information in statistical physics. *Studies In History and Philosophy of Modern Physics*, 36, 323-353, 2005.

[Batterman, 1990] R. W. Batterman. Irreversibility and statistical mechanics: A new approach? *Philosophy of Science*, 57, 395-419, 1990.

[Batterman, 1991] R. W. Batterman. Randomness and probability in dynamical theories: On the proposals of the Prigogine school. *Philosophy of Science*, 58, 241-263, 1991.

[Batterman, 2002] R. W. Batterman. *The devil in the details: asymptotic reasoning in explanation, reduction, and emergence* Oxford: Oxford University Press, 2002.

[Batterman, 2005] R. W. Batterman. Critical phenomena and breaking drops: Infinite idealizations in physics. *Studies in History and Philosophy of Modern Physics*, 36, 225-244, 2005.

[Becker, 1967] R. Becker. *Therory of heat* New York: Springer, 1967.

[Belot, 2006] G. Belot. This volume, chapter 2, 2006.

[Bennett, 2003] C. H. Bennett. Notes on Landauer's principle, reversible computation, and Maxwell's demon. *Studies In History and Philosophy of Modern Physics*, 34, 501-510, 2003.

[Berkovitz et al., 2006] Berkovitz, J., R. Frigg and F. Kronz. The ergodic hierarchy and Hamiltonian chaos. *Studies in History and Philosophy of Modern Physics*, 2006.

[Bertrand, 1889] J. Bertrand. *Calcul des Probabilités.* Paris: Gauthier-Villars. Reissued New York: Chelsea, 1889. no date.

[Bierhalter, 1994] G. Bierhalter. Von L. Boltzmann bis J. J. Thomson: die Versuche einer mechanischen Grundlegung der Thermodynamik (1866-1890). *Archive for History of Exact Sciences*, 44, 25-75, 1994.

[Birkhoff, 1931] G. D. Birkhoff. Proof of the ergodic theorem. *Proceedings of the National Academy of Sciences of the United States of America*, 17, 656-660, 1931.

[Bishop, 2004] R. C. Bishop. Nonequilibrium statistical mechanics Brussels-Austin style*Studies In History and Philosophy of Modern Physics*, 35, 1-30, 2004.

[Blatt, 1959] J. M. Blatt. An alternative approach to the ergodic problem. *Progress in Theoretical Physics*, 22, 745-756, 1959.

[Boltzmann, 1866] L. Boltzmann. Über die Mechanische Bedeutung des Zweiten Hauptsatzes der Wärmetheorie. *Wiener Berichte*, 53, 195-220 1866; in [Boltzmann, 1909] Vol. I, paper 2.

[Boltzmann, 1868a] L. Boltzmann. Studien über das Gleichgewicht der lebendigen Kraft zwischen bewegten materiellen Punkten. *Wiener Berichte*, 58, 517-560, 1868; in [Boltzmann, 1909] Vol. I, paper 5.

[Boltzmann, 1868b] L. Boltzmann. Lösung eines mechanischen Problems. *Wiener Berichte*, 58, 1035-1044, 1868. In [Boltzmann, 1909] Vol. I, paper 6.

[Boltzmann, 1871a] L. Boltzmann. über das Wärmegleichgewicht zwischen mehratomigen Gasmolekülen. *Wiener Berichte*, 63, 397-418, 1871. In [Boltzmann, 1909] Vol. I, paper 18.

[Boltzmann, 1871b] L. Boltzmann. Einige allgemeine Sätze über Wärmegleichgewicht. *Wiener Berichte*, 63, 679-711, 1871. In [Boltzmann, 1909] Vol. I, paper 19.

[Boltzmann, 1871c] L. Boltzmann. Analytischer Beweis des zweiten Haubtsatzes der mechanischen Wärmetheorie aus den Sätzen über das Gleichgewicht der lebendigen Kraft. *Wiener Berichte*, 63, 712-732, 1871. In [Boltzmann, 1909] Vol. I, paper 20.

[Boltzmann, 1872] L. Boltzmann. Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen. *Wiener Berichte*, 66, 275-370, 1872. In [Boltzmann, 1909] Vol. I, paper 23.

[Boltzmann, 1877a] L. Boltzmann. Bermerkungen über einige Probleme der mechanische Wärmetheorie *Wiener Berichte*, 75, 62-100, 1877. In [Boltzmann, 1909] Vol. II, paper 39.

[Boltzmann, 1877b] L. Boltzmann. über die Beziehung zwisschen dem zweiten Haubtsatze der mechanischen Wärmetheorie und der Wahrscheinlichkeitsrechnung resp. dem Sätzen über das Wärmegleichgewicht *Wiener Berichte*, 76, 373-435, 1877. In [Boltzmann, 1909] Vol. II, paper 42.

[Boltzmann, 1878] L. Boltzmann. Weitere Bemerkungen über einige Probleme der mechanischen Wärmetheorie. *Wiener Berichte*, 78, 7-46, 1878. In [Boltzmann, 1909] Vol. II paper 44.

[Boltzmann, 1881a] L. Boltzmann. über einige das Wärmegleichgewicht betreffende Sätze. *Wiener Berichte*, 84 136-145, 1881. In [Boltzmann, 1909] Vol. II paper 62.

[Boltzmann, 1881b] L. Boltzmann. Referat über die Abhandlung von J. C. Maxwell: "über Boltzmann's Theorem betreffend die mittlere verteilung der lebendige Kraft in einem System materieller Punkte". *Wiedemann's Annalen Beiblätter*, 5, 403-417, 1881. In [Boltzmann, 1909] Vol. II paper 63.

[Boltzmann, 1884] L. Boltzmann. über die Eigenschaften Monocyklischer und andere damit verwandter Systeme. *Crelle's Journal für die reine und angewandte Mathematik*, 98, 68-94 1884 and 1885. In [Boltzmann, 1909] Vol III, paper 73.

[Boltzmann, 1887a] L. Boltzmann. Ueber die mechanischen Analogien des zweiten Hauptsatzes der Thermodynamik. *Journal für die reine und angewandte Mathematik*, *100*, 201-212, 1887. Also in [Boltzmann, 1909], Vol. III, paper 78.

[Boltzmann, 1887b] L. Boltzmann. Neuer Beweis zweier Sätze über das Wärmegleichgewicht unter mehratomigen Gasmolekülen. *Wiener Berichte* 95, 153-164, 1887 in [Boltzmann, 1909] Vol. III, paper 83.

[Boltzmann, 1887c] L. Boltzmann. über einige Fragen der Kinetische Gastheorie. *Wiener Berichte* 96, 891-918, 1887. In [Boltzmann, 1909] Vol. III, paper 86.

[Boltzmann, 1892] L. Boltzmann. III. Teil der Studien über Gleichgewicht der lebendigen Kraft *Münch. Ber.* 22, 329-358, 1892. In [Boltzmann, 1909] Vol. III, paper 97.

[Boltzmann, 1894a] L. Boltzmann and C. H. Bryan. über die mechanische Analogie des Wärmegleichgewichtes zweier sich berührende Körper. *Wiener Berichte*, 103, 1125-1134, 1894. In [Boltzmann, 1909] Vol. III, paper 107.

[Boltzmann, 1894b] L. Boltzmann. On the application of the determinantal relation to the kinetic theory of gases. Appendix C to an article by G. H. Bryan on thermodynamics. *Reports of the British Association for the Advancement of Science*, pp. 102-106, 1894. In [Boltzmann, 1909] Vol. III, paper 108.

[Boltzmann, 1895] L. Boltzmann. On certain questions in the theory of gases. *Nature* 51, 413-415, 1895. In [Boltzmann, 1909], Vol. III, paper 112.

[Boltzmann, 1895b] L. Boltzmann. On the minimum theorem in the theory of gases. *Nature* 52, 221, 1895. Also in [Boltzmann, 1909] Vol. III, paper 114.

[Boltzmann, 1896] L. Boltzmann. *Vorlesungen über Gastheorie* Vol I. Leipzig, J. A. Barth, 1896. Translated, together with [Boltzmann, 1898] by S. G. Brush, *Lecture on Gas Theory* Berkeley: University of California Press, 1964.

[Boltzmann, 1896b] L. Boltzmann. Entgegnung an die wärmetheoretischen Betrachtungen des Hrn. E., Zermelo. *Wiedemann's Annalen*, 57, 772-784, 1896. In [Boltzmann, 1909] Vol. III, paper 119.

[Boltzmann, 1897a] L. Boltzmann. Zu Hrn Zermelos Abhandlung "über die mechanische Erklärung irreversibler Vorgänge". *Wiedemann's Annalen*, 60 392-398, 1897. in [Boltzmann, 1909], Vol. III paper 120.

[Boltzmann, 1897b] L. Boltzmann. über einige meiner weniger bekannte Abhandlungen über Gastheorie und deren Verhältnis zu derselben *Verhandlungen des* 69. *Versammlung Deutscher Naturforscher und Ärzte*, Braunschweig 1897, pp. 19-26. *Jahresberichte der Deutsche Mathematiker-Vereinigung*, 6, (1899), 130-138. Also in [Boltzmann, 1909] Vol. III, paper 123.

[Boltzmann, 1898] L. Boltzmann. *Vorlesungen über Gastheorie* Vol II. Leipzig, J. A. Barth, 1898. Translated, together with [Boltzmann, 1896] by S. G. Brush, *Lecture on Gas Theory* Berkeley: University of California Press, 1964.

[Boltzmann, 1898] L. Boltzmann. über die sogenannte *H*-Kurve. *Mathematische Annalen* 50, 325-332, 1898. In [Boltzmann, 1909] vol. III, paper 127.

[Boltzmann and Nabl, 1904] L. Boltzmann and J. Nabl. Kinetisch theorie der Materie. *Encyclopädie der Mathematischen Wisenschaften*, Vol V-1, pp. 493-557, 1904.

[Boltzmann, 1905] L. Boltzmann. *Populäre Schriften*. Leipzig: J. A. Barth, 1905. Re-issued Braunschweig: F. Vieweg, 1979.

[Boltzmann, 1909] L. Boltzmann. *Wissenschaftliche Abhandlungen* Vol. I, II, and III. F. Hasenöhrl (ed.) Leipzig, 1909. Reissued New York: Chelsea, 1969.

[Borel, 1914] E. Borel. *Le Hasard*. Paris: Alcan, 1914.

[Boyling, 1972] J. B. Boyling. An axiomatic approach to classical thermodynamics. *Proceedings of the Royal Society of London*, 329, 35-71, 1972.

[Bricmont, 1995] J. Bricmont. Science of chaos or chaos in science? *Physicalia*, 17, 159-208; also in P. R. Gross, N. Levitt and M. W. Lewis (Eds.), *The flight from science and reason*, New York: New York Academy of Sciences, pp. 131-176, 1995.

[Brown and Uffink, 2001] H. Brown and J. Uffink. (2001). The origins of time-asymmetry in thermodynamics: the minus first law. *Studies in History and Philosophy of Modern Physics*, 32, 525-538, 2001.

[Brush, 1966] S. G. Brush. *Kinetic theory* Vol. 1 and 2. Oxford, 1966: Pergamon.

[Brush, 1976] S. G. Brush. *The Kind of motion we call heat.* Amsterdam: North Holland, 1976.

[Brush, 1977] S. G. Brush. Statistical mechanics and the philosophy of science: some historical notes. In F. Supper & P. D. Asquith (Eds.) *PSA* 1976 *Proceedings of the Biennial meeting of the Philsopophy of Science Association* 1976, *Vol* 2, East Lansu = ing, Mich.: Philosophy of Science Association, pp. 551-584, 1976.

[Brush, 1999] S. G. Brush. Gadflies and geniuses in the history of gas theory. *Synthese* 119, 11-43. Also in [Brush, 2003], pp. 421-450, 1999.

[Brush, 2003] S. G. Brush. *The kinetic theory of gases.* London: Imperial College Press, 2003.

[Bryan, 1894] G. H. Bryan. Letter to the editor. *Nature*, 51, 175, 1894.

[Bryan, 1894] G. H. Bryan. The Assumption in Boltzmann's Minimum Theorem, *Nature*, 52, 29-30, 1894.

[Burbury, 1894a] S. H. Burbury. Boltzmann's minimum theorem. *Nature*, 51, 78-79, 1894.

[Burbury, 1894b] S. H. Burbury. The kinetic theory of gases. *Nature*, 51 175-176, 1894.

[Butterfield, 2006] J. N. Butterfield. This volume, chapter 1, 2006.

[Callender, 1999] C. Callender. Reducing thermodynamics to statistical mechanics: the case of entropy. *Journal of Philosophy*, 96, 348-373, 1999.

[Callender, 2001] C. Callender. Taking Thermodynamics Too Seriously. *Studies In History and Philosophy of Modern Physics*, 32, 539-553, 2001.

[Callender, 2004] C. Callender. Measures, explanations and the past: Should 'special' initial conditions be explained? *British Journal for Philosophy of Science*, 55, 195-217, 2004.

[Campisi, 2005] M. Campisi. On the mechanical foundations of thermodynamics: The generalized Helmholtz theorem *Studies In History and Philosophy of Modern Physics*, 36, 275-290, 2005.

[Casetti et al, 2003] L. Casetti, M. Pettini and E. G. D. Cohen. Phase Transitions and Topology Changes in Configuration Space. *Journal of Statistical Physics*, *111*, 1091-1123, 2003.

[Carathéodory, 1909] C. Carathéodory. Untersuchungen über die Grundlagen der Thermodynamik, *Mathematische Annalen*, 67, 355-386, 1909. Translation by J. Kestin, "Investigation into the foundations of thermodynamics" in [Kestin, 1976], pp. 229-256. This translation is not quite accurate.

[Carnot, 1824] S. Carnot. *Réflexions sur la puissance motrice du feu*. Paris: Bachelier, 1824. Re-edited and translated in E. Mendoza (Ed.) Reflections on the motive power of fire, New York: Dover, 1960.

[Cercignani, 1998] C. Cercignani. *Ludwig Boltzmann, the man who trusted atoms* Oxford: Oxford University Press, 1998.

[Cercignani *et al.*, 1994] C. Cercignani, R. Illner, and M. Pulvirenti. *The mathematical theory of dilute gases*. New York: Springer-Verlag, 1994.

[Chang, 2003] H. Chang. Preservative realism and its discontents: revisiting caloric. *Philosophy of Science*, 70, 902-912, 2003.

[Chang, 2004] H. Chang. *Inventing temperature. Measurement and scientific progress*. Oxford: Oxford University Press, 2004.

[Clausius, 1857] R. Clausius über die Art von Bewegung die wir Wärme nennen, *Poggendorff's Annalen*, *100*, 253-280, 1857. English translation in [Brush, 1966], Vol. 1. pp. 111-134.

[Clausius, 1862]R. Clausius. 'Ueber die Anwendung des Satzes von der Aequivalenz der Verwandlungen auf die innere Arbeit', *Viertelsjahrschrift der Züricher naturforschenden Gesellschaft*, 7, 48, 1862. Also in *Annalen der Physik und Chemie*, *116*: 73-112 (1862), English translation in *Philosophical Magazine*, 24, 81-97, 201-213, also in [Kestin, 1976], pp. 133-161.

[Clausius, 1865] R. Clausius. über verschiedene für die Anwendung bequeme Formen der Haubtgleichungen der mechanische Wärmetheorie, *Poggendorff's Annalen*, *100*, (253), 1865. English translation in [Kestin, 1976].

[Cohen, 1996] E. G. D. Cohen. Boltzmann and statistical mechanics, cond-mat/9608054v2, 1996.

[Compagner, 1989]A. Compagner. Thermodynamics as the continuum limit of statistical mechanics *American Journal of Physics*, 57, 106-117, 1989.

[Cornfeld *et al.*, 1982]I. P. Cornfeld, S. V. Fomin, and Ya. G. Sinai. *Ergodic theory* New York: Springer-Verlag, 1982.

[Culverwell, 1894]E. Culverwell. Dr. Watson's proof of Boltzmann's theorem on permanence of distributions *Nature*, *50*, 617, 1894.

[Culverwell, 1895]E. Culverwell. Boltzmann's minimum theorem *Nature*, *51*, 246, 1895.

[Curd, 1982] M. Curd. Popper on the direction of time. In R. Sexl and J. Blackmore (eds.), *Ludwig Boltzmann, internationale Tagung: anlässlich des 75. Jahrestages seines Todes, 5-8. September 1981: ausgewählte Abhandlungen*, pp. 263-303, 1982. Graz: Akademische Druckund Verlagsanstalt.

[Davies, 1974]E. B. Davies. Markovian master equations. *Communications in Mathematical Physics*, *39*, 91-110, 1974.

[Davies, 1976a]E. B. Davies. Markovian master equations II *Mathematische Annalen*, *219*, 147-158, 1976.

[Davies, 1976]E. B. Davies. *Quantum theory of open systems.* New York: Academic Press, 1976.

[Denbigh and Denbigh, 1985]K. G. Denbigh and J. Denbigh. *Entropy in relation to incom-*

plete knowledge. Cambridge: Cambridge University Press, 1985.

[Dias, 1994] P. M. C. Dias. Will someone say exactly what the *H*-theorem proves? A study of Burbury's Condition A and Maxwell's Proposition II. *Archive for History of Exact Sciences 46*, 341-366, 1994.

[Dugas, 1959] R. Dugas. *La théorie physique au sens de Boltzmann et ses prolongements modernes.* (Neuchâtel: Griffon), 1959.

[Earman, 2006] J. Earman. The past-hypothesis: not even false. *Studies in History and Philosophy of Modern Physics*, 2006.

[Earman and Rédei, 1996] J. Earman and M. Rédei. Why ergodic theory does not explain the success of equilibrium statistical mechanics. *British Journal for the Philosophy of Science*, *45*,63-78, 1996.

[Earman and Norton, 1998] J. Earman and J. D. Norton. Exorcist XIV: The wrath of Maxwell's demon. Part I. From Maxwell to Szilard. *Studies in History and Philosophy of Modern Physics*, *29*, 435-471, 1998.

[Earman and Norton, 1999] J. Earman and J. D. Norton. Exorcist XIV: The wrath of Maxwell's demon. Part II. From Szilard to Landauer and beyond *Studies in History and Philosophy of Modern Physics*, *30*, 1-40.

[Edens, 2001] B. Edens. Semigroups and symmetry: an investigation of Prigogine's theories. http: //philsci-archive. pitt. edu/archive/00000436/, 2001.

[Ehrenfest and Ehrenfest-Afanassjewa, 1912] P. Ehrenfest and T. Ehrenfest-Afanassjewa. Begriffliche Grundlagen der Statistischen Auffassung in der Mechanik. *Enzyclopädie der Mathe-matischen Wissenschaften* Vol. 4. F. Klein and C. Müller (eds.). Leibzig: Teubner, pp. 3-90, 1912. English translation *The conceptual foundations of the statistical approach in mechanics.* Ithaca N. Y.: Cornell University Press, 1959.

[Einstein, 1902] A. Einstein. Kinetische Theorie des Wärmegleichgewicht und des zweiten Hauptstzes der Thermodynamik. *Annalen der Physik*, *9*, 417-433, 1902.

[Einstein, 1910] A. Einstein. Theorie der Opaleszenz von homogenen Flüssigkeiten und Flüssigkeitsgemischen in der Nähe des kritischen Zustandes *Annalen der Physik*, *33*, 1275-

1298, 1910.

[Ellis, 1850] R. L. Ellis. Remarks on an alleged proof of the method of least squares, contained in a late number of the Edinburgh Review, in W. Walton (Ed.), Mathematical and otherWritings of R. L. Ellis. Cambridge: Cambridge University Press, 1863, pp. 53-61.

[Emch, 1965] G. G. Emch. On the Markov character of master-equations. *Helvetica Physica Acta*, *38*, 164-171, 1965.

[Emch, 2005] G. G. Emch. Probabilistic issues in statistical mechanics. *Studies In History and Philosophy of Modern Physics 36*, 303-322, 2005.

[Emch, 2006] G. G. Emch. This volume, chapter 10, 2006.

[Emch and Liu, 2001] G. G. Emch and C. Liu, *The Logic of thermostatistical physics.* Berlin: Springer, 2001.

[Farquhar, 1964] I. E. Farquhar. *Ergodic theory in statistical mechanics*, London, Interscience, 1964.

[Feshbach, 1987] H. Feshbach. Small systems: when does thermodynamics apply? *Physics Today*, *40*, 9-11, 1987.

[Fine, 1973] T. L. Fine. *Theories of probability: An examination of foundations.* New York: Academic Press, 1973.

[Fisher, 1964] M. E. Fisher. The free energy of a macroscopic system. *Archive for Rational Mechanics and Analysis 17*, 377-410, 1964.

[Fowler and Guggenheim, 1939] R. H. Fowler and E. Guggenheim. *Statistical Thermodynamics.* Cambridge: Cambridge University Press, 1939.

[Fox, 1971] R. Fox. *The caloric theory of gases: from Lavoisier to Regnault.* Oxford: Clarendon Press, 1971.

[Friedman, 1976] K. S. Friedman. A Partial vindication of ergodic theory. *Philosophy of Science*, *43*, 151-162, 1976.

[Frigg, 2004] R. Frigg. In what sense is the Kolmogorov-Sinai entropy a measure for chaotic behaviour? —bridging the gap between dynamical systems theory and communication theory. *British Journal for the Philosophy of Science*, *55*, 411-434, 2004.

[Galavotti, 2004] M. C. Galavotti. *A philosophical introduction to probability.* Chicago: University of Chicago Press, 2004.

[Gallavotti, 1994] G. Gallavotti. Ergodicity, ensembles, irreversibility in Boltzmann and beyond http: arXiv: chao-dyn/9403004. *Journal of Statistical Physics*, *78*, 1571-1589, 1994.

[Gallavotti, 1999] G. Gallavotti. *Statistical mechanics: A short treatise.* Berlin: Springer, 1999.

[Gantmacher, 1959] F. R. Gantmacher. *Matrizenrechnung* Vol 2. Berlin: Deutscher Verlag der Wissenschaften, 1959.

[Garber *et al.*, 1986] E. Garber, S. G. Brush, and C. W. F. Everitt, (Eds.). *Maxwell on molecules and gases* Cambridge Mass.: MIT Press, 1986.

[Garber *et al.*, 1995] E. Garber, S. G. Brush, and C. W. F. Everitt, (Eds.). *Maxwell on heat and statistical mechanics* Bethlehem: Lehigh University Press, 1995.

[Gearhart, 1990] C. A. Gearhart. Einstein before 1905: The early papers on statistical mechanics *American Journal of Physics*, *58*, 468-480, 1990.

[Gibbs, 1875] J. W. Gibbs. equilibrium of heterogenous substances. *Transactions of the Connecticut Academy*, *3*, 103-246 and 343-524 (1878). Also in [Gibbs, 1906, pp. 55-353].

[Gibbs, 1902] J. W. Gibbs. *Elementary principles in statistical mechanics*, New York, Scribner etc, 1902.

[Gibbs, 1906] J. W. Gibbs. *The Scientific Papers of J. Willard Gibbs*, Vol. 1, Thermodynamics, Longmans, London, 1906. Reissued by Ox Bow Press, Woodbridge, Connecticut, 1993.

[Giles, 1964] R. Giles. *Mathematical foundations of thermodynamics* Oxford: Pergamon, 1964.

[Gold, 1956] T. Gold. Cosmic processes and the nature of time. In R. Colodny (Ed.), *Mind and Cosmos.* Pittsburgh: University of Pittsburgh Press, pp. 311-329, 1956.

[Goldstein, 2001] S. Goldstein. Boltzmann's approach to statistical mechanics. In J. Bricmont, D. Dürr, M. C. Galavotti, G. Ghirardi, F. Petruccione, and N. Zanghi (Eds.) *Chance in physics: foundations and perspectives*, Lecture Notes in Physics 574. Berlin:

Springer-Verlag, pp. 39-54, 2001. Also as e-print cond-mat/0105242.

[Gorini et al., 1976] V. Gorini, A. Kossakowski, and E. C. G. Sudarshan. Completely positive dynamical semigroups of *N*-level systems. *Journal of Mathematical Physics*, *17*, 8721-8825, 1976.

[Grimmett and Stirzaker, 1982] G. R. Grimmett and D. R. Stirzaker. *Probability and random processes.* Oxford: Clarendon Press, 1982.

[de Groot, 1951] S. de Groot. *Thermodynamics of irreversible processes.* Amsterdam: North-Holland, 1951.

[de Groot and Mazur, 1961] S. R. de Groot and P. Mazur. *Non-equilibrium thermodynamics.* Amsterdam: North-Holland, 1961.

[Gross, 1997] D. H. E. Gross. Microcanonical thermodynamics and statistical fragmentation of dissipative systems. The topological structure of the N-body phase space. *Physics Reports*, *279*, 119-201, 1997.

[Gross and Votyakov, 2000] D. H. E. Gross and E. V. Votyakov. Phase transitions in "small" systems. *European Physical Journal B*, *15*, 115-126, 2000.

[Grünbaum, 1973] A. Grünbaum. Is the coarse-grained entropy of classical statistical mechanics an anthropomorphism? In *Philosophical problems of space and time*, 2nd ed., Dordrecht: Reidel, pp. 646-665, 1973.

[Guthrie, 1874] F. Guthrie. Molecular motion. *Nature*, *10*, 123, 19 =874. Also in [Garber *et al.*, 1995, p. 143-145].

[Greven *et al.*, 2003] A. Greven, G. Keller, and G. Warnecke, (Eds.) *Entropy.* Princeton: Princeton University Press, 2003.

[ter Haar, 1955] D. ter Haar. Foundations of statistical mechanics. *Reviews of Modern Physics*, *27*, 289-338, 1955.

[Hacking, 1975] I. Hacking. *The emergence of probability.* Cambridge: Cambridge University Press, 1975.

[Hacking, 1990] I. Hacking. *The taming of Chance.* Cambridge: Cambridhge University Press, 1990.

[Van Harn and Holewijn, 1991] K. van Harn and P. J. Holewijn. *Markov-ketens in diskrete tijd.* Utrecht: Epsilon, 1991.

[Herschel, 1850] J. F. W. Herschel. Quetelet on Probabilities*Edinburgh Review.* Also in *Essays from the Edinburgh and Quarterly Reviews with addresses and other pieces*, London: Longman, Brown, Green, Longmans and Roberts, 1857, pp. 365-465, 1850.

[Hertz, 1910] P. Hertz. über die mechanischen Grundlagen der Thermodynamik. *Annalen der Physik*, 33, 225-274; 537-552, 1910.

[Hill, 1987] T. L. Hill. *Thermodynamics of small systems.* New York: Benjamin, 1987.

[Höflechner, 1994] W. Höflechner. *Ludwig Boltzmann Leben und Briefe.* Graz: Akademische Druck- und Verlagsanstalt, 1994.

[Huang, 1987] K. Huang. *Statistical mechanics.* New York: Wiley, 1987.

[Huggett, 1999] N. Huggett. Atomic metaphysics. *Journal of Philosophy*, *96*, 5-24, 1999.

[Hutter and Wang, 2003] K. Hutter and Y. Wang. Phenomenological thermodynamics and entropy principles, 2002. In [Greven et al., 2003, pp. 55-78].

[Illner and Neunzert, 1987] R. Illner and H. Neunzert. The concept of irreversibility in kinetic theory of gases. *Transport Theory and Statistical Physics*, *16*, 89-112, 1987.

[Janssen, 2002] M. Janssen. Dog fleas and tree trunks: the Ehrenfests marking the territory of Boltzmann's *H*-theorem. Unpublished, 2002.

[Jauch, 1972] J. Jauch. On a new foundation of equilibrium thermodynamics. *Foundations of Physics*, *2*, 327-332, 1972.

[Jauch, 1975] J. Jauch. Analytical thermodynamics. Part 1. Thermostatics-general theory. *Foundations of Physics*, *5*, 111-132, 1975.

[Jaynes, 1965] E. T. Jaynes. Gibbs vs. Boltzmann entropies. *American Journal of Physics*, 33, 391-398, 1965. Also in [Jaynes, 1983, pp. 77-86].

[Jaynes, 1967] E. T. Jaynes. Foundations of probability theory and statistical mechanics. In: M. Bunge (Ed.) *Delaware seminar in the foundations of physics.* Berlin: Springer-Verlag, pp. 77-101, 1967. Also in [Jaynes, 1983, pp. 89-113].

[Jaynes, 1983] E. T. Jaynes. *Probability, statistics and statistical physics.* R. Rosenkrantz

(Ed.) Dordrecht: Reidel, 1983.

[Jaynes, 1992] E. T. Jaynes. The Gibbs paradox. In C. R. Smith, G. J. Erickson, & P. O. Neudorfer, (Eds.) *Maximum entropy and Bayesian methods.* Dordrecht: Kluwer, pp. 1-22, 1992.

[Jepps *et al.* , 2000] O. G. Jepps, G. Ayton, and D. J. Evans. Microscopic expressions for the thermodynamic temperature. *Physical Review E*, *62*, 4757-4763, 2000.

[Kadanoff, 2000] L. P. Kadanoff. *Statistical physics : statics, dynamics and renormalization* Singapore: World Scientific, 2000.

[van Kampen, 1962] N. G. van Kampen. Fundamental problems in the statistical mechanics of irreversible processes. In E. G. D. cohen (Ed.), *Fundamental problems in statistical mechanics* Amsterdam: North-Holland, pp. 173-202, 1962.

[van Kampen, 1981] N. G. van Kampen. *Stochastic processes in chemistry and physics.* Amsterdam: North-Holland, 1981.

[van Kampen, 1984] N. G. van Kampen. The Gibbs paradox. In W. A. Parry (ed.), *Essays in theoretical physics in honour of Dirk ter Haar*, Oxford: Pergamon Press, pp. 303-312, 1984.

[van Kampen, 1994] N. G. van Kampen. Models for dissitpation in quantumn mechanics. In J. J. Brey, J. Marro, J. M. Rubi, M. San Miguel (Eds.) 25 *years of non-equilibrium statistical mechanics* Berlin: Springer, 1994.

[van Kampen, 2002] N. G. van Kampen. The road from molecules to Onsager. *Journal of Statistical Physics*, *109*, 471-481, 2002.

[van Kampen, 2004] N. G. van Kampen. A new approach to noise in quantum mechanics*Journal of Statistical Physics*, *115*, 1057-1072, 2004.

[Karakostas, 1996] V. Karakostas. On the Brussels school's arrow of time in quantum theory. *Philosophy of Science*, *63*, 374-400, 1996.

[Kelly, 1979] F. P. Kelly. *Reversibility and stochastic networks* Chichester: Wiley, 1979. Also at http: //www. statslab. cam. ac. uk/ afrb2/kelly book. html.

[Kestin, 1976] J. Kestin. *The second law of thermodynamics.* Stroudsburg, Pennsylvania:

Dowden, Hutchinson and Ross, 1976.

[Keynes, 1921] J. M. Keynes. *A treatise on probability.* London: Macmillan, 1921.

[Khinchin, 1949] A. Khinchin. *Mathematical foundations of statistical mechanics.* (New York: Dover), 1949.

[Kirchhoff, 1894] G. Kirchhoff. *Vorlesungen über mathematische Physik, Vol IV: Theorie der Wärme* M. Planck (Ed.). Leipzig: Teubner, 1894.

[Kirsten and Körber, 1975] C. Kirsten and H. -G. Körber. *Physiker über Physiker*, Berlin: Akademie-Verlag, 1975.

[Klein, 1970] M. J. Klein. Maxwell, his demon and the second law of thermodynamics, *American Scientist*, *58*, 82-95, 1970; also in [Leff and Rex, 1987, pp. 59-72].

[Klein, 1972] M. J. Klein. Mechanical explanation at the end of the nineteenth century. *Centaurus*, *17*, 58-82, 1972.

[Klein, 1973] M. J. Klein. The Development of Boltzmann's Statistical Ideas. In E. G. D. Cohen & W. Thirring (eds.), *The Boltzmann equation*, Wien: Springer, pp. 53-106, 1973.

[Klein, 1974] M. J. Klein. Boltzmann, monocycles and mechanical explanation. In R. J. Seeger & R. S. Cohen (Eds.), *Philsophical foundations of science*; *Boston Studies in the Philosophy of Science XI* Dordrecht: Reidel, pp. 155-175, 1974.

[Klein, 1978] M. J. Klein. The thermostatics of J. Willard Gibbs: atyransformation of thermoidynamics. In E. G. Forbes (Ed.) *Human implications of scientific advance.* Edinburgh: Edinburgh University Press, pp. 314-330, 1978.

[Kroes, 1985] P. A. Kroes. *Time: its structure and role in physical theories.* Dordrecht: Reidel, 1985.

[Kurth, 1960] R. Kurth. Axiomatics of classical statistical mechanics. Oxford: Pergamon Press, 1960.

[Ladyman *et al.*, 2006] J. Ladyman, S. Presnell, T. Short, and B. Groisman. The connection between logical and thermodynamical irreversibility. http://philsciarchive. pitt. edu/archive/00002374/.

[Landau and Lifshitz, 1987] L. E. Landau and E. M. Lifshitz. *Fluid Mechanics* 3rd Edition, Oxford: Butterworth-Heinemann, 1987.

[Lanford, 1973] O. E. Lanford. Entropy and equilibrium states in classical statistical mechancs. In A. Lenard (Ed.) *Statistical mechanics and mathematical problems*. Berlin: Springer-Verlag, pp. 1-113.

[Lanford, 1975] O. E. Lanford. Time evolution of large classical systems. In J. Moser (Ed.) *Dynamical Systems, Theory and Applications*, Lecture Notes in Theoretical Physics Vol. 38, Berlin: Springer, pp. 1-111, 1975.

[Lanford, 1976] O. E. Lanford. On a derivation of the Boltzmann equation. *Astérisque*, 40 117-137, 1976. Also in J. L Lebowitz & E. W. Montroll (Eds.) *Nonequilibrium phenomena I: the Boltzmann Equation*. Amsterdam: North-Holland, 1983.

[Lanford, 1981] O. E. Lanford. The hard sphere gas in the Boltzmann-Grad limit. *Physica*, 106*A*, 70-76, 1981.

[Lavis, 2004] D. A. Lavis. The spin-echo system reconsidered. *Foundations of Physics*, *34*, 669-688, 2004.

[Lavis, 2005] D. A. Lavis. Boltzmann and Gibbs: An attempted reconciliation. *Studies In History and Philosophy of Modern Physics*, *36*, 245-273, 2005.

[Lebowitz, 1994] J. Lebowitz. Time's arrow and Boltzmann's entropy. In J. J. Halliwell, J. Pérez-Mercader & W. H. Zurek (Eds.). *Physical origins of time asymmetry*. Cambridge: Cambridge University Press, pp. 131-146, 1994.

[Lebowitz, 1999] J. L. Lebowitz. Statistical mechanics: A selective review of two central issues. *Reviews of Modern Physics*, 71, S346-S357, 1999. math-ph/0010018.

[Lebowitz and Penrose, 1973] J. Lebowitz and O. Penrose. Modern ergodic theory. *Physics Today*, *26*, 23-29, 1973.

[Leeds, 1989] S. Leeds. Discussion: D. Malament and S. Zabell on Gibbs phase averaging. *Philosophy of Science*, *56*, 325-340, 1989.

[Leff and Rex, 1987] H. S. Leff and A. F. Rex. *Maxwell's Demon* 2. Bristol: Institute of Physics, 1987.

[Leinaas and Myrheim, 1977] J. M. Leinaas and J. Myrheim. On the theory of identical particles. *Il Nuovo Cimento*, *37B*, 1-23, 1977.

[Lieb, 1976] E. H. Lieb. The stability of matter. *Reviews of Modern Physics*, *48*, 553-569, 1976.

[Lieb and Lebowitz, 1973] E. H. Lieb and J. L. Lebowitz. Lectures on the thermodynamic limit for Coulomb systems. In A. Lenard (Ed.) *Statistical mechanics and mathematical problems*. Berlin: Springer-Verlag, pp. 136-162, 1973.

[Lieb and Yngvason, 1999] E. Lieb and J. Yngvason. The phyics and mathematics of the second law of thermodynamics, *Physics Reports*, *310*, 1-96; erratum 314, (1999), p. 669, 1999. Also as e-print: http://xxx.lanl.gov/abs/cond-mat/9708200.

[Lindblad, 1976] G. Lindblad. On the generators of quantum dynamical semigroups. *Communications in Mathematical Physics*, *48*, 119-130, 1976.

[Lindblad, 1983] G. Lindblad. *Non-equilibrium entropy and irreversibility* Dordrecht: Reidel, 1983.

[van Lith, 2001a] J. van Lith. Ergodic theory, interpretations of probability and the foundations of statistical mechanics *Studies in History and Philosophy of Modern Physics*, *32*, 581-594, 2001.

[van Lith, 2001b] J. van Lith. *Stirr in Stillness*. Ph. D. Thesis, Utrecht University, 2001.

[Liu, 1999] C. Liu. Explaining the emergence of cooperative phenomena. *Philosophy of Science* 66 (*Proceedings*), S92-S106, 1999.

[Lorentz, 1887] H. A. Lorentz. Über das Gleichgewicht der lebendige Kraft unter Gasmolekülen. *Wiener Berichte*, *95*, 115-152, 187, 1887. Also in: *Collected Papers* The Hague: Martinus Nijhoff 1938, pp. 74-111.

[Lorentz, 1916] H. A. Lorentz. *Les theories statistiques en thermodynamique*. Leipzig: Teubner, 1916.

[Loschmidt, 1876] J. Loschmidt. über die Zustand des Wärmegleichgewichtes eines Systems von Körpern mit Rücksicht auf die Schwerkraft. *Wiener Berichte*, *73*, 128-142, 366-372 (1876), 75, 287-298, 76, 209-225 (1877).

[Maes and Netočný, 2003] C. Maes and K. Netočný. Time-reversal and entropy. *Journal of*

Statistical Physics, *110*, 269-310, 2003.

[Mackey, 1992]M. C. Mackey. *Time's arrow: the origins of thermodynamic behavior.* New-York: Springer, 1992.

[Mackey, 2001]M. C. Mackey. Microscopic dynamics and the second law of thermodynamics. In: *Time's Arrows, quantum Measurements and Superluminal Behavior.* C. Mugnai, A. Ranfagni & L. S. Schulman (Eds.) (Roma: Consiglio Nazionale delle Ricerche), 2001.

[Malament and Zabell, 1980]D. B. Malament and S. L. Zabell. Why Gibbs Phase Averages Work—The Role of Ergodic Theory. *Philosophy of Science*, *56*, 339-349, 1980.

[Mandelbrot, 1956]B. Mandelbrot. An outline of a purely phenomenological theory of statistical thermodynamics: 1. canonical ensembles. *IRE Transactions on Information Theory*, *IT-2*, 190-203, 1956.

[Mandelbrot, 1962]B. Mandelbrot. The role of sufficiency and of estimation in thermodynamics. *Annals of Mathematical Statistics*, *33*, 1021-1038, 1962.

[Mandelbrot, 1964]B. Mandelbrot. On the derivation of statistical thermodynamics from purely phenomenological principles. *Journal of Mathematical Physics*, *5*, 164-171. 1964.

[Mãné, 1987]R. Mãné. *Ergodic theory and differentiable dynamics.* Berlin: Springer, 1987.

[Maroney, 2005]O. J. E. Maroney. The (absence of a) relationship between thermodynamic and logical reversibility. *Studies In History and Philosophy of Modern Physics*, 36, 355-374, 2005.

[Martin-Löf, 1979]A. Martin-Löf. *Statistical mechanics and the foundations of thermodynamics.* Berlin: Springer-Verlag, 1979.

[Maxwell, 1860]J. C. Maxwell. Illustrations of the dynamical theory of gases. *Philosophical Magazine*, *19*, 19-32; 20, 21-37, 19 =860. Also in [Garber *et al.*, 1986, pp. 285-318].

[Maxwell, 1867]J. C. Maxwell. On the dynamical theory of gases. *Philosophical Transactions of the Royal Society of London*, *157*, 49-88 (1867). Also in [Brush, 1966]and [Garber *et al.*, 1986, pp. 419-472]

[Maxwell, 1872]J. C. Maxwell. *Theory of heat.* (2nd ed.) London: Longmans, Green and Co, 1872.

[Maxwell, 1873] J. C. Maxwell. Molecules. *Nature*, *8*, 437-441 (1873). Also in [Garber *et al.*, 1986, pp. 138-154]

[Maxwell, 1873] J. C. Maxwell. On the final state of a system of molecules in motion subject to forces of any kind. *Nature*, *8*, 537-538 (1867), 1873. Also in [Garber *et al.*, 1995, pp. 138-143]

[Maxwell, 1875] J. C. Maxwell. On the dynamical evidence of the molecular constitution of bodies. *Nature*, *11*, 357-359, 374-377, (1875). Also in [Garber *et al.*, 1986, pp. 216-237].

[Maxwell, 1877] J. C. Maxwell. A treatise on the kinetic theory of gases by Henry William Watson, *Nature*, *18*, 242-246, 1877. Also in [Garber *et al.*, 1995, pp. 156-167].

[Maxwell, 1878a] J. C. Maxwell. Constitution of bodies. *Encyclopaedia Brittanica*, ninth edition, Vol. 6, pp. 310-313, 1878. Also in [Garber *et al.*, 1986, pp. 246-254].

[Maxwell, 1878b] J. C. Maxwell. Diffusion. *Encylopedia Brittanica* ninth edition, Vol. 7, pp. 214-221, 1978; also in [Garber *et al.*, 1986, pp. 525-546].

[Maxwell, 1878c] J. C. Maxwell. On stresses in rarified gases arising from inequalities of temperature *Proceedings of the Royal Society of London*, *27*, 304-308; *Philosophical Transactions of the Royal Society of London*, *170*, (1880) 231-256, 1978. Also in [Garber *et al.*, 1995, pp. 357-386].

[Maxwell, 1879] J. C. Maxwell. On Boltzmann's theorem on the average distribution of energy in a system of material points. *Transactions of the Cambridge Philosophical Society*, *12*, 547-570, 1879. Also in [Garber *et al.*, 1995, pp. 357-386].

[Mehra, 1998] J. Mehra. Josiah Willard Gibbs and the Foundations of Statistical Mechanics *Foundations of Physics*, *28*, 1785-1815, 1998.

[Mehra and Sudarshan, 1972] J. Mehra and E. C. G. Sudarshan. Some reflections on the nature of entropy, irreversibility and the second law of thermodynamics, *Nuovo Cimento B*, *11*, 215-256, 1972.

[Meixner, 1969] J. Meixner. Processes in simple thermodynamic materials *Archive for Rational Mechanics and Analysis*, *33*, 33-53, 1969.

[Meixner, 1970] J. Meixner. On the foundations of thermodynamics of processes. In B. Ga-lOr, E. B. Stuart & A. Brainard (Eds.), *A Critical Review of Thermodynamics*. Baltimore: Mono Book Corporation, pp. 37-47, 1970. Also in [Kestin, 1976], pp. 313-323.

[Moran, 1961] P. A. P. Moran. Entropy, Markov processes and Boltzmann's H-theorem. *Proceedings of the Cambridge Philosophical Society*, *57*, 833-842, 1961.

[Morrison, 2000] M. Morrison. *Unifying scientific theories: physical concepts and mathematical structures*. Cambridge: Cambridge University Press. 2000.

[Morrison, 1966] P. Morrison. Time's arrow and external perturbations. In A. de Shalit, H. Feshbach, & L. van Hove (Eds.), *Preludes in Theoretical Physics in honor of V. F. Weisskopf*. Amsterdam: North Holland, pp. 347- , 1966.

[Müller, 2003] I. Müller. Entropy in non-equilibrium. In [Greven et al., 2003, pp. 79-107].

[Navarro, 1998] L. Navarro. Gibbs, Einstein and the foundations of statistical mechanics. *Archive for History of Exact Science*, *53*, 147-180, 1998.

[Nemytskii and Stepanov, 1960] Nemytskii, V. V., & V. V. Stepanov. *Qualitative theory of differential equations*. Princeton: Princeton University Press, 1960.

[Norton, 2005] J. D. Norton. Eaters of the lotus: Landauer's principle and the return of Maxwell's demon. *Studies In History and Philosophy of Modern Physics*, *36*, 375-411, 2005.

[von Neumann, 1932] J. von Neumann. Proof of the quasi-ergodic hypothesis. *Proceedings of the National Academy of sciences of the United States of America*, *18*, 70-82 and 263-266, 1932.

[Obcemea and Brändas, 1983] Ch. Obcemea and E. Brändas. Analysis of Prigogine's theory of subdynamics. *Annals of Physics*, *147*, 383-430, 1983.

[Olsen, 1993] E. T. Olsen. Classical mechanics and entropy. *Foundations of Physics Letters*, *6*, 327-337, 1993.

[Penrose, 1970] O. Penrose. *Foundations of statistical mechanics: a deductive treatment*. Oxford: Pergamon Press, 1970.

[Penrose, 1979] O. Penrose. Foundations of statistical mechanics. *Reports on Progress in*

Physics, *42*, 1937-2006, 1979.

[Penrose and Percival, 1962] O. Penrose and I. Percival. The direction of time. *Proceedings of the Physical Society*, *79*, 605-616, 1962.

[Pešić, 1991] P. D. Pešić. The principle of identicality and the foundations of quantum theory. I. The Gibbs paradox. *American Journal of Physics*, *59*, 971-974, 1991.

[Petersen, 1983] K. Petersen. *Ergodic theory*. Cambridge: Cambridge University Press, 1983.

[Pitowsky, 2001] I. Pitowsky. Local fluctuations and local observers in equilibrium statistical mechanics. *Studies In History and Philosophy of Modern Physics*, *32*, 595-607, 2001.

[Plancherel, 1913] M. Plancherel. Foundations of statistical mechanics. *Annalen der Physik*, *42*, 1061-1063, 1913.

[von Plato, 1991] J. von Plato. Boltzmann's ergodic hypothesis. *Archive for History of Exact Sciences*, *42*, 71-89, 1991.

[von Plato, 1994] J. von Plato. *Creating modern probability*. Cambridge: Cambridge University Press, 1994.

[Poincaré, 1889] H. Poincaré. Sur les tentatives d'explication mécanique des principes de la thermodynamique. *Comptes Rendus de l'Académie des Sciences (Paris)*, *108*, 550-553, 1889. English translation in [Olsen, 1993].

[Poincaré, 1893] H. Poincaré. Le mécanisme et l'expérience. *Revue de Métaphysique et de Morale*, *1*, 534-537, 1893. English translation in [Brush, 1966].

[Poincaré, 1896] H. Poincaré. *Calcul des Probabilités*. Paris: Carré, 1896.

[Popper, 1982] K. Popper. *Quantum theory and the schism in physics*. London: Hutschinson, 1982.

[Price, 1996] H. Price. *Time's arrow and Archimedes' point*. New York: Oxford University Press, 1996.

[Prigogine, 1955] I. Prigogine. *Introduction to the thermodynamics of irreversible processes*. New York: Interscience, 1955.

[Ray, 1984] J. R. Ray. Correct Boltzmann counting. *European Journal of Physics*, *5*, 219-

224, 1984.

[Redhead, 1995] M. Redhead. *From physics to metaphysics.* Cambridge: Cambridge University Press, 1995.

[Reichenbach, 1956] H. Reichenbach. *The direction of Time.* Berkeley: University of California Press, 1956.

[Ridderbos, 2002] T. M. Ridderbos. The coarse-graining approach to statistical mechanics: how blissful is our ignorance? *Studies in History and Philosophy of Modern Physics*, *33*, 65-77, 2002.

[Ridderbos and Redhead, 1998] T. M. Ridderbos and M. L. G. Redhead. The spin-echo experiment and the second law of thermodynamics. *Foundations of Physics*, *28*, 1237-1270, 1998.

[Rosenthal, 1913] A. Rosenthal. Beweis der Unmöglichkeit ergodischer mechanischer Systeme. *Annalen der Physik*, *42*, 796-806, 1913. English translation in [Brush, 2003, pp. 505-523].

[Rovelli, 2006] C. Rovelli. This volume, chapter 12, 2006.

[Ruelle, 1969] D. Ruelle. *Statistical mechanics: rigorous results.* New York: Benjamin, 1969.

[Rugh, 1997] H. H. Rugh. Dynamical approach to temperature *Physical Review Letters*, *78*, 772-774, 1997.

[Rugh, 2001] H. H. Rugh. Microthermodynamic formalism *Physical Review E*, *64*, 055101, 2001.

[Saunders, 2006] S. Saunders. On the explanation for quantum statistics. *Studies in History and Philosophy of Modern Physics*, *37*, 2006.

[Schrödinger, 1950] E. Schrödinger. Irreversibility. *Proceedings of the Royal Irish Academy*, *53*, 189-195, 1950.

[Simányi and Szász, 1999] N. Simányi and D. Szász. Hard balls systems are completely hyperbolic. *Annals of Mathematics*, *149*, 35-96, 1999.

[Sinai, 1963] Ya. G. Sinai. On the foundation of the ergodic hypothesis for a dynamical system of statistical mechanics *Soviet Mathematics Doklady*, *4*, 1818-1822, 1963.

[Sinai and Chernov, 1987] Ya. G. Sinai and N. I. Chernov. Ergodic properties of certain systems of 2—*D* discs and 3—*D* balls. *Russian Mathematical Surveys*, *42*, 181-207. Also in Ya. G. Sinai (Ed.) *Dynamical systems*; *collection of papers* Singapore: Wold Scientific (1991) pp. 345-371, 1987.

[Shenker, 2000] O. Shenker. Interventionism in statistical mechanics: some philosophical remarks. http://philsci-archive. pitt. edu/archive/00000151/ 2000.

[Sklar, 1973] L. Sklar. Statistical explanation and ergodic theory. *Philosophy of Science*, *40*, 194-212, 1973.

[Sklar, 1993] L. Sklar. *Physics and Chance. Philosophical Issues in the Foundations of Statistical Mechanics.* Cambridge: Cambridge University Press, 1993.

[Sklar, 2002] L. Sklar. *Theory and thruth.* Oxford: Oxford University Press, 2002.

[Spohn, 1980] H. Spohn. Kinetic equations from Hamiltonian dynamics: Markovian limits. *Reviews of Modern Physics*, *52*, 569-615, 1980.

[Spohn, 1991] H. Spohn. *Large Scale Dynamics of interacting Particles.* Berlin: Springer, 1991.

[Streater, 1995] R. F. Streater. *Statistical dynamics*; *a stochastic approach to non-equilibrium thermodynamics.* London: Imperial College Press, 1995.

[Styer, 2004] D. F. Styer. What good is the thermodynamic limit? *American Journal of Physics*, *72*, 25-29; Erratum p. 1110, 2004.

[Sudarshan *et al.*, 1961] E. C. G. Sudarshan, P. M. Mathews, and J. Rau. Stochastic dynamics of quantum-mechanical systems, *Physical Review*, *121*, 920-924, 1961.

[Szósz, 1996] D. Szósz. Boltzmann's ergodic hypothesis, a conjecture for centuries? *Studia Scientiarum Mathematicarum Hungaria*, *31*, 299-322, 1996.

[Szilard, 1925] L. Szilard. über die Ausdehnung der phänomenologischen Thermodynamik auf die Schwankungserscheinungen. *Zeitschrift für Physik*, *32*, 753-788, 1925.

[Tabor, 1989] M. Tabor. *Chaos and integrability in nonlinear dynamics*: *an introduction.* New York: Wiley, 1989.

[Thomsen and Hartka, 1962] J. S. Thomsen and Th. J. Hartka. Strange Carnot cycles; thermodynamics of a system with a density extremum *American Journal of Physics*, *30*, 26-33, 1962.

[Thompson, 1972] C. Thompson. *Mathematical Statistical Mechanics.* Princeton: Princeton University Press, 1972.

[Tisza, 1966] L. Tisza. *Generalized Thermodynamics.* Cambridge, Mass.: MIT Press, 1966.

[Tisza and Quay, 1963] L. Tisza and P. M. Quay. The statistical thermodynamics of equilibrium. *Annals of Physics*, *25*, 48-90, 1963. Also in [Tisza, 1966, pp. 245-287].

[Tolman, 1938] R. C. Tolman. *The principles of statistical mechanics.* London: Oxford University Press, 1938.

[Touchette, 2003] H. Touchette. *Equivalence and nonequivalence of the microcanonical and canonical ensembles: a large deviations study.* Ph. D Thesis, McGill University, Montréal, 2003.

[Touchette *et al.*, 2004] H. Touchette, R. S. Ellis, and B. Turkington. An introduction to the thermodynamic and macrostate levels of nonequivalent ensembles. *Physica A 340*, 138-146, 2004.

[Truesdell, 1961] C. Truesdell. Ergodic theory in classical statistical mechanics. In P. Caldirola (ed.), *Ergodic theories*, New York: Academic Press, 1961.

[Truesdell, 1969] C. Truesdell. *Rational thermodynamics.* New York: McGraw-Hill, 1969.

[Truesdell, 1980] C. Truesdell. *The tragicomical history of thermodynamics 1822-1854.* New York: Springer, 1980.

[Uffink, 1995] J. Uffink. Can the maximum entropy principle be regarded as a consistency requirement? *Studies in History and Philosophy of Modern Physics*, *26*, 223-261, 1995.

[Uffink, 1996] J. Uffink. The constraint rule of the maximum entropy principle. *Studies in History and Philosophy of Modern Physics*, *27*, 47-79, 1996.

[Uffink, 1996b] J. Uffink. Nought but molecules in motion. *Studies In History and Philosophy of Modern Physics*, *27*, 373-387, 1996.

[Uffink, 2001] J. Uffink. Bluff your way in the second law of thermodynamics. *Studies in History and Philosophy of Modern Physics*, *32*, 305-394, 2001.

[Uffink, 2003] J. Uffink. Irreversibility and the second law of thermodynamics. In [Greven et al., 2003, pp. 121-146], 2003.

[Uffink, 2004] J. Uffink. Boltzmann's Work in Statistical Physics. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2004 Edition), http://plato.stanford.edu/archives/win2004/entries/statphys-Boltzmann.

[Uffink, 2005] J. Uffink. Rereading Ludwig Boltzmann. In P. Hájek, L. Valdés-Villanueva & D. Westerståhl (Eds.) *Logic, Methodology and Philosophy of Science. Proceedings of the twelfth international congress.* London: King's College Publications, pp. 537-555, 2005.

[Uffink, 2006] J. Uffink. Insuperable difficulties: Einstein's statistical road to molecular physics. *Studies in History and Philosophy of Modern Physics*, 2006.

[Uffink and van Lith, 1999] J. Uffink and J. van Lith. Thermodynamic uncertainty relations. *Foundations of Physics*, *29*, 655-692, 1999.

[Uhlenbeck and Ford, 1963] G. E. Uhlenbeck. and G. W. Ford. *Lectures in statistical mechanics.* Providence, Rhode Island: American Mathematical Society, 1963.

[Vranas, 1998] P. B. M. Vranas. Epsilon-ergodicity and the success of equilibrium statistical mechanics. *Philosophy of Science*, *65*, 688-708, 1998.

[Wald, 2001] R. M. Wald. The thermodynamics of black holes. *Living Reviews in Relativity*, 4, 6. http://www.livingreviews.org/lrr-2001-6.

[Winsberg, 2004] E. Winsberg. Laws and statistical mechanics. *Philosophy of Science*, 71, 707-718, 2004.

[Yourgrau *et al.*,] abc W. Yourgrau, A. van der Merwe, and G. Raw. *Treatise on irreversible thermophysics*, New York: Macmillan, 1966.

[Zermelo, 1896a] E. Zermelo. Ueber einen Satz der Dynamik und die mechanische Wärmetheorie. *Annalen der Physik*, *57*, 485-494, 1896. English translation in [Brush, 2003, pp. 382-391].

[Zermelo, 1896b] E. Zermelo. Ueber mechanische Erklärungen irreversibler Vorgänge. *Annalender Physik*, *59*, 793-801, 1896. English translation in [Brush, 2003, pp. 403-411].

第十章　量子统计物理

杰拉德·埃姆什

1. 引言

在 1804 年至 1806 年探险期间，刘易斯和克拉克(Lewis and Clark)在寻找密苏里河源头的过程中，发现它是由杰斐逊、加勒廷和麦迪逊河交汇而成的，并最终成为宽广的密西西比河的一个主要支流。

类似地且具有一定任意性的是，我们可以用三个主要标志性成果来标记量子统计物理(QSP)的开端：普朗克在其 1900 年的论文[Planck，1900a；1900b]中提出的“量子假说”、吉布斯(Gibbs)在 1902 年出版的有关“统计力学”的书[Gibbs，1902]，以及现在已经很明确的爱因斯坦(Einstein)于 1905 年提出的“布朗运动”(Borwnian motion)[Einstein，1905b]。在今天看来，QSP 的影响已经表现在凝聚态物理学(从固体物理到天体物理)中。该领域的研究方向，虽然经常是试探性的，但已经能被用来给出一些预测，这些预测已被足够多的精确实验所证实，从而要求一种一致性的解释。本章的目的是指出可能发现这种解释的方向。我通过简单地回顾上面提到的三个成果的发展轨迹，即通过确定 QSP 最初的发展动机来开始探讨。

普朗克长期的犹豫不决表明他不仅领先自己的时代太多，甚至或许都在自己的思维之前，例如，最初，他曾根据实体(壁面之间的小型振荡器)的性质提出他的黑体辐射定律，而不是根据辐射的性质得到的。由普朗克转向用玻耳兹曼(Boltzmann)曾在材料物质的热物理中所用的论证来描述电磁波可以知道，他最初曾保持这个问题的开放性，即这是否仅是一个单纯的形式类比，还是一个可以从公认的辐射与物质之间的相互作用中得到证明的结论，还是，这种推测性的类比有更深层次的原因。普朗克的为难还体现在他在 1913 年所写的推选年轻的爱因斯坦为普鲁士科学院院士的推荐信中，他这样写道："他可能有时错过了他的目标，正如在他的光量子假说中那样，但真的不能因此而反对他。"虽然这可能被视为是对[Einstein，1905a]的讽刺，但需要注意的是，普朗克的量子假说的提出并不是昙花一现的意外。在用词方面他是很严谨的，例如他在 1911 年写给德国化学学会的信中对"理论""定理"及"假说"的使用[Planck，1911]。此后不久，世界各地的学者们就克服了他们自己的顾忌：诺贝尔奖在 1918 年被授予普朗克，因为他在"发现能量子"方面的贡献；在 1921 年授予爱因斯坦，因其在"发现光电效应定律"方面的贡献。在这里请注意，他们二人都对早期 QSP 的发展有贡献，具体就是普朗克的黑体辐射和爱因斯坦的固体比热容。详见下面的第 2.1 和 2.3 节。

吉布斯的书[Gibbs，1902]则专注于经典统计物理。然而德国学者克劳修斯(Clausius)、奥地利学者玻耳兹曼和英国学者麦克斯韦(Maxwell)对其中基本概念的理解并不相同，美国学者吉布斯认为该领域已经接近成熟期，有必要对其基础进行整合。关于其他领域的公理化，可见于希尔伯特(Hilbert)[Hilbert，1900；1899；1918]和爱因斯坦[Einstein，1921]体系的比较。即使在经典背景下，吉布斯也不愿意采用玻耳兹曼的遍历假说，这表明吉布斯用其著作的标题所言的"热力学的合理基础"这个未解难题仍然存在，被简单介绍的那些吉布斯的工作中，和本文的目的最相关的可参见[Uffink(尤菲克)，2006，Section 5]。本章所研究的是在量子领域中研究在何种程度与之前的研究是截然不同的，在何种程度上这种相异性与 QSP 的解释目标相关。

爱因斯坦关于布朗运动的论文在概念上仍属于经典物理学领域。尽管忽视了 19 到 20 世纪的交替之时许多数学家仍将概率理论作为基础的事实，参见

[Hilbert, 1900, Problem 6]，但爱因斯坦的方法见证了一个事实，即随机论证(即涉及随机过程的论证)在物理学家的领域内已经占有了一定的地位。爱因斯坦的结论被广泛地(如果不是普遍地)就其字面含义被理解为支持分子存在的经验证据，分子不仅仅被认为是方便计算的小实体或小单元，而且被认为是具有固定尺寸的物体[Einstein, 1906b]。此外，爱因斯坦后来的论文表明，他并不像他第一篇论文的烦琐标题可能表现出的那样是一个非凡的天才[Einstein, 1905b]。一方面，从物理学家的角度来看，必须注意的是，爱因斯坦在他第二篇论文的一开始就表明他忽略了西登托夫和古埃(Siedentopf and Gouy)的早期贡献，他们将“所谓的布朗运动”(爱因斯坦的原话)解释为是由分子的不规则热运动所引起的[Einstein, 1906c; Gouy, 1888]。另一方面，有了诸如卡克(Kac)和钱德拉塞卡(Chandrasekhar)等学者的后见之明，现代数学家们认识到，斯莫罗科夫斯基(Smolukowski)同时还从经验资料中抽取出了数学直觉，这使得他提出了后来被称为随机过程理论的论断[Smolukowski, 1906a; 1916]。然而，仅仅到了 1933 年，柯尔莫哥洛夫(Kolmogorov)就弄清楚了基础句法的本质，即概率的数学理论[Kolmogorov, 1933]。即便如此，关于严格意义上的语义当今仍存在一个悬而未决的问题：究竟是选冯·米塞斯(von Mises)的集合[von Mises, 1928]还是德菲尼蒂(de Finetti)的主观说明[de Finetti, 1937]。当考虑到概率理论在量子领域中的扩展，特别是 QSP 的具体要求时，我支持(见第 3.1 节)后者。

正如这篇文章一开始提出的那个问题，这三方面成果的交汇是否使得它们各自的基础问题复杂化了，或是相反，它们是否可以互相提供信息呢？我的论证将会沿后一种观点进行，虽然我并不是不知道一个普遍存在的问题，即质疑什么样的现实实体应该或者不应该被归为微观量子系统。作为用统计力学简化热力学这个大问题中的一部分，我会具体地考虑 QSP 是否以及能够如何解释多体系统的集合属性的问题：它确实假设了微观层面上的量子描述，但到目前为止它不能从本体论角度理解这些系统的个体组分。在我的论述中，我将遵从爱因斯坦的格言：“如果你想从理论物理学家身上找到关于他们所用方法的任何事实……那就不要听他们说了什么，而要看他们做了什么。”参见[Einstein, 1933]。

2. 早期成就

马克思·雅默(Max Jammer)在[Jammer, 1966]中给出了有关量子理论开端的特殊历史评论；并且他讨论了一些由此产生的辩论[Jammer, 1974]。在这里，我从讨论 QSP 的早期实用性成就开始，特别要提到两个方面：它们在高温区域的经典问题以及理解波粒二象性下的 QSP。这两方面都说明了在 QSP 中而不是在量子理论中问同一个问题，会由于背景的不同而得到不同的答案，比如说玻尔(Bohr)原子或散射过程，请与马拉·贝勒(Mara Beller)在理解量子革命时的视角做对比[Beller, 1999]。

2.1 普朗克关于黑体辐射的插值公式

普朗克拥有的实验数据是单位体积内电磁辐射能量的谱密度 $\rho_T(\nu)$，当电磁辐射在一个温度为 T 的黑体内处于平衡态时，它是关于频率 ν 的一个函数。在[Planck, 1900a; 1900b]中，普朗克为适应这些数据提出这个公式：

$$\rho_T(\nu) = A\frac{h\nu}{e^{\frac{h\nu}{kT}}-1},\quad 其中\ A = \frac{8\pi\nu^2}{c^3} \tag{1}$$

其中 c 为光速，$k=R/N_{Av}$ 为玻耳兹曼常数，R 是一般气体常数，N_{Av} 是阿伏伽德罗常数。此外，这个公式引入了一个新的常数 h，就是现在我们所知的普朗克常数。而普朗克自称公式(1)只是一个“幸运的猜测”，这样一个公式在概念真空中不可能存在。

维恩(Wien)[1894]、斯特藩(Stefan)[1879]和玻耳兹曼[1884]确定了两个定性的定律，维恩位移定律指出：

$$\rho_T(\nu) = \nu^3 f\left(\frac{\nu}{T}\right) \tag{2}$$

其中 f 是某种不确定的常数，满足下式积分收敛这个条件：

$$\frac{1}{V}E(T) = \int_0^\infty d\nu\rho_T(\nu) \tag{3}$$

它表示在温度 T 下每单位体积内辐射能量的密度。将(2)带入(3)中，就会得到

斯特藩—玻耳兹曼定律：

$$E(T)=\sigma T^4 \tag{4}$$

其中，σ 为常数。普朗克的提议符合上述定律。

后来又给出两个解析表达式（或称“定律”），具体指定式（2）中的函数 f。其中一个定律由维恩[1896]给出，为：

$$\rho_T(\nu)=\alpha\nu^3\exp^{-\gamma\frac{\nu}{T}} \tag{5}$$

其中 α 和 γ 是两个常数，当 ν/T 很大时，该定律已在经验上得以证实。与此相反，另一条定律由瑞利（Rayleigh）[1900]提出，请参阅金斯（Jeans）[1905a]：

$$\rho_T(\nu)=\frac{8\pi}{c^3}\nu^2kT \tag{6}$$

当 ν/T 很大时，该定律已在经验上得以证实。

显然，式（1）是介于维恩和瑞利—金斯公式之间的修改后的解析式，它提供了条件 ν/T“大”（或“小”）的定量意义，即 $\nu/T\gg k/h$（或 $\nu/T\ll k/h$）。在中间范围内，普朗克的插值公式非常好地符合了实验结果，无论从定性还是定量的角度来说都是如此。

不可能不给普朗克的同事们留下深刻印象的是普朗克的伟业即将完成，因为他能从第一原理出发解释其公式，至少在式（2）—式（6）可被理解的程度予以解释。然而，普朗克不得不采取“一种绝望的行为”——他自己的话[Max Jammer，1966]——在经过多次尝试后，他构造了一个启发式模型，在这个模型中辐射以离散的量子方式与壁面间处于热力学平衡态的假设的“谐振器”交换能量。该模型有几个缺点——其中之一是普朗克调整了玻耳兹曼的计数法——且其理论地位也具有很大的不确定性：

> 沃尔特·内斯特（Walter Nernst）……最初并不喜欢量子理论，声称它“除了一个插值公式外什么也没有……仅仅是为了计算而设定的一条规则……但普朗克的工作却证明它是如此有效……并且……视科学为责任的爱因斯坦很认真地对待它且仔细研究它”[Jammer，1966，59]。

后来物理学群体达成的共识是，任何希望从第一原理推导出式(1)的企图——也包括普朗克的企图——注定会失败。说明式(1)是一个基本的或初始定律，即它是不需要解释的，但应该对它的结论加以探讨。

2.2 爱因斯坦的涨落方程和波粒二象性

从一开始，爱因斯坦就注意到普朗克推导过程的两个缺点。首先是形式上的，但仍然是本质性的，正如普朗克表明的那样，普朗克的计数不符合玻耳兹曼统计的计数法。第二点需要指出的是，[Einstein，1906a]中提出，普朗克的方法在以下两方面会导致不一致：(a)他使用(经典)麦克斯韦电磁理论计算谐振器中辐射场的平均能量；(b)谐振器中的能量只能不连续地改变的假设。连同其他经验问题——其中就包括光电效应[Einstein，1905a]——这些困难使得爱因斯坦提出，普朗克辐射表达式(1)具有无可争辩的经验价值，因此“量子化”自身应该在辐射场中寻找原因，而不是在与壁面的相互作用的一个可疑机制中。同时，爱因斯坦还批判地理解了这个问题，即光是否应该在麦克斯韦电磁理论中被认为是类波的；或是否应该被看作是类粒子的，就像在19世纪初干涉实验之前的牛顿理论所认为的那样。

爱因斯坦的涨落方程[Einstein，1909a]指出，光应该被看作同时具有粒子性和波动性，特别地：

批注1。将式(1)中普朗克谱密度$\rho_T(\nu)$解释为当温度为T时，处于热平衡态的辐射频率为ν的量子振荡器的平均能量$\langle u_T(\nu)\rangle$。那么对于$h\nu/kT$的每个值，能量涨落$\langle(\Delta u)^2\rangle = kT^2\partial_T\langle u_T(\nu)\rangle$是下面两项的和：

$$\langle(\Delta u)^2\rangle = \langle(\Delta u)^2\rangle_p + \langle(\Delta u)^2\rangle_w，这里$$

$$\left\{\begin{aligned}&\langle(\Delta u)^2\rangle_p = \langle u_T(\nu)\rangle h\nu\\&\langle(\Delta u)^2\rangle_w = \langle u_T(\nu)\rangle^2\frac{c^3}{8\pi\nu^2}\end{aligned}\right\} 且 \langle(\Delta u)^2\rangle_p \ /\langle(\Delta u)^2\rangle_w = \exp^{\frac{h\nu}{kT}} - 1 \qquad (7)$$

因此，当$h\nu/KT\gg 1$时，粒子性$\langle(\Delta u)^2\rangle_p$占支配地位；当$h\nu/kT\ll 1$时，类波性$\langle(\Delta u)^2\rangle_w$占统治地位。因此在这个解释中，波粒二象性是一个程度的问题，而不是在相互排斥的两种性质之间二选一的问题。

除了在QSP中是一个概念问题外，这种二象性问题在其他经验背景下会变

得更加困难，因为在这些背景下人们可能更愿意将光子看作是粒子或者看作是波包。此外，这种二象性已经被扩展到所有(亚原子)粒子中，如通常与波有着密切关系的那些现象，例如光的衍射现在在电子束或中子束中已经观测到了，参见[Jammer，1966，pp. 249—253]，或更近期的[Rauch(劳赫)，2005]。在其他情况下，人们更喜欢使用粒子语言，例如在描述光电效应时[Einstein，1905a]，正如在大多数QM(量子力学)书中所讲的，光子撞击金属表面导致电子被发射出来，或者在原子光谱学中，一个粒子(原子)会发出一束光，云室和气泡室让我们直观地看到粒子碰撞，但他们在描述散射定理时使用了所谓的波的方式，例如[Amrein *et al.*(艾瑞恩等)，1977]。光的这种双重性，在爱因斯坦和德布罗意(de Broglie)的推测下，导致物理学家们学会了如何针对他们想强调的方面调整他们的语言。然而，关于"自干涉"的持续论证表明，仍然还有一些残存的歧义有待解决，参见[Taylor(泰勒)，1909]和[Aichele *et al.*(艾驰乐等)，2005]等论著对此进行的长期辩论，而且很显然在将来辩论还会持续。

以上回顾了QSP的早期发展，应该提一下上述爱因斯坦涨落方程(7)，以及由固体比热来对温度做出的解释——见下文第2.3节——促使埃伦费斯特夫妇认为[Ehrenfest and Ehrenfest，1911]在统计力学中，量子行为本身大多表现在低温条件下，而经典行为一般在高温条件下出现。事实是，在许多表达式中，如在普朗克分布式(1)中，普朗克常数 h 和温度 T 一起出现在因子 h/T 中，或以后来使用的形式 $\hbar\beta$ 出现，因此在这些表达式中"经典极限" $h\to 0$，和"高温限制" $T\to\infty$ 都被包括在 $(\hbar\beta)\to 0$ 中。所有涉及相关能或能量密度情况的系统，在由普朗克常数的数值决定的范围内测量时都是非常大的。

2.3 低于经典温度下固体的德拜(Debye)比热

本节的目的是要阐明在这个领域中当前的情况，杜龙和佩蒂特(Dulong and Petit)(1819年)曾提出一个论点，认为比热的效果——以卡/每度每摩尔为单位——对所有固体来说应该都是相同的：即 $3R$，其中 R 为气体常数。然而，后来证明这个"常数"是会随着温度降低而急剧降低的，以至于在19世纪末期，实验数据导致了一个假说，即固体比热会随着温度降至绝对零度0 K而变得微乎其微地小。同一时期，伦琴(Roentgen)(1895年)发现了X射线，之后几个

实验物理学家——埃瓦尔德(Ewald)(1911 年)在冯·劳厄(von Laue)、弗里德里克(Friedrich)和克尼平(Knipping)(1912 年)的建议下——利用 X 射线获得了衍射图样，从而证实了晶体是规律性的格状体的假设，原子都位于顶点处。

由于没有出现针对所观测到的依赖于比热的温度变化的经典解释，爱因斯坦和德拜提出了下面的模型，参见[Einstein，1907；Einstein，1911b；Debye，1912]。

其出发点是上面的式(1)，即黑体辐射的普朗克公式，现在重新诠释为温度 T 下固体的振动模型：

$$U(T)=\int d\nu g(\nu)U(\nu,T)\text{，其中 }U(\nu,T)=\frac{h\nu}{e^{\frac{h\nu}{kT}}-1}\text{ 且}\int_0^\infty d\nu g(\nu)=3N \tag{8}$$

这里 N 是固体中三维振子的数量。爱因斯坦假设 g 集中于一个固定频率 ν_0 上，德拜为 g 选择了一个最简单的振动分布，这个分布可以看作是在体积为 V 的晶体中，振动在晶格中原子间距离的顺序有一个最小的波长：

$$g(\nu)=G\begin{Bmatrix}1 & \text{如果} & 0\leqslant\nu\leqslant\nu_o\\ 0 & \text{如果} & \nu>\nu_o\end{Bmatrix}\text{，其中 }G=\frac{12\pi\nu^2}{s^3}V \tag{9}$$

现在考虑的振动为声波，而不是电磁波，因此 G 由上式得到——与式(1)中的 $A=\frac{8\pi\nu^2}{c^3}$ 相比较——可知代替光速 c 的 s 现在是声速。用 $12\pi=(2+1)\cdot4\pi$ 替换 $8\pi=2\cdot4\pi$ 反映了这样的事实，即在固体中声波，除了也存在于光波中的两个横向偏振方向外，还具有第三个自由度，即纵向模式。这些假设导致了以下结果。

批注 2。存在一个温度 Θ，使得比热满足：

$$C_V\simeq\begin{Bmatrix}3R\cdots\cdots\cdots\cdots\cdots\cdots\text{当 }T\gg\Theta\\ \frac{12}{5}\pi^4R\left(\frac{T}{\Theta}\right)^3\qquad\text{当 }T\ll\Theta\end{Bmatrix} \tag{10}$$

因此，德拜模型在两个温度区域间是有区别的：在高温下德拜模型遵循杜龙—佩蒂特定律，可以预测的是随着温度接近于 0K，比热会随着 $C_V\sim T^3$ 而逐渐变为零。在此模型中，温度 Θ(现在被称为德拜温度)取决于所考虑的固体的临界频率 ν_o，因此取决于在该固体中的声速及其密度 N/V。Θ 的数值给出了对

晶体比热在高低温下不同这一现象的定量分析——有关详细信息，请参阅第6.1节。此外，在德拜模型中，C_V 在整个温度 $T\in\mathbb{R}^+$ 范围内为连续单调递减函数。

本小节对式(1)—式(8)的最后评论为，类比于将光子看作光量子，固体中基础的声音振动也可被视为量子，现在被称为声子。

2.4 玻色—爱因斯坦凝聚：长程

当我们严肃对待宏观世界的微观图景可能就是量子图景这个想法时，最迫切的问题就是要找到理想量子气体的相应描述，这后来被称为玻色—爱因斯坦气体，或简称为玻色气体[Bose，1924；Einstein，1924]。出发点是在温度 T 和化学势 μ 下，处于平衡态的质量为 m 的大量全同粒子系综的巨正则配分函数 $Z(\Lambda,T,\mu)$，这个系综是封闭在一个体积为 $\Lambda\subset\mathbb{R}^3$ 的立方体中的，它满足周期性边界条件。由于这些粒子之间没有相互作用，因此总能量为它们个体能量的总和 $\epsilon_k=\hbar^2|k|^2/2m$，其中 $k\in Z^3$。量子假说就是普朗克分布式(1)在这里适用，以使得下式成立(其中 $\beta=1/kT$)：

$$Z(\Lambda,T,\mu)=\prod_{k\in Z^3}(1-\exp^{-\beta(\epsilon_k-\mu)})^{-1} \tag{11}$$

由该式我们可根据在经典统计力学学到的定律计算出比容 v 和压强 P，所谓的活性被定义为 $z=\exp(\beta\mu)$：

$$v^{-1}=z\partial_z\frac{1}{|\Lambda|}\ln Z(\Lambda,T,\mu)\text{且}\beta P=\frac{1}{\Lambda}\ln Z(\Lambda,T,\mu) \tag{12}$$

该问题因此得以完整地陈述，虽然(11—2.12)的结论是不容易被直接计算出的。该解通过一些经典分析后反映为一个数学上的偏移，而且在物理学上也表现为一个偏离：在非常低的温度下相变起始于凝聚相，这不是经典的理想气体！

先驱们已经掌握了必要的经典分析方法——包括现在广泛使用的，参见例如[Whittaker and Watson(惠特克和沃森)，1927，280]、[Erdélyi(艾德林)，1953，I，pp. 27—30]以及一些历史性的论著如[Truesdell(特鲁斯德尔)，1945]中的分析——他们确实认识到在极限 $\Lambda\uparrow\mathbb{R}^3$ 下，这个和值约化为：

$$\left.\begin{aligned} v^{-1} &= 4\pi\int_0^{\infty} dp\, p^2 z[\exp(\hbar^2 p^2/2mkT) - z]^{-1} \\ \beta P &= 4\pi\int_0^{\infty} dp\, p^2 \ln[1 - z\exp(-\hbar^2 p^2/2mkT)] \end{aligned}\right\} \tag{13}$$

在经典分析中这被看作阿贝尔(Abelian)积分，即：

$$\left.\begin{aligned} v^{-1} &= \lambda^{-3} g\left(\frac{3}{2}, z\right) \\ \beta P &= \lambda^{-3} g\left(\frac{5}{2}, z\right) \end{aligned}\right\},\quad \text{其中}\left\{\begin{aligned} \lambda^2 &= 2\pi\hbar^2/mkT \\ g(s,z) &= \frac{z}{\Gamma(s)}\int_0^{\infty} dt\, \frac{t^{s-1}}{\exp^t - z} \end{aligned}\right. \tag{14}$$

对于满足 $Re(s)>0$ 的每个 s，g 定义一个 z 的函数，它在复平面截面 $C\backslash[1, +\infty)$ 上是解析的。对 $|z|<1$ 和 $Re(s)>0$，我们得到了已被充分研究的勒奇(Lerch)ζ 函数，它可由幂级数展开为：

$$g(s,z) = z\zeta(s,z)\text{, 其中}\ \zeta(s,z) = \sum_{n=0}^{\infty} z^n (n+1)^{-s} \tag{15}$$

对于 $z=1$ 和 $Re(s)>1$，上述级数收敛于黎曼(Riem)ζ 函数 $\zeta(s)$。请注意，值 $s=\frac{3}{2}$ 和 $s=\frac{5}{2}$ ——这是式(14)所需要的——落在此范围内。此外 $g\left(\frac{3}{2}, \cdot\right)$：$z\in(0, 1)\to R^+$ 是平滑的、严格递增的，其中 $\lim_{z\to 1} g\left(\frac{3}{2}, z\right) = \zeta\left(\frac{3}{2}\right) = 2.612\cdots$ 因此，该问题在数学上是完全可控的。

现在讨论物理学方面。这可以分为两个步骤。

第一步很简单，只考虑高温和低密度区域，这里 $\lambda^3 v^{-1} < g\left(\frac{3}{2}, 1\right) = \zeta\left(\frac{3}{2}\right)$。特别地，直接做一阶展开为：

$$\text{对于}\ \lambda^3 v^{-1} \ll 1\text{：}\ Pv = kT[1 - 2^{-5/2}(\lambda^3 v^{-1}) + \cdots] \tag{16}$$

因此，在这种高温和低密度区域内，量子气体逐渐表现得像博伊尔/马里奥特/盖—吕萨克(Boyle/Mariottel Gay-Lussac)的经典理想气体一样。这是另一种确认埃伦费斯特说法的方式，根据这个说法在 QSP 中得到的经典极限可被看作是高温极限。注意式(14)中出现的热力学波长 λ 满足 $\lambda \sim \hbar\beta^{\frac{1}{2}}$，即再回到这个问题，极限 $T\to\infty$ ($\Leftrightarrow \beta\to 0$) 和 $\hbar\to 0$ 在形式上具有相同的结果。

处理这个问题的第二个步骤是找到物理上的偏离是在哪里被发现的。现在的问题是如何超越上面的区域，即超越非自然极限：

$$\lambda^3 v^{-1} = \zeta\left(\frac{3}{2}\right) \tag{17}$$

没有实际气体能超越这个限制。在数学上，这个限制性条件似乎可被视为从 $z = 1$ 开始的式(14)的解析性被破坏的结果。在物理上，这个问题的出现是因为极限 $|\Lambda| \to\infty$ 的选取太过草率。

因此，当 $|\Lambda| < \infty$ 时让我们返回到v^{-1}的表达式。那么我们有下式，其中$\langle n_k\rangle$表示状态下的平均粒子数：

$$\frac{1}{|\Lambda|}\sum_{k\in Z^3}\langle n_k\rangle = \frac{1}{|\Lambda|}\sum_{k\in Z^3,k\neq 0}\langle n_k\rangle + \frac{1}{|\Lambda|}\frac{z}{1-z} \tag{18}$$

由于 $k\neq 0$ 时$\langle n_k\rangle$随着 $z\to 1$ 不会发生突变，因此式(18)可分为两项这个结果表明，我们同时取极限 $|\Lambda| \to\infty$ 和 $z\to 1$，式(18)中的第二项会接近一个有限的极限，比如说 v_0^{-1}，这导致式(14)可由下式替换：

$$\left.\begin{aligned} v^{-1} &= \lambda^{-3}\zeta\left(\frac{3}{2}\right) + v_o^{-1} \\ \beta P &= \lambda^{-3}\zeta\left(\frac{5}{2}\right) \end{aligned}\right\} \tag{19}$$

上述的取极限过程，可以体现为：

$$v_o^{-1} = \lim_{|\Lambda|}\frac{1}{|\Lambda|}\langle n_o\rangle \tag{20}$$

这导致在宏观上会体现出来基态 $k = 0$ 被占据的情况，该理论无法预测 v_0 的值：它也可能依赖于温度。需要注意的是式(19)中的压强 P 只(通过 λ)依赖于温度。由式(19)描述的系统的状态被称为凝聚相，从正常相 $\lambda^3 v^{-1} < \zeta(\frac{3}{2})$ 过渡到这个相的过程被称为玻色—爱因斯坦凝聚(或 BEC)，其在低温下的表现需要完全从量子层面来预测，在经典世界没有对应的原因。

在这里举一个实验室中的实例来说明。在低温下超流相出现在^{4}He 中。这种现象开始出现时的密度大约为 $\rho\approx 0.178\text{g/cm}^3$。同时考虑阿伏伽德罗常数的值，我们取 $v^{-1}\approx 2.7\times 10^{23}\text{cm}^{-3}$，由此式(17)给出了一个热力学波长 $\lambda\approx 4.6\times 10^{-8}\text{cm}$，这样取不是没有道理的，这个量被认为是粒子间距离的量度。由式

(14)中 λ 的定义，这个值对应的温度为 $T \approx 3.2\mathrm{K}$。^{4}He 中超流相开始出现时实验温度为 $T \approx 2.2\mathrm{K}$，考虑到这个模型是如此粗糙，因而它非常地合适。此外，可以计算出模型的热力学，参见[Huang(黄昆)，1965]。并且它表明，比热 $C_v(T)$，最初从 $C_v(0)=0$ 开始单调增加，直至超过经典值3/2，然后经历了一个峰值——一阶导数不连续——从这里单调递减至 $\lim_{T\to\infty} C_v(T)=3/2$。^{4}He 的比热也表现出这样的一个奇点，尽管更明显：它是对数性的，因此它的名称 λ 点是由于作为温度函数的比热的曲线看起来像希腊小写字母 λ 而得到的。

所有这一切在20世纪20年代中期都被描述为伟大的成就。当人们试图用理论解释^{4}He 不是气体而是液体的事实时，下面的问题出现了，就这一点来说模型所需的理想气体假设是相当不现实的：液体并不是由非相互作用粒子构成。将相互作用加入理论中被证明是一个很大的问题，长期的实验事实表明^{4}He是唯一被确认为可以展现玻色—爱因斯坦凝聚的物质：理论物理学家没有可变的参数用以指导和调整他们的推测。继20世纪50年代中后期提出的方案之后，随着20世纪80年代到90年代之间微开尔文技术的出现，情况发生了巨大的变化，这项技术使得在谐振子势阱内的原子气体中能观测到BEC现象。参见两篇有深度但截然不同的综述性文章[Lieb(利布)，2001]和[Pitaevskii and Stringari(皮托斯基和斯丁格雷)，2003]，简单的综述请参见[Emch and Liu(埃姆什和刘)，2002，§14.2.2]。

本节主要限于讨论BEC处于起始阶段时在宏观和热力学方面的内容。在下面的第5.2节中，主要讨论了与在QSP中出现的模数结构相关的玻色—爱因斯坦模型的 C^* 代数，这种模数结构与平衡KMS条件相关。

2.5 超越玻尔原子：托马斯—费米(Thomas-Fermi)模型

标志着量子力学开始的薛定谔(Schrödinger)波动力学(1926年)取得了巨大的成功，物理学界立刻就认可了在这个新领域应用当时才刚建立的这种方法，氢原子能谱的理论解释被简化用来解决微分方程中的本征值问题。每本量子力学的入门级教材中都提出了这个推导过程。

然而，在玻尔原子之外，求解拥有哪怕稍多一些电子的原子的薛定谔方程也是一个不可逾越的任务：电子是带电粒子，一个单一电子与原子核之间的相互作

用在氢原子中已经得以解释，但人们无法解析地处理电子之间的电磁相互作用。

此后很快，托马斯[1927]和费米[1927]想出了一个半经典模型，它主要包含两个组分。首先是基态电子密度 ρ，它被假设为球对称的，且由下面的条件归一化：

$$4\pi\int_0^\infty drr^2\rho(r) = Z \tag{21}$$

其中，eZ 是原子核的电量。第二组分是原子的平均电势 $\Phi(r)$。假设这两种组分满足经典方程，静电学中的泊松(Poisson)方程：

$$\Delta\Phi \equiv \frac{1}{r}\frac{d^2}{dr^2}(r\Phi) = 4\pi e\rho,\text{其中}\lim_{r\to 0}\Phi(r) = eZ \tag{22}$$

然而，该模型具有一个量子特性可用来解释泡利(Pauli)不相容原理，这就是之前那一年刚提出的所谓费米—狄拉克(Fermi-Dirac)统计[Fermi，1926]。在这里，将其表示为如下形式：

$$n(r,\ p) = \begin{cases} 2h^{-3} & \text{如果} \quad \epsilon := \frac{1}{2m} - e\Phi < \epsilon_o \\ 0 & \text{如果} \quad \epsilon > \epsilon_o \end{cases} \tag{23}$$

通过对 p 进行积分(设 $\epsilon_0 = 0$)，我们得到 ρ 满足：

$$\rho(r) = \begin{cases} \frac{8\pi}{3h^3}(2me\Phi)^{3/2} & \text{如果} \cdots\cdots\cdots\cdots \Phi > 0 \\ 0 & \text{如果} \cdots\cdots\cdots\cdots \Phi < 0 \end{cases} \tag{24}$$

显然，该模型在概念上是不一致的，它徘徊于经典和量子领域之间。然而，我在学生时代时，这个模型是用来讲授量子力学课程的主要方式[Schiff(希夫)，1955；Landau and Lifshitz(朗道和利弗席兹)，1958a；Messiah(弥赛亚)，1960]，因为除了上面列出的假设以外，它不需要更多的假设就能解决问题。解是精确的，这取决于一个事实，即在可控的近似范围内它很好地实现了数值计算。

当使用普朗克常数 h、电荷 e 和电子质量 m 的数值时，该模型预测原子半径(包含了所有电子的球体的半径)是单调递增的，从 $Z=25$ 的 2.2×10^{-8} 厘米增大到 $Z=100$ 的 2.8×10^{-8} 厘米，这个量级是正确的。因此，这可以算作是量子理论的一个早期成就。

不过，我们应该想到，这样一个粗略的模型并不能说明全部的事实。事实上：(1)预测的增大在 $Z=55$ 时(对应于铯原子)会停止，之后半径将会减小，尽管速度缓慢；(2)当进行更严密的讨论时，我们会发现该模型产生的电子密度具有不合理的属性，即电子离原子核既非常近，又非常远。此外，该模型需要被再次认真地考虑，以解释稳定分子的存在性，或与相对论的方法相适应的问题。这些问题从来没有彻底离开过理论物理学家的视线，但近半个世纪以来它们或多或少都是以背景的方式存在的，直到严密的分析方法明确了在什么意义上模型是渐近精确的，并且可被用于研究原子、分子甚至星体的稳定性，参见[Lieb and Simon(利布和西蒙)，1977；Lieb，1982a；Lieb，1990]，还可参见[Catto *et al.*(卡托等)，1998；Le Bris and Lions(勒布里和莱恩斯)，2005]。

2.6 白矮星：钱德拉塞卡边界

回到第2.4节中讨论的理想量子气体，现在让我们来看看费米气体。与式(11)不同，我们从配分函数开始：

$$Z(\Lambda,T,\mu) = \prod_{k\in Z^3}(1+e^{-\beta(\epsilon_k-\mu)}) \tag{25}$$

上式需要在极限 $\Lambda\uparrow Z^3$ 下才能成立，我们将式(13)替换为：

$$\left.\begin{aligned} v^{-1} &= 4\pi\int_0^\infty dp\, p^2 z[\exp(\hbar^2p^2/2mkT)+z]^{-1} \\ \beta P &= 4\pi\int_0^\infty dp\, p^2\ln[1+z\exp(-\hbar^2p^2/2mkT)] \end{aligned}\right\} \tag{26}$$

在高温和低密度范围内，即 $\lambda^3v^{-1}\ll 1$，我们又得到了渐近展开式，其中首项对应于经典理想气体：

$$\text{对于 } \lambda^3v^{-1}\ll 1\text{：} P\,v\simeq kT[1+2^{-5/2}(\lambda^3v^{-1})+\cdots] \tag{27}$$

同样，取决于修正项，这与玻色—爱因斯坦的结果式(16)非常相似：当 T 变大时，它也逐渐地接近经典理想气体。

在低温和高密度区域，即 $\lambda^3v^{-1}\gg 1$，情况与2.4节完全不同：在这个区域玻色子往往凝聚在一起，然而根据泡利不相容原理，没有任何两个费米子被允许处于相同的态上。从化学的角度看，这个原理就是对门捷列夫元素周期表的量子解释。在QSP中泡利原理由式(26)表达，在系统的基态上，费米子占据

尽可能低的能态，这个态具有有限的能量，被称为费米能：

$$\epsilon_F = \frac{\hbar^2}{2m}[(3\pi^2)v^{-1}]^{\frac{2}{3}} \tag{28}$$

令温度满足 $kT \ll \epsilon_F$，动量分布则为：

$$\langle n_p \rangle = \begin{cases} 1 & 对于 \quad (|p|^2/2m), \ \widetilde{<} \epsilon_F \\ 0 & 对于 \quad (|p|^2/2m), \ \widetilde{>} \epsilon_F \end{cases} \tag{29}$$

在 ϵ_F 附近的很窄的 kT 范围内是一条尖锐曲线，这一范围内的气体被称为简并费米气体。为了描述这一范围的特征，用式(28)中 ϵ_F 的形式将 $kT \ll \epsilon_F$ 改写为：

$$\beta v^{-\frac{2}{3}} \gg \left[\frac{\hbar^2}{2m}(3\pi^2)^{\frac{2}{3}}\right]^{-1} \tag{30}$$

这给出了一个表示低温和高密度区域的定量方式；例如，这会产生一个关于常温下金属中的电子气体的有用一级近似。条件 $kT \ll \epsilon_F$ 满足 $\lambda^3 v^{-1} \gg 1$，而且在式(26)的范围内要求：

$$Pv \simeq \frac{2}{5}\epsilon_F\left[1 + \frac{5\pi^2}{12}\left(\frac{kT}{\epsilon_F}\right)^2 + \cdots\right], \ 即 \lim_{\frac{kT}{\epsilon_F}\to 0} Pv^{\frac{5}{3}} = \frac{2}{5}(3\pi^2)^{\frac{2}{3}}\frac{\hbar^2}{2m} \tag{31}$$

因此，在确定的密度下，当 $T\to 0$，压强严格地趋向于一个正的常数，这与 $T\to 0$ 时经典理想气体的表现形成了鲜明对比，参见式(27)，因而意味着 $P\to 0$。

关于这一点，宇宙天体能提供的例子不多，典型代表如白矮星和中子星。白矮星具有类似于太阳的温度，即中心温度在 10^7K 到 10^8K 之间，质量与太阳在同一个数量级，但密度却非常高，约为太阳的 10^6 至 10^7 倍。这些星体的所有氢燃料已经燃尽，因此，它们由完全电离的氦原子构成。在对白矮星的构成成分及形成条件这些假设下，我们计算出电子气体的密度，那么由 $T_F = \epsilon_F/K$ 可知，与式(28)中的费米能 ϵ_F 对应的 T_F 的结果为 $T_F \simeq 10^{11}K$。因此 $T \ll T_F$，且与假设一致——R. H. 福勒(R. H. Fowler)已于 1926 年做出了这个假设[Fowler, 1926]——白矮星内的电子系综可以被描述为简并费米气体，正是这种气体中巨大的压强使得星体不会在引力的作用下崩塌。然而真实的情况是，在这样的密度和压力下，电子必须被看作是相对论性的，即用 $\epsilon = \sqrt{(pc)^2 + (mc^2)^2}$ 代替 $\epsilon = p^2/2m$。这就造成了一系列的分析困难，其中之一就是式(31)中 v 的量值从

5/3 变为 4/3。在他对这种效应的计算过程中，钱德拉塞卡[1931a]注意到，由于重力造成的压强由星体的质量决定，因此如果质量太大的话后者就会崩塌。事实上他计算出了这个临界质量 $M_{\max}$ 为：

$$M_{\max} \simeq (3\pi)^{\frac{1}{2}} \left(\frac{\hbar c}{G}\right)^{3/2} (\mu m_N)^{-2} \simeq 1.4 M_{\odot} \tag{32}$$

这里(cgs 单位制)$\hbar = h/2\pi$，$h \simeq 6.62 \times 10^{-27}$ ergs · cm 是普朗克常数，$c \simeq 3 \times 10^{10}$cm/s是光速，$G \simeq 6.67 \times 10^{8}$dyn · cm^2 · g^{-2}是牛顿的万有引力常数，$m_N \simeq 1.66 \times 10^{-24}$g，$\mu$ 是每个电子的核子数。这里 $\mu = 2$ 是因为星体被假设已经用尽了氢原料，是由$^4_2$He 构成的。最后，为了在天文学单位下简化这个结果，$M_{\odot} \simeq 1.99 \times 10^{33}$g 取的是太阳的质量。今天天文学家将这个最大质量 $M_{\max}$ 称为钱德拉塞卡极限(数学家会说是一个“边界”)。

钱德拉塞卡的原始推导在数学上是正确的，但是多少有些烦琐，直到 1932 年末，朗道[1932]给出了一种更基本的论证。此外，在得知发现中子的消息后，他应用上面的公式推测出存在中子星。

包括这些预测在内的 QSP 的早期“成就”只是在事后看来才能明白是合理的。当它们出现在 20 世纪 30 年代初时，这些理论及其结果引起了不小的轰动，其中爱丁顿(Eddington)，一个权威的天文学家，在谈到归谬法时要求加入一个当时还未知的基本理论，对他而言，一个大质量恒星($M > M_{\max}$)塌缩为一个黑洞是异端邪说，而且他直言不讳地谈论这些。爱丁顿对钱德拉塞卡这个年轻学者的强烈抨击并没有使得他公开认错，直到 20 世纪 50 年代末[Chandrasekhar, 1958 年]和 60 年代初，虽然钱德拉塞卡决定转向研究其他天文学问题，但在那时他的预测和朗道的预测最终在观测上得到了确认。

钱德拉塞卡—爱丁顿争论如何解决的讨论，参见[Shapiro and Teukolsky(夏皮罗和特科斯基)，1983]，其题目已经说明了钱德拉塞卡最终被证明完全是正确的。关于钱德拉塞卡边界最初的、非技术化的物理学表述请参见[Thorne(索恩)，1994，Chapter 4]；关于中子星的描述，在[Thorne，1994，Chapter 5]中也有大量涉及；基础的技术支持，参见[Weinberg(温伯格)，1972，Chapter 11]。

3. 公理化修整

一般来说，有两个原因推动了公理化过程。第一个原因是为了寻找隐藏在身体内的偶而会呈现出的灵魂——有些人可能会说是骨架，是一种提炼过程。第二个原因是当一个理论面临着越来越无法克服的局限性时，它需要发生根本性的变化。在本节中，我将追溯这两个推动公理化进展的原因，在这里我想起了[Segal(西格尔)，1990]描述过的必要的矛盾。

一个有趣的巧合是：20 世纪 30 年代初，几乎同时出现了——尽管是相互独立的——QSP 的两部分内容的公理化：冯·诺伊曼(von Neumann)论文中的量子力学[Von Neumann，1932c]和柯尔莫哥洛夫论文中的《统计过程》(*Statistics*)，又名《概率与随机过程》(*Probability and stochastic processes*)。见[Kolmogorov，1933]。由于这两部分内容在本书的其他章节中会讨论，因此这里只提几句话就足够了。

3.1 柯尔莫哥洛夫和冯·诺伊曼的形式主义

概括地说，柯尔莫哥洛夫关于概率的句法起始于对测度理论的一种开创性的描述。给定一个三元组$\{\Omega, \mathcal{E}, \mu\}$，其中$\mathcal{E}$是集合$\Omega$的可测量子集的一个$\sigma$代数，$\mu$是一个可数可加性函数：

$$\mu:\ E\in\mathcal{E}\mapsto\mu(E)\in\mathrm{IR}^+,\ \text{其中}\ \mu(\Omega)=1 \tag{33}$$

即μ是$\{\Omega, \mathcal{E}\}$的一个概率测度。μ自然展开为所有本质有界函数$A:\Omega\to\mathbb{C}$的代数$\mathcal{A}=\mathcal{L}^\infty(\Omega, \mathcal{E}, \mu)$上的一个泛函：

$$\mu:A\in\mathcal{A}\mapsto\mu(A)=\iint_\Omega d\mu(\omega)A(\omega)\in\mathbb{C} \tag{34}$$

后面，我会把这个展开看作一个经典态。

同样，冯·诺伊曼的说法涉及一个三元组$\{\mathcal{H}, \mathcal{P}, \psi\}$，其中$\mathcal{P}$是希尔伯特空间$\mathcal{H}$上的所有闭子空间的正交模格，$\psi$是一个可数可加性正函数：

$$\left.\begin{array}{l}\psi:P\in\mathcal{P}\mapsto\psi(P)\in\mathrm{IR}^+\quad\text{有}\quad\psi(I)=1\quad\text{和}\\ \psi(\sum_n P_n)=\sum_n\psi(P_n)\quad\forall\{P_n\}\subset\mathcal{P}\quad\text{使得}\ n\neq m\models P_n\perp P_m\end{array}\right\} \tag{35}$$

我将会把这类函数 ψ 看作是一个量子态。特别地，盖里森(Gleason)定理断言——对此下面会有完整的陈述——对每个量子态 ψ 都存在一个密度算符，即单位迹的一个正算符 ρ，使得 ψ 扩展为从$\mathcal{H}$到其自身的所有有界线性算符的 W^* 代数$\mathcal{B}=\mathcal{B}(\mathcal{H})$：

$$\psi: B\in\mathcal{B}\mapsto\psi(\mathcal{B}) = \mathrm{Tr}\,\rho B\in\mathbb{R} \tag{36}$$

当采用冯·诺伊曼的形式时，我会认为任意闭子空间 $P\subseteq\mathcal{H}$和该子空间上的投影 $P\in\mathcal{B}(\mathcal{H})$是等同的，我会将 ψ 或 ρ 都看作$\mathcal{B}$上的态，我会将 ψ 对$\mathcal{P}$的约束看作一次量子测量。我也会遵循物理学家的定义，将 ρ 看作密度矩阵，从而忽略数学家们对算符和其在特定(正交归一)基上的表达式之间所做的区分。

经典力学的库普曼(Koopman)形式强调了经典和量子领域之间的数学相似性和差异性，参见[Emch and Liu，2002，255，267]。这种形式——实际上是 GNS 结构的前身——将$\{\Omega, \mathcal{E}, \mu\}$与所有关于 μ 的平方可积函数 $\Psi: \omega\in\Omega\mapsto\psi(\omega)\in\mathbb{C}$ 的希尔伯特空间$\mathcal{H}=\mathcal{L}^2(\Omega, \mathcal{E}, \mu)$联系了起来。因此每个元素 $A\in\mathcal{A}=\mathcal{L}^\infty(\Omega, \mathcal{E}, \mu)$可被视为$\mathcal{B}=\mathcal{B}(\mathcal{H})$的元素，即函数 $A: \omega\in\Omega\mapsto A(\omega)\in\mathbb{C}$ 等同于相乘运算 $A: \Psi\in\mathcal{H}\mapsto A\Psi\in\mathcal{H}$，这里$(A\Psi)(\omega)=A(\omega)\Psi(\omega)$。在这种等同性下，$\mathcal{A}$成为$\mathcal{B}$的一个最大阿贝尔 W^* 子代数，而$\mathcal{B}$的中心，即$\{C\in\mathcal{B} \mid \forall B\in\mathcal{B}: [B, C]=0\}$是平凡的，即构成了恒等算符的倍数。还请注意，每一个元素 $B\in\mathcal{B}(\mathcal{H})$可以被看作是由可数可加态占据的所有迹类算符的巴拿赫(Banach)空间$\mathcal{T}(\mathcal{H})$上的一个连续线性泛函，即 $B: T\in\mathcal{T}(\mathcal{H})\mapsto\mathrm{Tr}TB\in\mathbb{C}$，相反每一个$\mathcal{B}(\mathcal{H})$上的规范连续线性泛函都是以这种方式得到的，即$\mathcal{B}(\mathcal{H})$是$\mathcal{T}(\mathcal{H})$的巴拿赫空间对偶，等价地，$\mathcal{T}(\mathcal{H})$是$\mathcal{B}(\mathcal{H})$的预对偶。同样地，$\mathcal{L}^\infty(\Omega, \mathcal{E}, \mu)$的预对偶是$\mathcal{L}^1(\Omega, \mathcal{E}, \mu)$的巴拿赫空间，它由关于 μ 绝对连续的概率分布所占据。

从经典概率角度对量子态 ψ 进行解释需要分别对相容量子事件的每个族$\{P_n\}$应用式(35)。由式(35)和式(36)描述的对象之间的双射等价为盖里森定理的实际内容，参见[Emch and Liu，2002，225]。每个量子态可以被唯一地表示为式(36)的形式，并且每一个密度算符 ρ 通过式(36)定义一个满足式(35)的函数 ψ，即一个量子态 ψ。对于态的字面解释，即经验(常常对比于主观的)解释，前者体现在经典概率理论中，后者体现在量子理论中，参见例如[Jaynes(杰恩斯)，1967；Emch and Liu，2002；Emch，2005]，特别请参看[Uffink，

2006]针对CSP中关于概率解释的理论更替的演变过程做出的讨论。

再次概括地说，我支持“主观的”而不是上述学者的“频率的”解释，即我将一个物理系统的状态——无论是经典的还是量子的，宏观的还是微观的——看作是对我们关于系统所经历过程的知识的完整总结，这种观点在这篇文章的大多数篇幅中符合我的目的。特别是，这种关于量子态的意义能够很好地适应从冯·诺伊曼的量子力学所考虑的有限多自由度的系统，到QSP所考虑的无限多自由度的系统的转变，参见3.4节到6.3节。特别地，冯·诺伊曼的独立、同一系统的粒子束或“系综”[Von Neumann，1932c，note 156]能够描述他同时代的散射实验或原子光谱，我这里选择采用的理解量子态的视角能够更好地描述单个宏观系统——如一杯咖啡或一个测量装置。

3.2 冯·诺伊曼形式下的QSP

冯·诺伊曼形式下的平衡态QSP的核心是下面的一些结果[Von Neumann，1932c]：

定理3。令$\mathcal{H}$为希尔伯特空间，H为作用在$\mathcal{H}$上的自伴算符，使得对所有$\beta>0$，配分函数$Z:=\mathrm{Tr}\exp(-\beta H)$是有限的。并且，当$k>0$时，令$\mathcal{B}(\mathcal{H})$上的任意态$\rho$满足：

$$S[\rho]=-k\mathrm{Tr}\,\rho\log\rho \tag{37}$$

由于H具有离散谱，且具有下界，令ϵ_o为其最小本征值。如果H有上界，令s为其最大本征值，如果没有则为∞。那么，对于任意给定的$\epsilon_o<E<s$，遵从约束条件$\mathrm{Tr}\,\rho H=E$的态$S[\rho]$的最大值为：

$$\rho=Z^{-1}e^{-\beta H}\text{，其中 }Z=\mathrm{Tr}e^{-\beta H} \tag{38}$$

这里β的值是由约束条件E的值确定的。

证明该定理的第一步主要是表明最大值出现于形式为$\rho=\sum_n\lambda_nP_n$的那一类态中，其中$\sum_n\epsilon_nP_n$是H的谱分解。然后，应用拉格朗日乘子进行经典论证就能得到所要的结果，其中拉格朗日乘子涉及变量集$\Lambda=\{\lambda_n\}\subset\mathrm{IR}^+$，即通过确定函数$S[\Lambda]=-k\sum_n\lambda_n\log\lambda_n$的最大值得到结果，这里该函数同时受$\sum_n\lambda_n\epsilon_n=E$和$\sum_n\lambda_n=1$两个条件约束。

请注意，这个变分原理可以被改述为：将式(38)中的态ρ定义为使亥姆霍兹(Helmoholtz)自由能量最小化的态，现在可由单个约束条件 $\mathrm{Tr}\,\rho=1$(即$\sum_n\lambda_n=1$)来实现，这里亥姆霍兹自由能量的定义为$F:=E-TS$，其中E和S与定理中的意义相同，而$\beta=kT$，这里k为玻耳兹曼常数(见下文)。

还要注意的是，无论表示为这两种形式中的任何一种，变分原理都是源于玻耳兹曼和吉布斯的经典统计物理(CSP)的，参见[Uffink，2006]。从概念上讲，冯·诺伊曼 QSP 的结果在下述两个问题上同在 CSP 中的结果是一致的。第一个问题是将S解释为熵。有两种方法可以做到这一点。

(i) 第一种方法是，如同在 CSP 中那样，人们可以计算一些容易控制的模型，如理想气体中的S，并且在考虑的每一个具体案例中指出通过上述定理得到的$S_{\max}$的值与热力学熵的值一致，从而我们就可以将S定义为宏观热物理学中的平衡熵。只有在该阶段，k可以等同于通常的玻耳兹曼常数$k\simeq 1.3810^{-23}$ J/deg。注意该常数的单位，即[能量]/[温度]，对于热力学熵来说是适当的，这里T是积分因子，它使我们能将“热”差η转变为精确微分$dS=\eta/T$。如同在平衡态 CSP / QSP 中可以应用得很好那样，这种等同性也可以在非平衡态的情况下将S解释为熵。

(ii) 对S做出解释的第二种方法是，表明$I(\rho)=-S(p)$是对态ρ的信息内容的一次测量，即找到表达“信息内容”直观概念的经验性意义条件，并表明——取决于一个倍数常数——精确地存在一个S满足这些条件。辛钦(Khinchin)[1957]提供的关于经典概率分布的论证包括(特别是)改进后的一致性公理。这个论证过程由西凌(Thirring)[1983b]转换到了量子情况下。

定理 4。泛函$S[\rho]=k\mathrm{Tr}\,\rho\log\rho$是唯一满足下列条件的泛函：

1. 对ρ来说$S[\rho]$是连续的，从它是ρ的本征值的一个连续函数的意义上来看。

2. 对于希尔伯特空间$\{\mathcal{H}_n\mid n=1, 2, \cdots, N\}$的一个有限集上的每个有限概率分布$P=\{p_n\mid n=1, 2, \cdots, N\}$和每个态的有限集$\{\rho_n\mid n=1, 2, \cdots, N\}$，令$\rho$是通过$\rho=\bigoplus_{n=1}^N p_n\rho_n$定义在$\mathcal{H}=\bigoplus_{n=1}^N\mathcal{H}_n$上

的态。那么我们有 $S[\rho]=S[P]+\sum_{n=1}^{N}p_nS[\rho_n]$，其中 $S[P]$ 为概率分布 P 的辛钦泛函的值。

3. $S\left[\begin{pmatrix}\frac{1}{2} & 0\\ 0 & \frac{1}{2}\end{pmatrix}\right]=k\log 2$。

定理 4 的三个条件中的第一个是明确的，态的一个任意小的变化，会导致它传达的信息发生任意小的改变。第二个条件表示了在一个特定分割类下的细化，而第三个条件是归一化。正如在 CSP 中那样，由这些条件唯一确定的量子信息内容($-S$)，并正式被用于定义量子熵 S。

第二个问题涉及的是定理 3 的概念相关性，对比于[Uffink，1995]，它可以证明变分原理具有特殊用途。在我看来，无论是在经典还是量子情况下，选择态的主观解释，而不是频率解释，是非常自然的。事实上，如果我们想用态来说明我们所拥有的关于系统的知识，那么为 ρ 选择的这个态也不比明确表述的约束条件包含更多的信息。

当定理 3 中的算符 H 被用来表示系统的能量时，态(38)被称为——借用 CSP 中的吉布斯正则平衡态这个术语——常温 $\beta=1/kT$ 下的量子正则平衡态。需要特别注意的是，在薛定谔绘景下，由 H 导致的演化，即：

$$\forall t\in\mathbb{R}:\ \rho(t)=U(t)\rho U(-t)\text{，其中 }U(t)=\exp^{-i\frac{1}{\hbar}Ht} \tag{39}$$

就如预期的那样，当我们希望将能量算子与系统的哈密顿(Homiltonian)量等同起来时，会保持正则平衡态不变。

乍看之下，冯·诺伊曼的形式体系为量子遍历理论的发展提供了一个良好的开端。为了尽可能地保持简单性，我们考虑由标量积 $(X,Y)=\mathrm{Tr}\,X^*Y$ 表示的希尔伯特空间 $\mathcal{L}=\{X\in\mathcal{B}(\mathcal{H})\mid \mathrm{Tr}\,X^*X<\infty\}$。该空间被数学家称为作用于 $\mathcal{H}$ 的希尔伯特—施密特(Hilbert-Schmidt)算符空间。特别地，每个密度矩阵都是 $\mathcal{L}$ 的一个元素，因此该空间也被物理学家称为在 $\mathcal{H}$ 上描述的量子系统的刘维尔(Liouville)空间。将注意力限制在这个空间上是有好处的，式(39)可以延伸为 $\mathcal{L}$ 上的一个幺正操作：

$$V:(t, X)\in \mathbb{R}\times\mathcal{L}\mapsto V(t)[X]=U(t)XU(-t)\in\mathcal{L} \tag{40}$$

以同样的方式，如同连续幺正群的自伴生成算符$\{U(t)\mid t\in\mathbb{R}\}$被称为所考虑的量子系统的哈密顿量那样，连续幺正群的自伴生成算符$\{V(t)\mid t\in\mathbb{R}\}$被称为这个系统的刘维尔量。那么我们就有如下定理。

定理 5。令$\mathcal{H}\in\mathcal{B}$有纯离散谱，即$H$可以被记作$H=\sum_n\epsilon_nP_n$的形式，其中$P_n$是和为$I$的相互正交的投影算符。那么下面的极限存在：

$$E_{erg}[X]=\lim_{T\to\infty}\frac{1}{T}\int_0^T dtV(t)[X],\text{其中 } X\in\mathcal{L} \tag{41}$$

且有$E_{erg}[X]=\sum_n P_nXP_n$ 和 $\forall t\in\mathbb{R}$：$V(T)[E_{erg}[X]]=E_{erg}[X]$。

特别地，密度矩阵ρ的遍历平均$E_{erg}[\rho]$存在，它也是一个密度矩阵，而且具有时间演化不变性。

尝试着将定理5看作是伯克霍夫(Brikhoff)[1931]或冯·诺伊曼[1932a]经典遍历定理的适当量子版本是很有意义的。事实上，当分别从由下面两式定义的(可数可加)“态”来解读这些经典定理和定理5时，它们的结论是类似的：

· $A\in\mathcal{L}^\infty(\Omega,\mu)\mapsto\int_\Omega d\mu fA\in\mathbb{C}$，其中$f\in\mathcal{L}^1(\Omega,\mu)$，$f$为正且$f$由$\int_\Omega d\mu f=1$来归一化(经典情况)。

· $A\in\mathcal{B}(\mathcal{H})\mapsto\mathrm{Tr}\rho A\in\mathbb{C}$，其中$\rho$为密度矩阵，即它是一个正迹类算符，且$\rho$由$\mathrm{Tr}\rho=1$来归一化(量子情形)。

且对应的时间平均也相似。

需要注意的是，在经典定理之后，人们通常会得到一个涉及(准)遍历假说的推论，并继续进行一些有关CSP基础结果的讨论——参见[Uffink, 2006, Section 6.1]——而我在这里则不打算尝试沿着这条路线继续去探讨，从下面的定理7和定理8的角度看，定理5关于哈密顿量H所做的假设会给QSP的要求蒙上一层阴影。关于量子遍历定理能否更好地适应QSP所要求的内容，请参见下面的定理25。

尽管如此，可以给出关于定理5的两点有意义的注释。

(i)如果在这个定理中，H是非退化的，也就是说，如果$\forall n$：$\dim P_n=1$，那么$E_{erg}[\rho]$与下式一致：

$$Q_o[\rho] = \sum_n \mathrm{Tr}(\rho P_n)P_n = \sum_n (\rho\Psi_n,\Psi_n)P_n \tag{42}$$

这里 $P_n\Psi_n = \Psi_n$，其中 $(\Psi_n,\ \Psi_m) = \delta_{mn}$，因此这里的 $Q_O[\rho]$ 是由冯·诺伊曼量子测量过程导致的密度矩阵[Von Neumann，1932c，351]，也可参见下文第6.3节。特别地，如果 ρ 是一个纯态，即它是矢量 $\Psi = \sum_n c_n\Psi_n$ 上的一个投影算符 P_Ψ，那么 $Q_o[P_\Psi] = \sum_n |c_n|^2 P_n$ 就已经失去了编码于系数为 c_n 的相对相位中所有的信息。

（ii）在论文[Von Neumann，1932c，380 ff]中，冯·诺伊曼表明，一个态的熵 S 不会减少——在一般情况下，会增加——作为测量的结果，在幺正演化式(40)下它是恒定的。因此，他认为，由

$$S[Q_O[\rho]] \geqslant S[\rho] \tag{43}$$

可知，量子测量一般是一个不可逆过程。那么类似地，编码在一个（非简并）密度矩阵 ρ 中的信息只可以减少，作为对其进行时间平均的结果，这确实是一个合理的属性。

然而，当定理5被视为量子遍历理论的一个萌芽时，QSP 中的单调不可逆性则显得令人难以理解，这将是下一小节论证的内容。

3.3　超越冯·诺伊曼形式的一些原因

非平衡 QSP 不得不面对的一些问题由一个简单的自旋—晶格模型得以说明，我在头脑中产生的这个模型是源于一个真实的实验，即所谓的核自由感应弛豫实验，参见［Emch and Liu，2002，Section 15.3］。

该系统由 N 个相互作用自旋 $\{\sigma_k = (\sigma_k^x,\ \sigma_k^y,\ \sigma_k^u) \mid k = 1,\ \cdots,\ N\}$ 的一个线性链构成，这里 N 为偶数（并且在下文中的特定意义下，N 很大），并令

$$\sigma_k^x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},\ \sigma_k^y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix},\ \sigma_k^z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{44}$$

为作用于 $\mathcal{H}_k \simeq C^2$ 的泡利矩阵。那么系统的希尔伯特空间为 $\mathcal{H} = \otimes_k \mathcal{H}_k \simeq \mathbb{C}^{2N}$。在这个链中，$k$ 和 $k+n$ 位置上的两个自旋发生相互作用，其相互作用能为 $-J_n\sigma_k^z\sigma_{k+n}^z$，其中 $J_n > 0$，这个能量较低，是因为这些自旋的 z 分量是平行的，而不是反平行的。整个系统被放置在一个 z 方向上的均匀磁场 B 中。总哈密顿量为：

$$H_N = -B\sum_{k=1}^{N}\sigma_k^z - \sum_{k=1}^{N}\sum_{n=1}^{N/2} J_n\sigma_k^z\sigma_{k+n}^z,\ \text{其中}\ J_n = 2^{-n}J_o > 0 \tag{45}$$

该系统的初始态为：

$$\rho_N = Z_N^{-1}\,\mathrm{e}^{-\beta B\sum_{k=1}^{N}\sigma_k^z} \tag{46}$$

其中 $Z_N = \mathrm{Tr}\,\exp^{-\beta B\sum_{k=1}^{N}\sigma_k^z}$，对于三个"宏观"可观测量：

$$S_N^\alpha = \frac{1}{N}\sum_{k=1}^{N}\sigma_k^\alpha \tag{47}$$

其中 α 表示 x,y,z。

我们容易从式(39)和式(40)计算得到下式，其中 $H = H_N$ 由式(45)给出：

$$\left.\begin{aligned} \mathrm{Tr}(V_N(t)[\rho_N]S_N^x) &= \mathrm{Tr}(\rho_N S_N^x)\cos(2Bt)f_N(t) \\ \mathrm{Tr}(V_N(t)[\rho_N]S_N^y) &= \mathrm{Tr}(\rho_N S_N^y)\sin(2Bt)f_N(t) \\ \mathrm{Tr}(V_N(t)[\rho_N]S_N^z) &= \mathrm{Tr}(\rho_N S_N^z) \end{aligned}\right\} \tag{48}$$

这里有：

$$\left.\begin{aligned} &f_N(t) = f(t)/W_N(t),\ \text{其中} \\ &f(t) = \left[\frac{\sin(J_o t)}{J_o t}\right]^2 \quad \text{和} \quad W_N(t) = \left[\frac{\sin(2^{-N/2}J_o t)}{2^{-N/2}J_o t}\right]^2 \end{aligned}\right\} \tag{49}$$

备注 6。

1. 为了讨论假定存在的模型的不可逆性，磁场 B 沿 z 方向的磁化强度的(保守)拉莫尔进动$\{\cos(Bt),\ \sin(Bt)\}$应该很小或为零。

2. 如果支持模型的"不可逆性"观点，我们首先会注意到：

$$\forall t,\ \text{其中}\ |t| \ll T_N = 2^{N/2}\pi J_o^{-1}:\ f_N(t) \simeq f(t) \tag{50}$$

因此 $|\mathrm{Tr}V_N(t)[\rho]S_N^\alpha|$ 的衰变是由 t^{-2} 决定的。因此，在这段时间内，磁化强度式(48)具有明显的向平衡态演化的趋势。

3. 然而，相悖于模型将会表现出的不可逆地趋向平衡态的这种论断，我们观察到：

$$\lim_{t \to T_N} f_N(t) = 1 = f_N(0) \tag{51}$$

因此，在整个漫长的过程中，系统是周期性的。所以量子模型似乎证实了经典模型的策梅罗回归异议(Wiederkehreinwand)，关于后者请参阅[Uffink, 2006, Section 4.5]。

4. 然而，可取之处在于周期 T_N 随着系统尺寸 N 的增大而指数性增大，参见式(50)。这种指数性增大的行为在 CSP 中也出现过，在下文 6.1 节中将会简要提到的埃伦费斯特的狗蚤模型中我们将会看到这种现象。因此，现代伽利略会提出他自己的辛普利西奥论证，即对于宏观的大系统，在接近 T_N 之前，就设定好这种无须计入的扰动，从而掩盖这种难以修改的周期性。请将此与玻耳兹曼对策梅罗异议的回复作比较，参见[Uffink，2006，Section 4.5]。

5. 在评估了这种意义之后，萨尔维亚蒂(Salviati)将会引用一条来历不明的禁律的现代版本，其大意为“极限不可交换”，因为：

$$\lim_{N\to\infty}\lim_{t\to\infty} f_N(t)\text{不存在，但}\lim_{t\to\infty}\lim_{N\to\infty} f_N(t)\text{存在，其结果为}0 \tag{52}$$

6. 本模型还提出了经典洛施密特(Loschmidt)可逆性异议的一个量子版本(Umkehreinwand)——参见［Uffink，2006，Section 4.3］——这里有：

$$f_N(-t)=f_N(t)\quad\text{甚至}f(-t)=f(t) \tag{53}$$

我们证实了经典杰纳斯断言，通过说明溯测(postdiction)的安全性与预测的安全性是相同的。因此，这个模型表明，如果可逆性异议确实是针对 QSP 的一个真正的异议——我不相信它真的是——那么热力学极限将无法回避它，而上述备注(4)和(5)则说明了它是怎么回应回归异议的。

7. 最后，本模型——不是关于上述考虑的特殊实验的一个模型，而是为了解释转移系数而设计的，体现 QSP 中朝向平衡态的演化趋势的一个模型——存在的一个严重的不足之处是，即使在 $N\to\infty$ 的极限下，演化仍受时间上可逆的能量定律支配，而不是受指数定律支配，因为需要用该模型的演化来描述宏观情况中会遇到的那类行为，即牛顿冷却定律、傅里叶热方程或者更普遍的所有具有线性系数的宏观微分输送方程所描述的行为。

该模型明确地说明了关于 QSP 的冯·诺伊曼形式的一些本质局限，表现为以下两个一般性结果。定理 3 和 5 的主要假设，即哈密顿算符 H 具有离散谱，虽然在涉及 QSP 平衡时看似是没问题的(实际上是必然的)，但当我们试图将该形式推广到非平衡态的情况时，则会出现一个潜在的灾难性后果：正如我们将会看到的，经典反驳会转移到量子领域。

第一个结果是策梅罗经典回归定理的量子版本。用数学语言精确地回顾就是贝斯克维奇(Besicovitch)的标准原文[Besicovitch，1954]，即函数 f：$t\in\mathbb{R}$

$\mapsto f(t)\in\mathbb{C}$ 在哈拉尔德·玻尔的意义上被认为是近似周期性的，如果对于 T 的无限多个值(这些值分布在整个实数线上，以这样的方式保证实数线上不存在任意长的空区间)，$f(t+T)$(以一个任意的准确度)约等于 $f(t)$ 的话。

定理 7。 如果哈密顿量 $H=H^*\in\mathcal{B}(\mathcal{H})$ 有纯离散谱，即 $H=\sum_n \epsilon_n P_n$；且如果 $\{V(t)\mid t\in\mathbb{R}\}$ 是刘维尔空间 $\mathcal{L}$ 上式(40)定义的幺正行为，那么：

$$\forall X,\ Y\in\mathcal{L}:\ f_{X,Y}(t)=\mathrm{Tr}(V(t)[X]Y) \tag{54}$$

在 H. 玻尔的意义上是 t 时刻的一个近似周期函数。

证明。函数 $f(t)$ 是一个傅里叶级数 $\sum_{n,m} a_{n,m}\exp^{-\frac{i}{\hbar}(\epsilon_n-\epsilon_m)t}$，其中 $a_{n,m}=\mathrm{Tr}(P_nXP_my)$。由施瓦茨(Schwartz)不等式在 $\mathcal{L}$：$\sum_{n,m}|a_{n,m}|^2$ 上收敛可知——参见[Besicovitch，1954]——在 t 时刻 f 是 H. 玻尔意义上的一个近似周期函数。

那么，有人可能会试图通过假设——在量子力学的冯·诺伊曼形式下，给定 $\dim\mathcal{H}=\infty$，这是完全被允许的——哈密顿量的谱是纯连续的以便摆脱周期性。然而，从 QSP 的视角来看，这种方法会造成下面的新困难，即遍历的态可能不是可数可加的，即可能不能用密度矩阵来表示。

为了描述这种现象，考虑由通常的算符标准表示的巴拿赫空间 $\mathcal{B}=\mathcal{B}(\mathcal{H})$，并用 $\mathcal{B}^*$ 表示二元空间，即 $\mathcal{B}$ 上所有连续线性泛函的巴拿赫空间。那么：

$$\mathcal{B}^*=\mathcal{A}^*\oplus\mathcal{A}^{\perp} \tag{55}$$

这里：

i) $\mathcal{A}$ 是 $\mathcal{H}$ 上紧致算子的空间，即 $\mathcal{A}=\{A\in\mathcal{B}(\mathcal{H})\mid \Psi_n\rightharpoonup\Psi\Rightarrow A\Psi_n\to A\Psi\}$，$\rightharpoonup$ 和 $\to$ 分别表示 $\mathcal{H}$ 上的弱收敛和强收敛。当 * 代数 $\mathcal{A}$ 由继承自 $\mathcal{B}(\mathcal{H})$ 的算符标准表示时，$\mathcal{A}$ 在 $\mathcal{B}(\mathcal{H})$ 内是闭合的，因此它是自身的巴拿赫空间。事实上 $\mathcal{A}$ 是 $\mathcal{B}$ 的唯一非平凡闭合双向 * 理想化。

ii) 对于每一个 $\varphi\in\mathcal{A}^*$，$\mathcal{A}$ 的二元化，存在一个唯一的迹类算符 $R\in\mathcal{T}=\{B\in\mathcal{B}\mid \mathrm{Tr}(B^*B)^{\frac{1}{2}}<\infty\}$，使得 $\forall A\in\mathcal{A}$：$\varphi(A)=\mathrm{Tr}(RA)$。特别地，对于 $\mathcal{A}$ 上使得 $\sup_{A\in\mathcal{A},\|A\|\leq 1}\|\psi(A)\|=1$ 的每个正连续线性泛函 ψ，对应地存在一个唯一的密度矩阵，反之亦然。

iii) $\mathcal{A}^{\perp}=\{\varphi\in\mathcal{B}\mid A\in\mathcal{A}\Rightarrow\varphi(A)=0\}$。

请注意，当且仅当 $\mathcal{H}$ 是无限维时，$\mathcal{T}\subseteq\mathcal{L}\subseteq\mathcal{A}\subseteq\mathcal{B}$ 中的每个包含符号可严格

取得，当我们想要避免回归时，需要满足一个条件，因为 dim $\mathcal{H}<\infty$ 显然要求 H 的谱是纯离散的，因此定理 7 适用。

现在，我们可以精确地阐述上面提到的在可数可加态情况下描述遍历态的困难。

定理 8。令 $H\in\mathcal{B}(\mathcal{H})$是作用于$\mathcal{H}$的任意强连续幺正群$\{U(t)\mid t\in R\}$的自伴生成算符；令 $\rho\in\mathcal{T}\mapsto\rho(t)=U(t)\rho U(-t)\in\mathcal{T}$描述任意密度矩阵 ρ 的演化，其中 t 属于 IR；进一步，令 ψ_t 表示相应的可加(可数)态 ψ_t：$B\in\mathcal{B}(\mathcal{H})\mapsto\psi_t(B)=\mathrm{Tr}(\rho(t)B)\in\mathbb{C}$。那么，就有：

a. 对于每个紧致可观测量 $A\in\mathcal{A}$，遍历极限

$$\lim_{T\to\infty}\frac{1}{T}\int_o^T dt\ \psi_t(A) \tag{56}$$

存在，并且定义$\mathcal{A}$上的一个正线性泛函 $E_\infty[\psi]$。

b. 此外，如果 H 的谱是纯连续的，那么 $E_\infty[\psi]$不能扩展为$\mathcal{B}(\mathcal{H})$上的一个可数可加态。

证明　对于(a)部分，经济性的策略是利用两个密度定理，即：(i)当$\mathcal{L}$由希尔伯特—施密特标准表示时，它包含$\mathcal{T}$作为其致密子空间；(ii)当$\mathcal{A}$由继承自$\mathcal{B}(\mathcal{H})$的算符标准表示时，它包含$\mathcal{L}$作为其致密子空间。因此，我们可以唯一地将演化从$\mathcal{T}(\mathcal{H})$上升为唯一的希尔伯特空间$\mathcal{L}$中的幺正演化，在其中我们就可以应用经典遍历定理——[von Neumann，1932a]或[Emch and Liu，2002；Uffink，2006]——来断言对于任意的$(X, Y)\in\mathcal{L}\times\mathcal{L}$，特别是对于任意的$(\rho, A)\in\mathcal{T}\times\mathcal{L}$存在遍历极限。回顾对偶性$\mathcal{A}=\mathcal{T}^*$，那么遍历结果就可由连续性从$\mathcal{T}\times\mathcal{L}$扩展到$(\rho, A)\in\mathcal{T}\times\mathcal{A}$。

为了证明(b)部分，我们注意到，一方面，对每个 $A\in\mathcal{A}$，这个极限是由 $E_o[\psi](A)=\mathrm{Tr}(\sum_n(P_n\rho P_n A))$给定的，其中$\{P_n\}$是 H 的谱族中对应于非连续跃度的所有投影的集合。因此，当 H 具有连续谱时，这个集合为空集，且 $\forall A\in\mathcal{A}$：$E_o[\psi](A)=0$。另一方面，$E_o[\psi]$一定能扩展到比$\mathcal{A}$更大的空间中，例如对于任意 $X\in\{H\}'$，遍历极限明显存在，即对所有的有界可观测量，它是运动常数，特别地，$E[\psi_o](I)=1$。因此，即使 $E[\psi_o]$可以扩展为$\mathcal{B}$上的一个态，那么这个态就属于$\mathcal{A}^\perp$，且因此这个态在直和分解式(55)中就没有可数可

加部分。

该定理的证明过程表明，针对任意哈密顿量可以提出同样的异议，只要该哈密顿量的谱包含哪怕一个连续性区间。将哈密顿量看作是自伴的但不像上面那样是有界的，则只会增加更多的技术性问题，不能为定理8揭示出的基本极限提供一个解。

因此，针对QSP的冯·诺伊曼形式导致非平衡态QSP陷入一个两难境地：要么演化是近似周期的，要么遍历态不是可数可加的。特别地，非平衡态不能接近于由密度矩阵描述的渐近态。

使这个糟糕的状况变得更糟的是，泽(Zeh)发现——无可否认，在冯·诺伊曼的著作出版之后以及本节讨论的相关观点尚未出现之前的很长一段时期——理解孤立量子系统这个观念具有严重的经验性困难[Zeh，1970；Wigner(维格纳)，1984]。确定性会减弱吗？参见[Prigogine(普里高津)，1997]。泽的初始可观测量导致了退相干概念的发展，参见[Landsman(兰兹曼)，2006，Section 7.1]。在下文第6节中，我将简单介绍这一问题，以及与之相关的一些问题。

即使在平衡态QSP中，冯·诺伊曼的形式体系也受到了质疑：它不能解释热力学相位共存的现象，对这种异议的回应，请参阅下文第5.7节。

相应于这些问题来说，需要解决的一个基本问题是：冯·诺伊曼的形式体系不足以描述典型多体系统，这种系统将无限多的自由度带到了自然图景中。我们将在下一小节中讨论相应的补救措施。

3.4 哈格—卡斯特勒(Haag-Kastler)的公理系统和武田(Takeda)的归纳极限

本小节概述一个为了有效解决QSP中非相对论多体问题而提出的形式化系统。这个形式化系统的诞生是在对相对论量子场论[QFT]中出现的危机做出分析之后，学者们给出的公理化的回应[van Hove(范·霍夫)，1952；Friedrichs(弗里德里克斯)，1953；Wightman and Schweber(怀特曼和施威伯)，1955]，并由哈格[1955]盖棺定论，例如著名的哈格定理，以及[Barton(巴顿)，1963，Section 14]、[Streater and Wightman(斯蒂特和怀特曼)，1964，Section 9.5]和/或[Emch，1972a，Section 3.d]。这里所述的代数公理化具有足够多的基本细节，足以避开人们有时候会给它贴上的“专制”标签。

这里主要的思想是考虑可无限扩展系统的局域结构。在最初的计划中，哈格和卡斯特勒[1964]提到公理化 QFT 中的几个范例，参见［Haag(哈格)，1959a；Haag and Schroer(哈格和施罗尔)，1962]以及[Haag，1959b]。在日内瓦举行的研讨会上，我第一次听说了这种代数化方法，当时荒木(Araki)提到了他在苏黎世所做的报告[Araki，1961/2]中的一些相关方面的论述。这里还要提到的是西格尔关于代数化方法的早期主张[Segal，1947]。

本小节可分为两个互补的部分：第一部分对其一般结构进行了描述；第二部分则用一维量子自旋—晶格这个例子呈现了这种结构。

第一部分：基本结构

我们首先在空间中选择可有限扩展的区域 Λ 上的一个吸收定向网$\mathcal{F}$，通常情况下，该空间为相对论 QFT 研究领域的闵可夫斯基(Minkowski)空间$\mathbb{M}^{n+1}$，或非相对论 QSP 研究领域的欧氏空间 $\mathbb{R}^n$ 或晶格$\mathbb{Z}^n$。物理上感兴趣的案例是 $n=3$ 的情况，但研究模型通常构造在 $n=1$，2 的情况下。回想一下，定向网是一个部分有序集，这里的顺序关系是指通常的集合理论中的包含关系，因此，对于$\mathcal{F}$中的每一对元素 Λ_1，Λ_2，至少存在一个元素 $\Lambda\in\mathcal{F}$使得 $\Lambda_1\subseteq\Lambda$ 且 $\Lambda_2\subseteq\Lambda$。我们说这个网是吸收性的，意思是说，对于空间中的每一个点 x，至少存在一个元素 $\Lambda\in\mathcal{F}$使得 $x\in\Lambda$。符号 $\Lambda_1\bowtie\Lambda_2$ 将被用来表示两个区域 Λ_1 和 Λ_2，它们是因果不相交的，即在 QFT 中，这些区域是彼此类空的；在非相对论 QSP 中，它们在集合理论的意义上是不相交的，即 $\Lambda_1\cap\Lambda_2=\emptyset$。$G$ 表示空间中刚体运动的群，即针对$\mathbb{M}^{n+1}$的非齐次洛伦兹(Lorentz)群；或针对 $\mathbb{R}^n$ 的欧几里得(Euclid)群；或针对$\mathbb{Z}^n$ 的格变换群。

其次，对每个 $\Lambda\in\mathcal{F}$，我们分配一个 C^* 代数$\mathcal{A}_\Lambda$。不失一般性地，我们可以假设，$\mathcal{A}_\Lambda$有一个单位元 I_Λ。这种分配遵循以下三个假设。

假设 9(保序性)。当 $\Lambda_1\in\mathcal{F}$和 $\Lambda_2\in\mathcal{F}$满足 $\Lambda_1\subseteq\Lambda_2$ 时，可以给定一个单射*同态 i_{21}：$\mathcal{A}_{\Lambda_1}\to\mathcal{A}_{\Lambda_2}$，使得：

1. $i_{21}(I_{\Lambda_1})=I_{\Lambda_2}$；

2. $\Lambda_1\subseteq\Lambda_2\subseteq\Lambda_3\Rightarrow i_{32}\circ i_{21}=i_{31}$。

武田[1955]证明了下面的结果。

定理 10。设$\mathcal{F}$是一个定向网，且$\{\mathcal{A}_\Lambda\mid\Lambda\in\mathcal{F}\}$满足保序性假设。那么则存

在一个具有单位元 I 的 C^* 代数$\mathcal{A}$，以及一个单射 * 同态$\{i_\Lambda: \mathcal{A}_\Lambda \to \mathcal{A} \mid \Lambda \in \mathcal{F}\}$的族，使得：

1. $\forall \Lambda \in \mathcal{F}$: $i_\Lambda(I_\Lambda) = I$;
2. $\Lambda_1 \subseteq \Lambda_2 \Rightarrow i_{\Lambda_1}(\mathcal{A}_{\Lambda_1}) \subseteq i_{\Lambda_2}(\mathcal{A}_{\Lambda_2})$;
3. $\cup_{\Lambda \in \mathcal{F}} i_\Lambda(\mathcal{A}_\Lambda)$是$\mathcal{A}$的一个规范稠密子*代数。

C^*代数$\mathcal{A}$被称为网$\{\mathcal{A}_\Lambda \mid \Lambda \in \mathcal{F}\}$的 C^* 归纳极限。以下我们将使用符号：

$$\mathcal{A}_o := \bigcup_{\Lambda \in \mathcal{F}} i_\Lambda(\mathcal{A}_\Lambda) \text{和} \mathcal{A} = \overline{\mathcal{A}_o}^{n}$$

假设 11(局域交换性)。当 Λ_1，$\Lambda_2 \in \mathcal{F}$满足 $\Lambda_1 \bowtie \Lambda_2$，且 $\Lambda_3 \in \mathcal{F}$时，则 $\Lambda_1 \subseteq \Lambda_3$ 和 $\Lambda_2 \subseteq \Lambda_3$：

$$A_1 \in \mathcal{A}_{\Lambda_1} \text{且} A_2 \in \mathcal{A}_{\Lambda_2} \Rightarrow i_{31}(A_1) i_{32}(A_2) = i_{32}(A_2) i_{31}(A_1)$$

我们可以直接得到下面的结果。

推论 12。如果 $\Lambda_1 \bowtie \Lambda_2$，那么：

$$A_1 \in \mathcal{A}_{\Lambda_1} \text{和} A_2 \in \mathcal{A}_{\Lambda_2} \Rightarrow i_{\Lambda_1}(A_1) i_{\Lambda_2}(A_2) = i_{\Lambda_2}(A_2) i_{\Lambda_1}(A_1)$$

对于这里考虑的 QSP 来说，将上面的假设简单地看作局域性假设是语言上的一种无害的滥用。

假设 13(协方差)。给定操作 ν：$(g, A) \in G \times \mathcal{A}_o \to \nu_g[A] \in \mathcal{A}_o$使得对于每个区域 $\Lambda \in \mathcal{F}$，ν_g 引起$\mathcal{A}_\Lambda$和$\mathcal{A}_{g[\Lambda]}$之间的一个*同构，这里 $g[\Lambda]$表示点变换 g 下区域 Λ 的图像。

应用定理 10，这就可被提升到$\mathcal{A}$，即 $\nu_g[i_\Lambda(\mathcal{A}_\Lambda)] = i_{g[\Lambda]}(\mathcal{A}_{g[\Lambda]})$。

推论 14。操作 G 由连续性可扩展为一个模连续群表征 ν：$g \in G \to \mathrm{Aut}(\mathcal{A})$。

定义 15。在上面的符号体系中，$\mathcal{A}_o$被称为局域可观测量的代数，$\mathcal{A}$被称为准局域可观测量的代数。此外，如果 φ 是$\mathcal{A}$上的一个态，使得$\forall g \in G$: $\varphi \circ \nu_g = \varphi$，令 π_φ 为相应的 GNS 表征。冯·诺伊曼代数$\mathcal{N}_\varphi = \pi_\varphi(\mathcal{A})''$被称为相对于态 φ 的全局可观测量的代数。

需要注意的是准局域可观测量涉及标准极限，因此，它们是可以抽象定义的一般代数对象，即不需要涉及任何特定的希尔伯特空间表征。相反，不是准局域的全局性可观测量则涉及弱操作极限，从而依赖于取得这些极限的希尔伯特空间表示。在 QSP 中，这些可观测量通过 GNS 结构，取决于定义它们时系

统所处的物理状况，即取决于它们被关注的态。在3.5中我将详细讨论这方面的理论，这对于处理相变将是必不可少的，因为它们涉及可观测量的平均值，例如铁磁体的自发磁化，见下文第5.7节。

第二部分：可观测量代数网格的一个具体例子

这个例子呈现了无限量子自旋—晶格系统的可观测量代数结构，该结构是在上述第3.3节讨论的有限系统的热力学极限下得到的。考虑在每个节点(或“位点”)具有量子$\frac{1}{2}$自旋的一个无限大1维晶格$\mathbb{Z}$，因此具有复数元的2×2矩阵的C^*代数$\mathcal{M}(2,\mathbb{C})$的一个副本$\mathcal{A}_k$与每个位点$k\in Z$相关，即$\mathcal{A}_k$由泡利矩阵式(44)所产生，即由对应于位于位点$k$的$\frac{1}{2}$自旋的三个组成部分的三个可观测量生成。

现在令$\mathcal{F}$为所有有限子集$\Lambda\subset Z$的网。对于它们中的每一个子集来说，Λ是与“局域”C^*代数$\mathcal{A}_\Lambda=\otimes_{k\in\Lambda}\mathcal{A}_k$相关的，该$C^*$代数是$\mathcal{M}(2^{|\Lambda|},\mathbb{C})$的一个副本，其中$|\Lambda|$表示$\Lambda$中位点的数量。

现在令Λ_1和Λ_2为两个有限区域，且$\Lambda_1\subseteq\Lambda_2$。$\mathcal{A}_{\Lambda_1}$到$\mathcal{A}_{\Lambda_2}$的一个单射*同态由线性关系从对单项式的限制中得到，即：

$$i_{21}(A_1\otimes A_2\otimes\cdots\otimes A_{|\Lambda_1|})=B_1\otimes B_2\otimes\cdots\otimes B_{|\Lambda_2|}$$

$$\text{以及}\ \forall k\in\Lambda_2:\ B_k=\begin{cases}A_k\cdots\cdots\cdots\cdots & \text{如}\ k\in\Lambda_1\\ I_k & \text{如}\ k\notin\Lambda_1\end{cases}$$

这些包含关系满足假设9。

这里，当$\Lambda_1\cap\Lambda_2=\emptyset$时，两个有限区域的关系为$\Lambda_1\bowtie\Lambda_2$。由于当这两个位点不同时，与各个位点相关的可观测量的对易子会消失，因此与不相交区域相关的任意两个可观测量一定对易。从形式上看，这就是说假设11得以满足。

最后，令$G:=\mathbb{Z}$表示格转换的可加群。为了定义对局域可观测量进行的操作G，需要注意对所有的$g\in G$和$\Lambda\in\mathcal{F}$：$|g[\Lambda]|=|\Lambda|$，因此$\mathcal{A}_{g[\Lambda]}$和$\mathcal{A}_\Lambda$是相同的矩阵代数，即$\mathcal{M}(2^{|\Lambda|},\mathbb{C})$：局域可观测量及其变换的图像仅仅是相同矩阵的不同副本，以这一方式定义ν_g的确满足了假设13。

3.5 量子遍历理论与宏观观测值

经典遍历理论关心的是在群 G 及其混合属性下不变的测量 μ，而量子遍历理论讨论的是 G 不变态的属性及其群集属性。因此，在本小节中，我将在说明 QSP 的成功(至少是部分成功的)之处时，讨论空间平均和/或时间平均的作用。请与[Uffink，2006]以及[Earman and Rédei(厄尔曼和瑞德)，1996]这两篇论文进行比较。

因此，本节必须加以解决的一个问题是，遍历理论是否以及如何可以作为建立统计力学的基石。在传统上说是埃伦费斯特对玻耳兹曼遍历假说(或更确切地说是其测量理论版本和准遍历假说)的强调，促使群 G 被看作是由时间演化决定的群 IR。然而，部分是由于玻耳兹曼自己的著作中存在的一些摇摆不定的观点，关于这一假说与 CSP 的基础的相关性这一问题仍然是悬而未决的，特别地，请参见[Uffink，2006，Section I.3，and Subsections I.4.3，I.6.1]。很显然吉布斯[Gibbs，1902]已经选择了强调混合属性的作用，即比度量可递性更强的属性，并更精确地假设动力学是“不稳定的”，可参见[Uffink，2006，Sections I.4.1 and I.5]，其他问题请参见[Emch and Liu，2002，pp.317—330]，勒波维茨(Lebowitz)和西奈(Sinai)学派针对 20 世纪后半叶在此方面的研究给出了中肯的评价，请参阅[Szasz(萨斯)，1996]。

因此，我将集中讨论两个子问题：(i)在何种程度上经典遍历理论的数学形式能推广为量子理论的形式；(ii)在何种程度上这种推广可以更好地将 QSP 的基础的某些特定方面形式化。

第一个问题的答案是，只要做出一些小的调整，大部分的数学形式都可以。第二个问题的答案较为复杂。一方面，只要关注点仍然是时间的演化，那么主要的问题就仍然存在，其中之一就是缺乏逼真的模型。另一方面，群 G 讨论的是几何结构问题，而量子遍历理论——尤其是平均值的作用，以及与极值不变性和群集属性相关的定理——是与 QSP 结合起来用于区分微观描述的量子方面和宏观世界的经典方面。因此，按照遍历被认为是与时间演化相关还是与空间对称性相关来划分，可以将其分为两部分进行讨论。

A. 关于时间的遍历

从[Ford *et al.*（福特等），1965]首先提出的模型中，我们可以得到一些深刻的见解，它后来发展成了两个版本，经典版本和量子版本。[Davies（戴维斯），1972]已将量子版本进行了数学化修正。他已经证明对于链中的一个单一的一维量子振子来说，存在一个弱耦合一维量子谐振子的无限一维链可以作为热容器，并且扩散方程可以用来确定一维量子振子的演化。这是通过对范·霍夫弱耦合/长时极限的严格讨论得到的。在后面的6.1节中会就此进行详细讨论。由于我们稍后会给出一些精确的数学定义（参见5.3.C），因而现在我们足以指出，在范·霍夫极限下，约化演化可以得到哪一个半群是位于我们所考虑的单个振子位置上的作用于冯·诺伊曼代数$\mathcal{N}_o \simeq \mathcal{B}(\mathcal{L}^2(\mathbb{R}, \mathbb{C}))$完备正映射$\{\gamma_s \mid S \in \mathbb{R}^+\}$的可压缩半群。此外，当我们从单个振子的二维相空间$\{\xi P+\eta Q \mid \zeta=(\xi, \eta) \in \mathbb{R}^2\}$的任意一维子空间$\{x\mathbf{u} \mid x \in \mathbb{R}\}$上看，这种演化描述的是一个经典分布$\mu(x, s)$，它对于所有$s \in \mathbb{R}^+$满足扩散方程：

$$\partial_s\mu(x, s)=D[\partial_x^2+\beta(V'(x)\partial_x+V''(x))]\mu(x, s) \tag{57}$$

这里$\beta=1/kT$是自然温度，$V=\dfrac{1}{2}\omega x^2$是谐振势，而扩散常数D和频率ω为数值，它们的值仅依赖于方向$\zeta/|\zeta| \in \mathbb{R}^2/S^1$。需要注意的是相应的不变测度是正则平衡高斯测度$\mu(x)=Z^{-1}\exp(-\beta V(x))$，其中$Z=\int_R dx\mu(x)$，即$Z^{-1}=\sqrt{2\pi\beta\omega}$。

这里模型的关键点是确定由马尔可夫（Markov）扩散过程的可收缩半群$\{\gamma_s \mid s \in \mathbb{R}^+\}$描述的耗散系统会允许一个守恒动力学系统发生正则扩张。事实上，存在一个冯·诺伊曼代数$\mathcal{N}=\pi_\varphi(\mathcal{A})''$的自同构群$\{\alpha_s \mid s \in \mathbb{R}\}$，可以描述处于平衡态$\varphi$的振子的满链，当不存在相互作用时，它相当于温度$\beta$。与3.4节中相一致，这里的准定域可观测量$\mathcal{A}$是$C^*$代数$\otimes_{k\in Z}\mathcal{N}_k$，这里$\mathcal{N}_k$是$\mathcal{N}_o$的拷贝。自由平衡态具有$\varphi=\otimes\varphi_k$的形式，其中$\varphi_k$为$k$位置上振子的冯·诺伊曼正则平衡态。现在令$i$为$\mathcal{N}_o$到$\mathcal{N}$的单射，$\varphi_o$表示$\varphi$对$\mathcal{N}_o$的约束，即$\forall N_o \in \mathcal{N}_o$：$\varphi_o(\mathcal{N}_o)=\varphi(i[N_o])$。进一步令$E$：$\mathcal{N} \to \mathcal{N}_o$为一个关于态$\varphi$的正则条件期望，它满足$\varphi_o \circ E=\varphi \circ \{\mathcal{N}, \alpha, E\}$是$\{\mathcal{N}_o, \gamma\}$的一个扩张的意义是：

$$\forall (s, N_o) \in \mathbb{R}^+ \times \mathcal{N}_o: \gamma_s[N_o] = E \circ \alpha_s \circ i[N_o] \tag{58}$$

详细的内容请参阅[Emch, 1976]，尤其是这里的这一结果与海达(Hida)构建的布朗运动的经典流非常相似[Hida, 1970]，他还证明了在下面这个定义的意义上，这个流是一个经典的柯尔莫哥洛夫流。

定义 16。由一个概率空间$\{\Omega, \mathcal{E}\}$，一个概率测量μ，以及一个$\{\Omega, \mathcal{E}\}$的自同构群$\{\alpha^*_t \mid t \in \mathbb{R}\}$构成的经典动力学系统$\{\Omega, \mathcal{E}, \mu, \alpha^*\}$，使得$\forall t \in \mathbb{R}$: $\mu \circ \alpha_t^* = \mu$，那么它被认为是一个经典的柯尔莫哥洛夫流，当存在一个σ环$\mathcal{A} \subset \mathcal{E}$,使得在概念$\mathcal{A}_t = \alpha_t^*[\mathcal{A}]$下：

$$(1)\ \forall t > 0: \mathcal{A} \subset \mathcal{A}_t;\ (2)\ \vee_{t \in \mathbb{R}} \mathcal{A}_t = \mathcal{E};\ (3) \wedge_{t \in \mathbb{R}} \mathcal{A}_t = \{\emptyset, \Omega\}$$

我们在经典动力学系统中通过它们具有的严格的正动力学熵来描述柯尔莫哥洛夫流的特征，因此，它们具有很高的经典遍历层次，在勒贝格(Lebesgue)谱条件之上，从而也在较弱的混合及遍历条件之上。详细的介绍请参见[Arnold and Avez(阿诺德和阿瓦兹)，1968；Cornfeld *et al.*(科尼菲德等)，1982]。

上述被描述为可收缩半群的正则扩张的守恒量子动力学系统，满足了定义 16 的量子推广，即：

定义 17。由冯·诺伊曼代数$\mathcal{N}$上一个单射正交态φ，$\mathcal{N}$的自同构群$\alpha = \{a_t \mid t \in \mathbb{R}\}$，其中$\forall t \in \mathbb{R}$: $\varphi \circ \alpha_t = \varphi$，组成的量子动力学系统$\{\mathcal{N}, \varphi, \alpha\}$被称为一个广义柯尔莫哥洛夫流，如果存在一个冯·诺伊曼子代数$\mathcal{A} \subset \mathcal{N}$使得：

(1) $\forall t > 0$: $\mathcal{A} \subset \mathcal{A}_t$;

(2) $\vee_{t \in \mathbb{R}} \mathcal{A}_t = \mathcal{N}$;

(3) $\wedge_{t \in \mathbb{R}} \mathcal{A}_t = \mathbb{C}\, I$;

(4) $\forall t \in \mathbb{R}$: $\tau_t[\mathcal{A}] = \mathcal{A}$。

其中符号$\mathcal{A}_t = \alpha_t[\mathcal{A}]$，这里$\{\tau_t \mid t \in \mathbb{R}\}$是与$\varphi$正则相关的模态群。

备注 18。

1. 条件(2)中的$\vee$涉及一个弱算符闭包，即(2)意味着$\mathcal{N}$是包含所有$\mathcal{A}_t$的最小冯·诺伊曼代数，条件(3)中的$\wedge$就是常用的交集，因此(3)表示没有算符属于所有的$\mathcal{A}_t$，除非它是单位元的倍数。

2. 模群τ将会在第 4 节中引入，在这里只要说一句就够了，即如果我们处

理的是有限系统，那么相应于冯·诺伊曼正则平衡密度矩阵来说，τ 是与哈密顿量相关联的$\mathcal{N}$的自同构群。

3. 当$\mathcal{N}$被认为是存在于希尔伯特空间$\mathcal{H}=\mathcal{L}^2(\Omega, \mathcal{E}, \mu)$中的阿贝尔·冯·诺伊曼代数$\mathcal{L}^\infty(\Omega, \mathcal{E})$时，定义 17 包含了定义 16，在$\forall t\in R$: $\tau_t = id$ 的情况下，条件(4)就能被满足。

4. 在一般情况下，对于确保条件期望 E: $\mathcal{N}\to\mathcal{A}$的存在性来说，条件(4)是必要的。

5. 除了动力学熵的正性——它依赖于一个共识，即量子动力学熵的物理意义仍有待给出，请参见[Narnhofer and Thirring(奈霍芬和西凌)，1994；Tuyls(图伊斯)，1998]等参考文献——所有经典柯尔莫哥洛夫系统的遍历属性都会直接从经典领域延续到量子领域[Emch，1976]。在上面所描述的模型中，这些属性以量子$\{\mathcal{N}, \varphi, \alpha\}$三元组的形式表现了出来。

6. 定义 17 是由埃姆什[Emch，1976]首次提出的。[Narnhofer and Thirring，1989]一文中探讨了这个定义从 W^* 代数到 C^* 代数的推广。

7. 这条备注的素材源于[Arnold and Avez，1968]，我们这里插入这条备注只是为下一条备注做准备。在经典遍历理论中，高于柯尔莫哥洛夫流的上一级遍历，是阿诺索夫流。阿诺索夫流将哈达玛于 1898 年做出的观测形式化了，即相对于通常的平坦流形上自由流的线性灵敏度负曲率流形上的测地线表现出对初始条件的指数灵敏度。此外如果流形是紧致的，那么我们可以直观地期望，哈达玛的观测必然导致某种混合行为。情况的确是这样的：一旦哈密顿流被证明是遍历的——常负曲率的一个紧致表面上的测地线流——它就是一个阿诺索夫流。这些流在与流动方向的垂直方向上表现出指数性收缩和扩张，从而预示着经验观测到的李雅普诺夫(Lyapunov)系数存在一种微观解释。离散时间的原型是关于环面 $T^2 := \mathbb{R}^2/\mathbb{Z}^2$ 的阿诺德 CAT 映射操作。我们应该注意到，取决于柯尔莫哥洛夫流，经典遍历理论可以被看作是概率论的一部分。此外，阿诺索夫流涉及了对微分几何的一个本质诉求，这在 20 世纪后半期才通过俄罗斯学派的工作被人们认识到。

8. 为了研究阿诺索夫流概念的可能的量子推广，笔者通过与奈霍芬、休厄尔(Sewell)和西凌合作提出了阿诺索夫流的量子等价概念[Emch *et al.*，

1994a]。关于其前提，请参阅[Benatti *et al.*（贝纳蒂等），1991a]；在该背景下关于动力学熵的讨论，请参阅[Andries *et al.*（安德里斯等），1995]；关于评论及一般性的观点，请参见[Narnhofer，2001；Narnhofer，2005]。

这个推广的一个本质特征是，现在这个量子 CAT 映射的相空间是非交换环面 T^2_θ，它是孔耐(Connes)非交换几何中普遍存在的主要部分，参见[Connes，2000，Section XIII]或[Garcia-Bondia *et al.*（加西亚—邦迪亚等），2003，Chapter. 12]。这些环面在几何量子化方案中的地位，参见[Emch，1998b]。对于量子遍历理论，在[Emch *et al.*，1994a]中已经指出，扩张和压缩极限圆的生成元会形成导数的二维特异空间内的一个基，这个二维特异空间不是内部近似的，即它不能由内部导数一致地逼近[Garcia-Bondia *et al.*，2003，Section 12.3]。

在经典和量子阿诺索夫流中扩张和收缩方向的存在性为经典混沌和量子混沌之间架设了一个桥梁。量子混沌是什么——或者它应该是什么——这个问题已经在不同的方面受到了关注，参见[Gutzwiller(古兹维勒)，1990]。在哲学视角方面，请参见[Belot and Earman(贝洛特和厄尔曼)，1997]，对此最近的综述，请参见本书[Landsman，2006，Section 5.6]。

这个备注中所提到的研究，以及在 QSP 中的应用，在 QFT 中也存在数学化的等价研究，参见[Borchers(博彻斯)，1999；Wiesbrock(威斯布鲁克)，1997]以及下面的第 5.5 节。

总结和警示。似乎可以直接得出这样的结论：从经典遍历运动学到量子领域的数学化推广完成得相当不错。然而，在面对哈密顿力学时，在量子案例中基于物理动力学的讨论，在经典案例中进行得并不顺利。在本小节开始时所讨论的模型的帮助下，一些概念性问题可能已经得以阐明。在那里，耗散动力学系统$\{\mathcal{N}_o, \gamma\}$可以被视为两个不同的保守动力学系统的约化动力学，它们二者都作用于相同的无限谐振子系综。从第一个系统来说，约化动力学只能通过范·霍夫极限获得，该极限在单一子系统中导致了溶液中的完全弱耦合长时效应。但是，支配初始保守系统的动力学对应的时间尺度，与关于其他保守系统，即作为耗散系统的正则扩张的系统对应的时间尺度之间，没有任何共同之处。因此，几乎没有理由相信后者的遍历行为会反映出有关前者的任何全局性

动力学性质。

虽然在原始模型方面，它可能会受到责难，但需要强调守恒微观描述对应的时间尺度和突现宏观描述对应的时间尺度可能显著地不同。在更复杂的模型中，这必须被纳入到考虑范围内，且必须——以这种或那种方式——消除微观描述的复杂行为，这需要在清晰的遍历行为在宏观层面显现出来之前完成。看起来，范·霍夫的想法是希望找到能做到这一点的一种合理方法，参见下面的6.1节。

从玻耳兹曼研究的初始动机出发，经典遍历理论的大多数描述关注于时间演化的特性，特别是度量可递性和可观测度量的时间平均。将其推广到量子领域后，则经典遍历理论的其他方面也受到了关注，即关于支配演化的那些群以外的其他的群行为的空间平均。这将在本小节的第二部分中进行介绍。

B. 关于空间的遍历

如同已经被哈格[1959b]视为QFT的群那样，与QSP最直接相关的“其他”群是空间变换群，它被看作是哈格—卡斯特勒公理中协方差假设的一部分，可对照于上面的假设13。如果 $n=1, 2, \cdots$，令$\mathbb{X}^n$表示欧几里得空间$\mathbb{R}^n$或“立方”晶格$\mathbb{Z}^n$，并令 $|x|$ 表示矢量 $x\in\mathbb{X}^n$的长度。后面，我们将专注于所有从$x\in\mathbb{X}^n\mapsto x+a\in\mathbb{X}^n$的变换的阿贝尔群 $G\simeq\mathbb{X}^n$，这里 $a\in\mathbb{X}^n$。进一步令$\{\mathcal{A}_\Lambda \mid \Lambda\in\mathcal{F}\}$是定域代数的相应哈格—卡斯特勒网，$\mathcal{A}$为它们的 C^* 归纳极限，$\mathcal{A}$的形式为推论14中定义的自同构群$\{\nu_a \mid a\in X^n\}$。再令$\mathcal{A}_o\subset\mathcal{A}$表示定域可观测量的代数。对于任意确定的$\mathcal{F}$中的元素对$(\Lambda_1, \Lambda_2)$，存在 $a_{12}\in G$，使得对所有 $a\in G$ 有 $a[\Lambda_1]\bowtie\Lambda_2$，其中 $|a|>|a_{12}|$ 成立。因此，根据定域性(参见假设11)，当 $a\in G$，其中 $|a|>|a_{12}|$，$A_1\in\mathcal{A}_1$和 $A_2\in\mathcal{A}_2$，我们有 $\nu_a[\mathcal{A}_1]\mathcal{A}_2=\mathcal{A}_2\nu_a[\mathcal{A}_1]$。由连续性，我们要求：

推论19。对于所有 $A, B\in\mathcal{A}$：$\lim_{|a|\to\infty}\|\nu_a[A]B-B\nu_a[A]\|=0$，即变换群 G 以规范渐近阿贝尔的方式作用到准定域可观测量的代数$\mathcal{A}$上。

此属性在针对有限系统量子力学而建立的最初的冯·诺伊曼架构下是没有意义的。在为无穷系统设计的广义哈格—卡斯特勒框架下，这个陈述在空间变换下显然是正确的，但它很难由可控现实模型中的时间演化满足。

这就提出了三个问题：第一个问题是这个属性是否可推出一些有用的结

论；第二个问题是在不破坏这个属性可能导出的结论的前提下，是否可以弱化这个属性；第三个问题是是否存在一个该属性的弱化形式能够由时间演化满足。我认为，前两个问题的答案都是“是”。具体而言，关于第一个问题，请参阅下面的推论30；而第二个问题的答案，请参见后面的定理25。不过，在这里，我将指出我们要小心可以确保第三个问题有肯定回答的那些假设的诱惑，它们可能很难满足特定的模型，参见5.4.B中的最后一段以及后面的备注。

定义20。当$\forall(a, A)\in G\times\mathcal{A}$：$\varphi(\nu_a[A])=\varphi(A)$时，准定域可观测量的代数$\mathcal{A}$上的态$\varphi$被称为是平移不变的，记作$\varphi\circ\nu=\varphi$。态$\varphi$被认为是极值平移不变的，如果它是平移不变的，且不能被记为不同平移不变态的凸和。

用G表示变换群$\mathbb{X}^n=\mathbb{R}^n$或$\mathbb{Z}^n$，$G$一般被认为等同于$\mathbb{X}^n$。令$\mathcal{C}=\mathcal{C}(G)$是所有复值连续有界函数$f$：$G\to\mathbb{C}$的集合。此后，这个集合包含逐点相加和相乘函数，满足sup模$\|f\|=\sup_{x\in G}|f(x)|$。这些操作使$\mathcal{C}$具有了（阿贝尔）$C^*$代数的结构。因此可由$a[f](x)=f(x-a)$来定义$G$的作用。

定义21。由上面的概念可知，$\mathcal{C}$上的一个不变平均是$\mathcal{C}$上的使得$\forall(a, f)\in G\times\mathcal{C}$：$\eta(a[f])=\eta(f)$的一个态$\eta$。

给定$\mathbb{X}^n$，存在几个这类平均。例如，$\mathbb{R}$上的遍历平均可以定义如下。令$\mathcal{C}_e=\left\{f\in\mathcal{C}\mid\lim_{a\to\infty}1/2a\int_{-a}^{a}dxf(x)\text{ 存在}\right\}$，那么$\forall f\in\mathcal{C}_e$，令$\eta_e(f):=\lim_{a\to\infty}1/2a\int_{-a}^{a}dxf(x)$，这可将连续性扩展到$\mathcal{C}$上，使得给定一个不变平均，就是后来我将会用到的平均。我们可能会希望在$\mathcal{C}_+=\{f\in\mathcal{C}\mid\lim_{x\to\infty}f(x)\text{存在}\}$上定义类似的平均$\eta_+$。类似地，另一个平均$\eta_-$由承认极限为$x\to-\infty$的函数来得到。

为了定义态的平均和可观测量的平均，注意对类定域可观测量的代数$\mathcal{A}$上的每个态φ以及任意A，$B\in\mathcal{A}$，函数$\varphi(\nu_\bullet[A]B)$：$a\in G\mapsto\varphi(\nu_a[A]B)\in\mathbb{C}$——这里符号$\bullet$用于标记变量$a$的位置——是连续且有界的，即以$\|A\|\ \|B\|$为边界。因此函数$\varphi(\nu_\bullet[A]B)$属于$\mathcal{C}$。当$B=I$时，我们将$\varphi(\nu_\bullet[A]I)$简记为$\varphi(\nu_\bullet[A])$。有了这些概念，下面的定义就有意义了。

定义22。给定$\mathcal{C}$上的一个不变平均η以及类定域可观测量的代数$\mathcal{A}$上的任

意一个态 φ，那么 φ 的平均 $\eta[\varphi]$ 被定义为下面的变换不变态：

$$\eta[\varphi]: A \in \mathcal{A} \mapsto \eta(\varphi(\nu_{\bullet}[A])) \in \mathbb{C}$$

这个变换不变态 φ 被称为 η 群，如果：

$$\forall A, B \in \mathcal{A}: \eta(\varphi(\nu_{\bullet}[A]B)) = \varphi(A)\varphi(B)$$

需要注意的术语：

1. η 群集也被称为"弱群"。

2. η 群集不应该与"弱混合"的强属性相混淆，即：

$$\forall A, B \in \mathcal{A}: \eta | \varphi(\nu_{\bullet}[A]B) - \varphi(A)\varphi(B) | = 0$$

这里对于每个复数 z，$|z|$ 表示 z 的绝对值。这里"弱混合"符合经典遍历理论中的用法，参见[Arnold and Avez，1968，21]。

3. 简称为群的属性不涉及平均，因此是更强的，即：

$$\forall a \in \mathbb{R}^n \text{ 以及 } \forall A, B \in \mathcal{A}: \lim_{\lambda \to \infty} \varphi(\nu_{\lambda a}[A]B) = \varphi(A)\varphi(B)$$

这个属性在经典遍历理论中被称为"混合"，参见[Arnold and Avez，1968，20]。

4. 在下面的定义 27 中会引入一个更强的属性。

5. 上述表示当该系统处于态 φ 时，$\nu_a[A]$ 和 B 之间相关性程度的每一个属性随距离的增大而减弱。附属于这些特性的也用于 QFT 的项"群集"似乎是从散射理论那里继承而来的，它表示分离的散射产物或"群集"的渐近独立性。

可观测量平均定义的是有点复杂的。一般的数学框架，请参见[Emch，1972a，Subsection 2. 2. d]。特别地，与下面的注释 23 和定理 25 相关的一般陈述和证明，请参见[Emch，1972a，Lemma，pp. 174—175]和[Emch，1972a，Theorem 8，pp. 183—184]。需要注意的是，这里空间变换群的行为的渐近阿贝尔性(即上述推论 19)可由下面更简单的表述来说明。这里需要将全局可观测量纳入研究视域——参见上面的定义 15。

令 φ 为类定域可观测量的代数 $\mathcal{A}$ 上的变换不变态，$\{\pi_\varphi, \mathcal{H}, \Phi\}$ 为与 φ 相关联的 GNS 三元组。进一步令 $\mathcal{N}_\varphi = \pi_\varphi(\mathcal{A})''$ 以及 $\mathcal{Z}_\varphi = \mathcal{N}_\varphi \cap \mathcal{N}_\varphi'$。

对于给定的 $a \in G$，以及遍历 $\mathcal{A}$ 的 A，映射 $\pi_\varphi(A)\Phi \in \mathcal{H} \mapsto \pi_\varphi(\nu_a[A])\Phi \in \mathcal{H}$ 唯一地扩展为一个幺正算符 $U_a \in \mathcal{U}(\mathcal{H}) := \{U \in \mathcal{B}(\mathcal{H}) \mid U^* U = UU^* = I\}$。这定

义了一个使得$\forall(a, A) \in G \times \mathcal{A}$：$U_a\pi_\varphi(A)U_a^* = \pi_\varphi(\nu_a[A])$成立的连续归一化表征 U：$a \in G \mapsto U_a \in G \mapsto \mathcal{U}(\mathcal{H})$。

像通常那样，令 $U(G') := \{B \in \mathcal{B}(\mathcal{H}) \mid \forall a \in G: U_aB = BU_a\}$ 表示 $U(G)$ 的换位矩阵。等价地，这里 $U(G)' = \{B \in \mathcal{B}(\mathcal{H}) \mid \forall a \in G: U_aBU_a^* = B\}$。

最后，令$\mathcal{P} := \{\Psi \in \mathcal{H} \mid \forall a \in G: U_a\Psi = \Psi\}$，并用 P 表示从$\mathcal{H}$到$\mathcal{P}$的正交投影。

注释 23。对于$\mathcal{C}$上的每一个不变平均 η，映射 η_φ：$A \in \mathcal{A} \mapsto \eta_\varphi[A] \in \mathcal{Z}_\varphi \cap U(G)'$对于所有 $A \in \mathcal{A}$，由下式定义：

$$\forall \Psi_1, \Psi_2 \in \mathcal{H}: (\Psi_1, \eta_\varphi[A]\Psi_2) = \eta(\Psi_1, \pi_\varphi(\nu_\bullet[A])\Psi_2)$$

这是一个*同态，且满足 $\eta_\varphi[A]P = P\eta_\varphi[A] = P\eta_\varphi[A]P$。

定义 24。令 η 为$\mathcal{C}$上的一个不变平均；$\mathcal{C}$为 $A \in \mathcal{A}$的代数$\mathcal{A}$上的一个平移不变态；$\mathcal{N}_\varphi = \pi_\varphi(\mathcal{A})''$为通过 GNS 三元组$\{\pi_\varphi, \mathcal{H}, \Phi\}$与 φ 相关的全局可观测量的代数；$\mathcal{N}_\varphi^G = \{N \in \mathcal{N}_\varphi \mid \forall a \in G: U_aNU_a^* = N\}$为平移不变的全局可观测量的代数。那么一个类定域可观测量 $A \in \mathcal{A}$的平均被定义为平移不变的全局可观测量 $\eta_\varphi[A] \in \mathcal{N}_\varphi^G$。

我们现在已经准备好阐释与空间转换群的行为有关的中心量子遍历定理。

定理 25。令 ν：$a \in G \to \mathrm{Aut}(\mathcal{A})$表示准定域可观测量的代数$\mathcal{A}$上的空间变换群，并令 η 为$\mathcal{C}$上的任意一个不变平均。那么关于$\mathcal{A}$上的一个平移不变态 φ 的下列条件是等价的：

1. φ 是极值平移不变的；
2. φ 是 η 群集，即$\forall A, B \in \mathcal{A}$：$\eta(\varphi(\nu_\bullet[A]B)) = \varphi(A)\varphi(B)$；
3. 将 φ 扩展为与 φ 相关的全局可观测量的冯·诺伊曼代数$\mathcal{N}_\varphi$的正则扩展 $\tilde{\varphi}$：$N \in \mathcal{N}_\varphi \mapsto (\Phi, N\Phi) \in \mathbb{C}$ 是该代数上的唯一平移不变标准态；
4. 不变子空间$\mathcal{P} \subset \mathcal{H}$是一维的；
5. 每个类定域可观测量 $A \in \mathcal{A}$的平均 $\eta_\varphi[A]$是单位元的一个倍数，即 $\eta_\varphi[A] = \varphi(A)I$；
6. 所有平移不变的全局可观测量 $N \in \mathcal{N}_\varphi^G := \mathcal{N}_\varphi \cap U(G)'$都是单位元的倍数；
7. $\mathcal{Z}_\varphi \cap U(G)' = \mathbb{C}I$，其中$\mathcal{Z}_\varphi := \mathcal{N}_\varphi \cap \mathcal{N}_\phi'$。

备注 26。

1. 回顾在定义 15 中引入的三类可观测量。与有限区域 Λ 相关的定域可观测量由初始冯·诺伊曼形式来描述[von Neumann, 1932c]，其中，典型地，$\mathcal{A}_\Lambda = \mathcal{B}(\mathcal{H}_\Lambda)$ 且 $(\mathcal{H}_\Lambda) = \mathcal{L}^2(\Lambda, dx)$。因此，我们将定域可观测量看作 $\mathcal{A}_o = \cup_{\Lambda \in \mathcal{F}} \mathcal{A}_\Lambda$ 的自伴元素。准定域可观测量就是被抽象地定义为定域可观测量的模极限的可观测量，是对空间中无限延伸的多体系统的微观描述，下面的第 5 节将以三个具体的 QSP 例子开始。这些“准定域”可观测量属于 C^* 代数 $\mathcal{A} = \overline{\mathcal{A}_o}^{n}$。第三类可观测量，全局可观测量出现在宏观层面，当人们研究物质的主体属性时；它们属于公认为对应于($\mathcal{A}$中的)态 φ 的($\mathcal{A}$的)GNS 表征 π_φ 的弱闭包冯·诺伊曼代数 $\mathcal{N}_\varphi := \pi_\varphi(\mathcal{A})''$，特别地它们是通过被称为热力学极限的过程得到的，在下面的章节中我会讨论其中的几个例子。

空间平均是这类全局性可观测量的例子。铁磁性的一个具体例子可以表现为磁化强度三个分量中的任一个。第三类可观测量依赖于所考虑系统的全局状态，从而反映了系统的制备情况。例如，如果状态是极值平移不变的，那么这些可观测量是单位算符的倍数——回顾定理 25 中条件(1)和(5)的等价性——因此当它们的结构在定域范围内与给定的状态不同时，它们的值是相同的。当它们的结构在全局范围内彼此不同时，它们的值是不同的，这可以作为不同热力学相位存在的证据，参见第 5.7 节。

2. $\mathcal{A} = \overline{\mathcal{A}_o}^{n}$ 上的一个全局态 φ 通常由下面的连续性定义，其中 $\mathcal{A}_o = \cup_{\Lambda \in \mathcal{F}} \mathcal{A}_\Lambda$：

$$\forall \Lambda \text{ 且 } \forall A_\Lambda \in \mathcal{A}_\Lambda:\ \varphi(A_\Lambda) = \lim_{\substack{|\Omega| \to \infty \\ \Omega \in \mathcal{F}, \Omega \supset \Lambda}} \varphi_\Omega(A_\Lambda) \tag{59}$$

这里 $\{\varphi_\Omega \mid \Omega \in \mathcal{F}\}$ 是定域态的一个相容族。定域态本身是相对于某种一致性边界条件定义的，例如每一个 Λ 上的周期性边界条件。因此，全局态 φ 和全局可观测量的冯·诺伊曼代数 $\mathcal{N}_\varphi := \pi_\varphi(\mathcal{A})''$ 可能取决于我们选定的边界条件。这尤其会发生于长程序列存在的情况中，它往往是伴随着相变的开始的。即使在热力学极限下，这种对初始条件的依赖性，也是一种普遍存在的现象，这在经典统计物理学中我们就知道了。

事实上，在一个后来被证实了的论点中，请参见[Emch and Liu, 2002, pp. 416—417］及派尔斯（Peierls）［1936］，即二维空间内的伊辛模型，在温度足够低的情况下，将敏感性发展为边界条件：一个相位——比如说一个严格正磁化的相位——可以通过在边界上将所有自旋固定为“向上”来选择得到。

3. 这里需要再一次提到的是，在$\{\mathcal{N}, \varphi\}$是$\{\mathcal{L}^{\infty}(\Omega), \mu\}$的这种特殊情况下，上述定理简化为已知的经典情况。但是请注意，这里所陈述的是空间变换定理，而不是时间演化定理，原因是证明过程使用了空间变换满足的渐近阿贝尔性——参见上面的推论 19——或如下面的备注 31 中式(61)的一些弱化形式。然而，对于量子动力学模型的时间演化来说，即使是阿贝尔性的这种弱化渐近形式也很难得到。

当表征π_{φ}为主要的时，即当中心$\mathcal{Z}_{\varphi} := \pi_{\varphi}(\mathcal{A})'' \cap \pi_{\varphi}(\mathcal{A})'$满足$\mathcal{Z}_{\varphi} = \mathbb{C}\,I$时，定理中的群集条件(2)可能被加强。特别地，对于任意区域$\Lambda \in \mathcal{F}$，令：

$$\mathcal{A}_{\Lambda}{}^{c} := {}^{n}\overline{\cup_{\Omega \in \mathcal{F}; \Omega \bowtie \Lambda} \mathcal{A}_{\Omega}}$$

这里，对于任意子集$\mathcal{B} \subset \mathcal{A}$，${}^{n}\overline{\mathcal{B}}$表示$\mathcal{A}$的规范拓扑中的闭包$\mathcal{B}$。作为定域性$A \in \mathcal{A}_{\Lambda}$和$B \in \mathcal{A}_{\Lambda}{}^{c}$满足$AB - BA = 0$的一个推论。现在令$\mathcal{N}_{\varphi,\Lambda}{}^{c} := \pi_{\varphi}(\mathcal{A}_{\Lambda}{}^{c})''$。

定义 27。准定域可观测量的代数$\mathcal{A}$上的一个态φ被称为是一致群集的，当对于任意的$A \in \mathcal{A}$和$\epsilon > 0$，存在一个取决于A和ϵ的区域$\Lambda \in \mathcal{F}$，使得：

$$\forall B \in \mathcal{A}_{\Lambda}{}^{c}:\ |\varphi(AB) - \varphi(A)\varphi(B)| \leq \epsilon \|B\| \tag{60}$$

定义 28。冯·诺伊曼代数$\mathcal{N}_{\varphi}{}^{\infty} := \cap_{\Lambda \in \mathcal{F}} \mathcal{N}_{\varphi,\Lambda}{}^{c}$的元素在关于$\varphi$的无穷大处被称为可观测量。

注释 29。对于每个个别的态φ来说，无穷大处的可观测量是重要的，即$\mathcal{N}_{\varphi}{}^{\infty} \subseteq \mathcal{Z}_{\varphi}$。进一步讲，关于态$\varphi$的以下两个条件是等价的：

1. 所有无穷大处的可观测量是单位算符的倍数，即$\mathcal{N}_{\varphi}{}^{\infty} = \mathbb{C}\,I$；

2. φ是一致群集。

注意定义 27、28 和注释 29 并不要求φ是空间平移不变的，虽然它们在本质上涉及了$\mathcal{A}$的定域结构。对于空间平移不变态我们还有：

推论 30。下面两个条件：

1. φ是类定域可观测量的代数$\mathcal{A}$上的一个平移不变态；

2. 全局可观测量的代数$\mathcal{N}_{\varphi}$是一个因数，即$\mathcal{Z}_{\varphi} = \mathbb{C}\,I$

将二者结合起来可以得到：

a. φ 是极值平移不变的（因此满足定理 25 中提到的等价性条件）；

b. φ 是一致群集。

备注 31。

1. 当 φ 是一个极值 KMS 态时，推论 30 中的条件(2)得以满足，参见下面的第 5.6 节。

2. 定理 25、注释 29 及推论 30 的证明并不是不重要的，但它们都是到 20 世纪 70 年代初才给出的，参见[Emch, 1972a, Theorem II. 2. 8 and Theorem IV. 1. 7]。

3. 特别地，定理 25 的证明表明，其七个条件的等价性可以在更普遍的情况下得到，即空间变换群的作用可由满足 η 阿贝尔性条件的不变态 φ 的作用替换的情况，这里的不变态 φ 满足下面条件：

$$\forall A,\ B,\ C \in \mathcal{A}:\ \eta\{\varphi(C^*[\nu_g[A]B - B\nu_g[A]]C)\} = 0 \tag{61}$$

该条件比推论 19 中证明的有关变换群作用的正态渐近阿贝尔性弱得多。

4. 因此，尝试将上述过程转换到由量子动力学系统的时间演化决定的 IR 群中进行分析是很吸引人的事。事实上，如果 φ 是一个极值 IR$^-$ 不变态，那么，这样的动力学系统在式(61)的意义上将是 η 阿贝尔的，如果提供了 GNS 表征的矢量 Φ 的话——它对于 $\pi_\varphi(\mathcal{A})'$是循环的——那么它对于冯·诺伊曼代数 $\pi_\varphi(\mathcal{A})'$也是循环的，这个条件等价于要求 Φ 对于冯·诺伊曼代数$\mathcal{N}_\varphi := \pi_\varphi(\mathcal{A})''$是分离的，即 $N \in \mathcal{N}_\varphi$和 $N\Phi = 0$ 意味着 $N = 0$。在讲到一个由标准形式表示为冯·诺伊曼代数时，我们会涉及一个条件，即这个冯·诺伊曼代数$\mathcal{N}$允许一个矢量 Φ 对于$\mathcal{N}$和$\mathcal{N}'$都是循环的。在当前情况下关于此条件的相关性，请参见后面的定义 36 和定理 39。然而，这只是重新提出了是否 φ 在趋向平衡的演化下会取到极值的这个问题。在这方面，我们可能会注意到，这是本小节一开始讨论的关于弱耦合谐振子链中的扩张演化的案例，以及一般的关于广义柯尔莫哥洛夫流的演化 α 的案例。参见定义 17，或备注 18 后的“警示”或第 5.4(B)小节。

5. 在数学方面，人们更普遍地认为量子遍历理论是与群行为而不是与空间或时间变换相关。事实上，定理 25 和上面的第三条备注无须修改就可以扩展为对顺从群的行为的描述，这里顺从群就是定义 21 的意义上具有一个不变平

均的群$\mathcal{G}$(其中$G=\mathbb{R}^n$或$\mathbb{Z}^n$由$\mathcal{G}$替代)。关于顺从群理论的一般性描述,请参见[Greenleaf(格林利夫),1969],关于其在QSP中应用的简要回顾,请参见[Emch,1972a,pp. 164—172]。这里我们仅限于关注定域紧致群,这足以让我们注意到,紧致群、阿贝尔群以及它们的半直集是顺从群,特别地是有限维欧氏空间中的旋转群,平移群以及欧氏群是顺从群。然而,非紧致半单李群都不是顺从群,因此,特别地,四维相对论量子场论中的洛伦兹群不是顺从群。

6. 要使该理论涵盖的内容比顺从群行为更广,我们可以考虑一个C^*代数$\mathcal{A}$的"自同构大群",即α:$\mathcal{G}\to \mathrm{Aut}(\mathcal{A})$,对于每个自件$A\in\mathcal{A}$和$\mathcal{A}$上的每个$\mathcal{G}$不变态$\varphi$,它满足下式:

$$^{\omega-op}\overline{co\{\pi_\varphi(\alpha_g[A])\mid g\in\mathcal{G}\}}\cap\pi\varphi(\mathcal{A})'\neq\emptyset \tag{62}$$

这里对于任何一个矢量空间的子集S,$co\{S\}$表示S的"凸包",即S中元素的所有凸组合的集;对于任何一组$\mathcal{B}\subset\mathcal{B}(\mathcal{H})$,$^{\omega-op}\overline{\mathcal{B}}$表示$\mathcal{B}(\mathcal{H})$的弱算符拓扑结构中$\mathcal{B}$的闭包。斯托默(Størmer)在1967年引入了自同构大群的概念,之后他很快用该概念证明了经典概率理论中的德菲尼蒂可交换性定理的量子等价定理[Størmer,1969]。对量子理论语言基础的回顾和应用,请参见[Emch,2005]及其中的参考文献。请注意,任何顺从群都是系统的一个自同构大群,只要它们的作用是令系统对一些平均η表现出η阿贝尔性。

在这里,再一次地人们很难违背这样一个结论,即量子遍历理论现在是用来寻找关于QSP的进一步物理应用的一个成熟的数学理论,最显而易见的是通过这种理解它提供了本节中描述的各种群集(或混合)属性,参见后面的5.4和5.7小节。

4. KMS平衡条件

将KMS条件看作平衡态的一个典型特征的观点来源于两种思潮的交汇。

第一个方面的思想来源是由久保(Kubo)[1957]以及马丁和施温格(Martin and Schwinger)[1959]提供的,他们指出在凝聚态物理学中起核心作用的所谓热力学格林(Green)函数——参见[Bonch-Bruevich and Tyablikov(伯尼施—伯鲁维奇和泰伯利科夫),1962]——具有明显的解析特征。具体的例子,请参见后

面的注释32。

第二个灵感来源是莫雷(Murray)和冯·诺伊曼[1936]的原始论文，最后它成为了冯·诺伊曼代数理论，以及该理论先驱之一的莫雷的回忆录，见莫雷[1990]。大量的理论是构建在如下事实基础上的：存在一个矩阵代数$\mathcal{N}$，及其换位$\mathcal{N}'$，满足以下特征：

(i)它们是因数，即具有平凡中心：$\mathcal{N}\cap\mathcal{N}'=\mathbb{C}\,I$；(ii)$\mathcal{N}$和$\mathcal{N}'$确定一个公共的循环矢量$\Phi$；(iii)存在一个对合反幺正算符$J$，使得$J\Phi=\Phi$且$N\in\mathcal{N}\mapsto JNJ\in\mathcal{N}'$是一个双射。一个具体且简单的例子，请参见后面的式(71)。

对该理论的两个方面的讨论——分析和代数方面的讨论——都涉及了一些数学上的复杂性，因此本节分为两个小节：第一小节介绍一个简单的例子，第二小节介绍一般性理论。

4.1 维格纳方法

在本小节中，我希望遵循维格纳的要求[Wigner, 1962]："请用2×2矩阵来解释。"因此，我继续描述一个正则平衡态下的具有1/2自旋的量子粒子在自然温度$\beta>0$处，以及平行于z轴在磁场B中会发生什么。可观测量是具有复杂元的2×2矩阵的代数$\mathcal{M}$的自伴元素。其哈密顿量为：

$$H=-B\sigma^z=\begin{pmatrix}\epsilon_1 & 0\\ 0 & \epsilon_2\end{pmatrix}，其中\ \epsilon_1=-B,\ \epsilon_2=+B \tag{63}$$

根据冯·诺伊曼特性式(38)，正则平衡态为：

$$\psi_H:\ M\in\mathcal{M}\to\mathrm{Tr}(\rho_H M)，其中\ \rho_H=\begin{pmatrix}\lambda_1 & 0\\ 0 & \lambda_2\end{pmatrix} \tag{64}$$

这里$\lambda_n=Z_H^{-1}\exp(-\beta\epsilon_n)$，其中$Z_H=\exp(-\beta\epsilon_1)+\exp(-\beta\epsilon_2)$表示系统的分区函数。

在海森堡(Heisenberg)的概念中，与薛定谔的概念式(39)共轭的演化为：

$$\left.\begin{aligned}&\alpha_t:\ M\in\mathcal{M}\to\alpha_t[M]=U^*(t)MU^*(-t)\in\mathcal{M}\\ &其中\qquad U^*(t)=\begin{pmatrix}e^{i\epsilon_1 t} & 0\\ 0 & e^{i\beta\epsilon_2 t}\end{pmatrix}\end{aligned}\right\} \tag{65}$$

为了使计算更容易，且能立即推广到更高维，我们考虑矩阵：

$$E_{mn}:\ \Psi \in \mathbb{C}^2 \mapsto (\Psi_n,\ \Psi)\Psi_m \in \mathbb{C}^2$$

这里$\{\Psi_n \mid n=1,\ 2\}$是H的本征矢量，即：

$$\Psi_1=\begin{pmatrix}1\\0\end{pmatrix};\quad \Psi_2=\begin{pmatrix}0\\1\end{pmatrix}:$$

$$E_{11}=\begin{pmatrix}1&0\\0&0\end{pmatrix};\ E_{12}=\begin{pmatrix}0&1\\0&0\end{pmatrix};\ E_{21}=\begin{pmatrix}0&0\\1&0\end{pmatrix};\ E_{22}=\begin{pmatrix}0&0\\0&1\end{pmatrix}$$

这些矩阵形成了$\mathcal{M}$上的一个基，且式(64)和式(65)中的φ_H和α_t满足：

$$E_{kl}E_{mn}=\delta_{lm}E_{kn},\ \psi_H(E_{mn})=\lambda_m\delta_{mn},\ \alpha_t(E_{mn})=e^{i(\epsilon_m-\epsilon_n)t}E_{mn}$$

从这些关系和等式$\exp[-\beta(\epsilon_m-\epsilon_n)]\lambda_n=\lambda_m$中，我们得到解析函数

$$f_{klmn}:\ z\in\mathbb{C}\to\lambda_n e^{i(\epsilon_m-\epsilon_n)z}\delta_{lm}\delta_{kn}$$

满足$\forall t\in\mathbb{R}$：$f_{klmn}(t)=\psi_H(E_{kl}\alpha_t[E_{mn}])$和$f_{klmn}(t+i\beta)=\psi_H(\alpha_t[E_{mn}]E_{kl})$。此外，在带状区域有：

$$\Omega_\beta:=\{z\in\ \mathbb{C}\mid 0\leqslant \operatorname{Im} z\leqslant\beta\}$$

解析函数f_{klmn}是有界的，其边界为$\exp(|\epsilon_m-\epsilon_n|\beta)$。

正则平衡态ψ_H的这两个属性可线性地扩展为时间关系函数：

$$f_{MN}(t)=\psi_H(M\alpha_t[N])\text{和}f_{MN}(t+i\beta)=\psi_H(\alpha_t[N]M) \qquad (66)$$

其中M和N为$\mathcal{M}$中的任意量。

相反，假设φ是$\mathcal{M}$中的一个态，使得对于$\mathcal{M}$中的每一对M和N存在一个函数$f_{M,N}$：$z\in\Omega_\beta\mapsto f_{M,N}(z)\in\mathbb{C}$，使得：

(i)$f_{M,N}$在带状区域Ω_β上是有界的且连续的；

(ii)$f_{M,N}$在带状区域中是解析的；

(iii)对于所有的$t\in\mathbb{R}$：$f_{M,N}(t)=\varphi(M\alpha_t[N])$且$f_{M,N}(t+i\beta)=\varphi(\alpha_t[N]M)$。

那么，特别地，当$M=I$时，函数$f_{I,N}$是一个周期为$i\beta$的周期函数。因此它可被扩展为整个复平面上的有界解析函数。从而经典刘维尔定理——参见[Churchill and Brown, 1990, Thm. 43.1]——使得该函数必须是常数函数，即对于所有的$(t,\ N)\in\mathbb{R}\times\mathcal{M}$：$\varphi(\alpha_t[N]):=\operatorname{Tr}U^*(-t)\rho U^*(t)N$等价于$\operatorname{Tr}\rho N=\varphi(N)$，因此：

$$\rho=\begin{pmatrix}\mu_1&0\\0&\mu_2\end{pmatrix}$$

这里μ_1，μ_2的值为正，且$\mu_1+\mu_2=1$。通过比较每对指数(m, n)的解析延拓$f_{mn}(t)=\varphi(E_{nm}\alpha_t[E_{mn}])=e^{i(\epsilon_m-\epsilon_n)t}\mu_n$和$f_{mn}(t+i\beta)=\varphi(\alpha_t[E_{mn}]E_{nm})=e^{i(\epsilon_m-\epsilon_n)t}\mu_m$，我们得到$\exp[-\beta(\epsilon_m-\epsilon_n)]\mu_m=\mu_n$，因此由归一化要求$\varphi(I)=1$，即$\mu_1+\mu_2=1$可得：$\mu_n=\frac{e^{-\beta\mu_n}}{e^{-\beta\mu_1}+e^{-\beta\mu_2}}=\lambda_n$，因此，事实上$\varphi=\psi_H$。

总之，我们通过基本方法得到一个关于该理论第一个方面的简单例子，其解析特点如下。

注释32。令$H=-B\sigma^z$描述一个磁场B中的1/2自旋的哈密顿量。那么对于$\mathcal{M}=\mathcal{M}(2, \mathbb{C})$上任意态$\varphi$，下面的条件是等价的：

（Ⅰ）φ是关于哈密顿量H的正则平衡态ψ_H；

（Ⅱ）对于$\mathcal{M}$中的每一对元素(M, N)，存在一个函数$f_{M,N}$：$z\in\Omega_\beta\to\mathbb{C}$，使得：

$$\left.\begin{array}{l} f_{M,N}\text{在 }\Omega_\beta\text{ 上是有界且连续的；} \\ f_{M,N}\text{在 }\Omega_\beta\text{ 内部是解析的；} \\ \forall t\in\mathbb{R}\text{：}\begin{cases} f_{M,N}(t)=\varphi(M\alpha_t[N]) \\ f_{M,N}(t+i\beta)=\varphi(\alpha_t[N]M) \end{cases} \end{array}\right\} \tag{67}$$

现在我们将注意力转移到理论的代数方面，我们继续考虑同一个简单模型的演化，令φ为$\mathcal{M}$中的单射态，即那些使$M\in\mathcal{M}$且$\varphi(M^*M)=0$意味着$M=0$的态。不失一般性地，我们可以选择这样的一个基，在其中对应于φ的密度矩阵ρ是对角化的，其本征值$\lambda_n(n=1, 2)$严格为正，因为φ被认为是单射的。考虑由下式给出的$\mathcal{M}$的表征π：

$$\forall M=\begin{pmatrix} a & b \\ c & d \end{pmatrix}\in\mathcal{M}\text{：}\pi(M)=\begin{pmatrix} a & b & 0 & 0 \\ c & d & 0 & 0 \\ 0 & 0 & a & b \\ 0 & 0 & c & d \end{pmatrix}=M\otimes I \tag{68}$$

作用到具有标准标量积的希尔伯特空间$\mathbb{C}^4$上，其中Ψ_{kl}可定义为：

$$\Psi_{11}=\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix},\ \Psi_{21}=\begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix},\ \Psi_{12}=\begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix},\ \Psi_{22}=\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

它们是正交归一基。矢量

$$\Phi = \sum_k \lambda_k^{\frac{1}{2}} \Psi_{kk} = \begin{pmatrix} \lambda_1^{\frac{1}{2}} \\ 0 \\ 0 \\ \lambda_2^{\frac{1}{2}} \end{pmatrix} \tag{69}$$

满足 $\Psi_{mn} = \lambda_n^{\frac{1}{2}} \pi(E_{mn})\Phi$。从其中我们知道：

$\mathbb{C}^4 = \{\pi(M)\Phi \mid M \in \mathcal{M}\}$ 且 $\forall M \in \mathcal{M}$：$(\Phi, \pi(M)\Phi) = \varphi(M)$

因此 $\{\mathcal{H} := \mathbb{C}^4, \pi, \Phi\}$ 是与态 φ 相关的正则 GNS 三元组。此外，因为 φ 被假设为是单射的，因此 $\| \pi(M)\Phi \| = 0$ 意味着 $M = 0$，即 Φ 也是从 $\pi(\mathcal{M})$ 中分离出来的。现在的关键步骤是引入定义在 H 上的两个符号 J 和 Δ，并规定 J 是反线性的而 Δ 是线性的，且它们满足：

$$J\Psi_{mn} = \Psi_{nm} \text{以及 } \Delta\Psi_{mn} = \frac{\lambda_m}{\lambda_n}\Psi_{mn}$$

请注意，由于这里的 Δ 是随其光谱分辨率一起给出的，因此这个算符的函数可以由线性来定义为 $f(\Delta)$：$\Psi_{mn} \in \mathbb{C}^4 \mapsto f\left(\frac{\lambda_m}{\lambda_n}\right)\Psi_{mn} \in \mathbb{C}^4$。特别地，$\{\Delta^{is} \mid s \in \mathbb{R}\}$ 是作用在 $\mathbb{C}^4$ 上的幺正算符的连续群。

我们立即通过上述定义来验证算符 J 和 Δ 满足下列属性。

第一，J 是等距的，$J^2 = I$，Δ 是自伴的，$J\Delta J = \Delta^{-1}$，$J\Phi = \Phi = \Delta\Phi$。 (70)

第二，

$$J\begin{pmatrix} a & b & 0 & 0 \\ c & d & 0 & 0 \\ 0 & 0 & a & b \\ 0 & 0 & c & d \end{pmatrix} J = \begin{pmatrix} a^* & 0 & b^* & 0 \\ 0 & a^* & 0 & b^* \\ c^* & 0 & d^* & 0 \\ 0 & c^* & 0 & d^* \end{pmatrix} \in I \otimes \mathcal{M} \tag{71}$$

即对于由 $\mathcal{N}$ 表示的通过 π 表征的 $\mathcal{M}$ 的图像 $\pi(\mathcal{M})$，我们有 $J\mathcal{N}J = \mathcal{N}'$，因此式(71)给出了一个由 $\mathcal{N}$ 到其换位矩阵 $\mathcal{N}'$ 的明确的双射。式(71)是一般的富田—竹崎对偶性的一种特殊的情况(见下面的定理 39)。

第三，对于任意的但确定的 $\beta > 0$，我们有 $\forall t \in \mathbb{R}$：$\Delta^{-it/\beta}\Psi_{mn} = \exp[i(\epsilon_m - \epsilon_n)t]\Psi_{mn}$。因此 $\Delta^{-it/\beta}\pi(E_{mn})\Delta^{it/\beta}\Psi_{kl} = \exp[i(\epsilon_m - \epsilon_n)t]\pi(E_{mn})\Psi_{kl}$，其中 $\epsilon_n =$

$c-(1/\beta)\ln\lambda_n$，这里 c 是任意实常数。由此推出，幺正群 $\{\Delta^{it/\beta} \mid t\in\mathbb{R}\}$ 代表$\mathcal{N}$的一个自同构群，即：

$$\tau_t:\ N\in\mathcal{N}\to\tau_t[N]=\Delta^{-it/\beta}N\Delta^{it/\beta}\in\mathcal{N} \tag{72}$$

其中，对所有 $(t,\ M)\in\mathbb{R}\times\mathcal{M}$，$\tau_t[\pi(M)]=\pi(\alpha_t[M])$，其中 $\alpha_t[M]=\exp^{iHt}M\exp^{-iHt}$，且 $H=\sum_n\epsilon_n E_{nn}$。总之，这意味着 φ 是自然温度 β 下关于上述哈密顿量 H 的正则平衡态。

第四，算符 $S=J\Delta^{\frac{1}{2}}$满足 $S\pi(E_{nm})\Phi=\pi(E_{nm})\Phi$，因此由 J 可知 S 是反线性的：

$$\forall N\in\mathcal{N}:\ SN\Phi=N^*\Phi \tag{73}$$

最后，$\mathbb{C}^4=\mathbb{C}^2\otimes\mathbb{C}^2$ 上的幺正群 $\{\Delta^{it/\beta}\mid t\in T\}$ 的生成元 L 为：

$$L=H\otimes I-I\otimes H \tag{74}$$

因此 L 的谱是关于 0 点对称的：$Sp(L)=\{\epsilon_2-\epsilon_1,\ 0,\ 0,\ \epsilon_1-\epsilon_2\}$。

注释 33。令 $\{\mathbb{C}^4,\ \pi,\ \Phi\}$ 为 2×2 矩阵代数$\mathcal{M}$上的与单射态 φ 正则相关的 GNS 三元组，并令$\mathcal{N}$为作用在$\mathcal{H}=\mathbb{C}^4$ 上的冯·诺伊曼代数 $\pi(\mathcal{M})=\{\pi(M)\mid M\in\mathcal{M}\}$。那么：

1. $\mathcal{N}$同构于$\mathcal{M}$，且 φ 可以被视为$\mathcal{N}$上的一个单射态；

2. Φ 在$\mathcal{N}$中既是循环的也是分离的；

3. 由 S：$N\Phi\in\mathcal{H}=N^*\Phi\in\mathcal{H}$定义的反线性算符具有极性分解 $S=J\Delta^{\frac{1}{2}}$，这里 J 是一个从$\mathcal{H}$到其自身的对合、反线性等距，Δ 是一个作用在$\mathcal{H}$上的正算符；

4. J 建立了$\mathcal{N}$与其换位矩阵之间的一个对偶性；特别地：$N\in\mathcal{N}\mapsto JNJ\in\mathcal{N}'$ 是一个反线性双射；

5. $\{\Delta^{-it/\beta}\mid t\in\mathbb{R}\}$ 是实现$\mathcal{N}$的自同构性 τ_t 的一个群，相对于它态 φ 满足注释 32 中的解析性条件；

6. $J\Phi=\Phi=\Delta\Phi$，$J^2=J$ 且 $\forall s\in\mathbb{R}$：$J\Delta^{is}=\Delta^{is}J$。

注释 34。回顾注释 32 和 33 中的证明，我们证实它们可以从$\mathcal{M}=\mathcal{M}(2,\ \mathbb{C})$扩展到$\mathcal{M}=\mathcal{M}(n,\ \mathbb{C})$中，其中 n 是任意有限正整数。这两个注释可以进一步扩展到$\mathcal{M}=\mathcal{B}(\mathcal{H})$中，其中$\mathcal{H}$是一个（具有可数基的）希尔伯特空间，它满足：

(i)哈密顿量 H 满足 $\mathrm{Tr}(-\beta H)<\infty$；

(ii)态 φ 是可数可加的，它保证 φ 是单射的这个条件成立。

事实上，在这种情况下，我们可以再次利用对该注释的论证过程来讨论算符 $\pi(E_{mn})$ 的所有有限线性组合的 * 代数 $\mathcal{E}=\mathrm{Span}\{\pi(E_{mn})\mid m, n=1, \cdots\}$，其中 E_{mn}：$\Psi\in\mathcal{H}\to(\Psi_n, \Psi)\Psi_m\in\mathcal{H}$，这里 $\{\Psi_n\mid n=1, 2, \cdots\}$ 是 $\mathcal{H}$ 中的一个正交基。我们可以得到从 $\mathcal{E}$ 到冯·诺伊曼代数 $\mathcal{B}(\mathcal{H})$ 的扩展，因为 φ 是可数可加的且是单射的这个假设允许我们使用标准的连续性论证，例如[Dixmier(狄克斯梅尔)，1957，Theorem I. 3. 5，lemma I. 4. 4，Proposition I. 4. 1]或[Kadison and Ringrose(卡迪孙和林罗斯)，1983/1986，Volume ii，Chapter 7]。特别地，$\mathcal{N}=\pi(\mathcal{B}(\mathcal{H}))=\{\pi(\mathcal{M})\mid M\in\mathcal{B}(\mathcal{H})\}$ 是一个冯·诺伊曼代数——即 $\mathcal{N}=\mathcal{N}''$——并且它与 $\mathcal{B}(\mathcal{H})$ 是同构的。由于 $\mathcal{B}(\mathcal{H})$ 是一个因数，因此 $\mathcal{N}$ 也是，即该冯·诺伊曼代数的中心是平凡的：$\mathcal{N}\cap\mathcal{N}'=\mathbb{C}\,I$。此外，$\mathcal{N}$ 可被等同于 $\mathcal{B}(\mathcal{H})\otimes\mathbb{C}\,I$，$\mathcal{N}'$ 可被等同于 $\mathbb{C}\,I\otimes\mathcal{B}(\mathcal{H})$。

量子力学的冯·诺伊曼形式[von Neumann，1932c]允许我们做到这些，但不能再进一步了。回顾一下，在3.3节中我们给出了为什么需要更进一步的原因。下一小节为实现这一目标提供了重要工具。

4.2 久保—马丁—施温格条件和富田—竹崎(Tomita-Takesaki)理论

上述结果给出了三个定义，前两个只是数学术语方面的定义，第三个定义是本节的重点。

定义35。冯·诺伊曼代数 $\mathcal{N}$ 上的一个态 φ 被认为是正常的，只要它是可数可加的，即 $\varphi(\sum_n P_n)=\sum_n\varphi(P_n)$ 对于 $\mathcal{N}$ 上由相互正交投影组成的每个可数族 $\{P_n\}$ 成立。

这可被简单地扩展为一般的冯·诺伊曼代数条件式(35)，在[von Neumann，1932c]中被认为是概率测度完全可加性的量子等价属性。下面这个定义中的一些概念与上一小节中一些例子中的概念会发生冲突。

定义36。作用于希尔伯特空间 $\mathcal{H}$ 的一个冯·诺伊曼代数 $\mathcal{N}$ 被称作是标准形式的，当存在一个矢量 $\Phi\in\mathcal{H}$，它在 $\mathcal{N}$ 中既是循环的也是分离的，即 $\mathcal{N}\Phi$ 在 $\mathcal{H}$ 中

是规范密集的，且对于 $N\in\mathcal{N}$，$N\Phi=0$ 意味着 $N=0$ 时。

备注 37。这个概念已经存在了很长一段时间，但我们似乎可以公正地说，直到富田—竹崎模理论[Tomita，1967；Takesaki，1970a]的提出，人们才充分肯定了它在冯·诺伊曼代数的一般理论中的重要性。在最基本的层次上，注意到如果$\mathcal{N}$是标准形式，那么我们可以不失一般性地假设 $\|\Phi\|=1$，以使得 φ：$\mathcal{N}\in\mathcal{N}\to(\Phi,\ N\Phi)\in\mathbb{C}$ 是$\mathcal{N}$中的一个单射的正常态。

相反，从上面注释 34 中所使用的相同的连续性论证可知，如果 φ 是冯·诺伊曼代数$\mathcal{N}$中的一个任意正常态，那么与 φ 对应的 GNS 表示 π 就是一个冯·诺伊曼代数；如果 φ 是单射的，那么$\mathcal{N}$是同构于 $\pi(\mathcal{N})$ 的。因此，正则 GNS 矢量 Φ 不仅是循环的，也是分离的。因此，每当 φ 是一个单射正常态，那么$\mathcal{N}$是同构于以标准格式呈现出来的一个冯·诺伊曼代数 $\pi(\mathcal{N})$。

第三个定义适合作为本节的核心。这是对[Kubo，1957；Martin and Schwinger，1959]的工作的改进，[Haag *et al.*，1967]将其作为对全局性 C^* 代数中正则平衡态定义的扩展，它被认为是与无限系统相关的。

定义 38。令$\mathcal{A}$为一个 C^* 代数，并令 α：$t\in\mathbb{R}\to\alpha_t\in\mathrm{Aut}(\mathcal{A})$ 为一个$\mathcal{A}$的自同构群。$\mathcal{A}$上的一个态 φ 被认为在自然温度 β 下相对于 α 满足 KMS 条件，如果对于$\mathcal{A}$上的每对元素 $(A,\ B)$，存在一个定义在带 $\Omega_\beta=\{z\in\mathbb{C}\mid 0\leqslant\mathrm{Im}z\leqslant\beta\}$ 上的函数 $f_{A,B}$，使得 $f_{A,B}$ 在 Ω_β 上是有界的且是连续的，在 Ω_β 内部是解析的，以及 $\forall t\in\mathbb{R}$：$f_{A,B}(t)=\varphi(A\alpha_t[B])$ 和 $f_{A,B}(t+i\beta)=\varphi(\alpha_t[B]A)$。

现在可以说明本节中的主要数学结果了，它们主要来自从富田—竹崎模理论[Tomita，1967；Takesaki，1970a]。

定理 39(富田—竹崎)。令$\mathcal{N}$是作用于希尔伯特空间$\mathcal{H}$上的一个冯·诺伊曼代数，并允许一个循环和分离的单位矢量 Φ 存在。那么闭合反线性算子 S 被认为是映射 $N\Phi\to N^*\Phi$ 的闭包，其中所有的 $N\in\mathcal{N}$，它具有极性分解 $S=J\Delta$，其中 $J=J^2$ 是从$\mathcal{H}$到其自身的反线性等距，满足 $J\mathcal{N}J=\mathcal{N}'$；且 Δ 是一个正的自共轭算子(不一定有界!)，并使得 $J\Delta^{it}=\Delta^{it}J$；且对于任意 $\beta>0\ \forall\ (t,\ N)\in\mathbb{R}\times\mathcal{N}$：$\tau_t[N]=\Delta^{-it/\beta}N\Delta^{it/\beta}$ 定义一个$\mathcal{N}$的 * 自同构群 $\{\tau_t\}$，涉及它，单射正常态 φ：$N\in\mathcal{N}\to(\Phi,\ N\Phi)\in\mathbb{C}$ 对 β 满足 KMS 条件。此外 $\{\tau_t\mid t\in\mathbb{R}\}$ 是唯一的$\mathcal{N}$的 * 自同构群，当涉及它，φ 满足这个条件。

备注 40。

1. 该定理可推广到标准形式下的任意冯·诺伊曼代数，我们——在备注 34 中——将这个结果描述为与其任意单射正常态有关的$\mathcal{B}(\mathcal{H})$的 GNS 表征。

2. 这次回顾的主要目的是为了强调该定理并不要求$\mathcal{N}$是一个因子。

3. 该定理断言动力学τ是由 KMS 条件唯一确定的，但反过来却是不正确的。当$\mathcal{N}$是一个因子时，还存在$\mathcal{N}$上的一些其他正常态，涉及相同的动力学时，也满足 KMS 条件。事实上，当$\mathcal{N}$是一个因子时，我们可以证明对于每个 $Z\neq 0$，它属于中心$\mathcal{Z}=\mathcal{N}\cap\mathcal{N}'$，$\psi$：$N\in\mathcal{N}\to[\varphi(Z^*Z)]^{-1}\varphi(Z^*NZ)$定义了一个正常态，涉及$\tau$，对于相同的$\beta$，它再次满足 KMS 条件。我们将在 5.6 节中证明这个注释，它对于 5.7 节中的论证是至关重要的。

4. 除了其数学吸引力以外，KMS 条件在宏观系统的 QSP 中，可以被视为正则平衡态的定义，这个猜想的合理性也会在下一节中讨论。

5. 最后，数学的真实性要求我们提到——无论是否相关——证明定理 39 的主要困难在于说明映射 $N\Phi\to N^*\Phi$ 是封闭的。解决这个问题的方法，参见[Tomita，1967；Takesaki，1970a]这一最原始的论文。我们需要提醒读者，即使是[Kadison and Ringrose，1983/1986，Chapter 9]中的数学式推导也已经超出了我们这篇文章的范围。但为了表达定理中所涉及的关于结构的观点，我采用了注释 32 和 33 中的模型，在那里它被视为一种基本的数学工具。但缺点是这些模型，以及在备注 34 中描述的将它们从$\mathcal{M}(2,\mathbb{C})$到$\mathcal{B}(\mathcal{H})$的常规扩展，只涉及了因子，事实上在$\mathcal{B}(\mathcal{H})$的单射表征中，那不足以包含 QSP 的宏观作用，在那里无限多的自由度会起作用。正如哈格、胡戈霍兹(Hugenholtz)以及威尼克(Winnink)[Haag *et al.*，1967]正确设想的，实际上在物理应用中需要的正是定理 39 中涉及的概括性。这种物理直觉和富田—竹崎[1967；1970a]的数学理论的提出在时间上非常一致，这真是一件了不起的事，对此事件的生动讲述请参见[Kadison，1990，pp. 77—79]。

5. KMS 条件、QSP 和热力学

本节提供一些证据，以说明[Haag *et al.*，1967]中提出的对 KMS 条件所作的物理解释可被看作 QSP 中平衡态的另一个定义。我们已经看到，对于有限系统，冯·诺伊曼正则平衡态满足 KMS 条件，且只有这些态满足。这里，我将在 5.1—5.3 节中描述一些模型以表明，由[Tomita，1967；Takesaki，1970a]提出与发展的分子结构——我们认为(参见批注 33)在正则平衡态的有限系统中是实现了的——也出现在无限系统的平衡态 QSP 中，因而允许人们超越冯·诺伊曼体系[von Neumann，1932c]。在 5.4 节中将展示不同的稳定条件，以给出在 QSP 中 KMS 态下的热力学特征。在 5.5 节中将进行一次简短回顾以阐明那些后来被认为在相对论 QFT 中起作用的 KMS 条件的重要性，这种作用被预言者称为“革命性的”。5.6 节致力于探讨极值 KMS 态的代数特征的数学部分。在 5.7 节中我们回到 QSP，讨论存在相位跃迁的系统，并表明将任意正则平衡态按照其纯热力学相位进行的唯一分解都可以通过将 KMS 态按其极值 KMS 态进行的唯一分解接近地模拟。特别地，此节致力于证明如下重要观点，即 KMS 条件为无限系统的热力学提供了概念框架，其中相位跃迁的发生伴随着自发对称破缺。

5.1 超越福克(Fock)空间：BCS 模型

在 QSP 中，表明在应用冯·诺伊曼体系的过程中存在差错的第一个迹象出现在超导的巴丁—库珀—施里弗(Bardeen-Cooper-Schrieffer)模型中，即 BCS 模型中。确实，在对此模型的最初处理[Bardeen *et al.*，1957]中，选择用以描述有限但很大的金属体中电子间具体相互作用的哈密顿量，在第一类规范变换下是不变的。之后提出了一种近似，该近似被认为在无限体积极限下会变得准确，但在此形式化过程中，上述对称性丧失了。此外，所得哈密顿量的谱线呈现出依赖于温度的能量差距。人们可能会认为，经验主义者可能不希望涉及对称性的破缺，但能量差距不能够被忽略，经验主义者要在实验室测量它。因此，数学物理学家想到，他们应当去理解——如何或在多大程度上——哈密顿

量本身可能确实依赖于温度。在五年内，哈格[1962]找到了问题的原委，那就是整个处理据称是在 CCR 的一个固定的不可约化表征下进行的，因而普遍存在的福克表征和上述约束破坏了这个模型。

特别地，初始哈密顿量为：

$$H_{\Lambda} = \sum_{p,s} \epsilon(p) a_s(p)^* a_s(p) + \sum_{p,s} b(p)^* \tilde{v}(p, q) b(q) \tag{75}$$

其中 Λ 是系统包含于其中的空间区域，往往是有限体积 $|\Lambda|$ 的立方体；p 与 q 表示动量，它们是 $2\pi|\Lambda|^{-\frac{1}{3}}$ 的整数倍；$s=\pm\frac{1}{2}$；$a_s(p)^*$ 和 $a_s(p)$ 是自旋为 S 和动量为 p 的电子的生成和湮灭算符；$\epsilon(p) = -\mu + \frac{1}{2}p^2/2m$ 是动量为 p 的自由电子的能量；$b(p)^* = a_{\uparrow}(p)^* a_{\downarrow}(-p)^*$ 是所谓库珀对的生成算符；$b(p)^* \tilde{v}(p, q) b(q)$ 是两个库珀对，即四个电子之间的相互作用能，因而哈密顿量(75)在原先的场算符中并不是二次的。$\tilde{v}(p, q)$ 的形式随后会讨论。

近似哈密顿量为：

$$\tilde{H}_{\Lambda} = \sum_{p,s} E(p) \gamma_s(p)^* \gamma_s(p) \tag{76}$$

其中 $\gamma_s(p)^*$ 和 $\gamma_s(p)$ 是由博戈留波夫—瓦拉廷(Bogoliubov-Valatin)变换给出的初级激发的生成和湮灭算符：

$$\left.\begin{aligned} \gamma_{\uparrow}(p) &= u(p) a_{\uparrow}(p) + v(p) a_{\downarrow}(-p)^* \\ \gamma_{\downarrow}(p) &= -v(-p) a_{\uparrow}(-p)^* + u(-p) a_{\downarrow}(p) \end{aligned}\right\} \tag{77}$$

其中：

$$\begin{aligned} E(p) &= \{\epsilon(p)^2 + [\Delta(p)\Delta(p)^*]\}^{\frac{1}{2}} \\ D(p) &= \{[E(p) - \epsilon]^2 + [\Delta(p)\Delta(p)^*]\}^{\frac{1}{2}} \end{aligned} \quad ; \quad \begin{aligned} u(p) &= \Delta(p)^* / D(p) \\ v(p) &= [E(p) - \epsilon(p)] / D(p) \end{aligned} \tag{78}$$

其中 Δ 满足非常重要的自洽性方程：

$$\Delta(p) = -\sum_q \tilde{v}(p, q) \frac{\Delta(q)}{2E(p)} \tanh\left(\frac{1}{2}\beta E(q)\right) \tag{79}$$

显然 $\Delta=0$ 总是一个解，在此情形下 H 与 $\tilde{H}$ 的谱相重合，这是常规相，在该相位中不会有特别有趣的现象发生。模型的本质是存在一个关键的温度 T_c

(回想$\beta=1/kT$)，低于它我们会得到热力学更感兴趣的解$\Delta\neq0$。这对应于超导相位，我们今后继续讨论$0<T<T_c$的情形。

这是我们感兴趣的相位，因而有必要用物理学术语回顾一下物理学家们第一次在式(76)—式(79)中发现了什么。BCS设计了一个有限的过程——涉及热力学极限和一个“平均场近似”(弱的，但是非常长过程的相互作用)——通过它初始哈密顿量式(75)和新的哈密顿量式(76)变得可交换了，在它们被认为会导致相同的极限这一意义上。虽然式(75)是用电子的生成和湮灭算符$a_s^b(p)$表示的，但新的哈密顿量式(76)却与初级激发γ_s^b无关。这些激发的能量谱为$\{E(p)\}$，且在实验室中的可观测到的依赖于温度的“能级差”方面它与自由电子的能量谱$\{\epsilon(p)\}$不同——见式(78)，由这些能级差得到的数值结果与式(79)的预言符合得很好，参见[Schrieffer，1974，Figure 1—3]。

然而我们需要对数学图像做出一些解释。因为：(i)对于任意的$\theta\in(0,2\pi]$，初始哈密顿量式(75)在定义为$a_s(p)\to\exp(i\theta)a_s(p)$的规范对称下不变，而哈密顿量式(76)则不是；(ii)哈密顿量(76)的能量谱$\{E(p)\}$是依赖于温度的，而在式(75)中不存在温度依赖性。

因此问题就是我们需要解释人们如何能——正如在主流传统中那样——宣称这些近似值在热力学极限下会变为精确值。对此，人们不得不考察Δ的出处，即被看作是算符$\hat{\Delta}(p)=\sum_q\tilde{\nu}(p,q)b(q)$的近似的$\Delta(p)$是单位算符的标量倍数。对此的论证建立在如下论点之上，即在对$\tilde{\nu}$做出的适当假设下，我们可以认为$\hat{\Delta}(p)$的极限$|\Lambda|\to\infty$——在弱算符拓扑学中——存在且与所有生成和湮灭算符$a_s(q)^*$和$a_s(q)$对易，这二者产生的代数被默认为是不可约化的。在此极限下，算符$\Delta(p)$会被单位元(identity)的标量倍数所取代。当$\tilde{\nu}$是双傅里叶变换$\tilde{\nu}(p,q)=\int_\Lambda \mathrm{d}x\mathrm{d}yf(p,x)\nu(x,y)f(q,y)^*$，其中$f(p,x)=\begin{cases}|\Lambda|^{-\frac{1}{2}}e^{ipx} & x\in\Lambda\\ 0 & x\notin\Lambda\end{cases}$，某些“适当”的假设似乎是成立的，其中非定域势$\nu$满足$\nu(x,y)^*=\nu(y,x)$，$c=\int\mathrm{d}x\mathrm{d}y|\nu(x,y)|<\infty$和$\sum_q|\tilde{\nu}(p,q)|<\infty$，因

而有$\lim_{\Lambda\to\infty}|\tilde{\nu}(p,q)|=0$ *and* $|\tilde{\nu}(p,q)|\leqslant c/|\Lambda|$。

在这里专家们会注意到对平均分子场类型所做的近似，这是在20世纪的前10年中，由外斯(P. Weiss)和奥恩斯坦(L. S. Ornstein)在相变的经典理论中引入的一种启发性工具。然而，在没有某些进一步的讨论时，这里的近似不会被接受，因为它导致了上面提到的悖论。

我们现在认识到了哈格的巨大贡献[Haag，1962]，场代数不可约化的默认假设是站不住脚的。放弃该假设允许人们求解悖论：Δ和博戈留波夫—瓦拉廷变换式(77)中的系数u和v——而非单位元的倍数——现在属于由与对应于系统平衡态的GNS解释正则相关的表征的非平凡中心$\mathcal{Z}$。现在规范群以一种不平凡的方式作用于$\mathcal{Z}$，因而重建了理论的对称性。且在所考虑的极限下，时间演化被很好地定义为是由表征产生的冯·诺伊曼代数上的自同构群。在随后的研究中，这些技术细节被连续地改进与确证，参见如[Emch and Guenin(埃姆什和归宁)，1966；Thirring and Wehrl(西凌和维尔利)，1967；Thirring，1968；Dubin and Sewell(杜宾和休厄尔)，1970；Sewell，1982b]。

5.2 超越福克空间：玻色气体

即使在数学家们从形式上理解分子结构之前，它们的第一个实例就出现在了QSP中。人们确实可以在荒木和伍兹(Araki and Woods)[1963]为理想量子气体所做的玻色—爱因斯坦模型的开创性的重新审视中辨认出这些结构。关于模型的最初版本，请参见2.4节。本节将总结荒木—伍兹方法的主要方面。

假定读者熟悉具有可数无限多自由度的正则对易关系(CCR)的外尔(Weyl)形式的定义，即作用于(玻色子)福克空间$\mathcal{F}:=\oplus_{N=0}^{\infty}{}^{s}\mathcal{H}^{N}$上且满足$\forall f,g\in\mathcal{D}(\mathbb{R}^3)$：$W(f)W(g)=\exp\{-i\mathrm{Im}(f,g)/2\}$的幺正算符族$\{W(f)\mid f\in\mathcal{D}(\mathbb{R}^3)\}$，其中$\mathcal{D}(\mathbb{R}^3)$是所有无限可微函数$f$：$\mathbb{R}^3\to\mathbb{C}$的空间，其具有紧致支集，且${}^{s}\mathcal{H}^{N}$是只包含自身的单粒子空间$\mathcal{H}^1=\mathcal{L}^2(\mathbb{R}^3)$内的对称的$N$重张量积，参见[Emch，1972a]或[Halvorson(霍尔沃森)，2006]。

温度为$T>T_c$的玻色气体，其中T_c是由玻色与爱因斯坦发现的临界温度，对应于气体的正常相位的GNS表征π_g——在热力学极限下和确定的密度ρ和化学活性z下——由下式给出。表征π_g的希尔伯特空间可以等同于$\mathcal{H}=\mathcal{F}\otimes\mathcal{F}$，

其循环矢量为 $\Phi=\Phi_o\otimes\Phi_o$，其中 Φ_o 是$\mathcal{F}$中的真空矢量。则：

$$\pi_g[W(f)]=W(\zeta_+f)\otimes W(K\zeta_-f) \tag{80}$$

其中完备性条件要求我们明确指出$(\zeta_+f)^{\sim}(k)=[1+\rho(\beta,\ z;\ k)]^{\frac{1}{2}}\tilde{f}(k)$，以及 $(\zeta_-f)^{\sim}(k)=[\rho(\beta,\ z;\ k)]^{\frac{1}{2}}\tilde{f}(k)$，$(Kf)^{\sim}(k)=\tilde{f}(k)^*$；$\rho(\beta,\ z;\ k)=z[\exp(\beta\epsilon(k))-z]^{-1}$，其中 $\epsilon(k)=|k|^2/2m$，且 z 是由 ρ 和 β 通过 $\rho=(2\pi)^{-3}\int d^3k\rho(\beta,\ z;\ k)$确定的。

冯·诺伊曼代数$\mathcal{N}_g=\{\pi_g[W(f)]\mid f\in\mathcal{D}(\mathrm{IR}^3)\}''$是一个因数，其对易量为 $N'_g=\{\nu_g[W(f)]\mid f\in\mathcal{D}(\mathrm{IR}^3)\}'$，其中：

$$\nu_g[W(f)]=W(K\zeta_-f)\otimes W(\zeta_+f) \tag{81}$$

注意 ν_g 同样给出了外尔 CCR 的一个表征。

我认为第一次面向大多数主流物理学家提出这个方案的是[Haag *et al.*, 1967]，即在 1966 年关于统计力学的 IUPAP 哥本哈根会议上，威尼克[Winnink, 1967]确实是由概括上述结果来开始其报告的。正如冯·诺伊曼代数与其交换子之间的二元性已经是有限系统的一个特征那样——见上文注释 33 和评论 34——威尼克强调，这一特征可能普遍存在于具有无限多自由度的系统中，正如在——玻色气体——这个具体例子中那样，正则平衡态的热力学极限是可控的。强调对无限系统的处理——同样提出于威尼克的演讲之前［Verboven(沃博芬)，1967]——震惊了哥本哈根会议中的许多物理学家，“这样来说吧，人们难道不会认为转向研究无限系统的动机是为了得到比在有限系统中更为简单的结果吗?”[Uhlenbeck(乌伦贝克)，1967]，或有人更直接地认为：“这与统计力学有什么关系?”[van Kampen(范·坎朋)，1967]。已经出现的一种假设是形式体系对于充分描述相位跃迁是有用的，我将在5.6 与5.7 节中分析该假设。

回想起来，不同寻常的是荒木与伍兹[1963]已经发掘了诸多后来在普遍理论背景下的特征，这些理论建立在富田和竹崎后来关于数学形式体系的工作[Tomita，1967；Takesaki，1970a]以及哈格等人[1967]关于其在 QSP 中应用的研究基础之上。在荒木与伍兹的结果中，人们可能会注意到，他们针对 $T>T_c$ 的情况构造的冯·诺伊曼系数$\mathcal{N}_g$属于类型Ⅲ——这类参数人们已知其存在，但即使是在纯粹的数学文献中，也非常难以找到这样的例子——且这是此类型的

参数在 QSP 中第一次出现，虽然其普遍性随后在 QSP 与 QFT 中都被认识到了，且在纯粹的数学中也找到了，不过那都是后话了。此外，荒木与伍兹发现，在冯·诺伊曼代数中执行时间演化和空间平移的幺正算符$\mathcal{N}_g$并不属于这种代数。他们还讨论了在温度 $0<T<T_c$ 下超流相位的表征，且他们发现了相关的 GNS 表征是参数表征的积分。顺便值得一提的是，他们也提到这在形式上类似于哈格在其对 BCS 模型的研究中发现的数学结构，见上文 5.1 节。

5.3 KMS 条件和海森堡模型

KMS 条件自身事实上是实无穷量子系统的第一个证据，由荒木[1969]在研究一维量子自旋格的一类模型时给出，该模型包含的原始模型——最初由海森堡提出，是关于铁磁性的想象模型——由定域性的所谓的“交换”哈密顿量定义：

$$H_{\Lambda}=-J\sum_{k=a}^{b-1}\sigma_k\cdot\sigma_{k+1} \tag{82}$$

其中 J 用来描述位于一个规则的一维有限弦 $\Lambda=\{a,b\}\subset Z^1$ 上相邻量子自旋 $\sigma_k=(\sigma_k^x,\sigma_k^y,\sigma_k^z)$ 的相互作用的耦合常数，$\sigma_k\cdot\sigma_{k+1}$ 表示 $\sigma_k^x\sigma_{k+1}^x+\sigma_k^y\sigma_{k+1}^y+\sigma_k^z\sigma_{k+1}^z$。

确定此量子模型是否会在其热力学极限下支持铁磁性这个问题——即使在这个一维版本下——要比只考虑相互作用 $J\sigma_k^z\sigma_{k+1}^z$ 的经典伊辛(Ising)模型更加难以掌握。

对于经典模型，存在一种被称为转换矩阵的方法——事实上是针对二维伊辛模型提出的[Kramers and Wannier(克莱默和瓦尼尔)，1941]——它允许人们在有相互作用相邻最近的少数列中探讨此经典模型的一维版本，甚至在严格有限程相互作用下，即当超过这一有限距离时，自旋间的相互作用严格为零(对所有对都一样)的情况下进行讨论。这方面的大量研究[Ruelle(吕埃勒)，1968b]表明，该方法可以拓展到无限相互作用中，当衰减足够快以使得系统具有有限动量，或使得表面能量具有独立于体积的边界时。

正如[Emch，1972b]中某些细节所强调的，即使相邻最近的量子化海森堡模型也要求扩展类似于在无限区域经典情形下用到的那种方法。因而荒木

[1969]试图足够好地控制正则平衡态的热力学极限 φ 及其时间关联函数，来证实在所有正温度 β 下态 φ 满足 KMS 条件；且相对于该条件它是极值的——即不能分解为满足 KMS 条件的混合态——并表现为不会自发磁化。因而，当物理学家们——根据他们在经典情形下对集约行为初始点的理解——预言此量子一维模型将不会展现任何铁磁相位跃迁现象时，荒木则证明了它会。

荒木得到上述结果所基于的那类模型强烈地依赖于“点阵” $\mathbb{Z}$的一维性。不过，论证过程并不要求相互作用是各向同性的，即自旋不同分量间的耦合并不需要在所有方向上都相同。此外，论证过程也不要求点阵上自旋之间的相互作用局限于最近的邻域内，在论证的最初版本中，只要求在自旋之间超过一个固定的(但任意的)有限距离时相互作用会消失，但即使此限制也被放宽，以涵盖与相应的经典模型相同的范围，该论证也成立。最后，当在海森堡模型中代数 $\mathcal{M}(2,\mathbb{C})$ 的拷贝可用来描述单个 1/2 自旋时，该论证过程也可用于 $\mathcal{A}_k \simeq \mathcal{M}(n,\mathbb{C})$ 的情形，其中 $n<\infty$。

因此荒木的结果支持了这样的推测，即纯热力学相位可以用极值 KMS 态来描述，更多的论证过程见下文 5.7 节。

此外，还需补充的是，如果点阵平移不变性是通过局域哈密顿量 H_Λ 在理论中建立的话，即如式(82)中的例子那样，则极值 KMS 态 φ 在点阵平移群 $\mathbb{Z}$ 下是不变的，且——由于其 GNS 表征产生一个参数——φ 在此条件下同样是极值函数，从而自旋之间的空间相互作用随着其距离的增加而衰减得非常快。这里从技术上讲，φ 在空间中是指数性地统一地集中的，即对于任意准定域可观测量 $A\in\mathcal{A}_o$，存在正常数 γ 和 δ，使得对于所有有限的 N 和所有的 $B\in\mathcal{A}_{\mathbb{Z}\setminus[-N,N]}$：$|\varphi(AB)-\varphi(A)\varphi(B)|\leqslant\delta\,\|B\|\exp(-\gamma N)$。

在结束讨论海森堡模型之前，注意在 $T=0$ 情形下，它同样为 QFT 提供了一个很好的简单模型，参见[Streater(斯蒂特)，1967]。

5.4　KMS 条件与稳定性

如下五点概括了我之前讨论过的关于 KMS 的历史发展的情况。

1. 冯·诺伊曼关于在限定温度下量子正则态的定义仅限于有限系统，且将此限制应用于 QSP 的形式体系时(往好的方面说)会显得很累赘——见 3.3 节。

2. 对于有限系统，冯·诺伊曼平衡态刚好是那些满足形式分析条件，即所谓 KMS 条件的态——见 4.1 节。

3. KMS 条件可以扩展到由冯·诺伊曼主张的数学形式体系之外——见 4.2 节。

4. 在某些无限系统的具体模型中，KMS 条件能够得以满足，对于像温度平衡态那样具有可靠解释的态来说——见 5.2 和 5.3 节。

5. 在纯数学背景中，KMS 也表现为关于模代数的富田—竹崎理论，它被证明是非常富有成效的。虽然这方面的探讨超出了本文的范围，但在 4.2 与 5.2 节我还是对此进行了简单的讨论。

KMS 理论至少还应当满足另外两个要求，才能被看作是一个物理学理论，即：(i)该形式体系应该允许在超出冯·诺伊曼形式体系之外对量子现象进行数学描述；(ii)KMS 态应该是稳定的。上述的 5.2 与 5.3 节表明了如何满足前一个要求，更多的案例参见 5.7 节。本节论述第二个要求，讨论不同的稳定性标准——从 A 到 E。论述的顺序是为了将读者的注意力引向形式体系的不断发展上来，由此我们将会逐渐用变分原理的术语来描述 KMS 态。

A. 剪切与粘贴稳定性

我们首先讨论一个足够简单的模型，它能够提供精确的结果以支持这样一种期望，即处于正则平衡态的大系统应该能作为其“任何”部分的热容器。该模型是所谓的 X—Y 模型的变异，该变异模型由[Emch and Radin(埃姆什和雷丁)，1971]提出，且给出了解释。更多的参考文献将在本节的最后给出。

X—Y 模型本身——下文中被当作“未分割系统”——是具有有限程相互作用的一维量子自旋点阵气体。具体地，对于任意区域 $\Lambda=\{k\in\mathbb{Z}\mid a\leqslant k\leqslant b\}$，其中 $-\infty<a+1<b<\infty$，哈密顿量为：

$$H_{[a,b]}=-\sum_{k=a}^{b-1}(1+\zeta)\sigma_k^x\sigma_{k+1}^x+(1-\zeta)\sigma_k^y\sigma_{k+1}^y \tag{83}$$

从荒木的工作中(见上文 5.3 节)我们知道，下述两个对象的热力学极限($a\to-\infty$，$b\to\infty$)存在：(i)在有限自然温度 $\beta>0$ 下的正则平衡态 $\varphi_{[a,b]}$；(ii)时间演化 $\alpha_{[a,b]}$，且得到的无限系统的态 φ 和演化 α 满足 KMS 条件。

现在我们将总系统分成两个不会产生相互作用的部分：一个我们加下标 S

表示有限区域，另一个我们加下标 R 表示无限区域，它是$\mathbb{Z}$中 Λ_S 的补，即：

$$\Lambda_S = [c, d] \text{和} \Lambda_R = (-\infty, c-1] \cup [d+1, \infty)$$

其中 $-\infty < a < c-1$；$c < d-1$；$d+1 < b < \infty$。

这个被分割的系统可以被看作是具有下述哈密顿量的有限系统的热力学极限：

$$\widetilde{H}_{[a,b]} = H_{[a,c-1]} + H_{[c,d]} + H_{[d+1,b]} \tag{84}$$

显然原始系统和被分割的系统的 C^* 代数相同，为 C^* 归纳极限$\mathcal{A} := \otimes_{k\in\mathbb{Z}}\mathcal{A}_k$，其中$\mathcal{A}_k$是具有复数输入的 2×2 矩阵的代数$\mathcal{M}(2, \mathbb{C})$的拷贝。因而有$\mathcal{A} = \mathcal{A}_S \otimes \mathcal{A}_R$，其中$\mathcal{A}_S := \otimes_{k\in\Lambda_S}\mathcal{A}_k$和$\mathcal{A}_R := \otimes_{k\in\Lambda_R}\mathcal{A}_k$。

再有，对于原始(未分割的)系统，正则平衡态的热力学极限以及由(84)定义的分割后系统的演化的热力学极限，都是存在的且满足 KMS 条件，这里用 $\widetilde{\varphi}$ 和 $\widetilde{\alpha}$ 来表示它们。

注意 φ 和 $\widetilde{\varphi}$ 是不同的。例如，φ 相对于沿链轴的平移是不变的，而 $\widetilde{\varphi}$ 则不是。不过，此模型的第一个稳定性特征是在[Emch and Radin，1971]中得到的，即：

$$\forall A \in \mathcal{A}: \lim_{|t|\to\infty} \widetilde{\varphi}(\alpha_t[A]) = \varphi(A) \tag{85}$$

所以，如完全演化 α 所揭示的，原来由于分割导致断裂的 S 和 R 之间的关联被重新建立：分割被取消。

此外，令 $\widetilde{\varphi}_S$ 表示 $\widetilde{\varphi}$ 对$\mathcal{A}_S$的限制，对 R 中的操作类似，我们有 $\widetilde{\varphi} = \widetilde{\varphi}_S \otimes \widetilde{\varphi}_R$。演化 $\widetilde{\alpha}$ 会保持分割存在，即 $\forall A \in \mathcal{A}_S$[对于$\mathcal{A}_R$]：$\widetilde{\alpha}[A] \in \mathcal{A}_S$[对于$\mathcal{A}_R$]。所以，我们有 $\widetilde{\alpha} = \widetilde{\alpha}_S \otimes \widetilde{\alpha}_R$，即两个系统相互独立地演化。$S$ 和 R 再次分别满足 KMS 条件。

在此次分割之后，我们交换 S 和 R 的温度，从而：(i)涉及演化 $\widetilde{\alpha}_S$，$\widetilde{\varphi}_{S,\beta_S}$是$\mathcal{A}_S$中在特定温度 β_S 下的正则平衡态；(ii)$\mathcal{A}_R$中，在温度 β_R(可能不同于 β_S)下，有 $\widetilde{\varphi}_{R,\beta_S}$(写作 $\widetilde{\alpha}_R$)。因而令 $\widetilde{\varphi}_{S,\beta_S} \otimes \widetilde{\varphi}_{R,\beta_R}$为被分割系统的初态。涉及最初的相互作用演化 α 时，将温度 β 时整个系统的正则平衡态记作 φ_β。那么，对于所有 β_S，$\beta_R > 0$ 以及所有 $A \in \mathcal{A}$，下式成立[Emch and Radin，1971]：

$$\lim_{|t|\to\infty}\widetilde{\varphi}_{S,\beta_S}\otimes\widetilde{\varphi}_{R,\beta_R}(\alpha_t[A])=\varphi_\beta(A)\text{ ，其中 }\beta=\beta_R \tag{86}$$

因而其名称为“剪切与粘贴稳定性”。系统首先被分割为两部分，一个被无限系统 R 包围的有限系统 S，二者并不发生相互作用：跨越边界(即介于 $c-1$ 处和 c 处，以及 d 和 $d+1$ 处)的相互作用被“切断了”。在此结构中，有限系统 S 和无限系统 R 分别被置于(不同的)温度 β_S 和 β_R 下。当这些系统被“粘贴”在一起时，人们发现联合演化使得总系统 $S\cup R$ 的温度达到了 β，这是 R 最初所处的温度，即 $\beta=\beta_R$。在此意义下，R 成为了 S 的热容器。

导致这种结果的模型的特殊性质是，它满足我现在要描述的显著条件。

令 γ 是$\mathcal{A}$的自同构，由下式唯一地确定：

$$\forall k\in Z:\ \begin{cases}\gamma[\sigma_k^z]=\sigma_k^z\\ \gamma[\sigma_k^x]=-\sigma_k^x\\ \gamma[\sigma_k^y]=-\sigma_k^y\end{cases} \tag{87}$$

特别注意初始哈密顿量 $H_{[a,b]}$ 和剪切哈密顿量 $\widetilde{H}_{[a,b]}$ 属于偶代数 $\mathcal{A}_e:=\{A\in\mathcal{A}\mid\gamma[A]=A\}$。这在热力学极限 $\varphi\circ\gamma=\varphi$ 和 $\gamma\circ\alpha\circ\gamma=\alpha$ 下是必然的；且类似地，对于所有在分割之后得到的相应客体来说也是这样。特别地，演化保持为偶代数，即 $\forall(t,A)\in\mathbb{R}\times\mathcal{A}_e$：$\alpha_t[A]\in\mathcal{A}_e$。

现在，模型的特殊性质可以明确地表示为：

$$\forall A,\ B\in\mathcal{A}_e:\ \lim_{|t|\to\infty}\|A\alpha_t[B]-\alpha_t[B]A\|=0 \tag{88}$$

即当限制为模型的偶可观测量时，演化是强渐近阿贝尔的。

上述式(86)的论证——和当时的推广——是量子遍历理论的直接结果(见3.5节)。首先，人们注意到 φ 在空间中是均匀聚集的，即对于每一个 $\epsilon>0$ 和每一个 $A\in\mathcal{A}$，都存在一个有限的区域 Λ，使得对于每个在此区域外的 B 有 $|\varphi(AB)-\varphi(A)\varphi(B)|\leqslant\epsilon\|B\|$。这意味着 KMS 态 φ 相对于该条件是极值的，即不能够分解为是其他 KMS 态的凸组合，见下文5.6节，特别是定义57。这些特征是通过将 φ_e 和 φ 限制为偶代数$\mathcal{A}_e$而获得的。因此演化的渐近阿贝尔性意味着[Araki and Miyata(荒木和宫崎)，1968]态 φ_e 不只是时间演化不变的——正如我们所知道的每个 KMS 态必须是时间演化不变的——而且同样它相对于该条件是极值的，即不能够分解为其他时间演化不变的态，也就是说 φ_e

不能被写为 $\varphi_e = \lambda\psi_e + (1-\lambda)\chi_e$，其中 $0<\lambda<1$，且 ψ_e，χ_e 是时间演化不变的，除非有 $\psi_e = \chi_e = \varphi_e$。

因为 φ，φ_e，$\tilde{\varphi}_S$，$\tilde{\varphi}_R$ 是相等的，通过在 $\mathcal{A}_e$ 中对式(85)和式(86)进行论证并不会失去信息，特别地，式(88)意味着式(86)里在 *LHS* 中 $(\varphi_S \otimes \varphi_R)\circ\alpha_t$ 的点方法的极限存在，那么上述论证表明它必须与 φ 相一致。

进一步注意在[Emch and Radin，1971]中得到证明的结果事实上是一个更强的结果，这意味着式(86)是特殊的情况，进而式(85)也是特殊的情况，即对所有系统 S 的偶态 ψ_S 来说，有：

$$\forall A \in \mathcal{A}:\ \lim_{|t|\to\infty} \psi_S \otimes \tilde{\varphi}_R(\alpha_t[A]) = \varphi_\beta(A) \tag{89}$$

因此这加强了在方程式(86)之后提出的“剪切与粘贴”解释。

这一结果可以用两种方式作进一步的推广。第一，可以摒弃掉式(89)中 ψ_S 是一个偶态的限制，参见[Arkai and Barouch(荒木和巴洛奇)，1982]。第二，正如在[Emch and Radin，1971]中已经注意到的，特别地，当用 η 表示群 $\mathbb{R}$ 上的一个不变平均时，式(89)的遍历的或平均的版本：

$$\forall A \in \mathcal{A}:\ \eta\{\psi_S \otimes \tilde{\varphi}_R(\alpha[A])\} = \varphi_\beta(A) \tag{90}$$

出现于[Emch and Radin，1971]中，即使仅有弱版本的式(88)成立，满足 η 渐近阿贝尔性条件(60)，即：

$$\forall A,\ B,\ C \in \mathcal{A}_e:\ \eta\{\psi(C^*[A\alpha[B]-\alpha[B]A]C)\} = 0 \tag{91}$$

我们要么通过前面给出的普遍论证，要么利用刚刚回顾的特殊模型，从而提出是否能够从某些普遍的稳定性论证中得到 KMS 条件这个问题，这取决于研究者的知识状况。这一问题是从下面(B)—(E)多个角度提出的。

上面给出的模型在[Robinson(罗宾逊)，1973]中被再次讨论，也可见于[Arkai and Barouch，1982]及其参考文献中。它是长久以来一直在进行的研究，该研究起始于从第一原理推出牛顿冷却定律的尝试，相关问题通常在“回到平衡态”这个一般性主题下进行理解。如在文献中，对此主题的最新综合性研究可能是[Bach *et al.*（巴奇等)，2000]，在那里人们运用了大量的“新技术方法”，读者也将会在其中找到对于此主题相关的大量文献进行的翔实的抽样分析。在一个广泛的意义上，本节中的几个——但不是所有——稳定性标准同样

将关注点放在了从小的或局域性偏差回到平衡态这个被长期讨论的问题上。然而这个问题的普遍存在性不应该遮蔽其他两个重要的且在很大程度上未解决的问题，参见下文6.4节。

B. 消解局域扰动的稳定性

卡斯特勒等人研究了渐近阿贝尔性的不同条件。概述可见于[Kastler, 1976]，它同样也提供了带注解的参考文献。就他们的主要稳定性定理来说，其选择用下文中的定义41和42来对概念进行描述。

定义41。令$\mathcal{A}$为一个C^*代数。当$\forall(t, A)\in \mathbb{R}\times\mathcal{A}_o$：$\alpha_t[A]\in\mathcal{A}_o$且$\forall A, B\in\mathcal{A}_o$：$\int_{-\infty}^{+\infty}dt\ \| B\alpha_t[A]-\alpha_t[A]B\| <\infty$时，演化$\alpha$：$\mathbb{R}\to \mathrm{Aut}(\mathcal{A})$被认为是在一个标准紧致的*子代数$\mathcal{A}_o\subset\mathcal{A}$中$L^1$渐近阿贝尔。

讨论下面定义42前需要先了解一些准备符号。$\mathcal{A}$和α与它们在定义41中的意义一致，令$\mathcal{A}_{sa}=\{A\in\mathcal{A}\mid A=A^*\}$，且令$\mathcal{S}$为$\mathcal{A}$上所有态的集合，且其具有弱拓扑。对于$\varphi\in\mathcal{S}$和满足$\varphi(h^2)>0$的元素$h\in\mathcal{A}_{sa}$，定义：

(i)满足φ^h：$A\in\mathcal{A}\to\frac{1}{\varphi(h^2)}\varphi(hAh)\in\mathcal{A}$的微扰态$\varphi^h$；

(ii)满足$\alpha^h{}_t$：$A\in\mathcal{A}\mapsto U^h{}_t\alpha_t[A]\,U^h{}_t{}^*$的微扰演化$\{\alpha^h{}_t\mid t\in\mathbb{R}\}$，其中$\{U^h{}_t\mid t\in\mathbb{R}\}$满足所谓的“同环(co-cycle)微分方程”(导数被写为范拓扑)：

$$\forall t\in\mathbb{R}:\ i\frac{d}{dt}U^h{}_t=U^h{}_t\alpha_t[h]\text{，其中初始条件为 }U^h{}_o=I$$

为了理解在什么意义上α^h可能被看作是相应于h的微扰演化，注意上述同环方程允许存在唯一的连续解$t\in\mathbb{R}\mapsto U^h{}_t\in\mathcal{A}$，显然它可以用标准收敛戴森级数来计算：

$$U^h{}_t=\sum_{n=0}^{\infty}C^h{}_{t,n}\text{，其中 }C^h{}_{t,n}=(-i)^n\int_o^t dt_n\int_o^{t_n}dt_{n-1}\cdots\int_o^{t_2}dt_1\alpha_{t_1}[h]\cdots\alpha_{t_n}[h]$$

此解满足：(i)$U^h{}_t$是幺正的；(ii)$\forall s, t\in\mathbb{R}$：$U^h{}_{s+t}=U^h_s\alpha_s[U^h_t]$。

因此，定义为$\{\alpha^h{}_t\mid t\in\mathbb{R}\}$的演化是$\mathcal{A}$的自同构群，特别有：

$$\forall_{s,t}\in\mathbb{R}:\ \alpha^h{}_{s+t}=\alpha^h{}_s\circ\alpha^h{}_t$$

将α^h解释为通过算符h对α干扰导致的演化，它可由生成元α^h与α之间的下述关系得到：

$$i\frac{d}{dt}\alpha^h{}_t\Big|_{t=0}=i\frac{d}{dt}\alpha_t\Big|_{t=0}+\delta^h，其中\ \delta^h：A\in\mathcal{A}\mapsto[h,\ A]:=hA-Ah\in\mathcal{A}$$

定义 42。有了上述符号，在$\mathcal{A}$上的 α 不变态 φ 被认为是相对于内在干扰稳定的，当存在一个 φ 的近邻$\mathcal{V}_\varphi\subset\mathcal{S}$使得 $\forall A\in\mathcal{A}$和 $\forall h\in\mathcal{A}_{sa}$时，其中 $\varphi^h\in\mathcal{V}_\varphi$，满足下面三个条件时：

1. $\forall t\in\mathrm{IR}$：$\varphi^h(\alpha^h{}_t[A])=\varphi^h(A)$；

2. 当 $\lambda\in\mathrm{IR}$：$\lim_{\lambda\to0}\varphi^{\lambda h}(A)=\varphi(A)$；

3. $\lim_{t\to\infty}\varphi^h(\alpha_t[A])=\varphi(A)$。

定理 43。有了定义 41 中的$\mathcal{A}$，$\mathcal{A}_o$和 α，假定 α 在$\mathcal{A}_o$上是 L^1 渐近阿贝尔的。令 φ 是在$\mathcal{A}$上的 α 不变态，且假设 φ 在定义 42 的意义上相对于内在干扰是稳定的。那么，在下面将要讨论的三个辅助条件下，对于某些自然温度 β，φ 满足相对于 α 而言的 KMS 条件。

评论 44。定理的辅助条件概括为如下三条。

1. 态 φ 被假定为不是一个迹，即存在 A，$B\in\mathcal{A}$满足 $\varphi(AB)\neq\varphi(BA)$。这是为了避免在无限温度极限下可能出现的经典情况，即 $\beta=0$，即 $T=\infty$。

2. 在与 φ 正则相关的 GNS 表征中，执行 $\alpha(\mathrm{IR})$变换的幺正群 $U(\mathrm{IR})$的生成元被假定为不是单侧的。这是为了避免出现相反的情况，在这种情况下 φ 是零温度的基态，即 $\beta=0$，即 $T=0$。

3. 态 φ 被假定是在 * 子代数$\mathcal{A}_0$上的 4 阶超群集(hyperclustering)。这一技术性条件要求下式成立：对于每个正整数 $p\leqslant4$ 和所有 A_1，…，$A_p\in\mathcal{A}_o$，存在正的常数 C 和 δ 满足：

$$\forall t_1,\ \cdots,\ t_p\in R：\varphi_p^T(\alpha_{t_1}[A_1]\cdots\alpha_{t_p}[A_p])\leqslant C\{1+\max|\ t_k-t_l\ |^{1+\delta}\}^{-1}\quad(92)$$

其中截关联(truncated correlations)φ_p^T 由 $0=\varphi_o^T$、$\varphi(A)=\varphi_1^T(A)$和 $\varphi(A_1,\ \cdots,\ A_p)=\sum_P\varphi_{n_1}^T(A_{k_1},\ \cdots,\ A_{k_{n_1}})\cdots\varphi_{n_j}^T(A_{q_1},\ \cdots,\ A_{q_{n_j}})$来递归地定义，且其和超出了满足条件 $S=\cup_jS_j$，$j\neq k\Rightarrow S_j\cap S_k=\emptyset$ 的子集 $S_j\subseteq S$ 中的 $S=\{1,\ 2,\ \cdots,\ p\}$的所有保序分区，且在每个 $S_j=\{k_1,\ k_2,\ \cdots,\ k_{n_j}\}$ 中 $k_1<k_2<\cdots<k_{n_j}$。读者立即就可证明 $\varphi(A_1,\ A_2)=\varphi_2^T(A_1,\ A_2)+\varphi_1^T(A_1)\varphi_1^T(A_2)$，从而我们会认识到，与明确写出的对 P 的求和相比，递归关系更好地解释了高阶截关联究竟是什么。

注意 φ_p^T 提供了一个分层，在其中已经考虑到了所有低阶关联。特别地在 CCR 的情形中，罗宾逊[1965]的显著结果表明，或者这一分层无限上升，或者，如果对所有 $n \geqslant N$(其中 $N>2$)，截 φ_N^T 消失的话，则对所有 $N>2$ 它们必然消失。

截 φ_n^T 并不是一个新概念。它是经典概率理论中“累积量”和经典统计力学中“乌泽尔函数”的量子类似概念。罗宾逊定理的经典等价理论给出了高斯分布的一个特性，转换到量子统计中它还是自由谐振子系综的正则平衡态的另一个特性。因此罗宾逊定理预示了为什么建构和/或控制非“准自由”的 QFT 和 QSP 模型如此困难。

总之，定理的第三个辅助条件致力于表明，在时间过程中，所有 $p \leqslant 4$ 阶的时间关联对于长时分离而言衰变得足够快。

上文说到的卡斯特勒等人的研究似乎针对时间渐近阿贝尔性，即定义 41，以及时间超群集，即评论 44(3)的条件做出了系统性预言。因而我们将这些研究与在第 3 和第 4 节中遇到的任意条件进行比较。对于空间平移来说后者自然能得以满足，但在构造特殊模型时，即使将这些条件直接强加于微观动力学，即强加于描述时间演化的哈密顿量之上，也是极困难的。这究竟是说明上述定理(43)背后的理论存在内在缺陷呢，还是表明模型构建者缺少想象力或技术熟练程度低呢，这个问题在那个阶段仍然没有答案。不过，似乎当人们想要从 KMS 态中区分只相对于此条件是极值的那些态和相对于时间演化不变性是极值的那些态时，一种弱形式的渐近阿贝尔性既不是充分的，也不是必要的。参见如[Emch，1972a，Corollary 2，p. 206]，或下文评论 63(6)。这里再次回顾经典统计哈密顿力学中的永久遍历梦想，我们将会倾向于希望这样的区分或许是可能的。正如我在本文其他部分中提到的，我依然不能预见这个问题的答案(我的水晶球仍然是模糊的)。

C. 热容器稳定性

考虑系统 R 可以被构造为温度 β 下的“热容器”这个直观想法，如果它使得适当设置的测试系统 S 与 R 耦合时达到温度 β 的话。科萨科夫斯基等(Kossakowski)[1977]提出要以如下方式对该观点进行形式化，也可参见[Swell，2002，pp. 114—116]。具体的动机，请参见在本节 A 部分中描述的具体的 XY 模型。

为了模型化这样的一种情形，在这种情形下我们期望 R 应该比 S 大得多以便忽略从测试系统 S 到热容器 R 的反馈，我们假设 R 是无限的，而 S 是有限的。

假设的热容器 R 由一个三元数$\{\mathcal{A}^R, \alpha^R, \varphi^R\}$来描述，其中$\mathcal{A}^R$是一个 C^* 代数；α^R 是与$\mathcal{A}^R$自同构的演化群；φ^R 是$\mathcal{A}^R$上的态，在演化 α^R 下不变。用 $\delta^R :=$ $i\dfrac{d}{dt}\alpha^R t\mid_t=0$来标记演化 α^R 的生成元，对 R 的某些附加条件将在随后详述。

测试系统 S 是冯·诺伊曼意义上的动力学系统，即是用$\{\mathcal{A}^S, \alpha^S, \varphi_\beta^S\}$描述的，其中$\mathcal{A}^S=\mathcal{B}(\mathcal{H})$；$\alpha^S$ 是用哈密顿量 H^S 生成的演化，满足对于所有温度 $\beta>0$，$Z:=\mathrm{Tr}\exp(-\beta H^S)<\infty$；且 φ^S 由下式给出：

$$\varphi_\beta^S(A^S)=\mathrm{Tr}\,\rho_\beta^S A^S，其中\ \rho_\beta^S=Z^{-1}e^{-\beta H^S} \tag{93}$$

$\delta^S=i[H^S, \cdot\,]$表示 α^S 的生成元。最后，$\mathfrak{S}^s$ 表示在$\mathcal{A}^S$上所有可数可加态的集合。

在 R 与 S 之间的动力学耦合族$(\alpha^\lambda\mid\lambda\geqslant 0)$是用$\mathcal{A}=\mathcal{A}^R\otimes\mathcal{A}^S$上的自同构群描述的，其生成元具有下述形式：

$$\left.\begin{aligned}&\delta^\lambda=\delta^R\otimes I+I\otimes\delta^S+\lambda\delta_V，这里\\&\delta_V:\ A\in\mathcal{A}\mapsto i\lambda[V, A]\in\mathcal{A}，其中\ V\in\mathcal{A}_{sa}\end{aligned}\right\} \tag{94}$$

正如 R 的附加条件所示，从而它具有了 V 的形式，见下文(97)与(98)。

模型化的下一步是计划强调这样一种意义，在这个意义下，S 对演化 α^λ 的长时累积效应被解释为何时 R 与 S 是相耦合的。对此科萨科夫斯基等[1977]诉诸于所谓的范·霍夫极限，其案例已经在 3.5 节中提到过，也可见于下面的评论 45。对于这里的系统，范·霍夫极限取如下形式。首先，它仅考虑简化了的演化，即仅考虑总演化中系统 S 所经历的演化，数学上这一简化通过 E：$\mathcal{A}\to\mathcal{A}_S$达到，对于所有 $A_S\otimes A_R$，条件期望值由 $E[A_S\otimes A_R]=A_S\varphi_R(A_R)$确定，然后通过线性和连续性扩展到$\mathcal{A}$中。其次，范·霍夫极限要求关注长时/弱耦合区域，这个区域是通过按相互作用强度的负幂重新标度时间来定义的。因而，说明范·霍夫极限过程在这里即证明下述极限对所有正“重新标度的”时间 s 存在：

$$\gamma_s^S:\ A_S\in\mathcal{A}^S\mapsto\lim_{\substack{\lambda\to 0;t\to\infty\\ s=\lambda^2 t}}\alpha_{-t}^S\circ E\circ\alpha_t^\lambda[A_S]\in\mathcal{A}^S \tag{95}$$

评论 45。这类型的极限已有很长时间的历史。我第一次是从范·霍夫[van Hove, 1955]那里看到的，在那里作者将它看作一个将宏观传输现象与被期望可作为其基础的微观动力学相关联起来的工具。它强调了在这样的讨论中，时间应该被以一种由相互作用强度 λ 决定的方式来重新标度。为取此极限所做的一些证明将在下文 6.1 节中讨论。

最后，给定两个 C^* 代数$\mathcal{A}$和$\mathcal{B}$以及非负整数 n，我们说映射 γ：$\mathcal{A}\to\mathcal{B}$是 n 正定的，当它是线性的，且它引发的映射 γ_n：$\mathcal{A}\otimes\mathcal{M}(n,\mathbb{C})\to\mathcal{B}\otimes\mathcal{M}(n,\mathbb{C})$ 也是线性的时，即$\mathcal{A}\otimes\mathcal{M}(n,\mathbb{C})$中任意正元素的图像是$\mathcal{B}\otimes\mathcal{M}(n,\mathbb{C})$中的正元素时。如果$\mathcal{A}$或$\mathcal{B}$是阿贝尔性的，正映射必然是 n 正定的，因而 n 正定性对于 QSP 的非对易情况而言是一个新的概念。此外，如果一个映射对于所有 $n\in\mathbb{Z}^+$ 是 n 正定的，它被认为是完备正映射。为了和类似于上述式(95)右边的表达式联系起来，请注意完备正映射的组分也仍是完备为正的，且自同构、态、单射和条件期望值都是完备正映射。当 γ_o 是恒等映射，且 $\forall(s,t)\in\mathbb{R}^+\times\mathbb{R}^+$：$\gamma_{s+t}=\gamma_s\circ\gamma_t$ 时，$\mathcal{A}$到其自身的映射集$\{\gamma_s\mid s\in\mathbb{R}^+\}$被认为构成了一个半群。

这应该完全包括了描述科萨科夫斯基等[1977]提出的稳定性标准所必要的一般性预备知识，即：

定义 46。系统$\{\mathcal{A}^R,\sigma^R,\varphi^R\}$被认为是温度 β 下的热容器，当存在一个关于测试系统$\{\mathcal{A}^S,\sigma^S,\varphi^S_\beta\}$和动力学耦合$\{\alpha^\lambda\}$的“足够大”集合满足下面条件时：

1. 范霍夫极限(95)存在，且定义了$\mathcal{A}_S$的完全正变换半群$\{\gamma^S_s\mid s\in\mathbb{R}^+\}$；
2. $\mathcal{A}_S$上的正则冯·诺伊曼平衡态 φ^S_β 是 α^S 和 γ^S 下不变唯一态 $\varphi\in\mathfrak{S}^S$；
3. $\forall(\psi^S,A^S)\in\mathfrak{S}^S\otimes\mathcal{A}^S$：$\lim\limits_{s\to\infty}\varphi^S(\gamma^S_s[A_S])=\varphi^S_\beta(A_S)$。 (96)

上述定义中的术语“足够大”确实需要进一步地精确：关于相互作用 V 和热容器 R 的辅助条件就是在这个过程中被添加进来的，并允许我们利用它们来证明下面的批注 47 和定理 48。

其中一个条件是式(94)中的相互作用 V 具有如下形式：

$$V=\sum_{k=1}^{n}B^R_k\otimes B^S_k,\ \text{其中}\left\{\begin{array}{l}n\text{ 是确定的}\\ B^R_k\in\mathcal{A}^R_{sa}\text{和 }\varphi^R(B^R_k)=0\\ B^S_k\in\mathcal{A}^S_{sa}\end{array}\right\}\qquad(97)$$

注意 V 的条件期望 $E[V]$ 不存在。

一个附加条件是，存在$\mathcal{A}_o^R \subseteq \mathcal{A}^R$，满足：(i)展开式$\{\mathcal{A}_o^R \cup I\}$(其中 I 是在$\mathcal{A}^R$中的单位元)在$\mathcal{A}^R$中是标准稠密的；(ii)对所有的 $B_k^R \in A_o^R$，函数 $t \mapsto \varphi^R(B_j^R \alpha_t^R[B_K^R])$在 L^1 中；(iii)对于要测试的态 φ^R，多重时间截断关联满足：

$$t_1 < \cdots < t_l,\ 有\ |t_j - t_k| \to \infty \Rightarrow \{\varphi^R\}^T(\alpha_{t_1}^R[B_{t_1}^R]\cdots\alpha_{t_n}^R[B_{t_n}^R]) \to 0 \qquad (98)$$

由文献[Davies, 1974, theorem 2.3]和[Kossakowski *et al.*, 1977]得到如下结果。

批注 47。这些辅助条件充分表明对所有的 S，定义 46 中的条件(1)都能被满足。

这保证了测试系统的集$\mathfrak{T}_\beta$确实会足够大。

定理 48。当上面概述的情况成为现实时，下述条件是等价的：

1. 对某些温度 β，态 φ^R 是$\mathcal{A}_R$上关于演化 α_R 的一个 KMS 态；

2. 在定义 46 中，且在批注 47 中将“足够大”进行精确化说明的意义下，态为 φ^R 的系统 R 是温度 β 下的热容器。

评论 49。

1. 因此，当 φ^R 满足温度 β 下的 KMS 条件时，$\mathfrak{T}_\beta$ 中的每一个测试系统 S，都会由于热容器 R 的作用而趋向达到此温度下的平衡态。

2. 正如之前提到的，这一结果很大程度上是与模型无关的，且并不涉及(至少不会明确地涉及)任何关于时间渐近阿贝尔性的假设。此外，与一个单独模型不同，利用此模型可以证明处于平衡态的特殊的无限系统充当了每个有限部分的热容器，当前的定理描述了测试系统 S 的集合$\mathfrak{T}_\beta$，无限系统 R 充当了其热容器的角色。因而，此定理是针对上述 **A** 部分中讨论过的特殊激励模型的改进版本。

3. 还有，如[Kossakowski *et al.*, 1977]注意到的，他们提出的有效多重时间关联的衰减容易让人想起由卡斯特勒等人在克服局域干扰稳定性的结果中提出的类似条件，见上文 **B** 部分中的式(92)。

4. 从实用的角度看，该定理可以看作是对将平衡态 QSP 中的温度概念从有限系统上升到无限系统的这个过程的明确说明。

5. 然而，必须注意的是，在某些情况下我们已知 γ^S 满足定义 46 中的条件

(1)以及(98)中的关联衰退，这些情况似乎确实涉及了某些聚集特性，这些特性可能会对此定理的应用范围做出限制，即使其只能应用于 $\pi_\varphi(\mathcal{A}_R)''$ 为一个参数，且 φ^R 在 KMS 条件与时间演化不变性条件下都是极值的这样的案例中。因此，渐近阿贝尔性在这里就出现得不那么明显了。

D. 钝性

在[Pusz and Woronowicz(普兹和瓦伦纳维奇)，1978]中，作者注意到了 KMS 态的一个特性，他们称为钝性，且在不涉及时间渐近阿贝尔性的假设下，他们找到了一些方式来表明此特性反过来必然导致 KMS 特征。

具体地，令 $\{\mathcal{A}, \varphi, \alpha\}$ 为一个动力学系统，其中 $\mathcal{A}$ 是 C^* 代数，φ 是 $\mathcal{A}$ 上的态，且 $\{\alpha_t \mid t \in R\}$ 是与 $\mathcal{A}$ 自同构的单参数群。那么令 $D(\delta)$ 表示演化 α 的生成元 δ 的区间，即 $D(\delta)$ 是所有 $A \in \mathcal{A}$ 的线性子空间，以使得导数 $\delta[A] := i\frac{d}{dt}\alpha_t[A]$ 存在。

现在考虑此系统在一有限时间段内与一外部系统相互作用的这种情况，因此它们与我们所感兴趣的系统相互作用的结果可以被描述为满足下面的微分方程的微扰动力学 α^h：

$$\forall A \in D(\delta): \begin{cases} i\dfrac{d}{dt}\alpha^h{}_t[A] = \alpha^h{}_t[\delta[A] + [h_t, A]] \\ \alpha^h{}_{t=0}[A] = A \end{cases} \tag{99}$$

其中 h 是所有连续可微函数集 $C^1_+(\mathrm{IR}, \mathcal{A}_{sa})$ 的一个元，它在 IR^+ 中有紧致支集，且在 $\mathcal{A}$ 的自伴部分中取值。因此在所有存在 h 的时刻 t，即在微扰 h 真正起作用的所有时刻，该系统都是开放系统。h 的支集是紧致的且包含在 IR^+ 中，这一条件确定了，对所有时刻 $T > \sup\{t \in \mathrm{IR} \mid h_t \neq 0\}$，外部条件与时刻 $t = 0$ 时相同。要求关于外部干扰的时间依赖性的平滑条件 $h \in C^1$ 成立是为了数学上的方便，这在物理上也是合理的。因此

$$L^h_T(\varphi) := \int_0^T dt\ \varphi\left(\alpha^h{}_t\left[\frac{d}{dt}h_t\right]\right) \tag{100}$$

描述了在时间区间 $[0, T]$ 中传输到系统上的能量，在该时间段内系统处于外部干扰 h 的影响之下。

定义 50。如果对所有 $h \in C^1_+(R, \mathcal{A}_{sa})$ 和所有 $T > \sup\{t \in \mathrm{IR} \mid h_t \neq 0\}$：$L^h_T(\varphi) \geqslant 0$，态 φ 就被认为是钝性的。

在提出此定义之后，普兹和瓦伦纳维奇[1978]证明了如下结果。

定理 51。令$\{\mathcal{A}, \varphi, \alpha\}$为一个 C^* 动力学系统，且考虑如下条件：(Ⅰ)φ 要么是一个在温度 $\beta > 0$ 下关于 α 的 KMS 态，要么是一个基态；(Ⅱ)φ 在定义 50 的意义上是钝性的。那么：

1. 不需要进一步的假设：(Ⅰ)⇒(Ⅱ)。

2. 若进一步假设：(i)$\mathcal{A}$允许操作 ν：$G \to \mathrm{Aut}(\mathcal{A})$，其中 G 是一个局域紧致的可控群；(ii)$\mathcal{V}$与演化 α 对易，即 $\forall(t, g) \in R \times G$：$\nu_g \circ \alpha_t = \alpha_t \circ \nu_g$；(iii)相对于操作 G 来说 φ 是 η 聚集的。那么将这些条件放在一起就能得到(Ⅱ)⇒(Ⅰ)。

评论 52。下述的评论关注于定理的第(2)部分，即被认为是具有钝性的 KMS 态的操作特征。

1. 在钝性条件(Ⅱ)中，φ 没有被假设成在未被干扰的演化 α 下是不变的，在第(2)部分中此特性被认为等价于 φ 被证明满足 KMS 条件。

2. 从下面要得到的这个结论来看的话，ν 与演化 α 对易的条件是自然的：若自同构性对一个 KMS 态保持不变的话，则它必须与此态为 KMS 态时所涉及的演化相对易。

3. 在 3.5 节中已经介绍过不变性方法与可控群，见定义 21 和评论 31(5)中。

4. 在(2)中列出的辅助性假设中，甚至 φ 是 G 不变的这个前提条件也是不必要的，这可由它是 η 聚集的这个明确的假设得到，即(见定义 22)：

$$\forall A, B \in \mathcal{A}: \eta^G(\varphi(\nu_g[A]B)) = \varphi(A)\varphi(B)$$

事实上，这一条件进一步要求 φ 不能够分解为其他 G 不变态的凸组合。

5. 在 QSP 中，G 的自然候选对象是空间平移群。因此，与之前研究过的稳定性条件相比较，假设的聚集特性并不需要是时间相关的。这允许我们考虑演化不具有渐近阿贝尔性的系统。这一开端是意义重大的，当它针对 QSP 的目的得到了具体模型时：人们或许不希望非得将关于群$\{\nu_g \mid g \in G\}$的弱聚集与关于演化$\{\alpha_t \mid t \in \mathrm{IR}\}$的任意假定的聚集等同起来。

6. 此外，普兹和瓦伦纳维奇[1978]提出了一个备选方案，他们将该定理第(2)部分中的所有辅助条件替换为钝性的强形式。特别地，与考虑一个单动力学系统不同，对于每个正整数来说，他们考虑的是，相同的无相互作用的系统的众多拷贝$\{\mathcal{A}_k,\ \varphi_k,\ \alpha_k \mid k=1,\ \cdots,\ N\}$，由此人们构造了聚集的动力学系统$\{\mathcal{A}^N,\ \varphi^N,\ \alpha^N\}$，其中$\{\mathcal{A}^N=\otimes_{k=1}^N\mathcal{A}_k,\ \varphi^N=\otimes_{k=1}^N\varphi_k$，且$\alpha^N=\otimes_{k=1}^N\alpha_k\}$。然而微扰$h$被认为是$C_+^1(\mathbb{R},\ \mathcal{A}^N)$中的普遍性元素，因此$\alpha^h$被认为不能独立地作用于组合系统中的每一个系统。那么当对于每个正整数N态φ^N是钝性的时，φ被说成是完全钝性的。现在，不需要额外再做什么——即不需要在定理51中强加条件(2)——就可以证明φ的完全钝性等价于φ满足KMS条件这个条件。对于QSP，如何在完备钝性条件和定理中的条件(2)之间做出选择，在很大程度上只是个人喜好问题。

E. 热力学稳定性

作为本节的结尾，我希望指出热力学稳定性概念是如何导致KMS态的另一个特征的，这一特征并不限制涉及KMS条件时所考虑的态是否是极值的。为了回避技术性细节，我用最简单的例子来呈现这些思考，这个例子就是系统为一个量子自旋点阵的情况，并且因此系统是由C^*代数$\mathcal{A}=\otimes_{k\in\mathbb{Z}^d}\mathcal{A}_k$来描述的，其中$\mathcal{A}_k$是有限矩阵代数的拷贝，比如说$\mathcal{M}(n,\ C)$，其中$n$和$d$有限。全部$\Lambda\subset\mathbb{Z}^d$表示了点阵$\mathbb{Z}^d$的有限子集的一个联系，$\varphi$表示在$\mathcal{A}$上的一个态，$\varphi_\Lambda$表示对$\varphi$所做的约束，即将其限制为有限矩阵代数$\mathcal{A}_\Lambda=\otimes_{k\in\Lambda}\mathcal{A}_k$；$\rho_\Lambda$是对应于$\varphi_\Lambda$的密度矩阵。此外这里方便地假设动力学适用于位点间的短程——或可能是适当温和的相互作用。对下面陈述的内容进展到了什么程度感兴趣的读者能在[Sewell，2002]中找到相关的回顾，在最初的论文中，人们只涉及了原定的目标[Araki，1974；Araki and Sewell，1977；Sewell，1977；Sewell，1980b；Ruelle，1968a；Robinson，1971；Araki and Moriya(荒木和莫里亚)，2002]。

热力学第二定律的一个版本——比较于定理3之后定义的变分原理的等价形式——定义了自然温度$\beta=1/kT$下相对Λ的局域自由能：

$$\text{当}\ \begin{cases}E_\Lambda(\varphi)=\varphi_\Lambda(H_\Lambda)\\ S_\Lambda(\varphi)=-k\mathrm{Tr}\,\rho_\Lambda\log\rho_\Lambda\end{cases}\ \text{时，}F_{\Lambda,\beta}(\varphi)=E_\Lambda(\varphi)-T\,S_\Lambda(\varphi)$$

当$\mathcal{A}$上的两个态ψ和φ在有限区域Λ_o外相一致时，它们被认为是满足等价关系$\overset{\Lambda_o}{\sim}$的。因此当存在$\Lambda_o$使得$\psi\overset{\Lambda_o}{\sim}\varphi$时，我们将其记作$\psi\sim\varphi$。对于这里考虑的量子点阵，人们则能够证明下述极限存在：

$$\forall\psi\sim\varphi:\ \Delta F_{\beta}(\psi\mid\varphi):=\lim_{\Lambda\uparrow\mathbf{Z}^{\nu}}(F_{\Lambda,\beta}(\psi)-F_{\Lambda,\beta}(\varphi))\tag{101}$$

对于在ΔF_{β}中自变量ψ和φ的顺序，回想一下数学家(和某些哲学家)们是从右往左读的，而绝大多数的物理学家似乎是从左往右读的。因而，正如上面所写的，$\Delta F_{\beta}(\psi\mid\varphi)$表示仅在有限区域内当从态$\varphi$转向不同于$\varphi$的态$\psi$时自由能的增加值。荒木和休厄尔[1977]引入如下定义证明了如下结果，也可参见[Sewell，1980b；Sewell，2002]。

定义53。如在式(101)中有$\Delta F_{\beta}(\varphi\mid\psi)$，$\mathcal{A}$上的态$\varphi$被认为在自然温度$\beta$下是局域热力学稳定的，如果：

$$\forall\psi\sim\varphi:\ \Delta F_{\beta}(\psi\mid\varphi)\geqslant 0$$

因此，要求此稳定性条件必须被满足其实是一个变分原理：态φ的自由能不能仅由在一个局域区域上得到一个不同于φ的态ψ而约化。

定理54。对于这里考虑的这类量子点阵系统上的一个态φ来说，如下条件是等价的：

1. φ在自然温度β下满足KMS条件；

2. φ在自然温度β下是局域热力学稳定的。

评论55。

1. 此结果以一种本质的方式涉及了所考虑系统的局域结构，即全局性代数$\mathcal{A}$是相对于空间有界区域的局域代数$\mathcal{A}_{\Lambda}$的C^*归纳极限，其中指标系统$\mathcal{F}:=\{\Lambda\}$是吸收的，即回顾3.4节，第Ⅰ部分——对于空间中的每个点x，都存在一有界区域$\Lambda\in\mathcal{F}$满足$x\in\Lambda$。另一个版本要求对于每个有界空间区域Ω，存在某个$\Lambda\in\mathcal{F}$满足$\Omega\subseteq\Lambda$。两种版本在公理化QSP中都是可接受的。

2. 同样相比于热容器稳定性——见定理(48)——这一结果也是具有内在一致性的，在它将从下述两种不同角度对同一个系统平衡态所做的定义等价起来的意义上：微观KMS条件和所考虑系统的热力学的局域特性。特别地，该论证并不涉及所考虑系统与任意测量系统间的耦合。

3. 对定理有效性范围的拓展是合理的。在此方面，具有适度长程相互作用的量子自旋点阵系统受到了约束。然而，当人们期望向连续系统拓展时总会遇到某些技术困难。往往这些困难源自于与有限区域相对应的希尔伯特空间$\mathcal{H}_\Lambda$的无限维度，且事实上相应的哈密顿量H_Λ是无界的，并且，还不得不采取预防措施以保证局域粒子密度仍然是有界的。

4. 对上述变分原理的一类拓展是有益的，即从局域稳定性拓展到全局稳定性条件。具体地，再次考虑定义在$\mathbb{Z}^d$上的一个量子点阵系统。进一步假设动力学在平移群$G=\mathbb{Z}^d$下不变，且将关注点限制于态ψ的集合$\mathfrak{S}^G$，其中每一个态都是G不变的。最后假设如下极限存在。

$$f_\beta(\psi)=\lim_{\Lambda\uparrow\mathbb{Z}^d}|\Lambda|^{-1}F_{\Lambda,\beta}(\psi);\ \Phi_\beta=\lim_{\Lambda\uparrow\mathbb{Z}^d}|\Lambda|^{-1}\log\operatorname{Tr}\exp^{-\beta H_\Lambda}\qquad(102)$$

态$\varphi\in\mathfrak{S}^G$现在被认为是全局热力学稳定的——或简写为GTS——当它使自由能密度最小化时，即当：

$$f_\beta(\varphi)=\min_{\varphi\in S_G}f_\beta(\psi)=\phi_\beta\qquad(103)$$

只要我们保证量子点阵上的G不变态具有仅存在短程相互作用的G不变性动力学，那么则有：

$$\varphi\text{是GTS}\Leftrightarrow\varphi\text{是KMS}$$

然而，即使当相互作用被允许拓展为适当长程时⇔仍然有效，但“短程”要求对于⇐仍是必不可少的。有学者[Sewell，1980b]认为KMS态不是GTS的，即并不会使自由能密度最小化，可能亚稳态是其模型。

5.5 QFT的简短回顾

回顾KMS态在数学物理学中的角色，我认为应(虽然简短地)提及超越非相对论QSP限制的模数结构，即在相对论QFT中涉及它们。关于代数QFT的一般框架，参见[Halvorson，2006]；关于在弯曲时空中专门对QFT的描述，也可参见[Wald(瓦尔德)，1994]；关于针对本节中材料引起的某些解释性问题的讨论，参见[Clifton and Halvorson(克里夫顿和霍尔沃森)，2001]。

从本节发展出的视角看，本小节要讨论的问题可通过在闵可夫斯基空间QFT中证明富田—竹崎二元性来自然地引入——回顾批注33或定理39。

比索尼娅诺和威赫曼(Bisognano and Wichmann)[1975]给出了公理化 QFT 中标准结果，即瑞赫—施列德(Reeh-Schlieder)定理，参见[Streater and Wightman，1964，168]或[Emch，1972a，290]以及其中引用的文献的意义，特别是它保证了当真空态 φ 被限制为楔 $W_R=\{(x, y, z, t)\in M^{3+1}\mid z>\mid t\mid\}$ 时，在相应的代数 $\mathcal{N}_R$ 上是单射的。因而，φ 对 $\mathcal{N}_R$ 的这个限制 φ_R 使后者拥有了富田—竹崎模代数结构。这里，富田—竹崎理论的正则对象拥有一个开创性的几何解释。对合反幺正算符 J——对应于将楔 W_R 投影到楔 $W_L=\{(c, y, z, t)\in M^{3+1}\mid z<\mid t\mid\}$ 的映射 $(x, y, z, t)\to(x, y, -z, -t)$——执行一个从 $\mathcal{N}_R$ 向 $\mathcal{N}'_R\simeq\mathcal{N}_L$ 的双射；且模群 $\{\Delta^{i\lambda}\mid\lambda\in R\}$ 实现 $\mathcal{N}_R$ 上的洛伦兹上升：

$$\begin{pmatrix}x\\y\\z\\t\end{pmatrix}\to\begin{pmatrix}1&0&0&0\\0&1&0&0\\0&0&\cosh(2\pi\lambda)&-\sinh(2\pi\lambda)\\0&0&-\sinh(2\pi\lambda)&\cosh(2\pi\lambda)\end{pmatrix}\begin{pmatrix}x\\y\\z\\t\end{pmatrix}$$

因为在楔 W_R 内部做均匀加速运动的观测者感受到了作为过去和未来视界的边界，比索尼娅诺和威赫曼的结果可以被解释为它表明了在这类观测者的宇宙(楔 W_R)中，在下述意义上 M^{3+1} 中的真空场表现为一个热的浴场。态 φ_R: $N\in\mathcal{N}_R\mapsto\varphi(N)\in\mathbb{C}$ ——其中 $\mathcal{N}_R\subset\mathcal{N}$ 是对应于楔 W_R 的代数，$\mathcal{N}$ 是相应于完全闵可夫斯基空间的代数，且 φ 是定义在 $\mathcal{N}$ 上的真空——是在温度 $\beta>0$ 下，涉及演化 $\{\tau_t: N\in\mathcal{N}_R\mapsto\tau_t[N]=\Delta^{-it/\beta}N\Delta^{it/\beta}\in\mathcal{N}_R\mid t\in\mathbb{R}\}$ 的一个 KMS 态(这里，像往常一样，自然温度 $\beta=1/kT$ 的数值依赖于测量时间 t 时使用的标度)。

此解释的物理意义在瑞德勒(Rindler)早期的一篇综述性论文[Rindler，1966]中得以加强，并达到了这样的效果，即在 W_R 中做均匀加速运动的观测者的宇宙类似于遍及爱因斯坦方程的史瓦西(Schwarzschild)解的宇宙，即一个静态“黑洞”周围的宇宙。

有了这一参考文献，比索尼娅诺和威赫曼发现的现象作为富田—竹崎理论的结果转变成了盎鲁(Unruh)[1976]的一个新发现，他是在试图说明后来发现的霍金效应(也被人们熟知为霍金辐射)[霍金(Hawking)，1975]时，独立发现这一效应的。后者描述了一个相关但不同的现象，即围绕一个塌缩黑洞的热分布粒子的创生现象。盎鲁和霍金效应间的相似性与不同之处在[Wald，1994， Chapters. 5，7]中被重点讨论；对于天体物理学文献中该论题的某些热力学方面，参见如[Davies，1978；Hawking and Page(霍金和派兹)，1983]或[Wald，1994，Chapter 6]；对于什么是真实可测量的、如何以及在哪里测量这些特殊问题，请参见[Unruh and Wald，1984]；一些相关的哲学问题，请参见[Clifton and Halvorson，2001]。

比索尼娅诺和威赫曼关于公理化 QFT 文献的发现最早受到的冲击来自于休厄尔[1980a；1982a]的工作，休厄尔将他们的结果推广到某些弯曲的流形中，他察觉到了分叉视界在盎鲁效应中的作用，并提出要将富田—竹崎模理论与霍金温度和 KMS 条件下的温度联系起来。后来将 KMS 结构引入到 QFT 中被证明是一次"革命"的先兆[Borchers，2000]，继而导致了许多理论的发展，其中一些是：公理化代数 QFT 在弯曲流形的推广；用广义相对论时空中相交楔型区域上的附属于吸收集的模结构间的一致性关系这一术语来解释时空的内在几何特征；相对论 QSP 的开始，在那里局域 KMS 条件是由确定局域静止参考系的未来方向型类时矢量这一术语来构造的，参见如[Summers and Verch(萨默斯和维尔驰)，1996；Buchholz *et al.*(巴克霍尔兹等)，2002；Ojima(雄岛)，2003；Wiesbrock，1997；Buchholz，2003；Summers and White，2003；Buchholz and Lechner(巴克霍尔兹和莱希纳)，2004]；更加接近于真正意义上的霍金效应可见于[Haag *et al.*，1994；Kay and Wald(凯和瓦尔德)，1991；Fredenhagen and Haag(弗雷登阿根和哈格)，1990]；新的框架可见于[Fredenhagen，2003]。

5.6 数学插曲：极值 KMS 态

在 5.7 节中主要研究极值 KMS 态在 QSP 中的作用。本节的目的是回顾一些数学方面的基础内容，如极值 KMS 态的定义，用 GNS 表示来描述的它们的特征，以及将一个 KMS 分解为其极值组分的分解方法。

要点 56。

1. 令$\mathcal{A}$为一个 C^*代数，$\beta>0$，以及 τ 为$\mathcal{A}$的自同构群。对于 τ 和 β 来说，在$\mathcal{A}$上满足 KMS 条件的所有 KMS 态的集合$\mathfrak{S}_\beta$是凸的，即对于任意两个$\mathcal{A}$上的 KMS 态 ψ 和χ，对于相同的 τ 和β，以及任意的 $\lambda\in(0,1)$：$\varphi=\lambda\psi+(1-\lambda)\chi$ 仍是关于 τ 和 β 的一个 KMS 态。

2. 集合$\mathfrak{S}_\beta$在它由$\mathcal{A}$得到的w^*拓扑中是闭合的，且它在度规拓扑中是有界的。因而它是w^*紧致的，且克莱因—米尔曼(Krein-Milman)定理要求$\mathfrak{S}_\beta$是其极值点集合$\mathfrak{S}_\beta$的w^*闭合的凸包[Dunford and Schwartz(邓福德和施瓦茨)，1964，Theorem V. 8.4]。这不仅保证了极值点的存在，同样也保证了存在大量这样的点：$\mathfrak{S}_\beta$中的每一个元素是$\mathfrak{S}_\beta$中有限凸元素和的极值，见下文定义57。

3. 此外$\beta_1 \neq \beta_2$要求$\mathfrak{S}_{\beta 1} \cap \mathfrak{S}_{\beta 2} = \emptyset$。另外，当$\beta_1 \neq \beta_2$时由态$\varphi_1 \in \mathfrak{S}_{\beta 1}$和$\varphi_2 \in \mathfrak{S}_{\beta_2}$构造的GNS表征是不相交的，如果从这些态中没有一个态的子表征是幺正等价于其他态的任意子表征的意义上说，参见[Takesaki，1970c]。

定义57。给定一个冯·诺伊曼代数$\mathcal{N}$，与$\mathcal{N}$自同构的群$\{\tau_t \mid t \in \mathbb{R}\}$，$\beta \in \mathbb{R}^+$以及上文要点56(1)中的$\mathfrak{S}_\beta$。态$\varphi \in \mathfrak{S}_\beta$在自然温度$\beta$下被称为是极值KMS，如果它不能分解为$\mathfrak{S}_\beta$中的态——即对于相同的$\tau$和$\beta$，满足KMS条件的态——的话。所有极值KMS态的集合被表示为$\mathfrak{E}_\beta$。

定理58。令φ为冯·诺伊曼代数$\mathcal{N}$上的一个单射正常态，τ是$\mathcal{N}$的自同构唯一群，涉及它时，φ在某些自然温度β下满足KMS条件。用$\mathcal{Z}$表示$\mathcal{N}$的中心$\mathcal{N} \cap \mathcal{N}'$，则：

A. 对于每个$(t, Z) \in \mathbb{R} \times \mathcal{Z}$，$\tau_t[Z] = Z$；

B. 对每个正非零元素$Z \in \mathcal{Z}$，其中$0 < Z < I$，

$$\psi(N) := \varphi(Z)^{-1}\varphi(ZN) \text{和} \chi(N): = \varphi(I-Z)^{-1}\varphi((I-Z)N)$$

定义了$\mathcal{N}$上的两个态φ和χ，在相同的$\mathcal{T}$和β下它们满足KMS条件，并提供了一个φ的凸分解。

C. 对于$\mathcal{N}$上将$\varphi = \lambda\psi + (1-\lambda)\chi$凸分解为态$\psi$和$\mathcal{X}$(它们对相同的$\tau$和$\beta$来说，满足KMS条件)的每个$\varphi$而言，存在一个唯一的正的非零元素$Z \in \mathcal{Z}$其中$\| Z \| \leqslant 1$，使得对于所有的$N \in \mathcal{N}$有：

$$\varphi(N) = \varphi(Z)^{-1}\varphi(ZN) \text{和} \chi(N) = \varphi(I-Z)^{-1}\varphi((I-Z)N)$$

证明：正如在评论37中指出的，我们可以不失一般性地假设$\mathcal{N}$是用标准形式表示的，从而存在一个关于$\mathcal{N}$的循环矢量和分离矢量$\Phi \in \mathcal{H}$，其中$\forall N \in \mathcal{N}(\Phi, N\Phi) = \varphi(N)$。

[A.]$Z \in \mathcal{Z} \Rightarrow \forall (t, N) \in \mathbb{R} \times \mathcal{N}$，$\varphi(N^* \tau_t[Z]) = \varphi(\tau_t[z]N^*)$，因此$\varphi$为

KMS 态就要求 $\varphi(N\tau_t[Z])$ 随 t 是不变的，从而 $\forall t\in\mathbb{R}$：$(N\Phi, [\tau_t[Z]-Z]\Phi)=0$。$\Phi$ 为循环矢量要求 $[\tau_t[Z]-Z]\Phi=0$，Φ 为分离矢量要求 $[\tau_t[Z]-Z]=0$。

[B.] φ 是单射的且 $0<Z<I$ 为正的且非零的，要求 $0<\varphi(Z)<1$，考虑到 Z 属于 $\mathcal{N}'$ 且因此 Z^{+} 属于 $\mathcal{N}'$，人们可证实 ψ 和 χ 是 $\mathcal{N}$ 上的态，且它们从 φ 中得到了其 KMS 特征。此外，人们可直接从其定义中得到 $\varphi=\lambda\psi(N)+(1-\lambda)\chi(N)$，其中 $0<\lambda=\varphi(Z)<1$。

[C.] 相反，根据 $\varphi=\lambda\psi(N)+(1-\lambda)\chi(N)$，其中 $0<\lambda<1$，我们可知 $\psi\leqslant\lambda^{-1}\varphi$，且因此存在一个元素 $X\in\mathcal{N}'$ 使得 $\forall N\in\mathcal{N}$：$\psi(N)=(X\Phi, NX\Phi)$，即 ψ 是 $\mathcal{N}$ 上的一个矢量态，因而是正交的且可通过正交函数 $\lambda^{-1}\varphi$ 来优化。所以酒井—拉东—尼科迪姆(Sakai-Radon-Nikodym)[Sakai，1971，proposition 1.24.4] 要求存在一个正的 $Y\in\mathcal{N}$，其中 $\|Y\|\leqslant 1$，使得：

$$\forall N\in\mathcal{N}:\ \psi(N)=\frac{1}{2}\lambda^{-1}\varphi(NY+YN)$$

假定存在另一个具有相同特征的元素 $\tilde{Y}\in\mathcal{N}$。那么令 $X=Y-\tilde{Y}$。则我们有 $0=\varphi(X^{*}X+XX^{*})$，且由于 φ 是正的线性函数，并且 $X^{*}X$ 和 XX^{*} 都是正的，因此 $\varphi(X^{*}X)=0$。因为 φ 是单射的，所以 $X=0$ 即 $Y=\tilde{Y}$，即 Y 是唯一的。

仍有待证明的是定理的假设要求 Y 也属于 $\mathcal{N}'$。因为 φ 和 ψ 是 KMS 态，它们满足对所有的 $t\in\mathbb{R}$：φ，且 $\psi\circ\tau_t=\psi$。所以：

$$\psi(N)=\psi(\tau_t[N])=\frac{1}{2}\lambda^{-1}\phi(\tau_t[N]Y+Y_{\tau_t}[N])=\frac{1}{2}\lambda^{-1}\phi(N\tau_{-t}[Y]+\tau_{-t}[Y]N)$$

根据刚刚证明的 $Y\in\mathcal{N}$ 的唯一性，我们有 $\forall t\in\mathbb{R}$：$\tau_t[Y]=Y$。因此 φ 为 KMS 态要求 $\forall N\in\mathcal{N}$：$\varphi(NY)=\varphi(YN)$，因而有 $\psi(N)=\lambda^{-1}\varphi(YN)$。将 KMS 条件同时应用于 ψ 和 φ，我们有 $\forall N\in\mathcal{N}$：$NY=YN$，即 $Y\in\mathcal{N}'$。从而显然有 $\lambda=\varphi(Z)$。相同的论证可通过用 χ 代替 ψ 和用 $(I-Z)$ 代替 Z 来完成。

下面的特性是上述定理的直接结果。

推论 59。有了定理 58 的假设，当且仅当 $\mathcal{N}$ 为一个因数，即当且仅当 $\mathcal{N}$ 具有平凡中心(trivial center)：$\mathcal{N}\cap\mathcal{N}'=\mathbb{C}I$，KMS 态 φ 是极值 KMS。

批注 60。在定理 58 的假设下，假定 φ 不是极值 KMS，但 $\mathcal{N}$ 的中心 $\mathcal{Z}$ 是由一族相互正交的投影算子 $\{P_k\in\mathcal{Z}\mid k=1, 2, \cdots\}$ 生成的。则存在一个将 φ 分解

为$\mathcal{N}$上的态φ_k的凸组合$\sum_k \lambda_k \varphi_k$的唯一分解，这里对于相同的动力学$\tau$和相同的自然温度$\beta$，$\varphi_k$是极值KMS。

论证。说φ是一个非极值KMS的KMS态，就是说存在KMS态ψ_j和标量$\mu_j \in (0, 1)$使得$\varphi = \sum_j \mu_j \psi_j$。根据定理的$C$部分，对于每个$\psi_j$都存在一个正的$Z_j \in \mathcal{Z}$使得$\forall N \in \mathcal{N}$: $\psi_j = \phi(Z_j)^{-1} \phi(Z_j N)$。因为$\mathcal{Z}$是一个具有离散谱的阿贝尔性冯·诺伊曼代数，每个$Z_j$可以写作$\sum_k z_k P_k$，其中$z_k \in \mathbb{R}^+$且$P_k$为$\mathcal{Z}$中的最小投影算子。因此$\varphi_k$: $N \in \mathcal{N} \mapsto \lambda_k^{-1} \varphi(P_k N) \in \mathbb{C}$，其中$\lambda_k = \phi(P_k)$是$\mathcal{N}$上的态。根据定理的$B$部分，对于相同的$\tau$和$\beta$来说，这些仍然是KMS态。因此，仍待证明的仅仅是，涉及KMS条件时，态φ_k是极值的。

为了表明这一点，我们考虑分解$\mathcal{H} = \oplus_k \mathcal{H}_k$，其中$\mathcal{H}_k$为子空间$\{\Psi_H \in \mathcal{H} \mid P_k \Psi = \Psi\}$。因为每个$P_k$都属于$\mathcal{Z}$，子空间$\mathcal{H}_k$在$\mathcal{N}$和$\mathcal{N}'$下是稳定的，即$X \in \mathcal{N}$或$X \in \mathcal{N}'$，所以我们有$\forall \Psi \in \mathcal{H}_k$: $X\Psi \in \mathcal{H}_k$。那么令$\mathcal{N}_k = \{P_k N P_k \mid N \in \mathcal{N}\}$，$\mathcal{N}'_k = \{P_k N P_k \mid N \in \mathcal{N}'\}$，且注意到这些是作用于空间$\mathcal{H}_k$的冯·诺伊曼代数，空间$\mathcal{H}_k$容许存在一个循环且分离的矢量，即$P_k \Phi$，使得$\forall N \in \mathcal{N}_k$: $\tilde{\varphi}_k(N) := (\Phi_k, N\Phi_k)$定义$\mathcal{N}_k$上的一个单射正常态，因而它是对态$\varphi$的代数的限制。此外还要注意对所有的$t \in \mathbb{R}$，$\mathcal{N}_k$在$\tau_t$下是稳定的。由于$\mathcal{N}_k \cap \mathcal{N}'_k = \mathbb{C} I_k$（这里$I_k$是$\mathcal{H}_k$中的单位算符），因此$\tilde{\varphi}_k$是极值KMS。由归谬法，我们假设$\varphi_k$自身不是极值KMS。那么在$\mathcal{N}$上会存在某些KMS态$\psi$和某些$\lambda \in (0, 1)$使得$\psi \leqslant \lambda^{-1} \varphi_k$。通过$\tilde{\psi}_k$来表示$\psi$对$\mathcal{N}_k$的限制。那么特别地，我们有$\lambda^{-1} \varphi_k(N^* N) \geqslant \psi([NP_k]^* [NP_k]) = \psi_k(N^* N)$，即$\lambda^{-1} \tilde{\varphi} = \tilde{\psi}_k$。因为$\tilde{\psi}_k$是极值KMS态，且$\tilde{\psi}_k$是KMS，等式必须成立，即$\lambda^{-1} \tilde{\varphi} \geqslant \tilde{\psi}_k$；且因为$\tilde{\varphi}_k$和$\tilde{\psi}_k$是态，$\lambda = 1$，即在$(N)_k$上：$\tilde{\psi}_k(N^* N) = \tilde{\varphi}_k(N^* N)$。由施瓦茨不等式，我们可知对于每个$N \in \mathcal{N}$，$\tilde{\psi}_k([P_k N P_k]^* [P_k N P_k]) \leqslant \psi(N^* N)$且因而有$\psi \geqslant \varphi_k$。连同初始条件，即$\psi \leqslant \varphi_k$（因为我们现在知道$\lambda = 1$），由这两个不等式可归纳出$\psi = \varphi_k$。所以$\varphi_k$确实是在$\mathcal{N}$上的极值KMS态。因为$\varphi_k$是$\mathcal{N}_k$上的极值KMS态，$\psi$对这个代数的限制$\tilde{\psi}_k$必须与$\tilde{\varphi}_k$相一致，且因此$\varphi_k$是$\mathcal{N}_k$上的极大值KMS态。故$\varphi$被分解为极值

KMS 态的一个凸组合。因此由矛盾可得到唯一性。

定义 61。一个凸集C被认为是一个单形，在C上的每个点都存在一个分解为C上极值点的唯一凸分解。

回顾在二维欧几里得几何学中，三角形是一个单形，确实三角形上的每一点都可被看作是三角形顶点的一个唯一凸组合。但一个圆并不是一个单形：其端点的集合是圆周，且给定圆内的任意点，所有过此点的割线能给出端点的不同凸组合。

评论 62。

1. 因此，批注 60 可以阐释如下，即我们假设中心$\mathcal{Z}$的谱 $S_p(\mathcal{Z})$是离散的，那么$\mathfrak{S}_\beta$ 是一个单形；且分解是关于离散概率测量的加权求和，这个测量是在给定的动力学 τ 和给定的自然温度β下，由所有正常 KMS 态的集合$\mathfrak{S}_\beta$ 的极点$\mathfrak{E}_\beta$ 来支持的。根据对批注的证明过程，人们确证了后面这一陈述确实可扩展到所有正常态，且不仅是那些单射态。

2. 若 $Sp(\mathcal{Z})$不是离散的，上述求和必须用一个积分来代替，且测量理论的一些调整也是必不可少的，以确定在哪种意义上 φ 定义了一个集中于$\mathfrak{S}_\beta$ 边界的唯一测量。一般数学背景请参见[Takesaki, 1970a; Kadison and Ringrose, 1983/1986]，在这个背景下这些分解出现在对中心测量的研究中。就本文的目的而言，刚刚描述的简化版本足以确定支配 KMS 态分解为其极值点组合的唯一分解的理论的概念结构。

3. 注意由冯·诺伊曼假设描述的在一个量子系统中的态的集合不是一个单形：如果一个密度矩阵至少具有一个本征值，其阶大于或等于 2，那么将其分解为纯态的分解并不唯一。因此在量子统计物理学中 KMS 态的集合具有一个在量子世界中未曾听闻的经典特性。

4. 有待证明的是这一特性是与 QSP 紧密相关的，且因此 QSP 要求考虑这样的情形，在这些情形下相关的表征并不产生因数，参见量子力学的冯·诺伊曼形式体系，在那里正则平衡态仅导致因数表征——回顾评论 34 的结尾处。这一问题是下一节的研究对象。

5.7 极值 KMS 态，纯热力学相

为了证明纯热力学相在 QSP 中被描述为极值 KMS 态，我们可以提出的主

要证据源自于三方面的背景。

第一方面的背景基于5.3和5.4节，在那里我给出了将正则平衡态辨识为KMS态的有力证据。

第二方面的背景是下述事实，即极值KMS态是KMS理论中的基本研究对象。这使人想起在光谱学中将原子层级等同于对系统对称群的不可约化表征，著名的“群论扩散症（德语：Gruppenpest）”集中体现在[Wigner，1931]中。在数学中，这一方案被拓展到了关于我们熟悉的所谓特殊函数的系统表征上，在那里这些函数现在被作为群的不可约化表征的基础，参见如[Talman（塔尔曼），1968；Vilenkin（维兰），1968]。与此文章的中心思想更为接近的是，早先[Murray and von Neumann，1936]将因数看作是冯·诺伊曼代数理论的奠基石，根据相同的原理：由冯·诺伊曼代数的中心分解所确证的方法论选择是因数的直接积分，参见如[Kadison and Ringrose，1983/1986，Theorem 14.2.2，pp. 1027—1028]。一直以来，群论方法始终有助于解决核光谱学中的定性分类问题，以及基本粒子高能物理学中的问题。

第三方面的背景是将纯热力学相描述为极值KMS态——即其GNS表征为因子的KMS态，参见上文中的推论59（是一个数学事实），即在极值KMS态中对KMS态的分解是唯一的，参见批注60和上述评论62（2）。在QSP背景下，这一事实自然地使我们将关注点放到在热力学中会遇到的情形上，在其中平衡态唯一地分解为纯热力学相。

因而，这一节可分为两个部分。在**A**部分中，上面的推断需面对一个针对QSP的模型，在其中任何量都能被明确地计算。在**B**部分中，将纯热力学相描述为极值KMS态对一个著名的论证造成了影响，即朗道用空间关联函数对固体与流体间的基本微观区别的论证。这举例论证了将KMS分解为其极值组分的唯一分解如何有助于描述QSP中纯热力学相和经历相位变换的系统中自发对称破缺是可共存的。对后者的进一步讨论请参见[Liu and Emch，2005]。

A. 铁磁体的量子外斯—伊辛（Weiss-Ising）模型

首先回顾在5.3节中荒木揭示的关于铁磁体相位跃迁不存在的结果：对于那里涉及的每个模型而言，唯一的KMS态是极值的。

为了确定这种符合性是否适应于确实显示出不同热力学相位的系统，我们现在转向研究在相位变换物理学中具有悠久历史的一类模型[Weiss，1907；Brout(布劳特)，1965]，且数学家们认为它经得起充分详细分析的检验[Kac，1968]，这就是铁磁体的外斯—伊辛模型。

考虑一维点阵$\mathbb{Z}$，其中每个点$k \in \mathbb{Z}$上都存在一个量子自旋σ_k。对于每个有限弦，$\Lambda \in \mathbb{Z}$与一个哈密顿量相关：

$$H = -\sum_{k\in\Lambda}[B + B_{\Lambda,k}]\sigma_k^z\text{，其中 } B_{\Lambda,k} = \frac{1}{2}\sum_{j\in\Lambda} J_{\Lambda,jk}\sigma_j^z \tag{104}$$

其中B被解释为与特定方向z平行的均匀外磁场；$B_{\Lambda,k}$是一个平均磁场，即所谓的“分子”场，由于区域Λ中的所有其他自旋导致的位置k处的自旋存在于该场中。施加于范·德瓦尔斯(van der Waals)和外斯型模型上的人为假设为：相互作用$J_{\Lambda,jk}$的强度随Λ的大小$|\Lambda|$的减小而减小，该假设使得上述模型在热力学极限下精确可解，将此与5.1节中的BCS模型中的相互作用特征$|v(p, q)| \leqslant c/|\Lambda|$相比较。

这里采用[Emch and Knops(埃姆什和诺普斯)，1970]的简化版本，我们假定：

$$J_{\Lambda,jk} = \begin{cases} |\Lambda|^{-1}J > 0 & \text{当 } j \neq k \text{ 时} \\ 0 & \text{其他情形} \end{cases} \tag{105}$$

控制热力学极限$|\Lambda| \to \infty$，我们发现当$T < T_c$时会出现两个极值KMS态，这里$1/kT_c = \beta_c = J^{-1}$。它们可由一个全局可观测量的下述特征而被识别出——参见定义15和批注23——即，磁化强度$\mathbf{M}$，其三个分量：

$$M^i = \text{weak op. limit}_{|\Lambda|\to\infty}\frac{1}{|\Lambda|}\sum_{k\in\Lambda}\sigma_k{}^i \quad (i = x, y, z)$$

是在相应的温度独立表征中定义的。它们满足：

$$\begin{cases} (\mathrm{i})\, M^x = M^y = 0 \\ (\mathrm{ii})\, M^z = \tanh[\beta(B + JM^z)] \end{cases} \tag{106}$$

评论63。

1. 对于横向分量，M^x和M^y，式(106. i)被认为是源于系统的对称性。有意义的部分是平行于外加磁场的分量M^z的结果：式(106. ii)为经典自洽方程，

该模型将一个相位跃迁解释为存在一个温度 T_c(其中 $\beta_c = J^{-1}$)，低于它 M^z 就不会在 $B\to 0$ 时消失，而是趋向于一个有限的、依赖于温度的值，即所谓的自发磁性。

2. 因此在热力学极限下，确定模型的极值 KMS 态的问题具有两个新的解，在上述的 T_c 时不存在，这些极值 KMS 态表现出两个相反的自发磁化特征，这些特征是经典情形如[Kac，1968]中经常会遇到的两个纯热力学相位所具有的。

3. 相位跃迁发生于 $T = T_c$ 时，经常伴随的是在 $T < T_c$ 时，局域哈密顿量(104)中触发对称 $\sigma_k{}^z \mapsto -\sigma_k{}^z$ 会发生自发破缺。

4. 在这点上我们应该注意[Kac，1968]中的论述是在几乎正统的经典统计力学精神下进行的：由极限 $|\Lambda| \to \infty$ 下配分函数的最速下降法进行分析。[Emch and Knops，1970]中的新颖性在于同样也考虑了量子自旋的 x 和 y 分量，并研究因此而产生的量子动力学，以便对极值 KMS 态的解释符合已经从经典论述得到的已知结果。

正如有了在 5.1 节中回顾的 BCS 模型，它同样也具有一个“分子”场模型的结构，这里涉及了一些技术性细节：在热力学极限下，仅对属于所考虑的表示的冯·诺伊曼代数而言，才存在演化的收敛。

5. 给出式(105)的这一模型的简化版本，允许我们证明本小节中研究的普遍特征。然而我们可能会提及在[Emch and Knops，1970]中 $J_{\Lambda,jk}$ 被认为是依赖于距离 $|j-k|$ 的，但仅在这样一种方式下对于每个 $k \in \mathbb{Z}$，都存在一个常数 c_k，使得对于包含 k 的每个有限 Λ 有 $\sum_k |J_{\Lambda,jk}| < c_k$，它服从条件 $\forall j, k \in \mathbb{Z}$: $\lim_{|\Lambda| \to \infty} J_{\Lambda,jk} = 0$，因此热力学相位的集合会变得更复杂，但其分布仍然说明针对自发对称破缺的分解方法的充分性。

6. 我们第一次在 4.1 节已经指出，作为关于复分析的刘维尔定理的结果，KMS 态必然是时间演化不变的。然而，正如在[Emch and Knops，1970]中证实的，当前的模型允许非极值时间演化不变的极值 KMS 态——即不同时间演化不变态的凸求和的 KMS 态，而非 KMS 态存在——虽然这些极值 KMS 满足一个关于空间平移的非常强的聚集特征。那些非极值时间演化不变的极值 KMS 态的存在反映了这样的事实，即时间演化不具有渐近阿贝尔性。这不是模型的

一个意料之外的独特性，因为经验表明时间平移群的渐近阿贝尔性在 QSP 中很少能被满足——虽然已知有少数例外，其中有 5.4 节一开始讨论过的 XY 模型的平衡部分，虽然存在这样的事实，即定域性要求空间平移群具有非常强的渐近阿贝尔性。

液体和气体的共存——比如说水蒸气和液态水——在形式上呈现出与方向相反的磁相位共存方面的相似性。经典统计力学中的晶格气体模型方法与经典物理学中的铁磁对应方法非常相似：我们不是赋予每个点一个正则 n 维伊辛模型，一个取值为 $+1/2$ 和 $-1/2$ 的经典自旋，而是考虑一个由点阵中的点来标示的随机变量，且其取值为 1 或 0 取决于该点是否被一个分子所占据，双重(或更高重的)占据性被这些模型中的法则排除掉了。从现象上看，在相位图上液气共存曲线的变化接近于磁性物质共存的曲线。特别是二者都呈现了一个临界点，恰好都位于发散涨落存在的相空间中。在高于临界温度的温度下，不可能对液体与气体做出任何区分，这时物质的态被描述为流体是最适合的。

B. 影响朗道论证的 QSP

流体与相同物质的晶体相位共存的情形——比如说处于流体相与冰这种固相的水——在现象学上非常不同于气体—液体相位跃迁所呈现出的情形。这里没有临界点：液体—固体共存描绘的曲线随着压力与密度的增加而无限延长。朗道对流体—固体共存曲线中无临界点存在提出了一种启发式的论证，参见[Landau and Lifshitz，1958b，260]。该论证被乌伦贝克在[Uhlenbeck，1968，17]所接受："因为固体和流体相对于长程序列来讲性质是不同的，不能存在一个临界点，因为经过该点将会意味着长程序列将逐渐显现，而这是不可能的。这是朗道的论证，我发现它完全可信。"并且甚至，乌伦贝克在同一页上警告说："人们不能够避免这样的事实(直觉上是显然的，虽然未经检验!)，即在固体相位本身中已经存在长程序列。"

在一系列令人印象深刻的论文中，卡斯特勒等[1967]接受了挑战，在这些论文中普遍存在的渐近阿贝尔性的各种假设，在[Emch *et al.*，1970]给出的版本中被表明是可有可无的，该版本见下文。

该方案是将出现在一个具有欧几里得不变性的 KMS 态的分解式中的极值 KMS 态进行分类。令$\mathcal{A}$是 C^* 代数，它作为定域代数$\mathcal{A}(\Lambda)$在有限(这里"有限"

意味着有限体积：$|\Lambda|<\infty$）区域 $\Lambda\subset\mathbb{R}^3$ 的一个吸收集上的 C^* 归纳极限而得到的。令 α：$t\in\mathbb{R}\mapsto\alpha_t\in\mathrm{Aut}(\mathcal{A})$ 描述了一个演化；令 ν：$g\in\mathbb{E}^3\mapsto\alpha_g\in\mathrm{Aut}(\mathcal{A})$ 描述欧几里得群 $\mathbb{E}^3$ 的作用；且令 φ 是在 $\mathcal{A}$ 上的关于温度 β 时的演化 α 的一个 KMS 态；φ 被假定为在欧几里得群的作用下不变，即 $\forall g\in\mathbb{E}^3$：$\varphi\circ\nu_g=\varphi$；这一条件是由基本的相互作用具有欧几里得不变性这一现象方面的预期结果所引出的。

在下述两个条件被满足的意义上，进一步假定 φ 关于 $\mathbb{E}^3$ 的作用量是强传递的。

1. 对于任意两个出现在将 φ 分解为极值 KMS 态的分解式中的态 ψ 和 ψ'，至少存在一个 $g\in\mathbb{E}^3$ 使得 $\psi'=\psi\circ\nu_g$。

2. 对于一个——因而所有的——出现在将 φ 分解为极值 KMS 态的分解式中的态 ψ，稳定子群 $G_\psi:=\{g\in\mathbb{E}^3\mid\psi\circ\nu_g=\psi\}$ 包含至少三个非共面的平移。

注意对任意的 $g\in\mathbb{E}^3$ 和任意出现的 φ 在极值 KMS 态的分解中的 ψ，态 $\psi_g:=\psi\circ\nu_g$ 同样也出现在那里，且有 $G_{\psi_g}=g^{-1}G_{\psi_g}$。因此，由于共轭性，所有出现在 φ 分解式中的元素都具有相同的对称性。这一共轭性集合记作 G^φ，且将其指代 φ 的内在对称性。正是 φ 的欧几里得对称性的一部分在当 φ 被分解为其极值 KMS 组分时得以保留。其结果是，条件(1)本质上是一种便利性：若它不被满足的话，人们首先不得不将在共轭元素集合中的分解态分离开来，然后对每个分开的集合做出下面概述性分析。条件(2)排除了在这里人们不想考虑的反常情形。数学上，它强化了条件(1)以保证每个极值态在 $\mathcal{A}$ 上的所有态的空间中平移群 $\mathbb{R}^3\subset\mathbb{E}^3$ 下的轨道是闭合的。

之后在[Emch *et al.*，1970]中证明了一个满足上述条件的具有欧几里得不变性的 KMS 态 φ 必然属于下述四种类型之一。

第一种类型出现于当 φ 已经是极值 KMS，即其内在对称性是群 $\mathbb{E}^3$ 自身的情况下，这一情形准确来说发生在当一个——因而所有的——下述等价条件被满足时：

1. φ 是极值 $\mathbb{R}^3$ 不变的，即不能够被分解为在 $\mathbb{R}^3$ 中所有平移不变的态的一个凸组合。

2. $\mathbb{R}^3$ 的幺正表征，通过 GNS 构造与 φ 正则关联，它的生成元 $\mathbf{P}$ 的谱精确

地说包括一个本征值，即 $\mathbf{k}=0$，且这一本征值是非简并的。

3. φ 是空间中的一致聚集，即对于每个 $\epsilon>0$ 和 $A\in\mathcal{A}$，存在空间的一个有限区域 $\Lambda\subset\mathrm{IR}^3$ 使得：

$$\forall B\in\mathcal{A}(\Lambda^c):\ |\varphi(AB)-\varphi(A)\varphi(B)|\leqslant\epsilon\|B\| \tag{107}$$

其中$\mathcal{A}(\Lambda^c)\subset\mathcal{A}$是局域代数$\mathcal{A}(\Omega)$的 C^* 归纳极限，有 $\Omega\in\mathcal{F}$和 $\Omega\bowtie\Lambda$（即 $\Omega\cap\Lambda=\emptyset$），见上文定义 27，批注 29 和推论 30。

从这些特征来看，属于此类集合的态 φ 被解释为一个流体相。

为了描述其他三种类型，即不能描述流体的具有强传递欧几里得不变性的那些 KMS 态，我们现在关注 φ 的内在平移不变性概念。对于出现在将 φ 分解为极值 KMS 态的分解式中的任意态，令 G_ψ 表示 ψ 的欧几里得对称子群，且令 $H_\psi=G_\psi\cap\mathrm{IR}^3$ 表示保留 ψ 的空间平移子群。正如对共轭性集合定义的回顾中所得到的，人们证明了该群确实具有初始态 φ 的特征。同样也注意到强传递性要求 $\mathrm{IR}^3/H\psi$ 是紧致的。

欧几里得的、强传递的 KMS 态的第二种类型现在通过下述等价条件来确定，在该条件中 ψ 是出现于将 φ 按其极值 KMS 组分进行的分解中的任何态。

1. G_ψ 是一个结晶体群。

2. φ 不是极值 IR^3 不变的，且 H_ψ 是通过三个非共面的平移生成的。

3. 有了$\chi=\eta^{\mathrm{IR}^3}[\psi]$——其中 η^{IR^3}是平移群 IR^3 引起的任意不变性——χ 是 η 聚集的（见上文定义 22），但既非弱混合的，也非部分均等地弱混合，即 χ 满足：

$$\forall A,\ B\in\mathcal{A}:\ \eta^{\mathrm{IR}^3}(\chi(\nu\cdot[A]B)-\chi(A)\chi(B))=0 \tag{108}$$

但并不满足下述任意一个更强的条件：

$$\forall A,\ B\in\mathcal{A}:\ \eta^{\mathrm{IR}^3}|\chi(\nu\cdot[A]B)-\chi(A)\chi(B)|=0 \tag{109}$$

$$\forall A,\ B\in\mathcal{A}:\ \eta^{\mathrm{IR}^1}\left|\eta^{\mathrm{IR}^2}(\chi(\nu\cdot[A]B))-\chi(A)\chi(B)\right|=0 \tag{110}$$

$$\forall A,\ B\in\mathcal{A}:\ \eta^{\mathrm{IR}^2}\left|\eta^{\mathrm{IR}^1}(\chi(\nu\cdot[A]B))-\chi(A)\chi(B)\right|=0 \tag{111}$$

评论 64。

1. 分别来看，条件（1—3）中的每一个都不包含 φ 是一个流体相的情况。

确实，一个流体相是极值 KMS 的，从而其内在对称性是欧几里得群 $\mathbb{E}^3$，这与(1)相冲突；流体相是极值 $\mathbb{R}^3$ 不变的，这与(2)相冲突；流体相是均匀聚集的，它意味着关系式(108)—式(111)中的每一个都能被满足，而式(109)—式(111)在当前的相位下不能被满足。

2. 其他两个 φ 所属的类型可被描述如下。若式(109)得以满足的话，则要求 $H_\psi = \mathbb{R}^3$，因而与条件(2)的第二部分相冲突。这将对应于如下情形，在其中旋转对称性被打破而态 φ 的平移对称性会在其分解为极值 KMS 组分的分解式中完全保留。虽然这可能会发生在表现出自发磁性的系统中，但它直接与一个目的相关联，即在基本相互作用在欧几里得群 $\mathbb{E}^3$ 下不变的世界中，确定在形式体系下区分流体与固体的方式。

类似地，如果式(110)或式(111)得以满足的话，会要求 H_ψ 在一个或两个方向上连续，但在互补的方向上是离散的。这样的情形同样也被设想过——早在 1930 年代中期，参见如[Landau and Lifshitz, 1958b, 410]——但此处再一次指出，它们被假设的存在性并不会直接影响当前探讨的问题。

3. 空间平均态 χ，它被构造为是 $\mathbb{R}^3$ 不变的，仍然保留了态 φ 的对称性。的确 $\mathbb{R}^3$ 的幺正表征群的生成子 P_χ 的谱的离散部分与对应于 χ 的 GNS 构造相关，相容于 H_ψ 的交互群，即有：

$$H_\psi{}^* = \{\mathbf{k} \in \mathbb{R}^3 \mid \forall \mathbf{a} \in H_\psi: \ \mathbf{k} \cdot \mathbf{a} = 0 \bmod 2\pi\} \tag{112}$$

在原理上，交互群在 X 射线衍射形态下是可观测量。

为了避免发生上述在流体与固体共存曲线中临界点不存在的现象，对将欧几里得不变正则平衡态分解为其纯热力学态相位组分的分解式进行的分析，为这些态的聚集特征和几何特征提供了一种固有联系。即流体相表现出均匀的聚集特性式(107)，而晶体相导致了弱聚集具有的更弱的特性式(108)，因而证明了朗道的论证。

6. QSP 何去何从?

最后这节可以作为一个总结，一个结论，作为深化未在正文中讨论的理论的某些方面内容的一个附录集，并且我希望，这也可作为对此论文范围之外研

究领域的一个建设性的简介。

我来简要总结一下前面的内容。首先，在第 1 到 3 节中我回顾了在传统 QSP 文献中被以不同方式讨论的 QSP 的一些主要特征。之后，在第 4 和 5 节中，我指出 KMS 条件的代数形式体系提供了一个清晰可辨的语法规则，其语义支持平衡态 QSP 中的如下关系。

· 正则平衡态是由 KMS 态描述的，该概念很自然地将有限系统转换为在热力学极限下考虑的系统。

· 纯热力学相位是由极值 KMS 态来描述的。

· 出现在正则平衡态唯一分解式中的纯热力学相位组分可能具有一个低于初始态的对称性：只有起不同作用的相位的流形才反映出最初的对称性，参见 [Liu and Emch，2005]，其中我们在相位跃迁的量子理论中描述了自发对称破缺的"分解说明"。

针对这一背景，本节的内容分为四个小节。我首先回顾了极限的数学概念及本文主要内容中用到的关于它的物理解释。之后我将再次讨论宏观可观测量的概念，这里选取的角度引出了下一小节讨论的内容：量子测量问题。最后，我会就数学物理学家和理论物理学家们有建设性的争议历史给出一些评论——有前景的和/或需修正的——以使得他们能令科学哲学家们更了解这个广泛的研究领域。

6.1 QSP 中的四类极限过程

在对刚才回顾的论题的处理中，且早在第 2 与 3 节中，我们就单独地或同时地至少遇到了四种不同类型的极限。

1. 经典极限 $h\to 0$；
2. 高温极限 $T\to\infty$；
3. 热力学极限 $|\Lambda|\to\infty$；
4. 范·霍夫极限 $\{\lambda\to 0$ 与 $t\to\infty\}$，其中 $\tau:=\lambda^2 t$ 保持为有限的。

由于在其他地方都有人分别质疑过这四类极限(或"极限过程")的哲学合法性，因此我应该再次给予详述——用口语化的方式，即不明确地用到正统的 $(\epsilon, \delta_\epsilon)$——在本章中这些极限被一致地理解为数学意义下的可控(controlled)

极限：你给我一个容忍度，我告诉你代价；容忍度越小，代价就越高；但不论你愿意容忍的误差有多小，总是存在一个代价，在此代价之下你能够得到一个保证，即约定误差将会在你决定想要容忍的范围内。此外数学物理学还增加了一个要求，即“代价”是用公认的实验专家们承认的通用形式来表示的。我们就以此视角来逐个分析上述四类极限。

1. 经典极限

普朗克常数是一个基本物理学常数，在厘米—克—秒单位制中 $h \simeq 6.62 \times 10^{-27}$尔格秒，这里用的是熟悉的记号$\hbar = h/2\pi$。说它小是一种“价值判断”，这反映出了你对与你希望讨论的问题相关的能量尺度的认识。为了揭示极限过程的作用和它们的物理意义，我们分析一个具体的案例，一个典型的量子现象的经典极限——隧穿效应，在其中能量为 E 的粒子“穿过”高度为 $V_0 > E$ 的壁垒。这一效应是 1928 年伽莫夫（Gamov）、哥乃和康登（Gurney and Condon）在说明重核会放射阿尔法粒子的研究中分别独立发现的［Gamov，1928；Gurney and Condon，1928；Gurney and Condon，1929］。约瑟夫森（Josephson）结——在两个超导体中夹着的氧化物层——是对该量子现象的一种最新的阐释，参见［Josephson，1982］。在这里我们考虑最简单的模型，即穿越一个正方形的一维壁垒的量子隧穿。

我们可以很快证实薛定谔方程：

$$\left[-\frac{1}{2m}\hbar^2\frac{d^2}{dx^2}+V(x)\right]\Psi(x)=E\Psi(x)，其中 \tag{113}$$

$$V(x)=\begin{cases}0，若 x<a\\ V_o，若 -a<x<a，其中 0<a<\infty 且 0<V_o<\infty\\ 0，若 x>a\end{cases}$$

当$0<E<V_o$时，具有如下形式的解：

$$\Psi(x)=\begin{cases}A_-e^{ikx}+B_-e^{-ikx}，若 x<a\\ Ae^{\kappa x}+Be^{-\kappa x}，若 -a<x<a\\ A_+e^{ikx}，若 x>a\end{cases} \tag{114}$$

$$其中：k=\left\{\frac{2mE}{\hbar^2}\right\}^{\frac{1}{2}}，和 \kappa=\left\{\frac{2m(V_o-E)}{\hbar^2}\right\}^{\frac{1}{2}} \tag{115}$$

这里式(114)中五个系数A_-，B_-，A，B，A_+的相对比率是由强加的四个条件来确定，即Ψ和其微商(即导数)在边界$x=\pm a$处是连续的。特别地，这些条件意味着：

$$A_-=A_+e^{ika}\frac{1}{4ik\kappa}[(\kappa+ik)^2e^{-2\kappa a}-(\kappa-ik)^2e^{2\kappa a}]$$

那么按照逆三角不等式$|a-b|\geqslant\max\{|a|-|b|,|b|-|a|\}$：

$$\left|\frac{A_-}{A_+}\right|\geqslant\frac{k^2+\kappa^2}{4k\kappa}(e^{2\kappa a}-e^{-2\kappa a})=\left[\frac{1}{2}\frac{V_o}{\sqrt{E(V_o-E)}}\right]\sinh 2\kappa a$$

因为术语[…]独立于$\hbar$，所以为了强调$\hbar$的地位，我们将上述公式重写成：

$$\frac{|A_+|^2}{|A_-|^2}\leqslant C[\sinh 2\kappa a]^{-2} \tag{116}$$

按照式(114)中$A_\pm$的定义，式(116)的左边被解释为隧穿的透射系数。在相对应的经典模型中，$0<E<V_o$要求这个系数不存在。因而，要求量子模型近似于经典对应模型就是要求量子透射系数任意地小，比如说：

$$\frac{|A_+|^2}{|A_-|^2}\leqslant C\,[\sinh 2K]^{-2} \tag{117}$$

其中K可以为任意大。为了保证式(117)能被满足，那么导出式(116)的计算就表明条件$\kappa a>K$是充分的，即：

$$\hbar<K^{-1}[2m(V_o-E)]^{\frac{1}{2}}a \tag{118}$$

因此，量子系统式(113)的经典极限现在就会被限制：

(i) 在数学上，通过式(117)和式(118)的合取来约束；

(ii) 在物理上，正如式(118)那样用描述该系统的物理量给出了有效范围的一个估算。

在这个意义上，经典极限类似于非相对论极限：经典描述在量子理论中出现的方式与牛顿力学在爱因斯坦狭义相对论中出现的方式相同。正确理解的关键在于对近似有效性范围的评价。这样做了之后，我会毫无顾虑地向保险经纪人保证我的车不会从车库中隧穿出去，在拥挤的交通中驾驶也不用担心相对论的红移。这些是我的车、车库、我所居住的小镇所涉及的物理参数，参见[Gamov，1940]，其中伽莫夫出于教学兴趣假设了$h\simeq1$尔格秒或$c\simeq15$千米/小时。

2. 高温极限

按照埃伦费斯特的方法，我在本节中重复论证了在 QSP 中，当温度足够高时系统就会表现为经典系统。通常，那些表明我们所处体系的物理量类似于 (βh)，其中 $\beta = 1/kT$（这里 k 是玻耳兹曼常数，$k \simeq 1.38 \times 10^{-16}$尔格/秒）。

为了说明这一点，我们来看黑体辐射的结果(2.1 节)和固体的比热(2.3 节)。

我们定性地看到，若 $h\nu > kT$，即符合 1900 年普朗克公式的情况，这里(1)就再次生成了(5)，后者在 1896 年由维恩给出[Vien, 1896]。定量地，帕邢和万纳(Paschen and Wanner)[Paschen and Wanner, 1899]在 1899 年已经证实了维恩定律在可见光范围内与实验室数据一致，即波长 $\lambda = c/\nu$ 在 4000Å 与 7000Å 之间，温度达到 4000K，这就是今天我们所说的“量子体系”。正如太阳表面的温度大约为 6000K，我们不可能达到更高的温度。然而，就比率 $h\nu/kT$ 而言，增加 T 或减少 ν 都有相同的效果，后者意味着将观测置于红外区域，这在当时是可能的。确实，一年后卢默尔和普利斯海姆(Lummer and Pringsheim)[Lummer and Pringsheim, 1900]记录了波长在 12μ 到 18μ 范围内时维恩定律的系统偏差($1\mu = 10^{-6}$ m，因而 $12\mu = 12 \times 10^4$ Å，与可见红光的 $\sim 7 \times 10^3 4000$Å 相比较)。这一观测推动了瑞利—金斯公式(6)的纯经典推导以及之后普朗克在 $h\nu \gg kT$(维恩)和 $h\nu \ll kT$(瑞利—金斯)之间做出的修正。在实验方面，从帕邢和万纳到卢默尔和普利斯海姆在时间上急剧地(少于两年)转变标志出了从量子到经典体系边界的跨越。在数值上，这两个体系通过其与波长 λ_{max} 的距离 $|\lambda - \lambda_{max}|$ ——或等价的频率 ν_{max}——来描述，在 λ_{max} 时普朗克分布(1)达到其最大值。

对于固体的比热，在对方程(8)和(9)重新审视的基础上，德拜已经证明我们可以通过改进结论(10)来给出精确结果：

$$C_V = 3R\left\{4D\left(\frac{\Theta}{T}\right) - 3\left(\frac{\Theta}{T}\right)\left[\exp\left(\frac{\Theta}{T}\right) - 1\right]^{-1}\right\} \tag{119}$$

$$\text{其中 } D(x) = \int_0^x dt \frac{t^3}{e^t - 1} \text{和 } k\Theta = h\nu_o$$

批注 2 的获得是通过注意到：

$$D(x) \simeq \begin{cases} 1 & \text{对} \quad 0 < x \ll 1 \\ \frac{1}{5}\pi^4 x^{-3} & \text{对} \quad x \gg 1 \end{cases}$$

为了走得比这更远一些，并且确定经典体系 $C_V = 3R$ 的起点，我们需要注意两件事。第一，(119)中的 C_V 是变量为 Θ/T 的广义单调增函数，虽然它不能够被表示为初等函数，但它在数量上是可计算的。因为已知 Θ 是由晶体振动频率的界点 V_o 表示的，它的值可以通过数学方法来确定，例如，在室温下，Θ（对于铅来说）大约是 100K，（对于铝来说）大约是 400K，银和铜介于二者之间。对于这些以及其他许多金属，比热的测量值显著地接近于理论预测式(119)，参见如[Wannier, 1966, fig. 13.9, p. 276]。这一曲线表明了随着温度的降低，经典体系会单调而平滑地过渡到量子体系。具体地说，我们现在可以定量地讨论经典体系的起点。精确表达式(119)必然导致 Θ/T 时 C_V 展开中的前两项由下式给出：

$$C_V \simeq 3R\left\{1 - \frac{1}{20}\left(\frac{\Theta}{T}\right)^2\right\} \tag{120}$$

所以在室温 $T \simeq 300\text{K}$ 下，对经典值 $C_V = 3R$ 的修正范围是从大约 0.6%（对于铅）到 9%（对于铝），两个修正都与实验数据吻合得很好。

对于高温 QSP 下经典体系突现的其他早期观点，请参见 2.4 节与 2.6 节，在那里我曾指出通过控制玻色与费米量子气体的高温极限，我们就能重新获得经典理想气体。

3. 热力学极限

正如概念名字所示，热力学极限是为了从微观力学模型推导出不同的宏观热力学行为而设计的。这里我发现在非平衡态和平衡态统计物理学中分别提出问题会比较简便。

a. 非平衡态物理学。早在经典领域内，人们就利用了所考虑系统规模很大这个特性，来避免在物理现象的理论模型（如趋向平衡态的热力学方法模型）中反复出现的具有迷惑性的表象。例如，为支持气体的玻耳兹曼动能理论，埃伦费斯特提出了所谓的狗蚤(dog-flea)模型，该随机模型随后被马克·卡克所修正。这一模型在[Emch and Liu, 2002, Section 3.4]中被讨论，其中报告了计

算机实验的结果，涉及在两只“狗”之间“随机”跳跃的 $N=100$ 个“跳蚤”，而在数以千计的试验中，重现的频率被观测到——卡克表明它随 N 会指数地增加——正如期望的那样有规则地发生着。

在量子领域内，核子自由张弛这个真实的实验模型在上文 3.3 节中已经得以解决。这里再一次提到，该模型说明了达到平衡态的方法，在现实中它并不受随 2^N 增大的“重现时间”的影响，其中 N 是处于宏观的 CaF_2 晶体系统中的点阵数。所以对极限 $N\to\infty$ 的经验验证是，对于实验者而言，相关的时间参数 $\sim 2^{10^{23}}$，它确实是极高的；相应地，我没能从实验[Lowe and Nordberg（洛维和诺德伯格），1957]中看到对公认的重现的任何关心。支撑性的分析证据是由方程式(49)给出的明确的尺度修正，在备注 6 中我对此做了详细讨论。

b. 平衡态物理学。在平衡态情形中，热力学极限被用来分析大量物质的属性，可以说它是为了使我们能在远离具有边界效应的浅滩的公海中航行而提供的导航。与 QSP 中一样，在 CSP 中，这往往需要一些复杂的设置。

粗略来讲，直到 20 世纪中叶，这个目标才通过用积分取代求和而实现，就像在方程(13—14)中那样。正如在数学物理学的其他分支中那样，这一数学方法通常被控制得很好，虽然物理学有时需要非同寻常的谨慎，如方程(50)中附加说明中详细表明的那样。

后来，特别是在对相变模型的建构中，当集体行为的出现被证明对于理解当前现象来说是本质的，且当存在的问题被提出时，我们就需要更多的经验。特别是，对于系统的大小被允许为无穷大的极限情形，就需要考虑维度，那么，特别地，所考虑区域的形状必须是这样的：即面积对体积的比率趋向于零。立方体满足，而海绵则不满足。正如在第 5 节中提出的不同模型所表明的，明确且成功地实现这样的极限过程是可能的。最简单的例子是晶格系统，比如在晶格$\mathbb{Z}^d$上的自旋。第 5 部分中也给出了连续系统的例子，然而，在一般情况下，这样的系统，比如说 $\mathbb{R}^d$ 上的系统，需要额外的技术支持以确保空间上的均匀性以及避免聚束效应，因而，在后一种情形中理论并不总是能如人们所设想的完全被控制，参见[Sewell，2002]。极长程的相互作用可能会在以下方面带来更多的问题：(a)极限态的定义；(b)极限时间演化的控制。这些情形出现在 5.1 和 5.7. A 小节中。

4. 范·霍夫极限

我们在3.5节与5.4节中遇到过该极限过程的特殊案例。但本节中给出了更为一般的讨论。

利用一次非常好的主题变换，范·霍夫看起来试图通过对散射理论中长时渐近行为的讨论来为玻恩(Born)近似做出辩护，他在[Van Hove，1955]中提出了一个变量，以便允许他描述这样一种体系，在该体系中，他强调不可逆宏观现象的时间尺度超出了对潜在可逆的，即哈密顿量的微观动力学时间尺度的范围。范·霍夫的原始表示是针对特殊模型通过微扰方法来给出的，他将其展开到所有阶，并选择其中“最发散的曲线”进行求和。最初，他的大师级表现遭到了相当多的质疑，参见[Van Kampen，1962]。主要的问题是如何得到这样的条件，即在其中可能成为理论的本质内容的部分是源于与特殊模型的解相关的偶然图像。系统化数学分析现在可被用来说明一个联合长时/弱耦合极限是如何导致研究对象从保守的幺正演化转变为收缩动力学半群的，参见[Martin，1979；Davies，1976a]。

就本节的关注点，即对极限的控制(这种控制允许人们能确保所关注的系统在预期的体系内运行)而言，对于这里的指数衰变，人们想要证明的是具有如下形式的结果，参见如[Martin and Emch，1975，Section 4]。

存在有限常数$\tau_o>0$和$C>0$使得对于$0\leqslant\lambda^2t\leqslant\tau_o$有：

$$\left|\lim_{|\Lambda|\to\infty}(\Phi,\ U^o_{-t}U_t\Psi)_\Lambda-(\Phi,\ \exp(-[\Gamma+i\Delta]\lambda^2t)\Psi)\right|\leqslant\lambda C \qquad (121)$$

其中$U^o_{-t}U_t$描述了在所谓的相互作用图景下的演化，其中$U^o_t=\exp(-iH_ot)$，$U_t=\exp(-i[H_o+\lambda V]t)$；$H_o$，$H+\lambda V$，$\Delta$是自伴算符，$\Gamma$不仅是自伴的，而且为了在时间范围$0\leqslant\tau:=\lambda^2t\leqslant\tau_o$内描述衰变，它还必须是正的。因而术语“长时/弱耦合极限”的意义为：当耦合常数λ足够小，即式(121)的RHS很小时，演化就可由收缩半群$S(\tau):=\exp(-[\Gamma+i\Delta]\tau)$来近似，其中$\tau=\lambda^2t\in[0,\ \tau_o]$，条件是时间$t$在可表示为$t\simeq\tau/\lambda^2$的范围应该足够大。

将总哈密顿量$H=H_o+\lambda V$分为“未受干扰的”或“自由”部分H_o和“相互作用的”部分λV的这种分割法必须被证明。范·霍夫提出，它可追溯到这样的事实，即不可逆过程中我们感兴趣的可观测量是宏观的(见下文6.2节)，因而它

确定一个联合谱分解，在此谱分解中，H_o 是 H 的"对角部分"。例如，$A=\int dkA(k)a^*(k)a(k)$ 和 $H_o=\int dk\epsilon(k)a^*(k)a(k)$。这个观点同样有助于为应用相互作用图景 $U^o_{-t}U_t$ 做出证明，因为它意味着我们感兴趣的可观测量在"自由"演化中是不变的。正如宏观可观测量是平移不变的，记号 $\int dk$ 被用来表明：动量表示对应于其中可观测量和自由哈密顿量是对角化的那类谱分解。

20 世纪 70 年代得到的对范·霍夫极限的理解已经得到确证并得以扩展，参见如[Bach *et al.*，2000；Derezinski and Fruboes(德利津斯基和弗里博斯)，2005]以及其中的参考文献；原始资料可见于[Davies，1976a]，[Emch and Liu，2002，Section 15.2]和[Alicki and Fannes(埃里克和法纳斯)，2001]。

虽然我无意详细说明下面的历史性观点，但我可能会偶然提到相互作用图景的使用有助于范·霍夫在其微扰展开中分辨出多体物理学的某些典型特征，由此他提出了非平衡态 QSP，这是不同于支持量子散射理论 QFT 的。但直到今天，我仍不确定范·霍夫意义不明的言论是否已经被正确地吸收到了当代数学物理学文献中。

我也应当在这里提到，CSP 中也考虑了耦合极限。案例之一是经典气体的格拉德(Grad)极限，在那里体积 V 保持不变，分子数 $N\to\infty$，分子的截面 $\sigma:=\pi d^2\to 0$，因而每个分子的体积 $v(\sim d^3)\to 0$，气体的密度 $\rho:=\frac{N}{V}\to\infty$，虽然平均自由程 $\lambda=\frac{V}{Nd^2}$ 保持为常数。在[Grad，1958]中，格拉德提议将此极限作为得到玻耳兹曼方程的一种途径。针对后一问题的更多文献，请参见[Uffink，2006，Section I.6.2]或[Emch and Lin，2002，Section 3.3]。对于二维洛伦兹气体(其中 $\sigma=2d$)请参见[Martin，1979]，其中指出了格拉德和范·霍夫极限(以其适合于此模型的形式出现)在这样一种重要的意义下是等价的：它们分别预测到了一个相同的比率，即由观测确证的宏观时间尺度与两次连续碰撞之间平均自由时间所提供的微观时间尺度之间的比率。

[Batterman(巴特曼)，2002a] 中肯地评论了与渐近推理在解释、约化和突现中的作用相关的哲学问题。上述四个极限过程或许为完成这一认识论任务提

供了额外的机会，参见[Grad，1967]。

在本节结束时，我至少应该提到粗粒化，即从经典向量子世界转变的另一种方法[Van Kampen，1954；Emch，1964]。在本论文中不涉及它的一个原因是对于这里讨论的内容来说，我并不需要它。其中的原因也可能是我开始相信粗粒化的优越性很大程度上已经被无限系统的句法规则所取代了，这个无限系统允许人们避开许多与微观和宏观世界之间关系有关的棘手问题，见下文6.3节。然而，由于有了热力学极限，因此粗粒化有助于探索那些我们希望在精细描述中出现的宏观特征，这样做同样还强调了，区分不同的尺度和速度，会使这种转变能够更加平滑。

6.2 宏观可观测量

回到一般形式体系，假设我们已得到热力学极限，并将注意力集中于空间平移上，那么3.5节就强调了一个对量子遍历理论来说是新的特征。在由[Wick *et al.*，1952]提出的超选择分支(sectors)理论的意义上，空间平均可观测量是基本可观测量，即它们与所有准局域可观测量对易，且它们相互之间也对易。这仍然是量子理论的另一个经典方面。以此方式产生的具体的经典描述依赖于系统的全局制备(但对局域干扰并不敏感)，正如依赖于我们所考虑系统的具有变换不变性的态 φ 的空间平均可观测量的定义及其值所表明的那样。量子遍历理论的这一方面是以哈格—卡斯特勒公理中假设的“定域性”的直接结果的方式出现的。

因而将量子微观描述之外的经典宏观描述的出现看作是平移不变性和定域性的结果是恰当的，参见3.5.B节和5.7节。正如我们看到的，通向热力学极限的过程和随之产生的宏观可观测量允许我们分辨同时存在的多个热力学纯相，就像在零磁场中非零磁化强度表明永磁体的出现的例子那样。类似地，如果我们取热力学极限的话，当麦克斯韦稳定状态快结束时，在等温线推导过程中出现的非连续性的实验观测就可以被更好地理解，否则，等温线一直是解析的，并且如果考虑的是一杯茶的话，那么对等温线是如此弯曲的理论进行描述既不简便也没有用处。而且，没有人会说，在他们拿到饮料时，仅仅是因为冰块的尺寸是无穷的，他们才会发现这些冰块……幸亏不是这样的。然而，将固

体与液体区分开来的朗道标准(见5.7.b节)仅在考虑热力学极限时才严格有效。仅当忘记极限的定义时它会是一个悖论，这里和在物理学其他分支中一样，对极限正确理解的关键在于证明它们出现在性质上的不同体系中。

对时间平均可观测量，情况要更复杂些。回想这样一些基本事实。对任意的时间不变态 φ，GNS结构将准局域可观测量的 C^* 代数$\mathcal{A}$的表征 π_φ 与时间演化的幺正表征联系了起来，在时间演化下作为 $\pi_\varphi(\mathcal{A})$ 的弱算符闭包的冯·诺伊曼代数，即$\mathcal{N}_\varphi := \pi_\varphi(\mathcal{A})''$是稳定的。虽然可观测量的时间平均总是像其空间平均一样属于$\mathcal{N}_\varphi$，但现在时间平均也属于此代数的对易子$\mathcal{N}_\varphi' = \pi_\varphi(\mathcal{A})'$，从而属于其中心点$\mathcal{Z}_\varphi := \pi_\varphi(\mathcal{A}'') \cap \pi_\varphi(\mathcal{A})'$，当且仅当演化是 η 阿贝尔的。后一条件——见方程(61)——可能在某些特殊模型中得以满足，但其地位仍然是不稳定的以至于不能在与“定域性”相同的基础上将此条件奉为普遍“公理”。

尽管在上面才刚描述了这个极限，但事实上在时间演化下可观测量的一些其他遍历特征和理论的某些应用在3.5.A节中就已经讨论过了。

6.3 从QSP视角看量子测量过程

量子测量的技术文献在1970年代经历了某些显著的发展——参见如[Hepp(亨普)，1972；Bell，1975；Whitten-Wolfe and Emch(惠滕—沃尔夫和埃姆什)，1976]，也可见于[Emch，2003；Sewell，2005]——导致这种结果的部分原因是QSP代数方法的出现。

现在，学者们关于问题是什么已经达成了一致意见，最初的学说在维格纳对他自己称之为冯·诺伊曼“正统”理论[von Neumann，1932c]的详细诠释中表达得最好，维格纳的论文收录于[Wigner，1997，Part II]与[Wigner，1995，Part II]中，维格纳对此论题的态度最终陈述于[Wigner，1984]中，[Dickson，2006]概括了一些哲学问题。

维格纳反复提出的两点批评意见在很大程度上促进了对该学说几个基本原则的重新理解和贯彻。第一点批评意见是“为了增加测量的准确性，人们不得不使用一个非常大的测量仪器”[Wigner，1995，177]或“仪器尺寸的大小对于测量能力来说似乎是本质性的”[Wigner，1995，178]。第二点批评意见是关于无限回归问题的，即所谓的维格纳的朋友的论证，参见如[Wigner，1995，

215]。它是由一种必要性得出的，即"把目前为止被称为仪器的系统看作测量对象。换言之，人们将会把此仪器引入到与新测量对象的相互作用中……"见[Wigner，1995，pp. 208—209]。当这还不成为一个问题的时候，人们通常在对经典测量的分析中考虑这个问题，维格纳反复重申一个陈述，他认为这是由福克提出的，但他说他自己相信该陈述是"哥本哈根学派"学说的一部分，即"测量装置必须以经典方式来描述"。与本文特别相关的是，此处引用出自于维格纳的题为"测量仪器是宏观的吗?"的论文中[Wigner，1995，205]。

我相信造成这种困难的核心原因是，在维格纳最活跃的时期，物理学家们仍然惊叹于已经被察觉到的经典与量子世界之间对立。因此当人们发现一个连续性的解能连接我们对所存在世界的这两种描述——量子描述与经典描述——时，关于量子测量问题的一个新的分支学科才得以发展。这发生于可控极限过程的物理地位得以概念化时，特别是当人们理解了宏观可观测量的概念时，请参见上文6.1和6.2节中的参考文献以及[Landsman，2006]。我认为，为了论述QSP而发展起来的概念有助于建构这样一个测量仪器，它是用量子术语描述的，但其测量行为发生于经典体系内。现在我将指出，至少这部分概念问题是如何与已经得以澄清的量子测量相联系起来的。

令$\mathcal{A}_S$为被测系统的可观测量的代数，令$\mathcal{B}\subset\mathcal{A}_S$为阿贝尔子代数，其自伴元素是我们感兴趣的可观测量。为了形式上的简单性，这里我们做出下述假设，其中的一些其实并不必要。

· $\mathcal{A}_S$包含一个单元I_S，且是一个有限维矩阵集。

· $\mathcal{B}$的谱是非简并的，因此每一个可观测量$B\in\mathcal{B}$的形式为$B=\sum_k b_k Q_k$，其中$Q_k=Q_k^*$；$Q_kQ_l=\delta_{kl}Q_k$；$\sum_k Q_k=I$;且$\dim Q_k=1$。

最初，我们感兴趣的系统处于态φ_S：$A\in\mathcal{A}_S\mapsto \mathrm{Tr}\rho A\in\mathbb{C}$上，我们需要测量过程来确定，对于所有$B\in\mathcal{B}$有值$\varphi_S(B)$，即对所有$k$有值$\lambda_k=\varphi_S(Q_k)$，由此我们能计算$\varphi_S(B)=\sum_k b_k\lambda_k$。

对于这种测量，一个量子工程师团队将会被要求构建一台由代数$\mathcal{A}_M$及其自伴"指针"M_k(它是与Q_k一一映射的)描述的专用测量仪器。他们令该仪器处于态φ_M。出于简单性考虑，他们假设$\mathcal{A}_M$包含单元I_M，并约定$\sum_k M_k=I_M$。最

后，他们试图建构一个相互作用的哈密顿量机制，使得当感兴趣的系统与仪器相互接触后，$\mathcal{A}_S \otimes \mathcal{A}_M$上的初态 $\varphi^o = \varphi_S \otimes \varphi_M$ 将以满足下述两个条件的方式来演化：

(a)就测量仪器而言：

$$\forall M_l: \begin{cases} \varphi_M(M_l) \to \varphi^p(M_l) = \sum_k \lambda_k \psi_k(M_l) & \text{其中} \\ \psi_k(M_l) = \delta_{kl} & \text{没有发散} \end{cases} \tag{122}$$

(b) 就被测系统而言：

$$\forall A_S \in \mathcal{A}_S: \begin{cases} \varphi_S(A_S) \to \varphi^p(A_S) = \sum_k \lambda_k \varphi_k(A_S) & \text{其中} \\ \varphi_k(A_S) = \begin{cases} \varphi_S(Q_k)^{-1} \varphi_S(Q_k A_S Q_k) & \text{当 } \lambda_k \neq 0 \\ \varphi_S(A_S) \cdots\cdots\cdots\cdots & \text{当 } \lambda_k = 0 \end{cases} \end{cases} \tag{123}$$

我将对这些设计要求作些评论。首先注意式(122)能得到值 $\lambda_k = \varphi_S(Q_k)$，由此人们可以计算那些可观测量的期望值 $\varphi_S(B)$，仪器正是被设计用来测量这些可观测量的。稍后我将详述——见式(125)——测量结果“不发散”的要求，即形式上 $\varphi^p([X - \varphi^p(X)I]^2) = 0$ 的要求意味着什么。

为了将要求(123)与教科书中流行的对测量过程的描述关联起来，我们简要考虑它在冯·诺伊曼框架中的特殊形式，在其中 φ_S 是代数$\mathcal{A}_S = \mathcal{B}(\mathcal{H}_S)$上的纯态，且 Q_k 是一维的；令$\{\Phi_k\}$为$\mathcal{H}_S$上的正交基，其中 $Q_k \Phi_l = \delta_{kl} \Phi_k$；以此为基础，在不失普遍性的情况下，我们可以记为 $\varphi_S(A_S) = (\Phi_S, A_S \Phi_S)$，其中 $\Phi_S = \sum_k c_k \Phi_k$；且 $\lambda_k = |c_k|^2$。那么式(123)取 $\varphi^p(A_S \otimes I_M) = \mathrm{Tr}(\rho^p A_S)$ 的形式，其中 $\rho^p = \sum_k |c_k|^2 Q_k$。因而，从$\mathcal{A}_S$来看，纯态向量 Φ_S 向混合密度矩阵 ρ^p 演化。在此意义下，式(123)是在系统初态不一定为纯态的情形下，冯·诺伊曼(非选择性的)塌缩假设的一般形式。

注意式(122)与式(123)是对态 φ^o 演化的约化描述，这些要求只需要关注特殊可观测量的演化，这些特殊的可观测量为：(a)仪器的指针 M_l；(b)属于系统 S 的所有可观测量 A_S。特别地，式(123)的要求不会与由复合系统∪仪器形成的幺正演化所给出的测量过程(我们将其记作→)不相容。

与冯·诺伊曼量子概率的“相对频率”观点相一致——显然受冯·米塞斯的

启发[von Neumann，1932c，fn. 156]——式(123)这种普遍形式最适合应用于针对一束粒子进行的测量，而非针对单位粒子分别进行的测量。因而——与3.1节中提到的“物理系统态”解释相一致——对测量过程的这种描述认为，系统S的初态被看作是对其制备过程的总结。例如，在具有历史性意义的斯特恩—盖拉赫实验中，入射银原子束在加热炉上由蒸发产生，参见[Jammer，1966，133]。因而，实验者们知道的是原子束的方向和炉子的温度：当然后者是一个宏观概念！同样，这里测量仪器的初态也被看作是其制备的结果，遵循这种实用主义的解释，人们不应该要求(大的!)测量仪器的初态为纯态：显然这会要求在其制备过程中加入大量的信息——这些信息事实上对于执行测量来说是不必要的，如果我们以收集由分布$\{\lambda_k\}$描述的简单微观信息为目的的话。

由于会出现各种各样的麻烦的情况——例如出现在有限系统的递归现象或(本节前面介绍的)“维格纳的朋友”闯入的现象——我们的仪器制造者将会面对严重的失败，除非他们被允许存在于足够的时间和空间中，以使下述的过程理想化——以提前选定的程度——足够近似于他们所执行的测量过程$\varphi^o\to\varphi^p$，即：

$$\varphi^p(X):=\lim_{t\to\infty}\lim_{|\Lambda|\to\infty}\varphi^o(\alpha_t^\Lambda[X])\text{，其中 }X=\begin{cases}A_S\otimes I_M\\ \text{或}\\ I_S\otimes M_k^\Lambda\end{cases}\tag{124}$$

通过选择指示器，使得在热力学极限下，$\lim_{|\Lambda|\to\infty}M_k$存在，并且——在上文6.2节的意义上——它可定义为“基本”可观测量，特别地，读者可能想要回顾其与超选规则——即正统理论将之解释为是那些经典的可观测量——的关系。现在可以明确式(122)中的“无发散”要求，即我们要求：

$$\lim_{t\to\infty}\lim_{|\Lambda|\to\infty}[\varphi^o(\alpha_t[M^k]^2)_\Lambda-\{\varphi^o(\alpha_t[(M)^k])_\Lambda\}^2]=0\tag{125}$$

允许式(124)中的热力学极限存在甚至有一个额外的好处，即人们可能要求实验装置如此以使得在制备仪器的初态φ_M时，测量结果式(124)对于局域微扰在实验上是不敏感的。这一要求意味着当仪器的初态φ_M由任意态ψ_M：$A\in\mathcal{A}_M\mapsto\varphi_M(D^*AD)\in\mathbb{C}$代替时，或者甚至更一般地，由冯·诺伊曼代数$\pi_{\varphi M}(\mathcal{A}_M)''$上的任意正常态$\psi_M$代替时，式(124)中的$\varphi^p(X)$并不改变，其中$D$

为满足归一化条件 $\varphi(D^*D)=1$ 的 $\mathcal{A}_M$ 的任意(准)局域元。这种稳定性符合实用主义的要求，即大的(!)测量仪器的制备应该相当简单。

至少对于本节的主要目的而言，这里完全介绍了量子工程师在建造测量装置时会提出的各种要求。

在求解量子测量问题时，代数 QSP 的贡献如下：在已经构造出满足所有上述规定的严格可控的具体模型的意义下，上述方案可以被彻底地实现。因此这些模型确立了代数方法对于维格纳所谓的正统理论之外的物理学基础的适用性。总之，这一方法包括了对在正统理论限定内被认为是未知经典体系的描述，参见[Hepp，1972；Whitten-Wolfe and Emch，1976；Emch，2003；Sewell，2005]以及[Landsman，2006，Subsection 6.6]中列出的其他文献。

对式(124)的反对意见(即真实世界中的实验室在空间与时间中是有限的)是有道理的。但它忽略了对取极限的主要认识，可回顾上文6.1节。这里极限同样定义了一个渐近区域，因此，对极限过程的控制允许我们考虑，好的实验确实会要求占据空间和消耗时间，每个实验都能以我们所期望的精确性来计算。测量过程涉及一般宏观现象的一个特殊例子，即“趋近平衡态”的例子。在上文的6.1(3)小节中，我再次评论了出现在此区域内的热力学极限 $|\Lambda|\to\infty$ 的作用。

极限 $t\to\infty$ 的作用在测量过程的情况中应该给予进一步的评论。这并不是说需要无限的时间以记录测量的结果，相反，根据我们对极限的一般理解，断言极限 $t\to\infty$ 的存在对于每个 $\epsilon>0$，都存在一个可以计算的时间 T_ϵ，并且使得当 $t>T_\epsilon$ 时，总能以要求的精度 ϵ 完成测量。在测量过程“结束”时，正统理论要求幺正演化突然中断，相较于这种限制，我们的量子工程师们则不需要关闭测量装置。现在，不需要首先取极限 $|\Lambda|\to\infty$，只要求他们重新考虑自己计算出的有限尺寸仪器的作用，根据计算，他们评估仪器应该有多大，才能允许在他们不得不终止测量并避免一些危险的反馈前有充足的时间 T_Λ。因此将可控极限 $|\Lambda|\to\infty$ 看作是针对理论有效性的一个实用限制，并不会比将其看作是为了从对日常咖啡杯冷却过程的描述中去除天文学上时间范围内的回归可能而在理论上执行的热力学极限($N\to\infty$，$|\Lambda|\to\infty$，其中 $D:=N/|\Lambda|$ 是确定的)更优越。相较于在有限系统中给出的描述，在热力学极限下得到的描述更接近于被观测到的对冷却的现实记录：前者的描述确实会过多地被不相干的细

节所限制。总而言之，在现实中构造测量过程的模型时，量子工程师所面临的问题并不是不能满足辅助条件 $T_{\epsilon} \ll \tau \ll T_{A}$，其中 τ 表示他们手表上显示的实验室尺度的时间范围。也可参见[Bell，1975]。

虽然模型确实证明了上述方案中的所有要求是相容的，但模型的本质不能证明：

(i) 方案中的条件对于理解测量过程来说是必要的；

(ii) 方案中的条件是充分的，正如所需的其他要求和条件那样。

关于要点(i)，上述提出的方案强调了 QSP 对于理解量子测量过程所做的贡献。一个具体的目标就是不要让有关测量过程的理论搁浅在介于量子与经典世界之间的概念沙滩上：该方案充分利用了这样的一些条件，即在这些条件下 QSP 表明了对世界的量子描述是如何涵盖了在概念上非常重要的经典描述。因而不可约化的量子/经典二元论现在退化为一些更具综合性的观点，QSP 是其中之一。经典行为在量子理论中的出现也是退相干方案的一个重要方面，虽然这两种方法很可能会融合的观点还未获得普遍的认同。关于后面这个问题的恰当描述，以及它们与测量问题的关联，我会推荐读者参阅[Landsman，2006]；在此情况下，关于退相干问题的生动的且多少有些针锋相对的争论，请参见[Anderson(安德森)，1994；Adler(安德拉)，2003]。

上述要点(ii)至少包含两方面意思。其一是，虽然那些证实本节所讨论方案的具有内在一致性的模型在数学上是严密的，但它们很难被看作是足够真实的，因而不能使实验室中的那些同事们满意。关于充分性的上述要点(ii)的另一方面意思是，我不知道代数 QSP 如何有助于系统地表达在量子测量理论中仍未解决的一些有挑战性的问题。如果非要让我选择其中一个的话，我首先会直接地关注当前“常规地”执行于单个量子系统上的测量，参见如[Rauch and Werner(拉什和维纳)，2000]或[Rauch，2005]。这里是否真正需要所谓的“多个世界”和“一致历史”方法的问题，就显得过于宽泛以至于不能在关于 QSP 的文章中进行讨论，参见[Dickson，2006；Landsmann，2006]。

6.4 数学物理学与理论物理学

由于存在大量的关于由热浴驱使的从小的或局域的偏离“回归到平衡态”的

文献，因而一些在很大程度上尚未解决的问题就显得并不重要了。关于这种耦合系统的模型，请参见6.4节中的A与C部分。

下面讨论的问题中的绝大多数同样也存在于经典统计物理学中，QSP对处理这些问题毫无作用，但它解决了其中很少一些问题，下面将对此做出阐述。

首先第一个问题是避免无限的回归。如果我们感兴趣的一个(小)系统通过一个(大的)热容器达到平衡的话，那么热容器会如何达到其自身的正则平衡与保持温度呢？人们并没有对此问题给出概念性回答，而是构想出了KMS条件，它最初被看作是从理论物理学到数学物理学的一个精巧且形式化的转换，它被证明是一个非常有用的构造工具。

KMS条件的成功要求我们对其进行深入的物理学证明。随后人们找到了本质性的答案，并分别将它们表达为多个稳定性条件。在5.4节中，这种条件被表示为存在一个这样的序列，在其中它们的公式化表述逐渐趋向真实的变分原理。因而这种发展符合一种人们广泛接受的观点，即“变分原理被认为是物理学定律的极值情形”[Itô(伊藤)，1987，Art. 441]。这很好，但是正如在其他领域中一样，如下哲学问题仍然存在，即任何科学是否应当像力学或18世纪以来的许多科学那样完全地或最终被建立在变分原理之上。关于这一点，理论物理学给出了某些其他因素，如“大爆炸”与“退相干”，但它们的解释价值、一致性与充分性仍有待证明。同时，人们偏好最新的变分原理也不是不合理的，有了它，代数QSP已经被证明能够令其传统版本得以深化。

在有关平衡态回归的物理文献中，人们提出的第二个问题是对全局传输现象如热传导以及由电子与光子或金属中随机杂质相互作用引起的电子电阻率的描述。范·霍夫提出一种方案——其中范·霍夫极限过程是其中的一部分，见上文6.1(4)——来处理此类问题。仍存在的问题之一是对给出的陈述提供一种在数学上清晰的论证。需要被完全控制的一个更大的问题是超越特殊后验模型的偶然事件。这要求人们用经得起数学描述检验的物理术语来说明普遍微观特征，这些微观特征给出了对时间尺度和/或体系的真实的描绘，在这些时间尺度和/或体系下人们才观测到各种宏观现象。我首先想到的案例就是牛顿的“冷却定律”与傅里叶的“热”理论，即温度的指数平衡和控制处于不同温度的源之间的物质的稳定温度分布的热流。在本文中，尤其是3.5节，例如方程(57)，

或5.4节，例如方程(86)，其中的材料是沿此方向走出第一步的例证。更进一步地，[Eckman *et al.*（埃克曼等），1999；Bonetto *et al.*（博内托等），2000；Bach *et al.*，2000]已经得到了有希望但仍是形式化的结果，因而我们仍然需要做许多工作来将这些结果与日常的观点联系起来，这些观点阐述的是人们对具体物质的传输系数值的计算。

第三个问题或许更麻烦。即时间是否会反演[Earman，2002；Fredenhagen，2003]，即使在梦中，我也没有见到过任何“宇宙的时间之箭”令人信服地出现在为QSP而发展出的 C^* 代数方法中……但也不能将这种可能作为异端邪说而排除出形式体系[Buchholz，2003]。

第四个方向是考虑那些远离平衡态的情形，在其中我们希望得到QSP方案的推广。

甚至有人开辟了第五个研究方向，在其中体现出了QSP与代数方法的关系。研究有关量子随机过程的成熟数学理论[Parthasarathy（帕塔萨拉蒂），1995；Hudson（哈德森），1998]是否以及如何为统计力学的约化过程带来新的启迪，这确实是一个有趣的事情。

最后，QSP当然在凝聚态物理学中及其向复杂现象研究领域的扩展过程中找到了大部分的现实证据。然而，在这里谈论那些对于完全控制这些现实应用来说是必不可少的具体技术细节，远远地超出了本文的范围。对此领域内的已有研究所完成的一个详尽的文献综述可见于[Anderson，1994]。然而，正如[Feynman，1998]所指出的，这些问题需要被重新研究，用更多的线与结对其进行编织，并促进新的哲学反思：

千百次探索，我们才能越来越接近真理……

（Vingt fois sur le métier remettez votre ouvrage…）

参考文献

[Adler, 2003] S. Adler. Why decoherence has not solved the measurement problem: a response to P. W. Anderson. *Studies in History and Philosophy of Modern Physics*, 34: 135-142, 2003. See already [Anderson, 2001].

[Aichele *et al.*, 2005] T. Aichele, U. Herzog, M. Scholz and O. Benson. *Single-photon generation and simultaneous observation of wave and particle properties*, pages 35-41 in [Khrenninkov, 2005].

[Alicki *et al.*, 1998] R. Alicki, M. Bozejko and W. A. Majewski, editors. *Quantum probability*. Banach Center Publications 43, 1998.

[Alicki and Fannes, 1994] R. Alicki and M. Fannes. Defining quantum dynamical entropy. *Letters in Mathematical Physics*, 32: 75-82, 1994.

[Alicki and Fannes, 2001] R. Alicki and M. Fannes. *Quantum dynamical systems*. Oxford University Press, Oxford, New York, 2001.

[Amrein *et al.*, 1977] W. O. Amrein, J. M. Jauch and K. B. Sinha. *Scattering theory in quantum mechanics*. Benjamin, Reading MA, 1977.

[Anderson, 1994] P. W. Anderson. *A career in theoretical physics*. World Scientific, Singapore, 1994.

[Anderson, 2001] P. W. Anderson. Science: a 'dappled world' or a 'seamless web'? *Studies in History and Philosophy of Modern Physics*, 32: 487-494, 2001.

[Andries *et al.*, 1995] J. Andries, M. Fannes, P. Tuyls and R. Alicki. The dynamical entropy of the quantum Arnold cat map. *Letters in Mathematical Physics*, 35: 375-383, 1995. See already [Alicki and Fannes, 1992].

[Arafune *et al.*, 2003] J. Arafune *et al.*, editors. *A garden of quanta, essays in honor of Hiroshi Ezawa*. World Scientific, Singapore, 2003.

[Araki, 1961/2] H. Araki. Einfürung in die Axiomatische Quantenfeldtheorie, unpublished lecture notes at the ETH-Zurich, 1961/2.

[Araki, 1969] H. Araki. Gibbs state of a one dimensional quantum lattice. *Communications in Mathematical Physics*, 14: 120-157, 1969. See also [Araki, 1970] and [Araki, 1975].

[Araki, 1970] H. Araki. *One dimensional quantum lattice system*, pages 75-86 in [Michel and Ruelle, 1970].

[Araki, 1974] H. Araki. On the equivalence of the KMS condition and the variational principle for quantum lattice systems. *Communications in Mathematical Physics*, 38: 1-10, 1974.

[Araki, 1975] H. Araki. On the uniqueness of KMS states of one-dimensional quantum lattice systems. *Communications in Mathematical Physics*, 44: 1-7, 1975. See also [Kishimoto, 1976].

[Araki, 1984] H. Araki. On the XY-model on two-sided infinite chain. *Publications of the Research Institute for Mathematical Sciences, Kyoto University*, 20: 277-296, 1984. See also [Araki and Barouch, 1983].

[Araki and Barouch, 1982] H. Araki and E. Barouch. On the dynamics and ergodic properties of the XY-mo-del. *Journal of Statistical Physics*, 31: 327-345, 1983. See also [Araki, 1984; Araki and Matsui, 1985; Narnhofer, 1998].

[Araki and Matsui, 1985] H. Araki and T. Matsui. Ground states of the XY-model. *Communications in Mathematical Physics*, 101: 213-245, 1985.

[Araki and Miyata, 1968] H. Araki and H. Miyata. On the KMS boundary condition. *Publications of the Research Institute for Mathematical Sciences, Kyoto*, A4: 373-385, 1968.

[Araki and Moriya, 2002] H. Araki and H. Moriya. Local thermodynamical stability of Fermion lattice systems. *Letters in Mathematical Physics*, 62: 33-45, 2002.

[Araki and Sewell, 1977] H. Araki and G. L. Sewell. KMS conditions and local thermodynamical stability of quantum lattice systems. *Communications in Mathematical Physics*, 52: 103-109, 1977.

[Araki and Woods, 1963] H. Araki and E. J. Woods. Representations of the canonical commutation relations describing a nonrelativistic infinite free Bose gas. *Journal of Mathematical Physics*, 4: 637-662, 1963.

[Arnold and Avez, 1968] V. I. Arnold and A. Avez. *Ergodic problems in classical mechanics.* Benjamin, New York, 1968.

[Bach *et al.*, 2000] V. Bach, J. Fröhlich and I. M. Segal. Return to equilibrium. *Journal of Mathematical Physics*, 41: 3985-4060, 2000.

[Bak, 1967] T. A. Bak, editor. *Statistical mechanics, foundations and applications; Proceedings of the IUPAP meeting, Copenhagen*, 1966. Benjamin, New York, Amsterdam, 1967.

[Bardeen *et al.*, 1957] J. Bardeen, L. N. Cooper and J. R. Schrieffer. Theory of superconductivity. *Physical Review*, 108: 1175-1204, 1957.

[Barton, 1963] G. Barton. *Introduction to advanced field theory.* Wiley, New York, 1963.

[Batterman, 2002a] R. W. Batterman. *The devil is in the details: asymptotic reasoning in explanation, reduction, and emergence.* Oxford University Press, Oxford, New York, 2002.

[Batterman, 2002b] R. W. Batterman. Asymptotics and the role of minimal models. *British Journal for the Philosophy of Science*, 53: 21-38, 2002.

[Bell, 1975] J. S. Bell, On wave packet reduction in the Coleman-Hepp model. *Helvetica Physica Acta*, 48: 93-98, 1975.

[Beller, 1999] M. Beller. *Quantum dialogue: the making of a revolution.* University of Chicago Press, Chicago, 1999.

[Belot and Earman, 1997] G. Belot and J. Earman. Chaos out of order: quantum mechanics, the correspondence principle and chaos. *Studies in History and Philosophy of Modern Physics*, 28B: 147-182, 1997.

[Benatti *et al.*, 1991a] F. Benatti, H. Narnhofer and G. L. Sewell. A non-commutative version of the Arnold cat map. *Letters in Mathematical Physics*, 21: 157-172, 1991. Erratum in [Benatti *et al.*, 1991b].

[Benatti *et al.*, 1991b] F. Benatti, H. Narnhofer and G. L. Sewell. Erratum: "A non-commutative version of the Arnold cat map". *Letters in Mathematical Physics*, 22: 81, 1991.

[Besicovitch, 1954] A. S. Besicovitch. *Almost periodic functions.* Dover, New York, 1954.

[Birkhoff, 1931] G. D. Birkhoff. Proof of the ergodic theorem. *Proceedings of the National Academy of Sciences*(USA), 17: 656-660, 1931.

[Birkhoff, 1942] G. D. Birkhoff. What is the ergodic theorem? *American Mathematical Monthly*, 49: 222-236, 1942. See also [von Neumann, 1932a].

[Bisognano and Wichmann, 1975] J. J. Bisognano and E. H. Wichmann. On the duality condition for a Hermitian scalar field. *Journal of Mathematical Physics*, 16: 985-1007, 1975. See also [Bisognano and Wichmann, 1976].

[Bisognano and Wichmann, 1976] J. J. Bisognano and E. H. Wichmann. On the duality condition for quantum fields. *Journal of Mathematical Physics*, 17: 303-321, 1976.

[Boltzmann, 1884] L. Boltzmann. Ableitung der Stefanschen Gesetzes betreffend die Abhängigkeit der Wärmestrahlung von Temperatur aus der electromagnetischen Lichttheorie. *Annalen der Physik und Chemie*, 22: 291-294, 1884.

[Bonch-Bruevich and Tyablikov, 1962] V. L. Bonch-Bruevich and S. V. Tyablikov. *The*

Green function method in statistical mechanics. North-Holland, Amsterdam, 1962.

[Bonetto *et al.*, 2000] F. Bonetto, J. L. Lebowitz and L. Rey-Bellet. *Fourier's law: a challenge to theorists*, pages 128-150 in [Fokas *et al.*, 2001].

[Borchers, 1993] H. J. Borchers. On modular inclusion and spectrum condition. *Letters in Mathematical Physics*, 27: 311-323, 1993.

[Borchers, 1995] H. J. Borchers. On the use of modular groups in quantum field theory. *Annales de l'Institut Henri Poincaré*, 67: 331-382, 1995.

[Borchers, 1999] H. J. Borchers. On the embedding of von Neumann subalgebra. *Communications in Mathematical Physics*, 205: 69-79, 1999. See already [Borchers, 1993].

[Borchers, 2000] H. J. Borchers. On revolutionizing quantum field theory with Tomita's modular theory. *Journal of Mathematical Physics*, 41: 3604-3673, 2000. See already [Borchers, 1995].

[Bose, 1924] S. N. Bose. Plancks Gesetz und Lichtquantenhypothese. *Zeitschrift für Physik*, 26: 178-181, 1924.

[Brout, 1965] R. H. Brout. *Phase transitions.* Benjamin, New York, Amsterdam, 1965.

[Browder, 1976] F. E. Browder, editor. *Mathematical developments arising from Hilbert's* problems, volume 28 in *Proceedings of Symposia in Pure Mathematics*, 1976.

[Brunetti *et al.*, 2003] R. Brunetti, K. Fredenhagen and R. Verch. The generally covariant locality principle—a new paradigm for local quantum field theory. *Communications in Mathematical Physics*, 237: 31-68, 2003. See also [Fredenhagen, 2003].

[Buchholz, 2003] D. Buchholz. On hot bangs and the arrow of time in relativistic quantum field theory. *Communications in Mathematical Physics*, 237: 271-288, 2003.

[Buchholz and Lechner, 2004] D. Buchholz and G. Lechner. Modular nuclearity and localization. *Annales de l'Institut Henri Poincaré*, 5: 1065-1080, 2004.

[Buchholz *et al.*, 2002] D. Buchholz, I. Ojima and H. Roos. Thermodynamic properties of non-quilibrium states in quantum field theory. *Annals of Physics*, 297: 219-242, 2002.

[Bunge, 1967] M. Bunge, editor. *Delaware seminar in the foundations of physics.* Springer, New York, 1967.

[Caldirola, 1961] P. Caldirola, editor. *Ergodic theories.* Academic Press, New York, 1961.

[Catto *et al.* , 1998] I. Catto, C. Le Bris and P. L. Lions, *The mathematical theory of thermodynamical limits: Thomas-Fermi type models.* Oxford University Press, Oxford, New York, 1998.

[Chandrasekhar, 1931a] S. Chandrasekhar. The maximum mass of ideal white dwarfs. *The Astrophysical Journal*, 74: 81-82, 1931. See also [Chandrasekhar, 1931b], [Chandrasekhar, 1931c], [Chandrasekhar, 1932].

[Chandrasekhar, 1931b] S. Chandrasekar. The highly collapsed configuration of a stellar mass. *Monthly Notices Royal Astronomical Society*, 91: 456-466, 1931.

[Chandrasekhar, 1931c] S. Chandrasekar. The density of white dwarf stars. *Philosophical Magazine*, 11: 592-596, 1931.

[Chandrasekhar, 1932] S. Chandrasekar. Some remarks on the state of matter in the interior of stars. *Zeitschrift für Astrophysik*, 5: 321-327, 1932.

[Chandrasekhar, 1958] S. Chandrasekhar. *An Introduction to the study of stellar structure.* Dover, New York, 1958. (1st ed. University of Chicago Press, 1939).

[Chrétien *et al.* , 1968] M. Chrétien, E. P. Gross and S. Deser, editors. *Statistical Physics, Brandeis Summer Institute*, 1966. Gordon and Breach, New York, 1968.

[Churchill and Brown, 1990] R. V. Churchill and J. W. Brown. *Complex variables and applications*, 5th ed. McGraw-Hill, New York, 1990.

[Clifton and Halvorson, 2001] R. Clifton and H. Halvorson. Are Rindler quanta real? Inequivalent particle concepts in quantum field theory. *British Journal for the Philosophy of Science*, 52: 417-470, 2001.

[Cohen, 1962/1968] E. G. D. Cohen, editor. *Fundamental problems in statistical mechanics, i, ii.* North-Holland, Amsterdam, 1962, 1968.

[Connes, 2000] A. Connes. A short survey of noncommutative geometry. *Journal of Mathematical Physics*, 41: 3832-3891, 2000.

[Cornfeld *et al.* , 1982] I. Cornfeld, S. V. Fomin and Ya. G. Sinai. *Ergodic theory*, Springer, Berlin, 1982.

[Davies, 1972] E. B. Davies. Diffusion for weakly coupled oscillators. *Communications in Mathematical Physics*, 27: 309-325, 1972.

[Davies, 1974] E. B. Davies. Markovian master equations I. *Communications in Mathemati-*

cal Physics, 39: 91-110, 1974.

[Davies, 1976a] E. B. Davies *Quantum theory of open systems*, Academic Press, London, 1976. See also [Davies, 1976b], [Davies, 1976c].

[Davies, 1976b] E. B. Davies. Markovian master equations II. *Mathematische Annalen*, 219: 147-158, 1976.

[Davies, 1976c] E. B. Davies. Markovian master equations III. *Annales de l'Institut Henri Poincaré*, 11: 265-273, 1976. See also [Gorini et al. , 1978].

[Davies, 1975] P. C. W. Davies. Scalar particle production in Schwarzschild and Rindler metric. *Journal of Physics*, A8: 609-616, 1975. See also [Rindler, 1966]; [Unruh, 1976]; [Davies, 1978]; [Hawking and Page, 1983]; [Unruh and Wald, 1984]; [Wald, 1994]; [Clifton and Halverson, 2001].

[Davies, 1978] P. C. W. Davies. Thermodynamics of black holes. *Reports on Progress in Physics*, 41: 1313-1355, 1978.

[de Finetti, 1937] B. de Finetti. La prévision, ses lois logiques, et ses sources objectives. *Annales de l'Institut Henri Poincaré*, 7: 1-68, 1937. See also [de Finetti, 1972].

[de Finetti, 1972] B. de Finetti. *Probability, induction and statistics.* Wiley, New York, 1972.

[Debye, 1912] P. Debye. Zur Theorie der spezifischen Wärme. *Annalen der Physik*, 39: 789-839, 1912.

[Derezinski and Früboes, 2005] J. Derezinski and R. Früboes. Stationary van Hove limit. *Journal of Mathematical Physics*, 46: 1-15, 2005.

[Dickson, 2006] M. Dickson, This volume, chapter 4, 2006.

[Dixmier, 1957] J. Dixmier. *Les algèbres d'opérateurs dans l'espace Hilbertien (Algèbres de von Neumann)*. Gauthier-Villars, Paris, 1957.

[Domb and Green, 1972] C. Domb and M. S. Green, editors. *Phase transitions and critical phenomena, I: exact results.* Academic Press, London, New York, 1972.

[Donkor *et al.* , 2003] E. Donkor, A. R. Pirich and H. E. Brandt, editors. *Quantum information and computation*, International Society for Optical Engineering (SPIE), (2003).

[Dubin and Sewell, 1970] D. A. Dubin and G. L. Sewell. Time-translations in the algebraic formulation of statistical mechanics. *Journal of Mathematical Physics*, 11: 2990-2998, 1970.

[Dunford and Schwartz, 1964] N. Dunford and J. T. Schwartz. *Linear operators, parti; General theory*. Interscience, New York, 1964.

[Earman, 2002] J. Earman. What time-reversal invariance and why it matters. *International Journal for the Philosophy of Science*, 16: 245-264, 2002.

[Earman and Rédei, 1996] J. Earman and M. Rédei. Why ergodic theory does not explain the success of equilibrium statistical mechanics. *British Journal for the Philosophy of Science*, 47: 63-78. 1996.

[Eckman *et al.*, 1999] J. P. Eckmann, C. A. Pillet and L. Rey-Bellet. Non-equilibrium statistical mechanics of anharmonic chains coupled to two baths at different temperatures. *Communications in Mathematical Physics*, 201: 657-697, 1999.

[Ehrenfest and Ehrenfest, 1911] P. Ehrenfest and T. Ehrenfest. *Begriffliche Grundlagen der statistischen Auffassungen in der Mechanik*, volume 4/22 of *Enzyklopädie der mathematischen Wissenschaften*, Teubner, Leipzig, 1911.

[Einstein, 1905a] A. Einstein. Ueber einen die Erzeugung und Verwandlung des Lichtes betreffenden heuristischen Gesichtspunkt. *Annalen der Phy-sik*, 17: 132-148, 1905; transl. in [Stachel, 1998] pp. 177-197. See also [Einstein, 1906a].

[Einstein, 1905b] A. Einstein. Ueber die von der molekularkinetischen Theorie der Wärme gefordete Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der Physik*, 17: 549-560, 1905; transl. in [Einstein, 1926] pp. 1-18; and in [Stachel, 1998] pp. 85-98.

[Einstein, 1906a] A. Einstein. Zur Theorie der Lichterzeugung und Lichtabsorption. *Annalen der Physik*, 20: 190-206, 1906.

[Einstein, 1906b] A. Einstein. Eine neue Bestimmung der Moleküldimensionen. *Annalen der Physik*, 19: 289-306, 1906; transl. in [Einstein, 1926] pp. 36-61; compare with Einstein's Ph. D. dissertation; transl. in [Stachel, 1998] pp. 45-63. For an update of the 1906 paper, see [Einstein, 1911a].

[Einstein, 1906c] A. Einstein. Zur Theorie der Brownschen Bewegung. *Annalen der Physik*, 19: 371-381, 1906; transl. in [Einstein, 1926] pp. 19-35.

[Einstein, 1907] A. Einstein. Die Plancksche Theorie der Strahlung und die Theorie der spezifischen Wärme. *Annalen der Physik*, 22: 180-190, 1907.

[Einstein, 1909a] A. Einstein. Zum gegenwärtigen Stand des Strahlungsproblems. *Physikalische Zeitschrift*, 10: 185-193, 1909. See also [Einstein, 1909b].

[Einstein, 1909b] Ueber die Entwicklung unserer Anschauungen über das Wesen und die Konstitution der Strahlung. *Physikalische Zeitschrift*, 10: 817-825, 1909.

[Einstein, 1911a] Berichtigung zu meiner Arbeit. *Annalen der Physik*, 34: 591-592, 1911; transl. in [Einstein, 1926] p. 62.

[Einstein, 1911b] A. Einstein. Elementare Betrachtungen über die thermische Molekularbewegung in festen Körpern. *Annalen der Physik*, 35: 679-694, 1911.

[Einstein, 1921] A. Einstein. *Geometry and experience*, (Lecture before the Prussian Academy of Sciences, January 27, 1921) pages 232-249 in [Einstein, 1954].

[Einstein, 1924] A. Einstein. Quantentheorie des einatomigen idealen Gases. *Sitzungsberichte der Deutschen Akademie der Wissenschaften, Berlin*, 1924: 261-267.

[Einstein, 1926] A. Einstein. *Investigations on the theory of the Brownian movement*, R. Fürth, editor, A. D. Cowper, transl., Methuen, 1926, reprinted by Dover, New York, 1956; this is a collection of five papers published by Einstein in the years 1905-1908.

[Einstein, 1933] A. Einstein. *On the method of theoretical physics*, Herbert Spencer Lecture, 1933; in [Einstein, 1954] pp. 270-276.

[Einstein, 1954] A. Einstein. *Ideas and Opinions*, S. Bargmann, transl. and rev., Bonanza Books, New York, 1954; based *inter al.* on *Mein Weltbild*, C. Seelig, editor. Querido Verlag, Amsterdam, 1934.

[Emch, 1964] G. G. Emch. Coarse-graining in Liouville space and master equations. *Helvetica Physica Acta*, 37: 532-544, 1964.

[Emch, 1972a] G. G. Emch. *Algebraic methods in statistical mechanics and quantum field theory*. Wiley, New-York, 1972.

[Emch, 1972b] G. G. Emch. *The C^*-algebraic approach to phase transitions*, pages 137-175 in [Domb and Green, 1972].

[Emch, 1976] G. G. Emch. Generalized K-Flows. *Communications in Mathematical Physics*, 49: 191-215, 1976. See also [Kümmerer and Schröder, 1983].

[Emch, 1998a] G. G. Emch. *On the need to adapt de Finetti's probabilistic interpretation to QM*, pages 157-166 in [Alicki et al., 1998].

[Emch, 1998b] G. G. Emch. *Chaotic dynamics in non-commutative geometry*, pages 193-205 in [Strasburger et al., 1998].

[Emch, 2003] G. G. Emch. *Quantum measurement: a bridge to classical physics*, volume 5105, pages 255-264 in [Donkor et al., 2003].

[Emch, 2005] G. G. Emch. Probabilistic issues in statistical mechanics. *Studies in History and Philosophy of Modern Physics*, 36: 303-322, 2005. See already [Emch, 1998a].

[Emch and Guenin, 1966] G. G. Emch and M. Guenin. Gauge invariant formulation of the BCS model. *Journal of Mathematical Physics*, 7: 915-921, 1966.

[Emch *et al.*, 1970] G. G. Emch, H. J. F. Knops and E. J. Verboven. Breaking of the Euclidean symmetry with an application to the theory of crystallization. *Journal of Mathematical Physics*, 11: 1655-1668, 1970.

[Emch and Knops, 1970] G. G. Emch and H. J. F. Knops. Pure thermodynamical phases as extremal KMS states. *Journal of Mathematical Physics*, 11: 3008-3018, 1970.

[Emch and Radin, 1971] G. G. Emch and C. Radin. Relaxation of local thermal deviation from equilibrium. *Journal of Mathematical Physics*, 12: 2043-2046, 1971.

[Emch *et al.*, 1994a] G. G. Emch, H. Narnhofer, W. Thirring and G. L. Sewell. Anosov actions on non-commutative algebras. *Journal of Mathematical Physics*, 35: 5582-5599, 1994.

[Emch *et al.*, 1994b] G. G. Emch, G. C. Hegerfeldt and L. Streit, editors. *On Klauder's path: a field trip.* World Scientific, Singapore, 1994.

[Emch and Liu, 2002] G. G. Emch and C. Liu. *The logic of thermostatistical physics.* Springer, Heidelberg, 2002.

[Erdélyi, 1953] A. Erdélyi. *Higher transcendental functions.* McGrawHill, New York, 1953.

[Fannes and Tuyls, 1999] M. Fannes and P. Tuyls. A continuity property of quantum dynamical entropy. *Infinite dimensional analysis, quantum probability and related topics.* 2: 511-527, 1999.

[Fermi, 1926] E. Fermi. Zur Quantelung des idealen einatomigen Gases. *Zeitschrift für Physik*, 36: 902-912, 1926.

[Fermi, 1927] E. Fermi. Uno modelo statistico per la determinazione dálcune proprietá dell' atomo. *Atti della Accademia Nazionale dei Lincei*, *Rendiconti*, 6: 602-607, 1927. See also [Fermi, 1928].

[Fermi, 1928] E. Fermi. Eine statistische Methode zur Bestimmung einiger Eigenschaften des Atomes und ihre Anwendungen auf die Theorie des periodischen Systems der Elemente. *Zeitschrift für Physik*, 48: 73-79, 1928.

[Feynman, 1998] R. P. Feynman. Statistical mechanics, *a set of lectures.* Addison Wesley, Reading MA, 1998.

[Fokas *et al.*, 2001] A. Fokas, A. Grigoryan, T. Kibble and B. Zegarlinski, editors. *Mathematical Physics 2000.* International Press of Boston, Somerville MA, 2001.

[Ford *et al.*, 1965] G. W. Ford, M. Kac and P. Mazur. Statistical mechanics of assemblies of coupled oscillators. *Journal of Mathematical Physics*, 6: 504-515, 1965.

[Fowler, 1926] R. H. Fowler. On dense matter. *Monthly Notices Royal Astronomical Society*, 87: 114, 1926.

[Fredenhagen, 2003] K. Fredenhagen. *Locally covariant quantum theory*, *in Proceedings of the International Conference on Mathematical Physics*, Lisbon, 2003.

[Fredenhagen and Haag, 1990] K. Fredenhagen and R. Haag. On the derivation of Hawking radiation associated with the formation of a black hole. *Communications in Mathematical Physics*, 127: 273-284, 1990.

[Friedrichs, 1953] K. O. Friedrichs. *Mathematical aspects of the quantum theory of fields.* Wiley, New York, 1953.

[Gamov, 1928] G. Gamow. Zur Quantentheorie des Atomkernes. *Zeitschrift für Physik*, 51: 204-212, 1928.

[Gamov, 1940] G. Gamow. *Mr. Tompkins in Wonderland.* Macmillan, New York, 1940.

[Garcia-Bondia *et al.*, 2003] J. M. Garcia-Bondia, J. Varilly and H. Figueroa. *Elements of noncommutative geometry.* Birkhäuser, Boston, 2001.

[Gibbs, 1902] J. W. Gibbs. *Elementary principles of statistical mechanics, developed with special reference to the rational foundation of thermodynamics.* Scribner, New York, 1902.

[Glimm *et al.*, 1990] J. Glimm, J. Impagliazzo and I. Singer, editors. The legacy of John von Neumann. *Proceedings of Symposia in Pure Mathematics*, volume 50, 1990.

[Gorini *et al.*, 1978] V. Gorini, A. Frigerio, M. Verri, A. Kossakowski and E. C. G. Sudarshan. Properties of quantum Markovian master equations. *Reports on Mathematical Physics*, 13: 149-173, 1978.

[Gouy, 1888] L. Gouy. Note sur le mouvement brownien. *Journal de Physique* (2), 7: 561-564, 1888.

[Grad, 1958] H. Grad. *Principles of the kinematic theory of gases*, volume 12 of Handbuch der Physik. Springer, Berlin, 1958. See also [Grad, 1963].

[Grad, 1963] H. Grad. Asymptotic theory of the Boltzmann equation. *Physics of Fluids*, 6: 147-181, 1963.

[Grad, 1967] H. Grad. *Levels of description in statistical mechanics and thermodynamics*, pages 49-76 in [Bunge, 1967].

[Greenleaf, 1969] F. P. Greenleaf. *Invariant mean on topological groups.* van Nostrand-Reinhold, New York, 1969.

[Gurney and Condon, 1928] R. W. Gurney and E. U. Condon. Quantum mechanics and radioactive disintegration. *Nature*, 122: 439, 1928. See also [Gurney and Condon, 1929].

[Gurney and Condon, 1929] R. W. Gurney and E. U. Condon. Quantum mechanics and radioactive disintegration. *Physical Review*, 33: 127-140, 129. See also [Gamov, 1928].

[Gutzwiller, 1990] M. C. Gutzwiller. *Chaos in classical and quantum mechanics.* Springer, New York, 1990.

[Haag, 1955] R. Haag. On quantum field theory. *Matematisk-Fysiske Meddelelser Danske Videnskabernes Selskab*, 29(12), 1955.

[Haag, 1959a] R. Haag. *Discussion des "axiomes" et des propriétés asymptotiques d'une théorie des champs locale avec particules composées*, *in Les problémes mathématiques de la théorie quantique des champs*, Colloque CNRS, Lille, 1958; CNRS, Paris, 1959.

[Haag, 1959b] R. Haag. The framework of quantum field theory. *Il Nuovo Cimento*, 14: 131-152, 1959.

[Haag, 1962] R. Haag. The mathematical structure of the Bardeen-Cooper-Schrieffer model. *Il Nuovo Cimento*, 25: 287-299, 1962.

[Haag and Schroer, 1962] R. Haag and B. Schroer. Postulates of quantum field theory. *Journal of Mathematical Physics*, 3: 248-256, 1962.

[Haag and Kastler, 1964] R. Haag and D. Kastler. An algebraic approach to quantum field theory. *Journal of Mathematical Physics*, 5: 848-861, 1964.

[Haag *et al.*, 1967] R. Haag, N. Hugenholtz and M. Winnink. On the equilibrium states in quantum statistical mechanics. *Communications in Mathematical Physics*, 5: 215-236, 1967.

[Haag *et al.*, 1974] R. Haag, D. Kastler and E. B. Trych-Pohlmeyer. Stability and equilibrium states. *Communications in Mathematical Physics*, 38: 173-193, 1974. See [Haag and Trych-Pohlmeyer, 1977].

[Haag and Trych-Pohlmeyer, 1977] R. Haag and E. B. Trych-Pohlmeyer. Stability properties of equilibrium states. *Communications in Mathematical Physics*, 56: 213-224, 1977.

[Haag *et al.*, 1994] R. Haag, H. Narnhofer and U. Stein. On quantum field theory in gravitational background. *Communications in Mathematical Physics*, 94: 219-238, 1994.

[Halvorson, 2006] H. Halvorson. This volume, chapter 8, 2006.

[Hawking, 1975] S. W. Hawking. Particle creation by black holes. *Communications in Mathe-matical Physics*, 43: 199-220, 1975.

[Hawking, 1981] R. W. Hawking. Interacting quantum fields around a black hole. *Communications in Mathematical Physics*, 80: 421-442, 1981.

[Hawking and Page, 1983] S. W. Hawking and D. N. Page. Thermodynamics of black holes in anti-de Sitter space. *Communications in Mathematical Physics*, 87: 577-588, 1983.

[Heisenberg, 1928] W. Heisenberg. Zur Theorie der Ferromagnetismus. *Zeitschrift für Physik*, 49: 419-436, 1928.

[Hepp, 1972] K. Hepp. Quantum theory of measurement and macroscopic observables. *Helvetica Physica Acta*, 45: 237-248, 1972.

[Hida, 1970] T. Hida. *Stationary stochastic processes.* Princeton University Press, Princeton NJ, 1970.

[Hilbert, 1899] D. Hilbert. *Grundlagen der Geometrie.* Teubner, Leipzig, 1899; translated by L. Unger: *Foundations of Geometry.* Open Court, La Salle IL, 1971.

[Hilbert, 1900] D. Hilbert. *Sur les problèmes futurs des mathématiques, pages* 58-114 *in Comptes rendus du deuxième congrès international des mathématiciens, Paris*, 1900, Gauthier-Villars, Paris, 1902. See also [Browder, 1976].

[Hilbert, 1918] D. Hilbert. Axiomatisches Denken. *Mathematische Annalen*, 78: 405-415, 1918.

[Ho *et al.*, 1993] T. G. Ho, L. J. Landau and A. J. Wilkins. On the weak coupling limit for a Fermi gas in a random potential. *Reviews in Mathematical Physics*, 5: 209-298, 1993.

[Huang, 1965] K. Huang. *Statistical mechanics.* Wiley, New York, 1965.

[Hudson, 1998] R. L. Hudson. *Quantum stochastic analysis after four decades*, pages 194-217 in *Probability towards* 2000, Springer Lecture Notes in Statistics 128, New York, 1998.

[Ingarden *et al.*, 1999] R. S. Ingarden, R. Smolukowski, S. Chandrasekhar and M. Kac. *Marian Smolukowski: his life and scientific work.* Polish Scientific Publishers PWN, Warsaw, 1999.

[Itô, 1987] K. Itô, editor. *Encyclopedic dictionary of mathematics.* The MIT Press, Cambridge MA and London England, 1987; 2nd edition of the English translation of the 3rd Japanese edition.

[Jammer, 1966] M. Jammer. *The conceptual development of quantum mechanics.* McGraw-Hill, New York, 1966.

[Jammer, 1974] M. Jammer. *The philosophy of quantum mechanics.* Wiley, New York, 1974.

[Jaynes, 1967] E. T. Jaynes. *Foundations of probability theory and statistical mechanics*, pages 77-101 in [Bunge, 1967]. See also [Jaynes, 1989].

[Jaynes, 1989] *Papers on probability, statistics and statistical mechanics.* Kluwer, Dordrecht, 1989. 1st ed. 1983.

[Jeans, 1905a] J. B. Jeans. On the partition of energy between matter and ether. *Philosophical Magazine*, 10: 91-98, 1905. See also [Jeans, 1905b].

[Jeans, 1905b] J. B. Jeans. Comparison between two theories of radiation. *Nature*, 72: 293, 1905.

[Josephson, 1982] B. D. Josephson. *The discovery of tunnelling supercurrents*, pages 157-164 in [Lundqvist, 1982].

[Kac, 1968] M. Kac. *Mathematical mechanisms of phase transitions*, pages in 241-305 in volume i of [Chrétien et al., 1968].

[Kadison, 1990] R. V. Kadison. *Operator algebras—an overview*, pages 61-89 in [Glimm et al., 1990].

[Kadison and Ringrose, 1983/1986] R. V. Kadison and J. R. Ringrose. *Fundamentals in the* *theory of operator algebras*, i. *Elementary theory*; *ii. Advanced theory*. Wiley, New York, 1983/1986.

[Kastler, 1976] D. Kastler. Equilibrium states of matter and operator algebras. *Istituto Nazionale di Alta Matematica*, *Symposia Mathematica*, 20: 49-107, 1976. See also [Haag et al., 1974] and [Narnhofer and Thirring, 1982].

[Kastler and Robinson, 1966] D. Kastler and D. W. Robinson. Invariant states in statistical mechanics. *Communications in Mathematical Physics*, 3: 151-180, 1966. See also [Robinson and Ruelle, 1967].

[Kasler *et al.*, 1967] D. Kastler, R. Haag and L. Michel. Central decomposition of ergodic states. *Séminaire de physique théorique*, *Marseille*, 1967. See also [Kastler et al., 1972], [Kastler, 1976], [Kastler and Robinson, 1966], [Robinson and Ruelle, 1967].

[Kastler *et al.*, 1972] D. Kastler, M. Mebkhout, G. Loupias and L. Michel. Central decomposition of invariant states with applications to the group of time translations and of Euclidean transformations in algebraic field theory. *Communications in Mathematical Physics*, 27: 195-222, 1972.

[Kay, 1993] B. S. Kay. Sufficient conditions for quasifree states on spacetimes with a bifurcate Killing horizon. *Journal of Mathematical Physics*, 34: 4519-4539, 1993.

[Kay and Wald, 1991] B. S. Kay and R. M. Wald. Theorems on the uniqueness and thermal properties of stationary nonsingular, quasifree states on spacetimes with a bifurcate Killing horizon. *Physics Reports*, 207: 49-136, 1991. See also [Kay, 1993].

[Khinchin, 1957] A. Khinchin. *Mathematical foundations of information theory*. Dover, New York, 1957.

[Khrenninkov, 2005] A. Khrenninkov, editor. *Foundations of probability and physics*. American Institute of Physics, New York, 2005.

[Kishimoto, 1976] A. Kishimoto. On uniqueness of KMS states of one-dimensional quantum lattice systems. *Communications in Mathematical Physics*, 47: 167-170, 1976.

[Kolmogorov, 1933] A. N. Kolmogorov. Grundbegriffe der Wahrscheinlichkeitsrechnung. Springer, Berlin, 1933; translated by N. Morrison: *Foundations of the theory of probability*. Chelsea, New York, 1956.

[Kossakowski *et al.*, 1977] A. Kossakowski, A. Frigerio, V. Gorini and M. Verri. Quantum detailed balance and the KMS condition. *Communications in Mathematical Physics*, 57: 97-110, 1977; Erratum [Kossakowski et al., 1978].

[Kossakowski *et al.*, 1978] A. Kossakowski, A. Frigerio, V. Gorini and M. Verri. Erratum: "Quantum detailed balance and the KMS condition". *Communications in Mathematical Physics*, 60: 96, 1978.

[Kramers and Wannier, 1941] H. A. Kramers and G. H. Wannier. Statistics of the two-dimensional ferromagnet, i, ii. *Physical Review*, 60: 252-262, 263-276, 1941.

[Kubo, 1957] R. Kubo. Statistical mechanics of irreversible processes. *Journal of the Physical Society of Japan*, 12: 570-586, 1957.

[Kümmerer and Schröder, 1983] B. Kümmerer and W. Schröder. A Markov dilation of a non-quasifree Bloch evolution. *Communications in Mathematical Physics*, 90: 251-262, 1983.

[Landau, 1932] L. D. Landau. On the theory of stars. *Physikalische Zeitschrift der Sowjetunion*, 1: 285-288, 1932.

[Landau and Lifshitz, 1958a] L. D. Landau and E. M. Lifshitz, *Quantum mechanics.* Pergamon, London, 1958.

[Landau and Lifshitz, 1958b] L. D. Landau and E. M. Lifshitz, *Statistical physics.* Pergamon, London 1958.

[Landsman, 2006] N. Landsman. This volume, Chapter 5, 2006.

[Le Bris and Lions, 2005] C. Le Bris and P.-L. Lions. From atoms to crystals: a mathematical journey. *Bulletin of the American Mathematical Society*, 42: 291-363, 2005.

[Lenard, 1970] A. Lenard, editor. *Statistical mechanics and mathematical problems*, volume 20 of *Lecture Notes in Physics.* Springer, New York, 1970.

[Lieb, 1982a] E. H. Lieb. Thomas-Fermi and related theories of atoms and molecules. *Reviews of Modern Physics*, 53: 603-641, 1982; reprinted in [Lieb, 1997] pp. 259-297. See also [Lieb, 1982b].

[Lieb, 1982b] E. H. Lieb. Erratum: "Thomas-Fermi and related theories of atoms and molecules". *Reviews of Modern Physics*, 54: 311, 1982; reprinted in [Lieb, 1997] p. 298.

[Lieb, 1990] E. H. Lieb. The stability of matter: from atoms to stars. *Bulletin of the American Physical Society*, 22: 1-49, 1990; reprinted in [Lieb, 1997] pp. 11-59.

[Lieb, 1997] E. H. Lieb. *Selecta: The stability of matter, from atoms to stars.* Springer, Berlin, Heidelberg, 1997.

[Lieb, 2001] E. H. Lieb. *The Bose-gas: a subtle many-body problem*, pages 91-111 in [Fokas et al., 2001].

[Lieb and Simon, 1977] E. H. Lieb and B. Simon. Thomas-Fermi theory of atoms, molecules and solids. *Advances in Mathematics*, 23: 22-116, 1977.

[Liu and Emch, 2005] C. Liu and G. G. Emch. explaining quantum spontaneous symmetry breaking. *Studies in History and Philosophy of Modern Physics*, 36: 137-163, 2005.

[Lowe and Nordberg, 1957] I. J. Lowe and R. E. Nordberg. Free-induction decay in solids. *Physical Review*, 107: 46-61, 1957.

[Lummer and Pringsheim, 1900] O. Lummer and E. Pringsheim. Ueber die Strahlung der schwarzen Körpers für lange Wellen. *Verhandlungen der Deutschen Physikalischen Gesellschaft*, 2: 163-180, 1900.

[Lundqvist, 1982] S. Lundqvist, editor. *Nobel lectures 1971-1980.* World Scientific, Singapore 1982.

[Martin and Schwinger, 1959] P. C. Martin and J. Schwinger. Theory of many-particle systems. *Physical Review*, 115: 1342-1373, 1959.

[Martin, 1979] Ph. A. Martin. *Modèles en mécanique des processus irréversibles*, volune 103 of *Lecture Notes in Physics.* Springer, Berlin, 1979. See also [Martin and Emch, 1975] and [Spohn, 1977].

[Martin and Emch, 1975] Ph. A. Martin and G. G. Emch. A rigourous model sustaining van Hove's phenomenon. *Helvetica Physica Acta*, 48: 48-59, 1975. See also [Ho et al., 1993].

[Messiah, 1960] A. Messiah. *Mécanique quantique.* Dunod, Paris, 1960.

[Michel and Ruelle, 1970] L. Michel and D. Ruelle, editors. *Systèmesàun nombre infini de degrés de liberté*, CNRS, Paris, 1970.

[Murray, 1990] F. J. Murray. *The rings of operators papers, pages* 57-60 in [Glimm et al., 1990].

[Murray and von Neumann, 1936] F. J. Murray and J. von Neumann. Rings of Operators. *Annals of Mathematics*, 37: 116-229, 1936.

[Murray and von Neumann, 1937] F. J. Murray and J. von Neumann. Rings of Operators ii. *Transactions of the American Mathematical Society*, 41: 208-248, 1937.

[Narnhofer, 1998] H. Narnhofer. *Control of weak asymptotic abelianness with the help of the cross-product construction*, pages 331-339 in [Alicki et al., 1998].

[Narnhofer, 2001] H. Narnhofer. Kolmogorov systems and Anosov systems in quantum theory. *Infinite Dimensional Analysis, Quantum Probability and Related Topics*, 4: 85-119, 2001.

[Narnhofer, 2005] H. Narnhofer. Roots for quantum ergodic theory in von Neumann's work. *Reports on Mathematical Physics*, 55: 93-107, 2005.

[Narnhofer and Thirring, 1982] H. Narnhofer and W. Thirring. Adiabatic theorem in quantum statistical mechanics. *Physical Review A*, 26: 3646-3652, 1982. See also [Thirring, 1983a].

[Narnhofer and Thirring, 1989] H. Narnhofer and W. Thirring. Quantum K-systems. *Communications in Mathematical Physics*, 125: 565-577, 1989.

[Narnhofer and Thirring, 1994a] H. Narnhofer and W. Thirring. Clustering for algebraic K-systems. *Letters in Mathematical Physics*, 30: 307-316, 1994.

[Narnhofer and Thirring, 1994b] H. Narnhofer and W. Thirring. *Dynamical entropy of quantum systems and their Abelian counterpart*, pages 127-145 in [Emch et al., 1994b].

[Ojima, 2003] I. Ojima. *How to formulate non-equilibrium local states in QFT-general characterization and extension to curved space-time*, pages 365-384 in [Arafune et al., 2003].

[Parthasarathy, 1995] K. R. Parthasarathy. *Quantum stochastic calculus*, pages 1024-1035 in *Proceedings, International Congress of Mathematicians, Zurich*, 1994, Birkhäuser, Basel, 1995.

[Paschen and Wanner, 1899] F. Paschen and H. Wanner. Eine photometrische Methode zur Bestimmung der exponentialconstanten der Emissionsfunktion. *Berliner Berichte*, 1899: 5-11.

[Peierls, 1936] R. Peierls. On Ising's model of ferromagnetism. *Proceedings of the Cambridge Philosophical Society*, 32: 477-481, 1932.

[Peierls, 1979] R. Peierls. *Surprises in theoretical physics.* Princeton University Press, Princeton NJ, 1979.

[Pitaevskii and Stringari, 2003] L. Pitaevskii and S. Stringari. *Bose-Einstein condensation.* Oxford University Press, Oxford, 2003.

[Planck, 1900a] M. Planck. Ueber eine Verbesserung der Wienschen Spektralgleichung. *Verhandlungen der Deutschen Physikalishen Gesellschaft*, 2: 202-204, 1900.

[Planck, 1900b] M. Planck. Zur Theorie des Gesetzes der Energieverteilung im Normalspektrum. *Verhandlungen der Deutschen Physikalishen Gesellschaft*, 2: 237-245, 1900; *ibid. Annalen der Physik*, 4: 553-563, 1901.

[Planck, 1911] M. Planck. Deux récentes théories thermodynamiques: théorème de Nernst et hypothèse des quanta (Conférence prononcée le 16 décembre 1911 à la Société chimique allemande de Berlin); appended to [Planck, 1913].

[Planck, 1913] M. Planck. *Leçons de thermodynamique*, 3rd ed. , Hermann, Paris 1913.

[Prigogine, 1997] I. Prigogine. *The end of certainty: time, chaos and the new laws of nature.* The Free Press, New York, 1997.

[Pusz and Woronowicz, 1978] W. Pusz and S. L. Woronowicz. Passive states and KMS states for general quantum systems. *Communications in Mathematical Physics*, 58: 273-290, 1978.

[Rauch, 2005] H. Rauch. *Quantum phenomena tested by neutron interferometry*, pages 331-339 in [Khrenninkov, 2005].

[Rauch and Werner, 2000] H. Rauch and S. A. Werner. *Neutron interferometry*, Oxford Scientific Publishers, Oxford, 2000.

[Rayleigh, 1900] J. W. S. Rayleigh. Remarks upon the law of complete radiation. *Philosophical Magazine*, 49: 539-540, 1900.

[Rayleigh, 1902] J. W. S. Rayleigh. The constant of radiation as calculated from molecular data. *Nature*, 72: 243, 1902.

[Rindler, 1966] W. Rindler. Kruskal space and the uniformly accelerated frame. *American Journal of Physics*, 34: 1174-1178, 1966.

[Robinson, 1965] D. W. Robinson. A theorem concerning positive metric. *Communications in Mathematical Physics*, 1: 89-94, 1965.

[Robinson, 1971] D. W. Robinson. *The thermodynamical pressure in quantum statistical mechanics*, volume 9 of Lecture Notes in Physics. Springer, Berlin, 1971.

[Robinson, 1973] D. W. Robinson. Return to equilibrium. *Communications in Mathematical Physics*, 31: 171-189, 1973.

[Robinson and Ruelle, 1967] D. W. Robinson and D. Ruelle. Extremal invariant states. *Annales de l'Institut Henri Poincaré*, 6: 299-310, 1967. See already [Kastler and Robinson, 1966].

[Ruelle, 1968a] D. Ruelle. *Statistical mechanics.* Benjamin, New York, 1968.

[Ruelle, 1968b] D. Ruelle. Statistical mechanics of a one-dimensional lattice gas. *Communications in Mathematical Physics*, 9: 267-278, 1968.

[Sakai, 1971] S. Sakai. *C^*-algebras and W^*-algebras.* Springer, New York, 1971.

[Schiff, 1955] L. I. Schiff. *Quantum mechanics*, 2nd ed., MacGraw Hill, London, 1955.

[Schrieffer, 1974] J. R. Schrieffer. *Theory of superconductivity.* Benjamin, New York, 1974.

[Schroer and Wiesbrock, 2000] B. Schroer and H. W. Wiesbrock. Modular theory and geometry. *Reviews in Mathematical Physics*, 12: 301-326, 2000.

[Segal, 1990] A. F. Segal. *Paul the convert: the apostolate and apostasy of Saul the Pharisee.* Yale University Press, New Haven, 1990.

[Segal, 1947] I. E. Segal. Postulates for general quantum mechanics. *Annals of Mathematics*, 48: 930-948, 1947.

[Segal, 1958] I. E. Segal. Distributions in Hilbert space and canonical systems of operators. *Transactions of the American Mathematical Society*, 88: 12-41, 1958.

[Segal, 1959] I. E. Segal. Foundations of the theory of dynamical systems of infinitely many degrees of freedom, i. *Matematisk-Fysiske Meddelelser Danske Videnskabernes Selskab*, 31: 2, 1959.

[Segal, 1961] I. E. Segal. Foundations of the theory of dynamical systems of infinitely many degrees of freedom, ii. *Canadian Journal of Mathematics*, 13: 1-18, 1961.

[Segal, 1962] I. E. Segal. Foundations of the theory of dynamical systems of infinitely many degrees of freedom, iii. *Journal of Mathematics*, 6: 500-523, 1962.

[Segal, 1963] I. E. Segal. *Mathematical problems of relativistic physics.* Lecture Notes in Applied Mathematics. American Mathematical Society, Providence RI, 1963.

[Sewell, 1973] G. L. Sewell. States and dynamics of infinitely extended physical systems. *Communications in Mathematical Physics*, 33: 42-51, 1973.

[Sewell, 1977] G. L. Sewell. KMS conditions and local thermodynamical stability of quantum lattice systems, ii. *Communications in Mathematical Physics*, 55: 53-61, 1977.

[Sewell, 1980a] G. L. Sewell. Relativity of temperature and the Hawking effect. *Physics*

Letters A, 79: 23-24, 1980. See also [Sewell, 1982a].

[Sewell, 1980b] G. L. Sewell. Stability, equilibrium and metastability in statistical mechanics. *Physics Reports*, 57: 307-342, 1980.

[Sewell, 1982a] G. L. Sewell. Quantum fields on manifolds: PCT and gravitationally induced thermal states. *Annals of Physics*, 141: 201-224, 1982.

[Sewell, 1982b] G. L. Sewell. W^*-dynamics of infinite quantum systems. *Letters in Mathematical Physics*, 6: 209-213, 1982. See already [Dubin and Sewell, 1970] and [Sewell, 1973].

[Sewell, 2002] G. L. Sewell. *Quantum mechanics and its emergent macrophysics.* Princeton University Press, Princeton and Oxford, 2002.

[Sewell, 2005] G. L. Sewell. On the mathematical structure of quantum measurement theory. *Reports on Mathematical Physics*, 56: 271-290, 2005.

[Shapiro and Teukolsky, 1983] S. L. Shapiro and S. A. Teukolsky. *Black holes, white dwarfs, and neutron stars: the physics of compact objects.* Wiley, New York, 1983.

[Smolukowski, 1906a] M. Smolukowski. Sur le chemin moyen parcouru par les molécules d'un gas et sur son rapport avec la théorie de la diffusion. *Bulletin International de l'Académie des Sciences de Cracovie*, 1906: 202-213; reprinted with translation, pages 29-42 in [Ingarden et al., 1999]. See also [Smolukowski, 1906b].

[Smolukowski, 1906b] M. Smolukowski. Zur kinetischen Theorie der Brownschen Molekular Bewegung und der Suspensionen. *Annalen der Physik*, 21: 755-780, 1906.

[Smolukowski, 1916] M. Smolukowski. Drei Vorträgeüber Diffusion, Brownsche Molekularbewegung und Koagulation von Kolloidteichen. *Physikalische Zeitschrift*, 17: 557-571, 585-599, 1916; reprinted with translation, pages 43-127 in [Ingarden *et al.*, 1999].

[Spohn, 1977] H. Spohn. Derivation of the transport equations for electrons moving through random impurities. *Journal of Statistical Physics*, 17: 385-412, 1977.

[Stachel, 1998] J. Stachel, editor. *Einstein's miraculous year.* Princeton University Press, Princeton NJ, 1998.

[Stefan, 1879] J. Stefan. Ueber die Beziehung zwischen der Wärmestrahlung und der Temperatur. *Wiener Berichte*, 79: 391-428, 1879.

[Stømer, 1969] E. Stømer. Symmetric states on infinite tensor products of C^*-algebras.

Journal of Functional Analysis, 3: 48-68, 1969.

[Strasburger *et al.*, 1998] A. Strasburger, S. T. Ali, J. -P. Antoine, J. -P. Gazeau and A. Odzijewicz, editors. *Quantization, coherent states and Poisson structures. Proceedings of the* 1995 *Bialowieza Workshop.* Polish Scientific Publications PWN, Warszawa, 1998.

[Streater, 1967] R. F. Streater. The Heisenberg ferromagnet as a quantum field theory. *Communications in Mathematical Physics*, 6: 233-247, 1967.

[Streater and Wightman, 1964] R. F Streater and A. S. Wightman. PCT, *spin & statistics, and all that. Benjamin*, New York, 1964.

[Summers and Verch, 1996] S. J. Summers and R. Verch. Modular inclusion, the Hawking temperature and quantum field theory in curved space-time. *Letters in Mathematical Physics*, 37: 145-158, 1996.

[Summers and White, 2003] S. J. Summers and R. K. White. On deriving space-time from quantum observables and states. *Communications in Mathematical Physics*, 237: 203-220, 2003.

[Szasz, 1996] D. Szasz. Boltzmann's ergodic hypothesis, a conjecture for centuries. *Studia scientiarum mathematicarum Hungarica*, 31: 299-322, 1996.

[Takeda, 1955] Z. Takeda. Inductive limit and infinite direct product of operator algebras. *Tohoku Mathematical Journal*, 7: 68-86, 1955.

[Takesaki, 1970a] M. Takesaki. *Tomita's theory of modular algebras and its applications*, volume 128 of Lecture Notes in Mathematics. Springer, New York, 1970.

[Takesaki, 1970b] M. Takesaki. *States and automorphisms of operator algebras-standard representations and KMS boundary condition*, pages 205-246 in [Lenard, 1970].

[Takesaki, 1970c] M. Takesaki. Disjointness of the KMS states of different temperatures. *Communications in Mathematical Physics*, 17: 33-41, 1970.

[Talman, 1968] J. C. Talman. *Special functions: a group theoretic approach; based on lectures by Eugene P.* Wigner. Benjamin, New York, 1968.

[Taylor, 1909] G. I. Taylor. Interference fringes in feeble light. *Proceedings of the Cambridge Philosophical Society, Mathematical and Physical Sciences*, 15: 114-115, 1909.

[Thirring, 1968] W. Thirring. On the mathematical structure of the BCS-Model, ii. *Communications in Mathematical Physics*, 7: 181-189, 1968. See already [Thirring and

Wehrl, 1967].

[Thirring, 1983a] W. Thirring. On the equivalence of adiabatic invariance and KMS condition. *Foundations of Physics*, 13: 409-415, 1983.

[Thirring, 1983b] W. Thirring. *Quantum mechanics of large systems.* Springer, New York, Wien, 1983.

[Thirring and Wehrl, 1967] W. Thirring and A. Wehrl. On the mathematical structure of the BCS-Model. *Communications in Mathematical Physics*, 4: 303-314, 1967. See also [Thirring, 1968].

[Thomas, 1927] L. H. Thomas. The calculation of atomic fields. *Proceedings of the Cambridge Philosophical Society*, 23: 542-548, 1927.

[Thorne, 1994] K. S. Thorne. *Black holes & time warps, Einstein's outrageous legacy*, Norton, New York, 1994.

[Tomita, 1967] M. Tomita. *Standard forms of von Neumann algebras.* 5th Functional Analysis Symposium, Mathematical Society of Japan, Sendai, 1967.

[Truesdell, 1945] C. Truesdell. A function which occurs in the theory of the structure of polymers. *Annals of Mathematics*, 46: 144-157, 1945.

[Tuyls, 1998] P. Tuyls. *Comparing quantum dynamical entropies*, pages 411-420 in [Alicki et al., 1998]. See already [Alicki and Fannes, 1994], and also [Fannes and Tuyls, 1999].

[Uffink, 1995] J. Uffink. Can the minimum entropy principle be regarded as a consistency requirement? *Studies in History and Philosophy of Modern Physics*, 26: 223-261, 1995.

[Uffink, 2006] J. Uffink. This volume, Chapter 9, 2006.

[Uhlenbeck, 1967] G. E. Uhlenbeck. *Discusssion comment*, page 56 in [Bak, 1967].

[Uhlenbeck, 1968] G. E. Uhlenbeck. *An outline of statistical mechanics*, pages 1-29 in volume ii of [Cohen, 1962/1968].

[Unruh, 1976] W. G. Unruh. Notes on black-hole evaporation. *Physical Review D*, 14: 870-892, 1976.

[Unruh and Wald, 1984] W. G. Unruh and R. M. Wald. What happens when an accelarating observer detects a Rindler particle? *Physical Review D*, 29: 1047-1056, 1984.

[van Hove, 1952] L. van Hove. Les difficultés de divergences sur un modèle particulier de champ quantifié. *Physica*, 18: 145-159, 1952.

[van Hove, 1955] L. van Hove. Quantum mechanical perturbations giving rise to a statistical transport equation. *Physica*, 21: 517-540, 1955.

[van Hove, 1957] L. van Hove. The approach to equilibrium in quantum statistics. *Physica*, 23: 441-480, 1957.

[van Hove, 1959] L. van Hove. The ergodic behaviour of quantum many-body systems. *Physica*, 25: 268-276, 1959.

[van Hove, 1961] L. van Hove. *The problem of master equations*, pages 155-169 in [Caldirola, 1961].

[van Hove, 1962] L. van Hove. *Master equation and the approach to equilibrium for quantum systems*, pages 157-172 in volume i of [Cohen, 1962/1968].

[van Kampen, 1954] N. G. van Kampen. Quantum statistics of irreversible processes. *Physica*, 20: 603-622, 1954. See also [van Kampen, 1988].

[van Kampen, 1962] N. G. van Kampen. *Fundamental problems in statistical mechanics of irreversible processes*, pages 173-202 in volume i of [Cohen, 1962/1968].

[van Kampen, 1967] N. G. van Kampen. *Discussion remark*, page 67 in [Bak, 1967].

[van Kampen, 1988] N. G. van Kampen. Ten theorems about quantum mechanical measurements. *Physica* A, 153: 97-113, 1988.

[Verboven, 1967] E. J. Verboven. *Quantum thermodynamics of an infinite system of harmonic oscillators*, pages 49-54 in [Bak, 1967].

[Vilenkin, 1968] N. Vilenkin. *Special functions and the theory of group representations*, volume 22 of American Mathematical Society Translations, Providence RI, 1968.

[von Mises, 1928] R. von Mises. *Wahrscheinlichkeit, Statistik und ihre Wahrheit*, Springer, Wien, 1928. See also [von Mises, 1957].

[von Mises, 1957] R. von Mises. *Probability, statistics and truth*, McMillan, New York, 1957; prepared by H. Geringer from the 3rd German ed. (1951).

[von Neumann, 1932a] J. von Neumann. Proof of the quasi-ergodic theorem. *Proceedings of the National Academy of Sciences* (*USA*), 18: 70-83, 1932. See also [von Neumann, 1932b].

[von Neumann, 1932b] J. von Neumann. Physical applications of the ergodic hypothesis. *Proceedings of the National Academy of Sciences* (*USA*), 18: 263-266, 1932. See also

[Birkhoff, 1931].

[von Neumann, 1932c] J. von Neumann. *Mathematische Grundlagen der Quantenmechanik.* Springer, Berlin, 1932; translated by R. T. Beyer: *Mathematical foundations of quantum mechanics*, Princeton University Press, Princeton NJ, 1955; page references are to this edition.

[Wald, 1977] R. M. Wald. *Space, time, and gravity: the theory of big bang and black holes.* University of Chicago Press, Chicago, 1977.

[Wald, 1994] R. M. Wald. *Quantum field theory in curved spacetime and black hole thermodynamics.* University of Chicago Press, Chicago, 1994. For an elementary introduction, see [Wald, 1977].

[Wannier, 1966] G. H. Wannier. *Statistical physics.* John Wiley, New York, 1966.

[Weinberg, 1972] S. Weinberg. *Gravitation and cosmology, principles and applications of the general theory of relativity.* Wiley, New York, 1972.

[Weiss, 1907] P. Weiss. L'hypothèse du champ moléculaire et la propriété ferromagnétique. *Journal de Physique*, 6: 661-690, 1907.

[Whittaker and Watson, 1927] A. N. Whittaker and G. N. Watson, *A course in modern analysis*, 4th ed., Cambridge University Press, London, 1927.

[Whitten-Wolfe and Emch, 1976] B. Whitten-Wolfe and G. G. Emch. A mechanical quantum measuring process. *Helvetica Physica Acta*, 49: 45-55, 1976.

[Wick *et al.*, 1952] G. C. Wick, A. S. Wightman and E. P. Wigner. Intrinsic parity of elementary particles. *Physical Review*, 88: 101-105, 1952.

[Wien, 1894] W. Wien. Temperatur und Entropie der Strahlung. *Annalen der Physik*, 52: 132-165, 1894.

[Wien, 1896] W. Wien. Ueber die Energieverteilung im Emissionspektrum eines schwarzen Körpers. *Annalen der Physik*, 58: 662-669, 1896.

[Wiesbrock, 1997] H. W. Wiesbrock. Half-sided modular inclusions of von Neumann algebras. *Communications in Mathematical Physics*, 184: 583-685, 1997. See also [Schroer and Wiesbrock, 2000].

[Wightman and Schweber, 1955] A. S. Wightman and S. Schweber. Configuration space methods in relativistic quantum field theory. *Physical Review*, 98: 812-837, 1955.

[Wigner, 1931] E. P. Wigner. *Gruppentheorie und ihre Anwendung auf die Quantenmechanik* *der Atomspektren.* Vieweg, Braunschweig, 1931; translated by J. C. Griffin: *Group theory and its application to the quantum mechanics of atomic spectra.* Academic Press, New York, 1959.

[Wigner, 1962] E. P. Wigner. *Oral admonition to young researchers.* Private communication (1962), which I understand many others have had to face, before and since.

[Wigner, 1984] E. P. Wigner. *Review of the quantum-mechanical measurement problem*, pages 63-82 in *Science*, *computers and the information onslaught*, D. M. Kerr et al., editors, Academic Press, New York, 1984; reprinted in [Wigner, 1995] pp. 225-244.

[Wigner, 1995] E. P. Wigner. *Philosophical reflections and syntheses*, volume vi of *The collected works of Eugene Paul Wigner.* Springer, Heidelberg, 1995 (soft-cover, 1997). Annotated by G. G. Emch.

[Wigner, 1997] E. P. Wigner. *Foundations of quantum mechanics*, part ii of volume iii of *The collected works of Eugene Paul Wigner.* Springer, Heidelberg, 1997. Annotated by A. Shimony.

[Winnink, 1967] M. Winnink. *On the equilibrium states of infinite quantum systems at* $T\neq 0$, pages 59-66 in [Bak, 1967].

[Zeh, 1970] H. D. Zeh. On the interpretation of measurement in quantum theory. *Foundations of Physics*, 1: 69-76, 1970.

第十一章　宇宙哲学中的问题

乔治·F. R. 埃利斯

1. 引言

宇宙学是研究宇宙大尺度结构的一门学问，而“宇宙”这个词是指一切“真实的存在”[Harrison(哈里森)，2000]。这个定义是为了区别于“可见宇宙”，即可以通过天文观测而感知的宇宙，它仅仅是宇宙的一小部分。宇宙学的研究领域非常广泛，从星系、星系团、类星体等，到通过望远镜可以观测到的宇宙中的一切，研究它们的特性、分布、起源以及在广阔背景下它们之间的联系。观测宇宙学旨在确定可观测宇宙的大尺度几何，并通过观测遥远天体发出的射线来确定宇宙中物质的分布[Hoyle(霍伊尔)，1960；Kristian and Sachs(克里斯蒂安和萨克斯)，1966；Gunn *et al.*(冈恩等)，1978；Sandage *et al.*(桑达吉等)，1993；Bothun(博瑟)，1998]。物理宇宙学旨在研究宇宙早期的大爆炸膨胀中的相互作用[Peebles(皮布尔斯)，1971；Sciama(夏默)，1971；Weinberg(温伯格)，1972；Silk(希尔克)，2001；Perkins(珀金斯)，2005；Dodelson(都德尔逊)，2003]。天文宇宙学重在研究大爆炸之后比如星系和星系团等大尺度结构的演化发展[Sciama，1971；Peebles，1993b；Padmanbhan(帕

德马纳班)，1993；Rees(里斯)，1995；Dodelson，2003]。各种形式的量子宇宙学[Hawking(霍金)，1993；Gibbons *et al.*（吉本斯等)，2003；Copeland *et al.*（科普兰等)，2005]和宇宙粒子物理学方面的研究［Kolb and Turner(科尔布和特纳)，1990；Peacock(皮考克)，1999；Allday(奥得)，2002；Perkins，2005；Dodelson，2003]则试图描绘大爆炸之前的宇宙景象。这些研究领域以一种共生共长、互惠互补的方式共同建立了一个描绘真实宇宙起源和演化的理论[Bondi(邦迪)，1960；Harrison，2000；Silk，1997]。

宇宙的独特地位在于它产生了恒星、行星以及星系赖以出现和成长的环境，进而在物理和化学的作用下，行星上可以演化出生命，比如地球。假如宇宙环境截然不同，那么局部条件也会大相径庭，人类几乎不可能出现，事实上也不会出现任何生物[Carr and Rees(卡尔和里斯)，1979；Davies(戴维斯)，1982；Barrow and Tipler(拜罗和提普勒)，1984；Tegmark(泰格马克)，1998；Rees，1999]。因此，对于那些追求科学的人，尤其对于那些宇宙观测者和科学家来讲，宇宙学引起了人们对科学最本质的兴趣，它为其他学科设定了框架。作为最终的历史和地理类的科学，它是独一无二的。

宇宙学被当作一个严肃的科学研究分支始于爱因斯坦于1917年提出的静态宇宙理论。随后，哈勃(Hubble)于1929年观测到线性的宇宙射线红移图像，指出了宇宙的膨胀，以及在罗伯森—沃克(Robertson-Walker)几何下对非静态弗里德曼—勒梅特(Friedmann-Lemaitre)模型的几何和动力学理论理解的发展[North(诺斯)，1965；Berendzen *et al.*（勃兰登等)，1976；Smith(史密斯)，1982，Ellis(埃利斯)，1989；Kragh(克拉格)，1996]。在过去的几十年里，由于核物理和粒子物理被用来研究可观测宇宙的性质[Barnett *et al.*（巴奈特等)，1996；Nilsson *et al.*（尼尔森等)，1991]，宇宙学已经演化成为物理学的一个重要分支。同时，通过新的地面望远镜，比如凯克(Keck)望远镜、气球或卫星天文台，以及哈勃空间望远镜(Hubble Space telescope)(光和紫外线)、红外 infrared天文卫星(红外线)、伦琴卫星(ROSAT)(X射线)和宇宙背景探测器(COBE和WMAP)(微波)等，人们观测得到了大量全新的天文学数据[Gunn *et al.*，1978；Harwit(哈维特)，1984；Bothun，1998]，因此，宇宙学也已经成为天文学的一个重要分支，天文学已经从一个数学甚至是哲学领域的小科目演化

成了当今主流科学的一个重要组成部分，具有了基于大量数据并已被广泛接受的标准的学科体系[Weinberg，1972；Peebles *et al.*，1991；Silk，1997；Peacock，1999；Dodelson，2003]。然而，它又不同于任何其他的自然科学，它是独一无二的，其独特特征在推断、理论和观测之间的相互作用中很自然地得到了体现。

宇宙学和其他科学的主要区别在于其研究对象的独特性——整个宇宙[McCrea(麦克雷)，1953；McCrea，1960；Munitz(穆尼兹)，1962]——以及其作为其他物理和科学背景的角色，结果问题由于极早期宇宙的浩瀚尺度和巨大能量而变得突出了。我们根本无法复制宇宙形成的条件，也无法观测到极其遥远的天区和非常远古的时期。此外，我们也没有能力探测远古时期的物理关系。因此，宇宙学理论不可避免地具备哲学的性质(其他历史性的科学也是如此)，尤其是当它超越了纯粹的描述解释性作用时[Matravers *et al.*(马特拉斯等)，1995]。这正是宇宙学近几十年取得巨大进展的关键所在。如果我们以更多的野心追求理论，那么哲学的思辨将强烈影响我们对宇宙的理解。

在第二节对当今的宇宙学给出一个大体的概观后，这些问题就会在随后的章节里得到探讨。这些探讨会围绕宇宙学9个关键方面的性质，基于34个系列论点展开论述。广义地讲，这些论述关乎几何学、物理学和哲学，它们构成了宇宙学面临的哲学问题的环境及其与局域物理学之间的关系。我相信这种构想有助于我们关注这种关系中具体问题的重要性。对于那些认为宇宙学只不过是确定一些物理参数的人，这似乎太过于复杂了，但撰写本文的主要目的正是针对这种对宇宙学过于简单的看法。至于其他关于宇宙学哲学的论述，请参看[McCrea，1970；Munitz，1962；Ellis，1991；Leslie(莱斯利)，1994；Leslie，1998]。

2. 宇宙学概论

今天，宇宙学的一系列基本观点已经确立。经过数十年的努力，我们已经知道了星系之间的距离、宇宙的巨大尺度、宇宙正在膨胀和演化的基本特征。以前关于宇宙是静态的那些观点被认为是错误的[Ellis，1990]。宇宙学是基于“宇宙演化过程中物理定律总是一致的”这个假设的。在天文尺度上，万有引

力尽管很小，但是它是已知的唯一主导宇宙运行的力(其他已知的长程相互作用力是电磁力，但是宇宙中正负电荷相同，因此不会产生大规模的效果)。宇宙学的理论基础是经典相对论引力理论，即爱因斯坦的广义相对论[Malament(马拉蒙特)，2006]，它指出时空曲率和宇宙的演化由物质决定，它可以描述除了远古宇宙的一切。这个理论在任何时候都成立，它取决于物质和场，由其有效状态方程和相互作用来描述。

本节讲述宇宙学的基本理论、热大爆炸、宇宙学观测，包括宇宙背景辐射各向异性谱、因果关系和视界以及它们的影响、最新的理论发展(包括宇宙膨胀)、极早期的宇宙和目前的一致性模型，其中包括暗物质和暗能量。

2.1 基本理论

宇宙学基于下面的基本假设：时空的大尺度演变都可以通过把爱因斯坦的引力场方程到处运用来描述：全域演化遵循局部的物理学。宇宙学标准模型基于如下假设：[Robertson，1933；Ehlers(埃勒斯)，1993；Weinberg，1972；Hawking and Ellis，1973]如果可以在足够大的物理空间内进行平均，宇宙对于所有的基本观测者(宇宙中优越系上与物质的平均运动相联系的观测者)来讲总是各向同性的。当这种各向同性是准确的，宇宙在空间上是均匀的和各向同性的[Walker，1944；Ehlers，1993；Ellis，1971a]，那么物质运动就沿着无旋转无切变的测地线进行，其正切矢量为 u^a，这暗示了正则时间变量 t 的存在，这个时间变量遵循 $u^a=-t_{,a}$。用来描述宇宙大尺度结构的罗伯森—沃克(“RW”)几何确切地包含了这些对称[Robertson，1935；Walker，1936]。因此它们是共形平坦的，也就是说，外尔(Weyl)张量为零：

$$C_{ijkl}:=R_{ijkl}+\frac{1}{2}(R_{ik}g_{jl}+R_{jl}g_{ik}-R_{il}g_{jk}-R_{jk}g_{il})-\frac{1}{6}R(g_{ik}g_{jl}-g_{il}g_{jk})=0 \quad (1)$$

这个张量代表自由引力场，导致了比如潮汐力和引力波等非局域的影响，而这些影响确实不会发生在严格的罗伯森—沃克几何空间。

可以选择共动坐标以使得度规具有如下形式：

$$ds^2=-dt^2+S^2(t)d\sigma^2,\ u^a=\delta_0^a(a=0,1,2,3) \quad (2)$$

其中 $S(t)$ 是时间依赖的比例因子，正切矢量为 $u^a=dx^a/dt$ 的世界线表示基本观

测者的历史。空间部分$\{t=常数\}$是同质性的面，它具有最大的对称性：它们是常曲率$K=k/S^2(t)$的三维空间，其中k是K的符号。正则化的度规$d\sigma^2$描述了一个规范化常曲率k的三维空间，可以选择坐标(r，θ，ϕ)而得到：

$$d\sigma^2 = dr^2 + f^2(r)(d\theta^2 + \sin^2\theta d\phi^2) \tag{3}$$

其中$k=\{+1, 0, -1\}$时，$f(r)=\{\sin r, r, \sinh r\}$。$t$时刻的膨胀速度由哈勃参数$H(t)=\dot{S}/S$来描述。

为了确定度规在时间中的演化，人们运用了爱因斯坦(Einstein)场方程，表明物质在时空曲率上对度规式(2)和式(3)的影响。因为局域熵，张量T必然有着一个相对于正切矢量为u^a的优越世界线完美的流体形式：

$$T_{ab} = (\mu + p/c^2)u_a u_b + (p/c^2)g_{ab} \tag{4}$$

其中c为光速。能量密度$\mu(t)$和压强项$p(t)/c^2$是T_{ab}的类时或者类空本征值。爱因斯坦场方程的可积条件是能量密度守恒方程：

$$T^{ab}_{;b} = 0 \Leftrightarrow \dot{\mu} + (\mu + p/c^2)3\dot{S}/S = 0 \tag{5}$$

当适当的态函数方程$w := pc^2/\mu$把压强p与密度μ以及温度T: $p=w(\mu, T)\mu/c^2$相关起来时，上式就确定了(w可能是常量，也可能不是)。重子有$\{p_{bar}=0 \Leftrightarrow w=0\}$，辐射有$\{p_{rad}c^2=\mu_{rad}/3 \Leftrightarrow w=1/3, \mu_{rad}=aT^4_{rad}\}$，由于式(5)，它们意味着：

$$\mu_{bar} \propto S^{-3},\ \mu_{rad} \propto S^{-4},\ T_{rad} \propto S^{-1} \tag{6}$$

比例因子$S(t)$遵循瑞查得符里(Raychaudhuri)方程：

$$3\ddot{S}/S = -\frac{1}{2}\kappa(\mu + 3p/c^2) + \Lambda \tag{7}$$

其中κ和Λ分别是引力常数和宇宙学常数。[①] 这表示物质和场的主动引力质量密度是$\mu_{grav} := \mu + 3p/c^2$。对于普通物质，它为正，即

$$\mu + 3p/c^2 > 0 \Leftrightarrow \omega > -1/3 \quad (\text{“强能量条件”}) \tag{8}$$

因此普通物质会导致宇宙膨胀减速($\ddot{S}>0$)。同样明显的是，正的宇宙学常数自身将会引起加速膨胀($\ddot{S}<0$)。当物质和宇宙学常数都出现时，占控制

① 宇宙学常数也可以被看作是一种流体，其压强p与能量密度μ通过$\{p=-\mu c^2 \Leftrightarrow w=-1\}$相关联。宇宙学常数的历史，见[Earman(厄尔曼)，2001；2003]

地位的因素决定出现的结果。当$\dot{S}\neq0$时方程式(5)和(7)的一次积分是弗里德曼方程：

$$\frac{\dot{S}^2}{S^2}=\frac{\kappa\mu}{3}+\frac{\Lambda}{3}-\frac{k}{S^2} \tag{9}$$

这就是把三维空间曲率和四维空间曲率关联起来的高斯(Gauss)方程，表明了物质是如何直接导致三维空间曲率的[Ehlers, 1993; Ellis, 1971a]。由于时空对称性，10个爱因斯坦场方程等价于两个方程式(7)和(9)。这种有着度规为式(2)和式(3)、动力学由方程式(5)、(7)和(9)支配的罗伯森—沃克几何结构的模型，被称为弗里德曼—勒梅特宇宙。弗里德曼方程式(9)支配着宇宙的膨胀，守恒方程式(5)支配着随宇宙膨胀物质的密度，当$\dot{S}\neq0$时，如果式(5)和式(9)都得到满足的话，方程式(7)必然会成立。

给定每一个物质组分的明确的物质描述，即明确或暗示地指定态$w=w(\mu, T)$的方程，解的存在性和唯一性对于单物质组分和不同物质的组合比如$\mu=\mu_{bar}+\mu_{rad}+\mu_{cdm}+\mu_{\nu}$都适用。这里我们包括了冷暗物质和中微子。在任意时刻(例如今天)，这些解的初始数据包括：

- 哈勃常数$H_0:=(\dot{S}/S)_0=100h$ km/sec/Mpc;
- 存在的每一类物质的无量纲密度参数$\Omega_{i0}:=\kappa\mu_{i0}/3H_0^2$(用$i$标记)；
- 如果$\Lambda\neq0$，那么就有：或者$\Omega_{\Lambda0}:=\Lambda/3H_0^2$，或者无量纲减速参数$q_0:=(\ddot{S}/S)_0H_0^{-2}$。

给定物质的状态方程，这些数据就决定了唯一解$\{S(t), \mu(t)\}$，即唯一相应的宇宙历史。整体的物质密度是存在的每一类物质的Ω_{i0}项的总和，例如：

$$\Omega_{m0}=\Omega_{bar0}+\Omega_{rad0}+\Omega_{cdm0}+\Omega_{\nu0} \tag{10}$$

整体密度参数Ω_0是物质和宇宙常数的密度参数的和：

$$\Omega_0=\Omega_{m0}+\Omega_{\Lambda0} \tag{11}$$

现在对瑞查得符里方程(7)进行评估，可以给出这些参数之间重要的关系。当相对于物质项μ，压强项p/c^2可以忽略(在当前时刻是可信的)，[1] 就有：

---

① 假设我们把“暗能量”(2.3.6节)表示为一个宇宙学常数。

$$q_0 = \frac{1}{2}\Omega_{m0} - \Omega_{\Lambda 0} \tag{12}$$

这表明宇宙学常数 Λ 可以引起加速（负的减速因子 q_0）。如果它等于0，表达式简化为：$\Lambda = 0 \Rightarrow q = \frac{1}{2}\Omega_{m0}$，表明了物质如何引起宇宙的减速。在 t_0 时刻对弗里德曼方程(9)进行评估，空间曲率为：

$$K_0 := k/S_0^2 = H_0^2(\Omega_0 - 1) \tag{13}$$

$\Omega_0 = 1$ 对应于空间平直的宇宙（$K_0 = 0$），它与具有正空间曲率的模型（$\Omega_0 > 1 \Leftrightarrow K_0 > 0$）和具有负空间曲率的模型（$\Omega_0 < 1 \Leftrightarrow K_0 < 0$）区别开来。

弗里德曼—勒梅特模型是现代宇宙学的标准模型，从几何结构的极其简单性来说，它们令人惊奇地有效。它们的一个突出优点是在明确物质的局域引力效应的方式和决定宇宙进化的辐射时所具有的良好的解释能力。作为整体，这转而形成了局域物理学（包括物质和辐射的演化）的动力学背景。

2.1.1 基础解

对于重子（无压物质）和无相互作用的辐射，弗里德曼方程(9)具有以下形式：

$$\frac{3\dot{S}^2}{S^2} = \frac{A}{S^3} + \frac{B}{S^4} + \frac{\Lambda}{3} - \frac{3k}{S^2} \tag{14}$$

其中 $A := \kappa\mu_{bar0}S_0^3$ 和 $B := \kappa\mu_{rad0}S_0^4$，行为依赖于宇宙学常数 Λ [Robertson, 1933; Rindler（瑞德勒）, 2001]。

当 $\Lambda = 0$ 时，宇宙开始于一个非常致密的初始态——根据经典理论，这是一个密度和曲率为无穷大的奇点（见2.1.2节）。它将来的命运依赖于空间曲率的值，或者等价地依赖于密度参数 Ω_0。如果 $\{k = 0 \Leftrightarrow \Omega_0 = 1\}$ 或者 $\{k < 0 \Leftrightarrow \Omega_0 < 1\}$，宇宙将永远膨胀下去，但是如果 $\{k > 0 \Leftrightarrow \Omega_0 > 1\}$，宇宙会坍缩为一个奇点。因此 $\Omega_0 = 1$ 对应于临界密度 μ_{crit}，它将 $\Lambda = 0$ 的将在未来塌缩的弗里德曼—勒梅特模型和永远膨胀下去的模型区分开。Ω_0 只是物质密度和临界密度的比值：

$$\{\Omega_{crit} = 1 \Leftrightarrow \kappa\mu_{crit} = 3H_0^2\} \Rightarrow \Omega_0 := \kappa\mu_0/3H_0^2 = \mu_0/\mu_{crit} \tag{15}$$

当 $\Lambda < 0$ 时，所有的解都开始于一个奇点然后再坍缩。

当 $\Lambda > 0$ 时，如果 $k = 0$ 或者 $k = -1$，那么所有解开始于一个奇点，并将一直膨胀下去。如果 $k = +1$，那么又可以存在以奇点开始的模型，或者永远膨胀

下去，或者将来坍缩为一个奇点。然而在这种情况下也可能会有一个静态解(爱因斯坦静态宇宙)，也会有这个静态宇宙在将来或过去的渐进模型。而且 $k=+1$ 的模型将会发生反弹(从无穷大坍缩到一个最小半径，然后再膨胀)。

这些模型的动力学表现已经被深入研究过：首先是尘埃加上一个宇宙常数[Robertson，1933；Rindler，2001]，接着出现理想流体、带有体相黏滞性的流体、运动学理论解，标量场解。当前的模型采用了物质组分的现实性混合(重子、辐射、中微子、冷暗物质和标量场，或许还有一个宇宙学常数)。包含很多信息的相平面清晰地表明了更高对称性(自相似)的模型作为其他模型的吸引子和鞍点起作用的方式[Madsen and Ellis(马德森和埃利斯)，1988；Ehlers and Rindler，1989]。

最简单的膨胀解如下所述。

1. 爱因斯坦—德西特(Einstein-de Sitter)模型，对于它 $\{p=0,\ \Lambda=0,\ k=0\}\Rightarrow\Omega_0=1$。

以下是最简单的膨胀非空解：

$$S(t)=Ct^{2/3} \tag{16}$$

它始于 $t=0$ 时刻的一个奇异态(C 是任意的常数)。当哈勃常数取值为 H_0 时宇宙年龄 $\tau_0=\frac{2}{3H_0}$(从宇宙开端的固有时)。这是一个很好的宇宙膨胀模型，因为目前这个阶段，宇宙学常数开始占主导地位控制着膨胀，辐射的控制已经结束。如果 $k=0$ 且 $\Lambda=0$，对于遥远未来的宇宙这也是一个很好的模型。

2. 米尔恩(Milne)模型，对于它 $\{\mu=p=0,\ \Lambda=0,\ k=-1\}\Rightarrow\Omega_0=0$，给出了一个线性膨胀的空解：

$$S(t)=Ct \tag{17}$$

这就是正如一组均匀膨胀的观测者可以看到的平直时空，在 $t=0$ 点奇异。它的年龄为 $\tau_0=\frac{1}{H_0}$。如果 $k<0$ 且 $\Lambda=0$，对于遥远未来的宇宙，这是一个很好的模型。

3. 德西特宇宙，对于它 $\{\mu=p=0,\ \Lambda\neq0,\ k=0\}\Rightarrow\Omega_0=0$，给出了一个稳

态膨胀空解:①

$$S(t)=C\exp(Ht) \tag{18}$$

C 和 H 为常数。由于膨胀速率永远是常数，因此不存在开端，它的年龄为无穷大。② 对于 $\Lambda>0$，宇宙一直膨胀，这是一个很好的遥远未来的宇宙模型。也可以理解为 $\Lambda=0$，同时包含具有 $\mu+p/c^2=0$ 态的特殊方程的物质的解。也有德西特宇宙的其他罗伯森—沃克形式：一种测地完整形式 $k=+1$，$S(t)=S_0\cosh Ht$（常规反弹），另外一种测地不完整形式 $k=-1$，$S(t)=S_0\cosh Ht$（一个奇异开端）。因为这是一个常曲率的时空，没有特定的类时方向或者空间区域，③ 所以解没有唯一性。[Schrödinger（薛定谔），1956；Hawking and Ellis，1973；Rindler，2001]

2.1.2 最初的奇点？

以上都是特殊的模型，什么是一般模型呢？当满足不等式(8)，可以直接从瑞查得符里方程(7)得到：

弗里德曼—勒梅特宇宙奇性定理。在任何时刻都有 $\Lambda\leqslant0$，$\mu+3p/c^2>0$ 的弗里德曼—勒梅特宇宙中，在任意 t_0 时刻，$H_0\equiv(\dot{S}/S)_0>0$ 存在有限的时间 t_*：$t_0-(1/H_0)<t_*<t_0$，使得 $t\to t_*$ 时 $S(t)\to0$。如果 $\mu+p/c^2>0$，宇宙始于那里的一个时空奇点，其中 $\mu\to\infty$，$T\to\infty$。

这不仅仅是物质的开始，也是空间、时间和物理本身的开始。它是宇宙历史中最戏剧性的事件：它是一切存在的开始。其背后的物理特征是爱因斯坦场方程的非线性本质：越回到过去，宇宙越坍缩，有效重力密度越高，就会引起它更加严重的坍缩。我们期望能够避免瓦解塌缩的压强 p 却使情况变得更糟，因为 p 出现在瑞查得符里方程(7)中而且和能量密度具有相同的阶数（导致爱因斯坦场方程的形式）。需要注意的是哈勃常数给出的宇宙年龄估计是：$\tau_0=t_0-$

① 邦迪、霍尔德(Hold)和霍伊尔的宇宙稳定态使用了这个度规[Bondi，1960]，但在他们放弃爱因斯坦场方程时，度规是非空的。

② 但它是突出的，因为它是测地不完全的。这个度规只覆盖了德西特双曲面的一半[Schrödinger，1956；Hawking and Ellis，1973]。

③ 还存在这个度规的静止（非罗伯森—沃克）形式——那是度规被发现的第一种形式。

t_*，因为宇宙的开始早于 $1/H_0$。

原则上这样的结论可以通过宇宙学常数避免，但是实际上这并没有什么影响，因为我们知道宇宙至少以 6 倍的速度在膨胀，正如我们观测得知物体以 5 倍速率红移。①从式(14)可知，宇宙学常数必须有一个有效值，至少是当前物质密度的 $6^3=216$ 倍，导致接下来或者在任何早些时候有一个转变，因此会比当前观测到的上限大(和当前的物质密度有相同的阶数)。因此，在经典物理学支持下没有转变的可能[Ehlers and Rindler，1989]。然而如果它们在足够早的时期能够占主导地位，能量违背的物质组分比如标量场(见 2.6 节)，就可避免这样的结论。但这只能在量子场具有主导意义，即当宇宙至少比现在小 10^{12} 倍的时候才成立。

$T_{rad}\propto S^{-1}$(方程(6))的一个主要结论是，热大爆炸必然已经发生；密度和温度必然已经上升到至少有足够高的使量子场有意义的能量，即某种类似于大统一理论的能量；宇宙必然已经达到那些极端的温度和能量，在那种温度和能量条件下，经典理论被打破了。

2.2 热大爆炸

当我们追溯到面对初始量子态的时期，在宇宙中的物质和辐射变得越来越热，因为它们被压缩到较小的体积。在宇宙早期的热大爆炸时代，我们可以使用标准物理规律来检验物质和辐射的混合物不断扩展分开的过程[Weinberg，1972；Perkins，2005]。一个关键的特征是，在热大爆炸开始约 30 万年后，原子核和电子结合形成原子。在更早的时期，当温度更高时，原子无法存在，这是因为辐射具有这么多的能量，从而辐射能量摧毁一切试图形成它们的稳定的组成部分(原子核和电子)的原子。因此，在更早时期，物质是电离的，由相互独立地运动的带负电荷的电子和带正电荷的原子核组成。在这些条件下，自由电子和汤姆逊散射描述的辐射剧烈地相互作用。因此在那些时期，物质和辐射在平衡中紧紧耦合，而宇宙对辐射是不透明的。当温度经过电离降到约 4000K

---

① t_e 时刻发射，t_0 时刻被观测到的光线的红移 z 是通过 $1+z=S(t_0)/S(t_e)$ 与膨胀相关的，见 2.3.3 节。

时，原子核和电子形成原子，因而这种散射停止，因此宇宙变得非常透明（今天，我们能够看到遥远的星系）。这种过渡发生的时期就是已知的退耦时期——就是物质和辐射不再彼此紧紧耦合的时期，其时红移为 $z_{dec} \simeq 1100$ [Dodelson，2003]。由文献(6)可知，在早期宇宙以辐射控制($\mu_{rad} \gg \mu_{mat}$)为主，而在晚期则以物质控制($\mu_{rad} \ll \mu_{mat}$)为主。[①] 在退耦前物质—辐射密度等价性显著（这种等价性发生时的温度 T_{eq} 满足 $T_{eq} \simeq 10^4$K，此时比例因子为 $S_{eq} \simeq 10^4 S_0$，其中，S_0 是现在的取值）。S_{eq} 之前和之后，背景模型及其微扰的动力学有着明显的不同[Dodelson，2003]。

2.2.1 宇宙黑体辐射

在退耦时期，辐射是由物质发出的，因此辐射通过物质之间的空间自由地传播到我们这里。当它被发射时，具有黑体辐射的形式，因为这是早期的物质和辐射在热力学平衡中的结果。因此，在早期宇宙中，物质在 $z = z_{dec}$ 形成最后的散射面，在 4000K 时散发出极广阔宇宙的黑体辐射，[②] 自那时以来宇宙黑体辐射自由地传播，它的温度 T 与宇宙的尺度函数成反比。[③] 当辐射传向我们时，宇宙以 1100 的膨胀系数扩展。因此当它传到我们的时候，它已经冷却至 2.75K（即绝对零度以上 3 度左右，微波频谱峰值区域），所以是非常难以观测的。然而，人们在 1965 年测到了它，此后对它的光谱集中进行了大量地探测，以很高的精确度证明了它的黑体性质[Partridge（帕特里奇），1995]。现在，它的存在被作为一个有力的证据，证明宇宙确实是从一个很热的早期膨胀而来，并且标准物理学在早期宇宙那个时代的应用并未改变。

2.2.2 粒子的相互作用和元素形成

辐射的热容能力远远大于物质的热容能力。在很早的退耦前时期，物质和辐射的温度是相同的（因为它们彼此处于相互平衡的状态中），尺度为 $1/S(t)$，见方程(6)。早期宇宙的温度要超过地球上甚至太阳中心曾经能够达到的任何温度；当它下降到现在的温度 3K 时，连续的物理反应发生了，这些反应决定

---

① 动力学支配的冷暗物质(2.3.6 节)与重子遵循同一个密度定律(6)。

② 这通常被称为"宇宙微波背景"，或简写为 CMB。但它只是当前时代的微波。

③ 这种自由传播辐射的尺度来自 2.3.3 节的讨论。

着我们今天看到的周围物质的性质。在非常早的时间和高温下，只有基本粒子能够存在，甚至连中微子都有一个非常小的平均自由程，当宇宙冷却下来后，中微子从物质中去耦而在空间中自由地移动。在这些时期，宇宙膨胀是以辐射为主的，我们可以通过$\{k=0,\ w=1/3,\ \Lambda=0\}$这一模型模拟宇宙，式(14)的解把时间和温度唯一地关联了起来：

$$S(t)=S_0t^{1/2},\ t=1.92\sec\left[\frac{T}{10^{10}K}\right]^{-2} \tag{19}$$

其中后一个方程中没有自由常数。

在极早期，甚至连中微子都紧密耦合并且处在具有辐射的平衡中。它们在约10^{10}K 的温度下退耦[Dodelson，2003，pp. 44—46]，如果它们没有质量，则将导致今天宇宙遗留的中微子背景密度为$\Omega_{\nu0}\simeq10^{-5}$(可能更高，这取决于它们的质量)。早期宇宙中的关键事件与失衡现象的出现有关[Dodelson，2003，58]。一个重要的事件是核合成时代，即轻元素形成的时期。温度大约在10^9K以上时核无法存在，因为辐射是如此的强大以至于原子核一旦形成就被分裂成它们的组成部分(质子和中子)。然而在这一温度以下，如果粒子带着足够导致原子核反应发生的能量互相冲撞，合成的原子核保持完整无损(辐射是比它们的结合能要弱，因此无法分裂它们)，因而轻元素(氘、氚、氦和锂)的原子核通过俘获中子而被创造了。当温度下降到约10^8 K(核反应的临界温度)以下时，该过程终止。在这种方式中，轻元素的比例在核合成期结束时就已经确定；自此它们实际上维持不变。这些反应的速率非常高，且全部反应都发生在宇宙膨胀后的最初 3 分钟内。宇宙大爆炸理论的重大成就之一是：假设重子的密度很低($\Omega_{bar0}\simeq0.044$)，理论和观测就完美地一致。那么，这些元素的预测丰度——25%(权重)氦气，75%的氢气，其他元素丰度小于1%——同所观测到的丰度非常接近一致。因此，标准模型解释了轻元素的起源，这是以已知的早期宇宙中发生的核反应[Schramm and Turner(施拉姆和特纳)，1998]为依据的。然而，更重的元素不能在有效时间(约 3 分钟)内形成。

以类似的方式，极早期宇宙(核合成期之前)的物理过程可以用来解释现今宇宙中物质和反物质的比例：物质在数量上相对于反物质的小量超越，必须在重子合成的过程中创造，否则我们今天就不能生存。如果没有这样的过剩，物

质和反物质会全部湮灭而只剩下辐射[Silk, 2005]。然而，即使在极端恶劣的早期宇宙的条件下，其他数量(如电荷量)也已被认为是守恒的，所以它们现有的值产生于宇宙起源所给定的初始条件，而不是产生于宇宙发展时所发生的物理过程中。在电荷的例子中，整个守恒量似乎为零：在夸克形成质子和中子的重子合成时期后，带正电荷的质子和带负电荷的电子数量相等，因此在退耦时期，刚好有足够的电子与原子核结合而形成不带电的原子(在巨大的天体物体如我们的银河系中似乎没有净电荷，如果这并不为真，将会是电磁力而不是引力主宰宇宙学)。

退耦后，物质由于引力的不稳定性形成了大尺度结构[Bothun, 1998, pp. 183—222]，这最终导致第一代恒星的形成[Silk, 2005]，并可能与物质的重新电离关联起来［Dodelson, 2003, 73]。然而那时行星还无法形成，这有一个很重要的原因：当时的宇宙中没有重元素。第一批星体通过万有引力吸引作用将物体聚集在一起，物质越来越集中，就开始变热，直到其温度超过热原子核反应的燃点，核反应开始燃烧氢，从而形成氦。最后，更复杂的核反应活动在围绕同心球的中心开始了，产生重元素(例如碳、氮、氧)，直到铁的逐步形成。这些元素可以在星体中形成，因为有很长一段时间(数百万年)用于发生这种反应。大质量恒星燃烧较快，并最终耗尽核燃料。该恒星变得不稳定，因为引力吸引其核心迅速坍塌。温度上升的结果是在一个巨大的爆炸中将它吹开，在这段时期内新的反应发生，生成比铁重的元素。这就是当超新星在空中突然猛烈燃烧时我们看到的超新星爆炸(“新星”)，那里原先只是一个普通的星体。这样的爆炸使星体内部聚集的重元素吹进太空，形成大量的灰尘细丝围绕着天体的剩余残骸，这是在后来能被渐渐累积起来的材料。在第二代星体的形成过程中，这些材料可以形成围绕那些恒星的行星系统。因此，组成我们的元素(例如碳、氮、氧和铁的原子核)是在星球内部极高温的条件下被创造的，它们由于超新星爆炸而变得可为我们所用。如果没有这些爆炸，我们可能就不存在。

2.3 宇宙观测

宇宙学模型只有在提到天文观测时才有意义[Hoyle, 1960；Sandage, 1961；Ellis, 1971a；Weinberg, 1972]。这表现为两点：遥远物质的天文观测告诉我们

在遥远的宇宙中以及(因为光速有限)很久以前发生了什么事情；另一方面，周围物体(例如地球上的物质、太阳系、附近的星体)的观测，当涉及起源理论时告诉我们很长一段时间以前我们过去的世界线附近发生了什么。第一组观测可以被描述为“零锥”观测，第二组作为地质学组，最重要的项目之一是确定局域元素丰度，这些涉及核合成计算(见 2.2.2 节)。

观测完全依赖于望远镜和探测器技术[Harwit，1984；Bothun，1998]。在 20 世纪 20 年代和 30 年代，距离标度和宇宙的均匀性以及膨胀的基本证据初步建立，一直到 60 年代观测从可见光延伸至整个电磁频谱期间，进展缓慢。在最近几十年间，宇宙学从一个资料贫乏的学科转变为一个资料充足的学科。现在有大量新的可用的资料库，这是因为望远镜、探测器和计算机技术在近几十年中有了惊人的改进，特别是诸如电荷耦合器件(CCD)和光纤(可同时测量数以百计的红移)等新探测器的出现。我们现在不仅能用可见光、紫外线和红外线对星系进行观测和以空前的感光灵敏度确定光度和光谱，也能用无线电、X 射线和伽玛射线进行巡天观测。我们探测到的星系的红移值达到 6，并且我们已经发现了许多类星体和伽玛射线暴以及得到了多张由引力透镜拍摄的非常遥远星系的图片［Harwit，1984］，已经确定了与大规模速度流相关联的大尺度结构(星系团、超星系团、壁垒、空洞)[Bothun，1998，pp. 85—137]。

除了大型数计数和红移探测，我们已经测量了在全部波长下的背景辐射光谱和各向异性。当辐射在很小的角尺度(相对于作为孤立对象出现的离散源)下为常数时，我们精确地确定了作为“背景”的辐射。这种辐射和星系间物质的密度以及热历史有相当复杂的关系。背景辐射的最重要的组成部分是上面(2.2 节)提到的宇宙黑体辐射。详细的观测已经将其温度映射到整个天空，其灵敏度超过十万分之一。然而，背景辐射的其他组成部分(特别是 X 射线和无线电)传达了关于温度和星际物质密度的重要信息，并因此强烈地限制其可能的热学历史。例如热物质发射 X 射线，所以 X 射线背景测量限制了热星系间物质所允许的量，而中性的氢原子强烈吸收紫外辐射，产生莱曼—阿尔法光谱吸收线，所以没有这样的吸收强力地限制了中性氢原子的数量以及星系际物质的温度。

2.3.1　各向同性

宇宙学观测的第一要点是，当我们在一个足够大的物理尺度进行平均时，

它们对我们来说，在统计上是各向同性的，并没有显著的方向指向宇宙的中心。宇宙黑体辐射的高度各向同性有力地支持了以下结论：在我们允许地球相对于宇宙的运动（大约 250km/s）之后，它的温度在相对于我们的各个方向上是一致的，而这将产生千分之一的温度偶极子。[①] 正如在 COBE 和 WMAP 卫星上十万分之一的极为灵敏的探测器的最近一次测量，任何物质分布的不均匀性或各向异性将导致这种辐射的各向异性。我们认为高度的各向同性是主要原因，且在一个较好的近似下（参看 4.2.2 节），宇宙本身是空间均匀、各向同性的，而这为弗里德曼—勒梅特宇宙模型作为可观测到的宇宙区域的精确模型提供了观测支持。

2.3.2 距离尺度和年龄

所有天文学的根本问题是确定观测对象的距离。这是通过"宇宙距离尺度"[Bothun, 1998, pp. 25—83]——即最近的物体之间的距离由视差决定（本质上是局部的三角学——来完成的；而更遥远物体的距离由一系列的连续距离指标（造父变星，天琴 RR 型变星，最亮的红超巨星）直到一个宇宙学距离给出，红移 z 是一个主要的距离指示器，但却被物质相对于宇宙的静止框架的局部速度干扰。其他距离指示器例如塔利—费希尔（Tully-Fisher）法、行星状星云的光度函数、星团光度函数、表面光亮度的涨落有助于改进估算值[Bothun, 1998]。

与距离尺度密切相关的是哈勃常数 H_0（现在的宇宙膨胀速度）的确定，这是因为宇宙可观测区域的规模与哈勃常数相当。但哈勃常数也决定了宇宙的年龄，因此它的确定强调了宇宙学关键的一致性条件：宇宙中物体（岩石、行星、恒星、星团、星系）的年龄必须小于宇宙的年龄。自从我们有了对年龄和哈勃常数的良好估计,[②] 这一条件已经引起人们的关注。当前的低红移物体的观测（当前对 H_0 的估计为 70 km/sec/Mpc）似乎并不违反这一条件，其给出宇宙的年龄约 15 亿年，而最古老的星团大约 14 亿岁。然而，对非常遥远（年轻许多）

---

① CBR 偶极子可以被解释为归因于主要的宇宙非均匀性，但在一定程度上却被解释为是归因于我们相对于一个空间均匀性的宇宙的运动（"特有速度"）。

② 事实上因为宇宙年龄的困难，哈勃自己从来没有完全接受膨胀宇宙理论。他更倾向于提到红移距离关系而不是速度距离关系[Hubble, 1936]。但后来通过哈勃常数的一系列修正，问题得到了缓解，这要归因于对主要的远距离指示器的一种更好的理解。

的物体[Jain and Dev(杰恩和戴维), 2005], 它甚至可能是有问题的。

2.3.3　观测关系

光在时空中沿零测地线 $x^a(\lambda)$ 传播(切线向量 $k^a := dx^a/d\lambda$ 是这样的: $k^a_{;b}k^b=0$, $k^ak_a=0$)。在罗伯森—沃克几何中, 仅考虑径向测地线即可(由于该模型的对称性, 它们分别等同于一般的测地线)。然后由式(2), 我们发现, 光在 t_e 时刻发射, 在 t_0 时刻接收, 共动发射器和接收器之间的共动径向距离 $u(t_0, t_1) := r_0 - r_1$ 由下式给出:

$$\{ds^2 = 0, d\theta = 0 = d\phi\} \Rightarrow u(t_0, t_1) = \int_{t_1}^{t_0} \frac{dt}{S(t)} = \int_{s_1}^{s_0} \frac{dS}{S\dot{S}} \tag{20}$$

$\dot{S}$ 由弗里德曼方程(9)给出。与宇宙学观测相关的关键量为红移、区域距离(或外观尺寸), 以及相应于距离增加(由计数确定)的局部体积[Sandage, 1961; Ellis, 1971a; Weinberg, 1972]。如果以波长为 λ_0 观测发射波长为 λ_e 的物体, 则红移 z 由下式给出

$$1 + z := \frac{\lambda_0}{\lambda_e} = (1 + z_c)(1 + z_v) \tag{21}$$

其中 z_v 是由物体的局部特定运动导致的红移(对于共动的物体 $z_v=0$), z_c 是宇宙学红移, 由下式给出:

$$1 + z_c = \frac{S(t_0)}{S(t_e)} \tag{22}$$

由式(21), 发射光观测到的相同比率对所有波长都成立: 这是红移的一个关键识别特性。使用物体的红移作为距离指标的问题是把宇宙学部分从多普勒部分(将导致红移空间扭曲)中分离出来[Dodelson, 2003, pp. 275—282]。对于一个星系团来说, 通过对星系团数目的合理平均能够完成(对于共动的系团 $\langle z_v \rangle = 0$)。在红移 z_c 下的物体其区域距离 r_0 和对角大小为 α 的线性尺寸 l 由下式给出:①

$$r_0(z_c) := \frac{l}{\alpha} = f(u)S(t_e) \tag{23}$$

因此, 如果源物理尺度已知, 外观尺度的测量将决定区域距离。从亮度为

① 这只取决于 z_c, 因为可见的形状和大小是独立于源的运动的。

L 的点源产生各向同性的辐射，其辐射通量 F 由在一个单位时间内从源发射并由望远镜接收的辐射能量的分数给出：

$$F=\frac{L}{4\pi}\frac{1}{f^{2}(u)S^{2}(t_{0})(1+z)^{2}} \tag{24}$$

红移的两个因素中占第一位的是在观测者和源之间观测到的时间膨胀，其次是由于光子红移导致的能量损失。源的视星等 m 由通量定义为：$m=-2.5\log_{10}F+$ 常数。应用式(22)和式(23)，式(24)化成：

$$F=\frac{L}{4\pi}\frac{1}{r_{0}^{2}(1+z)^{4}} \tag{25}$$

这表明如果源的固有光度已知，视星等的测量将决定面积距离。在式(25)中使用式(23)，逐点扩展物体的表面亮度(每单位立体角接收的通量)只依赖于红移[Kristian and Sachs, 1966; Ellis, 1971a]——在确定检测概率和引力透镜观测时的一个重要特征。它进一步遵从如下结果：在温度 T_e 发射的黑体谱，观测到一个红移 z 时，仍然是一个黑体谱，但观测到的温度为 $T_0=T_e/(1+z)$—— 分析宇宙黑体辐射观测结果时的一个重要特征。

运用弗里德曼方程和相关的物态方程，区域距离可确定为哈勃常数 H_0、减速因子 q_0 和宇宙学常数 Λ 的红移函数 z_c。在无压物质和宇宙学常数为零的情况下，从式(20)、式(9)、式(22)和式(23)出发，[①] 可以得到 Mattig 关系[Sandage, 1961]：

$$r_{0}(z_{c})=\frac{1}{H_{0}q_{0}^{2}}\frac{(q_{0}-1)(1+2q_{0}z_{c})^{1/2}+(q_{0}(z_{c}-1)+1)}{(1+z_{c})^{2}} \tag{26}$$

因此无论是对已知物理尺度的源的外观尺度，还是对来自已知固有亮度的源的辐射通量的测量都将确定减速因子 q_0。如果宇宙学常数或辐射存在，这个关系的推广将成立。有趣的是，在某个红移下 $z_c=z_*$(取决于密度参数和宇宙学常数)，对于固定物理尺寸的物件有一个最小表观尺度。观测者的过去光锥在 z_* 达到最大面积。对来自更远距离的物体，整个宇宙作为一个引力透镜将放大它的外观尺度，因此，非常遥远的物体可以有和附近物体相同的角大小

---

① 或者，更加优雅地，来自测地线偏差方程，见[Ellis and Van Elst(埃利斯和范·埃尔斯特), 1999b]。

[Hoyle，1960；Sandage，1961；Ellis，1971a]。对于爱因斯坦—德西特宇宙，最小的角直径是 $z_* = 1.25$，在低密度的宇宙中，它在高红移时出现。

距离增量为 du（由可观测变量 z 和 m 中的增量 dz 和 dm 刻画），立体角 $d\Omega$ 内可见的物体数量由下式给出：

$$dN = W(u)\rho(t_e)S^3(t_e)f(u)dud\Omega \tag{27}$$

其中 $W(u)$ 是检测概率或"选择函数"［Dodelson，2003，263］，$\rho(t_e)$ 是发射时的物体数密度（空间均匀性表现在其与空间坐标无关）。在一项调查中，观测到的物体总数 N 通过整合从观测者到调查的限制给出：根据源的径向坐标 r_e（这可涉及红移或幅度），$N = \int_{r_0}^{r_e} dN$。如果物体的数目守恒（如当星系处于既不产生也不湮灭的时代），$\rho(t_e) = \rho(t_0)(1+z)^3$。我们从式(27)发现，在理想化的情况下，当 W 与距离无关（对于相对较近的物体是一个合理的假设）时有：

$$N = W\rho(t_0)d\Omega\int_{r_0}^{r_e} f(u)du \tag{28}$$

对于 $k = +1, 0, -1$，应分别完成该简单积分［Sandage，1961］。

上述公式可用来确定观测变量之间的测量关系，例如已知内秉属性（如已知尺度或亮度）的物体的 (m, z)，(α, z) 或 (N, m) 关系，由此在观测上确定 q_0。如果存在银河间介质、引力透镜或者各向异性的辐射散射的吸收，那么这些关系就必须得到修正。具体的观测比较必须考虑源谱和源探测以及识别的可能性［Harwit，1984］。这里我们遇到了图像和实在之间的对比：可以有很多对象，我们既不去探测，也不去认识它们是什么［Disney（迪士尼），1976］。存在一种"观测映射"把源特性和它们图像的本质联系起来，从而给出了一种这是如何发生的的有用看法。

这里一个重要的特征是，特定的对象在不同波长（比如光波、无线电波、X射线）之下看起来会完全不同。事实上它可能在某一个波长下是可发现的，但是在另一个波长下就不可能被发现。这很清楚地表明了我们关于实在的图像是如何依赖于我们所用的探测器的。要得到存在什么的完整图景，我们需要运用多种研究模式——在所有的波长、密度、光谱和偏振测量下进行成像［Harwit，1984］，也寻找时间的变化。第二种重要特征是观测选择的影响，比如马姆奎

斯特(Malmquist)偏见——如果我们有一群对象，它们有着不同的光度，在大的距离上我们将只能看到更亮的对象(而不会发现光线微弱的)。因此，平均光度会看起来随着距离的增加而增加，但这只是一个观测效应而不是事物的真实态。运用不同的探测门槛，可以在某种程度上控制这种影响。

2.3.4 计数和可见物体密度

星系的计数作为红移或者光度的一个函数表明了宇宙的近似空间均匀性[Hubble, 1936]。但是对于辐射源和类星体的计数表明宇宙并没有处于像邦迪、霍尔德和霍伊尔[Bondi, 1960]所提出的那样的稳定态。事实上如果存在过源数目和/或光度的演化[Sciama, 1971]，那么计数只与罗伯森—沃克几何相容。

计数也给出了宇宙中可视(发光)物体的密度估计：$\Omega_{vm0} \simeq 0.015$。这相对于临界密度($\Omega_0 = 1$)来说非常低，也远远少于由核合成研究决定的重子数密度($\Omega_{bar0} \simeq 0.044$，见2.2.2节)。因此宇宙中大部分重物质是以某种隐藏(不发光)的形式存在的[Bothun, 1998, pp. 223—272]，比如烧毁的恒星[Hogan(霍根), 1999]。

2.3.5 表观光度及其大小：暗能量

如果已知固有源尺度或光度，那么外观尺度或光度作为红移的函数可以用来决定减速因子 q_0(见2.1节)。问题是，直到最近还没有已被人们足够了解的有着标准大小或光度的星系或其他对象可以用来决定 q_0，并且它们特定的分布导致了马姆奎斯特效应决定的对观测的偏见。但是这随着最近对远距离星系超新星光度衰退曲线的观测而发生了明显变化。已经证明 I_a 型超新星的最高光度与它们的光变曲线衰退时间紧密相关，第一次给出了远处星系以可靠的“标准烛光”[Perlmutter *et al.* (珀尔马特等), 1998]。从这些观测得到的结论是，宇宙的膨胀速度并非如预期的那样放缓了，而是以一个与 $\Omega_{\Lambda 0} = 0.7$ 的宇宙学常数对应的速度加速着。这个证据与从宇宙黑体辐射观测以及计数而来的证据相一致[Dodelson, 2003; Silk, 2005]。

引起这种加速的场或物质的本质还不清楚。它的态方程 $w := pc^2/\mu$ 还不为人们所知，在这一点上人们提出了很多物理的和非物理的想法。从式(7)，它必然要违背强能量原理(8)，因此必须有一个很大的负压。它事实上可以归因于一个宇宙学常数($w = -1$)，这可能支配了自红移 $z \simeq 0.33$ 以后的宇宙膨胀，

在更早的时期则可以忽略(并且在小尺度下也可以忽略——它并不影响局域天体物理学)。但是，也可能是其他形式的物质或者场，它有着有效的负压，导致 $w < -1/3$，就像在标量场中可以发生的那样，见后文中的方程(33)。在那种情况下，它被称之为“完美典型”。关于这可能是什么存在着许多猜想，但在物质上没有明晰性。这里应当注意，观测的可替代解释也是可能的，因为它们能够被一个球对称的非均匀宇宙模型确切地复制，在这种宇宙中我们差不多处于中心地位[Mustapha *et al.*（穆斯塔法等），1998]，或者至少部分地被归因于宇宙膨胀中非均匀性的逆反应的结果[Ellis and Buchert(埃利斯和布御特)，2006]或者是有效区域距离上的非均匀性的结果[Kantowski *et al.*（坎陶克思等)，1995；Kantowski，1998]。这些可替代的解释都被研究过，但最可能的原因还是某种未知的有着有效负能量的物质或者场。

总之，标准引力方程和超新星观测意味着一个宇宙学常数或某种等价形式的暗能量的存在，它有着巨大有效负能量密度 μ_{grav}(因负压而产生)，支配着宇宙的当前膨胀。它的物理本质仍然是未知的。还没有任何已知的物理学可以解释为什么这种力应当存在于这个恰好可以探测到的水平上——量子场论把宇宙学常数和真空零点能联系起来了，而且指出它应当要远远大于观测值[Weinberg，1989；2000a；2000b；Rugh and Zinkernagel(鲁和辛克纳吉)，2002；Zinkernagel，2002；Susskind(萨斯坎德)，2005]。它为什么会存在于观测所表明的这个很小(刚刚可以探测到)的水平上[Seife(塞费)，2003]，这仍是一个主要的谜团。当今宇宙学的关键是一方面要尽量在观测上决定“暗能量”的有效状态方程，即反推场方程以得到观测的 $w(z)$[Saini *et al.*（萨伊尼等），2000]，特别是要确定 w 是常数还是随时间不断变化的，另一方面则要尝试给它的物理起源一个看似合理的理论解释。

2.3.6 物质分布和运动：暗物质

星系的分布及其运动已经得到了详细研究。它们发生于星团之中，转而组成了超星系团，嵌在围绕相对稀疏的银河间的虚空的巨大的势阱中。星系的光度函数描绘了发生在每一个光度级中的星系的数量。协变函数描绘它们的空间群集[Peebles，1993b；Dodelson，2003]。星系发生在星团中的大尺度运动也在星团自身间发生。很简单就可以设想出难以探测的物质(比如分布于空间的小

石头)。银河系的转动曲线和星团中星系运动的研究[Bothun, 1998, pp. 139—181]说明了大量未发现的暗物质的存在，并支配着宇宙的动力学：它的密度是 $\Omega_{dm\,0}\simeq 0.3$，比可见物质 $\Omega_{vm\,0}\simeq 0.015$ 和重子 $\Omega_{bar\,0}\simeq 0.044$ 都大得多，但是比临界密度 $\Omega_0\simeq 1$ 又要小很多。因此暗物质是非重子的，意味着它有着某种奇异的本质，因而不是质子和中子那样的普通物质[Seife, 2003]。相对于上一节讨论的"暗能量"，暗物质在天体物理学尺度和宇宙学尺度上都是动力学有效的。为了确定它的本质，人们已经做了许多尝试。比如，它可能是轴子、已知粒子的超对称伙伴、夸克核，或者有质量的中微子[Gribbin and Rees(格里宾和里斯), 1991; Perkins, 2005]，但是它究竟是由什么构成的还是未知的。实验室研究已经开始尝试探测这种物质，目前还没有获得成功。一个关键的问题是它表观的存在是否是我们对这个尺度上的引力理论的错误运用得来的。这正在研究之中，研究人员提出了各种各样修正引力方程的想法，这些修正的引力方程可能不需要大量暗物质的出现就可以解释观测现象。这是一种理论可能性，但是目前的共识是暗物质确实存在。

一个重要的区别在于暗物质是否由以下东西构成。

(i)很快冷却的弱相互作用的有质量粒子，此后形成了冷暗物质，在结构形成的时候缓慢地运动（导致从较小尺度结构到大尺度结构形式的自下而上的过程）；(ii)低质量、冷却得很慢的粒子，因而很长一段时间才形成了热暗物质，在结构形成的时候非常快地运动(导致一个自上而下的星系形成图景)。

结构形成的研究目前支持冷暗物质假设，这一假设认为被引力包围的对象分级形式发生在一个复杂的自下而上的过程中，这一过程包括了冷暗物质、重子和辐射的相互作用，在最初就有着小星系的形成[Silk, 2005; Mouri and Taniguchi(毛利和泰尼古奇), 2005]，之后聚集成为一个更大的结构。这些研究的基础是大量的数字模拟，其初始条件是由我们下面 2.6 节将会讨论到的膨胀方案所给出的。和"暗能量"不同，冷暗物质有一个普通的重子态方程，是一个有着 $p_{cdm}=0\Leftrightarrow w_{cdm}=0$ 的理想流体(4)。

另一种在星团中探测暗物质的方式通过它的引力透镜效应进行[Schneider *et al.* (施耐德等), 1992]。光线因大质量对象而形成弯曲是广义相对论的经典预言之一。富星系团或者星系核能够导致更远物体的强引力透镜。在更远处，

远处星系和类星体的多重图像产生了，有时形成环或者弧。近一点的质量形成的较弱透镜的结果是远处物体图像扭曲的典型图像。多重图像的分析可以用来重构透镜质量分布，弱透镜扭曲图像的统计分析现在可以给出我们远处星系和星团中物质分布的详细信息。这些研究表明，要在一个几乎平直的宇宙($\Omega_0 \simeq 1$)中得到足够的镜头，需要宇宙学常数的出现——现在不可能有暗物质的临界密度[Dodelson, 2003; Silk, 2005]。

当今宇宙学的一个关键特征是要尝试认识这种暗物质的本质，而且如果可能的话，要在实验室条件下对其进行探测。虽然观测支持冷暗物质方案，但一些关于精细尺度结构的出现的残余问题仍有待解决[Silk, 2005]。

2.3.7　宇宙黑体辐射功率谱

宇宙黑体辐射角度各向异性由一个角功率谱描述。这个功率谱展示了最后的散射面中每一个物理学尺度上微扰的功率总和[Bennet *et al.*(本尼特等), 2003; Seife, 2003; Dodelson, 2003]。在从暴胀(inflation)结束到最后的散射面的时间段中，不同波长模式可以完成不同数目的震动。这就把波的典型周期转化成了最后的散射面上的典型长度，导致了最后的散射面上非均匀性中的一系列极大值(“声峰”)和极小值，因而也导致了宇宙黑体辐射角度各向异性功率谱中的极大值和极小值[Hu and Sugiyama(休和苏格亚马), 1995b; 1995a; Peacock, 1999; Perkins, 2005]。这些非均匀性后来形成了结构形成的种子，与后来形成的结构的物理尺度功率谱相关联。它们由最后的散射面上的一个(三维)空间功率谱描绘，因为我们在二维球面 $S_{2:LSS}$ 收到观测的宇宙黑体辐射，这个球面是我们的过去光锥与最后散射面相交形成的，是作为天空中各向异性的二维的功率谱(由所有方向矢量 e_a: $e^a e_a = 1$, $e^a u_a = 0$ 的单位球面 S_2 所刻画)为我们所看到的。

最大的宇宙黑体辐射峰值(大约1°)的明显角度大小可以用来估计最后的散射面区域的距离，因此也就可以用来估计不同宇宙学常数值下的宇宙中物质的密度，并且决定着主要的宇宙学参数[Spergel *et al.*(斯伯格等), 2003]：

“通过把 WMAP 数据与其他天文学数据集结合起来，我们限制了宇宙的几

何 $\Omega_{tot}=1.02\pm0.02$，暗能量的状态方程 $w<-0.78$(95% 置信限度)，以及中微子中的能量密度 $\Omega_\nu h^2<0.0076$(95% 置信限度)。对于三维退化的中微子，这种限制意味着它们的质量要小于 0.23eV(95% 置信限度)。早期再离子化的 *WMAP* 探测排除了热暗物质的可能性。”

但是这里存在一个问题，虽然理论在观测上对于所有小角尺度上的一致性是引人注目的，在最大角尺度上却存在着发散。观测表明功率要小于预期。特别地，四极和八极要远远低于理论的预期。四极和八极的轴也非常精确地相互关联，并且还存在着其他的角度异常[Starkman and Schwarz(斯塔曼和施瓦茨)，2005]。这些效应可能是因为：(i)星系对观测的污染(它妨碍了我们对最后的散射面的视线)；(ii)可能(“偶然”)事件(它表示“宇宙方差”，具体的讨论见第 3 节)；(iii)我们居住于一个比例均衡的“小宇宙”中，它在空间上是封闭的，因而存在可能涨落的最大尺度[Weeks *et al.*(维克茨等)，2003]；(iv)某些意外的新物理影响我们对早期宇宙的理解上更深刻的问题。哪一种是实际情况，现在仍悬而未决。这可能最终是宇宙黑体辐射分析的一个危机，但是另一方面，人们可以仍只说这是一个统计上的侥幸成功(这里的根本问题是宇宙的唯一性，如第 3 节所讨论的)。

在这种辐射的偏振谱中也存在着类似的预期峰值，偏振图应当有着与引力波相联系的模式。暴胀预言了引力波在极早期宇宙中的存在(2.6 节)。对这种模式的检测将会是对暴胀的一个关键检验[Dodelson，2003；Sievers *et al.*(西弗斯等)，2005b]。偏振的研究表明宇宙的再离子化发生得和红移为 17 时一样早，这与类星体研究的推论相反。对各向异性的更详细研究涉及桑亚维—泽尔多维奇(Sunyaev-Zeldovich)效应(由于星团中热物质散射而引起的观测温度的变化)和引力透镜。

在宇宙黑体辐射图中存在着大量的信息，对它们进行更加精确的测量和解释是当今宇宙学的一个核心内容[Steinhardt(斯坦哈特)，1995；Peacock，1999；Dodelson，2003；Perkins，2005]。

2.4 因果和可视视界

一种影响结构形成和我们观测情况的基本特征可归因于因果影响不能在大

于光速时传播而引起的限制。这样，能够因果地影响我们的区域被限制在我们的过去零锥之内。结合宇宙的有限年龄，这就导致了粒子视界的存在。这个粒子视界限制了我们能够拥有因果联系的宇宙部分。①

粒子视界在定义上是由与我们的过去零锥相交的最远物质限制的世界线构成的[Rindler, 1956; 2001]。这是在宇宙开端之后能够拥有哪怕丝毫因果联系的物质的极限，由共动辐射坐标值

$$u_{ph} = \int_0^{t_0} \frac{dt}{S(t)} \tag{29}$$

描述。目前构成视界的物质的物理距离为：

$$d_{ph} = S(t_0)\, u_{ph} \tag{30}$$

关键的问题是当我们转向初始奇点的极限 $S \to 0$ 时，积分(29)是收敛的还是发散的。对于所有普通物质和辐射，视界都会存在于标准弗里德曼—勒梅特宇宙中，因为 u_{ph} 在那些情况中是有限的。例如，在爱因斯坦—德西特宇宙(见2.1.1节)中，$u_{ph} = 3t_0^{1/3}$，$d_{ph} = 3t_0$。那么除非我们居于一个这样的宇宙中：空间紧致部分如此之小，以至于光线确实有时间可以从宇宙的初始就穿越整个过程，否则我们就只能看到存在的东西的一小部分。这对于有着标准单连通拓扑的宇宙来说不会是真的。彭罗斯对共形方法的强有力运用，参见[Hawking and Ellis, 1973; Tipler *et al.*, 1980]，给出了视界本质的一个非常明晰的几何图景[Ellis and Williams(埃利斯和威廉姆斯), 2000]。它们可能并不存在于非弗里德曼—勒梅特宇宙，比如比安基(Bianchi)(各向异性)模型[Misner(米斯纳), 1969]中。在有着封闭空间部分的宇宙中，出现了一个补充问题：封闭的尺度是否小于视界尺度？因果联通性可能是在一个有限时间之内实现的，粒子视界的存在也就终止了。在标准的 $k = +1$ 的弗里德曼—勒梅特模型中，这只有在宇宙到达了最终奇点的时候才能发生。但是如果存在一个正宇宙学常数或者其他有效的正能量密度场，它就会发生得更早。视界总是在增长，因为式(29)表明 u_{ph} 是 t_0 的单调增长函数。尽管文献中有很多相反意见，但物质一旦进入视

① 在某些宇宙模型中还存在事件视界和表观视界[Rindler, 1956; Tipler *et al.*, 1980]和[Rindler, 2001, pp. 376—383]。

界就不可能脱离。在一个(微扰的)弗里德曼—勒梅特模型中，因果联系一旦发生，就会保持下去直到宇宙结束。

视界的重要性是双重的，它们构成了在结构和统一性的起源中相关的因果限制[Misner, 1969; Guth(古斯), 1981]，也描绘了宇宙可检验的东西的绝对极限[Ellis, 1975; 1980]。

2.4.1 因果限制

关于因果限制，视界是非常重要的，对于宇宙在大尺度下的光滑性和宇宙在小尺度下的团块结构都很重要。光滑性的问题可以在视界问题中得到概括，如果我们测量在标准弗里德曼—勒梅特模型中从天空的反方向到达这里的宇宙黑体辐射的温度，它来自最后的散射面的区域，这些区域相互之间可能从宇宙最初就没有任何因果联系。在一个标度因子为式(19)且由辐射支配的早期宇宙中，粒子视界在最后散射时的尺度显示为今天天空中一个1°的角度盘，并且对应于一个今天评估出来大约400，000光年的共动物理距离。那么为什么在这些普遍分割的区域中，条件如此相似？参见[Misner, 1968; Guth, 1981; Blau and Guth(布劳和古斯), 1987; Kolb and Turner, 1990]。注意，这个问题的本质是哲学的而不是物理学的。即是说，这里不存在与任何实验的冲突。有人宣称这个问题可以通过下面2.6节要提到的暴胀宇宙方案得到解决。

联系视界之间的存在的是这样一种预言：宇宙不同区域的物理场在对称破缺发生之后应当是无关联的，因为它们不可能已经因果地相互作用了。因此，如果大统一的理论是正确的，那么诸如磁单极子和宇宙弦之类的拓扑缺陷可能被预期为极早期宇宙膨胀的遗产[Kobl and Turner, 1990]。在标准宇宙学中，太多的磁单极子得到了预言。这也同样被暴胀解决了。

关于团块结构，这里的问题是，如果我们相信存在一个宇宙的非常光滑的态——如退耦时所表明的，通过较低程度的宇宙黑体辐射各向异性由弗里德曼—勒梅特模型的罗伯森—沃克几何表示——那么可能从那以后通过因果物理过程得到增长的结构的尺度就有限，由可用时间中万有引力引起的运动的相对速度也有限。(例如，我们的星系相对于由被称为“巨引力源”的庞大过高密度引起的宇宙黑体辐射静止框架的特殊运动)。如果存在更大尺度的结构或者更高的速度，这些必定已经印在了最后散射时刻的微扰之中，因为从那以后它们

就不能被以一种因果的方式产生。它们被放置在初始条件中而不是随着出现于一个更加统一的形式中的物理因果性而显现。早期宇宙中的暴胀抑制了微扰的增长，而这是早期宇宙微扰增长理论的一个关键因素。决定暴胀宇宙中局域因果效应的相关物理尺度的量是共动哈勃半径 $\lambda_H := (SH)^{-1}$。波长为 λ 的微扰的发展取决于 $\lambda > \lambda_H$ 还是 $\lambda < \lambda_H$[Dodelson, 2003, pp. 146—150]。

事实上，因果影响的范围甚至受到比过去光锥所表明的更强的约束：从视界大小得来的限制就是对可能受到以光速运行的粒子和力的影响的东西是什么的限制。然而只有自由穿梭的光子、无质量中微子以及引力子可以以光速运行，而这些粒子从宇宙远处而来，对我们的银河系或者太阳系影响极小(实际上我们需要非常高精度的实验才能探测到它们)。任何大质量粒子，或者与物质相互作用的无质量粒子，运行的速度都会小于光速(比如在退耦之前，光线有一个非常小的自由路径，信息只有通过声波传播，在紧密对偶的物质——辐射流中扩散)。无压标量和矢量微扰的特性是类时曲线，以相对于物质为零的速度运动。虽然有压的密度扰动可以以声速运动，但仍然只有张量扰动可以以光速运动。因此，显著影响我们的真正范围要比粒子视界小得多。正是围绕我们的过去世界线的小区域在退耦之后由共动尺度描绘了。也正是在这个尺度上，物质被合并进入了我们的银河系：目前的距离大约是 1 到 1.95Mpc,[①] 对应于一个观测到的角度，在最后的散射面上大约是 0.64 弧分度。在退耦之前，它可能受到了声音界限而不是粒子视界的限制[Dodelson, 2003, 257]。

2.4.2　观测极限

很明显，我们不能得到关于粒子视界之外发生的事情的任何观测数据。实际上，我们甚至不能看到那么远，因为退耦之前的宇宙是不透明的(我们关于宇宙的观点受到可见视界的限制，这个可见视界由我们可观测到的最远物质——即最后散射时刻从宇宙黑体辐射发出的物质——的世界线组成)。见[Ellis and Stoeger(埃利斯和斯特格), 1988]。这发生于退耦时刻 $t = t_{dec}$($z_{dec} \simeq 1100$)，因此可见视界由 $r = u_{vh}$描述。这里：

---

① 见都德尔逊[Dodelson, 2003] 第 283 页及与 W · 斯特格(W · Stoeger) 私下的交流。

$$u_{vh}=\int_{t_{dec}}^{t_0}\frac{dt}{S(t)}<u_{ph} \tag{31}$$

事实上，最后的散射面以两种方式描绘了我们的可见视界：我们不能看到比它的发生更早的时刻(因为早期的宇宙在 $t<t_{dec}$时是不透明的)，也不能探测到比我们在最后的散射面上看到的更远的物质(我们不能接收到来自共动坐标系值 $r>u_{vh}$处的物质的辐射)。我们通过测量来自比如 COBE 和 WMAP 之类的人造卫星的宇宙黑体辐射所得到的最后的散射面的图像仅仅只是组成可见视界的物质的一道风景，我们看到从辐射退耦的时刻起很远的过去。可见视界的位置取决于退耦以后的几何。可见视界确实存在，除非我们居住于一个小宇宙，在空间上封闭且封闭的尺度如此之小以至于我们正好看到退耦以后的宇宙附近。这是一种可能性，我们在下面(4.3.1 节)会进行讨论。如果存在一个早期的暴胀阶段，那么这些可见视界没有发生任何变化，因为暴胀在后来的时期中并没有影响暴胀或者零测地线。可视视界存在的主要结果——比如混沌暴胀理论(2.6 节)——并非观测可探测的，因为人们得不到任何关于视界之外存在什么的明确信息[Ellis, 1975; 1980]。这是我们在尝试检验宇宙模型的真实性(4.3节)的时候所要考虑的主要限制之一。

2.5 理论的发展

爱因斯坦引力场理论在宇宙中的应用在过去几十年里也得到了很大的进步，涉及场方程的精确解和一般特征、近似解以及二者之间关系的理解。

2.5.1 *精确解和一般特征*

理论最初预测宇宙一定会有一个起源，但很长一段时间里人们并不清楚这是否仅仅是因为宇宙非常特殊的各向同性和空间均匀几何标准的弗里德曼—勒梅特模型。很可能更现实的有着旋转和加速的模型，会表明预言是从所使用的理想模型中得到的数学作品。彭罗斯和霍金发展的奇点定理[Hawking and Ellis, 1973; Tipler *et al.*, 1980; Earman, 1999]表明这并非事实：即便对于现实的几何学，如果假定引力理论满足通常的能量条件，它也预言宇宙在时空奇点有一个开端。这项研究导致对因果性和一般宇宙模型的拓扑的理解得到了很大的深入[Tipler *et al.*, 1980]，包括奇点可能有着与在罗伯森—沃克模型中非常

不同的本质，比如说是各向异性的[Tipler *et al.*，1980]或者在本质上是非标量的[Ellis and King(埃利斯和金)，1974]。

人们了解了一些精确宇宙解，如坎陶克思—萨克斯和比安基空间均匀各向异性的模型、托尔曼(Tolman)—邦迪球对称非均匀性模型以及“瑞士奶酪”非解析模型，使得更多一般模型而不仅仅是弗里德曼—勒梅特模型的动力学和可观测行为得到了理解[Ellis and van Elst，1999a]。对动力学系统的研究[Wainwright and Ellis(温赖特和埃利斯)，1996；Uggla *et al.*（乌格拉等)，2003]关系到全部各向异性模型在适当态空间中的行为，使得对行为的一般模式(固定点、鞍点和吸引子等)的确定成为可能，因此也就对更高和更低对称的宇宙动力学之间的关系的确定成为可能。这些研究有助于理解在什么样的程度上弗里德曼—勒梅特模型在可能的宇宙模型族中是一般的、哪些模型可以给出类似于弗里德曼—勒梅特模型中给出的可观测量。特别地，它们还关系到对宇宙在极早或者极晚期的可能几何的考虑。

2.5.2 微扰理论、规范问题以及反作用

细致的微扰理论已经被发展起来，成为正在膨胀的宇宙中结构形成理论的基础，从而来考察微扰弗里德曼—勒梅特模型的动力学。这些模型中的流体可以有切变、涡度和加速度，并且外尔张量 C_{ijkl}不为零，以至于可以出现密度变化、潮汐力、奇特速度以及引力波。详细的研究运用了物质(电子、质子、暗物质)和辐射(光子、中微子)的动力学理论近似以及它们的动力学，由玻耳兹曼(Boltzmann)方程描述，参见［Dodelson，2003，Chapter 4］和[Uffink(尤非克)，2006]，与微扰弗里德曼—勒梅特度规描述的时空非均匀性相互作用。这里关键的问题是规范问题——怎样选择微扰时空的背景模型[Ellis and Stoeger，1987]。如果处理不当，就有可能得到完全规范的(它们是数学的而非物理的)明显的微扰解，结果是只通过改变坐标就可以改变表观增长率。要解决这个问题，要么仔细了解剩余规范自由度和规范的可能变换的所有环节，要么运用规范不变量(在我看来更可取)。见［Bardeen(巴丁)，1980；Ellis and Bruni(埃利斯和布鲁尼)，1989；Challinor and Lasenby(查理诺和赖森蓓)，1998]。

大多微扰理论的文献都是处理线性情况，但也有些研究解决的是非线性问题，例如[Langlois and Vernizzi(朗格卢瓦和维尔尼兹)，2005]，还有些人考虑

诸如磁场的来源和银河系自转的原因等问题。这里关键是适当地把天体物理学动力学的相对性分析与天体物理学家最常用的牛顿方法关联起来，比如 [Bothun, 1998]第 183—222 页，这并不简单。① 一个更深刻的待解决问题是引力熵的本质问题[Penrose(彭罗斯)，1989b；Ellis，2002；Penrose，2004]。物理学教科书中很多关于熵本质的陈述，当说到引力是主导型的，导致了诸如恒星和星系的自然构成的时候，都是错的。到目前为止，关于引力还没有被普遍接受的一致定义，而依赖于熵概念的宇宙学争论也站不住脚。

宇宙中非均匀性的存在提出了适应和回应的问题。在什么程度上，完全光滑的弗里德曼—勒梅特模型反映了更加现实的“粗糙”宇宙学模型的动力学本质[Ellis and Stoeger，1987]？非均匀性导致理想背景模型演化方程出现了额外的项来表示微扰在它们的动力学上的反作用[Ellis，1984]。这可能很重要，但仍然是一个有争议的问题[Ellis and Buchert，2006]。

2.6 暴胀

粒子物理学的进程控制着最早的诸如夸克胶子等离子体凝聚以产生重子的奇异进程发生的时代，那个时候量子场论的影响很大，这导致了一个重要的可能性：产生讨厌的引力效应的标量场在那个时候可能已经支配了宇宙动力学。这导致了暴胀宇宙理论，由阿兰·古斯(Alan Guth) [1981；1997]提出：如果 $\mu_{grav}=\mu+3p/c^2<0$，这在标量场支配早期宇宙动力学的情况下是可以发生的，极短时间的加速暴胀会先于热大爆炸的时代发生[Blau and Guth，1987]。因此产生了一个非常冷且光滑的真空支配态，并结束于“重新加热”：标量场转化为辐射，开始了热大爆炸的时代。人们认为这个暴胀过程可以解释上述困惑(第 2.4.1 节)，为什么宇宙如此独特(有空间均匀性和各向同性的几何以及非常均匀的物质分布)，为什么空间的部分目前如此接近于平直(我们仍然不知道空间曲率的正负)，这要求极早期的初始条件非常巧妙。暴胀解释了这些特征，因

① 某些确切的广义相对论结果，那些必然要被应用到广义相对论的牛顿极限中去，在牛顿理论中不存在对应部分。举个例子，应用于无压物质的无剪切定理[Ellis，1967]。基础的问题是，存在十个场方程都可以满足广义相对论，它们有着二十个可积分条件(比安基恒等式)，但只有一个场方程可以满足牛顿理论(泊松方程)，它有着四个守恒方程。

为在暴胀弗里德曼—勒梅特模型中粒子的范围将会比在有着通常物质的标准模型中大得多，这就允许大于可视范围尺度上物质的因果联系。暴胀还将扫除可视范围之外的拓扑缺陷。

更详细的是，在一个单一的有着常密度类空面的标量场 φ 的情形中，在选择与这些面正交的 u^a 时，应力张量有着完美的流体形式（弗里德曼—勒梅特流体形式）：

$$\mu = \frac{1}{2}\dot{\phi}^2 + V(\phi)\text{，}\ p/c^2 = \frac{1}{2}\dot{\phi}^2 - V(\phi) \tag{32}$$

因此：

$$\mu + 3p/c^2 = 2\,\dot{\phi}^2 - 2V(\phi) \tag{33}$$

缓慢转动的例子如 $\dot{\phi}^2 \ll V(\varphi)$，导致了 $\mu + p/c^2 = 2\dot{\phi}^2 \simeq 0 \Rightarrow \mu + 3p/c^2 \simeq -2\mu < 0$。这使得暴胀弗里德曼—勒梅特模型中的界限问题变得可能有解：如果宇宙早期发生了充分暴胀，那么我们从中得到的宇宙黑体辐射的所有区域都因果地联系。事实上，如果宇宙是从一个暴胀态或者有着紧致空间部分的暴胀开始的，那就可能不存在任何因果界限。暴胀模型还使得初始微扰包括速度微扰走向消亡，因此就解释了大尺度宇宙观测上的光滑性。人们曾经期望这个过程能够产生一个后来有着相当平直的空间的宇宙：

$$\Omega_0 = \Omega_{dm\,0} + \Omega_{\Lambda\,0} \simeq 1 \Leftrightarrow \Omega_k \simeq 0 \tag{34}$$

这个理论引出了一个令人高兴的想法，初始的微小量子涨落通过暴胀扩张到如此大的尺度，它们通过诸如星系团那样大尺度结构的引力提供了宇宙暴胀的种子。这个理论解释了宇宙黑体辐射各向异性范围的观测预言，热气球和人造卫星的观测，比如 WMAP[Spergal *et al.*, 2003]已经令人吃惊地证实了它们。这样，暴胀已经向我们提供了我们关于结构形成的第一种一致的理论。非均匀性在暴胀时代是以量子涨落的形式开始的，之后随着暴胀扩张到了物理尺度，但是在大于当前的哈勃尺度之后，就在幅度上保持为常数了，它导致了在热大爆炸时代之始的高斯自由尺度微扰。从这些涨落开始，冷暗物质产生势阱以便宇宙扩张的种子可以落入，但是辐射（与电子和重子紧密联结）阻止塌缩。如果波长 λ 大于金斯长度（Jean's length）λ_J，引力就胜利了，它与声速成正比[Rees, 1995; Ellis and van Elst, 1999a]。当 $\lambda < \lambda_J$ 时存在声振（声波），这些振动在退

耦时停止，这导致 λ_J 的戏剧化缩小，使得由引力不稳定性引起的结构增长进入了一种“颠倒”的方式(见 2.3.6 节)。①

一种比较流行的暴胀形式叫作混沌暴胀[Linde(林德)，1990；Guth，2001；Susskind，2005]，其中的暴胀结束于不同位置不同时间，这样就以大量嵌入一个仍然暴胀的宇宙区域“小型宇宙”(暴胀的宇宙区域，可能很像我们看到的周围区域，也可能非常不同)告终并始于不同的时间。这整体形成了一个类似分形的结构。有人认为这是似乎可行的标量场势的本质的必然结果。

暴胀并不是必然的结论，因为还有一些可选择的方案[Hollands and Wald(霍兰兹和瓦尔德)，2002；Khoury *et al.* (库利等)，2001]，WMAP 的结果在任何适当振幅的高斯尺度不变性扰动发生于热大爆炸时代之始的情况下都可以再现。然而，暴胀被大部分宇宙学家看作是对于导致目前微扰出现的热大爆炸时代之前的历史时期可提供的最好建议。但是应当注意，这是对所发生事情的一个一般提议，而并非一个特定的物理学理论。虽然人们提出了很多可能性(比如它可能是一个由高阶引力效应引起的有效场，也可能包含了多个标量场)，目前来讲所提出的暴胀场的一致性并没有建立起来，也没有与任何已知的粒子或者场联系起来。早期宇宙动力学和粒子物理学之间的期望联系只是潜在的而不是真实的[Earman and Mosterin(厄尔曼和莫斯特瑞)，1999]。关于宇宙黑体辐射各向异性和结构形成的详细研究期望能在各种各样的可能性之间做出区分。例如，检测光谱指数 n 是有着尺度不变性的值($n=1$)，还是只存在一个倾斜功率光谱($n\neq1$)。引力波的唯一光谱也会产生于暴胀宇宙的极早期，那么对于这种引力波的探测，不管是直接地用提议的引力波探测仪进行探测还是间接地对宇宙黑体辐射偏振中的相关旋度模式进行测量，都会是对暴胀的重要探测，比如决定早期宇宙的标量和张量微扰的比率 r[Dodelson，2003]。

2.7 极早期宇宙

据推测，量子引力过程支配着极早期宇宙的先前暴胀。存在着很多关于宇宙的量子起源的理论，但哪一种都没有绝对优势。问题在于我们没有很好的量

① 这是一个非常简单的说明，更详细的版本可见[Dodelson，2003；Silk，2005]

子引力理论[Rovelli(罗韦利),2006],因此所有的尝试本质上都是把已知的物理学推测至未知领域的不同提议而已。其中一个关键的问题是，量子效应是否能够排除初始奇点，不依赖起始就能够得到可能的宇宙。有些基本结果暗示这种可能性是存在的[Bojowald(波究瓦德),2001;Rovelli,2004;Mulryne *et al.*(穆利尼等),2005]。

2.7.1　存在量子引力时代吗

有个基本的问题是，在暴胀之初能否存在一个非奇性黑洞，以至于可以不需要考虑之前的量子引力时代？在暴胀时期，一个有效标量场的存在导致了对强能量条件(8)的违背，因此，乍看之下似乎在开始暴胀时代之前可能会有一个反弹，并且避免了量子引力时代的必然性。

但是一系列定理都暗示着暴胀模型不可能反弹，它们被描述为是未来无限而不是过去无限的[Guth,2001]。这是一个重要的问题，所以值得仔细追究。要达到反弹，有两个重要要求。弗里德曼方程(9)把比例因子 $S(t)$、曲率常数 k 和有效的整体能量密度 $\mu(t)$ 联系了起来。此能量密度在这个方程中的定义无关运用了什么样的动力学(多标量场、更高级数的引力或可以导出有效四维理论的更高维理论等)。① 瑞查得符里方程(7)包含了有效整体压强 $p(t)$，而它又是被这个方程定义的。在这一节中，宇宙学常数 Λ 被表示为理想流体，有着 $\mu_\Lambda + p_\Lambda/c^2 = 0$。要达到反弹，首先要求比例因子中的 $S(t)$ 是一个时间的函数，也就是：

$$\frac{\ddot{S}}{S} \geqslant 0 \Leftrightarrow \mu + 3p/c^2 < 0 \tag{35}$$

这恰好违背了强能量条件(8)。如果 $\mu + p/c^2 = 0$(真空)，那么这就是实际情况。确实，根据方程(33)，比如对于任何缓慢旋转的标量场这是可能的。第二，在比例因子为最小量的时候，还需要一个时间。因此，就必须存在一个时间 t_* 使得 $\dot{S}(t_*) = 0$。从弗里德曼方程(9)，可知：

$$\dot{S}^2(t_*) = 0 \Leftrightarrow \frac{\kappa\mu(t_*)}{3} = \frac{k}{S^2(t_*)} \tag{36}$$

---

① 各种量子引力理论得到修正的 Friedmann 方程的方法，见[Copeland *et al.*, 2005]

当 $k\leqslant 0$ 时，只有在 $\mu(t_*)<0$ 的情况下，此式才可能成立。即使如方程(32)所示存在一个标量场，则只在具有负的势能的时候才能成立，这看起来是个非物理的要求。当 $k=+1$ 时，在 $\mu(t_*)>0$ 的情况下成立[Robertson，1933]，这与普通物质是一致的。

因此，如果你想要在一个暴胀宇宙中得到反弹，关注一下 $k=+1$ 的模型是明智的。如果一个真空区域出现的时间足够长，那么情况会变得完全不同。也就是说，在我们回到过去的时候，曲率终将会取得胜利[Ellis *et al.*，2002b；2002a]。上述定理并不包含这种情况，参见[Guth，2001]，它们只考虑 $k=0$ 和 $k=-1$ 时的暴胀宇宙。应当注意到，虽然零尺度 $k=0$ 指数的例子很明显是很多人解决这个问题的方法的基础，但它是非常独特的——它在所有暴胀弗里德曼—勒梅特模型空间中是零测度的。

明确的非奇异性模型可以得到构造。最简单的是 $k=+1$ 框架中的德西特宇宙(见 2.1.1 节)，这是一个确切的永恒解，在最小半径 S_0 上反弹。这个模型有问题，它并不退出暴胀(这对应于一个确定的常数势)，但如果可能退出，就会存在不同版本。还存在可行的非奇异性模型，其遥远的过去渐近于爱因斯坦静止宇宙，并避免了对量子引力时期的需要[Ellis and Maartens，2004]。正是因为它们的遥远过去渐近于爱因斯坦的静止宇宙，这些模型才始于一种非常特殊的态，这是一种可能的情形。看起来暴胀最初有以下选择：(i)避免量子引力时期，但代价是有着特殊的("调整的")初始条件；(ii)在暴胀时期之前存在量子引力时期。因此，关键问题在于宇宙的最初是否十分特殊。

2.7.2 量子引力效应：宇宙的起源

目前想要解释宇宙起源及形成其演化的特定初始条件的努力，通常都采用了各种各样的方法来把量子理论应用于宇宙的创生[Lemaître，1931]，在这一点上有很多新颖的尝试。在本文关注广义相对论及其在宇宙学中的应用的时候，要很长的篇幅才能公平对待通向量子宇宙学的各种方法[Rovelli，2006]。我只是简单地发表一下对这些方法的看法。

通往宇宙学的完全充分的量子引力方法的尝试当然会遇到缺乏完全充分的量子引力理论以及量子力学基础问题的难题的约束，如测量问题、波函数的塌缩等等，见[Isham(艾沙姆)，1997；Dickson(迪克森)，2006；Landsman(兰兹

曼)，2006]，这些约束在实验条件下可以忽略，但是在宇宙学背景中却必须面对[Perez *et al.*(佩雷斯等)，2005]。量子宇宙学时代的各种尝试分别在量子力学某些特定的方面得到深入发展。人们可能期望这些方面出现一种能够以整体应用到宇宙的成功量子引力理论，比如要基于下列条件之一：(i)惠勒—德维特(Wheeler-Dewitt)方程和宇宙波函数的思想；(ii)某种嵌入更高维度时空的版本(由弦论所启迪)；(iii)圈量子引力的一种适当应用。事实上，它们尝试着以下努力。

(a)给出一个创生虚无的真理论[Vilenkin(维兰金)，1982]，但是这样的努力不能真正"解决"创生的问题，因为它们依赖于某些结构或者其他(举例来说，量子场论的精致框架和粒子物理学标准模型的很多东西)宇宙起源之前存在的东西，因此它们自身就需要解释。

(b)要描述一种自我维持或者自参照的宇宙，可以绕过创生的问题。

(b1)凭借凤凰宇宙的重生思想[Dicke and Peebles(迪克和皮布尔斯)，1979]，源于一个永恒的前存态，正如在韦内奇亚诺(Veneziano)基于弦理论二元性或者诸如林德的混沌暴胀模型的自我重复宇宙的类似的"前大爆炸理论"中那样；创生于某些非常不同的前存结构中的涨落(例如从德西特时空中的突现，或者开始于一个更高维时空中的前存在的"膜"之间的碰撞的"火宇宙")；或者从一个永恒静止初始态的突现。

(b2)始于一种态，这种态有着与通常不同的时间特性(或者有着时间的突现观念)：正如在哈特—霍金无边界方案[Hawking，1987；1993]和戈特(Gott)因果破坏提议[Gott and Li(戈特和李)，1998]中，宇宙"创生了自己"并且在没有封闭类时曲线的区域内开始了正常的暴胀。

这其中的任何一种都可能与其他一种方案相结合，因为：

(c)宇宙的有效集合[Tegmark，2003]要在以下状况中才能实现。

(c1)在或者是一个更大的纠缠量子实体的一部分，或者是一个单一的经典时空的一部分但是实际上又与对方并不相连的时空区域中。

(c2)在真正不相连的形式中。

但是所有这些方案都有很强的推测性，没有一种是牢固地基于已经很好地建立起来并经过检验的物理学，而且都没有在严格意义上为任何观测证据所支

持。它们都是从已知向未知的推算，可能为真也可能不为真。有一件事是确定无疑的，它们不可能都为真！

2.8 一致性模型

从热大爆炸时期而来的暴胀的思想得到了强有力的观测支持，星系的线性红移关系证明了暴胀，[①] 源的数目统计和宇宙黑体辐射的存在都证明确实存在着从一个非常热的早期时代而来的演化。测量到的轻元素丰度和早期宇宙的核合成证实了这种解释。这个基础理论对于批判性探究而言是顽强的。当前的很多行动都尝试把宇宙暴胀极早期的粒子物理相互作用与很晚之后不稳定地发生的由引力引起的结构形成联系起来。早期的宇宙暴胀的种子涨落的痕迹都可以通过当前的宇宙黑体辐射各向异性模式达到。因此现在占优势的宇宙学模型是某种接着会发生暴胀的量子引力时期；一个热大爆炸的时代；物质和辐射的退耦；然后是导致从最后的散射面上发生的宇宙暴胀的种子密度微扰向星系团的形成的引力不稳定性。

连同超新星的数据，宇宙黑体辐射角度各向异性，特别是它们峰值的分析，给出了一个这种类型的一致性模型[Bennet *et al.*, 2003；Tegmark, 2002；Tegmark *et al.*, 2004；Dodelson, 2003；Scott(斯科特), 2005]，之后就被物质聚类[Eisenstein *et al.*(爱森斯坦等), 2005a]和引力透镜的观测以及物质的大尺度运动[Silk, 2005]的统计所证明。这个模型已经被一套宇宙学参数的特定值所描述[Liddle(利德尔), 2004]，特别是：

$$\Omega_{cdm\,0}\simeq 0.3,\ \Omega_{\Lambda 0}\simeq 0.7,\ T_{cbr\,0}=2.75K,\ H_0\simeq 65\ \text{km/sec/mpc},\ t_0\simeq 1.4\times 10^{10}\text{年} \tag{37}$$

同样，$\Omega_{bar\,0}\simeq 0.044$ 是重子的密度，$\Omega_{vis\,0}\simeq 0.015$ 是发光物质的密度，$\Omega_{\nu 0}\simeq 10^{-5}$是无质量中微子的密度，这就暗示着 $\Omega_0\simeq 0.3+0.7\simeq 1$ 与暴胀预言式(34)一致。k 的正负是不确定的，但是如果所有观测的联合证据都按照表面意义去采用，那么它就应该是正的，同时 $\Omega_0\simeq 1.02\pm 0.02$[Spergel *et al.*, 2003]。正

---

① 静止宇宙中的引力红移这种可供选择的解释是不成立的，因为观测到的红移—距离关系是线性的[Ellis *et al.*, 1978]。

如上面注意到的，存在一些疑虑，第一关注的是关于年代的问题(见 2.3.2 节)；第二关注的是大角度宇宙黑体辐射各向异性(见 2.3.7 节)；第三关注的是小尺度下的冷暗物质结构形成的细节(见 2.3.6 节)。但这些问题在目前看来都不重要。

2.8.1　一些误解

尽管有其简单性，标准宇宙模型还是存在着一些常见的错误理解，参见[Lineweaver and Davis(莱恩威弗和戴维斯)，2005]，它们会导致哲学上的错误理解。

错误理解 1：宇宙是在某种东西中暴胀的。事实并非如此，它就是所有的存在。它只是变得越来越大，但一直保持它就是所有的存在。应当注意到，罗伯森—沃克宇宙可以表示为一个四维弯曲时空，在一个五维的平直嵌入时空中暴胀[Robertson，1933]，但是并没有必要把这个表达中的五维时空看作是物理上为真的。而且在我们考虑微扰以后，这种嵌入就不再可能了。为了局域嵌入一个真实的(微扰)宇宙模型需要一个十维的平直时空(通常要这样做整体上需要多得多的维度)。

错误理解 2：宇宙是从一个特定的点开始暴胀的，那个点就是暴胀中心。所有的空间点在这些宇宙中都是平等的，宇宙对于它们全部都是平等地膨胀的。每一个基本的观测者在一个确切的罗伯森—沃克几何中看到的东西都完全一样。对弗里德曼—勒梅特宇宙来说，不存在中心。

错误理解 3：物质不能以超光速远离我们。这是可以的，在某个瞬间；一个面{$t=$常数}上远离的两个基本观测者可以有一个相对速度，如果它们之间的空间距离足够大，相对速度就可能大于 c[Rothman and Ellis(罗思曼和埃利斯)，1993；Davis and Lineweaver，2004]。这不会违背狭义相对论，因为这不是局域速度差。在以这种速度分离的遥远星系之间不存在信息交换。例如，现在在我们周围存在一个以光速相对于我们后退的物质的球面，[①] 这个球面之外的物质以大于光速的速度离我们而去。当情况是如此的时候，发射宇宙黑体辐射的物质以大约 61 倍于光速的速度离我们而去[Rothman and Ellis，1993]。

---

① 这个球面与粒子视界不同，有时会被人们提到，见[Rothman and Ellis，1993]。

错误理解 4：罗伯森—沃克坐标系（在其中宇宙看起来是各向同性的）的优先存在违背了狭义相对论认为所有的参照系都是平权的思想。事实上这种坐标系的等价性为真是对于方程而不是对于它们解来说的。几乎所有的特解都会有优先的世界线和面，这只是一个对称性破缺的特例——方程解的出现比方程具有更少的对称性。这个特征是现代物理学的一个重要主题[Brading and Castellani（布拉丁和卡斯特拉尼），2006；Harvey（哈维），2006]。

错误理解 5：如果 $k = 0$ 或者 -1，空间部分就必然是无限的。这只有在它们有着它们“自然”简单地联系的拓扑时才为真。如果它们的拓扑更加复杂（比如 3-torus），它们就可能是空间无限的[Ellis，1971a；Lachiéze *et al.*（拉基等），1995]。这有很多发生的方式，确实，如果 $k = -1$，就存在着无限多的可能性。

错误理解 6：暴胀确切地暗示着空间的平直性（$k = 0 \Leftrightarrow \Omega_k = 1$）。在暴胀理论中任何东西都不能决定空间曲率的正负。暴胀宇宙在后来的时间里非常接近于平直，这与确切的平直（这要求初始条件的无限协调，如果说在 Ω_k 值中第两百万位数字在任何时间都是非零的，那么宇宙就不是空间平直的）非常不同。暴胀理论没有强大到可以暗示宇宙有着确切平直的空间部分。因此，宇宙学的一个关键问题是通过观测决定空间曲率的正负，这对于极早期宇宙[Ellis *et al.*，2002b；Ellis *et al.*，2002a]和后来的宇宙都有着潜在的动力学重要性（它决定着如果再塌缩是可能的，暗能量就应该衰减）。

2.8.2 总结

宇宙学已经由一种推测性的课题变成了一个数据驱动的科学，成为标准物理学理论的一部分[Barnett *et al.*，1996]，很多观测都支持这个重要的理论[Peebles *et al.*，1991；Silk，1997；Perkins，2005]。然而有些极早期理论不存在观测支持。有时要得到这样的支持是不可能的，对于所提出的物理学和几何学来说都是如此。因此，在某些方面，它仍然是一个原则上的提法，观测相对于理论来说是次要的。

我们现在按照围绕主要问题的一系列论题来探讨宇宙学和哲学之间的关系。人们仅仅通过简单地考虑全部问题就可以得到对整体争论的一个概观。在文末的表中会有一个总结。

3. 问题 A：宇宙的唯一性

第一个也是最基础的问题是，只存在一个宇宙[Munitz，1962；McCrea，1960；Ellis，1991]。这种研究对象的本质唯一性就把宇宙学和所有其他科学分离开了。特别地，导致我们所看到的宇宙的特殊态的唯一初始条件是以某种方式被物理学定律如我们所知的开始支配宇宙及其内容的演化的那个时间所设定的，不管那个时间是什么时候。我们不能以任何方式改变这个唯一的初始条件——它们是以一种绝对的和不可改变的方式给予我们的，虽然它们被理解为只是可能的而不是必然的。也就是说，它们可能是不同的，但仍然要与所有已知的物理学定律一致。其含义如下。

论点 A1：宇宙自身不能经受物理学实验的检验。我们不能用相同的或者改变了的条件去重新运行宇宙以看到条件如果不同的话会发生什么，所以我们不能对宇宙自身进行科学实验。

论点 A2：宇宙不能与其他宇宙进行观测对比。我们不能把宇宙与任何相似的对象进行对比，也不能用决定一个已知的物理上存在的宇宙类的统计特性的观测来检测我们关于它的假设。

在宇宙变化的思想里，所有的这些都变得与观测相关[Dodelson，2003，pp. 241，343]。早期宇宙结构形成的理论只是做出了统计预言(它不能尝试预言实际上将会形成的具体结构)。检验理论是要把我们的宇宙和一种宇宙的理论集合作对比，并且宣称在真实宇宙和基于模型集的期望特性之间存在着某种不同。如果这种不同足够小，那么对于期望值的偏离就被称作是一种统计上的偏离，即没有物理学意义——我们不需要再解释它；如果它很大，那么就需要解释。这是一个关键的问题，比如在宇宙黑体辐射各向异性的观测中就是这样[White *et al.*(怀特等)，1993；Kamionkowski and Loeb(卡米科维斯基和洛布)，1997]。WMAP 测量到的宇宙黑体辐射的功率谱在大角度范围内要小于期望值(2.3.7 节)。有个学派宣称这只是一个统计学的涨落，而另一个学派则认为它需要解释，比如说它有可能是一个小宇宙的证据[Luminet *et al.*(卢米涅等)，2003；Luminet，2005]。这种争论能够出现是因为只存在一个宇宙，并且在大角

度尺度上，只能进行很少的测量(在小角度尺度范围内，我们可以作任何测量，因此这种不确定性就变得很小)。

作为 A1 和 A2 的结果会出现下面的论点。

论点 A3：只运用到一个对象之上的“物理学定律”这个概念是可质疑的。 我们不能科学地建立可能应用到所有这一类对象中的“宇宙的定律”，因为我们不能检测任何这样建立的定律，除非运用与其中一个对象(观测到的宇宙)相一致的术语。

这是不充分的，一个观测点不能确定一个因果关系的性质。事实上，“定律”的概念在只存在一个给定的应用对象的时候，就变得令人生疑[Munitz, 1962]。物理学定律的基本思想是，它可以应用到一组对象上，所有的这些对象都有着其行为之下的相同不变性(就如定律所规定的那样)，虽然它们在特定的瞬间有着明显的变化，这些变化是定律作用于其上的系统的各种不同初始条件的结果。这种理解得到了物理学实验的检验。在这些实验中，相似系统演化的初始条件不同，在此基础上探索了一组同类对象的统计性质。这在宇宙学中不可能做到。

在宇宙中，物理学定律局域地应用于对象之上，并且在运用适当的初始或者边界条件(比如在罗伯森—沃克模型中，通过弗里德曼方程)局域地应用于所有地方的时候，就决定着宇宙作为一个整体的演化。除此之外，我们不能像在其他层次的复杂性上那样建立更高级的可应用于所有宇宙并决定它们的结构的有效定律。我们在这个结构层次上能做的只有观测和分析存在的唯一对象。麦克雷是这样表达的：“在我们说到星球结构的其他解的时候，除了当前我们感兴趣的这一个，作为能够存在的代表系统，我们的意思是它们能够如我们所了解的那样存在于宇宙之中。很明显，对于宇宙自身，任何这样的看法都不可能[McCrea, 1953]。”

由于使全域解变为一个局域解的约束同样也是一个解，我们就有着成千上万的“迷你宇宙”，在其上我们可以检验支配宇宙局域性质的定律。但是，一个迷你宇宙并非宇宙自身，它只是整体的一个小部分。通过检测这些“迷你宇宙”并且查看它们是否在本质上到处相同，我们就能够在某种程度上首先检查物理学定律在宇宙中任何地方都相同(所有宇宙学分析中的一个重要特征，参考 7.1

节)，然后检查宇宙在空间上是均匀的(这在下面要深入讨论，见 4.2.2 节)。但是第二个特征是“宇宙的定律”所必须解释的，检验均匀性并不解释为什么它就是这样一种状况，它的发生是因为特定的初始条件，有人认为这要归因于假定的“宇宙的定律”是运用到了整体上而不是运用到它的部分。

论点 A4：概率的概念在只存在一个对象的情况中是可疑的。在把概率应用于作为整体的宇宙的时候，问题出现了——还并不明确在只存在一个单独对象因而无法与其他任何存在对象进行比较的情况中这是否有意义。

但是概率的概念存在于宇宙学的很多现代争论中。比如说，对“精细调节”的讨论就基于概率(一种阐述某种东西不大可能的方式)的应用。这就假设了情况本来有可能是不同的，同时我们可以以一种不变的方式把概率分配给一系列没有实现的可能性。这里的问题是，如果确实只存在一个宇宙，这个宇宙有着一套在物理学出现之前，或者当物理学出现的时候都为固定的初始条件，那么就要解释情况在什么意义上本来有可能不同，并且对不同的理论可能性分配了确定的概率。我们不能科学地建立可能的决定这样的初始条件的宇宙创生定律以及作为结果的概率。如果我们运用贝叶斯解释，有人认为贝叶斯解释可以被有意义地运用于只存在一个对象的情况[Garrett and Coles(加勒特和科尔斯)，1993]，其结果依赖于我们的“优先知识”，那么在这种情况下就可以通过改变我们的初始前物理学假设而得到不同结果。相关的问题在提到量子宇宙学核心的“宇宙波函数”的意义的时候也会出现。这个波函数对任何特定的单一宇宙都不会给出唯一的预言。

下面对上述内容作两点评述。

第一，在实验科学——比如物理学、化学、微生物学和历史学以及地理学科学——比如天文学、地质学、进化论之间做出区分是很有用的。在科学方法的讨论中常出现于我们脑海中的是前者。在这种情形中的理解是，我们观测并对一类相同或者几乎相同的对象进行实验，然后确定它们的共性。那么问题就转至关注那些对象如何相同。夸克、质子、电子都与对方完全相同，因此有着完全相同的行为(事实上这个特征是经受住了检验的量子统计的基础)。所有的 DNA 分子，青蛙的、人类的以及生态系统的，都以某种方式不同于对方，但是却足够的相似，因此同一种宽泛的描述和定律也可以应用于它们。如果不是这

样，那么我们从一开始宣称它们属于同一类对象就是错的了。水分子、气体、固体和液体属于中间类——几乎相同，当然可以通过特定的物理学和化学定律进行可靠地描述。

关于地理和历史科学，在这里是确定地研究只发生过一次(太阳系的起源、地球上生命的进化、SN1987a 大爆炸等)的独一无二的对象(里奥格兰德、南极洲大陆、太阳系、仙女座星系等)和事件。由于其独特性，上述评论 **A1** 也可以用在这些情形中：这些独一无二的对象或者事件不能被改变，也不能对其进行实验。然而，评论 **A2** 就不能在这里应用：至少在原则上，在别处会存在类似的对象(其他河流、大陆、行星系、银河系等)或者类似的事件(其他星系的起源、其他行星系的演化、其他超新星的爆炸等)，它们都可以被观测并且与我们的特定模型作比较，也都在很多这样的情形上贯彻着统计分析，从而决定其规律性的基本模式。在这个方面，这些论题不同于宇宙学。

如果我们真的不能进行这样的分析——也就是说，如果 **A2** 在某些特定的案例中也可以应用——那么那个主体就参与了宇宙本质的这一方面。有人可能声称这里的分界线在于如果我们能够让自己相信一些大尺度的物理现象在整个宇宙中只会发生一次，那么它就应当被看作是宇宙本身的部分，但是，如果我们相信即便我们不能通过观测接近它们，它也在很多地点很多时间上发生着(如我们相信其他星系中行星也围绕着许多恒星在进行演化)，那么对于那一类对象和事件的研究就可以与宇宙本身的研究精确地区别开来，因为存在一个它们的类需要研究。

第二点评述是，有些工作者尝试通过在根本上否定宇宙的唯一性来避开这些问题。这是通过提出“多个宇宙”的真实存在而做到的。对于这些宇宙，概率概念可以适当地被应用(参考 2.7.2 节)，或者宽泛地把它设想为一个更大的在每一个区域都有着不同特性的宇宙的分离区域(比如在混沌暴胀中)，作为量子结果的多样性实现，或者设想为完全不相连的宇宙的一个整体——在它们之间没有任何物理联系——在其中所有的可能性都实现了。我们将在 9.2 节中回到这个问题上。

4. 问题 B：宇宙在空间和时间上的巨大尺度

宇宙的唯一性产生的问题由于其在空间和时间上的巨大尺度而变得更加复杂，这成为观测宇宙学的主要问题。因此我们需要在观测之外再举出各种各样的原理来得到唯一的模型：理论是作为解释观测的基础而出现的。

4.1　大尺度宇宙中的观测

我们与最近的星系之间的距离大约是 10^6 光年，也就是大约 10^{24} cm，而地球的大小大约是 10^9 cm。目前可见宇宙的尺度大约是 10^{10} 光年，也就是大约 10^{28} cm。这个相对于我们的物理尺度(大约 10^2 cm)而言的巨大尺度为我们的观测范围加诸了严重的约束(当然也阻碍了我们对它进行实验)。宇宙在这个方面的独特性就是它要面对这个尺度：是我们可以拥有因果性或者观测联系的最大限度。

论点 B1：天文学的观测都被约束在过去零锥之中，随着距离的增大而衰退。我们可以有效地只观测宇宙，在宇宙学的尺度上，从一个时空事件开始考虑。可见的观测只有在我们的过去光锥中才是可能的，所以在我们观测更远的距离时不可避免地要朝过去看。随着距离和事件的增加，不确定性也在增加。

宇宙的巨大尺度暗示我们只有从一个时空事件(“这里和现在”)开始才能有效地观测它[Ellis, 1971a; 1975]。如果我们要从这一个空间位置以几乎光速的速度离开，比如说大约 10，000 年，那么我们不会成功地离开我们的星系，更别说要接近另一个星系。如果我们要开始一个长时间的天文学实验，这个实验要存储比如说 20，000 年的数据，然后对它进行分析，那么我们观测宇宙时所处的那个时间点在本质上应当是没有改变的(因为其时长可能为 10^{10} 年数量级，额外的时间所造成的影响可以忽略)。这与地理科学非常不同：我们可以在地球上任意穿梭并且看清楚存在什么。如果宇宙要小得多的话，那么情况就会发生很大的变化。考虑到其真实尺度，我们现在能看到的星系离我们的当前距离大概是 10^9 光年，其结果是我们似乎只能从一个山顶上观测地球，而且不得不仅仅从那些观测来推断其本质[Ellis, 1975]。

由于我们只能通过以光速向我们运行的粒子——光子、无质量中微子和引力子进行观测，通过在所有波长(光学的、红外线的、紫外的、无线电的、X射线的)下操作的望远镜对远距离的源和背景辐射进行的天文学观测都被约束在我们过去光锥的射线中。这些都使得物质与我们的过去光锥发生交叉时可以对它进行精细观测(包括可见图像、光谱信息以及偏振测量)。在观测远距离区域时，我们还可以运用中微子和引力波望远镜，还有可能运用宇宙射线，这些都代表着以光速或者小于光速向我们运行而来的信息。但我们所有关于远距离区域的详细数据都是沿过去光锥收集的。

作为结果，在解释天文学观测的时候出现了三个相互交叉的问题。第一，由于我们只能从一点去观测宇宙，我们只能得到宇宙中的三维物质分布在天空的一个二维投影。要重构真正的分布，我们需要对所看到的对象进行可靠的距离测量。然而，由于源特性的变化，在校准距离的时候用的都不是可靠的标准烛光或标准尺寸对象，在这些情形中，我们必须研究源类型的统计特性以估计距离。

第二，我们必须看到远处的星系以及它们历史早期的其他对象(在那里它们的世界线与这个过去光锥相交叉)。① 这样宇宙学既是地理科学又是历史科学，两者合二为一：我们看到远处的早期时代的源，那时它们的特性有可能是不同的。当我们向过去看的时候，必须考虑源的演化，它们在发射光时的特性与它们现在的特性可能有很大的不同。我们只有理解这个演化才能确定对象的距离，但是在实践中，它正是我们试图去确定的一个未知因素。

第三，远处的源看起来非常小非常微弱，既是因为它们的物理距离，也是因为它们的光是高度红移的(因为宇宙的膨胀)。仅仅是对它们进行探测就已经随着距离的增大迅速变难，更不用说要去决定它们的性质。而且，星际间物质的吸收也会干扰来自远处对象的光。我们越往更远的过去看，问题就变得越严重，因此我们关于宇宙的可靠知识也就随着距离增加而迅速地减少了[Ellis, 1975]。

---

① 比如我们看到二百万年以前的仙女座星系，这要远远早于人类在地球上存在的时间[Silk, 2005]。

但是，地质学数据使情况得到了改善[Hoyle, 1960]。就是说，岩石、行星、星云、星系等的当前状态包含了组成那些对象的物质的大量历史信息。这样如果能够可靠地解释这些数据的话，我们就可以得到时空中非常早时期接近我们过去世界线的条件的详细信息［Ellis, 1971a; 1975］。比如说，把结构形成的理论与源特性的统计研究关联起来。

论点 B2："地质学"类型的观测可以探索我们过去世界线的遥远过去。物理学和天体物理学的观测告诉了我们关于遥远过去物质世界线附近的条件。它们还可以被用来探索更远对象的遥远过去。

这就把我们卷入了物理宇宙学，也就是对宇宙结构演化的研究。这些研究通过与天文学观测进行对比而受到检验。特别有用的是元素丰度的测量，这是热大爆炸中核合成得到的结果，可以给我们提供关于退耦之前很早时候的条件(2.2.2 节)。如果我们能够得到对象在高红移下的这种类型的足够有效的数据，那么我们就可以用这些数据去探测它们历史上很早的与我们的过去世界线相距很远时期的条件。在这个意义上的激励因素是决定高红移情况下的元素丰度的可能性[Dodelson, 2003, pp. 11—12]；[Pettini(佩提尼), 1999]。

4.2　决定时空几何：观测极限

观测宇宙学的唯一核心任务是决定存在的所有事物的大尺度几何，或者至少是我们能够观测到的所有事物的大尺度几何。

4.2.1　*直接的决定 vs 基于理论的方法*

人们可以用直接的方式去做这件事：尝试直接从观测来决定宇宙的几何(假定人们对观测到的源有某种理解)。[Kristian and Sachs, 1966; Ellis *et al.*, 1985］已经充分描述了可以做到这一点的方式(逆方法)，确实这里存在一个有趣的结果：

观测宇宙学定理：原则上从天文学观测可以得到的我们过去零锥的数据只对于要决定那个零锥上的时空几何而言是充分必要的[Ellis *et al.*, 1985]。从这个数据出发，人们在原则上可以确定零锥过去的空间时间，并且如果假设没有干涉条件，甚至可以决定其未来。

但这很难实现，因为两个原因，第一，由于估计所有被观测源的距离的问

题要求对源的性质有所了解(4.2.3 节);① 第二，由于想要获得一些必要数据(包括所有远距离对象的明显变形以及所有观测到的物质的横向速度)时的严重困难。我们越往过去光锥的远处看，不确定性就越大。这种直接的观测方法不存在假定时空几何的优先模型，在某种程度上是已为人们所追求的方法(并且大体来说是发现了大尺度结构比如巨大的壁和宇宙空洞的观测研究的基础)。但它并没有被广泛地接受为宇宙学的一种总体方法，因为这些观测困难，也因为它几乎没有解释价值。它只是告诉我们几何分布和物质分布是什么样的，而没有告诉我们它为什么具有那样的性质。

宇宙学特有的一般选择是运用一种基于理论的方法。我们先优先地假定一个基于一种有着高度对称性的时空几何的模型(通常是弗里德曼—勒梅特模型，见2.1 节)，然后与天文学观测的理论联系进行对比，决定其基本的自由参数(2.3.3 节)。然后物质分布的详细观测和大尺度速度以及宇宙黑体辐射各向异性就可以帮助我们既从统计学上即从一种天体物理学的描述[Dodelson, 2003]上，也从细节的角度即一种天文学描述[Ellis and Stoeger, 1987]上，确定其对于真实模型的偏离。

4.2.2 非直接确定：解释弗里德曼—勒梅特几何

宇宙学标准模型是弗里德曼—勒梅特宇宙模型，这类模型是确切地空间均匀性的，在各处都是各向同性的(2.1 节)。它们很容易理解，并且有着极强的解释力，而且它们的主要物理学预言(宇宙黑体辐射和宇宙早期特定轻元素的产生)似乎已经被证实。问题在于，在什么程度上观测数据只唯一地指代膨胀宇宙几何的这些宇宙模型? 这里，有人假定空间均匀性的一种足够大的平均尺度是有效的，这个尺度应当被明确地指出[Ellis, 1984]，它目前大约是 100 Mpc[Dodelson, 2003]。② 这些是宇宙学的背景模型，微扰弗里德曼—勒梅特模型描述了在小尺度上预期出现的对于确切的弗里德曼—勒梅特模型的偏离的本

---

① 观测和模型之间的关系总是要求某种理论，且从来不是直接的。

② 存在着分级模型，在其中因为它们的不规则性质，在任何尺度上都既实现不了流体近似也实现不了均匀性[de Vaucouleurs(德沃古勒), 1970]。观测到的银河系运动的规律性——正如(m, z)关系所证明的——反对这些模型，正如反对物质分布的大尺度观测一样[Peebles, 1993a]。

质(2.5.2)。

这里的关键特征是观测到的我们所处区域的各向同性(2.3.1 节)。考虑一个足够大的角尺度，源观测和背景辐射的天文学观测对我们来说是十分接近于各向同性的。事实上，后者是令人吃惊地各向同性的。由于这可以应用到所有的观测中(特别地，在其他宇宙区域内不存在较大的观测到的物质浓度)，这就在很高的精度上确立了在宇宙的可观测区域，时空结构和物质分布相对于我们都是各向同性的。我们可以很简单地构造精确的球对称宇宙模型[Bondi, 1947; Ellis and van Elst, 1999a]，正如这些观测所指示的那样。总之，它们将是空间上不均匀的，我们的星系位于中心，或者接近于中心。目前这是一个哲学上不受欢迎的观点，但当然是有可能的。问题在于除了球对称之外我们是否能够再给出空间均匀性的有说服力的观测证据。为了这个目的，人们运用了很多论据。

(a)宇宙学原理[Bondi, 1960; Weinberg, 1972]：假定空间的均匀性，因为它是最简单的情形，你不需要在当前基础上再有任何更复杂的东西。我们简单地采用一种哲学的原理作为争论的基础。这在本质上是一种宇宙初始条件的优先描述(最初有着罗伯森—沃克几何的宇宙会在后来拥有那种几何，因为初始条件的对称性被爱因斯坦保留下来了[Hawking and Ellis, 1973])，但是通常它不会被以那种方式表达出来。

(b)弗里德曼—勒梅特观测关系：如果我们能够表明源观测关系有着唯一的弗里德曼—勒梅特形式(26)和(28)来作为距离的函数，这就可能在各向同性之外确立空间的均匀性，因此也就能确立罗伯森—沃克几何[Ellis *et al.*, 1985]。这基本上就是类似在确立空间均匀性时运用计数所做的[Hubble, 1936]。然而，由于前面的论点 B1，之前提到过的观测问题——具体地说是未知的源演化——阻碍了我们完成这一任务：我们不能完全可靠地测量距离。天体物理宇宙学在原则上可以解决这个问题，但在实践上做不到。事实上真实的情形是相反的：以表面值采用辐射源的计数数据，如果不考虑源演化则与罗伯森—沃克几何相矛盾。

面对这种情况，通常的程序是假定空间均匀性以某种其他方式被人们所了解，推论出所要求的源演化，使得观测与这个几何假设相一致[Ellis, 1975]。

总是有可能找到一个能够达到这一点的源演化[Mustapha *et al.*, 1998]。这样，这种从观测上证明空间均匀性的方法的尝试就失败了。事实上，会有一种替代的解释：这个数据是空间非均匀性的证据，即我们居住在一个球对称的非同质性宇宙中，处于某个接近于中心的地方，宇宙红移在一定程度上是引力的，参见[Ellis *et al.*, 1978]。类似地，超新星数据通常被理解为暗示了宇宙学常数的存在(见2.3.5节)，也可以以这种方式解释为非均匀性的证据，这样就不再需要“暗能量”。大部分人都很不习惯这种看法——但并不能证明他们就是错的。

(c)物理学争论：人们可以宣称诸如暴胀(2.6节)之类的物理过程使得近似罗伯森—沃克的区域的存在非常可能，事实上比球对称非均匀区域的存在具有更大的可能性。这是一个切实可行的观点。但我们必须清楚这里发生了什么——我们在用一个基于可能发生也可能并没有发生过的物理过程(因为并不存在暴胀确实发生了的确切观测证据)的理论论据来替代观测检验。这得到了强力支持，因为基于暴胀宇宙学理论的详细宇宙黑体辐射各向异性形式在小尺度上的预言[Hu and Sugiyama, 1995b]已经得到了证实[Perkins, 2005]。但是如果能够表明球对称非均匀性模型(暴胀或者不暴胀)不能产生类似的各向异性形式，那论据也就只能变得严格。但是它们也许可以，因为导致了被典型预言的各向异性形式的声振事实上在暴胀之后发生了。而且如果初始条件适当，即便没有之前的暴胀阶段，它同样也会发生。

那么其他的观测方法又会如何呢？比如下面一种看法。

(d)一致的热历史：其思想是运用我们在天空中看到的物质的本质的一致性(比如，我们在大距离上看到的是相同类型的星系)推论出它们都必然经历了本质上相同的热历史，然后再由此热历史的均匀性证明宇宙必然是空间均匀的。例如，观测表明在很多方面的高红移处与局部的元素丰度是相同的，这在约束非均匀性的时候是非常有用的，这种约束是通过表明极早期宇宙在核合成时期的条件在这些方向的远处必然相同而进行的。然而，目前想要把这种思想转化为一种对同时性的适当检测还并不成功：事实上，还不清楚是否能够做到这一点，因为已经发现了一些与此猜测相反的(非常特殊的)例子[Bonnor and Ellis(伯纳和埃利斯), 1986]。不过这种方法可以被用来给出反对空间均匀性的证据：例如，如果测量到的元素丰度在任何方向上与高度红移下的情况都不同

[Pettini, 1999; Sigurdson and Furlanetto(西格森和福勒顿), 2005], 或者如果远处对象的年代与本地估计的年代不一致的话[Jain and Dev, 2005]。

最后，空间均匀性的争论被大体上接受的有以下内容。

(e)处处各向同性：如果所有的观测者都看到一个各向同性的宇宙，那么空间均匀性就可以由此得出[Walker, 1944; Ehlers, 1993; Ellis, 1971a]，事实上均匀性只要在三个空间上分离的观测者能看到各向同性的时候就可以得出了。现在我们不能从任何其他点上观测宇宙，因此我们不能从观测上确立远处的观测者也可以看到各向同性的宇宙。因此，标准的争论是假定一个哥白尼原理：我们并非优越的观测者。这在宇宙的所有可观测区域看起来都很类似的情况下似乎是可行的：我们在任何地方观测的时候都看不到条件的重大变化。结合我们看到的关于我们自身的各向同性，这就暗示了所有的观测者都看到一个各向同性的宇宙，而且这建立了一个 RW 度规[Walker, 1944; Ellis, 1971a; Hawking and Ellis, 1973]。埃勒斯、格伦(Geren)和萨克斯证明了一个很大的进步[Ehlers *et al.*, 1968; Hawking and Ellis, 1973]，它们表明如果简单地假设膨胀宇宙领域中每一个观测者周围自由传播的辐射的各向同性,① 这里的结果就可以由爱因斯坦和刘维尔(Liouville)方程得到。

EGS 定理：所有在一个膨胀宇宙区域 U 的每一点上做测地运动的基本观测者的宇宙黑体辐射的确切各向同性表明了 U 中确定的罗伯森—沃克几何。

这样我们可以通过假定一个弱哥白尼原理来确立空间均匀性。我们并非处在一个宇宙黑体辐射正好偶然高度各向同性的优越位置上，因此所有的共动观测者都可以被设想去测量宇宙黑体辐射，都可以得到高度各向同性的结果。这是目前我们在空间均匀性上最有说服力的基于观测的论点。

这里有一个问题是，假定宇宙黑体辐射的确切各向同性，这是一个确定的结果，那么这个结果是否稳固？事实上确实如此，某些区域中处处存在的一簇膨胀测地运动基本观测者自由传播的宇宙黑体辐射的几乎各向同性证明了宇宙

---

① 这个结果在一个静止宇宙中并不成立，因为在那里辐射温度只依赖于发射器和观测者之间的势差，因此辐射是处处各向同性的，即便宇宙是非均匀性的。参见[Ellis, 1978]。

几何在那个区域里是近乎罗伯森—沃克的[Stoeger *et al.*, 1995]。因而这个结果就可以应用于真实宇宙——假如我们做了哥白尼假设，其他所有的像我们一样的观测者，就如我们一样看到的是几乎各向同性的宇宙黑体辐射。这是我们目前能够做到的极限。其他时空点上对宇宙黑体辐射各向同性所做的弱检测来自桑亚维—泽利多维奇效应[Goodman(古德曼)，1995]以及宇宙黑体辐射偏振测量[Kamionkowski and Loeb，1997]，这给予该论点以广泛的支持，但并不足以对时空非均匀性有很好的限制。

观测的情况很明显。

论点B3：为宇宙确立一个罗伯森—沃克几何要依赖于貌似合理的哲学假设。空间均匀性的推论并非直接从天文学数据而来，而是因为我们对观测加诸了一个貌似合理但却无法检验的哲学原理。

上述分析的目的不是要严格地支持观测所容许的宇宙是球对称且非均匀的观点，而是要清楚地表明基于观测的最好论点的本质，根据这个论点我们能够(非常合理地)为空间均匀性的假设辩护。

接受了这个论点，下一步的问题就是，类罗伯森—沃克几何是在哪一个时空区域中确立的？我们发现的宇宙黑体辐射探索了宇宙从物质和辐射退耦的时期(在大约1100的红移下)到现在可见范围内的态。来自宇宙黑体辐射各向同性的论点可以被合理地应用在那个时代。然而，它并不必然地暗示更早或更晚时期宇宙的各向同性。因为存在着空间均匀性的各向异性微扰方式，在时间的两个方向上都不稳定，而且它们会发生在一般情况下。事实上，如果检查比安基(空间均匀但各向异性的)宇宙，运用动态系统理论的强有力工具可以表明，可以发生中间各向同性化[Wainright and Ellis(温莱特和埃利斯)，1996；Wainright *et al.*，1998]，尽管在极早期和极晚期是高度各向异性的，这样的模型还是可以在任意长时间内以任意精度模拟罗伯森—沃克几何，因此可以在错误中重现任何一组类弗里德曼—勒梅特观测。我们可以从宇宙黑体辐射各向异性测量和元素丰度的数据得到对于这些各向异性模式的当前优点的强有力限制，元素丰度是一种很有力的探测，因为(“地质学”类型的)它们能够检测元素构成时代的条件，远远早于退耦时期。但是不论这些观测限制有多么低微，各向异性模式在甚至更早或者更晚(远远晚于当前)的时间都能够起到主要作用。如果

暴胀发生了，那么这个结论就会加强，它以一种非常有效的方式淘汰了任何关于极早期宇宙各向异性和非均匀性的信息。

除了我们在什么时间可以把均匀性看作是已经确立的时间限制之外，还存在重要的空间限制。上述论点的应用并没有超出可见范围很远，因为我们没有理由相信宇宙黑体辐射在那里是高度各向同性的。事实上，如果混沌暴胀是正确的，那里的条件就不同了。

4.2.3　*确定罗伯森—沃克参数*

鉴于罗伯森—沃克几何是对大尺度可观测宇宙很好的描述，那么进一步的问题就是选择我们从所有弗里德曼—勒梅特模型上观测的特定宇宙时描述它的最佳拟合参数是什么(2.1 节)。重要的观测问题如下。

• 确定哈勃参数 H_0，它设定了观测到的宇宙的整体尺度。

• 确定密度参数 Ω_0、减速因子 q_0 和宇宙学常数 Λ(或者等价的参数密度 Ω_Λ)的三元组，这是一个特定弗里德曼—勒梅特模型的主要定义参数。宇宙黑体辐射数据、超新星的观测、深空计数、源协变函数、速度测量以及引力透镜观测可以确定这些量。

• 确定曲率 k 的正负，表明宇宙是否有封闭的空间部分，以及它是否可能在未来再次塌缩。观测的分析应当总是尝试确定这个符号，并从不假定 $k=0$ (如其经常所做的那样)[Wright(怀特), 2006]。

• 不同的参数被用来描述暗物质(2.3.6 节)和暗能量(2.3.5 节)的性质。由于它们的动力学是未知的，这也必须通过观测确定。

我们只是通过上述观测关系(2.3.3 节)，运用我们观测到的物质的类的统计分析得到了这些量的很好的估计。问题的出现是由于我们缺乏它们历史发展的充分理论。

论点 B4：解释宇宙学观测要依赖于天体物理学的理解。观测分析依赖于对描述观测到的源和我们所做的观测的不同辅助函数的估计。这引入了更多必须被观测地或者理论地确定的参数，在选择与观测相适应的特定模型时容许了大量的自由度。物理宇宙学致力于描述微扰的弗里德曼—勒梅特模型(它可以解释结构形成)而不是仅仅描述背景确切光滑的弗里德曼—勒梅特模型。这引入了更多有待确定的参数。

这里，在以直接确定背景(零阶)弗里德曼—勒梅特模型的特性为目标的方法和以确定这些模型的微扰特性为目标的方法[Tegmark, 2002]之间做出区分是非常有用的。确定背景模型参数的方法(2.1节)依赖于对所使用的远处的指示器(星系、辐射、源等)的特性的看法。它们会有自身的特性(亮度、光度、物理尺度、光谱等)和动力学演化，但这些通常都没有被很好地理解，不得不以一种参数的形式表示(例如，用参数描绘光度演化)。在每一种情况下，我们最终都假设了观测对象的天体物理学和演化历史的重要方面，它们并非宇宙学模型特有的一部分。观测到的源的统计特性也是通过参数化的函数描绘的(例如，光度函数描绘每类光度中星系的数量)。为了分析观测，这些参数化的函数必须为人所了解。这种情形是拉卡托斯(Lakatos)关于科学项目如何运转的观点的一个范例，在核心理论思想和用于检测它的数据之间有一个辅助假设带的介入[Lakatos, 1980]。这使得分析相当低地依赖于模型。这里模型只是间接地与背景模型相关——它们的解释是天体物理学而不是宇宙学的目标。因此，如果观测结果与一个特定宇宙模型不一致，人们总是可以宣称这是对辅助假设的理解缺陷而不是模型本身有缺陷[Lakatos, 1980]。

相比之下，许多估计Ω_0(在某种程度上是Λ)的方法依赖于对宇宙中非均匀性的增长和性质的研究，也就是说，它们探索微扰的弗里德曼—勒梅特模型(2.5.2节)。微扰弗里德曼—勒梅特模型的特性当然依赖于背景模型，但是引入了一整套更多的函数和参数来描述微扰[Dodelson, 2003]，比如物质的角加速度函数(或者其傅里叶变形，二维功率谱)、密度涨落的功率谱[Tegmark, 2002]，红移空间相关函数[Peebles, 1993a; Eisenstein *et al.*, 2005a]，以及速度的相关函数[Dodelson, 2003]。关联参数包括一个标量光谱指数(描绘非均匀性的物理尺度的光谱)，偏差参数b(表示星系形成是如何与非均匀性之中的密度峰值产生偏差的[Dodelson, 2003, 280])，以及初始涨落幅度Q(结构构成之源)。确定这些参数是宇宙学特有任务的一部分：要充分描绘微扰宇宙模型，我们致力于确定背景参数和描述微扰的量。这样模型的选择就依赖于描述它们的参数——什么被看作是一致的，什么是要去确定的[Liddle, 2004; Scott, 2005]。例如，标准暴胀理论预言了一个尺度不变的高斯微扰光谱。我们是要检验那个假设，还是把它视为理所当然？这出现在以下问题中：在引领统计检

验的时候，什么被"优先地"假定了？

4.2.4 一致性检验

宇宙学的一个关键问题是什么类型的观测对标准弗里德曼—勒梅特模型提供了严格检验。如果不存在能够反驳它们的观测，那么主体就可能处于不可靠的科学状况下。一种重要的此类检验是要获得对宇宙 t_0 的年龄的估计，这要依赖于 H_0，Ω_0 和 Λ，并把它们与(确立于天文物理学的背景之上的)宇宙中物体的年龄的估计进行对比：

论点 B5：宇宙学的一种严格观测检验是，宇宙的年龄必须要大于行星的年龄。宇宙年龄与行星年龄之间的差别是标准模型的一个不一致的敏感区域，因此对于确立可靠的距离尺度的极其必要，对于 H_0、行星年龄以及 Λ 的有效限制都是非常基本的。其他一致性检验有助于确证标准模型和巩固宇宙学作为实验科学的立场。

目前这些年龄问题对于局域对象来讲是可接受的，因为对我们距离尺度的估量由于 Λ 为正的数据[Perlmutter *et al.*, 1998]而得到了修正[Harris *et al.*(哈里斯等), 1998]。但是在这一方面还需要后续的警惕性，特别是高度红移物体存在着有问题的迹象[Jain and Dev, 2005]。如果这曾经被严肃对待，我们可能就不得不诉诸于球对称非均匀性模型而不是空间均匀性的模型，"大爆炸时刻"(描述宇宙的开始)就依赖于与我们的距离了[Mustapha *et al.*, 1998]。

注意，这个问题与大角度宇宙黑体辐射各向异性的情形完全不同(2.3.7节)，大角度尺度下低宇宙黑体辐射各向异性可以作为一个最后的方法被当成统计上的侥幸成功而抛弃，但在年龄问题上不能这样做。这关乎到个别具体宇宙模型的内在而不是概率上的一致性。因此，对宇宙学来说年龄问题的存在是有利的。其他的一致性检验包括：

- 表明宇宙黑体辐射温度 T_{cbr} 根据 $T_{cbr}=2.75(1+z)$ 而随红移发生变化[Meyer(迈耶), 1994]；
- 证明氦丰度在所有方向上与远距离处(高红移下)25%的初始值一致，参见[Dodelson, 2003, pp. 11—12]，还有[Pettini, 1999；Sigurdson and Furlanetto, 2005]；

• 检验对于所有的宇宙学源存在着 2% 的计数偶极平行于宇宙黑体辐射偶极[Ellis and Baldwin(埃利斯和鲍德温), 1984]。

4.3 隐藏宇宙

如果我们不是居住于一个小的宇宙中，那么进一步的基本问题就是我们可以观测到的宇宙区域是有限的，首先是因为我们不能看到比最后的散射面更早的时期，在那之前的宇宙是很难理解的(见 2.2 节)；其次是因为一个有限的时间已经流逝了，因为宇宙相对于射线已经变得很清楚，光在那个时间内只能穿行有限的距离。由于任何信号都不能超光速地向着我们而来，我们不能从星系得到任何比我们的可见视界更远的信息[Ellis and Stoeger, 1988]。我们可以观测到的最远物质是发出宇宙黑体辐射的物质(见 2.4.2)。

论点 B6：观测范围限制了我们用观测决定宇宙极大尺度几何的能力。我们只能向过去看到物质和辐射退耦的时代，因此没有更早期的直接信息。并且，除非我们居住在一个"小宇宙"中，否则宇宙中大多数物质都隐藏在可见视界之后。像它的大尺度几何之类的猜测不能得到观测的检验。在小宇宙情形下，情况就完全不同了，在那里我们可以看到宇宙中的任何物质，包括我们自身所在星系的早期。

这里的关键点是除非我们居住在一个小宇宙中，宇宙自身要比可观测宇宙大得多。存在很多星系——可能有无数个——在比视界大得多的范围之内，我们通过任何电磁辐射都不能观测到它们。而且，超出我们粒子视界的物质对我们不会有任何因果影响——远处的光可能从宇宙创生之始就开始传播了，因此这可能就是我们能够拥有因果联系的最远的物质[Rindler, 1956; Hawking and Ellis, 1973; Tipler *et al.*, 1980]。我们可以期望通过中微子或者引力子辐射天文台得到可见视界和粒子视界之间的物质的信息，但是我们得不到任何关于粒子视界之外存在什么的可靠信息。原则上我们可以因为这个视界之外的物质所施加的力而感受到其引力效应(比如，视界外的物质可能会影响视界内物质的速度，虽然我们看不到它)。由于广义相对论理论的约束方程，这是可能的，

这个方程实际上是类空面上的有效瞬时方程。[1] 然而，我们不能单纯地分析那个信号来确定视界之外什么样的物质分布是它的原因：一个特定的速度场可能是由于视界附近某个相对小的质量引起的，也有可能是离视界很远的一个大得多的质量引起的[Ellis and Sciama, 1972]。关于在非常大的尺度——也就是远远大于哈勃尺度——下可能会是什么样的条件的断言是无法证实的[Ellis, 1975]，因为我们没有任何关于远超视界之外的地方条件如何的观测证据。这种情况就像是一个蚂蚁想要从撒哈拉沙漠一个沙丘的顶上来测量世界。它的世界模型将会完全由沙丘组成——尽管在它的视界之外还存在着城市、海洋、森林、寒漠、山川等。

通常人们断言如果我们居住在一个低密度的宇宙中，并且宇宙学常数为零，那么宇宙就有着无限的空间部分。但是，这种推论只有在以下情况才可用：第一，如果宇宙在过去光锥中的类罗伯森—沃克性质在光锥之外无限远处继续为真；第二，如果空间部分有着它们的“自然”简单联系的拓扑，而且我们没有任何方式可以获得这两个条件都为真的观测证据。与此相比较，在混沌暴胀模型(2.6 节)中，宇宙在极大尺度上不会成为罗伯森—沃克几何的，这是一个确定的预言——它或多或少会由很多类罗伯森—沃克的区域组成，每一个都有着不同的参数值，各自被我们视界之外的高度非均匀的区域所分离[Linde, 1990]，整体构成了一个类不规则的结构。这个预言与之前流行的宇宙在这样的尺度上是类罗伯森—沃克的[Bondi, 1960; Weinberg, 1972]的假设(基于宇宙学原理)一样不稳定。它既不能由观测证明也不能由观测否定。相同的问题出现于一个与多宇宙理念相关的甚至更加极端的形式中。我们在下面会提到，见 9.2 节。

4.3.1 小宇宙

存在一个没有此类空间观测限制的例子。这发生于一个小宇宙，也就是在一个自身因为拓扑的原因而空间封闭的宇宙[Ellis, 1971b]出现时，并且在我们

[1] 它们在后来的任何时间在爱因斯坦场方程的解中都是有效的，因为它们从一开始就有效——初始数据必须满足约束方程——并且它们一旦被满足，约束就为动力学场方程所保留了。

自从退耦以来在宇宙周围所看到的小尺度上也是这样。因此我们能够看到存在的所有物质，还有许多发生的物体的影像[Ellis and Schreiber(埃利斯和施赖伯)，1986]。这种可能性在观测上是可以通过考察源统计、通过大角度宇宙黑体辐射各向异性中的低能观测以及通过探测宇宙黑体辐射天空中不同圆上相同的变化[Lachièze *et al.*，1995]来检验的。在观测到的宇宙黑体辐射各向异性(大角度尺度下功率的缺乏)中有着微弱的暗示表明这可能就是实际情况[Luminet *et al.*，2003；Luminet，2005]，但是这并没有稳定地确立下来。考察宇宙是否是小宇宙是一项重要任务，如果是这样，我们相对于宇宙的观测关系的本质就根本不同了。

4.4 观测到的宇宙

宇宙的可观测部分(即回到可视视界)是严格有限的，我们已经看到了其大部分。我们只能通过所有波长的电磁辐射、中微子以及引力波观测远处物体。我们已经通过所有波长的向过去散射面的电磁观测而得到了对整个天空非常完整宽阔的覆盖范围，这种观测就是电磁辐射可能观测到的极限。详细的观测(比如哈勃)对于角度和深度限制的区域都是可行的。适当波长下的详细观测已经开始了，需要探测银河后面存在着什么，银河可能掩盖了天空的一部分。在这个可观测区域中不太可能存在太多未发现的新现象，虽然确定我们已经观测到的现象的更加详细的特征(比如暗物质和暗能量的性质)是非常重要的。

论点 B7：我们在观测的完善性上已经取得了很大的进步。我们已经看到了电磁辐射能够观测到的宇宙的大部分。很可能没有多少新的天文学现象有待我们观测了。我们要确定更多的细节(由此更多地理解关于我们看到的东西)并看到更多的物体，而不是去发现很多新种类的东西。

事实上哈维特[1984]已经运用详细的天文学现象发现的多样性来估计还有多少有待发现的此类新的根本上不同的现象。

中微子和引力波在原理上允许我们回看更早的时代(分别为中微子退耦时代和量子引力时代)，但是从根本上来讲要难观测得多，更不用说在细节上要有有效的研究方向。然而后者有潜力以其他方式打开我们通往不可观测时代的道路。接近极早期宇宙方面，它们有可能给予我们意外的信息。这将可以被看

作是物理宇宙学的新特征。

5. 问题C：早期宇宙中的无约束能量

物理宇宙学出现类似问题是因为在热大爆炸早期时宇宙的发生能量在根本上是无约束的，因此我们从粒子加速器中能够得到的最高能量不能达到极早期的相关水平。在这一点上，宇宙的独一无二性在于，它是唯一的，考虑经历了这样的高能时空区域的科学，虽然要卷入巨大的时间尺度，我们还是与之有着密切的因果联系——事实上那些早期发生的事件在很大程度上决定了我们今天在周围可以看到什么。核合成中潜在的核反作用已经被很好地理解了，它们的横断面也合理地为人们所知。重子合成和夸克—胶子再结合的过程被合理地理解了，而且处于可被探测的边缘。但是早期相关的物理过程还不能通过在实验室或者基于粒子加速器的实验得到检验。定义所限制的物理视界通过地球上或者太阳系中的高能实验，把我们所希望检验的物理学方面与那些可以合理地认为不会存在这类检验的想法分离开来。

论点 C1：物理学视界限制了我们关于极早期宇宙的物理学知识。我们不能从实验上检验在宇宙极早期非常重要的物理学，因为我们在地球上的加速器中得不到它所要求的能量。我们不得不从已知的物理学推算到未知，然后再检验其含义。要做到这一点，我们要假定一些已知低能物理学的特定特征是事物在高能下如何存在的关键。我们不能用实验检验我们所做的是否正确。

注意这是独立于设定宇宙的初始条件的问题的。下面会考虑（见6.2节），这个问题出现在初始条件设定之后。宇宙是根据各种不同的物理学定律运行的。我们不能确定我们假设的物理学的有效性。与其运用已知的物理学来预测宇宙的演化，倒不如我们通过探索这些物理学在早期宇宙的含义来检验它的方案，这是我们能够监测关于高能状态下基本物理学的一些思想的唯一“实验室”［Yoshimura（吉村石），1988］，对于量子引力的情形这尤其真实。问题在于我们不能在这样做的同时也坚持执行物理宇宙学的目标。

我们对于那个时代物理学的理解必须要基于已知物理学向远超它所能检验的环境的推算。其窍门是要识别哪一些关键特征是推算中要用到的，比如，变

化的原理、破缺的对称以及相改变、对偶不变性、熵极限都是可能的选择。如果我们可以证明我们通过对相关物理学在早期宇宙上令人满意的含义所进行的猜测，通过某种适当方式进行检验，那么这就是一个引人注目的进展，但是如果这时我们只能以唯一的方式检验已经提出的物理学，那么其适当性就是有问题的。如果这个假设只解决了早期宇宙要解决的特定问题，而在其他方面无所作为，那么事实上它就没有什么解释力可言，而只是已知情形的一种可替换的(有可能是理论上适当的)描述。只有在相关物理学特定方案预言了多个(而不是一个)已确证的结果时，我们才能说它得到了明确的观测支持。比如，预言了粒子，然后在实验室中确证了它的存在，这样一个单一的假设同时解决了好几个不同的观测问题。其中有些选择在不同的理论背景下要比其他选择更好，但是我们必须区别这与它们已有的观测支持的不同。如果没有其他可检验的结果，它们就会缺乏物理学影响。举一个特别的例子，暴胀宇宙方案(2.6节)：作为早期宇宙迅速暴胀时期的基础[Guth, 1981; Gibbons *et al.*, 1983; Kolb and Turner, 1990; Guth, 1997] 的假定暴胀场并没有被确定，更不能通过实验室的实验去表明它的存在。由于这个场 ϕ 是未知的，人们可以给它分配一个任意势 $V(\phi)$，这种任意性就反映了我们无法以实验的方式确定其相关行为。可以表明，事实上任何想得到的宇宙尺度演化 $S(t)$ 都可以通过对这个势的适当选择而得到[Ellis and Madsen, 1991]，而且几乎所有想要得到的微扰谱都可以通过(可能不同的)适当选择来得到[Lidsey *et al.* (林赛等), 1997]。事实上，在每一种情形中，我们都可以通过用期望的结果重推其数学以确定所需要的势 $V(\phi)$(下见9.3.1节)。作为宇宙学目的的期望形式的这样一种理论势的数学存在，其自身并不能证明一个有着那种有效势的粒子或者场的存在。

论点C2：暴胀的未知性质意味着暴胀宇宙的方案是不完备的。遵循相关宇宙学说法暴胀理论对粒子物理学的承诺还没有实现。这只有在暴胀的性质被固定在一个实验确证了的或者粒子物理学要求其存在的特定场的时候才能成为事实。

暴胀的一个令人瞩目的成就是其预言的宇宙黑体辐射各向异性谱得到了证实，并且与物质功率谱一致[Eisenstein *et al.*, 2005a]。但是那个预言只能以从物质和射线紧密耦合时代到现在的物理学为依据，假定了早期宇宙的适当的初

始涨落谱，而不是依赖于那个谱暴胀起源的特定假设。真正的解决可能是，如果一个暴胀场的特性是从宇宙学方面得到预言，然后在实验室中得到证实的话，事实上那将被当作是理论物理学的一个丰功伟绩。但这也可能不会发生，因为这里关注的实验问题之所以出现是因为我们不能在地球上重现极早期宇宙的相关环境。

使这个问题变得很重要的一个关键关乎混沌暴胀理论(2.6节)。如上所述(见4.3节)，它的几何预言是没有通过观测证实的。不过，如果它多少是已知和已检验的基础物理学结论的必然结果的话，它仍然是一个很好的物理学预言。但情况并不是这样，这里提出的基础物理学没有得到实验检验，事实上它甚至不是唯一地被确定的或者与任何特定的物理学例子或场相联系的。它不可避免地要随弦理论[Susskind, 2005]从而遇到弦理论并非物理学确定的或者经过检验的部分问题。

6. 问题D：解释宇宙——起源的问题

物理宇宙学的唯一核心要务是同时解释宇宙为什么形成并且在大尺度上演化到今天这个高度对称的弗里德曼—勒梅特几何，以及小尺度上的结构是如何形成的?

6.1　宇宙的开始

宇宙有没有开端?如果有，其性质如何?这在上面已经讨论过了(见2.7.2小节)，但问题并没有得到解决。主要的相关问题是膨胀的过程在宇宙的生命中只发生一次还是会重复发生。第一种选择是标准模型，在其中宇宙的整体演化是一次性事件，我们看到的所有对象以及宇宙自身，都是瞬时对象，最后都会像烟花表演之后的死烟花一样燃烧殆尽。既然这样所有发生过的事情都发生在宇宙的一个膨胀相中(可能随之会有一个塌缩相，如果 $k=+1$ 并且现在的“暗能量”场在未来自行消失的话，这是可能发生的)。这种演化在时空奇点上会有一个奇性的开始，也是时间性质改变的开始，一个从之前的单一塌缩相开始的非奇性反弹，或者是从非奇性静态初始条件发生的开始[Mulryne *et al.*,

2005]。一种替代的选择是，过去发生了很多这样的相，将来还会发生更多。宇宙是一个凤凰宇宙[Dicke and Peebles, 1979]，新的膨胀相重复地出现于旧的废墟之上。尽管一个或者多个反弹的思想是一种旧思想[Tolman, 1934]，但是可能允许这种反弹行为的实际机制还并没有以一种完全令人满意的方式得到阐明。一种不同的版本是，新暴胀宇宙区域的混沌暴胀思想源自旧膨胀区域中的真空涨落，它导致了一种宇宙，这种宇宙在大尺度下有着类似分形的结构，有许多膨胀区域，它们在一个永久存在的宇宙(2.6节)中显露出不同的特性。

如上讨论(见2.7.1节)，很可能(如果宇宙有着正的空间曲率)量子引力领域可以避免，不存在宇宙的开端。但是，这可能要求特殊的初始条件[Ellis and Maartens(埃利斯和马顿斯), 2004]。如果量子引力领域确实发生了，我们不能得到是否存在一个创始事件的确切结论，因为我们不知道量子引力的本质，也不知道如何可靠地把它应用于初始条件问题所出现的宇宙学情境中。圈量子引力指出，宇宙可能是无奇点的[Bojowald, 2001]，有着反弹或者一个非奇性的开始，但是这个理论并没有得到证实。经过检验的物理学不能给出确定性的答案。也许可检验的物理学也做不到。

论点D1：初始奇点可能发生也可能没有发生过。宇宙的开端可能发生于有限的时间以前，但是可以想象各种各样的选择方案：永恒宇宙，或者其时间如我们所了解的以一种或者另一种方式存在的宇宙。我们不知道实际上发生的是哪一种，虽然量子引力的思想暗示了奇点可能可以避免。

就宇宙的性质来说，这是一个关键问题。时空奇点是件引人注目的事情，在那里宇宙(空间、时间、事件)存在开端，而所有的物理学都被破坏了，因此在科学的基础上去理解发生了什么的能力就被终结了。但是，永恒的存在也有问题，比如导致了庞加莱(Poincaré)永恒轮回的思想：任何曾经发生的事情都会在未来无数次重复发生，而且在过去已经无数次发生过了[Barrow and Tipler, 1984]。这是与无限性思想(下面会详细讨论，见9.3.2节)联系的典型问题。不清楚最终哪一种是哲学上更可取的：奇点还是永恒存在。这个决定要依赖于人们运用的期望标准(这样的标准下面会讨论到，见8.1节)。

6.2 初始条件问题

虽然初始奇点的发生之所以引人注目是因为它是物理学和时空以及物质的开端，但是它是否发生过，在某种程度上与是什么决定了宇宙的性质这个关键问题是无关的。

论点 D2：可检验的物理学解释不了初始态，因此也解释不了宇宙的特殊性质。不同的可能概率之间的选择以某种方式发生了。基本的问题是，这个选择的基础是什么。为什么宇宙是这种特定的形式而不是另一种，如果其他与物理学定律一致的形式看起来也非常可能呢？宇宙的不同可能概率选择(为什么是这一种而不是其他)的原因不能通过科学得到探索。这是一个要通过哲学或者形而上学去检验的问题。

即便字面意义上的创生不会发生，就如现在的很多看法一样，这并不能解决是什么决定了宇宙为什么以这种方式存在这一基本问题，因为它也有可能会是其他方式。如果认为是从一种以前的永恒——比如闵可夫斯基(Minkowski)空间——态演化为而来的，那么为什么那就成为了一种存在，为什么宇宙就从那个真空开始像气泡一样膨胀的时候开始，而不是在之前存在的永恒中的某个时刻开始？不管它从什么时候开始，它都可以发生得更早！有人致力于通过回溯到某种形式的永恒性或周期性的初始态来避免真正的开端，比如托尔曼的膨胀和塌缩周期系列[Tolman, 1934]，宇宙像气泡一样创生的看法形成于一个平直时空中[Tryon(特赖恩), 1973]，林德的永恒混沌暴胀[Linde, 1990]，韦内奇亚诺的从之前的一个塌缩相开始的重新膨胀[Ghosh *et al.*(高希等), 1998]，火宇宙思想[Khoury *et al.*, 2001]，以及通过“全息原理”包含了信息的基本限制的理论[Susskind and Lindesay(萨斯坎德和林德赛), 2004]。这些都没有避开根本的问题。可以说它们仅仅是推迟了面对它的时间，因为现在人们必须询问关于大爆炸膨胀相的假定优先态的起源和唯一性的所有相同问题。哈特—霍金“无边界”方案[Hawking, 1993]避免了初始奇性的问题，因为时空奇点发生了变化，因此巧妙地绕开了创生时间的问题；戈特的早期宇宙中的因果违背[Gott and Li, 1998]以不同的方式做到了同样的事情。这些想法都不能克服根本上存在的问题：为什么发生的是某一种特定的态而不是任何其他可能态？如

何就决定了是这种特定类型的宇宙才是实际上例示了的那个？这个问题不能单靠物理学来回答，除非可以表明只有一种形式的物理学是自洽的，但是各种不同的观点给出了相反的证据。

初始条件的解释是人们可以共同标记为“量子宇宙学”的一族理论的目标[Hawking，1993；Gott and Li，1998；Gibbons *et al.*，2003]，然而，正如之前讨论过的，在这里我们不可避免地遇到了宇宙的科学研究可以说些什么的限制——如果我们假定这样的研究必须牵涉用观察或者实验检验我们的理论的能力的话。任何物理学实验在这里都不会有所帮助，因为宇宙的唯一性以及在这一开端之前（在因果的意义上）没有时空存在的特征。因此，要定义一个“创生的物理学”的勇敢尝试延伸了“物理学”的意义。在开端之前（如果存在开端的话）物理学如我们所知是不适当的，我们的日常语言辜负了我们，因为时间不存在，因此我们想要考虑“开端之前”存在或者发生了什么的自然倾向是极具误导性的——当时不存在“之前”，而事实上当时也不存在“然后”！就好像当时存在“之前”和“然后”的讨论很寻常，但是尝试理解“创生”的科学在概念上是非常具有误导性的[Grunbaum（格伦鲍姆），1989]。当物理学不存在时，我们立刻就要面对检验相关因果机制的不可能性。所有实验检验都不能确定任何可能运行于原因和结果的概念受到猜疑的环境中的机制的性质。这特别出现在提出“宇宙初始条件的定律”时——因为在这里我们很显然提出了一个只有单一对象的理论。物理学定律在本性上被认为是要支配不只一个事件的，如若不然，它们就无法检验（第3节）。

6.3 特殊或者一般

宇宙的目前状态非常特别。要解释宇宙当前的大尺度各向同性和均匀性意味着要确定与初始和最终条件相关的动力学演化轨迹，然后从根本上或者解释初始条件，对此我们已经步履维艰了（见6.2节）；或者表明它们是不相关的。已出现的问题是，宇宙是否是从一个非常特殊的几何态开始的。

论点D3：宇宙的初始态可能特殊也可能一般。是否存在宇宙几何初始条件的一般性或者特殊性，这是非常关键的问题。似乎宇宙观测到的部分的初始态并非一般性的。

宇宙在几何上是特殊的这一假设潜藏在宇宙学原理中，直到20世纪60年代都被当作是宇宙学中的基本原理，即被当作是对特别初始条件的一个“解释”[Bondi, 1960; Weinberg, 1972]。之后米斯纳引入了混沌宇宙学项目[Misner, 1968]，该项目基于一种有着一般性的通过诸如黏性之类的物理过程而在后来变为各向同性的初始条件的宇宙思想，使得初始条件成为不相关的。这种各向同性化的概念后来变成了暴胀理论的核心概念，其基础假设是初始条件的“微调”是非物理的并且需要回避。然而，两种尝试都只是获得了部分成功：人们可以通过它们的过程解释物理宇宙的相当大程度的各向同性化和均匀化，但是这并非在所有的环境下都起作用。暴胀能够消除大量的各向异性[Wald, 1983]，但如果暴胀在产生如我们今天所看到的这样的宇宙上获得成功，那么非均匀性一定会受到限制，而且暴胀在解决弗里德曼—勒梅特模型视界问题上的成功——在这个模型中存在着确切的均匀性作为开始——并不会必然地在各向异性模型中重复发生。开始时过于各向异性的宇宙可能永远都不会暴胀，如果确实如此，视界问题在这样的模型中可能无法得到解决。① 只有特别特殊的态才会导致一般的热力学[Penrose, 1989a; Penrose, 2004; Wald, 2005; Carroll and Chen(卡罗尔和陈), 2005]，这在暴胀物理学中被看作为真。

暴胀只能在当初始条件以某种方式受到约束时才能保证成功。某些程度的几何特殊性在宇宙的观测区域初始时就必定已经发生了。这个特殊的领域可能在一个条件随机变化、只有孤立区域导致暴胀以及比我们所看到周围的最终领域大得多的宇宙范围中发生。如此诱人的可能还有它是一个不可检验的假设（即根本上是一种多重宇宙思想，见9.2节）。

特殊的初始条件（暴胀所要解释的）可能恰好以那种方式发生。根本的问题是我们没有关于宇宙的初始条件是特殊的还是一般的的证据，它们都有可能发生。如果我们说这些条件必须是一般的，那么我们就是在做一个哲学的断言，因为它并非是一个可证明的物理学陈述。部分问题在于我们对于可能宇宙的空

---

① 大部分暴胀的研究表明几何界限问题只在非常特殊的罗伯森—沃克几何中得到了解决，但在那些几何中不存在物理视界的问题，因为它们的假设是一开始就是空间均匀性和各向同性的。

间并没有一致的测量，看上去是特殊的还是一般的需要依赖于测量的选择。

7. 问题E：作为存在背景的宇宙

宇宙中通过确定物理学定律从而被限制的初始条件，设定了所有局域物理学的边界条件并以此为所有科学提供了环境。连同复杂系统的物质或者结构方程态的适当方程，这些确定了物理学结果的性质。宇宙学的唯一性在于它考虑的是这样的条件的起源。

7.1 定律和边界条件

物理宇宙学的一个基本假设是物理学定律在物理宇宙中的各个方面都是相同的，那些我们此时此地在实验室中确定的定律在应用到很远的地方以及很晚的时期的时候也应该是一样的(例如在红移 $z=6$ 的地方确定类星体的天体物理学)。没有这个假设，解释理论就没有牢固的根基。然而，因为上面讨论的宇宙的唯一性(见第3节)，它不像那些其他的界限清晰和重要的物理学，在宇宙学背景中定律和边界条件之间的区别变得很模糊。

论点E1：物理学定律可能依赖于宇宙的性质。我们在区别宇宙开端的宇宙学背景下的物理学定律和边界条件时有一个根本困难。有效的物理学定律可能依赖于宇宙的边界条件，甚至可能在宇宙的不同空间和时间点上有所区别。

因为我们不能以任何方式改变初始条件，就我们而言，它们是必然的而不只是可能的——因此初始条件和定律之间的基本区别消失了。一旦宇宙成为存在，区别就变得明显了——但是我们关注的是与宇宙的创生和物理学定律的确切存在相关联的“前”条件。当然，任何要区分控制那些定律的解的自然定律和边界条件的想法在这个背景中都是不可检验的。考虑到宇宙是所有物理学的唯一背景这一特征，认为宇宙可能影响局域物理学定律而不只是影响它们的初始条件的想法并不牵强[Ellis and Sciama, 1972; Ellis, 2002]。这在几十年中已经在三个具体的案例中得到了检验。

(a)变化的“常数”：如狄拉克(Dirac)在引力常数 G 的情况中提出的[Dirac, 1938]，后来又被约尔丹(Jordan)、布兰斯(Brans)和迪克深入发展的那样

[Brans and Dicke, 1961]，与宇宙膨胀相关的自然的物理学常数可能存在随时间的变化[Barrow, 2003]。这种想法必须在其余的物理学中坚持发展下去，并且应当与无量纲常数相关联，否则它们就可能只是所运用的测量单元中的伪变化，而不是一个真实的物理学变化。光速"*c*"可能不同的各种断言就进入了这个范畴[Ellis and Uzan(埃利斯和乌赞), 2005]。这种看法近期已经受到了基于量子场论和弦理论的思想的推动，这些思想暗示了许多"自然常数"事实上可能是变化的，并依赖于真空态的性质[Susskind, 2003；Freivogel *et al.*(弗瑞沃格尔等), 2005a]。此类想法在某种程度上是对观测检验开放的[Cowie and Songaila(考伊和桑盖拉), 1995；Will(威尔), 1979]，并且在已探索的情形中，它似乎并不发生于宇宙的可见区域——自然常数事实上是不变的，只有一个可能的例外：好的结构常数，被看作天文学时间尺度中微小变化的证据[Barrow, 2003]。这个问题仍在探索中。检验这种不变性有着根本的重要性，是因为宇宙学通常以物理学在宇宙的任何地方都相同这一基本规则呈现。如若不然，局域物理学就不能就物质在其他地方或其他时间的行为给予我们适当的指导，而宇宙学也会变成一个任意猜测的游戏。为了在此类变化被提出时以一种科学的方式前进，人们需要猜想这些变化的方式。这样，*G* 是常数的旧定律被确定其时间变化的新定律取代了[Brans and Dicke, 1961]。由不变的物理学定律以及相关常数决定的自然原理得到了保留。① 这样，最终这种想法是要用复杂的新定律替代简单的旧定律。这些必须被假设是不变的，不然我们就不能在科学上前进。

(b)惯性和马赫原理：可能物质的局域惯性特征由宇宙中远处物质的分布决定，因此如果宇宙不同，惯性就应该不同。这是马赫原理的复杂思想[Barbour and Pfister(巴伯和菲斯特), 1995]，它是爱因斯坦宇宙学思想的主要动力。这个思想的实际重要性和含义仍然有争议。

(c)时间之箭：物理学——因而在化学、生物学、心理学和社会学——中

① "尽管可视世界有着持续的变化和动力学，宇宙的构造的有些方面在它们不可改变的恒久状态下还是难以理解的。正是这些难以理解的不变的东西使我们的宇宙成为它之所以是的样子，并且与其他我们可能想象的世界有所区别。"见[Barrow, 2003]，第3页。

时间及其宏观之箭的存在和方向——与宇宙过去和未来的边界条件相关。基本物理学定律自身是时间对称的，因而不能解释这个特征[Davies, 1974; Ellis and Sciama, 1972; Zeh(泽), 1992; Uffink, 2006]。最近彭罗斯断言，时间之箭的存在本质上取决于有着相当特殊初始条件的宇宙[Penrose, 1989b; Penrose, 1989a; Wald, 2005]。因此，在通常物理学中看起来是不可改变的自然定律的东西(即有着给定的时间之箭的热力学第二定律)可能正好是宇宙初始和结束时特定边界条件的结果。即便基础的物理学定律相同，这可能也不会在所有的宇宙中都为真。

在每一种情形中，都会有关于深层基础的不变定律的可能性质以及那个背景中宇宙的态和因而发生的有效定律之间的关系的各种想法。在弦理论的可能"景愿"中也有这样的想法[Susskind, 2005]。但由于上面解释的原因(3 节)，这些想法本质上都不可检验：我们改变不了宇宙的边界条件，看不到发生了什么，但它们确实是各种思想持续丰富的源泉。

7.2 可替代的物理学

无论如何，对于宇宙学来说重要的结论是，要考虑如果不仅宇宙开端时的边界条件且物理学的定律都与现在所知的不同的话，那么考虑宇宙可能发生什么当然是正确的[Susskind, 2005]。

论点 E2：我们不能想当然地去看物理学定律的本质。宇宙学感兴趣的是探索假设的宇宙，其中的物理学定律与我们所居住的真实宇宙中得到的定律不同——这可以帮助我们理解为什么物理学定律如其所是地存在着(真实物理学宇宙的一个基本特征)。

在宇宙学中，我们不能认为物理学(还有化学)定律的存在和本质是毫无疑问的——这在生物学对生命的起源和演化的讨论中似乎是一个惯常的行为。这与其他科学形成了鲜明对比。在其他科学中，我们满足于把描述物质基本行为的定律的存在和本质当作是给定的和不变的。宇宙学探索感兴趣的是有着不同物理行为的宇宙特性的假定。"过去可能是什么"的考虑是一个有用的宇宙学思索，可能有助于为理解实际情况提供线索。这是宇宙学中"理想实验"效能的一种陈述。

事实上，如果人们希望探索诸如为什么生命会存在于宇宙中的问题，对于这个更大的框架——本质上，一套假定的有着许多不同特性的宇宙——的考虑必不可少(这当然不会等同于假定一套这样实际存在的宇宙，参见下面9.2节的讨论)。然而，对于在这样一套假定的所有可能或者所有可想象的宇宙中运用宇宙的任何声称的统计时，我们需要非常小心。这常常并不明确，而且如果这一套东西确实存在，而不只是假定的，或者如果它是产生明确概率(一个不可检验的想法)过程的结果，那么它们在任何情况下都只与物理学过程相关。我们可以从这类考虑中学习到可能的替换理论的本质，但是并不必然能知道它们可能发生的概率(如果那个概念有着任何真实的意义的话)。

7.3 复杂性的突现

随着宇宙的进化，复杂性演变作为一种以前不存在的新型对象——原子核、原子、恒星和星系、行星、生命、意识和思维的产物，比如书和计算机——的形式发生在局域系统中[Morowitz(莫罗维茨),2002]。在很晚的时期形成的新型的物理态，比如玻色—爱因斯坦凝聚，似乎不能离开智慧生命而存在。

论点E3：物理新颖性出现在膨胀宇宙中。以前不存在的新型物理随宇宙演化的过程而形成。它们的存在是宇宙为局域系统提供的边界条件和基础物理学所产生的可能空间都容许的。尽管它们的物理存在是新颖的，每一种新形成的东西都在它们之前的可能结构中有所预示。

在宇宙的演化中，物理的存在是新的，但是这种新颖性在物理学所允许的可能空间中必须有其先兆，以便它们可以形成。在这个意义上，真正的新颖性并不是无中生有的突现而只是被发现。宇宙就是允许这些发生的环境。导致生命存在的特征的本质及其它们的可能原因，将在9.1节中讨论。

8. 问题F：明确的哲学基础

根据上述讨论，特别是B6、C2和D2项的讨论，可以得出以下论点。

论点F1：哲学的选择必然是宇宙学理论的基础。不可避免的形而上学问题

必然出现在观测宇宙学和物理宇宙学中。为了形成理论，需要哲学的选择。

当然任何科学分析，也就是采用一种基本的科学方法、一种尝试根据因果定律尽可能简单地解释我们所看到的东西的承诺，这些根本上是物理学的，也往往存在一个哲学的基础。这在宇宙学中也很明显是真实的。但是，为了得到明确的几何模型——比如哥白尼原理，如上 4.2.2 节所述——以及确定早期宇宙的物理宇宙学应当是什么形式的，比如决定在从已知物理学推测未知物理学（第 5 节）过程中的核心物理学原理是哪一个时，我们需要更加明确的哲学。两种选择的根基都是一种宇宙学模型的令人满意的标准，它有助于决定在形成理论时应当关注哪一个特征。特别重要的是，我们的宇宙学理论范围的选择，还有好理论的标准的选择，是一个哲学决定，它将会决定其余的分析。有些宇宙学家倾向于忽视作为理论基础的哲学选择，但即便是过分简单化的或者未经检验的哲学观点也仍然是哲学观点！

8.1 理论的标准

关于一个好的科学理论的标准[Kuhn(库恩)，1977]，常常有以下四种看法。

1. 令人满意的结构：(a)内部一致性；(b)简单性(奥卡姆剃刀)；(c)美学诉求(“美的”或者“优雅的”)。

2. 内在的解释力：(a)逻辑紧密性；(b)理论的范围——统一不同的分离现象的能力；(c)关于某些明确测量的理论或者模型的可能性。

3. 外在的解释力，或者关联性：(a)与科学其他部分的联系性；(b)可延伸性——为更进一步的发展提供基础。

4. 观测和实验的支持，根据(a)可检验性：在定量和定性上都做出可检验的预言的能力；(b)确证：理论在什么程度上受到这些已有检验的支持。

特别地，是后者描述了科学理论，这里的科学理论是对比于其他类型的解释宇宙的特征以及为什么事物如其所是地存在着的理论而言的。应当注意到，这些标准在本质上是哲学的，因为它们自身不能通过任何实验证明自身的正确性。在一定程度上它们的选择是基于过去的经历以及哲学考虑。人们可以尝试构想科学理论的好的标准，但显然这些也同样需要哲学的合理性。这种计划最

终会以无限的倒退结束，除非它通过简单地接受一组特定的标准来结束于某个时期。

论点 F2：理论满意度的标准不能科学地选择或者证实。满意度的标准对于选择好的宇宙学理论来说是必要的，这些标准必须在哲学考虑的基础上进行选择。它们应当包含理论令人满意的结构的标准、内在解释力、外在解释力以及观测和实验的支持。

这里的建议是，在探索宇宙学的时候，上面提出的标准是一组可以很好地应用的标准。它们包含了那些最经常被应用的标准，见[Kuhn，1977]，对这些标准的评价还可见于[Penrose，2004；Susskind，2005]。

8.1.1 标准之间的冲突

这些标准是公认可取的。那么关键就是，一般地在研究历史科学的时候，特别是在宇宙学背景下进行研究时，它们不能全部都在同一个程度上得到满足，甚至会导致相反的结论。

论点 F3：在把标准应用到令人满意的宇宙学理论时，冲突会不可避免地出现。令人满意的宇宙学理论的哲学标准相互之间一般会有所冲突，因此人们必须在某种程度上在它们之间做出选择，这个选择将会决定最终的理论［Ellis，1991］。

最新发展的推动力已经远离了观测检验而开始接近于有着牢固理论基础的思想，事实上，有时几乎不再完全相信观测检验。目前，这种现象通过一种向目标理论结果的详细观测分析的良性运动而有所纠正，使主体变得成熟。然而，由于所有的限制都是观测和检验方面的，即标准(4)，在宇宙学背景中我们仍然要高度信任其他标准，有些在大多数科学中很重要的标准可能并不真的有意义。这对于2(c)来讲特别如此，如上所讨论的(见第3节)。但是很多方法仍然给予了概率思想以重要的权重。在最低限度上，需要考察和解释才能知道这是如何有意义的。而且，有些标准的歧义会产生争议。1(b)就是一个很明显的例子：比如，一个存在的宇宙集是简单地(因为它是一个可以简短陈述的思想)还是非常复杂地(因为那个陈述隐藏了无限可能性中包含的所有复杂性和模糊性)显示了所有的可能行为？1(c)也同样是有争议性的(“每个人的审美观都有所不同”)，具体讨论可见[Susskind，2005]。

科学理解的进程可能会发生变化，从而打破什么被认为是好的解释和什么被认为是不好的解释之间的平衡。举一个例子[Ellis, 1990]，宇宙学在20世纪20年代强烈地抗拒演化宇宙的思想。那时生物学进化已经很好地确立了，但是大陆漂移的思想还是受到强烈抗拒。后来这两种理论都变得开始欣赏演化模型的解释力。但即便如此，在当时，在宇宙学情形中，不管是因为美学的考虑还是因为形而上学的原因，有些人仍然追求一种稳态的描述，反对宇宙有一个开端的思想。直到今天，在各种永恒形式的模型(比如某些形式的混沌膨胀)中，还存在着这种倾向。另外一个例子是，从对根本秩序的猜想向对一般混乱条件的广泛猜想的变化在宇宙学原理的思想中得到表达。与此相联系的，是一种从做出几何假设向提供物理学解释模型的转移。正是这种转移构成了目前暴胀理论的主要支撑的基础。

论点 F4：相信暴胀物理学的理由是它关于宇宙中结构增长的解释力。暴胀预言了早期宇宙中的高斯无尺度微扰，因而(假如冷暗物质出现)以一种令人满意的方式解释了基础结构形式。这个理论曾经通过宇宙黑体辐射的观测值和物质功率谱得到惊人的证明。正是这种解释力使得它如此受到物理学家的肯定，即便其基础的物理学既没有被完全确定也没有经过检验，并且其大尺度观测预言仍然无法检验。

就结构形成来说，由涨落谱上的观测数据所支持的暴胀的物理解释力，是很强大的。对于大多数物理学家来说，这优越于基础物理学在鉴别和实验证明方面的缺乏(第5节)。暴胀提供了一个因果模型，为宇宙学可以解释什么，如标准2(b)提供了更广泛的现象，而不只是假定了有着明确限定的形式的初始数据。对平直性(如暴胀所预言的 $\Omega_0 \simeq 1$)和均匀性的解释为此提供了更多的支持，即便这些只是哲学的而不是物理学的问题(它们不与任何物理学定律发生矛盾，事情可能一直就是那个样子的)。然而，这个模型对远离视界之外发生了什么(正如在混沌膨胀理论中那样)的断言产生于区分理论在观测和实验检验的可能性上的优先次序[Earman and Mosterin, 1999]，将永远都不可能证明这些断言是正确的。

8.2　宇宙学的范围

为了理智地选择上面所讨论标准的特性，我们需要回答一个问题：我们应当尝试解释多少？

论点 F5：宇宙学理论会有一个或宽或窄的调查范围。我们的宇宙学理论所设想的范围决定了我们所要寻求解答的问题。在构造我们的理论时，按照我们根据物理学、几何学和基础的基本因果性想要以多高的目标去解释理论，宇宙学的哲学基础变得几乎是起着统治作用了。

这是人们必须要做出的一个选择，既是关于基础的，也是关于结论的。假如在这里做出了决定，人们就能明智地争论在研究一个那样范围的宇宙学理论时可以采用的适当哲学立场是什么。对于宇宙膨胀和核合成时代到现在的结构形成的研究是非常重要的且为人们所熟悉的。与之适应的哲学立场是极少的，但也必须是非常合理的。对其早期时代的物理过程的理解，回溯到量子引力，就没有那么明确了。其哲学的立场就更加重要也更有争议。量子引力时代的发展是高度推测性的，这里所采用的哲学立场是非常重要的，因为这个理论缺乏实验和观测的限制。

人们可以选择自己在什么程度上会继续对更早的时间、更基本的因果问题的起源的研究[Fabian(法比安)，1989]，因而也就选择了在什么程度上特定的哲学选择在他的理论中起着举足轻重的作用。基本的宇宙学问题在于[Ellis，1991]。

1. 为什么物理学定律有着它们所拥有的形式？出现的问题有，比如使得特定的定律起作用的是什么？例如，是什么保证了质子的行为？是万有引力的牵引吗？是什么使得一组物理学定律而不是另一组“突出出来”？如果，比如有人把一种宇宙学理论建立在弦理论的基础上[Susskind，2005]，那么是谁或者什么决定了量子引力会有着由弦理论恰当描述的性质？如果人们考虑所有的可能性，那么单独考虑弦理论就意味着很多约束。

2. 为什么边界条件有着它们所拥有的形式？这里的关键(见 6.2 节)是如何在各种可能性之间做出特定的可能选择，比如宇宙是否存在一个起源？

3. 为什么物理学定律存在？这涉及与物理学定律的本质相关的未决问题：

它们是描述性的还是约定俗成的(第9.3.3节)? 物质的本质是否真的在某些意义上是基于数学的，或者说它是否只是在其行为可以以数学方式描述的时候才发生?

4. 为什么存在的东西存在? 不管我们采用什么方法，这种深奥的有关存在的问题都是一个谜。①

最后这些冒险的问题还包括更加深奥的有争议的有关人类的问题，参见[Carr and Rees，1979; Davies，1982; Barrow and Tipler，1984; Tegmark，1998; Susskind，2005]。

5. 为什么宇宙允许有智慧的生命存在? 这与其他问题略微有些区别，在很大程度上要依赖于其他问题，但是对于要产生的大量独立存在的争论来说，它很重要。

所有这些问题都是哲学的而非科学的，因为它们不能纯粹在科学上去解决。它们中有多少(如果有的话)是我们在建构和评价宇宙学理论的时候应当考虑的?

一种选择是，决定以一种严格的科学方式处理宇宙学，排除上述所有的问题，因为它们不能被科学地解决。有人以一种稳定的技术性主题作为结束。这种技术性主题在定义上排除了上述哲学问题。这是一种一致的和逻辑上切实可行的选择。但这种逻辑上牢不可破的立场几乎没有什么解释力，因此也最容易由于前面所说的标准2(b)和3被否定。

第二种选择是要确定这些问题如此地有意义和重要，因此人们要解决其中的部分或者全部，即便这会导致超出严格的科学范围之外。正是在这里，标准2和3在某种程度上与标准4有所冲突。因此，如果我们试着解释宇宙自身的起源，这些哲学的选择就变得很重要了，确切地因为对理论的实验和观测限制比较弱，这通过审视目前存在的提议的多样性就可以看出来。

8.3 表象的限制和实在的知识

从上述讨论可以得出，科学方法在解释中可以实现些什么是有限度的。我

---

① 但是可以在格伦鲍姆[2004]那里看到不同观点。

们需要尊重这个限度，并且在争论结论是基于某种哲学立场而不是纯粹地基于可检验的科学论据的时候明确地承认它。如果我们承认这一点并且使这种立场明确化，那么不同观点的基础就会很明确，可以理性地提出其他备选方案。

这里有一个关键的基本特征涉及从认识论到本体论的本质：我们如何把证据和我们的存在理论关联起来？一个更加关键的问题是模型和实在之间的关系：

论点 F6：不管是观测还是理论模型都不能完全地反映实在。问题出现于认识论（知识的理论）与本体论（存在的本质）之间的混淆：存在并不一直明确地展现于可得到的证据中。我们作为理解的基础所使用的实在理论和模型必然是实在的真正本质的部分和不完备的反映，这在很多方面都很有帮助，但是也不可避免地在其他方面有所误导。它们不应当与实在本身混淆起来！

认识论和本体论之间的混淆一直都在发生，比如说逻辑实证主义和极端相对主义的错误都是以此为基础的。特别地，不能假定某些实体缺乏存在的证据就是其不存在的证据。在宇宙学中，很明显，比如可能存在这样一个区域，我们从这里得不到任何证据（因为视界的存在），但是我们有时可能从一个可得到证据的声学外推（没有人相信物质会结束于可见视界或正好超越视界的地方）合理地推论出不可见物质或区域的存在。然而，必须小心另外一种极端，即假定存在是可以一直被设想的，因为不管是否存在证据，有些理论就是这么认为的。这在当今的宇宙学中，比如在对多元宇宙的情况的演示中时有发生，虽然基础物理学还没有得到实验证实。可能有人会提出，忽略了对理论的实验、观测证实的需要的那些争论最终出现是因为这些理论与实在被人们混淆了，或者至少它被当作了实在的完全可信的完整表示。这发生于以下情况中：

- 在计算机对实在的模仿实际上仅仅可以表示实际存在物的一个高度简化和程式化的版本时，混淆了它与实在本身；
- 混淆了物理学定律本身与它们的抽象数学表示（如果它们确实在本体论上为真，参见 10.1 节）或者混淆了人类思想（“物理学定律”）的建构和可测物质的可靠行为（如果它们并非本体论上为真的话）；
- 混淆了模型的基于理论的结论和已证明的结论（例如，宣称宇宙必然会有平直的空间部分 $\Omega_0 = 1$，这样就可以想当然地认为，Ω_0 的值能够而且应当在

观测上被决定，因为它可以检验那个预言）。

任何模型(字面上的、直观的或者科学的)都不能完美地反映实在。这些模型在它们表示了什么上往往是有选择性的，在完备性上则是有所欠缺的。唯一能够完全反映实在的模型是实在本身理想的和完全精细的复制品！这种对模型和理论的限制的理解并不会减少这些模型的用处，相反它有助于我们正确地使用它们。当我们考虑自然定律如何可能与宇宙本身的起源、膨胀宇宙中生命的存在和本质相关起来的时候，这特别重要。这种即便在我们的理论还没有经过检验时也完全依赖于它们的倾向有时候似乎会出现，因为我们相信它们等同于实在——而它们最多只是实在的描述。

9. 关键问题

存在一些相互关联的问题，在其中上述讨论的特征或者出于当前争论的中心，或者可能出现于未来争论的中心。它们是：宇宙条件允许生命存在的原因(有关人类的问题)、紧密相关的关于多元宇宙的可能存在的问题、存在的本质，包括无限的存在和物理学定律的本质的问题。我们在这一部分中依次对它们进行审视。在某种程度上，它们在前面都已经被考虑过了，但是在这里还需要特别地强调，因为在后面的讨论中它们可能还要扮演重要的角色。

9.1 问题G：有关人类的问题：生命的精细调节

宇宙学中最深刻的基础问题之一是有关人类的问题，参见[Davies，1982；Barrow and Tipler，1984；Earman，1987；Fabian，1989；Davies，1987；Balashov，1991；Rees，1999；Rees，2003；Barrow，2003]。宇宙为什么恰恰会有生命的存在所要求的这种特定的性质？重点是，要使生命成为可能，会要求许多“精细调谐”。物理学定律中有许多没有被物理学解释的关系，但它们却是使生命成为可能所必需的。特别地，如果我们所了解的生命要存在，各种基本常数的值是高度约束的：“一个对生命友好的宇宙——我们可能称之为热爱生命的宇宙——在很多方面必须是特殊的；许多方法会导致失败的宇宙，比如没有原子、没有化学、没有行星或者导致生存时间太短的宇宙或者过于空旷而不能演

化的宇宙[Rees，2003]。"

宇宙允许智慧生物在任何时间任何地点的存在和演化，这是如何发生的？"宇宙的什么特征对于像我们一样的生物来说是必不可少的，是偶然的还是某些更深刻的原因使得我们的宇宙具有这些特征？"。参见[Gribbin and Rees，1991]。如上所述(8.2节)，人们是否把这看作宇宙学的一个适当问题来讨论，要依赖于人们为宇宙学所设想的范围。这里所采用的观点可能是，这是人们可能想要解释的重要问题之一，事实上有一篇重要的文献专门探讨了这个问题。这里我们探索这种精细调节的本质，然后考虑它是如何出现的可能答案。我们要考虑三个方面，详细内容可参考[Susskind，2005]。

9.1.1　物理学定律和复杂性的存在

物理学和化学的定律是，比如承认活体细胞、个体、不可思议的复杂性和变化的生态系统的功能，正是这些使得演化成为可能。需要解释的是，为什么物理学定律是这样的，比如，承认这些复杂功能能够起作用，如果没有这些就不会发生任何演化。我们可以设想一个物理学(还有化学)定律在那里与我们的定律不相同的宇宙。这些定律中几乎任何变化都会阻碍从功能去了解生命。

第一个要求是保证可以作为生命存在的基础规则的物理学定律的存在。这些定律如我们所知是基于变化和对称性原理的。我们不知道是否其他类型的定律能够产生复杂性。如果定律在广泛的意义上是我们目前所认为的样子，那么为了我们所知的一般类型的生命存在，就特别需要下面的几点是正确的[Davies，1982；Gribbin and Bees，1991]。

- 通过泡利不相容原理使物质稳固并且承认化学存在的量子化。
- 中子—质子质量的差别必须是高度约束的。如果中子质量比它现在小一点点，那么质子衰变就可能已经发生，以至于到现在根本没有原子能存留下来[Davies，1982]。
- 电子—质子电量相同是为了防止巨大的静电力压倒更弱的支配化学的电磁力。
- 强烈的核力必须足够强以保证稳固的原子核能存在[Davies，1982]。事实上复杂的物质只有在原子核的强相互作用力的特性处于一个比电磁力更牢固约束的范围时才能存在[Davies，1982]。

• 人类身体所依赖的化学涉及错综复杂的折叠和黏结的方式。如果好的结构常数(它控制着化学黏结的性质)有一点点不同，这些方式就会受到破坏。

• 为了复杂性得以存在，大空间维度的数目 D 只能是 3 [Tegmark，2003；Rees，2003]。

霍根检验过粒子物理学标准模型的参数的自由度并得出结论，如果复杂性结构要存在，标准模型的十七个自由参数中的五个必须处于一个高度约束的领域[Hogan，2003]。这当然是把粒子物理学标准模型的基本性质看作是理所当然的；否则，就很难确定这些约束会是什么。然而，他的研究足以表明不管基础物理学特别是粒子物理学的本质是什么，在所有可能的物理学定律中，也许只有很少的一部分能够与复杂性的存在相容。

9.1.2 物理学定律和适宜环境的存在

通过天体物理学过程产生生命存在所需要的适当栖身之所(比如存在围绕稳定恒星的行星)的这种创造在某种程度上依赖于基础物理学定律的本质。如果这些定律在广泛意义上是我们现在认为它们所应该是的那样，那么这样的栖身之所要存在，要求包括以下内容。

• 引力必须产生能够供生命及其能源各自栖居的巨大稳定结构(行星和恒星)。这就要求引力相对于电力要弱很多。电磁力和引力的比率$\mathcal{N}$必须接近于观测值$\mathcal{N} \simeq 10^{36}$[Rees，1999，Chapter 3]。

• 弱力必须允许氦的合成过程能够留下足够的氢。它通过一个系数 10^{-11} 与引力相联系。为了使其有效，中子—质子质量的差别必须接近于电子质量[Davies，1982]。

• 星球的平衡应当允许恒星(比如太阳)的寿命很长，因此也就允许轻元素向重元素的变化。这要求核聚变效能$\mathcal{E}$要接近于观测值$\mathcal{E} \simeq 0.007$ [Rees，1999，Chapter 4]。

• 在通过恒星的核相互作用制造重元素时，人们需要克服铍“瓶颈”[Gribbin and Rees，1991；Susskind，2005]。恒星中碳和氧的产生要求仔细设置两种不同的核能水平以提供共振。如果这些水平有一点点不同，我们需要用来构成生命的元素就不会存在[Fabian，1989]。事实上正是在这个基础上霍伊尔预言了著名的碳 12 能级，后来得到了实验的确证。

• 为了构成在空间中传播重元素的超新星爆发的基础，人们需要某种类似于中微子和有着特定对偶常数的弱相互作用的存在，正如行星形成所追求的那样[Gribbin and Rees，1991]。

• 核力必须足够弱以使双质子不能存在，否则就不会有质子来使得重元素可以存在[Davies，1982]。

7. 中微子质量必须不能太大，否则宇宙就不能有足够长的寿命[Davies，1982]。

9.1.3 宇宙学边界/初始条件与适宜条件

最后，假定物理学定律按照前面两节的要求来说是适当的，那么宇宙本身按照它适宜生命存在的初始或者边界条件也必然是适当的。如果物理学定律基本上与我们现在相信的一样，这些宇宙学的要求就包括以下内容。

• 宇宙的尺度及其年龄必须足够大。可能存在膨胀然后又塌缩的宇宙，其总体寿命只有 10，000 年。我们需要一个足够老的宇宙，第二代恒星才能存在，然后才会存在有着足够长稳定生命的行星，演化才有可能导致智慧生命的出现。所以宇宙必须有 150 亿年的寿命来保证生命的存在[Gribbin and Rees，1991]，从而我们必须有 $\Omega_{matter} \simeq 0.3$[Rees，1999，Chapter 6]。

• 宇宙学常数的大小必须不能太大，否则星系就不能形成。为了星系的形成，我们需要 $|\Omega_{\Lambda}| < 1$，参见[Rees，1999，Chapter 7]和[Susskind，2005]。

• 那些存在于早期宇宙中，稍后会成为银河系中的涨落的种子必须大小合适，以使结构形成而不会塌缩为黑洞：因此描述了最后的散射面上的原始涨落(因此还有微扰宇宙模型的几何，见 2.5.2 节)的数字 Q 必须有 $Q \simeq 10^{-5}$的量级[Rees，1999，Chapter 8]。

如果一系列脆弱的条件得不到保持，人类身体中相互作用系统的复杂性有可能就不能运行。例如，背景辐射可能永远都不会低于 3000 K，因此物质总是被电离的(电子和原子核总是保持和对方分离)，生命的分子也就永远无法形成。黑洞有可能变得如此普遍以至于它们迅速地吸引了宇宙中所有的物质，永远也不会存在一个可供生命发展的稳定环境。宇宙射线有可能永远都大量存在，任何尝试性的组织结构都会在它们得以复制之前就被毁灭。

• 局域系统必须是不相干的。局域性的概念是基本的，允许局域系统相对

于宇宙其余部分的复杂结构独立有效地起作用。我们需要其中的宇宙和星系很大程度上为空的，引力波和潮汐力要足够弱，① 以便局域系统可以在很大程度上以一种孤立的方式起作用[Ellis，2002]。

• 夜晚的天空是黑暗的，参见"奥伯斯"佯谬 [Bondi，1960；Harrison，2000]，这是宇宙膨胀和重子吸收光子的结果。这个特征是生命存在的必要条件，地球上的生物圈通过把废弃的能量处理给黑夜天空的冷源而运行[Penrose，1989b]。因此，一种解释为什么天空在夜晚看到的是黑暗的的方法是，如果不是这样，我们就不可能在这里观测它。

• 时间之箭的存在，以及像热力学第二定律那样的定律的存在，对于演化和意识来说可能是必需的。这依赖于宇宙初始和结束时的边界条件(见7.1节)。

• 可能要求经典区域从量子态中的凸显。极早期的宇宙应当是一个量子物理学可以支配的领域，导致了完全的不确定性并且不能预言任何初始态的结果。我们需要这个可以演化为一个态，在其中经典物理学会导致规律的特性和允许秩序出现的可预言性。

• 行星上的物理条件必须足够长时间地处于半平衡态，以允许我们能够存在的脆弱平衡在经过极其缓慢的演化过程后得以实现。

因此，适当的局域系统作为生命的栖居之所的存在严格地依赖于遥远距离物质的大尺度特性。这提供了一个稳固的局域环境，在其中生命得以发展。

9.1.4 总体上的精细调节

因此，宇宙中的条件能够阻止生命的发生有许多方式。只有在下列条件下生命才可以发生：存在重元素；存在充分的时间使得高级生命形式的演化可以发生；宇宙中存在不过于热也不过于冷的区域；存在支配化学和局域物理学的基本常数的精确限定值，等等。这些条件在一般的宇宙中不会为真。总之，有如下几个论点。

论点G1：生命是可能的，因为物理学定律和宇宙的边界条件都具有非常特殊的性质。只有特定的物理学定律和宇宙中特殊的初始条件，才允许我们所

① 因此外尔张量 C_{abcd} 必须处处适当地小，大概地说明了一个几乎是罗伯森—沃克的几何，参见[Stoeger *et al.*，1995]。

知道的这一类智慧生命存在。如果这些定律和条件没有这种严格的形式，那么就不会有任何一种使生命成为可能的演化过程。

为什么是这样呢？应当注意到，这里我们只能有意义地提到“生命如我们所了解的那样”。反复发生的问题之一是，是否会存在其他非常不同的生命基础。如果你愿意，你可以推测生命可能以某种非物质的形式存在，或者只是基于轻元素，或者不需要恒星或行星来提供一个可能的居所就可以深刻地存在于空的空间中。人类学历史是基于这不可行的假定，但是在这一点上我们不能证明任何事情。我们对生命在什么样的基础上才可能成为存在，而不是对周围的生命中所看到的广泛原理一无所知。我们所理解的生命的基本原理很大程度上要求复杂的系统性来使它可以实现复杂多样的功能。据我们所知，这些功能只能取决于有着信息存储、能量运用、感知外部世界等的物质的存在，它们要求最小的重元素(比如碳、氮、氧、磷等)、长期的能量来源(比如来自太阳的能流)以及一个稳定的环境(比如一个行星的表面)。当我们抛弃了理解的这个基础——不提供任何可能结构的想法就说“是的，但是其他形式的生命可能存在”——那就进入了一个推测的无用范围。它似乎没有提供任何前进的有用方法。

9.1.4.1　弱人择原理

对于人类学问题，有两种纯粹的科学方法。[①] 第一种是弱人择原理，基于以下观点：观测到的宇宙允许生命的存在。这并不奇怪，因为除非宇宙中存在观测者，否则就不能被观测到[Barrow and Tipler，1984；Balashov，1991]。这个表面上看起来无内容的陈述在我们转换方向并且追问以下问题的时候，就获得了内容：在宇宙的什么时间什么地点生命可以存在，对于它的存在，决定性的相互联系又是什么？比如说，它不能在当前膨胀态的太早时期存在，因为那样的话夜晚的天空就会过热。而且，可以推论基本量之间各种必要的关系，以使得观测者应当存在(比如上面提到的)，因此，如果基本常数随着宇宙中的时

① 我省略了所谓的最后的人择原理(FAP)，它主张智慧生命必然会演化然后保持其存在直到宇宙结束，因为我觉得它不值得作为一个科学思想进行严格讨论。事实上，它导致了一篇对完全可笑的人择原理(CRAP)的著名书评[Gardner(加德纳)，1986]。

间和地点的变换而变化，那么生命就只有在约束的区域才成为可能，在约束区域中，它们可以取适当的与人类有关的值。

因此，这种观点基本上解释了人择原理作为一种选择原理：观测者要存在的必要条件限制了宇宙可以被观测到的时间和空间。因为非常有可能在一个任意选择的宇宙中，条件使得生命不能存在于任何地方，这也经常与上述9.2节中讨论的多元宇宙存在的思想相联系。这是一种有趣的也通常是启发性的观点。例如，不管是混沌暴胀宇宙思想(2.6节)还是任何其他多元宇宙思想都不能运行，除非我们在它们的解释中增加这样一种与人类相关的成分来解释为什么我们从生命存在于何处的观点来观测宇宙。现在有些物理学家用它来解释宇宙学常数的低值(量子场论预言它应当有一个比观测所得的大得多的值，见9.2.5节)，并且出现于弦理论图景的可能性的背景中[Susskind，2005]。

9.1.4.2 强人择原理

对比之下，强人择原理［Barrow and Tipler，1984；Balashov，1991］声称宇宙中存在智慧生命是必然的。为了宇宙模型有意义，生命的存在是必需的。这明显是一个有争议的主张，因为很难提供科学的理由来支持这种观点。人们可以说已尝试过的最有力的辩护要凭借以下主张：为了量子理论能够有意义，观测者的存在是必需的。但是，这种辩护的基础是量子理论的很多不同解释之一，这些量子基础的本质是有争议的，且并未得到解决[Isham，1997；Dickson，2006；Landsmann，2006]。

而且，如果我们要假定这种辩护是对的，下一步就是要问：为什么宇宙无论如何需要量子力学？论据只有在我们能够证明量子力学对每一个自洽的宇宙都是完全必要时的情况下才是完备的。但是目前这种理解的方向还不完备，特别因为量子力学自身不是一个完全自洽的理论。除了在其基础上的由未解决的测量问题引起的概念问题之外[Isham，1997]，它还遭受发散问题的威胁。对于发散问题，我们可以用自己的方式绕开它们来计算我们需要什么，但是不能消除它们。强人择原理没有已经被人们接受的物理学基础，并且还提出了悬而未决的哲学问题[Earman，1987]，因而在这里我就不进一步讨论了。

9.1.5 与基本物理学理论的关系

许多物理学家走得更远，拒绝任何与人类相关的形式的推理。他们把它看

作是一种物理学理论在不能给出所需结果时才采用的一种站不住脚的说法，他们要追求单从物理学就得到完整解决[Scott，2005；Susskind，2005]的理论。一种可能性是，存在一种包含一切的基础理论，可以完全决定物理学的本质，不需要任何任意的参数，而这样一种仍待发现的理论碰巧要有这样一种本质，比如允许生命的存在。

但是在这种情况下，人类问题报复性地回归了。这样一种比如基于粒子物理学的变化的原理和特定不变群的理论，如何能够正好导致热爱生命的参数值？没有一种明确的方式可以回答这样的问题。基本物理学的唯一性只有在以生成一个更深奥的谜题作为代价的时候才能解决参数的自由度，没有任何清晰的解决方式。事实上，统一的基本力的本质注定是允许甚至是鼓励生命的存在的。但是并没有明确的理由说为什么就应当是这样。

第二种可能性是，物理学允许许多有着不同参数的有效理论——某种形式的多元宇宙，就比如可以由弦理论所暗示的[Susskind，2003；Freivogel *et al.*，2005a；Susskind 2005]。如果这些不同的选择都等同为真，那么生命可以发生，因为在某些情况下参数将会处于约束的热爱生命的规则中。这样，从这种观点，人类思想与多元宇宙的存在紧密相连，这就为它们的应用提供了一个合法范围。我们在下一节中将转向检验多元宇宙，但在此之前，我们先来考虑一下解决人类问题的形而上学选择。

9.1.6 形而上学选择

要在人类问题上取得进步，我们必须认真考虑终极因果性的本质。我们所看到的现象的基本原因是什么？如果我们追求其结论的物理因果链，我们就仍然要面对这个问题：为什么发生的是这些而不是其他？不管原因如何，我们追求的是终极原因。注意，我们在这里将要离开科学本身的地盘，开始探索形而上学的领域——是科学，事实上也是存在的基础。正如上面所注意到的，人们可以简单地决定不去追求这样的问题。如果我们要继续这个问题，看起来对终极因果性问题，基本上有六种方法，也就是随机性、必然性、高度可能性、普遍性、宇宙学本质选择和设计。我们依次简短地考虑一下。

选择1：毫无意义的随机选择。宇宙中的初始条件发生并导致了事物是它们现在所是的样子，这纯属偶然现象，概率在其中并没有起到作用。这里用不

到更深层次的解释，研究“终极原因”没有意义。

这当然是逻辑上可能的，但是作为一种解释并不令人满意，因为我们从这种方法得不到任何思想的统一或者预言的力量。但有些人还是或含蓄或明确地坚持这种观点。

选择2：必然性。事物必须是它们所是的样子，没有其他选择。我们所看到的特征及其基础的定律是宇宙的统一性所要求的：连贯性和一致性要求事物必须以其所是的方式存在。貌似真实的其他选择都是错觉。只有这种物理学才是自洽的：所有逻辑上可能的宇宙必须遵循同一种物理学。

要真正证明这是一个非常有力的论据，潜在地导向一种自洽和完备的科学观点。但是我们可以想象其他可替代的宇宙！——为什么它们被排除在外？而且我们在这里遇到一个问题，我们还没有成功地想出一个完全自洽的物理学观点：不管量子物理学的基础还是数学的基础都没有一个真正稳固一致的根基。只有这些问题得到解决，这个方向才能追求成功的结论。

选择3：高度可能性。虽然宇宙的结构看上去不大可能，但因为物理学原因，它事实上是高度可能的。

即使是按照它们自己的方式，这些论据也只是部分成功的。如果我们考虑所有的可能性，它们会面临很多问题：提出这种观点的讨论事实上先天地或含蓄或明确地限制了考虑的可能性范围，否则的话宇宙不大可能会是我们看到的样子。而且，我们没有一种适当的方式可以应用于初始条件的集合中，使得我们可以接近这些可能性。此外，正如上面讨论的(见第3节)，可能性争论在宇宙本身中的应用就是有风险的，因为宇宙是唯一的。尽管有这些问题，这种方法在科学界受到了大量的支持，比如，它是混沌暴胀思想(2.6节)的基础。它在普遍性假设的背景中实现了其最大的影响。

选择4：普遍性。这是“所有可能发生的都会发生”的主张，宇宙的总体或者不连续的膨胀宇宙区域的整体在现实中实现了，在其中所有的可能性都发生了[Rees，1999；2003；Tegmark，2003]。在它的完整版本中，人择原理在它的强形式(如果所有可能的都能发生，那么生命就一定会存在)和弱形式(生命只有在某些实现了的可能性中才会发生，这些是由弱人择原理从所有包括其他的可能性中选出来的，弱人择原理被看作是一种选择原理)中都得到了实现。有

以下四种方式已被采用。

1. 空间变化。膨胀宇宙区域的变化通过随机的初始条件在空间中得到了实现，就像在混沌暴胀中(2.6节)一样。虽然这为可能性的应用提供了一个合理框架，但从终极解释的观点来说它并没有真正获得成功，因为仍然还是只有一个独一无二的宇宙，它的(随机)初始条件需要得到解释。初始条件在整体上可能是统计均匀的，但在某些物理量中还可能存在整体变化，以至于宇宙并不是统计上均匀的。这些条件可能被约束在某个不允许生命存在的领域。这是不完全执行了整体思想。就其现在的运行来说，它确实是上面提到过的“高度可能性”的一个变种。如果它是已经证明的物理学的几乎唯一的结果，那么就会提供一个很好的证明，但是作为这样的思想基础的物理学甚至根本没有被唯一地定义，更不用说检验了。简单地声称某个有着某种特定势的标量场存在并不能证明它就确实存在!

2. 时间变化。膨胀宇宙区域的变化可能在时间中实现，在一个有着许多膨胀相的宇宙中(凤凰宇宙)，不管是全局还是局域地发生。与前面一种情形中应用的评论大致相同。

3. 量子机制。它可能在量子宇宙学的埃弗雷特—惠勒(Everett-Wheeler)“多世界”的存在中发生。在这种“多世界”中，所有的可能性都通过量子分支[Deutsch(道成)，1998]发生。这是为数不多的可以替代量子力学哥本哈根解释的提议中的一种，哥本哈根解释导致了观测者存在的必然性，因此就导致了上述考虑到的强人择解释(9.1节)。多世界的思想是有争议的，它发生于各种不同的竞争形式中[Isham，1997]，这些竞争理论都没有被普遍接受。多世界的思想并没有为事实上发生的特定事件提供因果解释：如果我们坚持它，那么我们就仍然不得不解释我们观测到的特定历史的特性(比如，为什么我们的宏观宇宙有着高度的对称而几乎所有的分支都不会有?)。最重要的是，它很明显是不可检验的，没有办法在实验上证明所有那些其他分支宇宙的存在，确切地说是因为理论给出了与标准理论相同的观测预言。

4. 完全不连续的。它们可能作为完全不连续的宇宙发生：真的存在一个宇宙的总体，在其中所有的可能性都发生，不需要相互之间有任何联系[Lewis(刘易斯)，1986；Rees，2003；Tegmark，2003]。那么会出现的一个问题是，

是什么决定了可能发生什么？比如，逻辑定律自身怎么样？它们在考虑所有的可能性时是否是不可侵犯的？我们回答不了，因为我们没有办法接近这许多假定的世界。我们在后面再深入探讨(见9.2节)。

在所有的这些情形中，即主要的问题出现在把这种观点和可检验性联系起来的时候，我们必须质疑这些想法作为科学解释时的意义。它们都与奥卡姆剃刀相矛盾：我们“解决”一个问题的代价是要设想一个复杂得多的存在。而且它们并没有解决终极问题：为什么这个宇宙的总体存在？有人可能会提出，实在的这种终极解释甚至会比在单一宇宙中的情形更有问题。但这种方法有它自己的内在逻辑，有人发现这种逻辑非常令人信服。我们在下面会考虑这种方法，见9.2节。

选择5：宇宙的自然选择。如果适当地建立一个塌缩到黑洞之后的重新膨胀过程，它将会打开通往演化概念的道路。这种演化不仅仅是宇宙的结构和内容在时间中发展的意义上的演化，而且是宇宙区域膨胀的达尔文选择可以发生的意义上的演化，正如斯莫林(Smolin)所提出的[Smolin，1992]。其思想是，可能会有向黑洞的塌缩和随后的重新膨胀，但在每一个转变的过程中都会有物理学常数的改变，因此每一个时刻都存在一个膨胀相，物理学行为会有所区别。那么关键之处就在于，常数的某些值会导致更多黑洞的产生，而某些值将会导致更少的黑洞产生。这允许演化选择偏爱会产生更多黑洞的膨胀的宇宙区域(因为物理学常数的有利值在那些区域中最重要)，因为它们会有更多的“后代”膨胀宇宙区域。这样人们就可以设想自然选择偏爱那些产生最大数量的黑洞的物理学常数。

选择6：目的或者设计。我们所观测到的对称和脆弱的平衡特别要求条件的连贯性和因果的契合，说明在某种意义上它们已经被有目的地设计了。也就是说，它们在物理学定律的设定和宇宙边界条件的选择上都给出了意向性的证据。这种观点是犹太教与基督教共有的神学基础。和其他观点不一样，这种观点引入了赋予某些东西以意义的元素。在所有其他的选择中，生命都是偶然存在的，是一些盲目过程的偶然产物。

从科学的角度看，这种观点的主要不足在于它缺乏可检验的科学结果(“由于上帝的存在，我预言宇宙中物质的密度应当是 x，好的结构常数应当是 y”)。

这是科学家一般会尽力避免这种方法的原因之一。有些人会立即拒绝这种可能性，认为它毫无意义，不值得考虑。但它当然在逻辑上是可能的。其现代版本，与前面所有的科学讨论一致，会看到某种目的成为物理学定律和宇宙边界条件的存在和特定本质的基础，以这样一种方式，生命(最后是人类)会因为那些定律的运用而变为存在，然后导致特定种类的动物在演化的过程中得到发展，成为历史记录中的证据。假设接受了这种演化发展的思想，很明确的是，在特定的物理学定律和初始条件的选择和实现中，允许这样的发展，高级的有创造性的活动发生了，这可能就是人们所设想的设计发生的地方。①

然而，从物理科学本身的视角来看，没有理由接受这种论据。事实上，以这种观点，在设计和偶然性之间真的不存在区别，因为没有人能说明它们会导致不同的物理预言。

9.1.7 形而上学不确定性

在考虑作为有关人类问题基础的终极因果性时，最终我们要面对的是在上述两个选项中进行选择。正如康德(Kant)和休谟(Hume)已经指出的，虽然我们也许能够强烈地主张其中的一个或另一个，但是我们证明不了其中的任何一个是对的[Hume，1993]。

论点 G2：形而上学的不确定性仍然是关于宇宙学中的终极因果性的。我们不管是通过科学还是哲学，都不能得到基本的形而上学宇宙学问题的确定性。

如果从纯粹的科学基础上去看人类的问题，我们最终不会得到任何答案，根本上是因为科学是通过把对它的考虑限定到一个实在的有限方面而得到合理的确定性的。即便它偶尔走人终极因果性的领域，也不是要去处理因果性。通过自身，它不能在这些选项中做出选择。没有相关的实验或者观测能够令人信服地解决这个问题。因此，要前进就需要更广义的观点，既重视科学也重视更广泛的考虑。这个问题的本质是哲学的而不是科学的。然后会出现的一个重要问题是，除了那些可以被描述为纯科学的数据(9.3.4 节)，还有什么类型的数

① 这与“巧妙设计”运动提出的观点不同。它没有假定演化过程的结果是上帝微调出来的。

据是与这些哲学选择相关的。

9.2 问题 H：多元宇宙的可能存在

如果存在许多有着不同特性的宇宙的足够大的整体，那么可能可以说它们中的某些只是碰巧做对了，所以生命可以存在。这有助于解释许多要不然就不受物理学约束的参数[Rees，1999；Rees，2003]的精细调谐本质。正如前一节中讨论的，存在很多种方式，在其中理论上多元宇宙可以实现[Lewis，1986；Tegmark，2003]。他们提供了一种方式可以把概率应用于宇宙[Sciama，1971；Bostrom(博斯特罗姆)，2002](因为它们拒绝宇宙的唯一性)。但是，这个概念有很多问题，而且，这种想法在观测和实验上都不可检验。因此，它的科学地位是有争议性的。

9.2.1 定义

在为多元宇宙辩护时，常常有人说“所有可能发生的，都发生了”(或者类似的说法)。然而这种说法并不能充分地确定一个多元宇宙。要适当地定义一个多元宇宙，需要两步[Ellis *et al.*，2004]。首先，人们需要确定宇宙学通过定义概率空间构想了什么，所谓的概率空间是：所有可能宇宙的空间 M，每一个都能用一个态空间 S 中的一组态 s 来描述。M 中的每一个宇宙 m 都可以用一组不同的参数 p 来描绘，p 是 S 上的坐标。这里需要做出选择。按照几何学，它是只包含罗伯森—沃克模型，还是更一般的模型(比如比安基模型，或者不存在对称的模型)？按照引力，它是只包含广义相对论，还是也包含了膜理论、有着不同的 G 的模型、圈量子引力模型、弦理论模型以及与它们相关的可能性“图景”，还有给予宇宙波函数概念的模型？它是否会只允许标准物理学可以有不同的常数，或一个更宽的物理可能性的范围，比如没有量子理论的宇宙，有些有着五种而不是四种基本作用力，其他的有牛顿引力的宇宙？定义空间的可能性意味着做出某种关于物理学和几何学的假设，这些假设要被应用到多元宇宙中认为可能的所有类型的模型中去，还要排除其他所有的可能性。

第二，人们需要明确的是，哪一个可能的宇宙在多元宇宙中得到了物理上的实现，每一个宇宙发生了多少次。一个多元宇宙，如果它要具有真正的解释力，它就必须是一个在物理上实现的多元宇宙而不是一个假设的或者概念的宇

宙。因此需要一个分布函数$f(m)$来确定 M 中每一种类型的可能宇宙 m 实现了多少次。函数$f(m)$表达了任意一次实现中的可能情况。事情可能不同！因此，$f(m)$描述了一个特定的被想象为在一组可能性中得到实现的宇宙或者多元宇宙的整体。举例来说，$f(m)$可能对于所有参数 p 的所有可能值是非零的（“所有可能发生的都发生了”），但有可能是 f 描述了一个多元宇宙，在其中有 10^{100} 个某个特定宇宙的相同复制（实现的过程找到了一个特定的成功方法，然后就没完没了地进行复制）。

此外，我们需要一个微元 $d\pi$，使这个函数能够决定多元宇宙中不同特性的数目和可能性：宇宙的数目对应一组参数增长 dN，对于连续的参数，由下式给出：

$$dN = f(m)\,d\pi \tag{38}$$

对于离散的参数，我们把来自所有允许的参数值的贡献都算进去了。整体宇宙的总数 N 由下式给出：

$$N = \int_M f(m)\,d\pi \tag{39}$$

通常会有所偏离。这里的积分是对成员参数的所有允许值的整体积分，我们把它看作是包含了所有相关的离散的综合。定义在一组宇宙上的量 $p(m)$ 的期望值 P 由下式给出：

$$P = \int_M p(m)f(m)\,d\pi \tag{40}$$

为了对多元宇宙进行适当的确定[Ellis *et al.*，2004]，这三个元素（概率空间、微元以及分布函数）必须清晰地定义。这几乎永远做不到。

9.2.2　非唯一性：可能性

在两个步骤中都存在非唯一性。说“所有可能的都发生了”并不能决定什么是可能的。直到可能的宇宙空间被完全描述，多元宇宙的概念才会有明晰的定义。还很不清楚怎样才能唯一地做到这一点。关于什么能被看作是“所有可能的”宇宙的整体的问题，可以通过定义这个词语的意思而产生任何你想要的结果——标准物理学和逻辑对它们不会产生任何必然影响：我在这样一个整体中设想为“可能”的也许你并不会赞同。在执行对多元宇宙中可能性的控制的时候，高级别的原理是什么？又为什么是这样？这里关键的一点是，我们对概率

的理解总是从我们已知的东西进行推算必然能达到的，而且我的想象可能比你的要更加丰富，也并不需要对应于那里真实存在着什么——如果那里确实存在着什么东西的话。我们是否只包括了以下内容。

• 弱变化：比如，只允许物理学常数的值发生变化？这是一个很有趣的应用，但当然不是在执行“所有能发生的都发生了”的思想。这是变化的一组极端约束。

• 温和变化：不同的对称群，或者维数等。比如我们可能会把弦理论的可能图景[Freivogel *et al.*，2005b]看作是支配多元宇宙的现实指示[Susskind，2003；2005；Freivogel *et al.* 2005a]。但是这离“所有可能的”还确实很远，因为那当然包括不被弦理论支配的时空。

• 强变化：不同数量和类型的作用力、没有量子理论的宇宙，或者相对论在其中为错的宇宙(比如存在以太)；那些弦理论在其中是一个量子引力的好理论而其他的理论并不是的宇宙；那些物理学定律在其中有着非常不同的基础(比如没有变化原理)的宇宙。

• 极端变化：物理学在其中不能很好地被数学描述的宇宙、有着不同逻辑的宇宙、由地方神灵统治的宇宙；允许像哈利·波特系列书籍里那样的魔法世界、完全没有物理学定律的宇宙？甚至没有数学或者逻辑的宇宙？

哪一种是多元宇宙的特性？为什么？我们在这里可以通过一个自相矛盾的问题来表达我们的困境：逻辑定律是否在所有可能的宇宙中都是必要的？

9.2.3 非唯一性：存在和因果性

一个特定的多元宇宙是通过确定实际上实现了的宇宙分布函数$f(m)$定义的。还不清楚什么机制可以成为这种分布的基础，任何关于这个机制的想法都是完全不可检验的。关于是什么决定了由假定概率空间定义的可能性的存在，我们需要一些指示：是什么决定每一个发生了多少次？除非我们理解了假定的基本机制，否则我们就不能给出任何严谨的答案。而且检验任何假设的机制都不会有前景。假设的作为多元宇宙中任何规则的基础的机制必须不仅仅存在于这个宇宙之前，而且要存在于所有其他的宇宙之前。如果有人假定一个宇宙在大尺度上是相连的但是在局域上却分离为因果不相连的区域，有着不同的物理特性(就像混沌暴胀中那样)，就得到了一个可以作为有效宇宙的创生机制的基

础的合理图景——但是其代价是假定了未经检验的、甚至是有可能无法检验的物理学的有效性。因为这一点，人们没有得到对唯一多元宇宙的说明：物理学有可能与我们假设的并不一样。

9.2.4 解释力

反过来对于这些问题我们又获得了怎样的解释力呢？有人指出，它们解释物理学和宇宙学的参数，特别正是宇宙学常数的那个有问题的观测值［Weinberg，2000a；Weinberg，2000b；Susskind，2005］，其论据如下。假定存在一个多元宇宙，观测者只能存在于其中一个极其不可能与生物相关的外围区域，在那里宇宙学常数的值非常小［Hartle（哈特），2004］。类似的论据在中微子质量的争论中也提出过［Tegmark *et al.*，2003］。如果多元宇宙有着许多特性不同的不同位置，这可能非常有助于我们理解有关人类的问题：有些区域允许生命存在，其他的区域则不允许［Barrow and Tipler，1984；Leslie，1989］。这确实提供了一些有用的解释力，但是远达不到令人信服的程度。首先，还不清楚为什么在我们提到的各种分析中要假定多元宇宙学的宇宙学常数应当有被约束的各种变化。如果我们假定“所有能够发生的都发生了”，那么变化就不是被约束的，那些分析也就无法应用。

第二，终极问题依然存在。为什么这种唯一的巨大整体有着其所拥有的特性？为什么是这一个多元宇宙而不是其他？为什么它是一个允许生命存在的多元宇宙？许多多元宇宙都不会允许任何生命存在。要解决这一点，我们可以提出一个宇宙总体的整体，有着更大的解释力和更少的观测证实的前景，等等。一个无穷回归隐约出现了。事实上，如果我们宣称（正如我们文章一开头所提出的）“宇宙”是所有物理存在的总和，那么当一个膨胀的宇宙区域的总体存在的时候，不管是否因果地相连，这个总体自身都应当被叫作“宇宙”，因为它就是物理存在的实体的总和。一个单独存在的宇宙区域的所有基本问题在多元宇宙中重新发生了——因为如果适当考虑，它确实就是宇宙！

9.2.5 可检验性

如果一个整体存在，而其成员不以任何物理的方式与可观测宇宙相连，那么我们就既不能以任何方式与它们发生相互作用，也不能观测它们。因此我们

可以讲任何我们喜欢讲的关于它们的东西而不用担心被反驳。① 这样，我们作出的任何关于它们的陈述都不会有稳固的科学或者解释地位。对于任何其他主张有着不同特性的整体(例如，主张在它们的整体中可能有不同类型的基本逻辑，或者主张它们的整体中存在着许多物理有效的神或魔鬼)的人来讲，它们都很脆弱。

论点 H1：多元宇宙的提议是不能通过观测或实验证明的，但是自洽性检验是可能的。直接的观测不能证明或者证伪一个多元宇宙的存在，因为不存在允许观测或者检验它们的必要的因果关系。它们的存在不能从已知物理学预言，因为假设的因果或者前因果过程要么是不可证明的，要么事实上是不可检验的。然而，具体多元宇宙模型的有些自洽条件可以被检验。

任何被推荐作为多元宇宙思想基础的物理学，比如科尔曼—德·卢西亚(Coleman-de Luccia) 隧穿[Coleman and de Luccia，1980]，都将是已知物理学的一个推论。但这个主要推论对宇宙学的有效性是不可检验的。

有人尝试过证明多元宇宙的可检验性存在，首先运用了里斯的“光滑斜面”论据[Rees，2003]。其过程如下。我们可以合理地假设存在于视界之外我们看不到的星系(8.3 节)。为什么不一小步一小步地把这个论据延伸到可以表明完全不相连的宇宙的存在呢？问题是虽然假定了存在的连续性，但这不完全成立。我们视界之外的领域被假定存在且有着与那些视界之内的领域有着相似的特性，因为它们是它的连续延伸，并且在很大程度上有着共同的因果起源。它们的性质可以从我们所看到的东西进行推断。不连续的多元宇宙领域被假定有着非常不同的特性，它们的性质不能从我们所能看到的东西推断，因为不存在连续性或者因果的联系。

第二，不少作者，比如莱斯利[1989]、温伯格[2000a；2000b] 和里斯 [2003] 都运用了基于以下思想的论据：宇宙只不过必须那样特殊。这是一种形式的“特殊性论据”。根据里斯，如果我们的宇宙被证明比我们的存在所要求的更加特殊地协调的话，那么用多元宇宙的存在来解释这种“过分协调”就会受到反驳。但是

① 但是有莱布尼茨(Leibniz)[Wilson(威尔逊)，1989]和刘易斯 [Lewis，1986，Section 2.4，pp. 108—115]的反对争论。

真实的宇宙并不比这更特殊，因此多元宇宙就没有受到驳斥。

更详细地说，单纯的量子物理学预言了宇宙学常数 Λ 非常大。但我们在宇宙中的存在要求它足够小以便星系和恒星可以形成。因此 Λ 显然必须低于星系形成这一门槛。如果我们的宇宙从属于一个整体，在这个整体中 Λ 等效地可能在与生物相关的区域取任何值(均匀概率假设)，[1] 那么我们就不会预期它会远远低于这个门槛。这是因为，如果它远远低于这个门槛，随机选择整体中那个宇宙的可能性就变得非常之小——在这个整体的与生物相关的子集中，只有极少数有着这样小的 Λ 值的宇宙。也就是说，更可能的是，整体中任何能友好容纳生物的宇宙都会有一个与那个门槛值接近的 Λ 值。目前在这一值上的数据表明它并非远远低于门槛。因此，就 Λ 而言，我们的宇宙并非显著地特殊于它所需要成为的，因此要通过多元宇宙的存在来解释它的精细调节是合理的。

这个论据有说服力吗？它是对已知存在的多元宇宙一致性的合理检验，以便概率的考虑可以应用。但是如果不存在多元宇宙(第 3 节)，它们就没法应用。而且，概率的考虑永远都不可能是令人信服的。事实上，

论点 H2：基于概率的论据证明不了多元宇宙的存在。概率论据不能被用来证明多元宇宙的存在，因为如果多元宇宙存在的话，它们只是可应用的。而且，概率论据永远都不能确定地证明任何事情，因为它不可能违背任何概率预言，因此当只有一种要考虑的情况时，任何统计观测都是不可能的。

在概率论据的基础上我们能说的全部就是，某种特定态是不大可能的，但这并不能证明它就是不可能的。事实上，如果说它有非常低的概率，那么这就确切地是宣称它是可能的。因此，这样的论据即便是可应用的，它们充其量也只能给出一个合理的指示。假定概率论据可以是令人信服的，等同于宣称宇宙是一般的而不是特别的。但事实是否真的如此也正是争论中的问题(见论点 D3)。这个论据作为一个多元宇宙的合理论据是有用的，但并不能作为多元宇宙存在的证明。

最后，有人提出多元宇宙的存在是一个有着无限空间部分的宇宙的不可避

---

① Λ 的概率分布峰值可能要比热爱生命的区域高得多，其低值在那个狭窄区域可能近似为常数，参见[Hartle，2004]。

免的结论[Tegmark，2003；Seife，2004]，因为它导致了条件的无限空间复制，参考[Ellis and Brundrit(埃利斯和布伦德里特)，1979]。但这里空间无限的假定是一个未经检验的哲学假设，它的正确性当然不能被观测所证明。撇开视界的存在会阻碍对这个假设的证实，即便整个宇宙都是可观测的，要证明它的正确性仍然是不可能的，因为按照定义，计算无数的客体要花费无数的时间。这是一种不可检验的哲学论据而不是一个经验上可检验的论据。而且，也可以表明它是不大可能的(9.3.2 节)。事实上，目前的数据说明了这并非实际情况。这是人们可以针对一些多元宇宙思想采用的一个良好的一致性检验。

9.2.6 解释 vs 可检验性

可以说，这种无限的整体实际上存在的论据有着某种解释的经济性，虽然其他人也可以说，为了支持存在的多样性，夸张地进行假设以解释我们确实了解存在的这一个宇宙，奥卡姆剃刀已经完全被抛弃了。当然，其代价是不管是从观测上还是实验上都缺乏可检验性——这通常被当作是任何严肃的科学理论的一个基本元素。① 它并非唯一可定义的或者可确定的，并且完全缺乏可证实性。没有办法决定多元宇宙中其他宇宙的特性，如果它们确实存在的话，因为它们永远都处在观测范围之外。重点是存在的问题不仅仅只有要表明多元宇宙存在这一个问题。如果这是一个科学的命题，那么人们需要能够表明哪一个特定的多元宇宙存在，但是没有观测可以做到这一点。事实上，如果你不能表明哪一个特定的宇宙存在，毫无疑问你就已经表明了任何一个宇宙的存在。

这样一种存在的主张在这个情况中有什么意义？加德纳[2003]是这样讲的："没有任何微小的证据可以表明存在在我们所居住的宇宙之外还存在其他宇宙。就我们目前知道的，宇宙甚至都不如两个黑莓那样丰富。"②

论点 H3：多元宇宙是一种哲学而不是科学思想。多元宇宙的思想为精细调谐提供了一种可能路线。但它并非是唯一地定义的，除了一些可能的一致性检验之外，并不是在科学上可检验的，最后也只是拖延了终极的形而上学问题

---

① 假设有了这篇文章中严格的条件，[Stoeger *et al.*，2004]指出了那些多元宇宙假设在其基础下可以检验的框架和条件，这些条件并没有得到满足。

② 这与道成和刘易斯对概念的证明[Deutsch，1998；Lewis，1986]形成了鲜明对照。刘易斯证明了"模态实在"的论点：我们所在的世界是多个世界中的一个。

而已。

一些多元宇宙思想(9.2.7 节)的决定性的一致性检验对于那些具体的多元宇宙提议来说是必要的条件，但它们自身却很难表明多元宇宙的想法是对的。促使人们相信这一点的事例来自理论和哲学的考虑而不是数据[Susskind, 2005]。如果科学被看作是可检验的，那么整体的物理存在①的主张作为一种科学解释的想法是有问题的。事实上，采用这些解释是理论超越可检验性的一次成功[Gardner, 2003]，但是被假设的理论却不是可检验的。因此由于哲学的原因而存在一个形而上学的选择。这并不意味着它不合理(它可以为非常有说服力的论据所支持)，而是意味着它缺乏科学身份的情况应当得到澄清。

9.2.7　观测和反驳

尽管上述给出的预测令人沮丧，还是存在一些具体的例子可以反驳混沌暴胀(多领域)类型的设想(2.6 节)的存在。第一，如果我们居住在一个"小宇宙"中，我们已经看到宇宙的周围，然后宇宙自身封闭在一个单一的类弗里德曼—勒梅特领域中，这样就不会再有更多这样的在一个单连通时空中因果地与我们相联系的领域存在。这种"小宇宙"的情况是观测可检验的(4.3.1 节)，它的确会反驳通常的混沌暴胀设想，但并不是一个真正"不连通的"多元宇宙思想。因为任何观测都不能表明它是错还是对。第二，许多混沌暴胀的版本，比如那些包含了来自德西特时空的科尔曼—德·卢西亚隧穿[Coleman and de Luccia, 1980]的版本。其中要求 $k=-1 \Leftrightarrow \Omega_0<1$ [Freivogel *et al.*, 2005b; Susskind, 2005]。在 WMAP 观测与其他可以得到的最好的数据(2.3.7 节)结合起来时，这个要求受到了 Ω_0 上的 2σ 界限的略微反驳。目前最好的数据与 $K=-1$ 稍微一致，但是受到那个数据最强有力支持的是 $K=+1$，支持有限封闭的空间片断，而不是一个诸如萨斯坎德等人所支持的无限多元宇宙[Freivogel *et al.*, 2005b; Susskind, 2005]。

9.2.8　物理学或者生物学范式——适应性演化

既然多元宇宙的思想最终必须用哲学而不是用科学检验证明是合理的，那么这种思想是否有一种哲学上更可取的版本？可以认为有。通过把生物学的主

① 相对于一个清楚明白的假设的考虑，这种总体有可能实际上是有用的，见 7.2 节。

要建设性原理(也就是适应性演化，这是已知的可以产生以前不存在的有序结构的最强大的过程)引入宇宙学，有可能得到更加强大的解释力。基本上这在李·斯莫林的某种达尔文适者生存的思想(9.1.6 节)中已经意识到了：黑洞的塌缩会随后发生再膨胀，但是每一次，为了允许向着产生最大数目的黑洞的范围的演化选择，都会伴随着物理学常数的改变。这种思想还不完备，但是非常吸引人。

论点 H4：宇宙学的基本物理学范式可能被推广到包含生物学见解。宇宙学中占优势的范式是理论物理学的。通过囊括生物学见解，特别是达尔文的进化论，它可能会实现更强的解释力。斯莫林膨胀宇宙领域群的演化思想[Smolin，1992]就是这种考虑的一个例子。

这种结果以重要的方式与标准宇宙理论精确地不同，因为它在一个理论中包含了三种 19 世纪的重要思想，即：(i)达尔文的物竞天择种群进化论；(ii)宇宙在其与膨胀相联系的结构的重要变化的意义上的演化；(iii)量子理论，构成了唯一部分说明的被认为会引起黑洞塌缩后重新膨胀的机制的基础。这与只由考虑对称形成的变化原理支配的动力学的理论物理学范式形成了鲜明对照。看起来很值得作为一个对结构的创生①理解得非常不同的方法去追求。

9.3 问题 I：存在的本质

这所有的基础是存在的本质问题，它有许多方面与从纯粹的物理学到更形而上的问题相关。

9.3.1 *物理学存在：物质的种类*

宇宙学未解决的关键问题关系到什么种类的物质和/或场存在。虽然我们能很好地了解太阳系中的物质，目前我们却还不能了解大范围宇宙中存在的大多数东西。

论点 I1：我们不了解宇宙早期或晚期的主要动力学物质组成。物理宇宙学的一个关键目标是要确定暴胀、暗物质以及暗能量的本质。做不到这些，宇宙学的因果性理解就不算完成，特别是宇宙的遥远未来的命运还是未知的。

---

① 参考萨斯坎德[2005]的观点。

这是目前宇宙学中大多数研究的核心工作。它们得不到阐述，宇宙学就不能适当地与物理学相连接，支配宇宙动力学的物质本质就是未知的。它的阐明无疑是宇宙学所关注的关键问题之一［Durrer（杜雷尔），2002］。一个关键的要求是，即便我们不能用实验确证所提出的物质本质，至少它应当是物理上合理的。对于某些目前的提议，比如有着负的动能项的所谓的“幽灵物质”来说，情况似乎并不是这样。

宇宙在遥远未来的命运主要依赖于暗物质（解释能量密度随时空变化的动力学场）态的有效状态方程。但问题是，即便我们目前能决定这些特性（参数值的一个特定范围），也不能保证它们在将来是什么样子的（参数值的一个非常不同的范围可能超出实验检验的可能范围）。而且，调整一个“暗能量”模型以适应超新星的数据并不能决定基础物理学。人们可以为任何单调演化 $S(t)$ 配以一个适当的状态方程函数 $p(u)$ 的选择。特别地，对任何 $S(t)$ 和任何 k，我们通过：

$$\kappa\mu(t)=3\left[\frac{\dot{S}^2(t)}{S^2(t)}+\frac{k}{S^2(t)}\right],\ \kappa p(t)=\left[\frac{\dot{S}^2(t)}{S^2(t)}+\frac{k}{S^2(t)}\right]-2\frac{\ddot{S}(t)}{S(t)} \qquad (41)$$

定义 $\mu(t)$ 和 $p(t)$。那么式(9)和式(7)就会被完全满足，我们就已经“解决了”这一选择的单调演化 $S(t)$ 的场方程。如果我们能够在观测上决定 $S(t)$ 的形式，比如从与超新星数据相联系的(m，z)曲线，这在根本上就是我们如何能够决定那种演化会要求哪种类型的“暗能量”或者“解释一个能量密度随时空变化的动力学场”。我们可以找到能消除这两个方程之间的 t 所暗示的状态方程。然而，这不是一个物理学解释，除非我们已经在某个独立的实验检验中证明了这种形式的物质存在，或者已经在理论上表明为什么这种物质或者场有着它某种更加基本的形式而不只是现象上适合。如果我们将暗能量视为一个标量场，式(32)和式(41)所表示的动能项 $\dot{\phi}^2$ 可能为负——最近有些研究者提出的所谓“影子物质”模型就是这种情况。如果运用通常的物理学标准，这就是这种类型的物质是非物理学的一个证明，而不是一个关于暗能量本质的证明。

9.3.2　无限的存在

如果宇宙中的物质或者客体的数目是有限的，而并不是说存在的量是无限的，那么存在的本质就会显著地不同。有些提议认为在一个多元宇宙和许多有

着无限空间部分的宇宙学模型中可能会有无限数目的宇宙，意味着无限数目的粒子、恒星以及星系。然而，无限与巨大数目是完全不同的！追随希尔伯特(Hilbert)[Hilbert，1964]，人们可以提出这些不可证明的提议不能为真："无限"这个词指代一个永远达不到的量或者数目，因此永远不会发生在物理现实中。① 他说道：

> "我们的主要结果是无限在现实中无处可寻。它既不能存在于自然中，又不能为理性思考提供一个合理基础……无限所能扮演的角色只是一种思想……它超越了所有的经验并把具体事物作为完全的……"[Hilbert，1964，151]。

这就暗示了"无限"不能在一个具体的物理环境中被达到或者实现，相反地，概念本身暗示了它就不能被实现！②

论点I2：经常所说的无限的物理学存在是有问题的。宇宙学或者多元宇宙中所说的物理上现实的无限出现了棘手的问题。人们可以表明它们是非物理的，在任何情况下这种断言当然都是无法证实的。

这在原则上可应用于任何单一宇宙中的小尺度和大尺度中：

• 微观层次上由一个实(数)流形表示的物理时空连续体的存在与量子引力所宣称的普朗克(Planck)尺度下的离散时空结构形成了对比。人们可能假定这种离散时空是完全非线性量子引力理论的一个一般特征[Rovelli，2004]。根据物理实在，这意味着去除了真实的线性连续体在所有的物理学变量和场中产生的不可计数的无穷大。③ 没有实验可以证明时间或者空间中存在着一个物理连续统。我们所有能做的就是在越来越小的尺度上检验时空结构，但我们不能接

---

① 一个更深刻的有意思的问题是对偶问题：这个量会发生在物理实在中吗？这与虚无的物理存在的思想是相关的，就如空间的类比[Seife，2000]。空间不是空无一物！参见[Susskind，2005]。

② 对照的观点见伯纳黛特(Bernadete)[Bernadete，1964]

③ 为了完全避免无限，要求在物理实在中没有连续体(因为任何连续体间隔都包含了无限数目的点)。概念上，没有它就意味着对许多事情进行了完全的改写。在我看来，考虑如何与观测一致地做到这一点，是一项非常有价值的计划。

近普朗克尺度。

●宏观层次上无限大的空间部分提出了像希尔伯特所指出的那样的问题，并且导致了生命和所有事件的无限复制[Ellis and Brundrit, 1979]。我们可以假设空间在欧几里得(Euclid)几何以及多个宇宙模型中永远延展，但是我们永远都不能证明任何真实宇宙中的三维空间以这种方式继续——这是一个不可验检的概念，并且宇宙的真实空间几何是非欧几里得的。因此欧几里得空间是抽象的，在物理上可能并不真实。混沌暴胀模型中假定的无限来源于对前存在的无限欧几里得空间的部分推测，没有理由认为它必然存在。在物理宇宙中，空间无限性可以通过或者产生于正的空间曲率，或者产生于为零或者负空间曲率的宇宙中的紧致拓扑的紧致空间部分而避免。与物理学边界条件相关的马赫式考虑说明了这是高度可取的[Wheeler(惠勒), 1968]。如果有人把弦理论当作物理学的重要基础，那“维度”就表明了三个大的空间维度也应当是紧致的，因为小的(“紧致的”)维度都被看作是这样的。现在从宇宙黑体辐射和其他观测得到的最好数据(2.3.7节)实际上表明 $k=+1$，说明了最适当的弗里德曼—勒梅特模型的封闭空间部分。

●永恒宇宙的存在说明实际上存在无限的时间，这本身就有问题。如果一个事件发生于任何时间 t_0，那么人们就需要有一个关于为什么它没有发生于这个时间之前(因为存在着一个无限的之前时间供它发生)的解释。如果宇宙真的是周期性的，庞加莱永恒性的回归(6.1节提到过)就是可能的。在任何情况下都不可能证明宇宙作为一个整体(或者甚至是我们居住的宇宙的部分)是过去无限的。观测不能做到，而用以保证这会发生所需要的物理学(如果初始条件是正确的)也是不可检验的。甚至，想要证明它是未来无限的尝试也是有问题的(比如我们不能保证无限未来中真空的特性——它可能会退化为一个对应于负的有效宇宙学常数的态)。

●它适用于多元宇宙的可能性质。要确定一个一般宇宙的几何，需要无数的信息，因为要做到这一点，必要的量是时空上的场，一般地要求在每一点(或者等价地在无数的傅里叶系数)上进行确定：它们几乎总不是算法上可压缩的。在所有的组合中，所有这些组分的所有可能值将会发生在一个多元宇宙中，在这个多元宇宙中，“所有能发生的都发生了”。还有无数种拓扑可能性，

大大恶化了所有关于无限和整体的问题。只有在高度对称的情况下，比如弗里德曼—勒梅特的解，这个数据才会降低为一个有限数目的参数，其中每一个都会发生于所有的可能值（它们自身经常被用来贯穿一个无限集，即整条实线）之中。这个整体之中的多个宇宙自身可能有着无限的空间范围，并且包含无数的物质，它们有着所有必要的问题。要构想一个宇宙无限集的物理创生（大多需要一个无数的信息，而它们中的许多自身会是空间无限的）至少比确定有限可明确的客体的一个存在的无限要困难一个量级。

这里应当特别注意出现于多元宇宙背景中的特定问题，这些问题是由经典理论分配给物理量的值的连续体引出的。例如，假设我们识别了一个集合中的模型的对应时间，然后假定所有的密度参数的值和那一时刻每一空间点上的宇宙学常数相关。因为这些值位于实数连续体中，这是模型的一个双重不可数的无限集。假定这样一个不可计数的无穷宇宙的真实物理的存在有悖于奥卡姆剃刀原理。但是另一方面，如果得以实现的模型的几何或者是有限的或者是可数的无限的，那么几乎所有的可能模型就都不能实现了。并且在任何情况下，这种假设都是荒谬且不可证实的。我们不能在观测上证明一个单一的其他宇宙的存在[Gardner，2003]，更不用说无穷个宇宙。无限的概念在有些多元宇宙的讨论中被轻松地抛弃了[Knobe *et al.*（亚诺布等），2005]，既不关心任何与这个陈述相关的哲学问题[Hilbert，1964]，也不关心它完全无法证实的特征。这是一种过分的主张，应当极为小心地处理。

9.3.3 物理学定律的本质

上述所有讨论的基础，是物质有序行为的基本概念。这一概念是由在宇宙中到处都相同的[①]有着数学性质的物理学定律所描述的。这里有三个相互联系的问题。

（i）物理学定律的本体论本质是什么：是描述性的，只描绘事物存在的方式；或者是规定性的，迫使事物成为这样？参见[Carroll，2004]。如果它们是描述性的，那么出现的问题是，为什么所有的物质有着相同的特性，不管它们

---

① 有效定律可能处处不同，因为比如真空态就处于变化中[Susskind，2005]。但是作为这种行为基础的基本定律自身被看作是不变的。

存在于宇宙何处？如果定律只是描述性的，为什么宇宙中的电子都是相同的？如果它们是规定性的，那么物质就必然是到处相同的（假设定律自身是不变的）。那么出现的问题是，物理学定律以什么方式存在，并对宇宙中的物质来执行自身？比如它们是否有着某种柏拉图(Plato)式的空间的存在，控制着物质的本质和存在？人们可以通过转而考虑概率空间，严格地限制实际上会变成存在的东西的可能本质[Ellis，2004]来避免讨论物理学定律本身，这种概率空间是在物理上存在着的东西的基础。这种空间几乎唯一地与潜在的定律相关，其方式与微分方程的解空间与方程的本质相关的方式相同。这使得人们能够避免物理学定律的本体论问题，但并不能解决它。

(ii)为什么物理学定律能如此好地被数学描述所解释？如果它们是规定性的，那么这个深刻的问题就可能与有人提出的数学实在的空间的柏拉图本质[Penrose，2004]相关。如果它们是描述性的，那么我们所用来概述它们的这些数学表达就只是一个方便的描述而不反映它们的最终本质。许多物理学和宇宙学的作品似乎假定了它们的最终存在的本质事实上是数学的——可能是把现象和实在混淆了(见8.3节)。

(iii)它们是否存在于宇宙之前并且控制着它的形成，或者它们是否与宇宙一起形成？这一点正是这个问题与宇宙学的本质深刻相关的地方，并且清楚地与上述其他两个问题相关。多个宇宙创生的理论假定了所有的这些定律，或者至少一个基础的子集，在物理宇宙的产生之前存在，因为它们被假定为是创生过程的基础。比如，整个量子场论经常被想当然地看作是在我们的宇宙之前存在的(见第6节)。这当然是一个不可证明的主张。

论点I3：宇宙学本质中的一个深刻问题是物理学定律的本质。物理存在的概率空间的本质是由物理学定律描述的。然而，还不清楚这些定律是规定性的还是描述性的，它们是否与时空和物质一起形成，还是先于它们而存在。

9.3.4　“终极实在”

哲学家们已经争论了一千年，存在的终极本质是纯物质性的，还是包含了某种形式的理性(“哲学理性”)和/或目的(“历史目的”)。什么是它的最终基础？宇宙的终极本质是纯数学的，还是它以某种方式有着精神的元素(参考9.1.6节)？那种深刻的争论可以从物理宇宙学中获知，但不能单独地由物理科

学解决(9.1.7节)。这里，我只对这个深刻的问题作出两点评价。

第一，即便只为了理解物质世界，也可以主张需要考虑存在的形式而不是只考虑物质——比如数学的柏拉图世界和精神世界，它们都能被当作是存在，而且按照对物质世界的影响来说都是因果有效的[Ellis，2004；Penrose，2004]。如果不考虑它们，我们对局域因果性的理解就是不完整的。

第二，在检查这些问题的时候，人们需要考虑从我们的日常生活和人类的广泛历史得到的体验(比如我们的伦理学和美学体验)以及那些由科学方法所获得的发现。许多作品认为宇宙中不存在目的——它完全只是一个在基本层次上以无目的和无意义的方式进行的粒子的集合。但是我想回答的是，那些作者在论文中所写的东西削弱了他们自己的主张。他们忽略了他们自己的行为提供的证据。在宇宙中当然在某种程度上存在意义，他们不辞劳苦地写出这样的主张的事实就证明了他们认为争论这样的问题是有意义的。这种存在的特征已经出现于物理宇宙的本质之中(7.3节)。事实上，人类思想在真实的物理世界中确切地通过许多由人类思想所接受的意义所激发的活动而变得因果地有效。任何把物理学和宇宙学与终极问题相关起来的尝试都必须认真地对待这些真实世界的经验[Ellis，2005]，否则，它将会简单地忽略大量不可否认的数据。这个数据并不解决终极问题，但确实指出了真实发生的存在的维度。

10. 结论

宇宙的物理尺度是巨大的，我们从中得到信息的远处对象的形象非常模糊。我们可以如我们所做的那样理解宇宙，这很值得注意。非常有趣的是宇宙学哲学以一种方式在相当大的程度上受到宇宙本质的可能面貌——它的巨大尺度(第4节)导致了可见视界的存在(4.3节)和早期宇宙中极端能量的发生(第5节)导致了物理视界的存在——的影响。如果宇宙学的物理学结构差异很大的话，与它相关的哲学问题(第8节)也会差异很大。而且，为了哲学分析可以与宇宙学完全吻合，宇宙学中观测和理论之间关系的详细性质(第2节)意义重大。

10.1 存在宇宙学定律吗

正如我们详细讨论的，宇宙的唯一性表明了宇宙本质的唯一性。我们现在回到初始问题，存在宇宙学定律吗？(第3节)在某种程度上，宇宙学定律只不过是我们所知道和喜欢应用于整个事件的局域定律，比如麦克斯韦(Maxwell)定律、爱因斯坦的场方程，当然还存在着从局域向整体的推算的问题。但是，虽然在宇宙学中这种推算要大得多，它似乎与我们在科学中一直所做的并没有什么不同。在这个意义上，不存在宇宙演化的特殊定律。但是，这并不决定结果：为了决定未来，宇宙学需要一些边界或者初始条件的规定。那么存在决定这个结果的真正的"宇宙学原理"、宇宙初始条件的定律吗？

"宇宙初始条件的定律"的思想似乎不是一个可检验的思想(第3节)。在科学上，人们只能描述发生了什么而不是把它与一般的原理相关起来，因为这样的原理不能得到检验。事实上，宇宙的任何边界或者初始条件的描述似乎只是这些条件的描述，而不是关于它们必须如何的可检验的方法。"宇宙学原理"——宇宙必然是空间均匀和各向同性的(4.2.2节)，是这样的初始数据的方式的描述，而不是为什么是这样的基本原因。一些研究者对这种观点的辩护是基于哥白尼原理(我们所处的位置在宇宙中并没有优越性)，后来被加强为宇宙学原理[Bondi，1960；Weinberg，1972；Harrison，2000]。但这是一个哲学假设——在根本上是宇宙必须有非常特殊的初始条件的断言——它有可能对也有可能错，并且不尝试物理学解释。这种争论现在过时了，因为我们现在比较喜欢专业的概括性和几何方法的物理学争论。但是以前它被强烈推荐过，例如[Weinberg，1972，pp. 407—412]，目前哲学争论的方向已经改变了。

然而存在一种可能会被推荐的宇宙学定律追随麦克雷[1970]，即一种"宇宙学中的不确定性原理"，对应于量子理论中的不确定性原理。不确定性应用于最大的尺度，如我们上面比较详细地讨论过的，同样也应用于最小的尺度，在最小的尺度上，它是量子理论的一个深刻特征。在这两种情况下，它的基础是相当不同的。一方面(在量子理论中)在本质上是本体论的，另一方面(在宇

宙学中)在本质上是认识论的。[①] 但它是我们与宇宙的联系中的一个关键环节，这样为了强调它在宇宙学和哲学关系中的核心地位，(跟随麦克雷)我们也许会把它形式化。

不确定性论题：基本的不确定性是宇宙学的一个关键环节。科学的探索可以告诉我们关于宇宙的很多东西，但是并不是关于它的根本本质，或者哪怕是它的一些主要几何和物理特性的。这种不确定性部分可以解决，但是大部分仍然存在。宇宙学理论应当承认这种不确定性。

10.2 我们能够真正断言什么

宇宙学考虑的是物理起源问题，这种起源发生在唯一存在的物理宇宙(第6节)中，它提供了我们赖以生存的环境第(第7节、9.1节)。如果我们愿意，这些问题可以扩展至包含最终问题(8.2节)，但是物理学理论却无法解决它们(9.1.7节)。最终，各种各样的谜团隐含于宇宙的存在和本质中(9.3节)。宇宙学的科学研究有助于阐明它们的性质，但却不能解决它们。

除了庆祝宇宙学的成就，人们还应当充分重视这一章中考虑到的限制和问题，不去宣扬它实际上能够达到的成就，也不去宣扬它实际上可以达到的确定性。否则长期下来会削弱宇宙学作为一个有着牢固科学成绩的项目的合法地位。在宇宙学的“标准模型”方面，这种地位可以得到有力辩护，假定这种标准模型具有保守的特征，这样它就不会受到理论或者数据的相对详细的转移的威胁。这些转移事实上并不会威胁到宇宙学的核心。而且这种辩护必须充分认识到，如果有人把宇宙学理论的解释作用推到极限(第6节)，会出现很难的哲学问题。比如，对于多重宇宙的可能存在(9.2节)，就不应当作出过强的科学论断。如果是这样的话，它们的基于哲学的合理论据就是好的。如果宣称它可以得到比它事实上可以达到的更多的解释力，或者宣称它的断言已经被证实，而事实上其必需的证据又难以得到，而且在某些情况下会永远都是这样，那么宇宙学就仍不完善。

---

① 假定量子不确定性实际上是本体论的而不是认识论的。但人们对此应当保持一种开放的思想，因为正是当前的信条并不必然意味着它为真。

致谢

感谢比尔·斯特格、马丁·波究瓦德、马尔科姆·麦卡勒姆(Malcolm MacCallum)、亨克·范·埃尔斯特、杰里米·巴特菲尔德(Jeremy Butterfield)以及约翰·厄尔曼，感谢他们对改进此文所提出的建设性意见！

所用的缩写名词：

CBR：宇宙黑体辐射

CDM：冷暗物质

EFE：爱因斯坦场方程

FL：弗里德曼—勒梅特(宇宙模型)

HBB：热大爆炸

LSS：最后的散射面

RW：罗伯森—沃克(几何)

SAP：强人择原理

WAP：弱人择原理

参考文献

[Allday, 2002] J. Allday. *Quarks, Leptons and the Big Bang*. Institute of Physics, 2002.

[Balashov, 1991] Y. V. Balashov. Resource letter Ap-1: The anthropic principle. *American Journal of Physics*, 54: 1069-1076, 1991.

[Barbour and Pfister, 1995] J. Barbour and H. Pfister. *Mach's Principle: From Newton's Bucket to Quantum Gravity*. Birkhäuser, Basel, 1995.

[Bardeen, 1980] J. M. Bardeen. Gauge-invariant cosmological perturbations. *Phys. Rev.*, D22: 1882-1905, 1980.

[Barnett *et al.*, 1996] R. M. Barnett *et al.* (Particle Data Group) Particle physics summary. A digest of the 1996 Review of Particle Physics. *Reviews of Modern Physics*, 68: 611-732, 1996.

[Barrow and Tipler, 1984] J. Barrow and F. Tipler. *The Cosmological Anthropic Princi-*

ple. Oxford University Press, 1984.

[Barrow, 2003] J. D. Barrow. *The Constants of Nature*. Vintage, 2003.

[Bennet *et al.*, 2003] C. L. Bennet *et al.* (the WMAP team). First year Wilkinson Microwave Anisotropy Probe (WMAP) observations: Preliminary maps and basic results. *Astrophysical Journal Supplement Series*, 148: 1-27, 2003. astro-ph/0302207.

[Berendzen *et al.*, 1976] R. Berendzen, R. Hart and D. Seeley. *Man Discovers the Galaxies*. Science History Publications, New York, 1976.

[Bernadete, 1964] J. A. Bernadete. *Infinity, An Essay in Metaphysics*. Clarendon Press, Oxford, 1964.

[Blau and Guth, 1987] S. K. Blau and A. H. Guth. Inflationary cosmology. In Hawking, S. W. and Israel, W., editors, 300 *Years of Gravitation*, page 524. Cambridge University Press, 1987.

[Bojowald, 2001] M. Bojowald. Absence of singularity in loop quantum cosmology. *Physical Review Letters*, 86: 5227-5239, 2001.

[Bondi, 1947] H. Bondi. Spherically symmetric models in general relativity. *Monthly Notice of the Royal Astronomical Society*, 107: 410, 1947.

[Bondi, 1960] H. Bondi. *Cosmology. Cambridge University Press*, Cambridge, 1960.

[Bonnor and Ellis, 1986] W. B. Bonnor and G. F. R. Ellis. Observational homogeneity of the universe. *Monthly Notice of the Royal Astronomical Society*, 218: 605-614, 1986.

[Bostrom, 2002] N. Bostrom. *Anthropic bias: Observation selection effects in science and philosophy*. Routledge, New York, 2002.

[Bothun, 1998] G. Bothun. *Modern Cosmological Observations and Problems*. Taylor and Francis, 1998.

[Brading and Castellani, 2006] K. Brading and E. Castellani. This vol, Ch. 13, 2006.

[Brans and Dicke, 1961] C. Brans and R. H. Dicke. Mach's principle and a relativistic theory of gravitation. *Physical Review*, 124: 925, 1961.

[Carr and Rees, 1979] B. J. Carr and M. J. Rees. The anthropic principle and the structure of the physical world. *Nature*, 278: 605-612, 1979.

[Carroll, 2004] J. Carroll, editor. *Readings on Laws of Nature*. University of Pittsburgh Press, 2004.

[Carroll and Chen, 2005] S. M. Carroll and J. Chen. Does inflation provide natural initial conditions for the universe? gr-qc/0505037, 2005.

[Challinor and Lasenby, 1998] A. D. Challinor and A. N. Lasenby. Cosmic microwave background anisotropies in the CDM model: A covariant and gauge-invariant approach. *Physical Review D*, 58. astro-ph/9804301, 1998.

[Coleman and de Luccia, 1980] S. Coleman and F. de Luccia. Gravitational effects on and of vacuum decay. *Physical Review D*, 21: 3305-3315, 1980.

[Copeland *et al.*, 2005] E. J. Copeland, J. E. Lidsey, and S. Mizuno. Correspondence between loop-inspired and braneworld cosmology. gr-qc/0510022, 2005.

[Cowie and Songaila, 1995] L. Cowie and A. Songaila. Astrophysical limits on the evolution of dimensionless physical constants over cosmological time. *Astrophysics Journal*, 453: 596-598, 1995.

[Davies, 1974] P. C. W. Davies. *The physics of time asymmetry.* Surrey University Press, London, 1974.

[Davies, 1982] P. C. W. Davies. *The Accidental Universe.* Cambridge University Press, 1982.

[Davies, 1987] P. C. W. Davies. *The cosmic Blueprint.* Heinemann, London, 1987.

[Davis and Lineweaver, 2004] T. M. Davis. and C. H. Lineweaver. Expanding confusion: common misconceptions of cosmological horizons and the superluminal expansion of the universe. *Publications of the Astronomical Society of Australia*, 21: 97-109, 2004. astro-ph/0310808.

[de Vaucouleurs, 1970] G. de Vaucouleurs. The case for a hierarchical cosmology. *Science*, 167: 1203-1213, 1970.

[Deutsch, 1998] D. Deutsch. *The Fabric of reality: The science of parallel universes-and its implications.* Penguin, 1998.

[Dicke and Peebles, 1979] R. H. Dicke. and P. J. E. Peebles. The big bang cosmology-enigmas and nostrums. In Hawking, S. W. and Israel, W., editors, *General Relativity: An Einstein Centenary Survey*, pages 504-517. Cambridge University Press, Cambridge, 1979.

[Dickson, 2006] M. I. Dickson. this vol , Ch. 4, 2006.

[Dirac, 1938] P. A. M. Dirac. New basis for cosmology. *Proceedings of the Royal Society of LondoAn*, 165: 199-208, 1938.

[Disney, 1976] M. J. Disney. Visibility of galaxies. *Nature*, 263: 573-575, 1976.

[Dodelson, 2003] S. Dodelson. *Modern Cosmology*. Academic Press, 2003.

[Durrer, 2002] R. Durrer. Frontiers of the universe: What do we know and what do we understand? astro-ph/0205101, 2002.

[Earman, 1987] J. Earman. The SAP also rises: A critical examination of the anthropic principle. *American Philosophical Quarterly*, 24: 307-325, 1987.

[Earman, 1999] J. Earman. The Penrose-Hawking singularity theorems: History and implications. In Goenner, H., Renn, J., and Sauer, T., editors, *The Expanding Worlds of General Relativity*. Birkhäuser, 1999.

[Earman, 2001] J. Earman. Lambda: The constant that would not die. *Archive for History of Exact Sciences*, 55: 189-220, 2001.

[Earman, 2003] J. Earman. The cosmological constant, the fate of the universe, unimodular gravity, and all that. *Studies in History and Philosophy of Modern Physics*, 34: 559-577, 2003.

[Earman and Mosterin, 1999] J. Earman and J. Mosterin. A critical look at inflationary cosmology. *Philosophy of Science*, 66: 1-49, 1999.

[Ehlers, 1993] J. Ehlers. Contributions to the relativistic mechanics of continuous media. *General Relativity and Gravitation*, 25: 1225-1266, 1993.

[Ehlers *et al.*, 1968] J. Ehlers, P. Geren, and R. K. Sachs. Isotropic solutions of the Einstein-Liouville equations. *Journal of Mathematical Physics*, 9: 1344-1349, 1968.

[Ehlers and Rindler, 1989] J. Ehlers and W. Rindler. A phase-space representation of Friedmann-Lemaître universes containing both dust and radiation and the inevitability of a big bang. *Monthly Notice of the Royal Astronomical Society*, 238: 503-521, 1989.

[Eisenstein *et al.*, 2005a] D. J. Eisenstein *et al.* Detection of the baryon accoustic peak in the large scale correlation function of galaxies, 2005. astro-ph/0501171.

[Ellis, 1967] G. F. R. Ellis. The dynamics of pressure-free matter in general relativity. *Journal of Mathematical Physics*, 8: 1171-1194, 1967.

[Ellis, 1971a] G. F. R. Ellis. Relativistic cosmology. In Sachs, R. K., editor, *General Relativity and Cosmology, Proceedings of the XLVII Enrico Fermi Summer School*. Academic Press, New York, 1971.

[Ellis, 1971b] G. F. R. Ellis. Topology and cosmology. *General Relativity and Gravitation*, 2: 7-21, 1971.

[Ellis, 1975] G. F. R. Ellis. Cosmology and verifiability. *Quarterly Journal of the Royal Astronomical Society*, 16: 245-264, 1975.

[Ellis, 1980] G. F. R. Ellis. Limits to verification in cosmology. *Ann New York Acad Sci*, 336: 130-160, 1980.

[Ellis, 1984] G. F. R. Ellis. Relativistic cosmology: its nature, aims and problems. In *et al.*, B. B., editor, *General Relativity and Gravitation*, pages 215-288. Reidel, 1984.

[Ellis, 1989] G. F. R. Ellis. A history of cosmology 1917-1955. In Howard, D. and Stachel, J., editors, *Einstein and the History of General Relativity*, volume 1 of *Einstein Study Series*, pages 367-431. Birkhäuser, Boston, 1989.

[Ellis, 1990] G. F. R. Ellis. Innovation resistance and change: the transition to the expanding universe. In Bertotti, B., Balbinto, R., Bergia, S., and Messina., A., editors, *Modern Cosmology in Retrospect*, pages 97-114. Cambridge University Press, 1990.

[Ellis, 1991] G. F. R. Ellis. Major themes in the relation between philosophy and cosmology. *Mem Ital Ast Soc*, 62: 553-605, 1991.

[Ellis, 2002] G. F. R. Ellis. Cosmology and local physics. *New Astronomy Reviews*, 46: 645-658, 2002. gr-qc/0102017.

[Ellis, 2004] G. F. R. Ellis. True complexity and its associated ontology. In Barrow, J. D., Davies, P. C., and Jr., C. L. H., editors, *Science and Ultimate Reality: Quantum Theory, Cosmology and Complexity, honoring John Wheeler's 90th birthday.* Cambridge University Press, 2004.

[Ellis, 2005] G. F. R. Ellis. Physics, complexity, and causality. *Nature*, 435: 743, 2005.

[Ellis and Baldwin, 1984] G. F. R. Ellis. and J. Baldwin. On the expected anisotropy of radio source counts. *Monthly Notice of the Royal Astronomical Society*, 206: 377-381, 1984.

[Ellis and Brundrit, 1979] G. F. R. Ellis and G. B. Brundrit. Life in the infinite universe. *Quarterly Journal of the Royal Astronomical Society*, 20: 37-41, 1979.

[Ellis and Bruni, 1989] G. F. R. Ellis and M. Bruni. A covariant and gauge-free approach to density fluctuations in cosmology. *Physical Review D*, 40: 1804-1818, 1989.

[Ellis and Buchert, 2006] G. F. R. Ellis and T. R. Buchert. The universe seen at different scales. To appear, *Physics Letter A*, special Einstein issue, 2006. gr-qc/0506106.

[Ellis and King, 1974] G. F. R. Ellis and A. R. King. Was the big bang a whimper? *Communications in Mathematical Physics*, 38: 119-165, 1974.

[Ellis *et al.*, 2004] G. F. R. Ellis, U. Kirchner, and W. Stoeger. Multiverses and physical cosmology. *Monthly Notice of the Royal Astronomical Society*, 347: 921-936, 2004.

[Ellis and Maartens, 2004] G. F. R. Ellis and R. Maartens. The emergent universe: inflationary cosmology with no singularity and no quantum gravity era. *Classical and Quantum Gravity*, 21: 223-232, 2004.

[Ellis *et al.*, 1978] G. F. R. Ellis, R. Maartens, and S. D. Nel. The expansion ofthe universe. *Monthly Notice of the Royal Astronomical Society*, 184: 439-465, 1978.

[Ellis and Madsen, 1991] G. F. R. Ellis and M. Madsen. Exact scalar field cosmologies. *Classical and Quantum Gravity*, 8: 667-676, 1991.

[Ellis *et al.*, 2002a] G. F. R. Ellis, P. McEwan, W. Stoeger, and P. Dunsby. Causality in inflationary universes with positive spatial curvature. *General Relativity and Gravitation.*, 34: 1461-1481, 2002. gr-qc/0109024.

[Ellis *et al.*, 1985] G. F. R. Ellis, S. D. Nel, W. Stoeger, R. Maartens, and A. P. Whitman. Ideal observational cosmology. *Physics Reports*, 124: 315-417, 1985.

[Ellis *et al.*, 1984] G. F. R. Ellis, J. J. Perry, and A. Sievers. Cosmological observations of galaxies: the observational map. *Astronomical Journal*, 89: 1124-1154, 1984.

[Ellis and Schreiber, 1986] G. F. R. Ellis and G. Schreiber. Observational and dynamic properties of small universes. *Physics Letters A*, 115: 97-107, 1986.

[Ellis and Sciama, 1972] G. F. R. Ellis and D. W. Sciama. Global and non-global problems in cosmology. In O'Raifeartaigh, L., editor, *General Relativity*, *pages* 35-59. Oxford University Press, 1972.

[Ellis *et al.*, 2002b] G. F. R. Ellis, W. Stoeger, P. McEwan, and P. Dunsby. Dynamics of inflationary universes with positive spatial curvature. *General Relativity and Gravitation*, 34: 1445-1459, 2002. gr-qc/0109023.

[Ellis and Stoeger, 1987] G. F. R. Ellis and W. R. Stoeger. The Fitting problem in cosmology. *Classical and Quantum Gravity*, 4: 1679-1690, 1987.

[Ellis and Stoeger, 1988] G. F. R. Ellis and W. R. Stoeger. Horizons in inflationary universes. *Classical and Quantum Gravity*, 5: 207, 1988.

[Ellis and Uzan, 2005] G. F. R. Ellis and J. -P. Uzan. 'c' is the speed of light, isn't it? *American Journal of Physics*, 73: 240-247, 2005.

[Ellis and van Elst, 1999a] G. F. R. Ellis and H. van Elst. Cosmological models (cargese lectures 1998). In Lachièze-Ray, M., editor, *Theoretical and Observational Cosmology*, *volume* 541, pages 1-116. Kluwer, Nato Series C: Mathematical and Physical Sciences, 1999. gr-qc/9812046.

[Ellis and van Elst, 1999b] G. F. R. Ellis and H. van Elst. Deviation of geodesics in FLRW spacetime geometries. In Harvey, A., editor, *On Einstein's Path: Essays in Honour of Englebert Schucking*, pages 203-226. Springer, 1999.

[Ellis and Williams, 2000] G. F. R. Ellis and R. M. Williams. *Flat and Curved Spacetimes.* Cambridge Univesity Press, 2000.

[Fabian, 1989] A. C. Fabian, editor. *Origins.* Cambridge University Press, 1989.

[Freivogel *et al.*, 2005a] B. Freivogel, M. Kleban, M. R. Martinez, and L. Susskind. Observational consequences of a landscape. hep-th/0505232, 2005.

[Freivogel *et al.*, 2005b] B. Freivogel, M. Kleban, M. R. Martinez, and L. Susskind. Observational consequences of a landscape. hep-th/0505232, 2005.

[Gardner, 1986] M. Gardner. WAP, SAP, PAP, and FAP. *The New York Review of Books*, 23(8): 22-25, 1986.

[Gardner, 2003] M. Gardner. *Are Universes Thicker than Blackberries*? W W Norton, 2003.

[Garrett and Coles, 1993] A. J. M. Garrett and P. Coles. Bayesian inductive inference and the anthropic cosmological principle. *Comments Astrophys*, 17: 23-47, 1993.

[Ghosh *et al.*, 1998] A. Ghosh, G. Pollifrone, and G. Veneziano. Quantum fluctuations in open pre-big bang cosmology. *Physics Letter Bs*, 440: 20-27, 1998.

[Gibbons *et al.*, 1983] G. W. Gibbons, S. W. Hawking, and S. T. C. Siklos. *The Very Early Universe.* Cambridge University Press, 1983.

[Gibbons *et al.*, 2003] G. W. Gibbons, E. P. S. Shellard, and S. J. Rankin, editors. *The Future of Theoretical Physics and Cosmology: Celebrating Stephen Hawking's 60th Birthday.* Cambridge University Press, 2003.

[Goodman, 1995] J. Goodman. Geocentrism reexamined. *Physical Review D*, 52: 1821-

1827, 1995.

[Gott and Li, 1998] J. R. Gott and L.-X. Li. Can the universe create itself? *Physical Review D*, 58: 023501, 1998. astro-ph/9712344.

[Gribbin and Rees, 1991] J. Gribbin and M. Rees. *Cosmic Coincidences.* Black Swan, 1991.

[Grunbaum, 1989] A. Grunbaum. The pseudo problem of creation. *Philosophy of Science*, 56: 373-394, 1989. reprinted in [Leslie, 1998].

[Grunbaum, 2004] A. Grunbaum. The poverty of theistic cosmology. *British Journal of Philosophy of Science*, 55: 561-614, 2004.

[Gunn *et al.*, 1978] J. E. Gunn, M. S. Longair and M. J. Rees. *Observational cosmology: Saas-Fee Course* 8. Geneva Observatory, Switzerland, 1978.

[Guth, 1997] A. H. Guth. The Inflationary Universe: *The Quest for a new Theory of Cosmic Origins.* Addison Wesley, 1997.

[Guth, 1981] A. H. Guth. Inflationary universe: A possible solution to the horizon and flatness. *Physical Review D*, 23: 347, 1981.

[Guth, 2001] A. H. Guth. Eternal inflation. In *Proceedings of "Cosmic Questions" Meeting*, volume 95. The New York Academy of Sciences Press, 2001. astro-ph/0101507.

[Harris *et al.*, 1998] W. E. Harris, P. R. Durell, M. J. Pierce, and J. Seckers. Constraints in the Hubble constant from observations of the brightest red giant stars in a Virgo-cluster galaxy. *Nature*, 395: 45, 1998.

[Harrison, 2000] E. R. Harrison. *Cosmology: The Science of the Universe.* Cambridge University Press, Cambridge, 2 edition, 2000.

[Hartle, 2004] J. B. Hartle. Anthropic reasoning and quantum cosmology. In Allen, R. and Pope, C., editors, *The New Cosmology.* AIP, 2004. gr-qc/0406104.

[Harvey, 2006] J. Harvey. this vol, Ch. 14, 2006.

[Harwit, 1984] M. Harwit. *Cosmic Discovery.* MIT Press, 1984.

[Hawking, 1987] S. W. Hawking. Quantum cosmology. In Hawking, S. W. andIsrael, W., editors, 300 *Years of Gravitation*, page 631. Cambridge University Press, Cambridge, 1987.

[Hawking, 1993] S. W. Hawking. *On the Big Bang and Black Holes.* World Scientific, 1993.

[Hawking and Ellis, 1973] S. W. Hawking and G. F. R. Ellis. *The Large Scale Structure of Space-time.* Cambridge University Press, 1973.

[Hilbert, 1964] D. Hilbert. On the infinite. In Benacerraf, P. and Putnam, H. , editors, *Philosophy of Mathematics*, pages 134-151. Englewood Cliff, N J. : Prentice Hall, 1964.

[Hogan, 1999] C. J. Hogan. Where the baryons are. astro-ph/9912107, 1999.

[Hogan, 2003] C. J. Hogan. Why the universe is just so. *Reviews of Modern Physics*, 72: 1149-1161, 2003.

[Hollands and Wald, 2002] S. Hollands and R. M. Wald. An alternative to inflation. *Gen Rel. Grav*, 34: 2043-2055, 2002.

[Hoyle, 1960] F. Hoyle. Cosmological tests of gravitational theories. In *Rendiconti Scuola Enrico Fermi XX Corso*, pages 141-. Academic Press, New York, 1960.

[Hu and Sugiyama, 1995a] W. Hu and N. Sugiyama. Anisotropies in the CMB: An analytic approach. *Astrophysical Journal*, 444: 489-506, 1995.

[Hu and Sugiyama, 1995b] W. Hu and N. Sugiyama. Toward understanding CMB anisotropies and their implications. *Physical Review D*, 51: 2599-2630, 1995.

[Hubble, 1936] E. Hubble. *The Realm of the Nebulae.* Yale University Press, 1936.

[Hume, 1993] D. Hume. *Dialogues and Natural History of Religion.* World Classics. Oxford University Press, 1993.

[Isham, 1997] C. J. Isham. *Lectures on Quantum Theory: Mathematical and Structural Foundations.* Imperial College Press, London, 1997.

[Jain and Dev, 2005] D. Jain and A. Dev. Age of high redshift objects-a litmus test for dark energy models. astro-ph/0509212, 2005.

[Kamionkowski and Loeb, 1997] M. Kamionkowski and A. Loeb. Getting around cosmic variance. *Physical Review D*, 56: 4511, 1997.

[Kantowski, 1998] R. Kantowski. The effects of inhomogeneities on evaluating the mass parameter w_m and the cosmological constant λ. *Astrophysical Journal*, 507: 483-496, 1998.

[Kantowski *et al.* , 1995] R. Kantowski, T. Vaughan, and D. Branch. The effects of inhomogeneities on evaluating the deceleration parameter. *Astrophysical Journal*, 447: 35-42, 1995.

[Khoury *et al.* , 2001] J. Khoury, B. A. Ovrut, P. J. Steinhardt, and N. Turok. The ekpyrotic universe: Colliding branes and the origin of the hot big bang. *Physical Review D*, 64: 123522, 2001.

[Knobe *et al.* , 2005] J. Knobe, K. D. Olum, and A. Vilenkin. Philsophical implications of inflationary cosmology. phys/0320071, 2005.

[Kolb and Turner, 1990] E. W. Kolb and M. S. Turner. *The Early Universe*. Addison Wesley, 1990.

[Kragh, 1996] H. Kragh. *Cosmology and Controversy*. Princeton University Press, Princeton, 1996.

[Kristian and Sachs, 1966] J. Kristian and R. K. Sachs. Observations in cosmology. *Astrophysical Journal*, 143: 379, 1966.

[Kuhn, 1977] T. Kuhn. Objectivity, value judgment, and theory choice. In *The Essential Tension*, pages 320-339. University of Chicago Pressm 1977.

[Lachièze *et al.*, 1995] M. M. R. Lachièze and J. P. Luminet. Cosmic topology. *Physics Reports*, 254: 135, 1995.

[Lakatos, 1980] I. Lakatos. *The Methodology of Scientific Research Programmes*. Cambridge University Press, 1980.

[Landsman, 2006] N. P. Landsman. this vol, ch. 5, 2006.

[Langlois and Vernizzi, 2005] D. Langlois and F. Vernizzi. Conserved non-linear quantities in cosmology. astro-ph/0509078, 2005.

[Lemaître, 1931] G. Lemaître. The beginning of the world from the point of view of quantum theory. *Nature*, 127: 706, 1931.

[Leslie, 1989] J. Leslie. *Universes*. Routledge, 1989.

[Leslie, 1994] J. Leslie. Cosmology: a philosophical survey. *Philosophia*, 24: 3-27, 1994.

[Leslie, 1998] J. Leslie, editor. *Modern Cosmology and Philosophy*. Prometheus Books, 1998.

[Lewis, 1986] D. K. Lewis. *On the Plurality of Worlds*. Basil Blackwell, Oxford, 1986.

[Liddle, 2004] A. R. Liddle. How many cosmological parameters? astro-ph/0401198, 2004.

[Lidsey *et al.*, 1997] J. E. Lidsey, A. R. Liddle, E. W. Kolb, E. J. Copeland, T. Barriero, and M. Abney. Reconstructing the inflation potential-an overview. *Reviews of Modern Physics*, 69,: 373-410, 1997.

[Linde, 1990] A. D. Linde. *Particle Physics and Inflationary Cosmology*. Harwood Academic, 1990.

[Lineweaver and Davis, 2005] C. H. Lineweaver and T. M. Davis. Misconceptions about the big bang. *Scientific American*, 2005.

[Luminet, 2005] J. -P. Luminet. A cosmic hall of mirrors. *Physics World*, 18: 23-28, 2005.

[Luminet *et al.*, 2003] J. -P. Luminet, J. R. Weeks, A. Raizuelo, R. Lehoucq, and

J. P. Uzan. Dodecahedral space topology as an explanation for weak wide-angle temperature correlations in the cosmic microwave background. *Nature*, 425: 593-595, 2003.

[Madsen and Ellis, 1988] M. S. Madsen and G. F. R. Ellis. Evolution of Ω in inflationary universes. *Monthly Notice of the Royal Astronomical Society*, 234: 67-77, 1988.

[Malament, 2006] D. Malament. This vol ch. 3, 2006.

[Matravers *et al.*, 1995] D. R. Matravers, G. F. R. Ellis, and W. R. Stoeger. Complementary approaches to cosmology: Relating theory and observations. *Quarterly Journal of the Royal Astronomical Society*, 36: 29-45, 1995.

[McCrea, 1953] W. H. McCrea. Cosmology. *Reports on Progress in Physics*, 16: 321-363, 1953.

[McCrea, 1960] W. H. McCrea. The interpretation of cosmology. *Nature*, 186: 1035, 1960.

[McCrea, 1970] W. H. McCrea. A philosophy for big bang cosmology. *Nature*, 228: 21, 1970.

[Meyer, 1994] D. M. Meyer A distant space thermometer. *Nature*, 371: 13, 1994.

[Misner, 1968] C. W. Misner. The isotropy of the universe. *Astrophysical Journal*, 151: 431-458, 1968.

[Misner, 1969] C. W. Misner. Mixmaster universe. *Physical Review Letters*, 22: 1071-1074, 1969.

[Morowitz, 2002] L. Morowitz. *The emergence of everything: How the world came to be complex.* Oxford, 2002.

[Mouri and Taniguchi, 2005] H. Mouri and Y. Taniguchi. Hierarchical object formation in the peculiar velocity field. astro-ph/0508069, 2005.

[Mulryne *et al.*, 2005] D. J. Mulryne, R. Tavakol, J. E. Lidsey, and G. F. R. Ellis. An emergent universe from a loop. *Physical Review D*, 71: 123512, 2005.

[Munitz, 1962] M. K. Munitz. The logic of cosmology. *British Journal for the Philosophy of Science*, 13: 34-50, 1962.

[Mustapha *et al.*, 1998] N. Mustapha, C. Hellaby, and G. F. R. Ellis. Large scale inhomogeneity vs source evolution: Can we distinguish them? *Monthly Notice of the Royal Astronomical Society*, 292: 817-830, 1998.

[Nilsson *et al.*, 1991] J. S. Nilsson *et al.*, editors. *The Birth and Early Evolution of our Universe: Proceedings of Nobel Symposium* 19. World Scientific, 1991.

[North, 1965] J. D. North. *The Meaure of the Universe.* Oxford University Press, 1965.

[Padmanbhan, 1993] T. Padmanbhan. *Structure formation in the universe.* Cambridge University Press, 1993.

[Partridge, 1995] B. Partridge. *3K: The Cosmic Microwave Background Radiation.* Cambridge University Press, 1995.

[Peacock, 1999] J. A. Peacock. *Cosmological Physics.* Cambridge University Press, 1999.

[Peebles, 1971] P. J. E. Peebles. *Physical Cosmology.* Princeton University Press, Princeton, 1971.

[Peebles, 1993a] P. J. E. Peebles. *The Large-Scale Structure of the Universe.* Princeton University Press, 1993.

[Peebles, 1993b] P. J. E. Peebles. *Principles of Physical Cosmology.* Princeton University Press, 1993.

[Peebles *et al.*, 1991] P. J. E. Peebles, D. N. Schramm, E. L. Turner, and R. G. Kron. The case for the relativistic hot big bang cosmology. *Nature*, 352: 769-776, 1991.

[Penrose, 1989a] R. Penrose. Difficulties with inflationary cosmology. In Fergus, E., editor, *14th Texas Symposium Relativistic Astrophysics.* New York Academy of Science, New York, 1989.

[Penrose, 1989b] R. Penrose. *The Emperor's New Mind.* Oxford University Press, 1989.

[Penrose, 2004] R. Penrose. *The Road to Reality.* Jonathan Cape, 2004.

[Perez *et al.*, 2005] A. Perez, H. Sahlmann, and D. Sudarsky. On the quantum origin of the seeds of cosmic structure. gr-qc/0508100, 2005.

[Perkins, 2005] D. Perkins. *Particle Astrophysics.* Oxford, 2005.

[Perlmutter *et al.*, 1998] S. Perlmutter *et al.* Discovery of a supernova explosion at half the age of the universe. *Nature*, 391: 51, 1998.

[Pettini, 1999] M. Pettini. Element abundances at high redshifts. astro-ph/9902173, 1999.

[Rees, 1995] M. J. Rees. *Perspectives in Astrophysical Cosmology.* Cambridge University Press, 1995.

[Rees, 1999] M. J. Rees. *Just six numbers: The Deep Forces that shape the universe.* Weidenfeld and Nicholson, London, 1999.

[Rees, 2003] M. J. Rees. *Our Cosmic Habitat.* Princeton University Press, Princeton, 2003.

[Rindler, 1956] W. Rindler. Visual horizons in world models. *Monthly Notice of the Royal*

Astronomical Society, 116: 662, 1956.

[Rindler, 2001] W. Rindler. *Relativity: Special, General, and Cosmological.* Oxford University Press, 2001.

[Robertson, 1933] H. P. Robertson. Relativistic cosmology. *Reviews of Modern Physics*, 5: 62-90, 1933.

[Robertson, 1935] H. P. Robertson. Kinematics and world structure. *Astrophysical Journal*, 82: 284-301, 1935.

[Rothman and Ellis, 1993] T. Rothman and G. F. R. Ellis. Lost horizons cosmology. *American Journal of Physics*, 61: 883-893, 1993.

[Rovelli, 2004] C. Rovelli. *Quantum Gravity.* Cambridge University Press, 2004.

[Rovelli, 2006] C. Rovelli. This vol ch. 12, 2006.

[Rugh and Zinkernagel, 2002] S. E. Rugh and H. Zinkernagel. The quantum vacuum and the cosmological constant problem. *Studies in History and Philosophy of Modern Physics*, 33: 663-705, 2002.

[Saini *et al.*, 2000] T. D. Saini, S. Raychaudhury, V. Sahni, and A. A. Starobinsky. Reconstructing the cosmic equation of state from supernova distances. *Physical Review Letters*, 85: 1162-1165, 2000.

[Sandage, 1961] A. Sandage. The ability of the 200-inch telescope to distinguish between selected world models. *Astrophysical Journal*, 133: 355-392, 1961.

[Sandage *et al.*, 1993] A. Sandage, R. G. Kron, and M. S. Longair. *The Deep Universe: Saas-Fee Advanced Course* 23. Springer, Berlin, 1993.

[Schneider *et al.*, 1992] P. Schneider, J. Ehlers, and E. E. Falco. *Gravitational Lenses.* Springer, Berlin, 1992.

[Schramm and Turner, 1998] D. N. Schramm and M. S. Turner. Big-bang nucleosynthesis enters the precision era. *Reviews of Modern Physics*, 70: 303-318, 1998.

[Schrödinger, 1956] E. Schrödinger. *Expanding Universes.* Cambridge University Press, Cambridge, 1956.

[Sciama, 1971] D. W. Sciama. Astrophysical cosmology. In Sachs, R. K., editor, *General Relativity and Cosmology*, pages 183-236. Academic Press, New York, 1971.

[Scott, 2005] D. Scott. The standard cosmological model. astro-ph/0510731, 2005.

[Seife, 2000] C. Seife. *Zero: the Biography of a Dangerous Idea*. Penguin, 2000.

[Seife, 2003] C. Seife. *Alpha and Omega: The Search for the Begining and end of the Universe*. Bantam Books, London, 2003.

[Seife, 2004] C. Seife. Physics enters the twilight zone. *Science*, 305: 464-466, 2004.

[Sievers *et al.*, 2005b] J. L. Sievers *et al.* Implications of the cosmic background imager polarization data, 2005. astro-ph/0509203.

[Sigurdson and Furlanetto, 2005] K. Sigurdson and S. R. Furlanetto. Measuring the primordial deuterium abundance during the cosmic dark ages. astro-ph/0505173.

[Silk, 1997] J. Silk. *A Short History of the Universe*. Scientific American Library, 1997.

[Silk, 2001] J. Silk. *The Big Bang*. Freeman, 3 edition, 2001.

[Silk, 2005] J. Silk. *On the shores of the unknown: A short history of the universe*. Cambridge University Press, 2005.

[Smith, 1982] R. W. Smith. *The Expanding Universe: Astronomy's Great Debate*. Cambridge University Press, 1982.

[Smolin, 1992] L. Smolin. Did the universe evolve? *Classical and Quantum Gravity*, 9: 173-191, 1992.

[Spergel *et al.*, 2003] D. N. Spergel *et al.* First year Wilkinson microwave anisotropy probe (WMAP) observations: Determination of cosmological parameters. *Astrophysical Journal Supplement Series*, 148: 175-194, 2003. astro-ph/0302209.

[Starkman and Schwarz, 2005] G. D. Starkman and D. J. Schwarz. Is the universe out of tune? *Scientific American*, pages 48-55, 2005.

[Steinhardt, 1995] P. J. Steinhardt. Cosmology confronts the cosmic microwave background. *International Journal of Modern Physics*, 10: 1091-1124, 1995. astro-ph/0407329.

[Stoeger *et al.*, 1995] W. Stoeger, R. Maartens, and G. F. R. Ellis. Proving almost-homogeneity of the universe: an almost-Ehlers, Geren and Sachs theorem. *Astrophysical Journal*, 443: 1-5, 1995. gr-qc/0508100.

[Stoeger *et al.*, 2004] W. R. Stoeger, G. F. R. Ellis, and U. Kirchner. Multiverses and cosmology: Philosophical issues. astro-ph/0407329, 2004.

[Susskind, 2003] L. Susskind. The anthropic landscape of string theory. hep-th/0302219, 2003.

[Susskind, 2005] L. Susskind. *The Cosmic Landscape: String Theory and the Illusion of Intelligent Design.* Little Brown, New York, 2005.

[Susskind and Lindesay, 2004] L. Susskind and J. Lindesay. *An Introduction To Black Holes, Information And The String Theory Revolution: The Holographic Universe.* World Scientific, 2004.

[Tegmark, 1998] M. Tegmark. Is "the theory of everything" merely the ultimate ensemble theory? *Annals of Physics*, 270: 1-51, 1998.

[Tegmark, 2002] M. Tegmark. Measuring spacetime: from big bang to blackholes. *Science*, 296: 1427-1433, 2002. astro = ph/0207188.

[Tegmark, 2003] M. Tegmark. Parallel universes. In J. D. Barrow, P. C. W. D. and Harper, C., editors, *Science and Ultimate Reality: Quantum Theory, Cosmology, and Complexity*, pages 459-491. Cambridge University Press, 2003. astro-ph/0302131.

[Tegmark *et al.*, 2004] M. Tegmark (the SDSS collaboration). Cosmological parameters from SDSS and WMAP. *Physical Review D*, 69: 103501, 2004.

[Tegmark *et al.*, 2003] M. Tegmark, A. Vilenkin and L. Pogosan. Anthropic predictions for neuturinos masses. astro-ph/0304536, 2003.

[Tipler *et al.*, 1980] F. J. Tipler, C. J. S. Clarke, and G. F. R. Ellis. Singularities and horizons: A review article. In Held, A., editor, *General Relativityand Gravitation.* Plenum Press, 1980.

[Tolman, 1934] R. C. Tolman. Relativity, *Thermodynamics, and Cosmology.* Clarendon Press, Oxford, 1934. Dover Publications, 1987.

[Tryon, 1973] E. P. Tryon. Is the universe a quantum fluctuation? *Nature*, 246: 396, 1973.

[Uggla *et al.*, 2003] C. Uggla, H. van Elst, J. Wainwright, and G. F. R. Ellis. The past attractor in inhomogeneous cosmology. *Phys Rev.*, D68: 103502 (1-22), 2003.

[Uffink, 2006] J. Uffink. This vol. Ch. 9, 2006.

[Vilenkin, 1982] A. Vilenkin. Creation of universes from nothing. *Physics Letters B*, 117: 25, 1982.

[Wainright *et al.*, 1998] J. Wainright, A. A. Coley, G. F. R. Ellis, and M. Hancock. On the isotropy of the universe: Do Bianchi VIIh universes isotropize? *Classical and Quantum Gravity*, 15: 331, 1998.

[Wainwright and Ellis, 1996] J. Wainwright, and G. F. R. Ellis. *The dynamical systems* *approach to cosmology*. Cambridge University Press, 1996.

[Wald, 2005] R. M. Wald. The arrow of time and the initial conditions for the universe. grqc/0507094, 2005.

[Wald, 1983] R. M. Wald. Asymptotic behavior of homogeneous cosmological models in the presence of a positive cosmological constant. *Physical Review*, 28: 2118, 1983.

[Walker, 1936] A. G. Walker. On Milne's theory of world structure. *Proceedings of the London Mathematical Society*, 42: 90, 1936.

[Walker, 1944] A. G. Walker. Completely symetric spaces. *Journal of the London Mathematical Society*, 19: 219-226, 1944.

[Weeks *et al.*, 2003] J. Weeks, J. -P. Luminet, A. Riazuelo, and R. Lehoucq. Well-proportioned universes suppress CMB quadrupole. astro-ph/0312312, 2003.

[Weinberg, 1989] S. W. Weinberg. The cosmological constant problem. *Reviews of Modern Physics*, 61: 1-23, 1989.

[Weinberg, 1972] S. W. Weinberg. *Gravitation and Cosmology*. Wiley, New York, 1972.

[Weinberg, 2000a] S. W. Weinberg. The cosmological constant problem. astro-ph/0005265, 2000.

[Weinberg, 2000b] S. W. Weinberg. A priori distribution of the cosmological constant. *Physical Review D*, 61: 103505. astro-ph/0002387, 2000.

[Wheeler, 1968] J. A. Wheeler. *Einstein's Vision*. Springer, Berlin, 1968.

[White *et al.*, 1993] M. White, L. M. Krauss, and J. Silk. Cosmic variance in CMB anisotropies: From 1°to COBE. *Astrophysical Journal*, 418: 535-543, 1993.

[Will, 1979] C. M. Will. The confrontation between gravitational theory and experiment. In Hawking, S. W. and Israel, W., editors, *General Relativity: an Einstein Centenary Survey*, pages 24-89. Cambridge University Press, 1979.

[Wilson, 1989] C. Wilson. *Leibniz's Metaphysics: A Historical and Comparative Study*. Princeton University Press, 1989.

[Wright, 2006] E. L. Wright. A century of cosmology. Astro-ph/0603750.

[Yoshimura, 1988] M. Yoshimura. The universe as a laboratory for high energy physics. In Fang, L. -Z. and Zee, A., editors, *Cosmology and Particle Physics*, pages 293-340. Gordon and Breach, 1988.

[Zeh, 1992] H. D. Zeh. *The Physical Basis of the direction of Time*. Springer Verlag, Berlin, 1992.
[Zinkernagel, 2002] H. Zinkernagel. Cosmology, particles, and the unity of science. *Studies in History and Philosophy of Modern Physics*, 33: 493-516, 2002.

问题和论点汇总表

问题 A：宇宙的唯一性

论点 A1：宇宙自身不能经受物理学实验的检验。

论点 A2：宇宙不能与其他宇宙进行观测对比。

论点 A3：只运用到一个对象之上的"物理学定律"这个概念是可疑的。

论点 A4：概率的概念在只存在一个对象的情况中是可疑的。

问题 B：宇宙在空间和时间上的巨大尺度

论点 B1：天文学的观测都被约束在过去零锥之中，随着距离的增大而衰退。

论点 B2："地质学"类型的观测可以探索我们过去世界线的遥远过去。

论点 B3：为宇宙确立一个罗伯森—沃克几何要依赖于貌似合理的哲学假设。

论点 B4：解释宇宙学观测要依赖于天体物理学的理解。

论点 B5：宇宙学的一种严格观测检验是，宇宙的年龄必须要大于行星的年龄。

论点 B6：观测范围限制了我们用观测决定宇宙极大尺度几何的能力。

论点 B7：我们在观测的完善性上已经取得了很大的进步。

问题 C：早期宇宙中的无约束能量

论点 C1：物理学视界限制了我们关于极早期宇宙的物理学知识。

论点 C2：暴胀的未知性质意味着暴胀宇宙的方案是不完备的。

问题 D：解释宇宙——起源的问题

论点 D1：初始奇点可能发生也可能没有发生过。

论点 D2：可检验的物理学解释不了初始态，因此也解释不了宇宙的特殊

性质。

论点 D3：宇宙的初始态可能特殊也可能一般。

问题 E：作为存在背景的宇宙

论点 E1：物理学定律可能依赖于宇宙的性质。

论点 E2：我们不能想当然地去看物理学定律的本质。

论点 E3：物理新颖性出现在膨胀宇宙中。

问题 F：明确的哲学基础

论点 F1：哲学的选择必然是宇宙学理论的基础。

论点 F2：理论满意度的标准不能科学地选择或者证实。

论点 F3：在把标准应用到令人满意的宇宙学理论时，冲突会不可避免地出现。

论点 F4：相信暴胀物理学的理由是它关于宇宙中结构增长的解释力。

论点 F5：宇宙学理论会有一个或宽或窄的调查范围。

论点 F6：不管是观测还是理论模型都不能完全地反映实在。

问题 G：有关人类的问题：生命的精细调节

论点 G1：生命是可能的，因为物理学定律和宇宙的边界条件都具有非常特殊的性质。

论点 G2：形而上学的不确定性仍然是关于宇宙学中的终极因果性的。

问题 H：多元宇宙的可能存在

论点 H1：多元宇宙的提议是不能通过观测或实验证明的，但是自洽性检验是可能的。

论点 H2：基于概率的论据证明不了多元宇宙的存在。

论点 H3：多元宇宙是一种哲学而不是科学思想。

论点 H4：宇宙学的基本物理学范式可能被推广到包含生物学见解。

问题 I：存在的本质

论点 I1：我们不了解宇宙早期或晚期的主要动力学物质组成。

论点 I2：经常所说的无限的物理学存在是有问题的。

论点 I3：宇宙学本质中的一个深刻问题是物理学定律的本质。

不确定性论题：基本的不确定性是宇宙学的一个关键环节。

第十二章　量子引力

卡罗尔·罗韦利

1. 引言

量子引力(QG)是一个试图找到描述引力的量子效应的理论的问题。这些效应超出了目前已被接受的物理学理论范畴。我们目前关于基本动力学原理的知识是由量子力学、量子场论、广义相对论以及粒子物理学的标准模型提供的。这组理论在科学史中已经在经验上获得了几乎无与伦比的成功：迄今为止没有任何观测现象明显地超出这组理论或者与之矛盾——或者，哪怕需要对它们进行些微的改动。但是，在期望相对论量子引力效应变得具有重大意义的领域，这些理论就失去意义了。量子引力效应并非当前所能观测到的，在目前可以达到的尺度范围内它们是可忽略的，人们认为它们只会在非常极端的物理学领域才有意义。比如，它们应当控制黑洞蒸发的结果、大爆炸最初期宇宙生命的起源以及在极小尺度(约 10^{-33} cm)或者极高能量下进行的任何测量。"量子引力"是人们对那个想要寻找可以描述这些领域的理论所赋予的名字。

但是这个问题所关注的范围远远超越了对一些目前无法实现的物理学的描述。20 世纪早期的物理学已经改变了我们对物

理世界理解的基础。它改变的正是我们用以理解它的那些概念的意义。广义相对论，这个在我们可以忽略引力的量子特性时对引力进行描述的场理论，改变了我们对空间和时间的理解。量子力学用我们关于运动的一般性理论替代了经典力学，并改变了物质、场和因果性的观念。现在我们还没有发现一种一致的概念框架能让这些改变同时具有意义。因此，我们对物理世界的理解在目前是严重破碎的。尽管基本物理学在经验上有效，但它正处于一种深刻的概念混乱的状态。量子引力问题就是要把广义相对论和量子力学的见解结合成为一个概念系统，在其中它们可以共处。这是一个要找到一种新的世界图景的问题，希望这个新的世界图景能够结束20世纪的科学革命。为此，许多人认为量子引力是基础物理学中最重要的开放问题。

特别地，量子引力是对空间和时间本质的探索。人们预期物理空间的结构和本质在普朗克(Planck)尺度下会发生极大的变化，时间演变概念化的传统方式在这个尺度下将会失去其可行性。因此，理论可能要求修正我们对空间和时间的思考方式。

由于广义相对论对物理学其他部分影响很小，并且许多物理学家的兴趣都集中在量子理论和粒子物理学的发展上，量子引力的研究在20世纪的几十年内发展非常缓慢。在最后的十年中，一方面由于经验的确证和具体的天体物理学和宇宙学、甚至是广义相对论技术应用的大繁荣，另一方面由于粒子物理学"标准模型"背景中粒子物理学的大部分困惑得到了满意的解决方法，导致了人们对量子引力强烈的兴趣集中，它的发展进程变得迅速起来。这一研究目前非常活跃。

量子引力的一些尝试性理论已被提出。其中发展得最好的是弦理论[Green *et al.*(格林等)，1987]和圈量子引力[Rovelli(罗韦利)，2004]。其他探索的活跃方向包括非对易几何[Connes(孔耐)，1994]、动态三角化[Ambjorn *et al.*(安比约恩等)，1997]、自旋泡沫形式[Perez(佩雷斯)，2003](与圈方法紧密联系)和有效理论。目前这些方法都没有得到任何经验的确证，也都没有在理论上获得一致意见。

量子引力提出了基本方法论的问题，包括概念和基础的问题。其中有些问题类似于物理学在另外一些重大概念变化——经典力学、场论、相对论或者量

子力学中出现的基础问题。老问题需要根据 20 世纪的新视野做出新的回答。举个典型的例子，关于空间相关本质的笛卡尔—牛顿—莱布尼茨争论(cartesian-newtonian-leibnizian)的复兴。

在我们当前关于物理世界、特别是关于空间和时间本质的争论中，任何一个详尽的讨论都不能忽略量子引力的研究提出来的这些争议和问题。

1.1　量子时空

广义相对论(GR)和量子力学(QM)已经广泛地扩展了我们对物理世界的理解。它们被实验依据稳固地支持着，也已经有了大量的科学和技术上的应用。但是它们也摧毁了前相对论的经典物理学所提供的一致的世界图景，因为它们各自是在彼此相互矛盾的假设下形式化的。量子力学的形式化运用了一个外在的时间变量，即薛定谔(Schrödinger)方程的 t；在量子场论中是运用了一个静止的、非动力学的背景时空。外在时间变量和静止的背景时空都是和广义相对论不相容的。

反过来，广义相对论是按照黎曼(Riem)几何术语形式化的：引力场被假设为一个经典决定论的动力学场，可以等同于黎曼几何的度规场。但是量子力学要求所有的动力学场都具有量子特性。在小尺度下，场看起来是由离散量子构成的，并由概率原理支配。

因此，广义相对论和量子力学是在相互矛盾的假设下形式化的。尽管它们获得了实验上的成功，但它们提供的对物理世界的理解还是相当分裂和混乱的。

大体上说，广义相对论告诉我们时空是一个动力学场，而量子力学则告诉我们动力学场是量子化的。因此在小尺度下，我们可以预期一种由“空间量子”组成的“量子时空”，并且允许“空间的量子叠加”。量子引力就是要给这样一种量子时空的观念以数学和概念的意义。

有些关于量子时空的本质以及这种观念引发的问题的一般指示，可以从基于广义相对论和量子力学的基本考虑得到。量子力学效应的尺度由普朗克常数 $\hbar$ 决定，引力强度由牛顿常数 G 决定，而相对论的范围是由光速 c 决定的，通过这三个基本常数的结合我们得到一个长度，叫作普朗克长度 $l_P = \sqrt{\hbar G/c^3} \sim$

10^{-33}厘米。量子引力效应在远远大于 l_P 的尺度下可以被忽略，因为在这些尺度下我们可以忽略 G、$\hbar$ 或者 $1/c$ 量级的量。因此我们希望，把时空作为黎曼空间的广义相对论时空描述在大于 l_P 的尺度下成立，而在接近这个尺度的时候就崩溃了。在此尺度下量子时空的结构变得具有重大意义。因此，量子引力是对普朗克尺度下时空结构的研究。

有简单的论据指出，l_P 可能起着最小长度的作用。同样，c 是最大速率，$\hbar$ 是最小交换作用。比如说，海森堡(Heisenberg)原理要求质量为 m 的物体的位置只能由满足$mvx > \hbar$ 的不确定性 x 决定，其中 v 为速度不确定性。狭义相对论要求 $v < c$，而根据广义相对论，区域 x 中可以聚集的质量有一个上限，遵循 $x > Gm/c^2$。超过这个限度，这个区域自身就会塌缩成为一个黑洞，消失于我们的观测之中。把这些不等式结合起来我们得到 $x > l_P$。也就是说，引力理论、相对论和量子理论放在一起，位置的确定性似乎会受到普朗克尺度的限制。诸如此类的考虑已经表明，空间可能并不是无限可分的。它在普朗克尺度下可能会存在一个间隔尺度，类似于能量在量子振荡器中的间隔单位。在某些量子引力理论比如圈量子引力中，空间的间隔尺度已经被充分地意识到了；而在弦理论中，也存在同样的迹象[Amati *et al.*(阿玛蒂等)，1989]。由于它是一个量子间隔尺度，就避免了与空间的原子性本质传统冲突。

在引力的量子化中，时间受到的影响更为彻底。传统量子力学中，时间被看作是一个外在参数，跃迁概率在时间中改变。广义相对论中，不存在外在时间参数。坐标时是一个不可观测的规范变量。时钟所测量的物理变量是引力场的一个复杂函数。因此量子引力的基本方程写出来并非是一个可观测时间变量中的演化方程。严格来说，在经典广义相对论中，这已经是事实了：广义相对论并不描述物理变量随时间的演化——它描述的是物理变量之间的相对演化。但是在经典广义相对论中由于时空作为引力场动力学方程的解出现，时间仍然需要解释。然而，动力学方程的一个解就像量子理论中粒子的一个“轨迹”，那里不存在物理轨迹，只存在可观测本征值之间的概率跃迁。因此在量子引力理论中，也许不可能用时空描述世界，同样，量子电子的运动也不能用单一的轨迹来描述。为了在普朗克尺度下理解世界，也为了给广义相对论和量子力学寻求一个一致的概念框架，我们可能必须完全放弃时间的观念，转而寻找用非时

间术语去描述世界的方法。时间也许只有在对物理实在的近似描述中才会是一个有用的概念。

接下来这一节概述量子引力研究的历史发展，说明了其主要思想、研究方法以及目前的尝试性理论。第3节将阐述这些研究带来的各种各样的问题。第4节特别讨论由量子引力所引起的空间和时间观念的变化。第5节讨论量子引力的问题与基础物理学中其他主要未解问题之间的关系。

2. 方法

完整说明关于量子引力的大量观点和方法不在本文讨论的范围。这里只阐述几个主要的研究方法，额外的参考文献可见文章最后的文献列表。

2.1 研究的历史和方向

早期思想

引力场应该具有量子特性，因此人们很早就意识到需要用一个理论来描述这些特性。1916年，即广义相对论诞生的第二年，爱因斯坦(Einstein)就已经指出量子效应必然会导致广义相对论的修正[Einstein, 1916]。1927年，奥斯卡·克莱因(Oskar Klein)指出，量子引力最终可能修正空间和时间的概念[Klein, 1927]。

20世纪30年代早期，罗森菲尔德(Rosenfeld)[Rosenfeld, 1930b; Rosenfeld, 1930a]写出了量子引力的第一篇专业论文，随后受到菲尔兹(Fierz)和泡利(Pauli)的追随[Fierz, 1939; Pauli and Fierz, 1939]，再后来是古普塔(Gupta)[1952]。其思想是引入一个虚构的“平坦空间”来考虑其周围度规的微小涨落——平坦空间上运动的引力波由线性爱因斯坦方程描述——并且运用电磁场中的方法来对这些波进行量子化。更加精确地说，爱因斯坦方程中同时表示时空度规和引力场的度规场 $g_{\mu\nu}(x)$ 可以写为以下两项之和：

$$g_{\mu\nu}(x) = \eta_{\mu\nu}(x) + h_{\mu\nu}(x) \tag{1}$$

$\eta_{\mu\nu}(x)$ 可解释为静止背景时空的度规；$h_{\mu\nu}(x)$ 可解释为量子化了的引力场。这里引入了表示引力波量子态的态希尔伯特(Hilbert)空间，其中 $h_{\mu\nu}(x)$ 由场算子 $\hat{h}_{\mu\nu}(x)$ 表示。这就是量子引力的“协变方法”。场 $h_{\mu\nu}(x)$ 的量子作为光子的引力

类似，被称为“引力子”，这个名字在30年代早期就开始使用了。

1938年，海森堡指出，引力耦合常数具有维度这一事实很可能为引力场的量子理论带来了麻烦[Heisenberg，1938]。30年代中期，年轻的俄国物理学家马特维·彼得洛维奇·布朗斯坦(Matvei Petrovich Bronstein)[Bronstein，1936]意识到在考虑完全非线性理论的时候，引力的独特特征需要特殊对待。场量子化的技术必须推广到可以应用于不存在时空背景的情形。布朗斯坦知道，广义相对论所给出的质量密度的极限彻底将其与量子电动力学理论区分开来，并且最终导致其需要“拒绝黎曼几何”，可能还要“拒绝我们通常的空间和时间的概念”。

研究的第二种方法是在40年代由彼得·伯格曼(Peter Bergmann)和他的小组开展的[Bergmann，1949a；Bergmann，1949b；1958；1961；Bergmann and Komar(伯格曼和考莫)，1980]。其思想是研究和量子化整个广义相对论的哈密顿(Hamiltonian)函数，而不只是量子化其围绕平坦时空的线性化函数。这种方法的优势在于它并不需要假设背景时空来定义理论。其思想是，希尔伯特空间中的态表示时空本身的量子态，而整个时空度规(可能取决于它的非动力学组分)变成了一个量子算符：

$$g_{\mu\nu}(x)\longrightarrow\hat{g}_{\mu\nu}(x) \tag{2}$$

这种方法叫作量子引力的“正则方法”。狄拉克(Dirac)也曾独立开始过相同的研究，他为这项工作发展了他的约束哈密顿系统理论[Dirac，1950；1964；1958；1959]。

量子引力的第三种方法是20世纪50年代末由查尔斯·米斯纳(Charles Misner)[Misner，1957]遵从约翰·惠勒(John Wheeler)的一项建议提出的，是按费曼(Feynman)的方式对广义相对论进行量子化。这种方式形式上由“几何上的路径积分”

$$Z=\int Dge^{-iS[g]} \tag{3}$$

所定义。此处g为度规场，$S[g]$是广义相对论的作用量。

这三种研究方法——协变、正则和路径积分分别由方程(1)、(2)和(3)表示——今天仍在继续应用。它们经常彼此影响，也间或有部分融合。但是在超过半个世纪的研究里，它们还是保持了不同的风格，并仍然具有明显的区别。

这三种研究方法的基本研究项目在50年代末就已经清晰地建立起来了。最初项目的执行被证明是一个相当艰巨的任务，但还是在60年代完成了，特别是协变方法中全套费曼规则和正则方法中惠勒—德威特(Wheeler-DeWitt)方程的写出。然而，所有这些方法在70年代都遇到了严重的困难：协变方法中的不可重整性、正则方法和路径积分方法中方程的欠明确性。这些困难在80年代得到了克服，特别是协变方向弦理论的发现、正则方向和路径积分方向圈量子引力和(相关的)自旋泡沫形式的发现。

发展的主要进程在下面进行介绍。

费曼规则和不可重整性

协变形式在60年代由费曼[Feynman，1963]、德威特[DeWitt，1964a；1964b；1965]、法捷耶夫(Faddeev)和波波夫(Popov)[Feynman and Popov，1967]发展起来。技术上的困难源于爱因斯坦方程的规范不变性，随着“ghost”粒子的引入而得到了解决，导致了微扰量子广义相对论的完整而一致的费曼规则，见[DeWitt，1967b；1967c；Faddeev and Popov，1967]。

但是在70年代早期，特霍夫特('t Hooft)、韦尔特曼(Veltman)以及后来登泽(Deser)和范·努伊淮曾(Van Nieuwenhuizen)的工作['t Hooft，1973；'t Hooft and Veltman，1974；Deser and Van Nieuwenhuizen，1974a；1974b]发现有迹象表明这个理论并不可行，印证了海森堡早年的担心。原因在于重整化的程序——也就是量子场论中用来消除考虑任意小(“紫外”)场涨落效应时出现的无穷大的方法——在引力情形下失效了。而广义相对论协变量子化方法由于不可重整的紫外发散而失效的明确而严密的证据，是在后来80年代晚期由高若夫(Goroff)和桑格诺提(Sagnotti)[Goroff and Sagnotti，1985；1986]得到的。

这一失效的解释仍然是有争议的。存在两种可能性。第一种是这个错误是由于要假定从背景时空的存在开始，无穷大来源于量子场的短距离涨落。这种可能性只有当时空在任意小尺度下都连续的情况下才会存在。但是引力是量子化的这一事实质疑了这种任意小尺度的存在。如果代之以时空具有量子化的颗粒状的小尺度结构，无穷大就可能只是方程(1)中把$\eta_{\mu\nu}(x)$考虑为时空度规所采取的近似造成的人为结果。假若如此，走出困境的方式就是抛弃背景时空，将整个引力场量子化，就如正则方法或者路径积分方法中所做的那样。

另一种可能性是说广义相对论才是那个不正确的理论。广义相对论得到在实验上的强有力支持，但只是在巨大的尺度下。在短距离时，世界可能由修正了的具有更好的紫外行为的广义相对论所描述。举一个历史的范例：弱相互作用的费米理论是一个经验上成功但却不可重整化的理论，这个问题的成功解决方式是用格拉肖—温伯格—萨拉姆(Glashow-Weinberg-Salam)弱电理论取代费米理论。格拉肖—温伯格—萨拉姆弱电理论是可重整化的，并且在小尺度下修正了费米理论。

受这一类比的激励，对广义相对论进行小尺度修正以使其具有更好的有限特性的研究已经跨越了好几十年。在大量尝试之后——其中，比如超引力[Freedman *et al.* (弗里德曼等)，1976]和高阶导数理论[Stelle(斯特拉)，1977]带给人们很高的希望，但后来失望了——研究走向了弦理论。

惠勒—德威特理论

在50和60年代，伯格曼的小组和狄拉克分别独立澄清了整个广义相对论的哈密顿结构，这是一项相当复杂的任务。这个结构后来又在阿诺维特(Arnowitt)、登泽和米斯纳[Arnowitt *et al.*，1962]的工作中运用定时类时曲面、在阿西提卡(Ashtekar)的工作中[Ashtekar，1986；1987]运用类似于杨—米尔斯(Yang-Mills)场的连接场以及在其他人的工作中得到澄清。

60年代早期，在以上结果的基础之上，佩雷斯[1962]写出了广义相对论的哈密顿—雅可比(Hamilton-Jacobi)方程：

$$G^2(q_{ab}q_{cd}-\frac{1}{2}q_{ac}q_{bd})\frac{\delta S(q)}{\delta q_{ac}}\frac{\delta S(q)}{\delta q_{bd}}+\det qR=0 \qquad (4)$$

其中R是度规q_{ab}的里奇标量曲率，$S(q)$是哈密顿—雅可比函数。1967年，德威特和惠勒效仿薛定谔从哈密顿—雅可比方程中产生薛定谔方程所采用的步骤，也就是，把哈密顿—雅可比方程解释为波函数，用导数算子替代导数得到的程函近似，写出了“爱因斯坦—薛定谔”方程[DeWitt，1967a]：

$$\left((\hbar G)^2(q_{ab}q_{cd}-\frac{1}{2}q_{ac}q_{bd})\frac{\delta}{\delta q_{ac}}\frac{\delta}{\delta q_{bd}}-\det qR\right)\Psi(q)=0 \qquad (5)$$

如今这个方程被称之为“惠勒—德威特方程”。原则上，这个方程被期望用来描述引力的整个量子力学。事实上，这个方程还远远没有达到很好地定义，直到

80年代晚期，泰德·雅各布森(Ted Jacobson)和李·斯莫林(Lee Smolin)[Jacobson and Smolin, 1988]发现了这个方程的类圈解，用阿西提卡的联络形式重新表示，才开启了通往圈量子引力的道路。

注意坐标时间变量 t 在经典方程(4)和量子方程(5)中都没有出现。时间变量的缺失引起了一场激烈的争论，我们在下面讨论。

米斯纳—霍金的求和几何

在70年代，史蒂芬·霍金(Steven Hawking)[1979]与他的小组在黎曼度规

$$Z = \int Dg\ e^{-S[g]} \tag{6}$$

上的"欧几里得(Euclid)"积分形式下得到并发展了惠勒—米斯纳路径积分(3)。希望欧几里得函数积分会被证明是一个比惠勒—德威特方程更好的计算工具。哈特(Hartle)和霍金[Hartle and Hawking, 1983]引入了"宇宙的波函数"和霍金积分的"无边界"边界条件的观念，为量子引力和量子宇宙学开启了一种新的直觉。哈特[1995]在量子力学向广义协变背景的充分延伸中提出了一种量子引力的历史求和公式的想法。这种思想后来被克里斯·艾沙姆(Chris Isham)[Isham, 1991]发展并进行了形式化。

但是欧几里得积分并没有为量子引力提供一种比惠勒—德威特方程更好的计算真实场理论量的方法。在80年代中期，引力领域中的气氛还是相当令人沮丧的。

量子引力的历史求和定义的思想在90年代由于自旋泡沫形式而重新复活了。这种形式提供了一种关于积分(3)的离散化的定义，这种定义看起来要好一些，这要感谢圈量子引力中同样的小尺度时空离散性。

黑洞热力学

1974年，霍金[Hawking, 1974; 1975]宣布了黑洞辐射的一个理论来源。一个(宏观的)质量为 M 的史瓦西(Schwarzschild)黑洞在温度 $T = \hbar/8\pi kGM$ 下发出热辐射，k 是玻耳兹曼(Boltzmann)常数。这个结论令人惊奇，只有贝肯斯坦(Bekenstein)一年前的一项观测对此有所预见。这项观测认为熵与黑洞是自然地相联系的[Bekenstein, 1972; 1973; 1974]，这种相关性是巴丁—卡特尔—霍金(Bardeen-Carter-Hawking)通过对热力学定律和黑洞的动力学行为之间进行

类比的分析[Bardeen *et al.*, 1973]所得到的。在霍金的结论中，史瓦西黑洞的贝肯斯坦熵为：

$$S=\frac{kc^3A}{4\hbar G} \tag{7}$$

这里 A 是黑洞表面的面积。霍金的这个美妙的结果并没有和量子引力联系起来——它是量子场论在弯曲空间中的技巧性应用，也就是量子场论与一个静止的、非量子化的引力场相互作用——但是对量子引力领域产生了强烈的冲击。它开启了一个新的研究领域——“黑洞热力学”——并且它开启了理解熵(7)的统计根源的量子引力问题。这对于任何引力的量子理论来说都是一个挑战。

两年后，比尔·盎鲁(Bill Unruh)[1976]发表了一篇颇有影响力的论文，有说服力地提出，传统量子场论真空态下加速的观测者与量子场会发生相互作用，就像在一个热浴缸里一样。这就清楚地说明了黑洞辐射，因为一个离黑洞距离固定的观测者是处于匀加速状态的(为了避免自由落体)，因此黑洞辐射可以仅仅被解释为一种盎鲁效应。但同时这个结果也似乎表明存在着一种我们还不了解的普遍关系，把引力、热动力学和量子理论连接在一起。

近年来，量子引力的弦理论和圈量子引力方法都能够从其基本原理导出方程(7)[Strominger and Vafa(斯特罗明格和瓦法)，1996；Rovelli，1996a；Krasnov(克拉斯诺夫)，1997；Ashtekar *et al.*，1998]，这已经被看作是两种方法的重大成功。然而，两种方法都不能完全让人满意。弦理论对于传统黑洞比如说史瓦西黑洞来说并不适用，只适用于某种叫作极值或者近极值的黑洞，圈量子引力给出了一个有限的结果，但是这个结果依赖于理论的一个自由参数(叫作 Immirz 参数 γ)，这个参数必须适当地选择才能得到式(7)中的参数 1/4。

非对易几何

一个几何空间 M 可以有两种不同的描述。一种是描述为点 x 的集合，另一种是用 M 上函数的对易代数 A 来描述。特别地，盖尔范德(Gelfand)的著名成果表明，一个紧致的豪斯道夫(Hausdorff)空间 M 是由与 M 上的连续函数代数同构的抽象代数 A 所决定的。这种代数观通过非对易代数的考虑引起了空间观念的广义化。在此意义上，是非对易代数定义了“非对易空间”。

量子理论是这样一种发现：动力学系统的相空间(由其经典态组成的集合)

必须被非对易空间替代。事实上，系统的可观测量——代表着我们可能与系统进行相互作用的方式——形成了经典相空间上函数的对易代数，在量子力学中，它们为非对易代数所取代。

在物理空间的情形中，A 可以用坐标代数或者动量空间来定义。如果我们把 A 的元素解释为表示了物理测量，那么按照量子力学，很自然地就会考虑代数非对易的可能性。相应地，就会得到这样一种假设：物理空间的小尺度结构可能由非对易几何描述。这种思想得到了各种各样的探索[Doplicher *et al.*（多普利克尔等），1994；1995；1996]。

在阿兰·孔耐[Connes，1994]提出的方法中出现了与量子引力的联系。孔耐发现在代数框架中，距离的观念被很自然地包含在狄拉克算子 D 中。这是一个出现于狄拉克电子旋量场方程中的导数算子。令$\mathcal{H}$为给定黎曼（自旋）流形 M 上的旋量场构成的希尔伯特空间，D 为（弯曲的）狄拉克算子，A 为 M 上函数的一个代数，被看作是$\mathcal{H}$上的（乘积）算子。由被称为“光谱三元组”的三元组（$\mathcal{H}$，A，D），我们可以重构黎曼流形。特别地，可以得到 X 和 Y 两点之间的距离：

$$d(x,\ y) = \sup_{\{f\in A,\ \|[D,f]\|<1\}} |\, x(f) - y(f)\,| \qquad (8)$$

这个关于距离的代数定义非常漂亮并令人惊奇。非对易时空可能由光谱三元组描述，在其中 A 是非对易的。孔耐指出，这个代数可能基于标准模型的对称性而被选择，符合于这样一种思想：标准模型可能以与麦克斯韦（Maxwell）理论揭示闵可夫斯基（Minkowskian）时空的结构相同的方式揭示了时空的小尺度结构。孔耐—查姆斯丁（Connes-Chamseddine）“光谱作用”只是狄拉克算子 D 的迹：$S = Tr[f(D^2/(\hbar G))]$，其中 f 是区间[0，1]上的特征函数。引人注目的是，这种作用最后被证明包含了标准模型的作用，包括很少被人理解的希格斯区域以及量子引力的作用[Chamseddine and Connes，1996；1997]。非对易几何的非对易性和量子力学非对易性之间的确切关联还没有得到广泛的研究。

其他思想和方向

人们也提出了许多探索量子引力的思想和方向。其中一部分研究方向仍然很活跃，我们在下面谈到的只是少数。

一种得到广泛探索的项目是要按照离散点阵理论连续极限的语言来定义量子引力，这种技巧是在量子色动力学中提出的。过去，这个方向上的很多尝试

都失败了，因为它们所考虑的点阵理论最后被证明是没有连续体极限的。其中有一种版本叫作动力三角剖分，现在还非常活跃，虽然也还没有证据可以证明存在一个连续极限。

拉斐尔·索金(Raphael Sorkin)和他的小组已经探索了一种离散模型，在这种模型中，时空被一组离散的点所取代，这组点有着表示因果关系的序列[Sorkin，1983]。引人注目地，这个模型预言了一个很小但非零的宇宙学常数，它有着具体的数量级，最近这个预言被证实了。

罗杰·彭罗斯(Roger Penrose)和他的小组把扭量理论作为度规几何的再形式化，希望借此解决量子引力的问题[Penrose，1967]。目前，扭量理论的结果更多地是关于数学的而不是关于物理学的。

其他研究方向，包括哈特的时空量子力学[Hartle，1995]，量子雷吉(Regge)计算[Williams and Tuckey(威廉姆斯和塔基)，1992；Williams，1997]，特霍夫特['t Hooft，1996]的决定论方法，以及芬克尔斯坦(Finkelstein)[Finkelstein，1997]的理论。

“现象学”和洛伦兹(Lorentz)不变性

直到几年之前，各研究团体都确信量子引力效应无疑远远超越了我们当前的观测范围。这种信仰最近因为很多迹象发生了动摇。迹象显示，量子引力效应事实上可能处于可观测范围的边缘。有迹象表明特定的数据已经被观测到了，而且似乎很难用传统的物理学来解释，可能是受到了量子引力效应的影响。这些迹象涉及比如高能粒子的宇宙传播、宇宙密度谱中的精细细节等。

问题似乎关乎量子引力是否违背了洛伦兹不变性。小的洛伦兹非不变量子引力效应如果存在，可能会在观测范围之内或者接近观测范围。另一方面，据推测，洛伦兹不变性效应远远小于此，因为洛伦兹不变性不允许某些效应发生。比如，对支配光传播的洛伦兹不变性传播关系的微小偏离，可能在宇宙旅行时间中累积从而产生可观测的频率相关延迟。在洛伦兹不变性情况中，光的旅行时间是一个无意义的观念。

有人会天真地认为量子引力中最小长度的存在必然会打破洛伦兹不变性。其论据是，最小长度在惯性系的变换下必须是洛伦兹收缩的，因此就不可能是最小的。但这个论据是错误的，因为它忽视了量子理论[Rovelli and Speziale

(罗韦利和斯皮兹勒),2003]。在量子理论中,离散量作为可观测量的本征值出现,然而对称变换改变了态,因此这里意味着值而不是本征值。

要说明这个现象,回想一下在经典力学中角动量的 z 分量 L_z 在旋转下连续变换。在量子理论中,令系统处于 L_z 的本征态 $|\psi\rangle = |\hbar/2\rangle$。从一个转动参考系看,这个系统似乎是处于叠加态 $|\psi\rangle = \alpha|\hbar/2\rangle + \beta|-\hbar/2\rangle$,这里 α 和 β 随着转动角连续变化。因此,期望值 $\overline{L_z} = |\alpha|^2\hbar/2 - |\beta|^2\hbar/2$ 在转动中连续变化,但是本征值保持不变。按照物理学术语,我们在所有的参考系中总是观测到离散值 $L_z = \pm\hbar/2$——在转动中连续变换的是看到这一个或者那一个的概率。同样,在圈量子引力中,一个(非零的)最小面积是一个本征值。从一个改进的参考系看,这个区域上本征态中的表面会出现在不同区域本征态的叠加中。这个区域的表面积的期望值可以比最小面积更小,但是(非零的)测量结果就不可以。

2.2 目前主要的尝试性理论

目前两种发展最好、受到研究最多的量子引力理论是弦理论和圈量子引力。

弦理论

弦理论让人感兴趣的原因是它是一种关于世界的基础理论,包含了引力场,可能可以避免紫外发散,在一种自然的和严格统一的结构中囊括了我们可以在世界上发现的所有多种多样的组分。

弦理论的起点是这样一种假设:基本客体不是点状粒子,而是弦,即一维的客体。它最初被当作一种强相互作用的尝试性理论得到研究,后来被证明是错的。量子弦理论只有在时空有着特定维度(叫作临界维度)的时候才是一致的。这个维度对于玻色弦是26维,对于包含了费米子的超对称弦是10维。如何调和临界维度和我们的世界看起来是四维的这个事实还是一个尚未解决的问题。

1984年,格林和施瓦茨(Schwarz)提出了弦理论可能是所有相互作用包括引力作用的统一理论的思想[Green and Schwarz, 1984]。事实上,现在提出的弦中有一种振动模式自旋为2,并且可以等同于引力子。而且,弦理论的一个必要(一般来说非充分)条件是背景时空要满足一个方程,这个方程在大尺度极限的情况下可以约化为爱因斯坦方程。

一致性把弦模型限制在了几个可选范围之内。超对称模型定义在 10 维平直时空上，运用了一个很大的规范群，似乎包含了我们世界的所有组分：规范群包含了粒子物理学标准模型规范群的子群，并且弦的最低能量振动包含了费米子、规范玻色子以及引力子。虽然没有找到完整的证据，但这个理论似乎不会遇到紫外发散。

其思想如下：十维时空中有六维是我们看不见的，因为它们被包（“紧致化”）入了一个很小的空间（或者因为我们被约束而只能生活在一个四维的表面）。在四维可见维度上的有效物理学理论依赖于其余六维空间紧致化的方式。这可以有很多种不同的实现方法，造成了大量有效的四维理论。目前在这大量的可能性中并没有找到一个有效的选择原则。某些作为结果的低能模型与标准模型看上去有着高度的相似之处，但是目前为止都不能精确地给出我们在低能状态下观测到的物理学。

弦理论是根据十维固定时空背景上的微扰展开定义的。20 世纪 90 年代中期，好几种非微扰弦理论得到了探索。更高维度的激发，叫作“膜”［Polchinski（波钦斯基），1995］，可能是除了弦本身之外由于一致性而被需要的（有人提出我们居住于其上的四维表面可能是一个四维膜）。不同的弦模型（以及与十一维超引力）之间似乎都通过一种被称为“对偶”的简单变换相关起来，这表明所有不同的弦模型事实上是同一种未知基本理论的不同极限，姑且称之为“M 理论”。这个假设的基本理论的实际构造——预期是背景无关的——还没有找到，而且目前弦理论的存在形式只是一些在分配的背景空间上扩展的（松散）相关的模型。

1998 年，有人提出，按照反德西特（de Sitter）时空和反德西特球面的成果，某种共形场论包含了与超引力理论相关的领域。这导致这样一种猜测：弦理论在反德西特时空上的紧致化是“对偶于”时空边界上的场理论的。相应地这就为按照边界理论定义 M 理论自身提出了一个新的思路，这种思想是要运用边界理论的背景相关办法使 M 理论实现背景无关性。

理论存在的困难很多。现在还没有一种紧致化的选择机制——这是对“真空”的选择的问题。由于每一种紧致化都能给出不同的物理预言，而存在着成百上千可能的紧致化方法，弦理论实际上是大量不同理论的集合，每一种理论

都能给出不同的预言，都有着不同的物理参数。结果是，理论不能计算标准模型参数的值，而且因为它不能相容于任何未来的实验结果，因此几乎完全是不可预测的。根据某些批评，这种对预言性的缺乏破坏了弦理论作为一个科学理论的特定本质。

即便我们想要选择一种特别的紧致化，现在也不存在一种在低能极限下能够精确地得到标准模型的紧致化方法。理论要求超对称，可观测超对称粒子的存在已经再三被当作理论的与众不同的预言被提出了。虽然有一些初步的说法，称超对称粒子还没有在实验粒子物理学中被发现。类似地，探测不可见维度效应的可能性也被考虑了，但是实验结果也是否定的。理论要求大量新物理学(多余的维度、无限数量的有着任意质量和自旋的场、超对称粒子……)，但是目前在真实世界中这些新物理学都不存在，或者不存在可观测结果。

圈量子引力

圈量子引力最让人感兴趣的原因是，它的物理假定前提只有量子力学和广义相对论，也就是经验成功的理论。该理论是背景无关的，是合并广义相对论时空观念与量子场论的一个很好的尝试。这个理论并不宣称自己是最后的“万物理论”。它不是紫外发散的，不需要对广义相对论做出高能修正，不需要超对称、多余维度或者其他不可观测的物理学。

圈量子引力是 1988 年提出的，它是下面两种研究方法融合的结果，最后被证明可以解决每一种方法的困难[Rovelli and Smolin，1988；1990]。

第一种方法是惠勒—德威特理论。与在惠勒—德威特方法中一样，圈量子引力是广义相对论的直接量子化，保持了其传统的物质耦合，并且其基础除了广义相对论和量子力学再无其他物理学假设。延续了量子力学的基本规则，圈量子引力的量子态是从广义相对论场变量的代数表示得到的，它们的物理学解释是从对表示物理量的自轭算符的对角化得来的。它与旧理论的不同之处在于选用了圈变量作为量子化的基本变量。多亏了这个选择，不适当的惠勒—德威特理论变成了良定的形式，有限的物理量可以得到计算。

第二种方法是这样一种思想：规范理论可以自然地按照类圈激发来描述。这种思想可以追溯到场论的最开始——法拉第的一种直觉。法拉第用称为“法拉第线”的线来描述充满空间的电和磁现象。在电荷存在的情况下，法拉第线

可以始于电荷终于电荷，在电荷不存在的情况下，它们闭合起来形成“圈”。麦克斯韦把法拉第的直觉引入数学物理学，介绍了电场和磁场，它们是在任何地方都与法拉第线呈正切的矢量场。这样就开启了一条通往现代物理学的道路，且完全基于场的观念。很多科学家认为规范场在圈的语言下可以得到更好的理解，包括波利考夫(Polyakov)、曼德尔施塔姆(Mandelstam)、威尔逊(Wilson)等。单个法拉第线的量子激发被称作“圈态”。

在点阵近似的背景下，量子场论的公式按照圈态的语言描述是可行的而且得到了很好的理解，但它要在连续的时空背景上定义就会面临困难。然而，这些困难在背景无关的情况中消失了。因为在背景存在的情况下，圈态局限于背景时空：对于空间中圈的每一个位置，都存在着不同的态。相反在引力情形中，不存在背景时空。圈态自身是空间的量子激发。因此圈态并不拘泥于空间之中，它们自己“织出”了物理时空，就像一堆线可以织出一件T恤衫一样。

更加确切地说，量子引力的圈态有着自相交的点，叫作“节点”。节点表示空间或者空间的单个原子的基本量子激发。通过一个圈直接连接的两个节点表示空间的毗邻原子。节点和连接节点的联结形成了一个图，并且具有量子数。这些量子数决定着空间原子量子化的体积以及分隔相邻节点的基本表面的量子化面积。一个有着这些量子数的图叫作“自旋网络”，因为联结上的量子数被证明是半整数，或者自旋。

自旋网络态没有位置；有意义的只是定义图的组合关系，而不是它在空间中的形状和位置。事实上，自旋网络态并不存在于空间中，它就是空间。因此，尽管圈量子引力的基本假定前提是保守的(量子力学和广义相对论)，它还是导向了一个非常新颖的空间图景。

根据标准量子力学的规则，物理区域的体积或者一个物理表面的面积的可能取值是由对应算符的谱决定的。这在给出了空间普朗克尺度的粒状结构时被证明是离散的。这些谱已经被计算过了，并且表示了圈量子引力的定量物理预言：人们预言任何面积或体积的普朗克尺度上的精确测量都只是作为这些谱中的值的结果给出的。比如，面积的谱由下式给出[Rovelli and Smolin，1995]：

$$A = 8\pi\gamma\hbar G\sum_i \sqrt{j_i(j_i+1)} \tag{9}$$

其中 j_i 是半整数的 n 元组（穿过对应于测量面积的面的自旋网络态的联结量子数）。γ 是 2.1 节中提到的伊米尔齐（Immirzi）参数。

动力学是由自旋网络态空间上的惠勒—德威特方程决定的。它的紫外限制是空间粒状结构的结果。这个方程有着不同限制和明确定义的版本。如果其中有一个在物理上正确的话，目前还不清楚是哪一个。

这个理论的应用包括 2.1 节提到的贝肯斯坦黑洞熵的一个推导，比如黑洞中心的经典奇点在描述上的应用以及在宇宙学上的应用。这个理论似乎可以支配黑洞奇点和最初的大爆炸奇点。支持该理论预言的间接经验证据在天文学和宇宙学领域也得到了积极探索。

圈量子引力的主要困难在于重获低能现象。由于闵可夫斯基真空及其激发的量子态还没有构造出来，因而粒子散射幅度还没有计算出来。这些不足削弱了有限性要求的力量，也影响了引力量子理论的关键要求之一：对低能物理学的再现。人们对动力学仍然知之甚少，惠勒—德威特方程不止有一种版本存在。有些人质疑其缺乏统一的时间演化、理论结果概念彻底的新颖性、背景时空被完全抛弃等问题。

圈—弦之争

一个理论只有在其最初的预言具有合理的独特性并且得到新的实验证实的时候，才具有可信性。圈量子引力和弦理论——或者其他任何目前的量子引力试探性理论——在这个意义上都不具有可信性。而且，虽然做了很多努力，这两个理论现在都还非常不完整，远远没有被人们清晰地理解。因此量子引力的问题必须被看作是完全未解决的。

然而，这两种方向在近些年都得到了很大的发展，十年前看上去太难的问题很多现在都解决了，现在面临的是量子引力的不完备但是可能的解的困惑。

但是，这两个理论在它们的假设、成就、具体结果以及概念框架上都有着显著的不同。它们提出的问题是关于世界物理图景的基础的，它们之间的争论包括了概念、方法论和哲学的问题。

弦理论的教义看起来是为了消除广义相对论的微扰量子化的困难，我们必须把引力场和质量耦合起来。通过把量子场论的点状费曼顶点替换为延展的弦之间的非点状相互作用可以获得有限性。理论保留了量子场论的基本概念结构

(如背景时空、统一性、根据渐近S矩阵给出的预言……)，其代价是宣布放弃对描述广义相对论的广义协变性的彻底执行，放弃大量额外的包袱(额外的维度、超对称、无限场……)以及在预言性上的极大降低。

另一方面，圈量子引力植根于描述广义相对论的广义协变性。紫外有限性是空间粒状结构的结果，而空间的这种结构是当我们把广义相对论看作时空自身的理论而不是背景时空的小微扰的理论时一个标准的量子力学效应。因此，圈量子理论使人感兴趣的是，它致力于合并量子场论和我们在广义相对论中发现的世界观。而且，它导向了原则上可证伪的明确的物理学预言。然而，即便是忽略理论的不完备性，这个结果的代价也是很大的。理论放弃了幺正性、时间演化、基本层次上的庞加莱(Poincaré)不变性以及物理学对象是在空间中局域化的且在时空中演化的观念。

这些激进的概念跳跃是否可行？如果可行，它们是否合理？这是现在争论的热门话题。

3. 方法论问题

3.1 量子引力研究的理由

经验数据的缺乏

量子引力研究的第一个显而易见的问题是研究是否完全合理，因为量子引力相关领域彻底缺乏经验数据。我们在量子引力的研究中没有任何经验指导——比如说像原子光谱引导了量子力学的发现那样。

有些评论家认为量子引力的研究是无效的，因为在量子引力领域这一远离我们经验的尺度下可能发生任何事情。也许这个研究是不可能的，因为可能的理论空间太大了。

目前，可能还没有理由这样担忧。如果这是问题的话，我们会有大量完整的、可预言并且具体的量子引力理论，那样问题就会变成如何在它们中做出选择。但现在的形势是相反的，我们还没有很多理论。实际情况是我们确实有大量关于量子引力的信息，因为我们有量子力学和广义相对论。与量子力学和广义相对论相一致，加上内在的一致性，就形成了一个极其严格的约束集。目前

争论的问题是至少应该发现一种完整和一致的量子引力理论。如果能够发现更多，我们当然不得不诉诸于实验来选择一个物理上正确的理论。

引力是否应当量子化?

常有人提出量子引力效应不存在，引力从本质上讲可能是经典(非量子)的。其理由是，引力与其他相互作用力有着显著的不同，因为它可以用时空几何来描述。这种意见也产生了一个研究计划，其目标是探测经典引力和量子场论在其中发生作用的理论的一致性。

这种意见因其简单形式目前在很大程度上已经被抛弃了。原因是，正如在量子引力早期人们注意到的，经典变量和量子变量之间的相互作用总是不一致的。如果一个动力学变量违背了海森堡不确定性关系，那么其他所有变量也就都违背了。然而回避引力量子化的思想已经以各种形式重新出现了。

有一种意见认为引力场可能并不表示真正的微观自由度，而只是关于它们的集合或者“流体动力学”，是对这些的大尺度描述。这个假设受到诸如由盎鲁效应揭示的引力和热力学之间关系的现象的支持。雅各布森[Jacobson，1995]甚至从式(7)和标准热力学的关系导出了爱因斯坦方程，提供了支持这种思想的证据。但是即便引力场只是一个集体变量，也并不意味着它不会显示量子效应。量子力学并不仅仅支配基本自由度，它也支配所有的自由度，包含那些共同的自由度。因此，这种可能性即便是可以被实现的，也不能驳倒对引力量子理论的需求。

另外一种意见认为引力可能是其他量子场中包含的突发现象。其理由基于这样一个事实：量子场论在弯曲时空上的重整化过程产生了作用中的项，与黎曼曲面上的多项式成正比，而其最低阶项正是引力相互作用。这种意见的困难在于度规场的动力学地位是模棱两可的。变分原理指出，动力学的确定只取决于有关动力学变量的作用的变化，而不取决于其中出现的任何事情的变化。如果度规场被假定为动力学变量，那么它就与其他场一样是个动力学场，那么它的行为的非独立性就由于它与其他场的相互作用的重整化而被修正了，是这一事实改变了它的动力学，而不是说它是一个量子场的事实改变了它的动力学。如果另一方面，度规不是动力学场，那么对于它的作用就不会是变化的(就如在狭义相对论背景中对于它的作用是不变的一样)，因此爱因斯坦方程并不是

由作用中的新项产生的。在第一种情形中引力场需要被量子化，而第二种情形与经典广义相对论所满足的经验事实是矛盾的。

3.2 研究态度

关于量子引力理论研究的方法论，在物理学家阵营中存在着不同的态度。

(a)“悲观”态度，上面已经提到过有太多种可能的担忧，任何事情都有可能发生于这里和普朗克尺度之间，因此引力量子理论的研究是无意义的。

如前所述，这种担忧是没有根据的，因为我们确定没有太多完整的量子引力理论，甚至连一个都没有。

(b)经常有人说量子引力需要一些全新的、激进的、甚至是疯狂的假设。这种“狂热的”态度是基于这样一种发现：伟大的科学家有勇气打破人们已经认可的旧假设，去探索一些新的“奇怪”假设。基于此，“狂热的”科学断定任何奇怪的假设都值得探索，即便它违背了已经确立的事实。

站在历史的角度，这种想法可能缺乏牢靠的根据。凭空而来的狂热思想很少能使科学进步。物理学之所以已经成功采用激进假设一直是因为这些假设被新的经验数据——开普勒(Kepler)的椭圆、玻尔(Bohr)的量子化、普朗克的能量子——或者被严格的理论推演——麦克斯韦的位移电流、爱因斯坦的相对论——所推动从而得不采用它们。一般来讲，随意的新颖假设都不会有什么结果。这种考虑导致了下一种态度(c)。

(c)量子引力的部分研究是由这样一种希望支撑的：广义相对论和量子引力中关于世界的知识可能会引导我们发现一个理论，这个理论能够描述我们没有探索过的物理学领域。

这个希望的一种动力是，现在我们都很精确地处于一种常见的形势之中，理论物理学在其中已经做到前所未有的好。理论物理学中很多最显著的进步都源于要致力于为两种基本却显然矛盾的发现找到一种共同的理论框架。比如，想要合并狭义相对论和非相对论量子力学的目标导致了反粒子的发现；想要合并狭义相对论和牛顿引力的目标导致了广义相对论；想要合并开普勒轨道和伽利略物理学的目标导致了牛顿力学；想要合并麦克斯韦理论和伽利略相对性的目标导致了狭义相对论，等等。在所有这些例子中，主要的进步都是通过“认

真对待”显然矛盾的理论并且探索坚持两种理论的基本原则都为真的意义而得到的。今天我们就处于这样一种典型的状态之中。我们已经知道了关于自然的两种新的非常普遍的“真相”，分别由量子力学和广义相对论表达，我们“只是”要指出它们放在一起时的含义如何。

(d)关于这个问题的一种不同看法是由那些接受量子力学是一种概念革命但却不这样看待广义相对论的人所持有的。根据这种看法，广义相对论的发现“只是”写出了另外一种经典场论而已。这种场理论可能只是我们不了解的某种理论的近似，而且其含义不能被过高估计。根据这种意见，广义相对论不能被过于严肃地看作是理论发展的一种导向。

对这种观点的一种可能的反对意见是，它混淆了以下两点：(i)广义相对论的作用量和广义相对论场方程的具体形式；(ii)广义相对论产生的空间和时间观念的修正。广义相对论的作用量可能是其他什么东西的低能近似。但空间和时间观念的修正与理论的微分同胚不变性和背景无关性有关而与其具体形式无关。量子引力的挑战是把这些新奇性与量子场论合并，而不是与广义相对论作用的具体形式合并。

(e)一种共有的态度是物理学家的“实用”态度，倾向于忽略或者推迟这些基本问题而去发展和调整当前的理论。这种研究风格在20世纪60年代粒子物理学标准模型的研究中行之有效，当时对于存在的量子场论的调整的长期过程得出了一个非常有效的理论。

这种态度在目前基础混淆的情况下是否有效还是有疑问的。在60年代，经验数据日日变动以便随时追踪研究，而现在已经得不到新的数据了。这种“实用”态度可能误导研究，在极端的情况下，“实用的”物理学家仅仅关注眼前理论的发展，而不关心理论所预言的世界与我们看到的世界相似度越来越小。有时候他们甚至因为理论与世界看起来如此不同而感到兴奋，认为这就是他们在知识上远远领先的证据。但是更可能的是，理论和世界之间的不同正是他们有多离谱的唯一证据。而不幸的是，今天的理论研究也面临类似的情况。

科学知识的累积性和科学哲学的影响

上述“悲观的”“狂热的”和“实用的”态度可能受到一种不强调科学知识的累积性而去强调理论与取代它的新理论之间的“不可通约性”的科学哲学的影

响。科学哲学中这种长期争论多少影响了许多理论家的研究态度。

另一方面，上面提到的态度(c)是基于这样一种期望：量子力学和广义相对论的核心原则代表了我们通往量子引力这一未知领域的最好引导。在无论如何理论结果都预期会有激变演化（比如空间本质的变化）这种困难的研究形势下，基于知识累积性的信念的保守假设会起到重要作用。

这种对科学知识累积性的信仰的基础是那些“永恒的”发现的思想。例如，地球并非宇宙的中心，同时性是相对的，绝对速度是无意义的，下雨不是我们跳舞祈雨的结果。

一个理论在其被发现时所处的领域之外可能也会有价值，这一事实可能是理论物理学的历史有效性，特别是诸如麦克斯韦无线波、狄拉克的反物质或者广义相对论的黑洞等的重大预言的根本。这可能被理解为只是科学归纳：是自然有着规律性这一事实的结果。这里不是进入这一讨论的地方，但还是可以看到，这些规则的存在受到量子引力领域一些研究者的支持，并且作为他们信心的源泉——当然，虽然并不必然——在我们的经验还没有探索到的量子引力领域，量子力学和广义相对论发现的世界真相会被证实而不是违背。

4. 空间和时间的本质

广义相对论修正了我们理解空间和时间的方式。要把广义相对论和量子力学结合起来需要对这些观念进行进一步修正。但是，清楚地区分量子引力所要求的空间和时间观念的修正和广义相对论自身对空间和时间观念已有的修正是很重要的。这些在下面的4.1节中会简单地总结，然后4.2和4.3节开始讨论量子引力中空间和时间的观念。

4.1 广义相对论的物理意义

广义相对论是这样一种发现：时空和引力场是同一种实体。我们称为“时空”的那种东西自己是一个物理客体，在很多方面类似于电磁场。我们可以说广义相对论是这样一种发现：根本不存在时空。牛顿称为“空间”、闵可夫斯基称为“时空”的那种东西，在一种我们忽略它的动力学的机制下只是一个动力学

客体(引力场)。

在牛顿和狭义相对论物理学中，如果我们移开动力学实体——粒子和场——剩下的是空间和时间。在广义相对论物理学中如果移开动力学场，什么都不会留下。牛顿和闵可夫斯基的空间和时间被重新解释为引力场的结构。这意味着物理学实体——粒子和场——并非沉浸于空间中或运动于时间中。它们并非依靠于时空而存在。可以说，它们彼此相互依赖而存在。

在经典广义相对论中，人们习惯于延续“空间”和“时间”的表达来指示引力场的特点。但是在量子理论中，场可以有着量子化的“粒状”特征并且其动力学是量子化的，因此就只是概率性的。引力场的大多数“空间”和“时间”特征可能会丢失。

这种熟知的时空“舞台”的缺失被称为经典理论的背景无关性。技术上讲，背景无关性由主动微分同胚下广义相对论行为的规范不变性实现。微分同胚是一种变换，在四维坐标流形上平滑地拖曳所有的动力学场和粒子。转而，微分同胚(或微分同胚不变性)下的规范不变性是作用的两种特性结合的结果：第一种特性是它在任意坐标变换(或者广义协变)下的不变性，第二种特性是不存在非动力学“背景”。因此，背景无关性 = 微分同胚不变性 = (广义协变性 + 非动力学背景场的缺失)。这些观念会在下面更详细地论述。

微分同胚不变性

前广义相对论场理论是用时空流形 M 及其上的一组场 φ_1，…，φ_n 来形式化的。流形 M 是一个(伪)度规空间，点 $P\in M$ 表示空间的物理点。时空点由坐标 $x=(x^1, x^2, x^3, x^0)$ 标记，这一坐标表示的是时间测量工具(时钟)和距离测量工具(“量杆”)的读数。更加确切地说，M 具有了(伪)距离函数 $d(x, y)$，这个函数被解释为两点 x 和 y 之间的四维间隔：负的 $d^2(x, y)$ 给出了在 x 和 y 之间做惯性运动的钟测出的时间；正的 $d^2(x, y)$ 给出以 x 和 y 为两端的量杆的固有长度，这个量杆处于惯性运动状态，根据爱因斯坦同时性的定义，x 和 y 是同时的。类光 $d(x, y)$ 指的是光线在真空中从 x 运行至 y。注意，前广义相对论物理学涉及两种不同类型的测量(之间的关系)：(i)时空测量，测量的是时空可观测量，通过时钟和距离测量工具等进行；(ii)场测量，测量的是场可观测量，也就是场 φ_1，…，φ_n 的值(或函数)。

相同的解释框架也用在狭义相对性量子场理论中。唯一的区别是场可观测量可以被量子化。激发量子的数量有着粒子解释。比如在特别高能量的散射实验中，场可观测量(ii)是由粒子探测器(一种场测量工具)显示的粒子数，而时空可观测量(i)是粒子的动量，由测量探测器的时空位置决定。

理论预言的不只是场可观测量 φ 或者时空可观测量 x 的值，而是两者的结合，比如场 φ 在某个时空位置 x 的值 $\varphi(x)$。时空和场可观测量都是有着直接的操作解释的量，它们可以被称为“部分可观测量”。另一方面，可被理论预言的量，比如 $\varphi(x)$，对于一个给定的位置 x，可以被叫作“完全可观测量”。

广义相对性场理论的解释与此不同。在这样一种理论中，存在场 g 表示引力场，可能还有其他场表示其他的动力学变量。这些场定义在微分流形 M 上，通过坐标 x 坐标化。因此广义相对性场理论的形式结构类似于前广义相对性场理论的结构，但是两个主要区别导致了不同的解释。首先，场定义于其上的流形 M 不是一个度规场。引力场 g 为 M 配备了度规结构 $d_g(x, y)$。①因此时钟和距离测量工具测量引力场 g 的特性。这样(i)类时空可观测量和(ii)类场可观测量之间的差别就模糊了。这两种部分可观测量之间的差别的模糊性是广义相对论的一个关键概念创新之处。

第二，场方程在所谓的主动微分同胚的场变换下是不变的。一个主动微分同胚 $g \to \tilde{g}$ 决定于一个光滑可逆函数 f: $M \to M$(但不应该与之混淆)。在微分同胚变换下，场 g 和所有其他场都是被 f“拖曳”的 M。比如，变换的场 $\tilde{g}$ 定义了一个新的距离函数 $d_{\tilde{g}}$，它与 g 通过如下

$$d_{\tilde{g}}(f(x), f(y)) = d_g(x, y) \tag{10}$$

定义的函数相关。

简而言之，由场 $\tilde{g}$ 定义的两点 $f(x)$ 和 $f(y)$ 之间的距离与由 g 定义的两点 x 和 y 之间的距离是相同的。

场方程不变性具有这样的重要性，其原因如下。一个主动微分同胚可能会

---

① 曲线 $\gamma^\mu(s)$ 在 M 上的长度是 $d_g(\gamma) = \int ds \sqrt{g_{\mu\nu}(\gamma(s)) \frac{d\gamma^\mu(s)}{ds} \frac{d\gamma^\nu(s)}{ds}}$，$d_g(x, y)$ 是 $d_g(\gamma)$ 在连接 x 和 y 的曲线上的一个局域极值。

修正未来某个时间面t_0上广义相对论方程的一个解，但在t_0之前则不作任何修正，因此运动方程的两个不同解在过去可能相等但在未来却不同。这个事实为我们提供了选择：(i)我们把理论解释为非决定论的理论，未来不由过去决定；(ii)我们把主动微分同胚解释为一个“规范变换”，即我们假设理论的完全可观测量只由这种变换下不变的量给定。选项(i)并不切实可行，因为经验表明经典引力物理学是完全决定论的。我们因此被推向选项(ii)，该选项有大量的解释结果。

要理解这些结果，令P为M上的一点，$\varphi(P)$为P点场的任意特性。比如，$\varphi(P)$可能表示电磁场在P点的值，或者P点的时空标量曲率，或者诸如此类的东西。这些特性在主动微分同胚下都不是不变的。因此从上面的论证得出，这些特性都不能从理论中得到预言。所以理论并不决定时空点$P \in M$处的物理学。

乍看这个结论可能显得令人困惑，如果物理学并不预言时空点上发生了什么，那么它能预言什么？事实是，在历史上爱因斯坦自己最初也被这种发现搞得很困惑和泄气，甚至到了从他之前预想的广义相对论会具有微分同胚不变性的思想退缩的地步[Norton(诺顿)，1984]。上述爱因斯坦的论证叫作“洞论证”(因为爱因斯坦只在时空的一个有限区域考虑微分同胚的影响，这个区域不存在物质，或称之为“洞”)，这个论证出现在[Einstein and Grossmann(爱因斯坦和格罗斯曼)，1914]中。关于这个论证及其引发的讨论，参考[Earman(厄尔曼)，1987；1989；Belot(贝洛特)，1998；Earman，2001；Pauri and Vallisneri(包里和瓦利斯纳里)，2002]及相关文献。但后来爱因斯坦改变了他的想法，同时接受了微分同胚不变性和结论(ii)，意识到这个结论完全地印证了他对广义相对性概念革命最核心物理意义的直觉。

走出困境的方式是要理解在广义相对性背景中，流形的点并不表示独立于场的物理实体。询问P点场的特性是什么没有意义。时空位置只能决定于场自身或者我们考虑的任何其他动力学客体。比如，如果我们考虑的理论包含了两个粒子，这两个粒子的轨迹碰巧有一次相交，那么粒子的相遇确定了一个时空点。理论能够预言粒子相遇所确定的时空点上场的值和任何其他物理特性。然而，这个点并不能被简单地确定为是M的点，因为同一个物理态可以用通过主

动微分同胚得到的一组动力学变量表示，在此粒子相遇于不同的点，比如M上的Q。在粒子相遇的点上，场值在这样一种变换下是不变的，因为粒子的轨迹和场同时是拖曳的M。爱因斯坦用理论本身的动力学客体(场和粒子)的语言把这种确定位置的方式叫作“时空巧合”。

因此，广义相对性的理论不处理给定时空点上的动力学量的值，它处理的是其他由动力学量确定其处于“什么位置”和“什么时间”的动力学量的值。严格说来，前广义相对论背景物理学就处理了由其他物理量决定其在“何处”和“何时”的动力学量的值，因为用于确定位置的时间和距离是物理量。但是在前广义相对论背景中，我们可以在以下两者之间做出严格的划分：(i)“时空”，被看作是背景实体，由人们看作是非动力学的时钟和量杆测量；(ii)动力学变量。另一方面，在广义相对论背景中，这种划分消失了，时间和距离测量被重新解释为对引力场的测量，与其他场测量的地位相同，并且在非动力学背景和动力学物理变量之间没有任何区别。

坐标的物理意义

上述内容的一个结果是，广义相对论背景中坐标x的物理解释与它们在非广义相对论背景中的解释是不同的。在广义相对论背景中，坐标x完全没有解释：广义相对论中的可观测量对应于独立于坐标x的理论的量。回想一下，非广义相对论坐标表示的是时钟和量杆的读数：在广义相对论背景中，时钟和量杆的读数是由引力场的非局域函数$d_g(x, y)$表示的。而非广义相对论坐标x和广义相对论坐标x以同一种方式得到指示的事实仅仅是一个不幸的历史偶然事件。

要说明可观测量在哪一种意义上独立于坐标x，考虑一个典型的广义相对论测量。广义相对论标准应用的一个例子是太阳系动力学的精确测量和精确模型。在这个情况中，部分可观测量是地球和其他不同行星之间的“瞬时”距离d_p，被定义为雷达信号从地球到行星p再返回地球经过的固有时(由地球上的静止时钟测量)。固定地球上一个任意的初始事件，一个额外的部分可观测量可以作为从此事件沿地球轨迹的固有时τ被人们获得。这样完全可观测量就是行星在不同局域固有时τ的距离的值$d_p(\tau)$。在太阳系的一个广义相对论模型中适当地选择初始数据，可以对有的p和τ预言$d_p(\tau)$，这些预言可以与经验

进行对比。在建立这个模型的时候，我们选择太阳系区域中的任意坐标x，在这个坐标系中表示引力场、电磁场和行星的位置。预言的量 $d_p(\tau)$ 是场的复杂非局域函数和行星的位置，它独立于选择的坐标 x。不可否认，这里引入可观测量 τ 只是为了方便，我们可以等同地把理论的预言只表示为部分可观测量 d_p 之间必然包含的一组关系 $f(d_p)=0$。

广义协变性和克雷茨曼(Kretschmann)异议

广义相对论在主动微分同胚下的不变性源于广义相对论场方程的两个特征。第一，它们是广义协变的。即它们在 M 上的任何坐标变换 $x\to x'(x)$ 下保持相同的形式。这意味着在 M 上不存在优越坐标系。第二，在场方程中不存在静止的非动力学场。

第一个特征，即广义协变性，是爱因斯坦最坚持的特征，该特征引导他发现了广义相对论。对广义协变性的要求在广义相对论革命后选择与我们对世界的理解相容的物理学理论时仍然起到主要作用。

然而，单独的广义协变性并不是很有意义的。事实上任何场方程都能在任意坐标系中写出。这个事实在爱因斯坦写出广义相对论方程不久后就为克雷茨曼[1917]所指出，并且引起了很多讨论。举个例子，考虑场方程：

$$(\partial_T^2-\partial_X^2-\partial_Y^2-\partial_Z^2)\varphi(X,\ Y,\ Z,\ T)=0 \tag{11}$$

如果引入任意坐标 x(有着组分 x^μ)作为函数 $x^\mu=x^\mu(X,\ Y,\ Z,\ T)$(反函数为 $X(x)$，$Y(x)$，$Z(x)$，$T(x)$)，那么波函数(11)变成广义协变方程：

$$\Box_g\varphi(x)\equiv\partial_\mu\ \sqrt{\det-g(x)}g^{\mu\nu}(x)\partial_\nu\varphi(x)=0 \tag{12}$$

在这个方程中未知的是 $\varphi(x)$，而 $g^{\mu\nu}(x)$ 和 $\det g(x)$ 是场方程

$$g_{\mu\nu}(x)=\frac{\partial X(x)}{\partial x^\mu}\frac{\partial X(x)}{\partial x^\nu}+\frac{\partial Y(x)}{\partial x^\mu}\frac{\partial Y(x)}{\partial x^\nu}+\frac{\partial Z(x)}{\partial x^\mu}\frac{\partial Z(x)}{\partial x^\nu}-\frac{\partial T(x)}{\partial x^\mu}\frac{\partial T(x)}{\partial x^\nu} \tag{13}$$

的反函数和决定因素。方程(12)定义的标量场 φ 的场论并非微分同胚不变的，因为它们取决于不同的函数 $g_{\mu\nu}(x)$，在此意义上不同坐标系中未知的 φ 的场方程是不同的。

另一方面，方程(12)可以变为微分同胚不变性理论的方程之一，在这个理论中，$g_{\mu\nu}(x)$也是未知的量之一，即一个动力学场。因此，理论是否是微分同

胚不变的，并不仅仅取决于方程的形式，还取决于动力学量及其运动方程的完整说明。

但是，爱因斯坦对广义相对论的独自坚持，可能不应该被解释为缺乏明晰性，而只应该被解释为他在强调写出一个成功的相对性场论的一个必要而不充分的步骤的重要性方面所做的努力。

另外，就像物理学理论的所有形式特征，即便是完全微分同胚不变性，可能也不应该被解释为一个刚性的选择原则，只通过自己就能选择物理学理论。运用充分的技巧，任何理论都可能以微分同胚不变性的语言重新表述。这也适用于任何其他形式的不变特征。比如，任何理论都可以被重写为转动不变的理论，或者具有任何其他想要的不变性特征，只需要增加变量就可以了。

举个例子，方程(11)可以被看作是物理上等价于微分同胚不变性系统：

$$\Box_g\varphi(x)=0,\ Riem[g]=0 \tag{14}$$

这里 $Riem[g]$是黎曼曲率。现在未知的就是两个场 φ 和 g 了，但是这也要付出代价。首先，这个理论有一个“伪”动力学场，因为 g 通过系统的第二个方程被约束为取决于规范的单一解。g 没有物理自由度，在物理上是一个固定的背景场，虽然我们用了一个小技巧把它称作一个变量，然后把这个变量约束为一个单一解。第二，我们可以坚持理论(14)的拉格朗日形式[Sorkin, 2002]，但是要做到这一点我们必须引入一个额外的场，这样就可以论证说作为结果的有着额外场的理论与式(12)是不同的[Earman, 1989]。

微分同胚不变性是用来表述从广义相对论的关键概念转换的数学语言的关键特征：世界并不是由一个静止的非动力学时空结构形成的，这个结构定义了局域化且动力学场位于其上。世界只是由与其他场相互作用的动力学场形成的；局域化只是对于场自身而相关地定义的。

关系论和实体论

牛顿用了一个非动力学的背景空间。在《原理》(*Principia*)的第一部分牛顿就非常明确地主张我们必须假设空间是一种实体的存在性。这个部分可以被解读为对之前长期占支配地位的笛卡尔空间关系论理解的反驳。

关于空间的两种传统观点，绝对(“空间是实体”)论和关系(“空间是实体之间的关系”)论，被适当地修正以在当代科学哲学的实体论和关系论名义下继

续考虑科学进步。

我们可以说爱因斯坦“揭露了”牛顿所引入的实体(它严重扰乱过莱布尼茨)的“真面目”。牛顿的空间只是一个与其他场类似的场，想想牛顿是在它的动力学能够被忽略的领域中考虑它的。在空间和时间中的局域化，即被牛顿用来反对笛卡尔的相关局域化，被爱因斯坦揭示出对于特别选择的实体——即引力场——归根结底仍然是笛卡尔意义上一个相对的位置。在某种意义上，我们可以由此认为广义相对论在牛顿实体论的插入之后实现了一种向空间和时间的关系定义的回归。

换言之，在前相对论物理学中，时空是一种有着结构的容器，它是世界的存在之所。而在广义相对论物理学中不存在这类东西。只存在相互作用的场(包括引力场)和粒子，理论中表示的局域化的观念是相对的，动力学客体(场和粒子)只是相对于对方局域化的。这是亚里士多德和笛卡尔所论证的相对空间的观点，牛顿在《原理》的第一部分所写的与它们相反。牛顿的观点中赞同两点：惯性效应的物理实在性，比如他著名的水桶实验中水面的凹度，以及他基于绝对空间的理论的巨大经验成功性。爱因斯坦为凹度的产生原因提供了一种可能的解释：局域引力场——以及一个基于相关联的空间的在经验上比牛顿理论更加有效的理论。爱因斯坦因此重新开启了空间和时间相关联的理解的可能性，这种理解曾经由于牛顿的水桶实验而中断。

笛卡尔关系论的基础是“邻近性”观念。两个客体是邻近的，如果它们与对方毗邻的话。空间是事物关于这样的邻近关系的秩序。广义相对论时空结构的基础是一个非常类似的观念：爱因斯坦的“时空一致性”严格地类似于笛卡尔的“邻近性”。

理解这种新颖的空间和时间相关联的观点的关键是法拉第场是物理实体这一革命性思想。回想一下，法拉第把场设想为一个充满所有物体的真实线簇。爱因斯坦的整个理论工作彻底地执行了这种对场的实在性解释。在爱因斯坦的科普作品中，引力场是一个巨大的“海蜇”，这是一个比法拉第更好的隐喻。不仅仅粒子是实体，场也是，且引力场是很多场中的一种。这些实体只是相对于对方而局域化的。

不过，实体论的立场可以仍然——而且事实上它确实仍然——得到支持。

爱因斯坦关于牛顿时空和引力场是同一实体的发现也可以这样表述：不存在引力场，时空具有动力学属性。这种选择在文献中并不罕见。这里所用语言的不同只是语言的选择问题。这样根据广义相对论，实体论者就可以声称运用“空间是一种实体”：事实上，即引力场是一种实体。由于局域化的定义可能要关乎引力场，实体论者也可以说“时空是一种定义局域化的实体”。

然而，这是一种极度弱化的实体论立场。广义相对论时空在什么程度上与任何其他实体不同，我们可以对于它们哪一个来定义关系的局域化？我们可以用任何用以定义局域化的东西来命名“时空”。牛顿的实体论的深刻形式包含了对“空间”的精确描述[Newton，1962]：

> 因此，如果空间被看作与运动物体真正不同，那么位置的定义和由此给定的局域运动的定义就要参照某种无运动的东西，诸如单独的广延或“空间”，这是很有必要的。

根据这种实体论的宣言，空间的特征是与“运动客体真正不同的”。在现代这个时期以及法拉第和麦克斯韦概念革命之后，我相信这只能被翻译为“真正地与诸如粒子或者场的动力学实体不同”。这并非广义相对论时空的情况。现代实体论者可以放弃牛顿的强实体论(“时空是一个非动力学实体”)，而采用弱得多的论点“我们把时空称为引力场，一个动力学实体”。但是如果不只是语言的选择不同的话，这个立场和关系论立场之间的区分又如何？

不可否认，广义相对论的关系论与传统的关系论也并不能很好地相符。举例来说，广义相对论的可观测量被看成是非实体的符合量，因为它们不需要时空点作为支持。但是它们也不是传统意义上的关系论的，传统的关系论包含了“材料”物体或事件之间在历史上的关系。

传统的实体论—相对论的抉择在法拉第—麦克斯韦概念革命之前就已经形成了，并没有考虑场的存在。在法拉第和麦克斯韦之后，我们也按照一个新的动力学实体，也就是场的语言来理解世界。一旦我们接受了场的存在，并认为爱因斯坦的发现和牛顿的空间是一种场，实体论和关系论之间的距离就大大缩小到了仅仅是字面意义上的不同。

当长期争论中的两方之间变得如此的接近，它们的区别降低到字面意义的时候，就有可能说这个问题已经解决了。在这个意义上，可以认为广义相对论解决了空间的关系论和实体论解释长期以来的问题。

4.2 背景无关性

量子力学与上述的广义相对论空间和时间是否相容呢？是的，但是必须运用量子力学的足够一般的形式。比如薛定谔的绘景只在有着全域可观测时间变量 t 的理论中才可行，这就与广义相对论冲突，因为广义相对论中不存在这样的变量。因此，薛定谔绘景在背景无关的情况中没有什么意义。但是人们认为量子力学的表述比薛定谔绘景更加普遍。比如见[Hartle，1995]和[Rovelli，2004]。此类形式有时候指示“广义化的量子力学”，虽然它在与人们用“经典力学”来命名像牛顿的、拉格朗日的、哈密顿的或者辛量子的力学那样有着不同程度概括性的形式同样的意义上也叫作“量子力学”。

另一方面，大部分微扰量子场论的传统体系与广义相对论框架是极度不相符的。这有很多理由如下。(i)量子场论的传统形式依赖于庞加莱不变性。特别地，依赖于能量观念和产生单一时间演化的非零哈密顿算子的存在。例如，真空是能量最低的态。但是在广义相对论理论中一般不存在全域庞加莱不变性，并不存在能量的一般概念和非零哈密顿算子。(ii)量子场论的根基是粒子的物理概念。量子场论关于弯曲空间[Fulling(弗林)，1989]及量子场论中加速度和温度之间的关系的理论经验表明，在一般的引力情况下，粒子的概念可以相当脆弱。(iii)考虑传统的重整化量子场论。理论的物理内容可以按照其 n 点函数 $W(x_1, \cdots, x_n)$ 的方式表述。在广义相对性理论中，任意坐标变换 $x \to x' = x'(x)$ 下的不变性直接就表明了 n 点函数必须满足：

$$W(x_1, \cdots, x_n) = W(x'(x_1), \cdots, x'(x_n)) \tag{15}$$

由于 n(不同的)点 $(x_1, \cdots, x_n)$ 的任何集合都可以通过一般的坐标变换而变为其他集合，因此 W 是恒定的！这确实依赖于它的论据！很明显，我们正处于一个由传统量子场论引起的非常困难的框架中。

有一个逃避策略可能可以回避这些困难，如在方程(1)中那样把引力场写为两项的和，并假定时空和因果关系由第一项而不是由整个引力场定义。这种

逃避策略恢复了背景时空。不采用这种逃避策略且因此延续了广义相对论的整个对称的量子引力公式被称为背景无关性的。

分裂

不同的研究方向以4.1节中讨论的广义相对论时空概念革命的不同评价为导向。如果认真对待这些概念革命，把它们理解为我们所学到的关于这个世界的一个特征，那么量子引力的问题就变成了理解如何去定义和解释一个背景无关的量子场论的问题。这种观点指向了圈量子引力以及类似研究中的很大部分。这种研究方法受广义相对论研究传统的影响更为强烈并不足为奇。

另一方面，如果广义相对论概念变化被看作是偶然的，那么发展量子引力的动力就更多地来自其他未解决的问题，比如说统一性的问题。这种观点的一种常用论据是，由于量子引力影响微观物理学，我们总可以选择一个尺度，这个尺度足够小可以忽略宏观弯曲效应，同时也足够大可以忽略量子引力效应。在这个尺度上，世界是洛伦兹不变的。因此，在量子引力中，我们总可以假定一个渐近洛伦兹不变区域的存在。这种考虑方法在弦理论阵营中占据主导，更多的是受到粒子物理学传统的影响。传统物理学与庞加莱不变性有着紧密的联系，在整个20世纪通常都忽略了引力。

有时候，文化的分裂非常明显，尽管不断地有人在努力填充其中的鸿沟。论战的双方都觉得对方没有意识到某些基本和本质的东西：量子场论的结构正如它在半个世纪的探索中为人们所理解的那样，站在粒子物理学的一边，随着广义相对论出现的对于空间和时间的新颖物理学理解则站在相对论一边。双方都期望对方的观点能在最后被证明是没有意义的。一方的理由是广义相对论仅仅是更复杂理论的一个低能极限，不能被过分认真地当作是自然的深层结构的表示。另一方的理由是，量子场论是建立在静止的度规时空上的，因此在真正的背景无关情况中是无意义的。

4.3 时间的本质

有很多文献都写过这样一个事实，非微扰量子引力，也就是惠勒—德威特方程(5)并不包含时间变量 t。但是这种“量子引力中的时间问题”的表达是有误导性的，因为它混淆了这个问题中特别针对量子引力的方面和经典广义相对

论中已经表述的方面。事实上经典广义相对论可以完全在哈密顿—雅可比形式中按照方程(4)的语言阐述，在其中也不存在时间变量。

经典广义相对论背景中的时间概念与狭义相对论背景中运用的时间概念非常不同(与前相对论中运用的更不相同)。在前相对论情况中，延续了牛顿的做法，我们假定存在普遍的物理变量 t，可以用时钟来测量。所有的物理现象都可以用含该自变量 t 的演化方程来描述。在前狭义相对性概念中，时间的这种概念被弱化了。时钟并不测量通常的时间变量，而是测量沿惯性轨迹经过的固有时。但是如果我们确定一个洛伦兹框架，我们仍然可以按照自变量 x^0 中的演化方程来描述所有的物理现象，即便这种描述隐藏了系统的协变性。

在广义相对性背景中，我们必须分辨两个常被混淆的问题。首先，我们可以考虑物质与给定的引力场相互作用的动力学或等价地考虑物质与给定的时空几何相互作用的动力学的问题。在这种情况下，固定的引力场仍然决定沿任何(类时)时空轨迹的固有时 τ 的值，这种固有时由沿时空轨迹运动的时钟测量。也就是，给定的引力场决定时间的局域概念。

另一个不同的问题由引力场自身的动力学或者引力和物质之间的相互作用动力学给出。在这种情况下，不存在外在的时间变量来扮演可观测独立演化变量的角色。场方程是用一个演化参数给出的，这个演化参数是时间坐标 x^0，但是这个坐标，正如在上述4.1中解释的，并不对应于任何可观测量。一般地，沿时空轨迹的固有时 τ 也不能被用作自变量，因为 τ 是引力场自身的一个复杂非局域函数。因此，适当地讲，广义相对论并不承认系统按照可观测时间变量演化的这种描述。这在广义相对论的哈密顿—雅可比公式(4)中表现得尤为明显。这并不意味着广义相对论缺乏预言性。简单地说，广义相对论所预言的是部分可观测量之间的关系，这些部分可观测量并不能一般地表述为优越时间自变量上因变量的依赖性。

不可否认，时间变量 t 的本体论地位远远不是牛顿物理学中那样简单。在牛顿物理学中，我们按照 t 中演变的物理变量 $A(t)$，$B(t)$，…来描述世界。人们可能注意到，在某种意义上我们永远不能直接接近 t 而只能接近物理变量 A，B，…，因为用来测量 t 的时钟工具自身是有着可观测时间变量 $C(t)$(比如表针的位置)的物理系统。因此我们实际上观测到的总是可观测变量 $A(C)$，$B(C)$，

$A(B)$，…，而永远不是 t 自身的相对演变。牛顿在《原理》中清晰地表达了这个观点，但也同样观测到了明显的运动 $A(C)$，$B(C)$，$A(B)$，…的直接数学化由于实体化了 t 的存在，并且按照 t 来表述所有的演化而极大地被简化了。这在非相对论和非引力的情况中当然非常有效。但是牛顿的策略在某些领域可能会失效也并非不合逻辑。事实上，它在相对论引力领域中就失效了，因为在这里不存在普遍时间，我们只能描述可观测量的相对依赖性。这就是广义相对论中发生的事情。

在某种意义上，任何部分可观测变量都可以被选择作为广义相对论中的独立量。一般地，任何东西都不能拥有牛顿时间 t 所假定的理想属性，这种理想属性单调地增长且与系统的态无关。例如，在封闭的宇宙学中，宇宙的容积 a 与大爆炸之后的固有时 t_c，沿着星系的世界线，经常被用作自变量。但是如果宇宙开始重新缩小，并且 t_c 只是在宇宙被假定为各向同性的近似中定义时，a 就表现得不好了。(如果从大爆炸而来的两个有着不同固有时的星系相遇，那么 t_c 的值又如何?)

经典广义相对论中时间观念的这种弱化很少得到强调，因为归根结底，在经典物理学中我们可能忽略动力学理论的完整动力学结构，只考虑其运动方程的单一解。如上所述，广义相对论运动方程的一个单一解决定一个时空，这里固有时的观念与每一个类时世界线相关。而在量子情况中，不存在单一的时空，就如不存在量子粒子的运动轨迹一样，时间的确切概念变得模糊起来。

对时间问题的各种态度

在文献中我们可以发现关于时间问题的不同态度。技术性的参考文献(到目前为止并不完全)可见[Isham，1992；Kuchar(库查尔)，1992]。

正如上面已经提到的，量子引力的研究很大一部分都忽略了这个问题，并且坚持闵可夫斯基空间、庞加莱不变性及其相关的时间演化(作为庞加莱群的一个子群)的观念都不应该在建立量子引力的时候被抛弃，尽管这是广义相对论的特征。

其他作者们坚持，即便庞加莱不变性在广义相对论中失效了且时间的概念变得更加复杂了，但世界*存在*于时间之中，且它的描述是对时间中演化的系统的描述的思想，仍然是我们不能放弃的基本观点。

这些作者中有些人提议对广义相对论作一些小的修改，使其能够重新引入一个可观测时间演化的基本观点。有一种可能性是要选择一种优先规范修正，在其中微分同胚不变性部分地被打破，时间坐标被规范地修正到等同于引力场的某些函数。举一个例子就是约克时间，它被定义为一个类空表面的外部曲率。或者修正广义相对论的动力学，以得到一个有着独立时间参数的理论。

其他人则完全接受了广义相对论试图在不使用时间和时间演化的基本观念的情况下概念化世界所带来的挑战，正如下一节将要阐述的。

没有空间和时间的物理学？

力学公式怎样才可以不使用空间和时间作为自变量？下面可以给出一个可说明的例子，可参见[Rovelli，2004]的第4和6章。考虑一个有限时空区域 R，以一个封闭的三维面 Σ 为界。令 (φ, g) 表示 Σ 上的所有场，包括引力场 g 的值，令 (P_φ, P_g) 表示 Σ 之外的场的法向导数。原理上，广义相对论的所有预言都可以被表达为集合 (φ, g) 和 (P_φ, P_g) 可以采用的可能值的约束。类似地，在原理上，量子引力的所有预言都可以按照测量场 (φ, g) 时的概率振幅 $W(\varphi, g)$ 的方式进行表达。这是费曼观测到的粒子的量子动力学包含于其传播者 $W(x', t'; x, t)$ 之中这一现象的广义化。微分同胚不变性意味着 $W(\varphi, g)$ 并不依赖于 Σ 嵌入 M 的方式。换言之，整个量子空间的和时间的依赖性是编码在 $W(\varphi, g)$ 对于引力场 g 的依赖之中的。比如，我们把 Σ 等同于一个散射实验的初始、最终和边界值的面，那么它就只是引力场在决定初始和最终面之间的时间间隔的边界上的值。回顾一下，事实上在广义相对论中，空间距离和时间间隔是引力场的函数。这样 $W(\varphi, g)$ 在原则上就可以被用来确定关于实验的所有可能概率性预言，不需要运用独立的空间或者时间变量。

为了理解量子引力场，可能要抛弃某些对几何的强调。几何学很好地描绘了经典引力场，但并非量子时空。这并非对爱因斯坦遗产的违背，相反地，它是在爱因斯坦思想的精确意义上向“相对论”迈出的一步。因此，量子引力的关键概念困难可能就是要寻找一种方式，在缺少熟悉的空间和时间舞台的基本概念的情况下理解物理世界。现在所需要的可能就是把我们从与把世界想象成为

“栖居的空间”和“时间中演化”的习惯联系起来的偏见中解放出来。①

在逻辑上，是否可能在缺少时间和时间演化的基本概念的情况下理解世界，而且这是否可能与我们关于世界的经验一致，这是一个尚未解决的问题。

幺正性

缺少时间演化的基本观念特别意味着在理论中不存在幺正时间演化。对于幺正性的缺乏，很多来自高能传统的物理学家持非常怀疑的态度。在高能传统中，对幺正性的要求不断地起到重要的历史作用。通常人们认为，如果一个或然性理论不存在幺正时间演化，那么它是不一致的。这并不正确，因为不一致性来源于在标准时间演化存在的情况下幺正性的缺乏，而并非时间演化的缺乏。如果说我们用宇宙的容积 a 把宇宙的演化描述为自变量，那么就没有理由要求宇宙存在的概率对于所有的 a 都统一。事实上存在有限的可能性，宇宙只是达到一个最大值 a，然后就重新缩小了。另一方面，在并不用外在变量 t 表示的演化中量子力学概率解释的一致性还并不明确。

5. 与其他未决问题之间的关系

在物理学的历史上，经常是两个未决问题都找到了共同的解决办法。比如，对光本质的理解问题和统一电和磁的理论问题在麦克斯韦理论中找到了共同的解。但是往往同时解决两个问题的期望也会破灭。比如，在20世纪60年代，人们期待找到一个强相互作用力的理论同时排除重整化理论，但是量子色动理论被证明是第一个理论的很好解决方法却并不涉及第二个问题。量子引力的问题被认为与理论物理学的所有未决问题都相关。

---

① 如果我们持这种极端的立场，一个问题就是要从一个不受时间影响的宏观理论恢复到时间演化的宏观观念以及宏观时间可观测量的明确特征。众所周知，要精确确定一个动力学系统中是什么描述了时间变量，这难得出乎人的意料。另一方面，物理学系统的热力学和统计学行为是强有力的时间性地描述的。因此，人们考虑了假设[Rovelli，1993a；1993b；Connes and Rovelli，1994]“时间流”是世界的一个特征，只出现于热统计力学描述的情况中。换言之，“时间”可能是我们对于世界微观态的忽视而造成的人为结果。

5.1 统一

目前对物理世界的描述由一些场论组成，一方面是广义相对论，另一方面是标准模型，它又是由弱电理论和量子色动力学组成的。另外，费米子在多处多重态出现，还有希格斯标量，理论拥有多于一打的基本常数。随着电和磁理论、电磁学和弱相互作用理论的成功统一，在很长时间内研究的目标都是要通过提供一个由更少基本常数支配的单一一致性理论来降低标准模型的复杂性。

对于这个“统一性”问题和量子引力问题的关系，没有一致的意见。先天地不存在严格的理由说明为什么引力的量子特性只有与其他场论关联起来才可以理解。比如电磁力的量子特性曾经在量子电动力学背景中得到理解，不需要参照其他相互作用力；强相互作用力的特性也是如此。

也有人提出其他的一些论据来支持两个问题必须同时解决的观点。我提三个，第一个是推测性的，我认为是不牢固的。第二个和第三个是技术性的，很有些分量。首先，人们普遍期望现在应该拥有一个最终的“包含一切的理论”。然而在历史上，这种期望通常出现在理论物理学中，并且到目前为止一直是错的。

第二种论据来自早期想要用一个重整化理论来替换广义相对论的历史。超引力表明引力的紫外发散由引力和物质之间的一种适当对偶(在超引力的情形中，是一个费米场)所消除(虽然最后并没有成功)。

第三种支持两种问题之间的关系的论据如下。在标准模型中，决定电磁力、弱相互作用力和强相互作用力强度的耦合常数依赖于所考虑的现象的尺度。在通常的尺度下，它们在大小上显著不同，但是在接近普朗克尺度的时候，它们就会趋于汇合。这就表明，统一性可能发生的尺度应当与量子引力效应变得清晰的尺度一样，标志着两种现象可能会是相关的。

具体地，量子引力在处于紧密统一性的背景中的弦理论中得到实现。而圈量子引力则为量子引力问题提出了一个与统一性无关的解。

5.2 量子力学的解释

虽然量子力学取得巨大的经验上的成功并且现在到处都在应用，但是很多

人认为它并没有被完全理解。在我们用它描述与外部系统(观测者)相互作用的物理系统时，理论的解释相对不会存在争议。但是当我们把观测者的量子特性纳人考虑时就会出现很多困难。物理学阵营中对于这个问题的态度多种多样，从完全否定问题的存在到为解决问题而给出各种各样的修正量子力学的建议都有。但是在过去十年中，越来越多的物理学家认为这个问题是一个真正的未决问题。

已经有很多人提议把量子力学的解释问题与量子引力问题联系起来。人们再次期望可以得到一个终极理论，而这个终极理论必须是完全自洽的。

罗杰·彭罗斯提出了一种具体的机制，通过这种机制量子的线性可能会被引力打破：引力可能会是导致物理上实现了的波函数发生塌缩的物理因素[Penrose，1986]。这种提议原则上是可以被经验地检验的。

在斯莫林和阿德勒(Adler)要从矩阵模型统计动力学的统计行为中得到量子力学的尝试中[Smolin，2002；Adler，2004]，已经提出可能能够从量子场论出现的物理现象的一个共同的更深层次上去追求引力和量子场理论的共同起源。

最后，有人提出，由广义相对论所揭示的时间空间结构的关系方面与由量子力学的“相关”解释作强调的相关方面[Rovelli，1996b]可能是有联系的。前者由系统之间的接近关系所决定，后者由系统之间的相互作用所决定。但是一方面，局域性暗示着相互作用只发生在连续系统之间；另一方面，接近性只是在通过物理相互作用才明显，暗示了两种关系之间的严格联系。然而这些思想还都只是建议而已。

5.3 宇宙学常数

当前对宇宙的描述中一个难以理解的方面是由宇宙学常数给出的，这个常数由爱因斯坦引入，描述的是一个可以在大尺度下修正引力的长距离引力耦合。这个常数在宇宙学、量子场论(在其中量子场理论效应倾向于使它变得不切实际地大)中起着重要的作用，并且近来的宇宙观测似乎表明它的值很小，但并不是像之前所期望的那样为零。

到目前为止，虽然量子引力的研究并没有涉及这个常数的任何明确结果，

但是必须注意到，拉斐尔·索金的量子引力理论在观测之前就预言了这个常数的一个很小值，并对其数量级进行了纠正，如2.1.7中所述。

5.4　量子宇宙学

“量子宇宙学”指的是把宇宙整体作为一个量子系统来研究的学科。关于这个说法存在两个不同的问题。

第一个是我们宇宙的动力学模型化的量子形式，特别是对宇宙是均匀的这一极度简化得来的动力学系统的量子特征的研究。这些经典的模型，比如弗里德曼—罗伯森—沃克(Friedmann-Robertson-Walker)模型，在宇宙学中起着重要的作用并且人们相信它可以给出对我们宇宙的大尺度特征的很好描述。人们对它们的量子化感兴趣有几点理由。第一，它们提供了一个简化的框架，在其中量子引力的很多概念困难都可以得到审查，解也都可以得到检验。第二，它们可以被用来研究什么样的引力量子理论能使我们涉及由量子引力效应支配的接近大爆炸和大爆炸时刻的物理学。

关于这些模型的研究早在60年代就由布赖斯·德威特[DeWitt，1967b；1967c]和查尔斯·米斯纳[Misner，1969]开始进行了，并且在接下来的几十年中得到了巨大发展。当然这些模型的局限性在于它们的基础限制量子引力的自由度不能是无限数，必须是有限的，因此它们错过了量子引力问题整体场论的方面。

弦宇宙学由加布里埃·韦内奇亚诺(Gabriele Veneziano)及其合作者所发展，他们希望找到弦理论的观测结果[Gasperini and Veneziano(加斯佩里尼和韦内奇亚诺)，1993]。圈量子引力在量子宇宙学上的应用(“圈宇宙”)最近得到了一个模型，它是有限的并且在初始奇点运行良好[Bojowald(波究瓦德)，2001；Bojowald and Morales-Tecotl(波究瓦德和莫拉莱斯—特克托尔)，2006]。

第二个问题叫作“量子宇宙学”，是关于描述形成整个宇宙的量子系统的概念问题。因此对它来说，不存在“外部”观测者，也就是说它是观测者处于系统之中的量子力学研究。

这第二个问题粗略地与量子引力问题相关。确实，它不可能对于引力场是“外部的”，但我们不应该混淆时间和空间意义上的“外部”和动力学意义上的

"外部"。一个观测者，在时间空间的意义上对于电磁场也不可能是"外部的"，但我们还是能够考虑电磁场系统，把它看作是一个量子系统，与被看作是经典观测者的外部系统相互作用。对于引力系统也可以来用同样的方法。因此，没有先天的理由阻止我们在量子引力背景中运用量子力学的标准哥本哈根解释(不管它是否令人满意)。换句话说，量子引力的问题和这量子宇宙学的第二个问题并不是必然联系的。

另一方面，由把观测者当作系统的一部分而引起的困难以及量子力学中由于微分同胚不变性引起的困难，特别是外部时间的缺乏，其本质都是类似的，并且都质疑了哥本哈根解释的可行性。吉姆·哈特提出了一种研究计划，希望能够同时解决两种困难，并且为量子力学制定一个不需要外部时间和外部观测者的普遍形式[Hartle，1995]。

6. 结论

量子引力经历了七十年的研究，仍不存在一个一致的、完全成立的理论。没有一个量子引力理论得到了任何直接或者非直接的实验支持。在这七十年的进程中，很多思想都得到了探索，研究方式来来往往，多次有人宣称发现了量子引力的圣杯，但后来的结果是都被否定了。

即便如此，量子引力的研究并没有被看作是无意义的。相反，一种一致的逻辑引导着理论的发展。从20世纪50年代问题的最早构想和研究方向到今天，研究计划的执行曾经很困难，但是却很完美。困难曾经出现，解决方式也被提出，在很多困难之后逐渐开始实现(至少是部分地实现)最初的愿望。

70年代早期曾经有人提出，广义相对论也许可以被看作是一个非不可控发散的理论的低能极限。30年之后，弦理论这样一种理论已为人们所知。1957年查尔斯·米纳斯指出，在正则框架中应当可以计算本征值；1995年，37年之后，本征值在圈量子引力中得到了计算。还有很多要理解的东西，现在的有些发展可能是行不通的。我们还没有走到最后，我们只是在丛林中的半道上。但是看看这个主题的整体发展，很难否认我们确实得到了重要的进展。

进展不可能只是技术性的。量子引力理论的研究重新提出了一些新的问题，

比如什么是空间？什么是时间？“在某处”意味着什么？“运动”的意义是什么？运动应当对于客体还是对于空间来定义？我们能不能不涉及空间或者时空来对物理学进行形式化？还有，什么是因果性？观测者在物理学中的地位如何？

诸如此类的问题在物理学的主要发展进程中起着重要的作用。比如它们对爱因斯坦、海森堡、玻尔以及他们的同事们来说就起着重要的作用。对于笛卡尔、伽利略、牛顿以及他们同时代的人，还有法拉第、麦克斯韦和他们的同事来说也是如此。今天，一些物理学家认为这种提出问题的方式“太过哲学了”。20 世纪后半叶的大多数物理学家确实认为这种性质的问题是无意义的。这种观点对于他们所面对的问题来讲是适当的。如果基础的明晰性来源于议题是在一个给定的概念框架中的问题解决的话，就没有理由对基础感到担心：实用的方法是最有效的。现在，基础物理学面对的困难类型发生了改变。要理解量子时空，物理学必须再次回归到那些基本问题上去。我们必须找到旧的基本问题的新答案。新的答案必须考虑我们在量子力学和广义相对论中所学到的东西。除非这些问题得到认真的重新考虑，否则量子引力的问题不会得到解决。

致谢

感谢克里斯琴·乌斯雷彻(Christian Wüthrich)，杰里米·巴特菲尔德和约翰·厄尔曼对初稿的仔细阅读以及他们非常有价值的建议。

参考文献说明

关于量子引力历史的详细介绍见[Rovelli，2004]中的历史附录。关于早期历史，见[Stachel，1999b；1999a]以及[Gorelik(高里克)，1992]。关于当前量子引力研究的方向，见评论文章[Horowitz(霍罗威茨)，2000；Carlip，2001；Isham，1991；Rovelli，1998a]。它所提出的问题和哲学讨论的观点的有趣概论可见于[Callender and Huggett，2001]一书的诸多参考文献中，同时也见于[Rovelli，1997；2000]。量子思想的一般介绍，见旧的经典评论，它们在思想上和目前各种不同观点上的都很丰富，比如约翰·惠勒于 1967 年的[Wheeler，1968]、史蒂芬·温伯格于 1979 年的[Weinberg，1979]、史蒂芬·霍金于 1979 和 1980 年的[Hawking，1979；1984]、卡雷尔·库查尔于 1980 年的

[Kuchar, 1984]，以及克里斯 · 艾莎姆的权威综合[Isham, 1984a; 1984b; 1997]。关于弦理论，经典的教科书有格林、施瓦茨和威滕(Witten)以及波钦斯基的作品[Green *et al.*, 1987; Polchinski, 1998]。关于圈量子引力，见[Rovelli, 1998b; 2004]。关于弦理论的困难和弦与圈理论结果的对比，见[Rovelli, 2003]。斯莫林的通俗读本[Smolin, 2000]提供了一个理解量子引力的通俗易懂的介绍。

参考文献

[Adler, 2004] S. Adler. *Quantum Theory as an Emergent Phenomenon.* Cambridge University Press, 2004.

[Amati *et al.*, 1989] D. Amati, M. Ciafaloni, and G. Veneziano. Can spacetime be probed below the string size? *Physics Letters B*, 216: 41-47, 1989.

[Ambjorn *et al.*, 1997] J. Ambjorn, M. Carfora, and A. Marzuoli. *The Geometry of Dynamical Triangulations.* Lecture Notes in Physics. Springer-Verlag, Berlin, 1997.

[Arnowitt *et al.*, 1962] R. Arnowitt, S. Deser, C. Misner. The dynamics of general relativity. In L. Witten (ed.), *Gravitation: An Introduction to Current Research*, page 227. Wiley, New York, 1962.

[Ashtekar, 1986] A. Ashtekar. New variables for classical and quantum gravity. *Physical Review Letters*, 57: 2244-2247, 1986.

[Ashtekar, 1987] A. Ashtekar. New Hamiltonian formulation of general relativity. *Physical Review D*, 36: 1587-1602, 1987.

[Ashtekar *et al.*, 1998] A. Ashtekar, J. Baez, A. Corichi, and K. Krasnov. Quantum geometry and black hole entropy. *Physical Review Letters*, 80: 904-907, 1998. gr-qc/9710007.

[Bardeen *et al.*, 1973] J. Bardeen, B. Carter, and S. Hawking. The four laws of black hole mechanics. *Communication in Mathematical Physics*, 31: 161-170, 1973.

[Bekenstein, 1972] J. Bekenstein. Black holes and the second law. *Nuovo Cimento Letters*, 4: 737-740, 1972.

[Bekenstein, 1973] J. Bekenstein. Black holes and entropy. *Physical Review*, D7: 2333-2346, 1973.

[Bekenstein, 1974] J. Bekenstein. Generalized second law for thermodynamics in black hole physics, 1974. *Physical Review*, D9: 3292-3300.

[Belot, 1998] G. Belot. Why general relativity does need an interpretation. *Philosophy of Science*, 63: S80-S88, 1998.

[Bergmann, 1949a] P. Bergmann. Non-linear field theories. *Physical Review*, 75: 680-685, 1949a.

[Bergmann, 1949b] P. Bergmann. Non-linear field theories II: Canonical equations and quantization. *Review of Modern Physics*, 21: 480-487, 1949b.

[Bergmann, 1958] P. Bergmann. Conservation laws in general relativity as the generators of coordinate transformations. *Physical Review*, 112: 287-289, 1958.

[Bergmann, 1961] P. Bergmann. Observables in general relativity. *Review of Modern Physics*, 33: 510-514, 1961.

[Bergmann and Komar, 1980] P. Bergmann and A. Komar. The phase space formulation of general relativity and approaches towards quantization. In A. Held (ed.), *General Relativity and Gravitation: One hundred years after the birth of Einstein*, pages 227-254. Plenum Press, New York, 1980.

[Bojowald, 2001] M. Bojowald. Absence of singularity in loop quantum cosmology. *Physical Review Letters*, 86: 5227-5230, 2001. gr-qc/0102069.

[Bojowald and Morales-Tecotl, 2006] M. Bojowald, H. Morales-Tecotl. Cosmological applications of loop quantum gravity. gr-qc/0306008. to appear in the proceedings of the Fifth Mexican School (DGFM): *The Early Universe and Observational Cosmology.*, 2006.

[Bronstein, 1936] M. Bronstein. Quantentheories schwacher Gravitationsfelder. *Physikalische Zeitschrift der Sowietunion*, 9: 140, 1936.

[Callender and Huggett, 2001] C. Callender and N. Huggett (eds.). *Physics Meets Philosophy at the Planck scale.* Cambridge University Press, Cambridge, 2001.

[Carlip, 2001] S. Carlip. Quantum gravity: A progress report. *Reports on Progress in Physics*, 64: 885, 2001. gr-qc/0108040.

[Chamseddine and Connes, 1996] A. Chamseddine and A. Connes. Universal formula for noncommutative geometry actions: Unification of gravity and the standard model. *Physical Review Letters*, 24: 4868-4871, 1996.

[Chamseddine and Connes, 1997] A. Chamseddine and A. Connes. The spectral action principle. *Communication in Mathematical Physics*, 186: 731-750, 1997.

[Connes, 1994] A. Connes. *Non Commutative Geometry.* Academic Press, New York, 1994.

[Connes and Rovelli, 1994] A. Connes and C. Rovelli. Von Neumann algebra automorphisms and time versus thermodynamics relation in general covariant quantum theories. *Classical and Quantum Gravity*, 11: 2899-2917, 1994.

[Deser and van Nieuwenhuizen, 1974a] S. Deser and P. van Nieuwenhuizen. One loop divergences of the quantized Einstein-Maxwell fields. *Physical Review D*, 10: 401-410, 1974a.

[Deser and van Nieuwenhuizen, 1974b] S. Deser and P. van Nieuwenhuizen. Nonrenormalizability of the quantized Dirac-Einstein system. *Physical Review*, D10: 411-420, 1974b.

[DeWitt, 1964a] B. DeWitt. In Gauthier-Villars (ed.), *Conf'erence internationale sur les th'eories relativistes de la gravitation.* Editions Scientifiques de Pologne, Warsaw, 1964a.

[DeWitt, 1964b] B. DeWitt. Theory of radiative corrections for non-Abelian gauge fields. *Physical Review Letters*, 12: 742-746, 1964b.

[DeWitt, 1965] B. DeWitt. *Dynamical Theory of Groups and Fields*". Wiley, New York, 1965.

[DeWitt, 1967a] B. DeWitt. Quantum theory of gravity. I. The canonical theory. *Physical Review*, 160: 1113-1148, 1967a.

[DeWitt, 1967b] B. DeWitt. Quantum theory of gravity. II. The manifestly covariant theory. *Physical Review*, 162: 1195-1239, 1967b.

[DeWitt, 1967c] B. DeWitt. Quantum theory of gravity. III. Applications of the covariant theory. *Physical Review*, 162: 1239-1256, 1967c.

[Dirac, 1950] P. Dirac. Generalized Hamiltonian dynamics. *Canadian Journal of Physics*, 2: 129-148, 1950.

[Dirac, 1958] P. Dirac. The theory of gravitation in Hamiltonian form. *Proceedings of the Royalyal Society of London*, A246: 333-343, 1958.

[Dirac, 1959] P. Dirac. Fixation of coordinates in the Hamiltonian theory of gravitation. *Physical Review*, 114: 924-930, 1959.

[Dirac, 1964] P. Dirac. *Lectures on Quantum Mechanics.* Belfer Graduate School of Science, Yeshiva University, New York, 1964.

[Doplicher, 1996] S. Doplicher. Quantum spacetime. New problems in the general theory of fields and particles, part II. *Annales de l'Institut Henri Poincare*, 64: 543-553, 1996.

[Doplicher *et al.*, 1994] S. Doplicher, K. Fredenhagen, J. Roberts. Spacetime quantization induced by classical gravity. *Physics Letters B*, 331: 39-44, 1994.

[Doplicher *et al.*, 1995] S. Doplicher, K. Fredenhagen, J. Roberts. The quantum structure of spacetime at the Planck scale and quantum fields. *Communications in Mathematical Physics*, 172: 187-220, 1995.

[Earman,] J. Earman. Private communication.

[Earman, 1987] J Earman, J. Norton. What price spacetime substantivalism? The hole story. *British Journal for the Philosophy of Science*, 38: 515-525, 1987.

[Earman, 1989] J. Earman. *World Enough and Space-time: Absolute versus Relational Theories of Spacetime.* MIT Press, Cambridge, 1989.

[Earman, 2001] J. Earman and G. Belot. Pre-Socratic quantum gravity. In C. Callander (ed.), *Physics meets philosphy at the Planck scale.* Cambridge University Press, Cambridge, 2001.

[Einstein, 1916] A. Einstein. Naeherungsweise integration der Feldgleichungen der gravitation. *Preussische Akademie der Wissenschaften (Berlin) Sitzungsberichte*, pages 688-696, 1916.

[Einstein and Grossmann, 1914] A. Einstein and M. Grossmann. Entwurf einer verallgemeinerten Relativitätstheorie und einer Theorie der Gravitation. *Zeitschrift für Mathematik und Physik*, 62: 225-261, 1914.

[Faddeev and Popov, 1967] L. Faddeev and V. Popov. Feynman diagrams for the Yang-Mills field. *Physics Letter B*, 25: 29-30, 1967.

[Feynman, 1963] R. Feynman. Quantum theory of gravitation. *Acta Physical Polonica*, 24: 697-722, 1963.

[Fierz, 1939] M. Fierz. Ber die relativistische Theorie kräfterfreier Teilchen mit beliebigem Spin. *Helvetica Physica Acta*, 12: 3-37, 1939.

[Finkelstein, 1997] D. Finkelstein. *Quantum Relativity.* Springer, Berlin, 1997.

[Freedman *et al.*, 1976] D. Freedman, S. Ferrara, and P. van Nieuwenhuizen. Progress toward a theory of supergravity. *Physical Review*, D13: 3214-3218, 1976.

[Fulling, 1989] S. Fulling. *Aspects of Quantum Field Theory in Curved Spacetime.* Cambridge University Press, Cambridge, 1989.

[Gasperini and Veneziano, 1993] M. Gasperini and G. Veneziano. Pre-bigbang in string cosmology. *Astroparticle Physics*, 1: 317-339, 1993.

[Gorelik, 1992] G. Gorelik. First steps of quantum gravity and the Planck values. In J. Eisenstaedt and A. Kox (eds.), *Studies in the history of general relativity*, volume 3 of *Einstein Studies*, pages 364-379. Birkhaeuser, Boston, 1992.

[Gorelik and Frenkel, 1994] G. Gorelik and V. Frenkel. *Matvei Petrovic Bronstein and the Soviet Theoretical Physics in the Thirties.* Birkhauser Verlag, Boston, 1994.

[Goroff and Sagnotti, 1985] M. H. Goroff, A. Sagnotti. Quantum gravity at two loops. *Physics Letter B*, 160: 81-86, 1985.

[Goroff and Sagnotti, 1986] M. H. Goroff, A. Sagnotti. The ultraviolet behavour of Einstein gravity. *Nuclear Physics B*, 266: 709-736, 1986.

[Green and Schwarz, 1984] M. Green and J. Schwarz. Anomaly cancellation in supersymmetric d = 10 gauge theory requires SO(32). *Physics Letters B*, 149: 117-122, 1984.

[Green *et al.*, 1987] M. Green, J. Schwarz, and E. Witten. *Superstring Theory.* Cambridge University Press, Cambridge, 1987.

[Gupta, 1952] S. Gupta. Quantization of Einstein's gravitational field: General treatment. *Proceedings Physical Society A*, 65: 608-619, 1952.

[Hall and Hall, 1962] A. Hall and M. Hall (eds.). *Unpublished papers of Isaac Newton.* Cambridge University Press, Cambridge, 1962.

[Hartle, 1995] J. Hartle. Spacetime quantum mechanics and the quantum mechanics of spacetime. In B Julia, J. Z. -J. (eds.), *Proceedings on the* 1992 *Les Houches School*, *Gravitation and Quantisation*, page 285. Elsevier Science, Paris, 1995.

[Hartle and Hawking, 1983] J. Hartle and S. Hawking. Wave function of the universe. *Physical Review*, D28: 2960-2975, 1983.

[Hawking, 1974] S. Hawking. Black hole explosions? *Nature*, 248: 30-31, 1974.

[Hawking, 1975] S. Hawking. Particle creation by black holes. *Communication in Mathematical Physics*, 43: 199-220, 1975.

[Hawking, 1979] S. Hawking. The path-integral approach to quantum gravity. In S. Hawking and W. Israel (eds.), *General Relativity: An Einstein Centenary Survey.* Cambridge University Press, Cambridge, 1979.

[Hawking, 1984] S. Hawking. Quantum cosmology. In B. DeWitt and R. Stora (eds.), *Relativity, Groups and Topology, Les Houches Session XL.* North Holland, Amsterdam, 1984.

[Heisenberg, 1938] W. Heisenberg. Die Grenzen der Anwendbarkeit der bisherigen Quantentheorie. *Zeitschrift für Physik*, 110: 251-266, 1938.

[Horowitz, 2000] G. Horowitz. Quantum gravity at the turn of the millenium. Plenary talk at the Marcell Grossmann conference, Rome, 2000. gr-qc/0011089.

[Howard and Stachel, 1989] D. Howard and J. Stachel (eds.). *Einstein and the History of General Relativity*, volume I of *Einstein Studies.* Birkhauser, Boston, 1989.

[Isham, 1984a] C. Isham. In B. DeWitt and R. Stora (eds.), *Relativity Groups and Topology. Les Houches* 1983. North Holland, Amsterdam, 1984a.

[Isham, 1984b] C. Isham. Quantum gravity: an overview. In *Oxford* 1980, *Proceedings, Quantum Gravity* 2. Oxford University Press, Oxford, 1984b.

[Isham, 1992] C. Isham. Canonical quantum gravity and the problem of time. Lectures presented at the NATO Advanced Institute: "Recent problems in Mathematical Physics, Salamanca June 15", 1992.

[Isham, 1994] C. Isham. Quantum logic and the histories approach to quantum theory. *Journal of Mathematical Physics*, 35: 2157-2185, 1994. gr-qc/9308006.

[Isham, 1997] C. Isham. Structural problems facing quantum gravity theory. In M. Francaviglia, G. Longhi, L. Lusanna, and E. Sorace (eds.), *Proceedings of the* 14*th International Conference on General Relativity and Gravitation*, pages 167-209. World Scientific, Singapore, 1997.

[Isham, 1991] C. J. Isham. Conceptual and geometrical problems in quantum gravity. In H. Mitter and H. Gausterer (eds.), *Recent Aspects of Quantum Fields: Proceedings of the Xxx Int. Universitatswochen Fur Kernphysik, Schladming, Austria February and March* 1991 (*Lectur*), page 123. Springer-Verlag, Berlin, 1991.

[Jacobson, 1995] T. Jacobson. Thermodynamics of spacetime: The Einstein equation of state. *Physical Review Letters*, 75: 1260-1263, 1995.

[Jacobson and Smolin, 1988] T. Jacobson and L. Smolin. Nonperturbative quantum geometries. *Nuclear Physics B*, 299: 295-345, 1988.

[Klein, 1927] O. Klein. Zur fünfdimensionalen Darstellung der Relativitaetstheorie. *Zeitschrift für Physik*", 46: 188-208, 1927.

[Krasnov, 1997] K. Krasnov. Geometrical entropy from loop quantum gravity. *Physical Review D*, 55: 3505-3513, 1997.

[Kretschmann, 1917] E. Kretschmann. Über den physikalischen Sinn der Relativitätspostulate. *Annals of Physics (Leipzig)*, 53: 575-614, 1917.

[Kuchar, 1984] K. Kuchar. Canonical methods of quantization. In *Oxford* 1980, *Proceedings*, *Quantum Gravity* 2. Oxford University Press, Oxford, 1984.

[Kuchar, 1992] K. Kuchar. Time and interpretations of quantum gravity. In G. Kunstatter, D. Vincent, and J. Williams (eds.), *Proceedings of the 4th Canadian conference on General Relativity and Relativistic Astrophysics*, Singapore. World Scientific, 1992.

[Misner, 1957] C. Misner. Feynman quantization of general relativity. *Review of Modern Physics*, 29: 497-509, 1957.

[Misner, 1969] C. Misner. Quantum cosmology. I. *Physical Review*, 186: 1319-1327, 1969.

[Newton, 1962] I. Newton. De gravitatione et aequipondio fluidorum. Translation in [Hall and Hall, 1962], 1962.

[Norton, 1984] J. Norton. How Einstein found his field equations: 1912-1915. *Historical Studies in the Physical Sciences*, 14: 253-316, 1984. Reprinted in [Howard and Stachel, 1989, 101-159].

[Pauli and Fierz, 1939] W. Pauli and M. Fierz. On relativistic field equations of particles with arbitrary spin in an electromagnetic field. *Helvetica Physica Acta*, 12: 297-300, 1939.

[Pauri and Vallisneri, 2002] M. Pauri and M. Vallisneri. Ephemeral point-events: is there a last remnant of physical objectivity? *Dialogos*, 79: 263-303, 2002.

[Penrose, 1967] R. Penrose. Twistor algebra. *Journal of Mathematical Physics*, 8: 345-366, 1967.

[Penrose, 1986] R. Penrose. Gravity and state vector reduction. In R. Penrose and C. Isham (eds.), *Quantum Concepts in Space and Time*, page 129. Clarendon Press, Oxford, 1986.

[Peres, 1962] A. Peres. On Cauchy's problem in general relativity. *Nuovo Cimento*, 26: 53-61, 1962.

[Perez, 2003] A. Perez. Spin foam models for quantum gravity. *Classical and Quantum Gravity*, 20: R43-R104, 2003. gr-qc/0301113.

[Polchinski, 1995] J. Polchinski. Dirichlet branes and Ramon-Ramon charges. *Physical Review Letters*, 75: 4724-4727, 1995.

[Polchinski, 1998] J. Polchinski. *String Theory*. Cambridge University Press, Cambridge, 1998.

[Rosenfeld, 1930a] L. Rosenfeld. Über die Gravitationswirkungen des Lichtes. *Zeitschrift für Physik*, 65: 589-599, 1930a.

[Rosenfeld, 1930b] L. Rosenfeld. Zur quantelung der Wellenfelder. *Annalen der Physik*, 5: 113-152, 1930b.

[Rovelli, 1993a] C. Rovelli. Statistical mechanics of gravity and thermodynamical origin of time. *Classical and Quantum Gravity*, 10: 1549-1566, 1993a.

[Rovelli, 1993b] C. Rovelli. The statistical state of the universe. *Classical and Quantum Gravity*, 10: 1567-1578, 1993b.

[Rovelli, 1996a] C. Rovelli. Black hole entropy from loop quantum gravity. *Physical Review Letters*, 14: 3288-3291, 1996a.

[Rovelli, 1996b] C. Rovelli. Relational quantum mechanics. *International Journal of Theoretical Physics*, 35: 1637-1678, 1996b.

[Rovelli, 1997] C. Rovelli. Halfway through the woods. In J. Earman and J. Norton (eds.), *The Cosmos of Science*. University of Pittsburgh Press and Universitäts Verlag-Konstanz, 1997.

[Rovelli, 1998a] C. Rovelli. Strings, loops and the others: a critical survey on the present approaches to quantum gravity. In N. Dadhich and J. Narlikar (eds.), *Gravitation and Relativity: At the turn of the Millenium (15th International Conference on General Relativity and Gravitation)*, pages 281-331. Inter-University centre for Astronomy and Astrophysics, Pune, 1998a. gr-qc/9803024.

[Rovelli, 1998b] C. Rovelli. Loop Quantum Gravity. *Living Reviews in Relativity*, 1998b, www. livingreviews. org/Articles/Volumel/1998-1rovelli.

[Rovelli, 2000] C. Rovelli. The century of the incomplete revolution: searching for general relativistic quantum field theory. *Journal of Mathematical Physics*, 41: 3776-3800, 2000. hep-th/ 9910131.

[Rovelli, 2003] C. Rovelli. A dialog on quantum gravity. *International Journal of Modern Physics D*, 12: 1509-1528, 2003. hep-th/0310077.

[Rovelli, 2004] C. Rovelli. *Quantum Gravity*. Cambridge University Press, Cambridge, 2004.

[Rovelli and Smolin, 1988] C. Rovelli and L. Smolin. Knot theory and quantum gravity. *Physical Review Letters*, 61: 1155, 1988.

[Rovelli and Smolin, 1990] C. Rovelli and L. Smolin. Loop space representation for quantum general relativity. *Nuclear Physics B*, 331: 80, 1990.

[Rovelli and Smolin, 1995] C. Rovelli and L. Smolin. Discreteness of Area and Volume in Quantum Gravity. *Nuclear Physics B*, 442: 593-622, 1995.

[Rovelli and Speziale, 2003] C. Rovelli and S. Speziale. Reconcile Planck-scale discreteness and the Lorentz-Fitzgerald contraction. *Physical Review D*, 67: 064019, 2003.

[Smolin, 2000] L. Smolin. *Three Roads to Quantum Gravity*. Oxford University Press, Oxford, 2000.

[Smolin, 2002] L. Smolin. Matrix models as non-local hidden variables theories, 2002. hep-th/ 0201031.

[Smolin, 2003] L. Smolin. How far are we from the quantum theory of gravity, 2003 hep-th/ 0303185.

[Sorkin, 1983] R. Sorkin. Posets as lattice topologies. In B. Bertotti, F. de Felice, A. Pascolini (eds.), *General Relativity and Gravitation: Proceedings of the GR10 Conference*, volume I, page 635. Consiglio Nazionale Delle Ricerche, Rome, 1983.

[Sorkin, 2002] R. Sorkin. An example relevant to the Kretschmann-Einstein debate. *Modern Physics Letters A*, 17: 695-700, 2002.

[Stachel, 1999a] J. Stachel. Early history of quantum gravity. In B. Iyer and B. Bhawal (eds.), *Black Holes, Gravitational radiation and the Universe*. Kluwer Academic Publisher, Netherlands, 1999a.

[Stachel, 1999b] J. Stachel. Early history of quantum gravity (1916-1940), 1999b. Presented at the HGR5, Notre Dame, July 1999.

[Stelle, 1977] K. Stelle. Renormalization of higher derivatives quantum gravity. *Physical Review*, D16: 953-969, 1977.

[Strominger and Vafa, 1996] A. Strominger and G. Vafa. Microscopic origin of the Bekenstein-Hawking entropy. *Physics Letter*, B379: 99-104, 1996.

['t Hooft, 1973] G. 't Hooft. An algorithm for the poles at dimension4 in the dimensional regularization procedure. *Nuclear Physics B*, 62: 444-460, 1973.

['t Hooft, 1996] G. 't Hooft. The scattering matrix approach for the quantum black hole, an overview. *International Journal of Modern Physics A*, 11: 4623-4688, 1996. gr-qc/9607022.

['t Hooft and Veltman, 1974] G. 't Hooft and M. Veltman. One loop divergencies in the theory of gravitation. *Annales de l'Institut Henri Poincare*, 20: 69-74, 1974.

[Unruh, 1976] W. Unruh. Notes on black hole evaporation. *Physical Review*, D14: 870-892, 1976.

[van Nieuwenhuizen, 1981] P. van Nieuwenhuizen. Supergravity. *Physics Reports*, 68: 189-398, 1981.

[Wald, 1994] R. Wald. *Quantum Field Theory on Curved Spacetime and Black Hole Thermodynamics.* University of Chicago Press, Chicago, 1994.

[Weinberg, 1979] S. Weinberg. Ultraviolet divergences in quantum theories of gravitation. In S. Hawking and W. Israel (eds.), *General Relativity: An Einstein Centenary Survey.* Cambridge University Press, Cambridge, 1979.

[Wheeler, 1968] J. Wheeler. Superspace and the nature of quantum geometrodynamics. In C. DeWitt and J. Wheeler (eds.), *Batelle Rencontres*, 1967, volume 242 of *Lectures in Mathematics and Physics.* Benjamin, New York, 1968.

[Williams, 1997] R. Williams. Recent progress in Regge calculus. *Nuclear Physics Proceedings Suppl*, 57: 73-81, 1997. gr-qc/9702006.

[Williams and Tuckey, 1992] R. Williams and P. Tuckey. Regge calculus: A bibliography and brief review. *Classical and Quantum Gravity*, 9: 1409-1422, 1992.

第十三章　经典物理学中的对称性与不变性

凯瑟琳·布拉丁

埃琳娜·卡斯特拉尼

1. 引言

术语“对称性”伴随着许多古老的含义，包括美、和谐、各部分之间的相似、平衡、平等、均衡和规律。该术语的这些意义显然互相关联，现代物理学中使用的对称性概念就产生自这一类概念。我们对身边的物理客体的近似对称性很熟悉——人脸的左右对称，雪花旋转 60 度的转动对称，等等。我们把给定几何图形的对称性定义为，当等同的组成部分在特定的操作（比如旋转）下交换时该图形的不变性。19 世纪群的代数概念的发展，使得该概念能够进行推广与深化，出现的关于对称性的精确数学概念，不仅能够用于物理客体和几何图形，也适用于数学方程——故也适用于用数学方程表达的物理学定律，这是我们所特别感兴趣的。对称性的群理论概念是在规定群变换下的不变性概念。“不变性”是一个数学术语：当某物在给定变换下没有被改变时它就是不变的。这一数学概念用来表达我们所感兴趣的物理对称性，即群变换下的不变性。对称性概念在现代物理学中被证明是非常成功的，接下来我们将要探讨它。

第 2 节从物体的对称性与定律的对称性之间以及对称性原

理与对称性论证之间的区别开始。该部分包括了对居里原理(Curie's principle)的讨论。第3节讨论物理学中研究的对称性与群理论的数学方法之间的重要联系。我们也简要回顾了在19世纪群理论是如何首先应用于几何学进而应用到物理学，并以此开启群理论方法在当代物理学中获得至关重要性的篇章。考虑到这些，第4节对物理学中的对称性意味着什么以及物理学中不同对称性种类的分类法进行了说明。第5节我们讨论了经典物理学中对称性的一些应用，开始是经典力学中的变换理论，然后转到爱因斯坦(Einstein)的狭义与广义相对论理论(见第6节)。我们关注这些理论中对称性的作用与意义，进而延伸至第7节对诺特定理(Noether's theorems)的讨论。最后，在第8节中我们给出了关于在经典物理学中对称性的地位、作用以及解释方面的评论。请注意，我们的重点始终是经典物理学。对于在量子物理学中对称性的影响与重要性，我们建议读者参阅本书的其他章节，如迪克森(Dickson)(第四章，3.3小节)，兰兹曼(Landsman)(第五章，4.1小节)，特霍夫特(t' Hooft)(第七章)和霍尔沃森(Halvorson)(第八章，5.2小节)。[①]

2. 物体的对称性与定律的对称性

为什么必须区分物体的对称性与定律的对称性，这一点可从以下论述中找到答案。考察特定物体的几何对称性(如上面提到的雪花60度的旋转对称性与人脸的左右近似对称性)与考察物体的非对称性(如椅子不具有旋转对称性)是一回事，而考察支配那些物体时间演化的定律的对称性是另一回事。我们能够把力学定律应用到椅子的演化上，将其考虑为一个孤立系统，则这些定律是旋转不变的(它们并不在空间中挑选一个优先的方向)，即使椅子本身不具有旋转不变性。换言之，我们应该区分态或解的对称性与定律的对称性。当然在区分了这两种对称性之后，我们可以进一步考察二者之间的关系，例如目前对居里原理的讨论，见下文2.2小节。

① 进一步的讨论见[Brading and Castellani(布拉丁和卡斯特拉尼)，2003]。

2.1 对称性原理与对称性论证

在对称性原理与对称性论证之间作出区分同样重要。对称性原理在定律中的应用对20世纪的物理学而言至关重要，正如下面我们将在爱因斯坦的狭义与广义相对论理论中所看到的。要求定律——不论其精确形式如何——满足特定的对称特性，成为理论物理学家寻求各种定律具体形式过程中的主要方法论工具。

另一方面，对称性论证涉及的是基于某特定现象的对称特性进而得到关于该现象的特定结论。对称性的这种应用已有很长的历史，案例包括阿那克西曼德(Anaximander)对地球不动的论证、阿基米德(Archimede)的杠杆平衡定律，以及布里丹的驴子(Buridan's ass)。[①] 每个案例的相关论证都可以理解为应用莱布尼茨(Leibniz)充足理由原则(PSR)的一个例子：若某事件而非其他事件的发生不具有充足的理由，则什么也不会发生(即最初的情形并没有改变)。上述案例还有其他共同之处：在每一案例中充足理由原则都应用于如下基础上——初始情形具有某种对称性。[②] 初始情形的对称性意味着所有备选项都完全等价。若备选项完全等价，则没有充足的理由在它们之间进行选择，从而初始情形保持不变。这种论证最常采取如下形式：具有某种对称性的情形这样来演化，即没有一个非对称的原因时，保留初始对称性。换言之，初始对称性的打破不可能没有原因：非对称不可能自发产生。这种方式的论证也可见于对“居里原理”新近的讨论，接下来我们讨论这一原理。

2.2 居里原理

皮埃尔·居里(1859—1906)通过对晶体的热、电、磁特性的研究，仔细思考了物理系统的物理性质与它的对称特性之间的关系，因为这些性质与晶体的结构及对称性有着直接的联系。更准确地说，他提出了如下问题：在一给定的、具有特定对称特性的物理介质中(例如晶体介质)，允许发生哪种物理现象

---

① 对这些例子的讨论，见[Brading and Castellani，2003，Chapter 1，Section 2.2]

② 第一种情况下是旋转对称，第二、三种情况是左右对称。

(如哪种电磁现象)? 他的结论系统地汇总于他1894年的工作《论物理现象中的对称原理》(*Sur la symétrie dans les phénoménes physiques*), 可以总结为:①

(a_1)当某种原因导致某种结果时，该原因的对称性元素必定会在结果中发现;

(a_2)当某种结果表现出某种不对称性时，该不对称性必定能在给出该结果的原因中找到;②

(a_3)在实践中，这两个命题的逆命题并不为真，即结果能够比原因具有更多的对称性;

(b)现象可能存在于具有同一特征对称性或其特征对称性的子群对称性的介质中。换言之，某种对称元素可以与某种现象共存，但它们不是必然的。必然的是，特定的对称元素并不存在，不对称性创造了现象。

结论(a_1)在文献中常常被称为居里原理。结论(a_2)逻辑上等价于(a_1)，结论则是，对称性必然是从原因传递到结果，而不对称性却不是这样。结论(a_3)阐明了这一结论，强调因为不对称性不需要从原因传递到结果，所以结果可能会比原因更对称。③ 结论(b)援引了一个在所有居里案例中都可以找到的在“介质”与“现象”之间的区别。设有一个具有已知对称特性的介质，居里原理关心介质中发生的现象与介质的对称特性——或者更确切地说是“不对称性”之间的关系。结论(b)表明，居里认识到了不对称性——现在称为对称破缺——在物理学中起到的重要作用。

居里原理的应用需要满足许多条件，原因与结果必须明确定义，二者间的因果关联必须保持成立，原因与结果的对称性也都必须明确定义(这同时涉及

---

① 居里论文的英文翻译，见[Curie, 1981]，翻译中有一些误导。

② 居里在他的论文里使用了术语“不对称(dissymmetry)”，这是当时的习惯。这与现代术语中的对称破缺(symmetry breaking)意思相同，现今常常等同于非对称性(asymmetry)的意思。为了更精确，人们应该对一个对称破缺过程(破缺的对称性，broken symmetry)的结果进行区分，是缺少了与所考虑情形相一致的一种可能对称性(不对称性，就像在19世纪的文献中所称的那样，特别是在路易斯·巴斯德(Louis Pasteur)关于分子不对称性的研究工作中)，还是缺少了与所考虑情形相一致的所有可能对称性(非对称性)。

③ 请注意对一些作者而言结论(b)是一个独立的原理。瑞德卡迪(Radicati)(1987)更进一步将结论(a_1)(a_2)(b)分别描述为三个不同的原理：居里第一、二、三原理。

了所考虑物理系统的物理和几何特性)。居里原理为给定现象的发生提供了必要条件，只有那些与原理描述的对称条件相容的现象才能发生，从而居里原理具有了重要的方法论功能，一方面，它提供了一种选择规则(给定具有特定对称性的初始情形，则只允许特定的现象发生)；另一方面，它为物理理论提供了反证标准(违背居里原理可能表明物理描述中的某些内容是错误的)。

当然，居里原理的这些应用依赖于我们对其真实性的认可，而这正是文献中受到质疑的部分，尤其是与自发对称破缺相关的部分。为了证明该原理，人们提出了不同的方案。居里本人似乎把它看作是某种因果性原理。对此，近期的文献质疑该原理能否从包含对“因”与“果”定义的前提中得到证实。沿此方向，现已逐渐把该原理理解为是由决定性物理定律的不变性特性产生的。这一方法的开创性论文是[Chalmers(查尔默斯)，1970]，该论文用系统早期态与后期态的对称性之间的关系，以及联系这些态的定律阐述了居里原理的构造。这一“公认的观点”因为提供了与居里的意图明显不同的重构(因而被称为“居里原理”并不恰当)，并且建立在可能会破坏该观点的意义与重要性的假设之上而会受到批判。我们在下述简要评论中会讨论到这些。①

由于公认的观点本身与时间上有序的因果对(或系统的态)相关，它提供了一种历时的或动力学的分析。事实上，居里本人关注的是共时的或静态的情形，考虑了发生在同一时间不同现象间的相容性，而非系统从一个态向另一个态的演化。换言之，居里采用“因—果”术语并不是试图表明所考虑现象的时间顺序。这点可以清楚地从他的案例中看到，也可以从他自己的分析不包括对定律的讨论这一事实上看出来，而对定律的讨论是历时版本的核心。因而历时版本贴上“居里原理”的标签是对居里最初原理和他对该原理讨论的曲解。

然而，历时版本真的有意义且重要吗？其解释可以理解为是对充足理由原则的应用，其中我们非常关注定律是否为系统按照那些规律从初态向终态演化中的对称性破缺提供一个“充足理由”。厄尔曼[2004]对历时版本的重构拥有精

① 详细的讨论见[Brading and Castellani，2006]。我们将其首先归功于查尔默斯的“公认的观点”在[Ismael(伊斯梅尔)，1997]和[Earman(厄尔曼)，2004]中得到了发展。也见厄尔曼在本书中的论述。

确性上的优势，并能够进行这样的论证：若初态拥有一给定的对称性，且定律决定性地保持该对称性，则终态同样拥有该对称性。但是，事情并不是看起来的那样简单，因为该论证选取了一个具有给定对称性的态。一般地，确定一个态的对称性要求诉诸于背景的结构——如空间或时空，或解空间。在某些情形中，所要求的结构可能并不重要或是最低限度的。但尽管如此，系统的动力学仍然不独立于该结构（考虑空间或时空结构的案例，或更深刻的解空间）。一般会产生这样的结果：态的对称所依赖的结构和动力学的对称所依赖的结构并不相互独立，并且出现任何与上述论证相反的情形都需要谨慎处理。事实上，我们认为，对历时版本是否有意义和重要与否这一问题的回答，部分地取决于对该独立性缺失及其在论证中角色的研究，在关于“居里原理”历时版本的文献中已经提供了这些内容。

居里原理的最初版本和历时版本，都开始于物理系统态的对称性。在当代物理学中，关注的焦点已转向到定律的对称性，物理系统的对称性和定律的对称性之间的重要联系与那些系统态之间的对称性无关，而与解的系综的对称性有关。① 一般而言，在一个动力学方程的对称性把一个给定解转变为另一个解的意义上，一个动力学方程的对称性不是个别解（更不用说态）之间的对称性，而是整个解集的对称性。考虑定律与解之间的这种关系，就产生了居里原理的另一个版本，也就是我们这里提出的版本。② 和居里原理的历时版本一样，我们的提议背离了居里原始的方案，但我们认为该提议仍忠于居里最初研究的主要动机。在这一新的版本中我们试图统一两件事：

1. 我们把居里的动机问题理解为“哪种现象在物理上是可能的”，把他的方案解读为我们能够用对称性来引导对该问题的解答；

2. 我们在应用定律对称性方面超越了居里，他对此只字未提，而这已成为当代物理学的核心论题之一。

将这两部分结合起来，将简单陈述一个居里原理的“最新”版本，即定律

---

① 这里的“解”我们指的是一个系统在时间上展开的历史，系统的“态”即“瞬时解”。

② 请注意这个版本没有涉及从原因到结果的时间演化（如历时版本中一样），也没有局限于某一给定的瞬时或一定时间段中的系统态（如共时版本中一样）；相反，它涉及的是作为整体的解的系综的结构。

(方程)的对称性存在于其解的系综中。这一版本表达了居里的基本观点——“对称性不会丧失(没有原因的话)”——从本质上说就是定律的对称性要在解的系综中寻找这一事实。我们如此定义对称性与定律之间的关系并不会使其相对于居里的动机问题失去意义。正好相反，其意义在于当不是所有解都已知时，我们能够把定律的对称性作为求解的向导，即确定哪种现象在物理上是可能的。按照居里的说法，我们可以问“哪种现象是可能的”，并且我们可以将定律的对称性与解的系综的对称性之间的联系作为寻找物理上可能现象的向导。这样一来，一方面得到了一个明确的陈述即定律(方程)的对称性要在解的系综中寻找；另一方面，在我们不知道所有解的时候又有认知的突破口。我们相信这对于居里的动机问题而言是真的，正如上述(1)所表述的。

3. 对称性与群论：早期历史

群论是研究对称性理论最强有力的数学工具。在本节中，我们会在一开始给出群的定义，概括该概念在数学代数方程中的产生；然后转向群论在19世纪被首先应用到几何学而后应用到物理学的过程。

3.1 群概念的引入与群论的最初发展

一个群是一族$\mathcal{G}$元素g_1，g_2，g_3，…，并为其定义了乘法，通过该乘法可以从群中的任意两个元素g_i和g_j得到群的第三个元素(它们的积)$g_k=g_ig_j\in\mathcal{G}$，在下述要求成立的情况下：①

- 对于所有的g_i，g_j，$g_k\in\mathcal{G}$，有$(g_ig_j)g_k=g_i(g_jg_k)$(乘法的结合律)；
- 存在一个单位元素$e\in\mathcal{G}$，使得对于所有的$g_i\in\mathcal{G}$，有$g_ie=eg_i$；
- 对于所有的$g_i\in\mathcal{G}$，存在一个逆元素$g_i^{-1}\in\mathcal{G}$，使得$g_ig_i^{-1}=g_i^{-1}g_i=e$。

群的概念是由埃瓦里斯特·伽罗瓦(Évariste Galois)(生于1811年，去世于

① 通过放宽这些条件，群的概念可以被弱化(例如取消逆的要求将产生幺半群的概念，仅保留结合律将产生半群的概念)。于是问题出现了：是否完全群的结构或一些更弱的结构与给定理论的对称特性有某种联系。

1832 年）引入的，在他短暂的生命里，他也为用根式求解方程的数学问题做出了贡献。[①] 文艺复兴时期的数学家们发现了三次与四次方程的预解式，[②] 而是否存在用根式求解五次及更高次一般方程的公式，在很长一段时间都悬而未解，且在刺激着代数的发展。尤其是 18 世纪后半叶的研究集中于根置换下不变的函数在方程求解中的作用，从而产生了置换理论。在置换理论最有影响力的文献——拉格朗日（J. L. Lagrange）的《关于方程的代数解法的思考》（*Réflexions sur la résolution algébrique des équations*）中，得到了置换理论一些基本的结果。[③] 拉格朗日的文章成为了后续代数学发展的基础。这些后续发展包括从鲁菲尼（P. Ruffini）1799 年对 $n \geqslant 5$ 时用根式求解 n 次方程的不可能性的第一个论证，到柯西（A. L. Cauchy）、阿贝尔（N. H. Abel）以及最后伽罗瓦的意义深远的工作。[④]

伽罗瓦的工作[⑤]标志着一个转折点，他通过采用新的方法与代数概念——主要是群的概念——为求解方程这一悬而未解的问题提供了答案。伽罗瓦引进的这一概念是与方程根置换集的特性相关的（置换构成了他所谓的“群”），与“群”一起引进的还包括其他群论的基本概念，如“子群”“正规子群”和“单群”。[⑥] 通过用“对称度”来刻画一个方程，伽罗瓦能够将方程求解问题转换为对其所涉置换群特性的研究。其中，对称度由保留其代数关系（就是后来所熟知的方程的伽罗瓦群）的根的置换群来确定。通过这种方式，他得到了用根式

---

① 就是说，对方程的系数进行有限次代数操作——加、减、乘、除、自乘和开方——来求解。

② 二次方程的预解式从巴比伦时代就为人所知。对于代数方程预解公式存在问题的历史回顾见[Yaglom（亚格洛姆），1988，3f.]

③ 在其他结果中，所谓的拉格朗日定理用现代术语表述则是：有限群子群的阶是该群的阶的因子。

④ 1815 年柯西（1789—1857 年）推广了鲁菲尼的结论；阿贝尔（1802—1829 年）在 1824 年发表了利用根式法不可能求解五次方程的证明，并在 1826 年发表了论文“关于高于四次的一般方程的代数解法不可能性的证明”。

⑤ 包括一些提交到科学院的“文献记录”，1830 年发表在科学院“院刊”上的三篇简短论文，以及一些信件——其中最后一封是在他死亡决斗前一天晚上写给他的朋友奥古斯特・谢瓦利埃（Auguste Chevalier）的。

⑥ 详细内容请见[Yaglom，1988，9f.]

求解方程的充分必要条件。

伽罗瓦在群论中的成就，最早于1846年由约瑟·刘维尔(Joseph Liouville)予以发表，后在约尔丹(Jordan)于1870年的《置换论及代数方程》(*Traité des substitutions et des equations algébriques*)中得到整理与扩展。这本书是第一本系统的群论教科书，对这一新理论在数学科学其他领域——如几何学与数学物理学中的应用起到了决定性的影响。

3.2 群论的应用：克莱因(Klein)与李的贡献

射影几何学，不变量理论与群论：克莱因和李的出发点

在《置换论及代数方程》面世的同一年，索菲斯·李(Sophus Lie)与菲利克斯·克莱因这两位即将成为扩展群论应用范围关键人物的年轻数学家，为了能够接触数学的巴黎学派从柏林搬到巴黎住了一段时间。那时候，李与克莱因刚好合作写了一篇论文，研究了射影变换群下某些曲线的不变特性。事实上，他们到巴黎主要是受到他们对射影几何学兴趣的驱使。射影几何学由彭赛列(J. V. Poncelet)建立，用以研究图形在中心射影下保持不变的特性。那时，射影几何学由于结合了建立在不变性概念之上的代数方法与几何方法，已经成为一个成果颇丰的研究领域。不变量理论本身是数学的一个繁茂分支，是对代数形式的不变量系统研究的核心。利用不变量理论，英国的代数学家阿瑟·凯利(Arthur Cayley)①在这之前不久澄清了欧几里得(Euclid)几何学与射影几何学的关系，表明前者是后者的特殊情形。在离开法国前，克莱因基于用建立在射影空间上的二次式来定义距离(一个度规)的可能性，试图将凯利的结果推广到非欧几何学中。

在巴黎的时候，李与克莱因不仅认识了约尔丹，也认识了微分几何学的专家加斯东·达布(Gaston Darboux)。达布激发了他们对于微分几何学与射影几何学间联系的兴趣。

---

① 凯利是"不变的三重奏"三成员中的一位，这是法国数学家厄米特(Hermite)给他们的绰号，其他两位是詹姆斯·西尔维斯特(James Sylvester，他发明了不变量理论中的绝大多数术语，包括词语"不变量")和乔治·萨尔蒙(George Salmon)。

克莱因的爱尔兰根纲领(Erlangen Program)

克莱因对当时几种不同的几何学系统之间的关系有着异常的兴趣。他致力于为几何学新近分析形成的不同分支找到一个统一的基础性原理。在这方面，他卓有成效地把不变量理论在几何特性研究中的应用，与他和李关于把代数群论应用到关于几何变换处理的理念结合起来。他在著名的《爱尔兰根纲领》(*Erlangen Program*)中宣布由此得到了一个关于几何学理论新的群理论概念，该纲领随着一篇题为“对近期几何学研究的比较考察”①的就职演讲而广为人知，该演讲是 1872 年 23 岁的克莱因在就任爱尔兰根大学教授时所作的。在几何学最终是统一的这一观念的引导下，对由于不同几何学的存在而引起的问题，克莱因的解决方式是利用变换群下的不变性概念(即对称概念)提出对几何学理论的总体刻画。根据他的刻画，对于一个给定的域(平面、空间或给定的流形)和一组作用于其上的变换群，几何学被定义为研究群变换下不变量的科学。因而，每种具体的几何学都由特征化的对称群来确定(例如平面欧几里得几何由作用于平面上的仿射变换群所确定)，几何学间的相互关系可以用相应群之间的关系来描述(例如两种几何学的等价相当于两种对应群之间的同构)。

有了克莱因对几何学的定义，几何特性与对称特性变得非常接近：一个图形的对称性，定义在给定“空间”②上，空间的几何特性在群$\mathcal{G}$的变换下不变，则图形的对称性由保持图形不变的$\mathcal{G}$的子群来确定。新的群论方法推动了从归纳法(从 19 世纪用晶体可见的与显著的对称特征对晶体结构的分类中所熟悉)③向更抽象和更演绎的方法的转变。外尔(Weyl)在关于对称性的经典著作《对称》(*Symmetry*)[1952]中将该过程描述如下。

无论何时人们同一个被赋予结构的具体 Σ 打交道，就要设法去确定它的自同构群，也就是不改变所有结构关系的那些元素变换的群……在此之后，你就可以开始研究元素的对称构形，即在所有自同构群的某个子群下不变的构形[Weyl, 1952, 144]。

---

① “*Vergleichende Betrachtungen über neuere geometrische Forschungen*”。

② 一组赋予了结构的点。

③ 这方面的经典教科书是[Shubnikov and Koptsik(舒布尼科夫和科尔斯基), 1974]，也见下面的第 8 节。

如此一来，对称性分类就能够扩展到空间中与常见的平面和“空间”不同的图形。

克莱因本人对图形对称群分类的贡献在于他对离散群的研究，特别是，他研究了与正多面体对称特性相关的变换群，这已被证明在代数方程的根式求解中非常有用。

李的连续群理论

在1872年之后，尽管克莱因特别关注离散群变换，李却把他全部的研究投入到建立连续变换群论，这方面的研究结果系统地收录于他与恩格尔(F. Engel)合作的三卷本著作《变换群论》(*Theorie der Transformation gruppen*)[Ⅰ：1888，Ⅱ：1890，Ⅲ：1893]中。李对连续群发生兴趣与微分方程理论有关，后者被他看作是“现代数学中最为重要的学科”。在巴黎的时候，李已经开始研究一阶偏微分方程理论，该理论由于在哈密顿(W. R. Hamilton)和雅可比(C. G. Jacobi)的力学框架中的核心地位而格外受关注。① 他的计划是将伽罗瓦求解代数方程的方法扩展到微分方程，即利用关于方程“伽罗瓦群”的认识(由从解到解的变换所形成的对称群)求解它或将其约化为一个更简单的方程。因而，李的指导思想是，连续变换群在微分方程求解中的作用类似于伽罗瓦在代数方程情形中使用的置换群的作用。

在一些早期的几何学研究中李已经考虑到了连续变换群。在他和克莱因对特定类型曲线(他们称之为“W曲线”)的研究中，他考察了那些连续关联的变换，连续关联是在它们都是通过重复一个无穷小变换生成的意义上而言的。② 无穷小变换与连续变换群之间的相关性成为他关于切触变换研究的核心。之所以称为切触变换是因为它们保留了与表面的接触或相切。李开始把切触变换和暗含在他“线到球面的映射”中的几何互反性结合起来研究。线到球面的映射是

---

① 对于经典力学的详细情况，我们建议读者参考巴特菲尔德(Butterfield)在本书中的论述以及其中的参考文献。

② 关于本部分，请见[Hawkins(霍金斯)，2000，Section 1.2]。在霍金斯看来，李和克莱因在W曲线方面的工作“不仅是第一次开始研究连续群，也是首次将无穷小变换是系统连续变换特有且有用的性质这一理念见诸文字”。

他在巴黎时发现的在线几何与球面几何之间的一种映射。① 当他转而考虑一阶偏微分方程时，他很快认识到他们需要承认切触变换是对称性变换(即从解到解的变换)。因而，切触变换能够形成一阶偏微分方程的"伽罗瓦群"。这启发了他去发展切触变换的不变理论，这代表了他关于连续群综合理论的第一步。

李的关键性结果发现每个连续变换群都能够赋予一个代数(即现在所谓的它的李代数)，这一发现使得他能继续他的项目。李表明，连续变换群的无穷小生成元遵循群定律的线性化版本，其中涉及对易子括号(或李括号)，这一线性化定律表征了李代数的结构。简而言之(用现代术语表达)，我们把一个连续群(现在称为一个李群)②的元素(变换)描述为一定数量 r 个连续参量 a_l($l=1$, 2, …, r)的函数。且这些群元素可以用相应数量的 r 个无穷小算子 X_l 来表示，其中群的生成元 X_l 满足用李括号表征的"乘法交换律"

$$[X_s, X_t] = c_{st}^q X_q$$

从而形成了所谓的群的李代数。系数 c_{st}^q 是刻画群结构的常数，也称为群的结构常数。③

有了这一结果，对连续群的研究与分类就能够以相应的李代数来进行。后续的发展证明了这是相当有成效的，这不仅体现在代数学与几何学方面，也体现在物理学方面。首先来看李所处时代的物理学。李按照他求解微分方程的方法，通过重新解释泊松(Poisson)与雅可比(特别是雅可比)关于力学中一阶偏微分方程积分得到的结果，得到了连续群与李代数之间的对应关系。④ 因而他的成就对于当时的学者们讨论的动力学问题求解有着重大意义。

但是直到 20 世纪，随着如赫尔曼·外尔，艾米·诺特和尤金·维格纳(Eugene Wigner)等人(仅仅提到那些最先为把李理论应用到现代物理学上做出贡献的核心人物)的工作，李群理论与李代数才在物理学现象描述中取得了重要的地位。如今，从李开始的对这一理论的应用已涵盖了整个理论物理学，同

---

① 见[Hawkins, 2000, Section 1.4]。
② 对于本段中此术语与其他术语的精确定义，见巴特菲尔德在本书中的相关论述。
③ 更多的细节，见巴特菲尔德在本书中的相关论述。
④ 详细内容见[Hawkins, 2000, Section 2.5]。

时包括大的和小的方面：经典与量子力学、相对论理论、量子场论以及弦理论。[①]

4. 什么是物理学中的对称性？定义与种类

4.1 物理学中的“对称性”意味着什么

我们可以在直观上理解对称性的科学概念从物理客体或几何客体到定律(依此类推)的推广。我们用数学方程写下定律，不同的方程会有不同的数学客体和算符。对一个特别的变换群而言，这些客体与算符根据由相关的客体或算符的数学性质确定的规则或(在数学不能确定变换规则时)由我们制定的规则进行变换。当我们通过群的任意元素来对出现在方程中的每个客体与算符进行变换时，若方程的“形式”保持不变，那么我们说该群是方程的对称群。

更确切地说，我们所指的物理学中定律的对称性变换可以用如下表述中的一种来构造，它们在选取相同的变换集合这一意义上是等价的：

(1)应用于所涉理论中独立与非独立变量，保持定律形式不变的变换。

(2)从解映射到解的变换。

对称变换可从主动和被动两个角度来审视。从被动的视角看，对称变换是我们对同一物理演化在两个不同坐标系统下的再描述。[②] 即如表述(1)所示，变换独立或非独立的变量。若在原先坐标系下的描述是给定方程组的解，则在新坐标系下新的描述是同一方程组的解。(若该变换不是一个对称变换，则一般而言，在新坐标系下的新的描述将不会是同一方程组的解，而是其他方程组的解。)通过对称变换，从同一方程组的一个解向另一个解的映射产生了对此变换的主动解释。在主动解释中，这两个解被看作是在同一坐标系下描述不同的物理演化。因而，表述(2)也就很自然地成为主动解释。

表述(1)中的“定律形式”指的是定律的函数形式，通过独立变量和非独立

---

① 在本书中，特别参见特霍夫特(第七章)、迪克森(第四章)和贝洛特(Belot)(第二章)。

② 这里的“坐标”我们指的是广义坐标。一般来说，系统的每个自由度都对应一个坐标。

变量来表达。一般地，对那些变量的变换将产生这样一种表达，该表达的函数形式不同于原先的表达(例如，x变为x^2)。对此谈一谈“不变性”与“协变性”是有益的。读者应当谨记关于这些术语是如何应用于物理学定律的讨论尚没有达成一致，特别是在物理学哲学文献中。术语“不变性”经常用于客体，而“协变性”则用于方程或定律。但是，这背后还有更为本质的区别，只有对它们有了正确的理解才能把不变性概念应用于定律。

我们认为欧汉尼安与鲁菲尼(Ohanian and Ruffini)[1994, Section 7.1]的讨论非常有用，它非常漂亮地提炼了物理学和物理学哲学文献中的精华，提炼结果叙述如下。当一个方程的形式在给定变换下保持不变时，我们就说该方程在此变换下是协变的。这就是他们书中定义(1)中的概念。在某种程度上，这一概念非常弱，对于一个在给定变换下不具有协变性的方程，我们总能够改写它以使它具有协变性。另一方面，这种改写可能会引入新的变量函数，正是这些新的量的物理解释使得协变性获得了物理意义。我们将在下文6.3小节给出对广义协变的特殊情形和爱因斯坦广义相对论的更多讨论。

正如欧汉尼安和鲁菲尼所刻画的，方程的不变性是比协变性更强的要求。不仅是方程的形式应该保持不变，而且任意非动力学量的值——包括如光速之类的“常数”——也要保持不变。至于“非动力学量”，我们指的是所有那些出现在方程中，但它们本身不满足运动方程的客体。这里我们进入了如何区分“绝对的”客体和“动力学的”客体的困境，正如安德森(Anderson)所讨论的[Anderson，1967]。①

在这两种情形(协变和不变)下，相伴变换——当采用主动解释分析时——从解变换到解。当采用表述(2)时，弄清楚一个解的意义是什么将是非常重要的。它并不是指瞬时解，即系统的一个瞬态，而是指所涉系统的整个历史，即可能的时间演化。②

① 这也见下面的6.3节。处理关于此问题文献的困难就在于协变性、协变原理、不变性、绝对的和动态的客体等诸如此类常用术语在用法和意义上的多样性。

② 这一区分非常重要——比如在2.2节对居里原理的讨论中。

4.2 对称性的种类

物理学中的对称性拥有许多不同的种类，用诸如“全域”和“局域”、“内部”和“外部”、“连续”和“离散”之类的术语进行区分。本节我们将简要评论这些术语和随之而来的差别。

最为熟悉的是全域时空对称性，如牛顿力学的伽利略不变性，狭义相对论的洛伦兹(Lorentz)不变性。全域时空对称性被认为对所有自然定律都有效，因为所有的过程都发生在时空中。拥有这种普遍特征的对称性被维格纳称为“几何的”对称性，见[Wigner，1967，17]。

20世纪物理学中引入的一些对称性并不具有这一普遍特征。它们大多数是全新的种类，在科学史上没有先例，在某些情形下被特意地引入以描述特定形式的相互作用——维格纳给它们的名称是“动力学的对称性”[Wigner，1967，pp. 15，17—18，22—27，33]。

现代物理学的各种对称性也可以根据第二种差异进行分类：全域对称性和局域对称性。物理学和物理学哲学中用到的术语“全域”与“局域”有着许多不同的意义。这里所指的差异是在随常数参量而定的对称性(全域对称性)和随时间及空间的任意光滑函数而定的对称性(局域对称性)之间的差异。虽然洛伦兹不变性是全域对称性的一个范例，而经典电磁学的规范对称性(一种内部对称性)[①]和广义相对论的微分同胚不变性(一种时空对称性)却是局域对称性的范例，因为它们可以用时空的任意函数来表示。[②] 回想维格纳给出的区分则是，洛伦兹不变性是一种几何对称性，适用于所有相互作用，而电磁学的规范对称性特别与电磁相互作用相关，因而是一种动力学对称性。

经典电磁学的规范对称性是一种内部对称性，这是因为向量势的变换发生在场系统的内部空间中，而不是在时空中。经典电磁学的规范对称性可以看作只是一种数学奇特性，是该理论特有的，但随着量子理论的出现，对内部自由

---

① 更多关于规范对称性和内部对称性的讨论，见下面的段落。

② 我们将在下面的6.1节进一步讨论广义相对论的局域对称性，见贝洛特在本书中的相关论述。

度及相关内部对称性的应用，规范对称性开始变得很重要。①

非齐次洛伦兹群的平移、旋转与加速都是连续对称性的案例，就它们而言任何有限的对称变换都可以用无穷小对称变换建立起来。与连续对称性相反，我们有电荷共轭、宇称和时间反演(CPT)的离散对称性，此外还有置换不变性。因而，牛顿力学和经典电磁学在宇称(左右倒置)和时间反演(概言之：定律在沿时间倒退方向上演化的态序列中成立，正如它们在前进方向上的态序列中成立一样)下保持不变。只要我们正确地进行相关的电磁场变换，经典电动力学同样在电荷共轭下保持不变。最后，在如下意义上，经典统计力学置换不变：所假设的粒子彼此相同，且它们的置换是从一个解到另一个解。然而，离散对称性的作用和意义仅在量子理论中才能够完全展现。

在下文第 8 节中，我们将讨论一些与经典物理学中这些不同种类对称性相关的解释问题。

5. 对称性在经典物理学中的应用

5.1　经典力学的变换理论

正如我们看到的，李对连续群兴趣的产生与他对一阶偏微分方程理论的研究有关，而这一理论在哈密顿和雅可比给出的力学体系中占据核心地位。建立在这一体系上的力学变换理论确实是动力学方程的不变性特性在物理学中系统应用的最初案例之一。这些对称性是按照如下策略被应用的：运动方程的积分通过将原先的动力学系统变换——通过对称变换——为具有更小自由度的另一个系统而得到简化。

历史上，是拉格朗日和欧拉(L. Euler)打开了应用上述“变换策略”求解动力学问题的可能之路。欧拉—拉格朗日的分析力学体系，建立在拉格朗日影响深远的著作《分析力学》(*Mécanique Analytique*)(1788 年)之上，用一种在所有

① 对于与经典电磁理论中的规范对称性相关的解释性问题，见贝洛特[1998]。规范对称性随着量子理论的发展而凸显出来。术语"规范对称性"本身来自外尔 1918 年的引力与电磁场理论。对于规范对称性所有这些方面的讨论，见[Brading and Castellani，2003]。

坐标变换下协变(参见4.1节)的形式表达了运动定律。这意味着我们能够更方便地选择适合于所考虑动力学问题的坐标系。特别地，我们希望找到一个包含“循环的”(也称“可忽略的”)坐标的坐标系统。可忽略坐标的出现意味着方程的部分积分：若所有的坐标都可忽略，则问题就彻底且轻易地得到了解决。因而方法是试图找到(通过应用动力学不改变的坐标变换)更多的可忽略变量，从而将运动方程的积分问题转变为找到合适坐标变换的问题。

力学的分析方法的后续发展——从哈密顿的“正则运动方程”到雅可比得到的这些方程的广义变换理论(正则变换理论)——展现了“变换策略”观点的诸多优势。对于更详细的内容，我们推荐读者参考本书中巴特菲尔德的那章，还有经典的参考文献如兰索斯(Lanczos)[1949；1962]和惠特克(Whittaker)[1904；1989]。巴特菲尔德在本书第一章通过详述经典力学中的辛约化理论，全面阐释了通过使用对称性将力学问题简化的策略。这一策略也是巴特菲尔德[2006]的主要论题，该文关注根据诺特第一定理对称性是如何产生守恒量(见下文第7节)并由此减少在求解问题时需要考虑的变量数。

在结束这些关于经典力学中的对称性和变换理论的评论时，我们要强调两点。

首先，我们注意到通过对称方法将一个动力学问题(方程)转换为另一个等价问题(方程)的问题求解策略，可以看作是居里原理现代版本的应用案例(见2.2小节)：通过利用一个特定的对称性将一个方程转换为另一个等价的方程，可以得到一个我们能够求解的方程；新方程的解通过该特定的对称性与旧方程的未知解相关联；即我们由此得到了一个等价的解。

第二，我们强调的是，在所有这些发展中，动力学方程的不变性特征，虽然毋庸置疑地相当重要，但都是作为一种专门的辅助手段来对待的，即研究正则变换的唯一目的就是求解手头的动力学问题。给出方程，然后为了便于找到它们的解而研究其不变性特征。20世纪初爱因斯坦狭义相对论的提出带来了对这种关于对称性与物理定律之间关系的思考方式上的倒置，正如下文我们将要看到的。

5.2　作为理论构建导向的对称性原理

正如爱因斯坦在他 1905 年提出狭义相对论的论文中表达的那样，相对性原理断言：

那些物理系统的状态据以变化的定律，同描述这些变化所参照的坐标系是两个匀速的相对运动坐标系中的哪一个并无关系。①

这进一步证明了，这些坐标系是惯性参考系，它们之间通过构成非齐次洛伦兹群的洛伦兹变换相关联。

如此陈述的相对性原理满足在第 1 节中列出的对称性原理的条件：

- 洛伦兹变换，应用于理论的独立或非独立变量，当描述从一个惯性系变换到另一个惯性坐标系中时，不改变定律的形式；
- 洛伦兹变换将相对于某个惯性坐标系给定的一个解映射到另一个解。

这一原理被爱因斯坦明确作为理论构建的指导性原理，不管理论的最终细节如何，该原理必须满足。② 确实，只用了相对性原理与光速不变假设，爱因斯坦就得到了包括洛伦兹变换的许多结果，正如上面注意到的，这表明了对自牛顿时代以来给予相对性原理较动力学定律优先性的反转。惠更斯(Huygens)把相对性原理作为导出动力学结果的基本假设，但在牛顿那里，最早作为独立假设出现在他手稿中的相对性原理，在《原理》中被降格为一个推论。③ 从那时起直到爱因斯坦，惯性运动的相对性被看作所考虑的定律的结论，而一旦某种特定的相互作用定律细节清楚以后它可能被证明是错误的。一般经典物理学也类似，对称性——如空间平移和旋转——以前被看作是定律的特性，作为那些特殊定律的结果而存在。而在爱因斯坦那里变了，对称性可以假设为优先于已知的定律细节，可以用来为将要提出的定律施加限制。因而，对称性获得了一种新的地位，独立于定律细节被提出，从而具有了强大的启示性。正如维格纳

---

①　米勒(Miller)的翻译版[Miller，1981]，参见其中的第 395 页。

②　对于爱因斯坦关于原理性/构建性理论区别的讨论，见[Brown(布朗)，2006，Chapter 6]和[Howard(霍华德)，2007]。

③　实际上，它并不是由牛顿运动三定律得出的——我们必须进一步假定速度独立于质量和受力，见[Barbour(巴伯)，1989，Section 1.2]。

写的，爱因斯坦关于狭义相对论的论文“标志着趋势的逆转”，在爱因斯坦的论文之后，“现在通过定律的不变性来导出自然定律并检验其有效性是很自然的，而不是从我们所认为的自然定律导出不变性的定律”[Wigner，1967，5]。

对爱因斯坦狭义相对论(STR)奏效的方法论在他提出广义相对论(GTR)时也起了作用。广义相对论中，爱因斯坦把许多不同的原理作为了最终理论具有的可能形式的约束条件。① 爱因斯坦所坚持的这些原理之一就是拓展在STR中建立的相对性原理，以涵盖相对于其他坐标系作加速运动的坐标系，实现该原理的方式是要求他的新理论中的方程是广义协变的。爱因斯坦那时正在寻找牛顿水桶问题的“马赫”解，他认为这要求不能有优先的参考系。因而，在他1916年的评论文章中，爱因斯坦写道：“*物理学定律必须具有这样的本性，即它们适用于任何运动种类的参考系*。沿此思路我们得到了对相对性假设的拓展。”(此处用斜体表示强调是原文中就有的。)

关于广义协变原理经常讨论的两个问题是：(a)它是否会使任意的平滑坐标变换成为一个对称变换；(b)它是否是相对性原理的一般化。问题(b)的回答显然是“否”，但当前对于(a)的回答尚未达成一致。② 我们将在下一节讨论(a)。这里我们用对(b)的一些简短评论作为结束。

即使GTR中的广义协变性是一个对称性原理，它也不是相对性原理的拓展。这就是说，广义协变性并不谈论不同参考系的观测等效性。③ 正如已经注意到的，爱因斯坦认为广义协变性可能提供了这样一个原理的想法与他为牛顿水桶问题寻求“马赫”解，以及他的等效原理都是相联系的。但是，等效原理并不意味着处于任意运动态的参考系具有观测等效性(爱因斯坦也一直如此认为)，并且爱因斯坦最后也认识到GTR无法证明牛顿水桶的解只与物质的相对运动有关。④

---

① 这些原理主要有：相对性原理——区别于后来(在1918年)爱因斯坦所指的“马赫原理”；等效原理；能量—动量守恒原理。

② 可参见[Torretti(托雷蒂)，1983，pp. 152—154]，诺顿(Norton)[1993]讨论了它与等效原理的关系，也可参见安德森[1967]。

③ 进一步的讨论，可参见[Norton，1993]和[Torretti，1983，Section 5.5]。

④ 简洁明了的讨论，见[Janssen(詹森)，2005]。

不论广义协变性是否以及在何种程度上是一个对称性原理的微妙之处，很显然它拥有巨大的启发力，不仅对爱因斯坦发展 GTR 如此，而且不止于此。比如希尔伯特(Hilbert)对物理学公理化的研究，见[Corry(考瑞)，2004]及其参考文献，以及外尔构建一个统一场理论的尝试，见[O'Raifeartaigh(奥瑞弗蒂)，1997]，对外尔 1918 论文“引力与电”的英文翻译，也见[Weyl，1922]。在所有这些情形下，广义协变性为理论构建提供了强有力的工具。在下一节中我们将进一步讨论 GTR 中广义协变性的意义与解释。①

6. 广义相对论中的广义协变性

在前一节中我们注意到了广义协变性原理在理论构建中的导向作用。本节中我们把注意力转向一些与 GTR 中广义协变性相关的更深入的问题，这些问题在哲学文献中引起了关注。我们将开始于前一节中提起的问题，即关于任意的平滑坐标变换作为对称变换的情形。然后我们将讨论与广义协变性相关的各种特征，包括那些爱因斯坦在所谓“洞问题”中提到的特征，之后转向广义协变性是否具有物理内容的问题。② 我们把对诺特定理的讨论放在第 7 节中。

6.1　广义协变性和作为对称变换的任意坐标变换

广义协变原理会使得任意平滑坐标变换成为对称变换吗？处理本问题的一种方式是，考虑主动变换而非被动变换(见上文 4.1 节)，并且比较 GTR 与 STR 中的情形。

在 STR 中，一个洛伦兹变换——按照主动解释——接管物质场，并根据用度规编码的时空结构重新分配它们。相对性原理对于这样的变换成立，这是因为物质场在两种情况下(通过洛伦兹变换相关联)的演化在观测上是不可区分的：在实践中或是原理上，没有观测能够区分这两种情形。在 GTR 中，主动

① 对于狭义与广义相对论的详细陈述，见马拉蒙特(Malament)在本书中的论述，也见罗韦利在本书中的论述。

② 也见贝洛特在本书中的论述。

广义协变是通过在时空流形上的主动微分同胚实现的，参见罗韦利在本书第十二章4.1小节的论述。这些涉及的不仅有物质场变换，也有度规场变换，在两种情形中都是根据时空流形来重新分配的。再次说明，"两种情形"在观测上"不可区分"，但这次笼统给出的原因是"两种情形"实则是一种。①

为什么在考虑到STR中的洛伦兹变换时我们要接受存在两种本质上不同的情形，而对于GTR中的微分同胚只存在一种情形？一种途径是宣称二者间的关键差别在于洛伦兹变换可以在一有效隔离的物质场子系统上实现，并产生在观测上的不同情景。不过，在该场景中若不参照所考虑子系统之外的物质场，则该子系统的演化是不可区分的。例如，在伽利略著名的船实验中，我们考虑两个观测上不同的情景——一种是船相对于河岸是静止的，另一种是船相对于河岸做匀速运动的——我们注意到两种情形下船舱中物理系统的行为是不可区分的。② 对于GTR的广义协变性而言，不具有与伽利略船实验的相似性。③

科索(Kosso)[2000]强调了可执行的对称变换对于产生观测上不同情景的重要性。在这种观点中，对称变换的观测意义在于将原则上可能的两种观测结合起来。首先，必须有可能在经验上确证实施了变换——进而确证能够通过子系统的变换产生观测上不可区分情景的重要性。其次，我们必须能够观测到，子系统随后的内部演化不受影响。GTR中任意的平滑坐标变换不满足这些要求中的第一条，这就标志着这些变换不同于洛伦兹变换。④

在这一方法下，虽然GTR的场方程对于任意的坐标选择都具有相同的形式，但这对于任意坐标变换具有对称性而言是不充分的。此外，主动解释的变

① 也见6.2小节。

② 该实验只能是近似的，这取决于所考虑的子系统能在多大程度上与"外界的"物质场隔离开来。

③ 一个可能的建议是：我们执行一个变换 T，此变换在某一区域 R 外是恒等变换，且不同于区域 R 中的恒等式，这不会得到期望的结果。至少在原理上，这两个情景必须具有可区分的观测后果。在伽利略船的情形下，如果我们允许子系统再次与其他物质发生相互作用，我们将看到：一种情况是船撞上礁石，另一种情况是船不会撞上。于是，原理上就有了观测的可区分性。变换 T 无法产生一个这样的情景：任何未来的事件都可以使我们能够将它与最初的情景区分开来。

④ 事实上这一结果普遍地适用于局域和全域对称性之间的不同。也见[Brading and Brown，2004]。

换必须拥有一物理解释——我们必须相对于其他事物来变换一个事物。当我们执行一个微分同胚时，我们得到同一个解而非一个新解，因为我们没有相对于度规来重新分布物质场。

我们要强调的是，这只是求解 GTR 中广义协变性是否应该被理解为是一个对称性原理这一问题的唯一方式。与此对立的立场可以在[Anderson，1967，Section 10—3]中找到，他论证说我们必须把爱因斯坦理解为他把广义协变性看作是对称性要求，并且试图找出在什么条件下它可以发挥其作用。

6.2　广义协变理论的特征

任何广义协变的理论都拥有哲学上值得关注的某些特征。首先是理论中的因果性和决定论有显而易见的问题，其次需要对初始数据的说明有所约束。爱因斯坦是在寻找他的引力理论的过程中认识到了第一种特征的各个方面。从1913 年直到 1915 年秋天这段时期，他认为他所谓的“洞问题”为不存在物理上可接受的广义协变理论这一结论奠定了基础。

在“洞问题”中，爱因斯坦考虑了一个没有物质场(洞)的时空区域，然后指出在一个广义上协变的理论中，关于洞之外的物质与引力场值的数据再多也不足以唯一地确定洞内引力场的值。由此，爱因斯坦推断出不存在物理上可接受的广义协变理论。①

这里需要了解的背景是，爱因斯坦那时正在寻找一个物质场加上场方程能够唯一地确定度规的理论。② 1915 年夏天，爱因斯坦在哥廷根讲授相对论，听众包括戴维·希尔伯特。若我们假定爱因斯坦的讲授中包括他“洞问题”的一个版本，那么我们能够合理地推断希尔伯特很快就对“洞问题”所指的问题进行重新解释，并且用广义协变理论是否允许柯西问题适定来指出该理论存在的问题。③

---

① 对“洞问题”的阐释和讨论见诺顿[1984，pp. 286—291]和[1993，Sections 1—3]，斯塔彻(Stachel)[1993]和里克曼(Ryckman)[2005，Section 2. 2. 2]，也见罗韦利在本书中的论述。

② 关于爱因斯坦对马赫原理(滥)擅用的更多内容，见[Barbour，2005]。

③ 见布拉丁和里克曼[2007]，也见[Brown and Brading，2002，Section Ⅳ]。

在 GTR 提出后的几年里，希尔伯特在指出任何广义协变理论面临的因果性与决定论问题中扮演了核心角色。他指出，在任何一个这样的理论(包括 GTR)中，场方程的数目都会比变量数少 4 个，这将导致理论的数学非充分决定性。正如希尔伯特所强调的，柯西问题不是适定的：给定初始数据一个说明，场方程确定不了变量的唯一演化。

接下来，我们能够看到非充分决定性问题与广义协变性之间的联系。为了使柯西问题适定，我们必须能够用初始数据来表达度规的二次时间微分(外加能够从初始数据计算得到的更进一步的空间微分)。但是，若我们重新表述 10 个(无源的)爱因斯坦场方程 $G_{\mu\nu}=0$，以便清楚地展示包含度规二次时间微分的所有项，我们就会看到对于 6 个未知的 $g_{ij,00}$ 有 10 个方程，剩下的 4 个二次时间微分 $g_{\mu0,00}$ 没能够出现在方程中。① 这是广义协变性的直接影响，我们总能够在初始数据曲面邻域进行坐标变换以使度规分量和它们的一次微分保持不变，而二次时间微分 $g_{\mu0,00}$ 在该曲面为零。因而，必须在所有坐标系中有效的场方程，也许不能够包含关于二次时间微分的信息。初始数据并不唯一地确定度规：存在 4 个我们可以自由选择的函数 $g_{\mu0,00}$。

现在，从一开始就习惯去断言，仅在对这 4 个任意函数选择上不同的爱因斯坦场方程的解在物理上是等价的。② 但在这里我们应当注意到，广义协变的这一“规范自由”解释有它自身的问题。③ 例如，在该框架下理论的可观测量必须是“规范不变的”量，但这样的量(到目前为止)都被证明不属于具有操作意义上的“可观测”的任何东西。广义协变性的规范自由解释有时也持有这样的观点，即规范自由——因此广义协变本身——缺少物理内容。我们将在 6.3 节中考虑该问题。

在上文所述的我们对非充分决定性问题的解释中，我们注意到爱因斯坦场方程为 6 个未知的 $g_{ij,00}$ 给出了 10 个场方程。因而，非充分决定性问题的另一面

① 对过充分和非充分决定性问题讨论的细节见[Adler, Bazin and Schiffer(阿德勒，巴赞和希弗)，1975，Chapter 8]。

② 回顾上文 6.1 节的讨论。

③ 见贝洛特在本书中的论述。

是一个关于 $g_{ij,00}$的过充分决定性问题，这意味着初始超曲面上的数据说明要有所约束。这是我们在本节一开始的评论中提到的所有广义协变理论的第二个特征。事实上，出现约束方程也是具有局域对称结构的其他理论如电磁学所共同具有的特征，从哲学上看，其意义存在于理论与初始数据之间的关系中。17 世纪时，笛卡尔描绘了一个处于无序态的世界的故事，通过对自然定律的常规操作，会从该世界中出现一个类似于我们所拥有的世界。① 当然，世界产生自初始混沌的这一图景已有很长的历史，但是通过对自然定律的操作涌现出有序则是一个新的转折。它涉及将初始条件从随后受规律支配的宇宙演化中分离出来，这些初始条件可以是任何事情。用现代的术语来说，这是一个没有约束的理论：理论决定系统的哪种特性必须被确定以给出恰当的初始数据，但之后我们就能自由地为这些特性赋予我们想要的值；理论的方程用于那些数据随时间向前的演化。相比之下，有约束的理论包括两种类型的方程：必须满足初始数据的约束方程，还有演化方程。在 GTR 中，10 个场方程中的 4 个把初始数据超曲面的曲率与超曲面上的物质—能量分布联系起来，剩下的 6 个场方程是演化方程。总之，在有约束的理论中，初始的“无序”毕竟不能够如此的无序，但它本身必须满足由理论定律设定的约束。

6.3 广义协变性有物理意义吗？

正如我们在第 5 节中看到的，爱因斯坦把广义协变性看作是引导他提出广义相对论的对称性原理。毫无疑问，广义协变性对爱因斯坦是一个有益的启发，但是在广义协变性是否事实上具有物理意义这个问题上争议从未停止过。早在 1917 年克雷茨曼(Kretschmann)就有力地提出了这一问题。到今天仍在继续的这一争论的重点是，在充分精巧的数学设计下任何理论都可以给出广义协变的公式化表述，因而广义协变原理并不约束理论的物理内容。事实上，诺顿[2003]开始他对这一问题的讨论时称这种广义协变性的负面观点已经成为主

---

① 写于 1633 年前后，《世界报》(*Le Monde*)并未在笛卡尔有生之年发表它。英文译本请见笛卡尔[1998]。“有序出于无序”的故事在《光的专著》(*Treatise on Light*)中第 6、7 章。是否自然定律的常规操作足以从混沌中产生出有序成为一个众说纷纭的议题。

流，讨论展开之前他给出了一种替代观点(见下文)。

对我们而言清楚的是，上面讨论的广义协变理论的特征性质至少在某些理论(包括GTR)中并非毫无意义，并且主流观点——那些实际宣称这些问题毫无意义的观点——应该被反对。那些试图反对主流观点的人们采取了两步走的策略：首先，表明在什么条件下广义协变性对理论的物理内容有所限制；其次，揭示对物理内容有什么影响。因此，承认任何理论都可以纳入广义协变形式这一普遍数学观点，但是，通过关注广义协变性在一给定理论或一类理论中的实行方式，要抵制广义协变性必定在物理上毫无意义的想法。

例如，安德森[1967]，欧汉尼安与鲁菲尼[1994]，诺顿[2003]与厄尔曼[2006]都曾试图解释在什么条件下广义协变性的纯粹数学特征开始具有物理内容。[1] 安德森对理论的对称性(它拥有物理意义)和方程的协变群(不需要有物理意义)进行了区分。安德森是区分"绝对的"与"动力学"客体[2]的经典参照，按照这种区分，当且仅当一个理论不包含绝对客体时理论中方程的协变群才成为一个对称群。欧汉尼安与鲁菲尼[1994]则诉诸于他们在理论方程的不变性与协变性之间做出的区分。[3] 他们认同协变性是一个数学特征(或许仅仅是当前理论一种特殊构造的人工产物)，但我们要求的不仅是方程的协变性，也要求对于独立于物质状态(如光速和普朗克常数之类)的理论中出现的任意客体(包括一种或更多的组分)而言，它们的值在广义坐标的变换下应保持不变。诺顿[2003]强调了在确定理论内容时物理考虑的作用，如此一来就限制了我们能够玩的形式游戏。厄尔曼[2006]开始尽力强调存在于"纯粹的坐标自由"(按照被动解释的任意坐标变换相关)与"微分同胚不变性是所涉理论的规范对称性这一实质要求"之间的差别。这是说，他提醒我们目前紧要的问题不是我们把一个

---

① 也见[Norton, 1993, Section 5]和罗韦利在本书中的论述。

② 事实证明很难将绝对客体和动力学客体进行准确区分，但是直观地区分却足够清晰。动力学客体满足场方程并与其他客体发生相互作用，然而绝对客体不受理论中其他场的动力学行为影响。对安德森的方法及其反例进行详尽的处理，请见[Pitts(皮兹), 2006]。这篇论文的结论是安德森的直觉可以足够的精确以应付到目前为止出现在文献中的所有反例(包括由皮兹自己提出的反例)，但是还是存在一个皮兹不能解决的反例，即杰拉奇(Geroch)提出的反例。争论仍在继续!

③ 见4.1节。

理论改写为广义协变形式的能力(承认我们总有给定充分精巧的数学设计的能力),而是通过微分同胚也就是(主动)点变换(见上文6.1节)相关联的物理情形之间的关系。当通过微分同胚关联的理论模型表征了对同一物理情形的不同描述时,"实质的广义协变性"成立。我们声称,GTR满足实质的广义协变性,而像STR这样的理论的一般协变表述(generally covariant formulations)就不需要被满足,我们的目标是表明这一要求给出了两类理论的分界线,在其中一类理论中广义协变性能够表征理论在物理上的重要特性,而在另一类中则不行。①

因而,安德森、欧汉尼安与鲁菲尼、诺顿,以及厄尔曼都试图通过对广义协变性在理论中的实现方式施加条件以赋予广义协变性"只不过是数学的"要求增加内容。一旦增加了这些要求,就会对理论的内容产生不同的影响,比如度规成为一个动力学客体。在每种情形下,目标都是把在GTR中实现的广义协变性提升为一个对称性原理。②

对引力理论中广义协变性意义的考虑形成了三条定理,这些定理对物理学中对称性的一般解释很重要。这些定理是由艾米·诺特和菲利克斯·克莱因完成的,这将在下节中讨论。

7. 诺特定理

任何在物理学世界中关于对称性意义的讨论,若不提到诺特定理都将是不完整的。这些定理把理论的对称性特征与其他重要的特征如守恒律联系了起来。

在物理学中,术语"诺特定理"是最常与在全域连续对称性与守恒量之间的联系相关联的。经典力学中熟悉的案例包括下述联系:空间平移与线性动量的守恒;空间旋转与角动量的守恒;时间平移与能量的守恒。事实上,这一定理是她在1918年的论文"变分问题的不变量(Invariante Variationsprobleme)"中提

① 诺特第二定理是将真正的"规范理论"区分于那些在其中所涉局域对称性仅是形式上的理论的重要手段,见第7节。

② 布朗和布拉丁[2002]试图通过诺特定理(见第7节)详细分析为得到GTR内容的特定方面需要在广义协变性上增加什么样的附加条件。

出的两个定理之一。①

在陈述两个定理之前，我们先给出下述告诫性的评论。在变分对称性(与作用的不变性相联系，诺特定理是通过该术语来构造的)与动力学对称性(与动力学定律相关，是我们这里讨论的主题)之间的关联很微妙，见[Olver(奥尔弗)，1993，Chapter 4]。诺特本人从未提到该联系，也从来没有在论文中用到“对称性”一词。她讨论了与拉格朗日物理学的作用量积分在数学上类似(但是其一般化)的积分，并用变分方法和群论来引出积分的变分对称与一组恒等式之间的成对的对应关系。

之后诺特证明了两个定理，第一个用于变分对称群依赖于常数参量的情形，第二个用于变分对称群依赖于变量的任意函数的情形。② 在接下来对她的定理的表述中，我们用术语“诺特对称”指代场方程的对称性，对此而言，产生于无穷小对称变换的作用变化至多是一个表面项。采用4.2节的说法，则第一种类型的对称对应于全域动力学对称，第二种对应于局域动力学对称。我们用适用于拉格朗日场理论的形式来表述这些定理；诺特自己对这些定理的表述并没有这样的规定。在有限维经典力学语境中对第一个定理的讨论见巴特菲尔德在本书中的相关论述。我们这样来表述这些定理，以使得我们能够参照它们来描述概念性的内容，但关于它们导出和内容的数学细节我们建议读者参照其他文献——尤其是[Olver，1993]和[Barbashov and Nesterenko(巴巴斯沃与内斯特兰科)，1983]。

对于依赖于场 $\phi_i(x)$ 及其一次微分的拉格朗日密度 L，我们可以将诺特的两个定理表述如下。

诺特第一定理

若一个平滑地依赖于 ρ 个常数参量 $\omega_k(k=1, 2, \cdots, \rho)$ 的连续变换群是与拉格朗日量 $L(\phi_i, \partial_\mu\phi_i, x^\mu)$ 相关的欧拉—拉格朗日方程的诺特对称群，则满足下面的 ρ 个关系式，对称群依赖的每一个参数都相应有一个关系式：③

① 英文译本见[Noether，1971]。

② 见[Brading and Brown，2007]。

③ 注意我们在用爱因斯坦求和约定对重复的希腊符号进行求和。

$$\sum_i E_i^L \xi_i^k = \partial_\mu j_k^\mu \tag{1}$$

式子左边是一个欧拉表达式的线性组合：

$$E_m^L \equiv \frac{\partial L}{\partial \phi_m} - \partial_\mu \left(\frac{\partial L}{\partial \phi_{m,\mu}} \right) \tag{2}$$

其中

$$E_m^L = 0 \tag{3}$$

是场 ϕ_m 的欧拉—拉格朗日方程。(ξ_i^m 依赖于所考虑的特定对称变换与场，其细节对我们当前的目的来说并不重要。)

式子右边是流的散度 j_k^μ。当左边为零时，流的散度等于 0，且该表达式能够转换为满足一定条件的守恒量。从而，诺特第一定理给出了全域对称性和守恒量之间的联系。①

诺特第二定理

若一个平滑地依赖于 ρ 个时空的任意函数 $p_k(x)$ ($k=1, 2, \cdots, \rho$) 和它们一阶微分的连续变换群，是与拉格朗日量 $L(\phi_i, \partial_\mu \phi_i, x^\mu)$ 相关的欧拉—拉格朗日方程的诺特对称群，则如下 ρ 个关系式满足，对称群依赖的每一个函数相应有一个关系式：

$$\sum_i E_i^L a_{ki} = \sum_i \partial_\nu (b_{ki}^\nu E_i^L) \tag{4}$$

a_{ki} 与 b_{ki}^ν 依赖于所考虑场的特定变换，并且这里也不需要关心细节，但出于下面的应用，我们注意到尽管 a_{ki} 甚至在对称变换是全域变换时也出现，b_{ki}^ν 却仅在对称变换是局域变换时才出现。② 从本质上说，我们在这里得到的是欧拉表达式与它们一阶微分之间的从属性。这种从属性是推导定理时用到的局域对称性的结果。当所有的场都是动力学场时（即满足欧拉—拉格朗日方程），会出现不是所有的场方程都相互独立的情形。这种形式上的非充分决定性是具有局域对称

① 这一定理得到了广泛的讨论。特别见[Barbashov and Neoterenko，1983]和[Doughty（道蒂），1990]。关于有限维经典力学背景下的诺特第一定理，我们推荐读者参考巴特菲尔德在本书中的论述和[Butterfield，2006]。

② 更多的细节请读者再次参考[Brading and Brown，2007]。

结构理论的特征。①

正如希尔伯特在引力的广义协变理论语境中所认识到的，这种非充分决定性独立于拉格朗日量的具体形式。② 在广义相对论情形下，一旦我们确定拉格朗日量并将其代入式(4)，我们将得到(缩并形式的)比安基恒等式(Bianchi idertities)。

对诺特本人而言，她的论文的灵感来自于希尔伯特、克莱因以及爱因斯坦对能量守恒在广义协变理论中地位的讨论。在此讨论期间希尔伯特评论说，物质场的能量守恒在广义协变理论中不再具有与在之前(非广义协变的)理论中相同的地位，因为它独立地遵循物质场的场方程。诺特的两个定理可以用来支持这一猜想，参见[Brading，2005]。关于能量守恒在广义相对论中地位的讨论仍在继续，总的说来，该问题的根源在于能量—动量不能够局域地定义。③

今天来看，诺特理论的意义在于它的普遍性。在诺特 1918 年的论文之前，人们已经知道了许多全域时空对称和与它们相关的守恒量之间的特定联系，并且在 GTR 形成的过程中及之后时间里，爱因斯坦与希尔伯特在他们对于能量守恒的研究中也都预见到了第二定理的某些方面。④ 但是，诺特系统的处理能够让我们认识到，这些关系其实并不依赖于一个特定理论的具体动力学，而是从拉格朗日理论结构和比理论上完整的动力学要弱很多的规定中得到的。例

① 第二定理所表达的依赖性有意义与否取决于局域对称性相对于其成立的场的状态。利用这种方式，诺特第二定理可用作尝试从那些局域对称性是“完全的数学产物”的理论中区分出“真正的规范理论”的一个有效工具，参见上面的6.3节和[Earman，2006]。对于一个“真正的规范理论”而言，从属性具有重要的物理含义。

② [Hilbert，1915]。

③ 广义相对论中的能量—动量守恒定律是从与物质场相关联的能量—动量张量的协变散度等于零来精确表述的。或者，我们可以通过物质场能量—动量的坐标散度加上引力场能量—动量的坐标散度等于零来表述。受散度算符的影响，引力场的能量—动量坐标散度不是唯一定义的，并且与非定域性问题相关，它在某些坐标系中等于零而在另外一些坐标系中则不等于零。通过对等效原理的思考，我们可以理解这种坐标依赖性：按照等效原理，区分惯性场—引力场以得到惯性力与引力的区分本身就是个依赖于坐标系的问题。进一步的讨论，可参见[Misner(米斯纳)，Thorne and Wheeler(索恩和惠勒)，1970，pp. 467—468]及[Wald(瓦尔德)，1984，70]。也见马拉蒙特在本书中的论述。

④ 关于爱因斯坦见[Janssen，2005，pp. 75—82]；关于希尔伯特见[Sauer(萨奥尔)，1999]。

如，假定满足引力场方程，但独立于那些方程的具体形式，也独立于物质场的场方程(事实上，和物质场是否满足欧拉—拉格朗日方程完全不相关)，广义协变性将引起 GTR 中的能量守恒。[①] 诺特定理成为研究理论结构的有力工具——产生理论的哪些方面需要哪些假设，如此等等。[②]

很值得提到第三个定理，它是由菲里克斯·克莱因[1918]在相同的语境下(即对引力的广义协变性与能量守恒的研究中)得到的，并与诺特的两个定理相联系。出于与导出它的方法相关的原因，我们称其为“边界定理”。[③] 与诺特第二定理一样，边界定理关系到的是局域对称性，并且得到了一系列等式，朱利亚和席尔瓦(Julia and Silva)[1998]称它们为“级联方程”。[④] 这里我们陈述一个简化版的边界定理：在无穷小对称变换下(即不允许有表面项的可能)，作用量保持不变。[⑤]

边界定理(约束形式)

若一个平滑地依赖于 ρ 个时空任意函数 $p_k(x)$ $(k=1,2,\cdots,\rho)$ 和它们一阶微分的连续变换群是与拉格朗日量 $L(\phi_i,\ \partial_\mu\phi_i,\ x^\mu)$ 相关联的欧拉—拉格朗日方程的诺特对称群[⑥]，则如下三组 ρ 个关系式满足，对称群依赖的每一个参数相应有一个关系式：

$$\sum_i \partial_\mu(b_{ki}^\mu E_i^L) = \partial_\mu\, j_k^\mu \tag{5}$$

$$\sum_i (b_{ki}^\mu E_i^L) = j_k^\mu - \sum_i \left[\partial_\nu\left(\frac{\partial L}{\partial(\partial_\nu\phi_i)} b_{ki}^\mu\right)\right] \tag{6}$$

① 见[Brading and Brown，2007]。

② 在广义协变情形下对这一问题的讨论见[Brading and Brown，2002]。

③ 边界定理也出现在赫尔曼·外尔的工作中，在那里具体化到他的统一场理论中，见[Weyl，1922，pp. 287—289]。第一次出现是在 1919 年第三版中，并由内山菱友(Utiyama)[1956；1959]以一种非理论规定的形式出版。

④ 当和诺特的第二定理合并，边界定理是尝试从那些局域对称性是“完全的数学产物”的理论中区分出“真正规范的理论”的一个有效工具，这一区分是通过检查由定理产生的等式，并通过这些等式的物理意义或其他方面而做出的。

⑤ 关于边界定理的更多细节，包括允许表面项的一般化，见[Brading and Brown，2007]。

⑥ 这里陈述的是边界定理的约束形式，以使得诺特对称群必须属于与一个不变作用量相关群的约束类。

$$\left(\frac{\partial L}{\partial(\partial_\mu\phi_i)}b_{ki}^{\nu}\right)+\left(\frac{\partial L}{\partial(\partial_\nu\phi_i)}b_{ki}^{\mu}\right)=0 \tag{7}$$

同样有，b_{ki}^{ν}依赖于所考虑场的特定变换，在这里我们不需要关心其细节。第一个方程是和与局域对称性相关的超势的存在相联系的。① 第二个方程可以用于研究场与其源的关系，例如，在经典电磁学中，我们可以研究理论的局域规范对称性与下述条件之间的关系：

$$j^{\mu}=\partial_{\nu}F^{\mu\nu} \tag{8}$$

即具有动力学源的麦克斯韦(Maxwell)方程。再次以经典电磁学的情形为例，第三个方程是这样的条件，即电磁张量是反对称的(表明了此条件与该理论的局域规范对称性之间的关系)：

$$F^{\mu\nu}+F^{\nu\mu}=0 \tag{9}$$

这些评论已经经过了必要的简化，详细的推导和对这些结果的讨论读者可以参考[Barbashov and Nesterenko，1983]与[Brading and Brown，2003；2007]。边界定理的等式与诺特两个定理的等式并不是完全相互独立的，哪一个最有用取决于所涉及的语境和问题。正如诺特定理一样，边界定理的成立独立于动力学方程的具体细节，它们都使我们能够研究与那些理论对称性相关的理论结构特征。

8. 经典物理学中对称性的解释

接下来的论述我们将从“维格纳层级”(Wigner's hierarchy)开始，它已经成为关于对称性、定律与事件之间关系的权威观点。在这之后，我们简要讨论对称性与无关性之间的联系，以及这是如何影响对上文4.2节中描述的各种对称性的解释的。

对物理学理论中对称性的一般解释能够采取许多互补性的方法。我们可以探讨各种对称性所扮演的不同角色，各种对称性具有的认识论、本体论或者其他方面的地位，以及对称变换下结构不变的意义。最后，我们分别对这些问题

① 例如见[Trautman(陶特曼)，1962，179]。

给予评论。

8.1 维格纳层级

当代关于物理学中对称性地位及其意义的哲学讨论开始于维格纳 1949 年的论文“物理学理论中的不变性”，还有他后来于 1964 年发表的三篇论文。[①] 在这些论文中，维格纳做出了上面提到的在几何对称性与动力学对称性之间的区分，下面我们还要讨论到这种区分。他还提出了他关于物理学知识的层级观点，按照这种观点对称性被看作是定律的特性：

> 我们关于周围世界的知识中存在着一奇特的层级。每一个瞬间都带来令人感到惊奇和不可预见的事情——未来真的不确定。尽管如此，我们周围的事件确有一个结构，即在我们认知的事件之间存在着关联。科学希望发现的正是这种结构和这些关联，或至少是精确和严格定义的关联……我们知道许多自然定律，我们希望并期待着发现更多，没人能预见到下一个将要发现的定律。尽管如此，在自然定律中存在一种我们称之为不变性定律的结构。在有些情形下这种结构的影响是如此的深远，以至于能够在假设自然定律满足不变性结构的基础上猜出它们……这种从事件到自然定律、从自然定律到对称性或不变性原理的递进，正是我想说的我们周围世界知识的层级[Wigner, 1967, pp. 28—30]。

这种关于对称性是定律特性观点的，已成为权威观点。

8.2 对称性与无关性

存在与对称性相关联的关于定律的或潜在事件的一种普遍特性：某些量的

---

① 维格纳的论文收录于论文集《对称与反射》(*Symmetries and Reflections*) [Wigner, 1967]中。

无关性，否则的话这些量可能会被认为具有物理意义。① 在4.2节中我们概述了在物理学中对称性的多样性，在每一种情形下对称性都与一种特性相关，这种特性被认为与描述系统受定律支配行为的目的无关。例如，左右对称性意味着一个系统是左手式还是右手式与它受定律支配的演化无关。著名的案例就是，这种对称性在弱相互作用中被打破，对某些过程而言，系统受定律支配的行为被证明是手性敏感的，参见[Pooley(普利)，2003]。

在4.2节中，我们从数学上描述了全域对称性与局域对称性间的区分，即它们分别依赖于常数参量和任意时间(或空间)的函数。这种区分的物理意义可以通过被认定为是无关性的这种相关特性来理解。一个全域对称性反映了一特定量绝对值的无关性：只有相对值是有意义的。例如，在牛顿力学中，空间平移不变性成立且绝对位置与系统的行为无关。② 只有相对位置是重要的，这一点通过方程的全域空间平移不变性反映在理论结构中——方程并不依赖于或诉诸于绝对位置的背景结构。

全域对称性是局域对称性的一种特殊情形。一个局域对称性不仅反映了绝对值的无关性，而且还反映了一定距离外确定的相对值的无关性：仅有局域的相对值(即在某一点处确定的相对值)是相关的。这一点通过运动方程的特性反映在理论结构中，也就是运动方程不依赖于确定一定距离外相对值的背景结构(即不存在与所考虑特性相关联的全域背景结构)。③

8.3 对称性的作用

随着量子理论的出现，对称性在物理学中所发挥的许多不同作用变得越发

① 关于对称性、等效性和无关性之间联系的分析，见[Castellani，2003]。

② 这里我们考虑的是不包含牛顿绝对空间的牛顿力学。

③ 取而代之的是，我们要求理论中明确出现一种联系，该联系提供了将两个远距离的客体置于一起从而能够在它们之间进行局域性对比的规则。

明显。[①] 不过，通过利用晶体明显不同的对称特性对晶体的分类，我们早就看到了其强有力的分类作用。事实上，正是阿羽依(René-Just Haüy)对于对称性的这种应用方式，使得结晶学直到1801年才作为一种不同于矿物学的学科而出现。[②] 此外，在狭义相对论和广义相对论二者(见上文第5节)的构建中，相对性原理的启发性的和/或规范性的作用很明显。如今在统一基本作用力的努力中非常突出的统一作用，早就出现(虽然在方法论上有些不同)在希尔伯特对引力与电磁学广义协变理论的构建中，参见[Sauer，1999]，也出现在1918年外尔关于引力与电磁学的统一理论中。对称性也可发挥各种解释作用。例如，在诺特第一定理(见第7节)的基础上，我们可以说，正是因为经典力学的平移对称性(加上对其他条件的满足)，理论中的线性动量才是守恒的。另一个例子则是通过维格纳的层级，把对称性原理作为关于下述二者的解释：(1)定律的形式方面；(2)为什么某些事件发生而其他事件不发生。

8.4　对称性的地位

对称性的地位是本体论的、认识论的还是方法论的？正如上文(第5节)在相对论情境中所讨论的，很明显对称性具有重要的启发性功能。这表明了一种方法论地位，其在量子理论的语境中得到了进一步的发展。我们也可以问是否应该赋予对称性一种本体论或认识论的地位。

根据本体论的观点，对称性被看作是“自然中存在的”，或刻画了物理世界的结构。认为对称性是自然属性的一个理由是所谓的时空对称性的几何解释。

---

① 19世纪20年代，应用群论及其表征来开发量子力学中的对称性，对理论基础和理论的现象学解释而言，都标志着对称性在物理学中重要性的一戏剧性变革。正如维格纳在许多场合中所强调的，“在量子理论中不变性原理日益增长的有效性”[Wigner，1967，47]的一个本质原因是量子物理系统态空间的线性性质——对应于量子态叠加的可能性。对于量子物理学中对称性应用的细节，我们推荐读者参考迪克森、兰兹曼以及霍尔沃森在本书中的论述。哲学的讨论见[Brading and Castellani，2003]。

② 在赫塞尔(J. F. Hessel)和布拉维(A. Bravais)的研究中，对离散对称性在结晶学中的应用持续于19世纪，得到了32点变换晶体类和14种布拉维晶格。这些类和晶格由费德罗夫(E. S. Fedorov)，申夫利斯(A. Schönflies)和巴洛(W. Barlow)在18世纪90年代并入到230种空间群中。离散群的理论在如固态物理学、化学和材料科学等领域仍然很重要。

按照该解释，物理定律的时空对称性被解释为是时空本身的对称性，是物理世界的“几何结构”。此外，通过把它们看作是其他类型空间——通常熟知的“内部空间”——的特性，这种看待对称性的方式可以扩展到非外部的对称性。① 基于这样一个内部空间的观点，一个实在论者会是什么立场仍然是个开放的问题，也是一个值得讨论的有意义的主题。

一种研究对称性的本体论立场局限性的途径是研究它们的经验地位或观测地位：所考虑的对称性能否直接被观测到？我们首先必须回答，一个可观测的对称性意味着什么，是否所有的对称性都具有相同的观测地位。科索[2000]得出这样的结论：不同种类的对称性有着显著不同的经验地位。特别是，尽管全域连续对称性能够直接被观测到——通过伽利略船之类的实验——而局域连续对称性却只能有间接的观测证据。②

人们熟悉的全域时空对称性的直接观测地位把我们引向对称性的认识论方面。按照维格纳的观点，时空不变原理扮演着发现自然定律可能性的先决条件：“若事件之间的关联天天在变，在空间每一点上都不一样，就不可能发现它们”[Wigner，1967]。对维格纳而言，对称性原理的这一概念本质上与我们的无知相关(若我们能够直接知道自然定律，我们就不需要在寻找这些定律时利用对称性原理了)。可以为该观点给出方法论的解释，按照该解释，预设这样的时空规则是为了使发现物理学定律这一事业得以实施。③ 另外一些人得到的观点是，对称性原理是作为康德意义上的“先验原理”起作用的，参见[Mainzer(迈因策尔)，1996]。需要注意的是，维格纳的出发点(如之前引文所述)并不意味着精确对称性——认识论上需要的只是对于适当的时空区域而言全域对称性近似成立的，这样在发现自然定律的事件中才会有充分的稳定性与规则性。

正如本讨论与前一节的讨论所表明的，在我们还未非常深入到解释性的问

① 对称性的分类见上文4.2节。

② 见上文6.1节，还有布拉丁和布朗[2003b]，他们论证了一种关于科索例子的不同解释。

③ 我们感谢布兰登·福格尔(Brandon Fogel)指出这一点，还有他提出的在时空对称性的这一观点和爱因斯坦可分离性概念的方法论特征的对比。

题时，各种对称性之间的差别十分重要。正因为此，近期许多关于物理学理论中对称性解释的工作并没有如上述所集中于一般性问题，而是关注于特定对称性的解释性问题。①

8.5 对称性，客观性和客体

现在转向在对称变换下保持不变的结构问题上来。客观性是什么的问题不应该依赖于它被考虑的特定视角这一古老而又自然的观念，而应该用如下群论术语来重新表示：客观性就是指在相关变换群下的不变性。这种在对称性与客观性之间的联系有着很长的历史，至少可以追溯到20世纪早期。外尔[1952]强调了这一点，他写道："我们发现客观性意味着相对于自同构群的不变性。"在客观性与不变性之间的联系主要是在相对论理论(包括狭义相对论和广义相对论)语境中讨论的。我们回顾闵可夫斯基(Minkowski)的名言[1908，75]，"从今以后空间本身与时间本身，都将注定消退为阴影而已，只有二者的统一体才是一种独立的实在"，此后他完成了对爱因斯坦狭义相对论的几何化并把时空间隔(而非空间间隔和时间间隔)确认为几何不变量。参与讨论广义相对论中客观性与不变性之间联系的还有希尔伯特和外尔，以及其他科学家，该问题至今仍在讨论。②

与这一问题相关的是用对称性来将物理客体刻画为不变量的集合。最初在量子理论语境下提出的这一方法，也可以应用于经典物理学。③ 它的基本思想是，不变量——如质量和电荷——是我们用以描述客体的量。因此，通过群论的应用，我们可以用对称性考虑来确定不变的量，并把客体"建构"或"组合"为这些不变量的集合。④

---

① 这些包括经典电磁学和量子理论中发现的各种规范不变性，除了GTR中的广义协变性(它们是连续的对称性)之外，还有宇称的离散对称性(在弱相互作用不满足)和置换不变性，后二者是在经典理论中发现的，但仍然需要在量子理论的视角下来重新思考。见[Brading and Castellani，2003]。

② 从上文中(6.2和6.3节)对爱因斯坦"洞问题"和GTR中可观测量的地位的讨论中我们看到了这一争论的某些方面。

③ 见马克思·玻恩(Max Born)，重印于[Castellani，1998]。

④ 更多的讨论见[Castellani，1988，Part Ⅱ]。

综上所述，与经典物理学中对称性相关联的哲学问题是非常广泛的。我们在这里提供的只不过是一个被我们自己的兴趣与困惑所影响的概览，我们希望它能有助于对这一在哲学与物理学上丰富的领域进行更深入地探索。

致谢

我们非常感谢主编杰里米·巴特菲尔德和约翰·厄尔曼给予的鼓励和具体评论。我们也同样感谢布兰登·福格尔、布莱恩·皮兹以及托马斯· 里克曼的评论和建议。

参考文献

[Adler *et al.*, 1975] R. Adler, M. Bazin, and M. Schiffer. *Introduction to General Relativity.* 2nd edition, McGraw-Hill Book Company, New York, St Louis, San Francisco, Toronto, London, Sydney, 1975.

[Anderson, 1967] J. L. Anderson. *Principles of Relativity Physics.* New York and London: Academic press, 1967.

[Barbashov and Nesterenko, 1983] B. M. Barbashov and V. V. Nesterenko. Continuous symmetries in field theory, *Fortschr. Phys.*, 31: 535-67, 1983.

[Barbour, 1989] J. B. Barbour. *Absolute or Relative Motion*? Vol. 1 *The Discovery of Dynamics.* Oxford University Press, 1989.

[Barbour, 2005] J. B. Barbour. *Absolute or Relative Motion*? Vol. 2 *The Deep Structure of General Relativity.* Oxford University Press, 2005.

[Belot, 1998] G. Belot. Understanding electromagnetism. *British Journal for the Philosophy of Science*, 49: 531-555, 1998.

[Belot, 2006] G. Belot. The representation of time and change in mechanics. This volume.

[Brading, 2005] K. Brading. A note on general relativity, energy conservation, and Noether's theorems. In J. Eisenstaedt and A. J. Kox, Birkhuser (eds.) *The Universe of General Relativity* (*Einstein Studies*), 2005.

[Brading and Brown, 2002] K. Brading and H. R. Brown. General covariance from the perspective of Noether's theorems. *Diálogos*: 59-86, 2002.

[Brading and Brown, 2003] K. Brading and H. R. Brown. Symmetries and Noether's theorems. In K. Brading and E. Castellani (eds.), pages 89-109, 2003.

[Brading and Brown, 2004] K. Brading and H. R. Brown. Are gauge symmetry transformations observable? *British Journal for the Philosophy of Science*, 55: 645-65, 2004.

[Brading and Brown, 2007] K. Brading and H. R. Brown. Noether's theorems, gauge symmetries and general relativity. Manuscript, 2007.

[Brading and Castellani, 2003] K. Brading and E. Castellani (eds.). *Symmetry in Physics: Philosophical Reflections.* Cambridge University Press, 2003.

[Brading and Castellani, 2006] K. Brading and E. Castellani. Curie's Principle, Encore. In preparation, 2006.

[Brading and Ryckman, 2007] K. Brading and T. A. Ryckman. Hilbert's axiomatic method and the foundations of physics: an interpretation of generally covariant physics and a revision of Kant's epistemology. Manuscript, 2007.

[Brown, 2006] H. Brown. *Physical Relativity: Spacetime Structure from a Dynamical Perspective.* Oxford University Press, 2006.

[Butterfield, 2004] J. Butterfield. Between laws and models: Some philosophical morals of Lagrangian mecha-nics; available at Los Alamos arXive: http://arxiv.org/abs/physics/0409030; and at Pittsburgh archive: http://philsci-archive.pitt.edu/archive/00001937/.

[Butterfield, 2006] J. Butterfield. On symmetry and conserved quantities in classical mechanics. Forthcoming in W. Demopoulos and I. Pitowsky (eds.), *Festschrift for Jeffrey Bub.* Kluwer: University of Western Ontario Series in Philosophy of Science, 2006; available at Los Alamos arXive: http://arxiv.org/abs/physics/; and at Pittsburgh archive: http://philsciarchive.pitt.edu/archive/00002362/.

[Butterfield, 2006] J. Butterfield. On symplectic reduction in classical mechanics. This volume.

[Castellani, 1998] E. Castellani (ed.). *Interpreting Bodies. Classical and Quantum Objects in Modern Physics.* Princeton University Press, 1998.

[Castellani, 2003] E. Castellani. Symmetry and equivalence. In K. Brading and E. Castellani (eds.), pages 321-334, 2003.

[Chalmers, 1970] A. F. Chalmers. Curie's principle. *The British Journal for the Philosophy* *of Science*, 21: 133-148, 1970.

[Corry, 2004] L. Corry. *David Hilbert and the Axiomatization of Physics (1898-1918)*. Dordrecht: Kluwer Academic, 2004.

[Curie, 1894] P. Curie. Sur la symétrie dans les phénomènes physiques. Symétrie d'un champ électrique et d'un champ magnétique. *Journal de Physique*, 3ᵉ série, vol. 3: 393-417, 1894.

[Curie, 1981] P. Curie. On symmetry in physical phenomena. Trans. J. Rosen and P. Copié, *Am. J. Phys.*, 49(4): 17-25, 1981.

[Descartes, 1998] R. Descartes. *The World and other Writings*. S. Gaukroger (ed. and trans.), Cambridge University Press, 1998.

[Dickson, 2006] M. Dickson. Non-relativistic quantum theory. This volume, 2006.

[Doughty, 1990] N. Doughty. *Lagrangian Interaction*. Addison-Wesley, 1990.

[Earman, 2004] J. Earman. Curie's principle and spontaneous symmetry breaking. *International Studies in Philosophy of Science*, 18: 173-199, 2004.

[Earman, 2006a] J. Earman. Two challenges to the substantive requirement of general covariance. *Synthese*, in press, 2006.

[Earman, 2006b] J. Earman. Aspects of determinism in modern physics. This volume.

[Halvorson, 2006] H. Halvorson. Algebraic quantum field theory. This volume.

[Hawkins, 2000] T. Hawkins. *Emergence of the Theory of Lie Groups: an Essay in the History of Mathematics* 1869-1926, New York: Springer, 2000.

[Hilbert, 1915] D. Hilbert. Die Grundlagen der Physik (Erste Mitteilung), Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen. *Mathematisch-physikalische Klasse*: 395-407, 1915.

['t Hooft, 2006] 't Hooft. The conceptual basis of quantum field theory. This volume.

[Howard, 2007] D. Howard. 'And I shall not mingle conjectures and certainties': The roots of Einstein's principle theories/constructive theories distinction, manuscript, 2007.

[Ismael, 1997] J. Ismael. Curie's principle. *Synthese*, 110: 167-190, 1997.

[Janssen, 2005] M. Janssen. Of pots and holes: Einstein's bumpy road to general relativity. *Ann. Phys.*, 14, Supplement: 58-85, 2005.

[Julia and Silva, 1998] B. Julia and S. Silva. Current and superpotentials in classical gauge invariant theories I. Local results with applications to perfect fluids and general relativity, 1998. gr-qc/9804029v2.

[Klein, 1918] F. Klein. Über die Differentialgesetze für die Erhaltung von Impuls und Energie in der Einsteinschen Gravitationstheorie. Königliche Gesellschaft der Wissenschaften zu Göttingen. Mathematisch-physikalische Klasse. *Nachrichten*: 469-492, 1918.

[Kosso, 2000] P. Kosso. The empirical status of symmetries in physics. *British Journal for the Philosophy of Science*, 51: 81-98, 2000.

[Lanczos, 1949] C. Lanczos. *The Variational Principles of Mechanics.* University of Toronto Press, [1949], 1962.

[Landsman, 2006] N. P. Landsman. Between classical and quantum. This volume.

[Mainzer, 1996] K. Mainzer. *Symmetries of Nature. A Handbook.* New York: de Gruyter, 1996.

[Malament, 2006] D. Malament. Classical relativity theory. This volume.

[Miller, 1981] A. I. Miller. *Albert Einstein's Special Theory of Relativity.* Addison-Wesley Publishing Company, 1981.

[Minkowski, 1908] H. Minkowski. Space and time. In *The Principle of Relativity. A Collection of Original Memoirs on the Special and General Theory of Relativity.* New York: Dover, pages 75-91, [1908], 1923.

[Misner *et al.*, 1970] C. W. Misner, K. S. Thorne, and J. A. Wheeler. *Gravitation.* New York: W. H. Freeman and Company, 1970.

[Noether, 1971] E. Noether. Noether's theorem. *Transport Theory and Statistical Physics*, 1: 183-207, 1971.

[Norton, 1984] J. Norton. How Einstein found his field equations: 1912-1915. *Hist. Stud. Phys. Sci.*, 14: 253-316, 1984.

[Norton, 1993] J. Norton. General covariance and the foundations of general relativity: eight decades of dispute. *Rep. Prog. Phys.*, 56: 791-858, 1993.

[Norton, 2003] J. Norton. General covariance, gauge theories, and the Kretschmann objection. In K. Brading and E. Castellani (eds.), pages 110-123, 2003.

[Ohanian and Ruffini, 1994] H. C. Ohanian and R. Ruffini. *Gravitation and Spacetime*, 2nd edition. London, New York: W. W. Norton and company, 1994.

[Olver, 1993] P. J. Olver. *Applications of Lie Groups to Differential Equations*, 2nd edition, New York: Springer, 1993.

[O'Raifeartaigh, 1997] L. O'Raifeartaigh. *The Dawning of Gauge Theory*. Princeton University Press, 1997.

[Pitts, 2006] J. B. Pitts. Absolute objects and counterexamples: Jones-Geroch Dust, Torretti constant curvature, Tetrad-Spinor, and scalar density. *Studies in History and Philosophy of Modern Physics*, forthcoming, 2006. Manuscript available at Los Alamos arXive: http://arxiv.org/abs/gr-qc/0506102/.

[Pooley, 2003] O. Pooley. Handedness, parity violation, and the reality of space. In K. Brading and E. Castellani (eds.), pages 250-280, 2003.

[Radicati, 1987] L. A. Radicati. Remarks on the early developments of the notion of symmetry breaking. In M. G. Doncel, A. Hermann, L. Michel, and A. Pais (eds.), *Symmetries in physics (1600-1980)*. Barcelona: Servei de Publicaciones, pages 195-206, 1987.

[Rovelli, 2006] C. Rovelli. Quantum gravity. This volume.

[Ryckman, 2005] T. Ryckman. *The Reign of Relativity*. Oxford University Press, 2005.

[Sauer, 1999] T. Sauer. The relativity of discovery: Hilbert's first note on the foundations of physics. *Arch. Hist. Exact Sci.*, 53: 529-575.

[Shubnikow and Koptsik, 1974] A. V. Shubnikow and V. A. Koptsik. *Symmetry in Science and Art*. London: Plenum Press, 1974.

[Stachel, 1993] J. Stachel. The meaning of general covariance: the hole story. In J. Earman*et al.* (eds.), *Philosophical Problems of the Internal and External World: Essays on the Philosophy of Adolf Grünbaum*. University of Pittsburgh Press, pages 129-160, 1993.

[Torretti, 1983] R. Torretti. *Relativity and Geometry*. Oxford: Pergamon Press, 1983.

[Trautman, 1962] A. J. Trautman. Conservation laws in general relativity. In L. Witten (ed.), *Gravitation: An Introduction to Current Research*. New York: John Wiley and Sons, pages 169-98, 1962.

[Utiyama, 1956] R. Utiyama. Invariant theoretical interpretation of interaction. *Physical Review*, 101: 1597-607, 1956.

[Utiyama, 1959] R. Utiyama. Theory of invariant variation in the generalized canonical dynamics. *Prog. Theor. Phys. Suppl.*, 9: 19-44, 1959.

[Wald, 1984] R. Wald. *General Relativity.* University of Chicago Press, 1984.

[Weyl, 1922] H. Weyl. *Space, Time, Matter.* Reprinted by Dover (1952), 1922.

[Weyl, 1952] H. Weyl. *Symmetry.* Princeton University Press, 1952.

[Whittaker, 1904] E. T. Whittaker. *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies.* Cambridge University Press, ([1904] 1989).

[Wigner, 1967] E. P. Wigner. *Symmetries and Reflections.* Bloomington: Indiana University Press, 1967.

[Yaglom, 1988] I. M. Yaglom. *Felix Klein and Sophus Lie. Evolution of the Idea of Symmetry in the Nineteenth Century.* Boston-Basel: Birkhäuser, 1988.

第十四章　现代物理学中的决定论

约翰·厄尔曼

1. 引言

这一章的目的是回顾决定论的几个方面，有些物理学家对它们很熟悉，但是在哲学文献中却很少讨论。并且我要表明这些方面是如何把决定论与物理学中的对称性问题、时空的结构和本体论地位问题、可预言性和可计算性问题关联起来的。①我们将会看到，决定论在某些方面是一个很强的信条，很难消除；但是在另一些方面，它却很脆弱，需要各种各样的假设来赋予其通过努力获得成功的机会。一般的非相对论量子力学是否承认可行的决定论基础仍然是一个有争议的话题。鲜为人知的是，在一些情况下，量子力学事实上比它的经典对应部分更加遵从决定论。量子场论为了构造量子可观测量的场代数，至少在其经典水平上假定了决定论。决定论在经典广义相对论最重要的未决问题，即宇宙监督的猜想中处于核心部分，并且决

① 最近对决定论问题的研究可见巴特菲尔德(Butterfield)［1998］、厄尔曼(Earmann)［2004a］以及霍夫(Hofer)［2003］。关于各种决定论方面的论文集，见安特曼斯潘彻和毕肖普(Atmanspacher and Bishop)［2002］。

定论的本质和地位问题在量子引力理论研究的关键问题中也处于核心位置。

2. 引论

2.1　决定论的形而上学

在本节一开始，我们首先领会前广义相对论和前量子力学所理解的决定论学说，然后试图去理解这些信条应该如何调整才能适应广义相对论和量子力学的这些理论。在前广义相对论中，时空作为静止的背景存在，物理学的剧情在此基础上上演。前量子力学的物理学中，同样假定存在一个真实物理量(也叫作“可观测量”)的集合$\mathcal{O}$，在每一个时刻，它们都有一个确定的取值，把这些称之为发生量值。其他物理量在性质上可能是有倾向性的，可能仅仅在适当的环境下才会呈现确定的值。但是大家普遍认为，这些有倾向性的量是附随于无倾向性量的。[①] 历史 H 是从$\mathbb{R}$ 到基本量值的元组的映射，这里，对任何 $t \in \mathbb{R}$，态 $H(t)$ 给出基本量在时刻 t 的行为的一个快照。世界只有在任何一对历史 H_1、H_2 满足以下物理学规律时才相对于$\mathcal{O}$是拉普拉斯决定论(Laplacian deterministic)的：如果在某个时刻 t 有 $H_1(t)=H_2(t)$，那么在所有 t 上，有 $H_1(t)=H_2(t)$。

这里给出一系列的注解。首先，上述定义中出现的“t”应当是一个全域时间函数。这个概念在某种程度上可以应用于经典时空、狭义相对论时空和广义相对论时空：全域时间函数是一个平滑映射 t：$\mathcal{M} \to \mathbb{R}$，这里$\mathcal{M}$是时空流形，对任何 p，$q \in \mathcal{M}$，只有在从 p 到 q 存在一个指向未来的类时曲线时，才有 $t(p)<t(q)$。[②] 在经典时空中，绝对(或者独立于观测者的)同时性概念的全部

---

① 附随性的大概意思是，当且仅当在可能的例子范围之内，X 附随于 Y，如果 Y 没有变化，那么 X 就没有变化。附随性的强度和类型取决于什么被看作是可能的例子。这里主要关注的是物理附随性，其中可能的例子是指那些与物理原理相容的例子。

② 这个定义假定时空是时间上可定向的，并且其中一个方向已经被挑出来作为时间的未来方向。第一个假定满足经典和狭义相对论时空。广义相对论时空可能不是时间可定向的，参见马拉蒙特(Malament)在本书中的论述，但是一个覆盖的时空总是可定向的，因为时间可定向性只有在时空不是单连通的情况下才不会成立。第二个假定暗示着已经发现了时间方向问题的一些解决办法，参见尤菲克(Uffink)在本书中的论述。

内容，一条类时曲线，是倾斜于绝对同时性平面的。而 $t=$ 常数的全域时间函数必须与同时性平面一致，因此，在经典背景下 t 由 $t \to t' = t'(t)$ 这个变换决定。在相对论背景下，类时曲线的任意一点的切线都在那一点的光锥之内。在因果关系上有问题的广义相对论时空中，即哥德尔(Gödel)时空，见 6.1 节，不存在全域时间方程，并且如上所定义的全域意义上的拉普拉斯决定论是无意义的。[①] 但是如果对于某个相对论时空来讲存在一个全域时间函数，那也就存在多个这样的函数。对于全域时间函数的不适当选择会导致前面定义的拉普拉斯决定论的失败。因此我们分析，在相对论背景中，决定论定义的应用要求一个适当选择的时间函数，其本质将在下面进行阐释。

第二，上述决定论的形式一方面假定了自然定律之间的区别，另一方面假定了初始/边界条件。当这种区别变得模糊时，决定论的教义也开始模糊起来。关于自然定律有着大量的哲学文献。[②] 由于在谈到理解物理学实践中定律的本质及其作用时，其大部分是没有启发意义的，在这里很大程度上就将其忽略了。为了当前的目标，我只简单地规定，一个定律可接受的解释必须受此定律附生于非模态的、特别的事实的整体之上的经验主义约束。[③] 哲学家喜欢思索非经验主义的定律，但是这样的实体是否应当存在似乎超出了科学的范围，因此它们与当前的目标不相关。我比较喜欢大卫·刘易斯(David Lewis)[1973, pp. 72—77]的经验主义约束方式，因为它把对定律的解释与物理学的实践联系了起来，物理学定律是出现于物理学理想理论中的公理或者假设，这里，理想理论是在真实理论的类别内部在简单性和信息量之间获得了最好的平衡的理论。我们拥有的历史上所有关于决定性定律的备选例子都包含了关于基本当前

---

① 相对论时空全域时间函数存在的一个充分必要条件是稳定的因果性，(大体上讲)它要求存在不会产生封闭类时曲线的零锥扩展。其精确定义参见[Wald(瓦尔德)，1984, pp. 198—199]。哥德尔时空不仅不允许全域时间函数，它也不允许任何全域时空片段(即无边的类空超曲面)。

② 对自然定律的不同说明的概述，见[Carroll(卡罗尔)，2004]。

③ 这就是大卫·刘易斯关于自然定律所称之为“休谟附随性(Humean Supervenience)”的东西，对其的证明见[Earman and Roberts(厄尔曼和罗伯茨)，2006]。

量的相对小的子集$\mathcal{B} \subset \mathcal{O}$，其假设是其余的量都附随于$\mathcal{B}$。[①] 如果如上所述，简单性是物理学定律的一个重要特征，这并不令人惊奇。赫尔曼·外尔（Hermann Weyl）[1932，42]坚信简单性必定被列入对定律的说明，但是他注意到，“由于它使得自然定律的意义要依赖于简单函数和复杂函数之间或者简单函数类和复杂函数类之间不断波动的区别，它往往会削弱决定论形而上学的力量”。我认为，这是一个必须被咽下并消化的结果。只对他们的扶手椅负责的哲学家们则可以随意持其他观点。

第三，这个世界将会被部分地确定，这一点具有概念上的可能性。也就是说，对于由当前物理量的一些适当子集$\mathcal{D} \subset \mathcal{O}$中的量值定义的部分历史是确定的，但是对于由另外一些适当子集$\mathcal{N} \subset \mathcal{O}$中的量值所定义的部分历史则是非确定的。但是很难想象这样一种状况，即如果$\mathcal{D}$和$\mathcal{N}$都是基本量，上述情况还会发生。因为，为了使元素$\mathcal{N}$的非决定性演化不扰乱$\mathcal{D}$的决定性演化，$\mathcal{N}$中的量绝不能与$\mathcal{D}$中的量发生相互作用，否则就会出现这样一种诡异现象：$\mathcal{N}$量对$\mathcal{D}$量的影响消失了，操作上也一样。然而，当$\mathcal{N}$不是基本量的时候，这种貌似合理的考虑就失效了。特别地，正如下面要讨论的，某个层次上的随机过程可以附随于另一个更低层次上的决定性过程，参见[Butterfield，1998]。这个事实保证了从观测所得的随机行为到非决定论的推断不出问题。

第四，物理学定律往往遵从偏微分方程的形式，在这种情况下，拉普拉斯决定论的问题就转为方程是否承认初值函数（initial value formulation），也就是说，对于任意一个初始值，是否存在一个唯一的解与之相对应的问题。[②] 什么可以算作初始值要依赖于细节，但是通常它与这些变量的有限个时间导数的瞬时值一起组成了方程中独立变量的瞬时值。“任意的”初始值可能被看作是包含了相关变量在任何运动学上可能的值——就如牛顿力学中粒子的位置和速度的初始值——但是“任意”必须被看作是意味着在与运动方程相容的界限中的任意，这可能对初始值加上了重要的约束。这就引出了下一节的评论。

---

① 例如，在经典粒子力学中，$\mathcal{B}$的元素是粒子的位置和动量，并且假设了任何其他真实的力学量可以表示为这些基本量的函数。

② 正如下面将要讨论的，还存在着更深层次的问题，比如解是否连续地取决于初始数据。

第五，在相对论情形中，场方程常常转化为约束方程，在初始数据和演化方程中置入了约束，这就控制着初始数据在时间演化中如何满足约束方程——麦克斯韦(Maxwell)的电磁学方程和爱因斯坦(Einstein)的引力场方程就是基本的例子。在这些例子中，演化方程确保一旦约束方程得到满足，它们就能够在时间演化中继续得到满足。这应当是确定性方程的一个特征，因为，如果在未来的时间中，唯一解中被初始数据选择的数据不再满足约束的话，那么定律就自行失效了。可以说，在相对论世界中时间的一个基本特征——可能是使时间维度和空间维度分离的关键特征——就精确地处于演化和约束方程的这种分离中。①

第六，只要不存在先天的保证使物理学理想理论的定律为确定的，物理学的历史就表明决定论被当作是“可废止的方法论规则”：假设决定论为真；如果目前发现的备选定律不是决定性的，那么假定存在其他还能够发现的定律，或者说目前发现的定律只是近似于正确的定律；我们只有在长时间和重复的失败中才能接受以下假设，不能找到决定性定律并不代表着缺乏想象或者不够努力，而是反映了自然是非决定论的这一事实。在量子力学奠基人之一马克斯·普朗克(Max Planck)的工作中可以找到这个观点的表述：决定论(也叫作因果定律)，是一个“启发性的原则，一个标志，我认为它是我们拥有的最重要的标志，在杂乱无序的实践中引导我们，指引着科学探索为了得到丰富成果所要坚持的正确方向”，见[Planck，1932，26]。②

2.2 多种多样的决定论

在哲学文献中存在着一种关注拉普拉斯决定论的倾向，但物理学中还存在其他种类的决定论。比如说，有些过程由延迟微分方程描述，而对这些方程来说瞬时初始数据可能并不足以得到唯一解。举一个简单的例子就是一阶常微分方程$\dot{x}(t)=x(t-C)$，其中延迟常数$C>0$。拉普拉斯决定论失败了，因为给定的初始数据$x(0)$与多个解相容。但拉普拉斯决定论的近亲(相似观点)仍然有

① 关于空间和时间的差异的相关观点的证明，参见[Callender(坎林德)，2005]和[Skow(斯科)，2005]。

② 关于量子力学奠基人对决定论地位争论的历史，参见[Cushing(库欣)，1994]和[Stöltzner(斯托尔兹纳)，2003]。

效，因为对于时间间隔 $t \in [-C, 0]$，$x(t)$ 的确定会确立唯一解。[①] 如果延迟常数 C 被无界的 t 的一个函数 $\tau(t)$ 所代替，或者说如果延迟微分方程的形式比我们举的例子更加复杂，那么即便是弱拉普拉斯决定论也可能会失败。后者可以用一种运动方程来说明：$\dot{x}(t)=f(t)x(t-1)$，这里 $f(t)$ 是一个连续函数，超出[0，1]之外它就为零，并且满足 $\int f(t)\,dt=-1$。拉杜(Raju)[1994，120ff]给出了一个 f 函数的例子，除非对于所有的 $t \geqslant 1$ 有 $x(t)$ 全部为零，否则 $t<0$ 时运动方程将无解；而如果对于 $t \geqslant 1$ 有 $x(t)$ 全部为零的话，那么运动方程在 $t<0$ 时就会有无数解。在把延迟项 $x(t-1)$ 替换为超前(advance)项 $x(t+1)$ 的例子中，结果是所有的历史都不能决定唯一的未来解。对于延迟/超前微分方程最重要的物理应用，即带电粒子在自身延迟/超前相互作用下的运动的初值问题，人们仍了解得很少。[②]

为了明确起见，需要着眼于拉普拉斯的决定论变化。在这一类决定论变化之中，未来决定论和过去决定论之间也存在区别。过去拉普拉斯决定论指的是，对于任何一对满足物理学定律的历史 H_1，H_2，如果对于某个时刻 t，$H_1(t)=H_2(t)$，那么对于所有的 $t'>t$，$H_1(t')=H_2(t')$。未来拉普拉斯决定论有着类似的定义。原则上，拉普拉斯决定论在一个时间方向上有效的话在另一个方向上就无效。然而，如果运动的定律是时间反演不变的，那么未来和过去决定论就会同时有效或者失效。时间反演不变性是指，如果 H 是一个满足定律的历史，那么它的“时间反演”历史 H^T 就会同样满足定律，这里 $H^T(t):={}^RH(-t)$，“R”为反演算子，定义在一个具体事例分析的基础上，通常是类比经典粒子力学。在经典粒子力学中，$H(t)=(\mathbf{x}(t), \mathbf{p}(t))$，$\mathbf{x}(t)$ 和 $\mathbf{p}(t)$ 分别为粒子在 t 时刻的位置和动量，并且 ${}^RH(t)=(\mathbf{x}(t), \mathbf{p}(t))$。[③]

---

① 关于延迟微分方程的相关结果参见[Driver(德里弗)，1977]。

② 德里弗[1979]研究了同一带电粒子在半延迟和半超前相互作用约束下做关于 x 轴的对称运动的特殊情况。他指出，假定两个粒子在停止移动时分开得足够远，那么它们的位置以及它们停止运动的时间就能确定一个唯一解。

③ [Albert(阿尔伯特)，2000，Chapter 1] 给出了时间反演不变性的一种不同的说明；但还需要参看[Earman，2002]和[Malament，2004]。

由于物理学基本定律的所有备选项，除了那些基本粒子弱相互作用的定律，都是时间反演不变的，因此过去决定论和未来决定论之间的区别往往被忽略了。

这是暗示了在决定论和对称特性之间存在着有意义联系的第一条线索。[①]从下一节开始，还会遇到更多的例子。

2.3 决定论和对称性：居里原理

现在被叫作“居里原理”的表述是在1894年由皮埃尔·居里(Pierre Curie)提出的：

(CP)当某些效应表明了一种特定的非对称性的时候，这种非对称性一定能在引起它的原因中找到[Curie，1894，401]。

有些评论者在这个原理中看到了深刻的真理，但其他人只看到了错误，还有些人认为它没有什么意义，可对比[Chalmers(查尔默斯)，1970]、[Radicati(拉德坎提)，1987]、[Van Fraassen(范·弗拉森)，1991，23-24]以及[Ismael(伊斯梅尔)，1997]。我对居里原理的解读是把它当作了必要的真理，把居里原理看作是坚持了一种条件性。

如果：

(CP1)支配系统的运动定律是决定论的；

(CP2)支配系统的运动定律在对称变换下不变；

(CP3)系统的初始态在上述对称下不变；

那么：

(CP4)系统的末态在上述对称之下也是不变的。

当前提中的第一项(CP1)成立时，第二项(CP2)可作如下理解：如果初始态(initial state)经历了一段时间Δt的演化，然后上述对称运用于(唯一的)演化态(evolved state)，结果与首先把对称运用于初始态然后让产生的态演化一段时间Δt是一样的。在这种理解下，读者很容易从(CP1)—(CP3)得到(CP4)。

---

① 关于现代物理中的对称性和不变性的讨论，参看布拉丁和卡斯特拉尼(Brading and Castellani)在本书中的论述。

经典物理学与相对论物理学中居里原理的具体实例可以在[Earman, 2004b]中找到。在6.3节，会提到一个广义相对论的例子。[①]

虽然(CP)是一个必然真理，但它远不是平凡的，因为它有助于把对非对称的因果解释的探索引导至被视为系统末态的结果：要么是非对称性已经在初始态中表现出来；或者初始态是对称的，非对称性随着时间流逝慢慢出现；要么是由于支配系统演化的定律并不遵守对称性；或者由于它们是非决定性的。如果像通常的情况一样，排除了后两种可能性，那么末态中的非对称性就一定源于在初始态中的非对称性。也要注意，由于在科学实在论中永不休止的争论，居里原理的运用也有分歧，因为初始态中的非对称性可能不仅对于肉眼来说感觉不到，而且对于任何宏观探测方式来说都感觉不到。[②]

3. 经典物理学中的决定论和非决定论

3.1 经典物理学中决定论的艰难之路

经典物理学被广泛地认为是为决定论提供了一个友好的环境。事实上，决定论要在这个环境中达到成功还必须克服许多障碍。首先，经典时空结构可能并不丰富到足以支撑粒子运动的拉普拉斯决定论。其次，即便时空结构是丰富的，如果力的函数不满足合适的连续性条件的话，唯一性在牛顿方程的初始值问题中也可能失效。第三，通常在经典粒子和经典场中都出现或者只在经典场中出现的运动方程，并不承认初始值公式，除非加上一些补充条件。第四，即便是在局域(在时间上)唯一性对初始值问题有效的情况下，在有限的时间之后，解也可能会不成立。

下一小节开始这些论题中的第一个——决定论与经典时空的结构和本体论之间的联系。其他的将会在适当时候进行阐述。

---

① 关于居里原理的其他评论，参见布拉丁和卡斯特拉尼在本书中的论述。

② 关于居里原理及其与量子场论中自发对称破缺之间的联系的更为详细的讨论，参见[Earman, 2004b]；对于在量子统计物理学中的自发对称破缺，参见埃姆什(Emch)在本书中的论述。

3.2 决定论、时空结构以及时空本体论

这里给出以下理由认为除非有足够的正确时空结构作为支撑，否则拉普拉斯决定论和它的任何相似观点都不会成功。假定(静止的)经典时空背景由一个可微流形$\mathcal{M}$和$\mathcal{M}$上多种多样的几何对象场O_1，O_2，…，O_M刻画，并且假定物理学定律遵从方程的形式，其变量是O_i的，并且附加场P_1，P_2，…，P_N描述了时空的物理内容。(为具体起见，读者可能想要考虑P_i场为矢量场的情况，矢量场的积分曲线被当作是粒子的世界线。)时空的对称性是$\mathcal{M}$对自己的一个微分同胚d，它保留了由O_i场给定的背景结构——象征性地，对于所有的i，$d^* O_i = O_i$。这里d^*表示由d产生的拖曳。[1] 根据定律形式的假设，一个时空对称的d也必须是运动定律在以下意义上的对称：如果$\langle \mathcal{M}, O_1, O_2, \cdots, O_M, P_1, P_2, \cdots, P_N \rangle$满足运动定律，那么$\langle \mathcal{M}, O_1, O_2, \cdots, O_M, d^* P_1, d^* P_2, \cdots, d^* P_N \rangle$也满足。[2]

现在，背景时空的结构越贫乏，时空对称性的级别就越高。并且，如果时空对称群足够丰富，它就会包含这样的元素：时空在某些时间片断$t=$常数之上或者之下的部分是恒定映射，但是超越这个片断就是非恒定的。我们可以将这样一种映射叫作“扼杀决定论的对称性”，因为当应用到运动方程的任何解的时候，它都会产生另一个解，这个解对于过去的所有时刻都与第一个解相同，但是对于未来时间，它就会不同于第一个解。这对于哪怕最弱的拉普拉斯决定论也是一种违背。

举一个莱布尼茨时空(Leibnizian spacetime)[3]的例子。莱布尼茨时空的结构包含了以下所有内容，事实上也只包含这些内容：绝对的或者独立于观测的同

---

① 流形$\mathcal{M}$的一个微分同胚d是$\mathcal{M}$向自身的一个一一映射，它保留了$\mathcal{M}$的微分结构。为了具体一点，假设d是C^∞的。

② 因为假设定律(比方说)是与O_i和P_j相关的微分方程，它们就不可能对$\mathcal{M}$上O_i和P_j处取某些给定值的一点的“裸同一性”敏感。定律的这种微分同胚不变性是被称为实体广义协变性的那个东西的组分之一(参见6.2节)。有人可能想通过引入个体时空点的名字来打破微分同胚不变性，但这些名字的出现可能会违背定律应该有的“普遍性”特征。

③ 各种经典时空结构的详细讨论，见[Earman, 1989]。

时性的概念、时间度规(假定了时间在非同时事件之间的流逝)以及一个欧几里得(Euclid)空间度规(假定了事件之间的空间距离取决于绝对同时性的平面)。在坐标系(x^α, t); $\alpha=1,2,3$ 中，适应于这个结构，时空对称性是：

$$x^\alpha \to x'^\alpha = R^\alpha_\beta(t)x^\beta + a^\alpha(t) \cdots\cdots\cdots\cdots \alpha,\ \beta=1,2,3 \tag{1}$$
$$t \to t' = t + const(\text{常数})$$

这里 $R^\alpha_\beta(t)$是正交含时矩阵，$a^\alpha(t)$是 t 的任意平滑函数，对称性(1)包含了扼杀决定论的对称。

值得注意的是，如果时空的结构变得非常小，似乎就不可能存在运动的有意义的定律，不管是决定论的还是非决定论的。比如说，假设从莱布尼茨时空中去除了时间度规和空间度规，只留下绝对同时性的平面；再假定，物理学定律确定世界充满了恒定质量的尘埃粒子，这些粒子的世界线是永不相交的平滑曲线。这样，要么尘埃的任何平滑的、非相交的运动都为运动定律所允许，要么全部不被允许，因为任何两个这样的运动都是由这个最小时空的一个对称联系的。

面对上述构造，有两种挽救决定论的不同策略可以尝试。它们对应于截然不同的时空本体论态度。第一种策略是补充背景空间的结构。增加一种转动的标准，除去 $R^\alpha_\beta(t)$中的时间依赖性，产生麦克斯韦时空。但是由于 $a^\alpha(t)$仍然是 t 的任意函数，仍然存在着扼杀决定论的对称。如果增加一种惯性运动或者直线运动的标准，把 $a^\alpha(t)$ 线性化为 $v^\alpha t + c^\alpha$(v^α 和 c^α 为常数)，就产生新的牛顿时空①，这种时空的对称由熟悉的伽利略变换

$$x^\alpha \to x'^\alpha = R^\alpha_\beta(t)x^\beta + v^\alpha t + c^\alpha \cdots\cdots\cdots\cdots \alpha,\beta = 1,2,3 \tag{2}$$
$$t \to t' = t + const\ (\text{常数})$$

给出。式(2)中指出的映射不包含扼杀对称的决定论，因为如果这样的映射是一个有限时间延伸的恒等映射的话，无论这个时间延伸多么短，它就是恒等映射时段。注意，这种挽救决定论的方式包含着对运动的“绝对”量的拥护：在新牛顿时空中，询问一个孤立例子是否在加速或者一个孤立的延展物体是否在旋转，都有

① 完整的牛顿时空添加了不同的惯性参照系——“绝对时空”——因此消掉了式(2)中的速度项。

着很好的意义。不可否认，这种绝对加速和旋转可以被称为"关系"量，但是关系中的第二个位置是由时空的结构所提供的——特别地，由惯性结构提供——而不是如那些运动关系论的拥护者所考虑的由其他的实在物体提供。

第二种挽救决定论的策略不是要补充背景时空的结构，而是要通过置疑上述构造——时空"容器观"的一个隐藏前提来进行。生动地讲，这个假设等于把时空看作是粒子和场位于其中的一种介质。更加精确地讲，按照上述方法，它等于这样一种假设：即使 d 是一种时空对称，$\langle \mathcal{M}, O_1, O_2, \cdots, O_M, P_1, P_2, \cdots, P_N\rangle$ 和 $\langle \mathcal{M}, O_1, O_2, \cdots, O_M, d^*P_1, d^*P_2, \cdots, d^*P_N\rangle$（这里 d 是 $\mathcal{M}$ 的任意微分同胚，对某些 j，$d^*P_j \neq P_j$）也描述不同的物理状态。也就是说，对于所有的 i，有 $d^*O_i = O_i$。对容器观的反对导致了（一种）时空关系论。冒着抛弃决定论的可能性的危险，时空关系论者会采取上述构造来表明，运动的关系论者也应该是时空的关系论者。运动的关系论者认为谈论绝对运动是无意义的，所有关于运动的有意义的讨论都应当被解释为对实在物体相对运动的讨论。这样，他们自己就不能使用挽救决定论的补充策略了。因此，如果他们希望得到决定论，他们就必须抓住时空关系论的生命线。

运动的关系论是一种严肃的立场，但是在历史上，它更多地是被人们用承诺而不是用推演的方式描述的。牛顿造就了一个关于陆地和空中物体运动的极其成功的理论。他的支持者保证他们能够提出和他的理论一样经验满足并且具有解释力的理论，但却不需要诉诸他所假定的运动的绝对量。但是，他们提出的观点却主要是夸夸其谈，而不是可行的理论。① 这样的理论是在 20 世纪才得以构建的，参见 [Barbour（巴伯），1974]、[Barbour and Bertotti（巴伯和布托提），1977] 以及 [Barbour，1999] 关于这些理论的历史发展。爱因斯坦的广义相对论之后静止背景时空的观念就被排除了，彻底改变了绝对和相对之争的表述。

3.3 决定论和规范对称

当哲学家听到"规范（gauge）"这个词，他们想到的是基本粒子物理学、

① 牛顿的反对者在一个方面是正确的：牛顿的绝对时空假定，在不同惯性参照系的意义上讲并不是支撑他的运动定律所必需的。

杨—米尔斯(Yang-Mills)理论等。这是一种浅显的看法。非平凡规范自由的重要例子甚至在经典物理学中都出现过——事实上我们在上一节刚刚遇到过。规范思想的出现是因为理论在其态描述与物理态为多对一的关系的意义上有着“冗余结构”，这是迈克·雷德黑德(Michael Redhead)命名的。对于这样一个理论，规范变换的定义是指把那些对应于同一个物理态的描述联系起来的变换。

物理学的历史表明，了解研究中的规范自由的主要原理是为了坚持决定论。这个论题有力地支持了与现代物理学最相关的这类情况，即运动/场方程可以由相互作用原理导出，因此运动方程就以欧拉—拉格朗日方程的形式存在着。① 当拉格朗日函数是非奇异的，适当的初始数据就会选出欧拉—拉格朗日方程的唯一解，拉普拉斯决定论就有效。② 但是，如果这种行动承认李群是一个变分对称(李群的参数是独立变量的任意函数)，那么我们就得到了非充分决定性的例子，因为诺特(Noether)第二定理告诉我们欧拉—拉格朗日方程必须满足一组数学恒等式。③ 当这些独立变量包含时间的时候，时间的任意函数会出现在欧拉—拉格朗日方程的解中，这就显然破坏了决定论。

这一点可以在上一小节中介绍的麦克斯韦时空中构造的粒子力学的帮助下得以说明。这种时空对称性下的适当拉格朗日不变量由下式给出：

$$L = \sum \sum_{j<k} \frac{m_j m_k}{2M}(\dot{\mathbf{x}}_j - \dot{\mathbf{x}}_k)^2 - V(|\mathbf{x}_j - \mathbf{x}_k|) \qquad M := \sum_i m_i \tag{3}$$

这个拉格朗日函数在海森矩阵 $\partial^2 L/\partial \dot{\mathbf{x}}_i \partial \dot{\mathbf{x}}_j$ 不存在反演的意义上是奇异的。欧拉—拉格朗日方程是：

$$\frac{d}{dt}\Big(m_i(\dot{\mathbf{x}}_j - \frac{1}{M}\sum_k m_k \dot{\mathbf{x}}_k)\Big) = \frac{\partial V}{\partial \dot{\mathbf{x}}_i} \tag{4}$$

这些方程并不能唯一地决定粒子位置的演化：如果 $\dot{\mathbf{x}}_i(t)$ 是一个解的话，$\mathbf{x}'_i(t) = \mathbf{x}_i(t) + \mathbf{f}(t)$ 也是。对于任意的 $\mathbf{f}(t)$，这就证实了上面对决定论的明显

---

① 关于拉格朗日和哈密顿(Hamiltonian)形式的说明，见巴特菲尔德在本书中的论述。

② 如果至少加上下面 3.5 节中讨论的连续性假设的话。

③ 对于诺特理论的说明可见[Brading and Brown(布拉丁和布朗)，2003]以及布拉丁和卡斯特拉尼在本书中的论述。

崩溃所给出的直观论据。决定论可以通过把变换 $\mathbf{x}_i(t)\to\mathbf{x}_i(t)+\mathbf{f}(t)$ 看作是一个规范变换而得到恢复。

这种通往规范的方法的系统化发展是由狄拉克实施的。① 一个奇异拉格朗日系统对应于一个约束哈密顿系统。主要的约束作为正则动量定义的结果出现。在一个一阶拉格朗日量 $L(q,\ \dot{q},\ t)$（q 代表位形变量，且 $\dot{q}:=dq/dt$，正则动量为 $p:=\partial L/\partial\dot{q}$）的简单情况中。第二约束作为以下要求的结果出现：主要的约束要为运动所保留。整组约束选出了哈密顿相空间 $\Gamma(q,\ p)$ 的约束表面 $C(q,\ p)$。第一类约束是那些在 $C(q,\ p)$ 上与所有约束交换的约束。正是这个第一类约束被看作是规范变换的基础。这里规范不变量（又叫"可观测量"）就是在规范轨迹上为常量的相函数 $F:\Gamma(q,\ p)\to\mathbb{R}$。

把此形式应用于我们所举出的麦克斯韦时空里粒子力学的例子，正则动量为：

$$\mathbf{P}_i := \frac{\partial L}{\partial \dot{\mathbf{x}}_i} = \frac{m_i}{M}\sum_k m_k(\dot{\mathbf{x}}_i - \dot{\mathbf{x}}_k) = m_i\dot{\mathbf{x}}_i - \frac{m_i}{M}\sum_k m_k\dot{\mathbf{x}}_k \tag{5}$$

这些动量不是独立的，但是必须满足三种基本约束，这些约束要求总动量

$$\phi_\alpha = \sum_i p_i^\alpha = 0,\ \alpha = 1,2,3 \tag{6}$$

中的 x，y 和 z 分量为零。这些主要约束也是仅有的约束，不存在第二类约束，它们都是第一类的。这些约束在每一个位移 $\mathbf{x}_i$ 中产生了相同的规范自由，也就是说，由时间的相同任意函数给出的欧几里得位移。规范不变量，比如相对的粒子位置和相对的粒子动量，确实是确定地演化的。

约束形式的技术细节是复杂的，但是不应该忽视这样的事实：挽救决定论的渴望是激励专家们的动力。这是一段来自于约束哈密顿系统的标准参考书[Henneaux and Teitelboim（享诺和泰特尔鲍姆），1992]相关内容：

> 整个哈密顿函数中任意函数的出现……告诉我们不是所有的 p 和 q 的量（结构变量和它们的正则动量）都是可观测量（即真实物理学

① 关于这些问题标准的参考文献是[Henneaux and Teitelboim，1992]。关于这种形式的温和处理，见[Earman，2003]。

> 量)。换言之，虽然当 q 和 p 的量的集合给定的时候，物理态是唯一定义的，但反过来却不正确——即是说，有不止一个正则变量值的集合可以表达一个给定的物理态。想知道这个结论是如何得出的，我们要注意，如果给定我们一个正则变量在时刻 t_1 的初始态并且因此完全确定了这个时刻的物理态，我们期望运动方程能完全确定其他时刻的物理态。这样，从定义上，正则变量在 $t_2 \neq t_1$ 的值的任何不明确性都应当是一种物理不相关的不明确性。

正如在引文中提出的，人们对决定论明显失效的标准反应是归咎于描述的冗余的出现：按原始变量(q 的和 p 的)得到的态描述与物理态之间的对应是多对一的；若去除这种描述的冗余，就可看到物理态在决定论性地演化着。在这个反作用的过程中可能存在技术的困难。例如，尝试产生一个约化相空间——它的态描述与物理态是一一对应的——通过找出规范轨迹可以导致奇异性。但是当没有遇到这样的技术难题时，正常(也就是非约束)哈密顿动力学可应用于约化相空间，而约化相空间变量是确定地演化的。

除了上述这种对决定论明显失效的这种标准反应，还有两种可能性。第一种非正统的看法把决定论的明显违背看作是真的。这等于是说：(a)把约束形式的内容当作是与真正的物理学量一样的规范不变量；(b)否认这些量由那些当与处于运行状态的定律结合起来的时候就恢复了决定论的定律所支配。第二种非正统的说法接受了正统的结论，认为决定论的失败仅仅是表面上的，但是它由于接受了(a)又违背了正统，且由于否认(b)而背离了第一种说法，并且因此把恢复决定论的附加定律的存在作为先决条件。这两种说法中，表面与(a)相符合的例子很容易从物理学著作的例子中构造出来，在那些作品中初值问题是通过对用规范不变量表达的运动方程补充规范条件以固定唯一解来解决的。例如，按照电磁势术语写出的麦克斯韦方程并不能为势及其时间导数的初值问题决定唯一解。加上洛伦兹(Lorentz)规范条件就把麦克斯韦方程变为了承

认初值问题公式的二阶双曲偏微分方程(见4.2节)。①类似的例子在广义相对论中也可以构造，广义相对论正统地把度规势作为规范变量处理(见6.2节)。在这些例子中，正统的目标就是通过选择规范来得到规范不变量的值。如果在思想里面没有这样一个清晰的目标，这个过程就会产生一种错觉，认为非规范不变量是具有物理意义的。上述两种非正统说法正是把这个错觉当作真的了。第二种说法没有把规范条件作为计算方便的问题，而是把其当成了附加物理学定律。据我所知，历史上这种说法在物理学中没有得到丰富的发展。

由于决定论不能先天地保证为真，那么认为约束形式的正统解读保证了运动方程承认初始值形式这一事实，就一定意味着支持决定论的实质性假设存在于形式之中。确实如此，因为拉格朗日/哈密顿形式在解的时空上强加了一个结构：在本书第一章和第二章解释的几何语言中，解的空间有着辛或者预辛结构。这种形式当然并不能保证可以运用于理论提出者可能选择的用来作为运动定律的所有运动方程。实际上，它甚至不能保证可以应用于所有牛顿类型的二阶常微分方程。19世纪80年代亥姆霍兹(Helmholtz)发现了从一种作用原理导出这种类型的方程所需的必要条件。这些条件后来被证明是(局域)充分的，也是必要的。在一个多世纪之后，为更加广义的运动方程寻找可以从一种作用原理导出的充分必要条件——不管是以偏微分方程还是以常微分方程的形式——仍然是一个热点研究话题。②

3.4 牛顿物理学中场和流体的决定论

牛顿引力理论可以被解释为一种场论。引力由 $\mathbf{F}_{grav} = -\nabla\varphi$ 给出，其中引力势 φ 满足泊松(Poisson)方程：

$$\nabla^2\varphi = \rho \tag{7}$$

ρ 是密度函数。如果 φ 是泊松方程的一个解，那么 $\varphi' = \varphi + g(\mathbf{x})f(t)$ 也是。其

---

① 其中 $\mathbf{A}$ 是一个矢量势，Φ 是标量势，洛伦兹规范要求 $\nabla\cdot\mathbf{A} + \frac{\partial\Phi}{\partial t} = 0$(这里将光速设定为单位1)。

② 数学家们在“反演问题”的标题下讨论此问题。关于此问题的精确形式化以及对结果的检查，见[Anderson and Thompson(安德森和汤普森)，1992]和[Prince(普林斯)，2000]。

中 $g(\mathbf{x})$ 是空间变量的一个线性函数，是 t 的一个任意函数。选择一个 f 使得 $t \leqslant 0$ 时 $f(t)=0$，但 $t>0$ 时 $f(t)>0$。探测粒子在预备解(primed solution)中 $t=0$ 之后经历的额外引力正比于 $f(t)$，且由过去的任何事情非充分决定。

如果要求引力必须要与引力源相关，就可以排除式(7)的决定论破坏解。但是要消除泊松方程的齐次解，就是要走向这样一个方向：把牛顿引力场看作是用来描述本质上是粒子间真实的直接相互作用的引力相互作用的单纯数学工具。① 这样，决定论有助于解决牛顿物理学的本体论问题：牛顿物理学中对决定论的坚持把场降级到了第二类地位。在相对论物理学中，场得到了承认，其原因之一是相对时空结构支持了保证场的确定解的场方程(见 4.2 节)。

在牛顿物理学中，自然出现的场方程是椭圆方程(比如泊松方程)或者抛物线方程，它们都不能独立地支撑决定论。举一个后一种方程的例子，即经典热方程：

$$\nabla^2\Phi=\kappa\frac{\partial\Phi}{\partial t} \tag{8}$$

其中 Φ 是温度变量，κ 是热导率。② 在有限时间之后式(8)的解可能就不存在了，因为温度“爆炸”了，唯一性也失效了。因为如果热方程任意快地热传导，可能构造出意外的解 Φ，有着以下特性：(i) Φ_s 无限可微；(ii) $t\leqslant 0$ 时 $\Phi_s(\dot{\mathbf{x}}, t)=0$，但是 $t>0$ 时 $\Phi_s(\mathbf{x}, t)\neq 0$，见[John(约翰)，1982, Section 7.1]。由于(8)是线性的，如果 Φ 是一个解，那么 $\Phi'=\Phi+\Phi_s$ 也是。并且由于 Φ 和 Φ' 在 $t\leqslant 0$ 时相同但是在 $t>0$ 时不同，意外解的存在就完全破坏了决定论。

式(8)的解的唯一性可以通过要求 $\Phi\geqslant 0$ 重新得到，这得益于把 Φ 解释为温度，因为维德尔(Widder)[1975, 157]已经表明，如果式(8)中 $\Phi(\mathbf{x}, t)$ 的一个解在 $t=0$ 时为零，并且在 $t\geqslant 0$ 时对于所有的 $\mathbf{x}$ 都为非负的话，那么它就必须恒为零。但是可以期望不用必须规定非负性来支撑，决定论就可以被建立起来，然后被用于表明如果在 $t=0$ 时温度分布对所有的 $\mathbf{x}$ 都是为非负的，那么唯

① 牛顿引力场地位的下降也可得到以下事实的支持：不像在相对论中遇到的场，牛顿引力场不携带任何能量或者动量。

② 此方程不是伽利略不变的这个事实不需要引起关注，因为 Φ 含蓄地指出了一种介质的温度，静止参照系是描述热扩散的首选参照系。

一被决定的演化就是保持温度为非负的。或者，式(8)的解的唯一性和存在性都可以通过限制 $|\mathbf{x}|\to\infty$ 时 $|\Phi(\mathbf{x}, t)|$ 的增长来得到。但是，还是可以期望这种增长的限制可以是决定论演化的结果，而不是必须要把它规定为使决定论成为可能的条件。

可以通过在一开始就提出运动学和动力学的区别来避免为了支持决定论而回避问题实质的现象。运动学是对可能态的空间进行确定。比如，量子力学波动方程的有限增长不能确定薛定谔(Schrödinger)方程提供的决定论，因为其限制来源于波动函数是希尔伯特(Hilbert)空间的一个元素这个条件。而希尔伯特空间是量子力学的运动学处理的一部分(见第5节)。由于这种处理是为了保证波函数的概率解释，所以我们得到讽刺的结果：概率的引入，似乎是要宣判决定论的失败，却同时支持了决定论。上面刚提到的例子，以及下一小节中的例子，以及本小节一开始的例子，都表明在经典物理学中运动学/动力学的区别有时可以是相对不固定的，确定的考虑被用于决定在何处适可而止。下面的例子会增强这种寓意。①

不可压缩流体在 $\mathbb{R}^N$($N=2, 3$)运动的纳维叶—斯托克斯(Navier-Stokes)方程为：

$$\frac{D\mathbf{u}}{dt} = -\nabla p + \nu\Delta\mathbf{u} \tag{9a}$$

$$div(\mathbf{u}) = 0 \tag{9b}$$

其中 $\mathbf{u}(x, t) = (u^1, u^2, \cdots, u^N)$ 是流体的速度，$p(\mathbf{x}, t)$ 是压强，ν = 常数且 $\geqslant 0$ 是黏滞系数，$D/dt := \partial/\partial t + \sum_{j=1}^{N} u^j \partial/\partial x^j$ 是迁移导数，参见[Foias *et al.*(福尔斯等)，2001]的综合考察。如果流体遭遇外力，式(9a)的右边就要另外加上一项。欧拉方程是 $\nu=0$ 的特例。式(9a－b)的初值问题通过给出初始值

① 对式(8)引出的决定论问题的另一种反应是，式(8)仅仅是一种唯象的方程，热是分子的运动，因此决定论的命运最终取决于在粒子的运动定律的基础上推延它们。然而，下面将会看到，为了确保粒子运动的决定论，有时需要规定无穷远处的边界条件。在任何情况下，推延策略的逻辑结论都意味着只有拥有了最终包涵一切的理论，才有可能对决定论做出评判。似乎对今天的物理学哲学来讲好一点的策略是，认识到今天最好的理论可能会在将来被一个更好的理论所代替，而这个更好的理论将会传递关于决定论的不同信息。

$$\mathbf{u}(\mathbf{x},\ 0)=\mathbf{u}_0(\mathbf{x}) \tag{9}$$

出现。其中 $\mathbf{u}_0(\mathbf{x})$ 是平滑(C^∞)自由发散矢量场。通过满足式(9)—式(10)的函数 $\mathbf{u}$, $p\in C^\infty(\mathbb{R}^N x[0,\ \infty))$ 得到解决。对于物理学的合理理解，要求 $\mathbf{u}_0(\mathbf{x})$ 不能随 $|\mathbf{x}|\to\infty$ 增长得太大，也要求流体的能量对于所有的时间都有界:

$$\int_{\mathbb{R}^N}|\mathbf{u}(\mathbf{x},t)|^2dx<\infty\text{，对全部的 }t>0 \tag{10}$$

当 $v=0$ 时，能量守恒，而 $v>0$ 时能量消耗。

对于 $N=2$，大家知道对于任何给定的 $\mathbf{u}_0(\mathbf{x})$ 都存在物理学合理的平滑解。对于 $N=3$，问题就变为开放的了。然而对于这种情况，大家知道如果解的存在所要求的时间区间 $[0,\ \infty)$ 被替换为 $[0,\ \mathrm{T})$(T 为依赖于 $\mathbf{u}_0(\mathbf{x})$ 的可能有限数)的话，这个问题就有正解。当 T 有限时，它被称为“爆发时间”，因为 $|\mathbf{u}(\mathbf{x},\ t)|$ 在 t 趋近于 T 时必然会变得无界。对于欧拉方程，有限的爆发时间意味着 t 趋近于 T 时旋度，即 $\mathbf{u}(\mathbf{x},\ t)$ 的弯曲变得无界。

纳维叶—斯托克斯方程的平滑解如果存在，则是唯一的。这个断言似乎被纳维叶—斯托克斯方程的对称证明为错的。因为如果 $\mathbf{u}(\mathbf{x},\ t)=\mathbf{f}(\mathbf{x},\ t)$, $p(\mathbf{x},\ t)=g(\mathbf{x},\ t)$ 就是一个解，那么其变换 $\bar{\mathbf{u}}(\mathbf{x},\ t)=\mathbf{f}(\mathbf{x}-\varepsilon\alpha(t),\ t)+\varepsilon\alpha_t$, $\bar{p}(\mathbf{x},\ t)=g(\mathbf{x}-\varepsilon\alpha(t),\ t)-\varepsilon\mathbf{x}\cdot\ \alpha_t+\frac{1}{2}\varepsilon^2\alpha_{tt}$ 也是一个解，这里 $\alpha(t)$ 是 t 的一个任意平滑函数，参见[Olver(奥尔弗)，1993，130，177]。选择 $\alpha(t)$ 使得 $\alpha(0)=\alpha_t(0)=\alpha_{tt}(0)=0$，但是对于 $t>0$ 的 $\alpha(t)\neq0$，同样的初始值却会产生不同的结果，除非 $\mathbf{f}(\mathbf{x}-\varepsilon\alpha(t),\ t)+\varepsilon\alpha_t=\mathbf{f}(\mathbf{x},\ t)$。然而，变换解违背了能量条件(11)的有限性。

解的存在可以进行如下证明。(9a－b)乘以一个平滑测试函数并在 $\mathbf{x}$ 和 t 上部分积分，会产生一个积分方程，它在任何分别是 L^2(平方可积)和 L^1(可积)的 $\mathbf{u}(\mathbf{x},\ t)$ 和 $p(\mathbf{x},\ t)$ 上都是确定的。如果存在这样的满足所有测试函数积分方程的解，该解就被称为弱解。从平滑解转到弱解要证明一个解在所有时间上都存在。但是这样一种转变重新提出了唯一性的问题，因为纳维叶—斯托克斯方程弱解的唯一性还没有解决。欧拉方程弱解的显著的非唯一性结果来自谢弗(Scheffer)[1994]和施尼雷尔曼(Shnirelman)[1997]构造的自激发/自毁灭弱

解：在紧致空间$\mathbb{R}^3$中，$t<-1$和$t>1$时$\mathbf{u}(\mathbf{x}, t)\equiv 0$，但在这段时间中间量是非零的。

非常引人注目的是，经典运动方程决定论的基本问题还没有解决，这些问题使人们对数学家认为值得关注的问题产生兴趣。对纳维叶—斯托克斯方程在$N=3$时的平滑解问题的解决为克雷(Clay)数学研究所赢得了一百万美元的奖金，参见［Fefferman(费弗曼)，2000］。

3.5 连续性问题

考虑在势为$V(x)(X\in\mathbb{R})$的一条实线$\mathbb{R}$上运动的质量为m的单粒子。常微分方程的初值问题的标准存在和唯一性定理可以用来表明牛顿运动方程

$$m\ddot{x}=F(x)\coloneqq -\frac{dV}{dx} \tag{11}$$

具有局域的(时间上的)唯一解，如果力函数$F(x)$满足利普希茨(Lipschitz)条件。[①] $-\frac{9}{2}|x|^{4/3}$是势在原点违背利普希茨条件的一个例子。对于初始值$x(0)=0=\dot{x}(0)$，式(12)有多个解：$x(t)\equiv 0$，$x(t)=t^3$，和$x(t)=-t^3$，这里为了方便，设m为单位质量。另外，当对于$t<k$时$x(t)=0$和$t\geqslant k$时$x(t)=\pm(t-k)^3$(这里k是任意正常数)的情况也存在解$x(t)$。这样的力函数并不出现在现实的物理情境中，这就表明自然是尊重决定论的。在量子力学中可以看出自然在与某些可能减弱牛顿决定论的非利普希茨条件势相容的情况下尊重决定论。

3.6 经典解的崩溃

再次考虑在势为$V(x)$的实线$\mathbb{R}$上运动的质量为m的单粒子，并假设$V(x)$满足利普希茨条件以保证对牛顿运动方程的初始值问题有暂时的局域唯一解。

---

① 如果存在一个常数$K>0$使得对于所有x_1，$x_2\in(a, b)$有$|F(x_1)-F(x_2)|\leqslant K|x_1, x_2|$，那么$F(x)$在区间$(a, b)$上满足利普希茨条件。关于这一点的一个充分条件是，$dF/dx$存在且连续，同时对于$(a, b)$上的某些$K>0$，有$|dF/dx|\leqslant K$。

但是，如果势能使粒子在有限的时间内加速到 $-\infty$ 或者 $+\infty$，决定论就失效了。[①]过去决定论被打破了，因为在保持与当前位置和速度相同的情况下，两个这样的解都可以与所有的未来时间(比如) $t \geq t^*$ 相一致，在这些时刻，空间中任何地方都不存在粒子，但是在过去的时间 $t < t^*$ 就不一致了。由于假定了势是不随时间变化的，运动方程是时间反演不变的，因此采用这些逃逸解的反演会产生一些解，在其中任何当前的空的空间都受到无限空间而来的粒子的干扰。这些干扰打破了未来决定论。把这些逃逸和干扰解集合起来会更加不利于决定论。

19 世纪 90 年代，保罗·潘立文(Paul Painlevé)猜想，对于$\mathbb{R}^3$ 中在牛顿引力作用 F 运动的 $N>3$ 点质量粒子，存在着牛顿运动方程的解展示了非碰撞奇点。也就是说，虽然粒子并不碰撞，解在有限的一段时间之后也不存在了。雨果·冯·泽佩(Hugo von Zeipel)[1908]表明在这样一个解中，粒子的位置在有限时间之内一定会变得无界。最后，20 世纪末，夏(Xia) [1992]表明对于 $N=5$ 的点质量粒子，牛顿运动方程允许那些在其中粒子并不碰撞但却会在有限时间内自己加速到空间的无限远处，从而证明了潘立文猜想。参见[Saari and Xia(萨里和夏)，1995]比较通俗易懂的探讨。

决定论能够通过上述对通常的初始条件增加一个无限远边界条件的策略而重获幸运，或者从两种评论获得安慰。第一种是在自然相空间测量中，导致夏类型的逃逸解的初始条件集合有着零测度性。但是不知道是否对于所有的非碰撞奇点都是相同的情况。第二种是非碰撞奇点来源于利用一个无限深势阱在有限时间内得到无限速度的点粒子这一理想。这个评论不足以在考虑有限大小粒子的无限性时挽救决定论，这个问题放到下一小节讨论。

有趣的是，对于在相互吸引的牛顿引力作用下运动的点粒子，量子力学同时解决了可以导致经典解崩溃的碰撞[②]和非碰撞奇点问题(见 5.2 节)。这不仅仅是数学上的求知欲，也因为它是氢原子的存在和稳定性解释中的一个重要

① 其充分必要条件见[Reed and Simon(里德和西蒙)，1975，Theorem X. 5]。

② 当两个或多个点粒子发生碰撞时会产生一个碰撞奇点，在碰撞时间中解是不连续的。

组分。

3.7 无限集

考虑一组只能沿欧几里得空间中的直线运动的桌球集。假设这些桌球只能通过接触产生相互作用，且只能发生两个桌球之间的相互碰撞，每一次碰撞都遵循弹性碰撞的经典定律。显然，读者会说，这样一个系统是决定性的。如果这个集合是有限的，那么这种说法就是对的。但是如果这个集合是无限的，速度也是无限的，那么决定论就失效了，因为即便是有着所有该有的约束，系统也可以自激发，参见[Lanford(兰福德)，1974]。佩雷兹·拉罗多古提亚(Pérez Laraudogoitia)[2001]表明了如何运用这样的无限集来产生上一节提到的逃逸解的类似物。在上一节中，所有的粒子在一个有限的时间总和中全部消失了。在这种场景的时间反演中，空间最初是空的，然后毫无预警地就从空间的无限远处涌入了无限长的桌球流。

如果可以把桌球做得无限小，那么无论是设定反对无限速度的规则，还是在无限远处加上边界条件，都不足以恢复决定论［Pérez Laraudogoitia，2001］。因为它们的可计算的无限集能被芝诺(Zeno)压缩到一个有限的空间区域（0，1]，把第一个桌球的中心放在1上，第二个的中心放在1/2上，第三个放在1/4上，依次类推。为了说明的简单性，假定所有的桌球质量都相等(恒等于1)。一个单位质量的小桌球以单位速度从右向左运动，与第一个桌球发生碰撞，通过芝诺链在单位时间内传递波动，最后所有的桌球都静止不动。所有的桌球在所有时间都静止不动的单调历史当然也是碰撞定律的一个解。把这个单调的历史与前面的相比就可以表明，过去的拉普拉斯决定论被打破了。①

决定论的失效带来的是对能量动量守恒的破坏，虽然只是在很弱的意义上讲。也就是说，在对象桌球最初静止于其中的惯性系中，总能量和总动量各自在碰撞前后有着不同的值。但是在其他惯性系中，不存在违背，因为碰撞前后

---

① 这种有趣的历史的时间反演始于所有的球最初都是静止的，然后它们自激发，从左向右发出一种碰撞波，弹出标记球。如果这种自激发的历史是物理上可能的，那么未来拉普拉斯决定论就被违背了。不过这可能被否定，因为它违背了牛顿第一运动定律。

的值都是无限的。[①] 佩雷兹·拉罗多古提亚[2005]已经表明了如何构造这样一种场景，在其中能量和动量守恒都被强烈地打破，这种打破可以发生在任何一种惯性系中。

3.8 依赖性的领域

带有一些人为性质，上面讨论的经典决定论的处理方式之一可以用一个概念来概括。这个概念在比较经典物理学和相对论物理学中决定论的命运时也很有用。在这里让我们用一个因果曲线来理解时空中表示能量/动量的物理可能演化的时空轨迹的平滑曲线。把对时空区域 S 的有依赖性的未来领域 $D^+(S)$ 定义为所有时空点 p 的集合，任何以 p 为未来终点而没有过去终点的过去方向的因果曲线都横穿 S。对 S 有依赖性的过去领域 $D^-(S)$ 也类似地定义。依赖性 $D(S)$ 的全域是 $D^+(S)\cup D^-(S)$。如果 $p\notin D(S)$，那么可能看起来 S 区域中的态不足以决定 p 时刻的态，因为存在一个穿过 p 但却永远不会出现在 S 上的可能因果过程。

由于经典物理学的运动学和动力学都不能为能量和动量传递速度定义一个上限，那么原则上似乎任何类时曲线——即任何与绝对同时性平面成斜角的平滑曲线——都可以被看作是因果曲线。并且即便 S 被看作是绝对同时性的整个平面时，它也是 $D(S)=\varnothing$ 的一个结果。3.4、3.6 和 3.7 节的结果表明了这个“原则”能够被一些满足牛顿运动原理的系统实现。

我们已经看到，通过补充经典时空的结构，可以得到一些对经典决定论的处理方式。因此我们目前考虑的处理方式也在此范围之内。全牛顿时空是通过对新牛顿时空增加一个重要的惯性系(“绝对空间”)形式的绝对空间得到的。在这个设定下，时空对称足够小以至于当前存在有限的不变速度(直观地是相对于绝对空间测量的速度)，这样就可以对定律形式化，从而为因果传递的绝对速度设定一个有限的上限。特别地，这个运动也不像显示的那么必要，例如，按照经典时空情形中麦克斯韦电磁学理论公式显然要求重要的惯性系这一

① 关于对无限动量/能量的可利用性如何不出意料地呈现非决定论的评论，参见[Norton(诺顿)，1999，1268]。

事实，光速 c 就是在这个惯性系中测量出来的速度。

但是，众所周知，由于绝对空间的运动无法探测，这样一种公式显得有些尴尬。这种尴尬为从经典到相对时空提供了一个直接(虽然过时了)的途径。把狭义和广义相对论中所用的相同的几何语言运用于经典时空，参见[Earman，1989，Chapter 2]，绝对空间由协变恒定的类时矢量场 A^a 表示，在这个矢量场中，积分曲线是绝对空间点的世界线。空间度规由一个衰减的二阶逆变张量 h^{ab}表示，它与 A^a 一起定义了一个张量，这个张量在形式上是闵可夫斯基(Minkowski)度规：$\eta^{ab}:=h^{ab}-A^aA^b$。绝对运动的不可观测性意味着没有什么优先的方式可以把 η^{ab}分割为一个 h^{ab}部分和一个 A^aA^b 部分，表明 η^{ab}在物理上和形式上都是洛伦兹度规。正如我们在4.1节中将会看到的，这将使得决定论在狭义相对论时空情形下局域或者全域时间片段为非空的独立领域植根于更坚实的基础。

3.9 决定论、可预言性以及混沌

拉普拉斯决定论的宇宙参考一个“智能体”(被称为“拉普拉斯小妖”)。

我们应该把宇宙的当前态看作是它的先前态的结果和以后态的原因。一个知道给定时刻自然界所有的相互作用力以及宇宙中所有事物的瞬时位置的智能体，能够在单个公式中包含世界上最大物体和最轻原子的运动，假定它的智力足以处理所有需要分析的数据。对于它来说没有什么是不确定的，过去和将来在它眼里都是现在。①

可能过于按照字面意思理解了拉普拉斯的观点，哲学家和物理学家一样混淆了决定论和可预言性。这种混淆使得他们这样推理：这里可预言性失效了，

① 见[Laplace，1820]，该文献的英文翻译见[Nagel(尼盖尔)，1961，pp. 281—282]。一个多世纪之前，莱布尼茨支持一种类似的观点：“人们看到，在这个广阔的世界中所有事物都按照数学方式运行，无懈可击。那么如果有人拥有看穿事物内部的洞察力，还有足够的记忆力和智慧去设想所有的情形并将它们考虑进来，那么他将成为一个先知并将在当下就如镜中所见一样，看到未来。”引自[Cassirer(卡西瑞尔)，1956，12]。

那么，这里决定论失效了。[1] 这是不正确的，其根源是没有能够区分决定论和可预言性。决定论是一种关于世界如何演化的本体论信条，可预言性是关于用各种各样的方式能从世界当前态的知识推出什么样的未来或者(过去)态的认识论信条。

但是，在承认初值问题的运动定律决定论和实践的可预言性之间存在着有趣的联系。初值问题是在以下意义上被设置的：在某种适当的拓扑中，解连续地依赖于初始数据。[2] 在粒子力学中运用常微分方程的初值问题的标准存在以及唯一性证据也提供了良定性的证据，这可以追溯到存在性证据是构造上的，因为它可以给出构造聚焦于由初始数据决定的解的一系列近似性的过程。

要说明可预言性的良定性的含义，考虑一个由牛顿运动方程组成的单粒子系统。如果满足适当的利普希茨条件，那么对于 $t=0$ 时刻粒子的任何给定位置 $q(0)$ 和速度 $\dot{q}(0)$，(在围绕 $t=0$ 的有限时间之内)都存在唯一解：象征性地表示为 $q(t)=F(q(0),\dot{q}(0),t)$。而且，由于这个初值问题是良定的，对任何固定的 $t>0$(在解确定可以存在于其中的区间内)，F 是 $q(0)$ 和速度 $\dot{q}(0)$ 的连续函数。假定实践的预言任务是要预言粒子在某个给定的 $t^*>0$，精确性为 $\epsilon>0$ 的时候的实际位置 $\bar{q}(t^*)$，假定对位置和速度的测量并不是没有误差，只是误差可以任意小。因为 F 是连续的，存在 $\delta_1>0$ 和 $\delta_2>0$ 使得如果 $|q(0)-\bar{q}(0)|<\delta_1$，$|\dot{q}(0)-\bar{\dot{q}}(0)|<\delta_2$，那么 $|q(t^*)-\bar{q}(t^*)|<\epsilon$。这样，$t=0$ 时测量的精确度为 $\pm\delta_1/2$ 和 $\pm\delta_2/2$ 的实际位置和速度分别保证了当测量值插入 F 时，$t=t^*$ 的函数值回答了分配的预言任务。(但是要注意，由于实际的初始态

---

① 站在哲学的角度，卡尔·波普尔(Karl Popper)是最好的例子。波普尔在1982年已经做到了将“科学决定论”的学说按照预言任务的语言公式化。物理学角度的一个例子是雷克尔(Reichl)[1992]：“现在我们知道牛顿方程是决定论的这个假设是一个谬论！当然牛顿方程是力学的起点，但是一般来说，它只允许我们确定(或是预言)可积力学系统的长时间行为，而这些行为在自然界基本没有被发现。”我很高兴在雷克尔著作的第二版中这一段被改成了：“现在我们知道牛顿方程能预言未来的假设是一个谬论！”引自[Reichl，2004，3]。

② 当拓扑是在一个用时间函数 $s(t)$ 表示的瞬时态上的规范 $\|\cdot\|$ 引起的，那么适定性要求存在一个非递减的非负函数 $C(t)$ 使得对于任何解 $s(t)$ 都有 $\|s(t)\|\leqslant C(t)\|s(0)\|$，$t>0$。

是未知的，要求的精确度 $\pm\delta_1/2$ 和 $\pm\delta_2/2$ 也是未知的，它可能会依赖于未知态和 ϵ，以及 t^*。如果 $\pm\delta_1/2$ 和 $\pm\delta_2/2$ 有非零的最小值对应于给定的预言任务，那么无论初始态是什么，这都可以克服。但是不能保证这样的最小值就存在。某个 $t^{**}<0$ 时刻有着已知位置和速度精确度的优先测量会给出限制，这可以从 $t=0$ 时刻的位置和速度上的 F 计算得出。然后，最小值就可以计算在满足限度范围内任何位置和速度所必须预言的 $t=0$ 时刻的测量精确度 δ_1 和 δ_2。）

雅克·哈达玛（Jacques Hadamard）曾经为偏微分方程的柯西（Cauchy）和初始值问题做出了巨大贡献，他非常正确地运用了"适定性"（也就是"适当地给定"）这个术语。因为他把它当作是物理系统的适当数学描述的一个标准，参见［Hadamard，1923，32］。这个物理系统运动方程允许数值函数，其解连续地依赖于初始值。然而，标准的科朗—希尔伯特（Courant-Hilbert）参考书《数学物理学的方法》（*Methods of Mathematical Physics*）指出：

> "适当给定"的问题并非目前唯一适当反映了真实现象的问题。很不幸，目前几乎没有什么数学上的进步可以解决这个重要任务，甚至不能确定这样的问题并非"适当给定"但仍然非常重要，并且为现实情况所激发［1962，Vol.2，230］。

在［Payne（佩恩），1975］及其参考文献中，可以发现一些进展。

哈达玛的意见是，如果一个系统的时间发展不能持续地依赖于初始条件，那么"在我们看来似乎它被纯粹的随机性所支配了，从庞加莱（Poincaré）[①]时代就知道它精确地构成了决定论中的非连续性，不遵从任何定律"［1923，38］。目前的意见是，经典系统中随机性的出现并不是由于适定性的失败，而是由于混沌的出现。

决定性混沌的引入并没有改变任何上述关于决定论和可预言性的结论。对于混沌的概念不存在普遍统一的一致意见，但是它们的目标类型可以用原因或结果的术语去分辨。原因的术语是解决由于如同正的李雅普诺夫指数这样的指

① 参见在庞加莱《科学与方法》（*Science and Method*）［1952］中的文章"偶然性"。

数而产生了对初始条件的敏感依赖；而结果的表达是各种不同的高阶各态遍历的特性，比如一个混合系统，一个 K 系统，一个伯努利(Bernoulli)系统等。①一般来讲，对初始条件的敏感依赖加上态空间的压缩性足以获得这样的特性。对初始条件的敏感依赖是混沌行为的根本原因，这与解和初始条件的持续依赖并不相互矛盾。假定测量初始条件时的误差可以任意小，因而它并不破坏以任何期待的有限精度预言一个固定未来时间的态的任务。但是，如果测量的精度存在一个固定的最小范围——也就是说，由于测量仪器是宏观的，而且不能在低于某个自然的宏观尺度时进行辨别——那么决定性混沌的出现就可以使得某些预言任务成为不可能。而且，混沌的出现意味着在确定初始条件时不管误差多么小(只要是非零的)，那么预言未来态时的精度就随时间迅速降低了。要通过以极小的误差 $\delta>0$ 确定 $t=0$ 时的初始条件来确保能运用某个在所有 $t>0$ 时刻都为给定的精度 $\epsilon>0$ 来预言，就必须不只要求适定性，还应该要求稳固性，这与混沌是矛盾的。②

经典混沌的案例也表明了宏观层次的决定论并不只与宏观层次的随机性相容，还要与可以成为宏观随机性基础的决定性宏观动力学相容。举例来说，最低阶的各态遍历特性(各态历经)有理由为宏观正则概率分布的应用做出辩护，并提出相对频率解释；因为它暗示了宏观正则分布是对于勒贝格(Lebesgue)测度来说绝对连续的唯一静止分布，而且对一个相体积的测量等同于限制相点在体积中所花费时间的相对频率。在这些情形中，在“客观的”和“认识论的”的概率之间似乎不存在有效对比。概率在如下意义上才是认识论的：对初始条件的一种数学上精确的知识的条件化把结果的概率降至 0 或者 1。但是概率并不仅仅在表达我们的无知时才是认识论的，因为它们附随于基础的宏观动力学。

帕特里克·苏佩斯(Patrick Suppes)[1991；1993]运用这样的例子争辩说，由于我们被限制在宏观层次，决定论对我们来说变成了一个“先验”问题，因为我们说不清楚我们是在处理一个不能简化的随机性，还是一个决定性混沌的案

① 关于这些概念的定义，参见尤菲克在本书中的论述，或者[Lichtenberg and Lieberman(利希滕贝格与利伯曼)，1991]。

② 态 $s(t)$ 上规范的稳定性要求存在一个常数 C，使得对于任意解 $s(t)$，有 $\|s(t)\| \leqslant C\|s(0)\|$，$t>0$。

例。尽管我觉得有必要争论，但我没有完全被说服。有两种相互竞争的假设可以解释观测到的宏观随机性：第一种归因于宏观决定性加上对初始条件的敏感依赖，第二种归因于不可简化的宏观随机性。近几十年对决定性混沌的研究提供了第一种假设生效的细节。第二种假设的细节有待补充。特别地，需要解释观测到的宏观随机性是如何附随于假定的宏观随机性的。[①] 然后必须证明这两种假设是不能由宏观层次所有可能的观测充分决定的。如果这两个条件都满足的话，我们就会面临一个理论被观测证据非充分决定而提出的对科学实在论的整体挑战。科学实在论争论中所有反复排练过的运动与反运动都会积极活动起来。但是这样的交战只有在第二种假设的一些具体版本出现之后才有意义。

3.10 拉普拉斯妖、预言以及可计算性

由于我们可以非常随意地想象此妖，那就让我们假设拉普拉斯小妖可以以绝对的数学精度确定目标系统的初始条件。至于可计算性，让我们假定这个小妖可以任意支配一个图灵机。即便这些能力那样引人瞩目，它们也不能使这个小妖有能力预言系统的未来态，即便它是决定论的。回到上一小节牛顿粒子的例子，如果在确定其解 $q(t)$ 的函数 $F(q(0), \dot{q}(0), t)$ 中考虑了粒子在 $t=0$ 时刻的位置和速度。结果就是一个 t 的函数 $\mathcal{F}(t)$。如果考虑不同的初始条件值，就会产生不同的 $\mathcal{F}(t)$——事实上，根据决定论的假设，$\mathcal{F}(t)$ 与不同初始条件的对应在任何有限时间段上都必须不同，不管这个时间段有多短。由于存在着不同初始条件的连续体，因此就存在着不同 $\mathcal{F}(t)$ 的连续体。但是只有可数的 $\mathcal{F}(t)$ 才能是图灵可计算函数。[②] 因此，对于妖所遇到的大多数初始条件，都不

---

① 微观随机性总会渗入宏观层次，这点并不明显。

② 关于图灵可计算或递归函数的一种熟悉的概念被公式化为自然数的函数。但它可以被推广应用到实数函数中。首先，一个可计算的实数 x 被定义为一个有效收敛的可计算有理数序列 $\{r_n\}$ 的极限，即存在一个递归函数 $f(n)$ 使得 $k \geqslant f(n)$ 要求 $|x - r_k| \leqslant 10^n$。接下来，一个实数序列 $\{x_n\}$ 应当是可计算的，当且仅当存在一个双序列 $\{r_{kn}\}$ 使得当 $k \to \infty$ 时在 k 和 n 中有 $r_{kn} \to x_n$。最后，实数序列的函数是可计算的，当且仅当它将它区域内的每个可计算序列映射到一个可计算序列上，而且它是有效一致连续的。详细讨论参看[Pourel and Richards（包瑞尔和理查兹），1989]。

能通过运用通常的图灵机来计算$\mathcal{F}(t)$在相关t时刻的值来预言对应的$t>0$时的粒子位置$q(t)$，按照皮托斯基(Pitowsky) [1996]的风趣措辞，这个小妖要做出确切的预言必须要有神谕才行。

但是，如果$q(0)$和$\dot{q}(0)$都是图灵可计算的实数，那么就不需要神谕，因为对应的$\mathcal{F}(t)$是一个图灵可计算函数。而且，如果t是一个图灵可计算实数，那么$\mathcal{F}(t)$也是。这源于一个事实：常微分方程的存在和唯一性证明给出了一个有效步骤，可以产生一系列近似，它们都有效地趋于解。因此，如果可计算的初始数据被考虑进步骤之中，那么结论就会是一个有效的可计算解函数。如果运动方程是偏微分方程，那就不需要类似的结论。超前于相对论背景，一个标量场的波动方程提供了一个例子，在其中初始条件的图灵可计算性不会被决定论的演化所保留(见4.4节)。

举一个更加有趣的例子，在其中我们的拉普拉斯妖的版本必须考虑摩尔(Moore)[1990; 1991] 和皮托斯基[1996]所讨论的神谕。摩尔构建了一个抽象的宇宙图灵机嵌入于一个由粒子组成的具体的经典力学系统，粒子弹跳于抛物柱面镜和平面镜之间，这样粒子的运动被约束在一个单位正方形中。运用这个嵌入，摩尔就能够表明递归无法解决的问题是如何可以被转化为关于粒子未来行为的预言任务的。而这样一个任务，小妖如果不能得到神谕的帮助就不能完成，即便它知道粒子的绝对精确的初始条件！例如，图灵定理认为没有一种递归算法能判定一个宇宙图灵机何时会停在一个已给出的输入上。尽管已被嵌入到粒子—镜面系统中的宇宙图灵机的停止态对应正在进入一个此后被约束的单位正方形的确定区域中的粒子，小妖也不能预测粒子何时会进入这个区域。莱斯(Rice)[1953]对图灵定理的概括中很多对处于无约束的未来宇宙图灵机的行为是递归无解的，并且这些逻辑性问题会转化成关于粒子在无约束的未来中行为的物理问题，而小妖如果没有查看神谕就无法回答它们。

读者可能会问，为何我们应当坚持图灵可计算性。不管抽象的图灵机是否能被嵌入到系统中，为何不将一个决定论性的机械系统看作一个类似的图灵机？比如以上例子中做决定性运动的牛顿粒子，为什么不认为粒子是一个类似的计算机，它的运动就是“计算”可能对于任何给定的初始条件$q(0)$，$\dot{q}(0)$的可解的非图灵函数$q(t)=F(q(0), \dot{q}(0), t)$？在字里行间我并没有发现去除

唬人的引用和发展一种类似可计算性的概念有错。但是此类概念的实用价值是可疑的。要决定 t 的哪种函数正在被计算，并且接近于各种 t 值下的计算值，需要以极大的精确度确定粒子的位置。

与拉普拉斯决定论的整体性版格格不入的非图灵可计算性和广义相对论之间的联系将会在 6.6 节提到。

4. 狭义相对论物理中的决定论

4.1 时空的相对论结构如何改善了决定论的命运

狭义相对性理论保留了静止时空背景的牛顿观点，在这种时空之上物理学的剧本得到演绎。但是它们用闵可夫斯基时空替代了经典时空来作为背景。这个替代给决定论的境况带来了巨大改善。闵可夫斯基时空的对称性由庞加莱小组给出，他们承认一个有限的不变速度 c，即光速，使得阐述一种运动/场方程的法则成为可能。这个法则要同时满足时空对称性也就是法则的对称性，以及能量—动量传播速度不能大于 c 这几项基本要求。多亏了这条法则，所有由无限速度派生的对经典决定论的威胁都被一扫而光。

最新的论点可以按照 3.8 节中介绍的概念来阐述。对于讨论的法则的类型，一条因果曲线是一条时空世界线，它在任意点的切线都位于那一点的零锥之内或者之上，其结果是依赖性区域现在都是非平凡的。闵可夫斯基时空承认全域时间函数是多余的。但是相对于经典时空，可以选择一个这样的函数 t，使得 t 的位势面的依赖性区域 D(t 为常数)非空。事实上，可以选择 t 使得对于每一个和所有的 t 为常数的情况，依赖性区域 $D(t=$常数$)$是一个柯西面，即 D $(t=$常数$)$是整个时空。实际上，任何惯性时间坐标都是全域时间函数的例子，所有时间函数的位势面都是柯西面。① 在狭义相对论背景中，之前 2.1 节中给出的拉普拉斯决定论的定义将被理解为将这种柯西特性应用于 t 中。

以上讨论的这些对决定论来说协调的特性并不是狭义相对论自身的自发结

① 留给读者的练习题：构造一个闵可夫斯基空间的全域时间函数 t，使得 t 的任何位势面都不是柯西面。

论，而是包含了大量附加假设，意识到这一点很重要。在狭义和广义相对论物理中都用到的应力—能量张量 T^{ab} 描述了物质—能量在时空中是如何分布的。有时被称为 T^{ab} 的局域守恒定律的 $\nabla_a T^{ab}=0$（其中 ∇_a 是由时空度规决定的协变微商），并不能保证任何观测者测量到的局域能量－动量流永远都是非类空的。这一种保证还要求对于任何指示未来的类时 U^a，$-T^{ab}U_a$ 是一个指示未来的非类空矢量。① 把此要求结合上一项更深的要求：要求任意观测者测量的局域能量密度是非负的，即对于任意的非类空矢量场 U^a，如 $T^{ab}U_aU_b\geqslant 0$，就会产生所谓的支配性的能量条件。毫不奇怪，这一条件与 T^{ab} 的局域守恒一起确实保证了产生 T^{ab} 的物质场在以下意义下不会比光传播得更快：当 T^{ab} 在某个类空区域 S 中为零，那么它在 $D(S)$ 中也必须为零，参见［Hawking and Ellis（霍金和埃利斯），1973，pp. 91—94］。人们认为支配性的能量条件为现实世界中所有遇到的物质场所满足，但偶尔在物理文献中也会声称有反例出现。

4.2 基本场

3.4 节中的例子表明场在经典时空中因为拉普拉斯决定论中的思想而遇到了怎样的困难。情况在闵可夫斯基时空中发生了戏剧性的变化，这种时空支持双曲偏微分方程形式的场方程。② 例如，质量 $m\geqslant 0$ 的标量场 ϕ 的克莱因—戈登（Klein Gordon）方程遵从如下方程：

$$\nabla_\alpha\nabla^\alpha\phi-m^2\phi=0 \tag{12}$$

这是一个线性、对角化、二阶双曲偏微方程的特例。这样的方程有着唯一全域存在的关于初值问题的证据：给定一个闵可夫斯基时空的柯西面 $\sum$ 和 C^∞ 初始数据，由 $\sum$ 上的 ϕ 值和遵守 $\sum$ 的 ϕ 的法向导数组成，存在唯一的式（13）贯穿时空的 C^∞ 解。而且，初值问题在特解连续地依赖于初始数据（适当的拓扑）条件下是适定的，并且最终克莱因—戈登场因果地传播。因为如果在一个封闭子集 $S\subset\sum$ 外初始数据改变了，$D(S)$ 上的特解不会变。注意，我们有一

---

① 减号来自于时空度规的标记（＋ ＋ ＋ －）的选择。

② 与物理应用相关的偏微分方程分类标准的参考文献见［Courant and Hilbert，1962，Vol. 2］，也可见［Beig（贝格），2004］。

个完全清晰的拉普拉斯决定论的例子——无须用无限远处的边界条件或是任何其他可做的测量来填补非决定论可以钻进的漏洞。相比之下，人们已经知道，给出闵可夫斯基时空类时超曲面上的初始数据会导致非适定性的柯西问题。实际上，不仅解不会连续地依赖于初始数据，而且还存在 C^∞ 初始数据，它没有对应解。这种类时和类空方向上的决定论命运之间的非对称性，有可能如上面我们所注意到的，有所好转并且被当作是挑选时间维度时的一个基础。

应该强调，人们只知道有限种类的双曲偏微分方程对初值问题是正定的。对于数学物理学，要表明物理学理论中出现的场方程可以写为一种属于这些种类的形式，这是一个挑战。相对而言，要表明当用势的形式表示时，如果对势的运用有一个适当的规范条件，那么磁场的无源麦克斯韦方程可以采用线性的、对角的、二阶双曲微分方程的形式就会简单多了。在其他情况中这项挑战需要真正的智慧。①

物理学家如此确信经典（=非量子的）狭义相对论物理中的决定论，以至于他们根据场是否遵循闵可夫斯基时空上初值问题的全域存在性和唯一性定理形式的拉普拉斯决定论去区分“基本的”和“非基本的”物质场。根据这个标准，克莱因—戈登场和无源麦克斯韦电磁场被归类为基本物质场。然而，尘埃物质场却不能做出这种区分，因为奇异性可以源于常规的初值。例如，在一个塌缩的尘埃球中，如果外壳向内塌缩得足够快，以至于突破内壳，那么尘埃的密度就会变得无穷大。这样的壳——穿透奇异性甚至对于物理上可行的麦克斯韦方程组的合理初始数据都是可以成立的。对于这样的合理初始数据，电磁场的源是由一个遵从洛伦兹法则的带电尘埃构成的。但是要在这样的例子中忽视失败，不需要对决定论的绝对忠诚。可以这样去驳斥它：尘埃物质是一种想象，就像其他所有的想象一样，它在某些情形下起作用。但是，当在式(13)的右边加项使克莱因—戈登方程转化为一个非线性方程时，就需要用决定论的信念来处理将会发生的事情。例如：

$$\nabla_a \nabla^a \phi - m^2\phi = \lambda\phi^2 \tag{13}$$

其中 λ 是一个常数。我们已经知道，式(14)对应于常规初始数据的解可以在一

① 例子参见[Beig，2004]。

个有限的 t 值下变为无限，并且这样的数据都是非零测量的，参见[Keller(凯勒)，1957]。

人们已经做了大量尝试，去修正经典耗散流体的纳维叶—斯托克斯方程(参见3.4节)，从而使它们在变为一个有着因果传播的双曲偏微分方程组系统的意义上与狭义相对论一致。典型地，在这里成功的准则是作为结果的系统要与初值公式相容，再一次巩固了狭义相对论中对决定论的信念。实施这个计划的一个难点在于引入动力学变量和动力学方程的必要性，并因此产生了经典方程的很多不同的相对论推广。杰拉奇(Geroch)[1995]已经指出，我们不需要为这种"富人的尴尬"所累，因为这些相对论推广中的不同之处在纳维叶—斯托克斯理论拥有的经验观察的层次上被洗刷干净了。

4.3 狭义相对论中的可预言性

闵可夫斯基时空的零锥结构可能使彻底的拉普拉斯决定论对那些嵌入的观测者起到反作用。这些观测者不是简单地"被给定"初始数据，而是必须通过与系统发生因果联系而搜索出这些初始数据。这些系统的未来是他们所要预测的。比如，考虑一个观测者 O 的困境，这个观测者的世界线在图1中用 γ 标注。在时空位置 p 点，这位观测者想要预测3天后她会发生什么事(正如在她的固有时中测量的)。

假定实际上，对 O 来讲3天后她对应时空位置为 p' 点。同时假定控制与预测相关的变量的方程使得一个类空超曲面 Σ 上的初始数据决定了 $D(\Sigma)$ 中的唯一解。然后为了用解相关方程的方法来实施预测，O 必须确定局部类空超曲面上成片分布于 p' 点过去零锥上的态，如图1中的 σ。当 O 的"现在"爬上她的世界线，"现在"的过去的光越来越多地扫过 σ 的部分，但是直到她的"现在"到达 p' 点，她才能够因果地接近 σ 的全部。对穿于 p' 点的过去锥上的所有其他片段，情况也是一样。因此，正

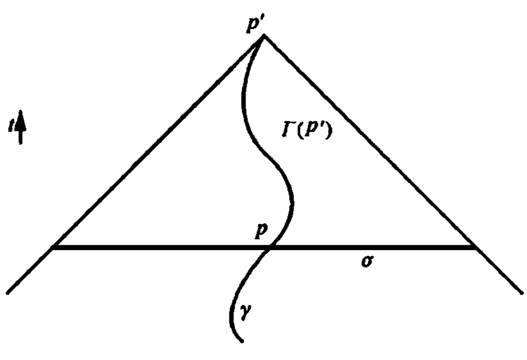


图1 闵可夫斯基时空中可预言性的失败

是那个为拉普拉斯决定论提供安全基础的类空结构，在将被预测的事情发生前阻止了O获得她所需要的信息。

这种可预测性的困境可用一种方法形式化。在广义相对论时空$\mathcal{M}$中研究g_{ab}的可预言性时，这种方法将会变得有用，这里$\mathcal{M}$是一个可微流形，g_{ab}是定义在所有$\mathcal{M}$上的洛伦兹度规，闵可夫斯基时空是一个特例，其中$\mathcal{M}=\mathbb{R}^n$，且g_{ab}是闵可夫斯基时空的度规。杰拉奇在 1977 年定义了一个$q\in\mathcal{M}$点的可预测区域$P(q)$，它包含所有点$p\in\mathcal{M}$使得：(i)每个未来终点为p却没有过去终点的过去指向的类时曲线都进入了q点的时间上的过去$I^-(q)$[①]；(ii)$I^-(p)\not\subseteq I^-(q)$。我们需要条件(i)来保证那个能够影响$p$点事件的因果过程对区域$I^-(q)$来说是很容易看到的。这个区域对“现在”为$q$的观测者来说是可以因果地进入的。我们需要条件(ii)来保证从q的角度看，在p点被预测的事件还没有发生。对闵可夫斯基时空可预言的困境可以陈述为一个定理，即对于每个闵可夫斯基时空的q点，$P(q)=\varnothing$。

注意，可预言性困境的出现不仅仅是因为闵可夫斯基时空的局部零锥结构，而且还因为它的全域拓扑结构。要讲清楚这一点，假定(1+1)闵可夫斯基时空中的空间被压缩而产生图 2 所示的圆柱空间$\mathcal{C}$。现在可预言性成为可能，因为任意q点的$I^-(q)$包含一个柯西面，比如图 2 中的Σ。结果是$P(q)=\mathcal{C}-I^-(q)$。

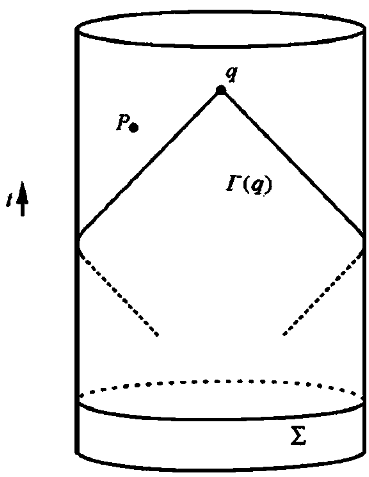


图 2　当闵可夫斯基时空的外表被压缩后，提升了的可预言性的命运

对于标准的闵可夫斯基时空以及其他对于每一个时空点q都有$P(q)=\varnothing$的时空而言，人们会好奇预言可能会有多么可靠。答案是，如果要求完全的可靠性，那么唯一可

① 对于相对论时空$\mathcal{M}$，g_{ab}上的一点q，时序的过去$I^-(q)$包含了所有的$p\in\mathcal{M}$使得存在一条p到q的未来指向的类时曲线。点q的时序未来$I^+(q)$也是类似地定义的。

靠的预测会有条件形式，其前提是不能从 q 点因果地得到事件。但是会有很多这样的带有不同前提和不同结果的条件，同时由于人们不会去了解哪一个前提得到了实现，那么能做得最好的就是进行一个包含很多条件的“预测”(像经济学预测一样)。另一方面，如果不要求完全的安全性，那么如果从过去的观测点到条件中的一个前提进行归纳推理，就会得到具有可能性而不是确定性的无条件的预测。

如果有人想要得到原理上可证实的预测，那就需要对预测范围的定义附加第三个条件，即(iii)$p \in I^{+}(q)$。图2中的点 p 满足条件(i)和(ii)，但不满足条件(iii)。

4.4　狭义相对论和可计算性

包瑞尔和理查兹[1981]构造了一个例子，在这个例子中决定性的演化并不保持图灵可计算性。所讨论的运动方程是相对性的波动方程，在惯性坐标系下可写为：

$$\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}+\frac{\partial^2 u}{\partial z^2}-\frac{\partial^2 u}{\partial t^2}=0, \qquad c\equiv 1 \tag{14}$$

包瑞尔和理查兹研究了对应于 $t=0$ 时刻的形式为 $u(x, y, z, 0)=f(x, y, z)$ 和 $\partial u(x, y, z, 0)/\partial t=0$ 的初始数据。他们表明，存在一个图灵可计算的 $f(x, y, z)$ 使得对应的解 $u(x, y, z, t)$ 在 $t=1$ 处不是图灵可计算的。然而，这样的解对于波动方程必然是一个弱解(在3.4节的意义上)，因为它必定不是可微的。而且非保留结果对用来定义收敛的标准是灵敏的。事实上，如果图灵可计算性是用能量标准①定义的，那么对于任意图灵可计算的函数 f 和 g，对应于 $u(x, y, z, 0)=f(x, y, z)$ 和 $\partial u(x, y, z, 0)/\partial t=g(x, y, z)$ 的解 $u(x, y, z, t)$ 都是图灵可计算的，参见[Pour-el and Richards, 1989, pp. 116—118]。

① 对初始条件 f 和 g，能量形式通过下式给出：$\|f, g\|^2 := \iiint[(\nabla f)^2+g^2]dxdydz$。同时对于 $\mathbb{R}^4$ 上的函数 u，规范是 $\|u(x, y, z, t)\| = \sup_t E(u, t)$，其中：

$$E(u,t)^2:=\iiint\left[\nabla u+\left(\frac{\partial u}{\partial t}\right)^2\right]dxdydz$$

如果 u 是波函数的解，那么 $E(u, t)$ 与 t 无关。

5. 普通量子力学中的决定论与非决定论

按决定论的传统观念，量子力学是非决定论的典型例子。正如大多传统观点认为的，这一小部分包含了真理的元素，但又像大多传统观点一样它忽略了重要的细微之处：事实上在这个例子中，量子力学在某些方面要比经典物理学更具有决定性且更具有可预言性。要从传统观点中提炼出真理的元素会花费相当大的努力，特别地，量子非决定论的提出是因为"波包塌缩"的传统观点是基于量子测量过程的一个有争议的解释。在转向这些问题之前，我将会在5.1节中讨论一个与前面提出的一些论题相关的问题，同时在5.2—5.4中所讨论的哲学文献中被不公平地忽略掉的一些问题。

5.1 量子力学中的决定论和伽利略不变性

这里举另外一个例子，它说明了将决定论和对称性的观点联系起来会在物理学的深刻见解上取得多么丰硕的成果。考虑在一条实线$\mathbb{R}$上的单个无自旋粒子的运动并且做一些类似于波函数的希尔伯特空间$\mathcal{H}$的处理，即$\mathcal{H}=L_c^2(\mathbb{R}, dx)$。量子粒子的态$\psi(x)\in\mathcal{H}$的演化满足薛定谔方程：

$$i\hbar\frac{\partial\varphi}{\partial t}=\hat{H}\psi \tag{15}$$

其中$\hat{H}$是哈密顿算子。这个演化是决定论的，或者据说是决定论的。但是把假定的决定论与假定的式(16)①的伽利略不变性结合起来，就会立刻产生一个困惑。由于式(16)在时间上是一阶的，那么，给定$t=0$时刻所有$x\in\mathbb{R}$的波函数$\psi(x, 0)$的值，就应该能确定所有$t>0$时刻$\psi(x, t)$的值。但是如果薛定谔方程是伽利略不变的，那又会如何？一个适当的伽利略变换$x\to x'=x-vt$，$v\neq0$是一个对$t=0$的恒等映射，但对于$t>0$就不恒等了。假设式(16)是伽利略不变的，那么这个映射必须在解到解之间也成立。由于我们讨论的映射在$t=0$时是恒等的，那么两个解应当有相同的初始数据$\psi(x, 0)$；但是因为映射在$t>0$

① 薛定谔方程的伽利略不变性的处理参见[Brown, 1999]。

不恒等了，那么初始解及其在伽利略改进中的影像在未来应当发生偏离，从而违反拉普拉斯决定论。这个小困惑的解决方法是，拒绝在伽利略变换下 ψ 的标量起作用的隐含假设。实际上，薛定谔方程的伽利略不变性可以用来暗示 ψ 的伽利略变换取决于粒子的质量。这反过来需要质量的“超选择定则”，这是由柏格曼(Bargmann)在1954年发现的，意思就是不同质量的态的叠加在非相对论量子力学中是没有物理意义的。

5.2 量子力学如何比经典力学更具有确定性

关于量子力学，物理教科书提供了一个量子化的过程，从感兴趣的系统的经典动力学的哈密顿公式开始，以含糊的排列算子为模，产生了对方程(16)中量子哈密顿量算子 $\hat{H}$ 的形式化表征。① 但是为了将这种形式化表征转变为一个真正的算子，必须明确定义的范围。因为通常来说，$\hat{H}$ 是一个无界算子，因此，最好定义在希尔伯特空间的稠密区域。一般在希尔伯特空间中不难找到一块稠密区域，在这个区域中 $\hat{H}$ 作为一个对称算子起作用。那么问题就变成了这个算子在本质上是否是自轭的，即有一个唯一的自轭(SA)延展，它也用$\hat{H}$ 表示。② 如果是这样，那么 $\hat{U}(t) := \exp(-i\hat{H}t)$ 对于所有的 $t \in \mathbb{R}$ 是幺正的，而且由于 $\hat{U}(t)$ 是对整个希尔伯特空间定义的，这样希尔伯特空间中每一个矢量的时间演化 $\psi(t) = \hat{U}(t)\psi$ 都是对所有时间定义的。薛定谔方程(16)只是这个演化方程的“无穷小”版本。因此，如果 $\hat{H}$ 在本质上是自轭的，则任何困扰经典态决定性演化的问题都不能对量子态的决定性演化产生困扰。

令人惊奇的是，在某些对应经典系统没有显示决定性演化的情况中，量子哈密顿算子在本质上可以是自轭的。回想3.5节的例子，一个粒子在电势 $V(x)$，$x \in \mathbb{R}$ 中的一条实线$\mathbb{R}$ 上运动。正如我们所看到的，当电势在原点附近

① 量子化的不同方法详细参见兰兹曼在本书中的论述。

② 一个定义在希尔伯特空间$\mathcal{H}$的稠密区 $D(\hat{O})$ 域上的线性算符 $\hat{O}$ 是对称的，这只在对所有 ψ，$\varphi \in D(\hat{O})$ 都有 $(\hat{O}\varphi, \psi) = (\varphi, \hat{O}\psi)$ 的情况下成立，其中 $(\cdot, \cdot)$ 是$\mathcal{H}$上的内积。$\hat{O}$ 是自轭的意味着 $\hat{O} = \hat{O}^*$，即 $\hat{O}$ 是对称的并且 $D(\hat{O}) = D(\hat{O}^*)$，其中 $\hat{O}^*$ 是 $\hat{O}$ 的共轭。这里 $D(\hat{O}^*)$ 被定义为 $\varphi \in \mathcal{H}$的集合，这样就存在一个 $\chi \in \mathcal{H}$，使得对于所有 $\psi \in D(\hat{O})$，都有 $(\hat{O}\psi, \varphi) = (\psi, \chi)$，那么，$\hat{O}^*\varphi := \chi$。

与 $-|x|^{4/3}$ 成比例时，牛顿运动方程的初值问题没有唯一解。但如果 $V(x)$ 是局部可积且有下界的，那么量子哈密顿算子 $\hat{H}=\frac{\hbar^2}{2m}\frac{d^2}{dx^2}-V(x)$ 本质上是自轭的。如果适当地调整经典非决定势，使其远离原点，那么它也可以满足这一条。①

如果初值问题在时间中有局域的唯一解，但这个解在有限时间之后会崩溃，那么经典非决定论的另一种形式就会出现。3.6 节中给出的例子是一个在相互吸引的作用力为 $1/r^2$ 影响下运动的经典质点粒子系统，人们注意到，解可能会因为碰撞奇点或者非碰撞奇点而崩溃。两种类型的奇点都不会在量子类似中发生，因为这种情况下的量子哈密顿算子还是自轭的。②

量子力学也消除了 3.7 节讨论的经典桌球的芝诺非决定论，至少在间接的意义上。一个假定存在的量子类似可能会通过把无数非重叠波包压缩在单位区间 $(0, 1]$ 来模仿无限多不同粒子的芝诺结构。这种压缩要求在接近原点的时候，在 Δx 与波包相关的不确定性变得无限小。通过不确定性原理，动量 Δp 的不确定性在接近原点的时候将变得无限大。如果相比于模拟经典行为所要求的某些特定宏观标准，Δx 和 Δp 都很小，那么这就表示了假定的量子类似的崩溃。③

5.3 量子力学如何不如经典力学具有确定性，即使它没有态矢量约化

量子态演变的决定论要么全对要么全错。如果 $\hat{H}$ 是本质上自轭的，那么就“全对”。因为正如已经注意到的，唯一自轭延伸的求幂会给出算子的幺正演化，这种演化是对希尔伯特空间的所有时间和所有矢量定义的。如果 $\hat{H}$ 不是本

① 一个适当的密集区域为 $\{\varphi\in L^2_{\mathrm{C}}(\mathbb{R}, dx):\psi, \psi'\in AC(\mathbb{R})$ 和 $\hat{H}\psi\in L^2_{\mathrm{C}}(\mathbb{R}, dx)\}$，其中 $AC(\mathbb{R})$ 代表绝对连续函数。

② 这个结果就是众所周知的加藤(Kato)理论，见[Kato, 1995, Remark 5.6]。本质自轭性及其对量子决定论的启示问题的更加详细的讨论，见[Earman, 2005]。

③ 要仿效一个在其中粒子的位置和动量都已给定的经典态，需要一个量子态 ψ，它不仅把给定值恢复为期望值，而且给出了与相关宏观标准相比要小的 $(\Delta x)_\psi$ 和 $(\Delta p)_\psi$。因为如果 $(\Delta x)_\psi$ 和 $(\Delta p)_\psi$ 分别来讲相对于标准都要大一些，那么就存在一种可预见的可能性，可能会发现粒子的位置(或者动量)的值与给定值不同。

质自轭的，就有两种可能性要考虑。第一种是 $\hat{H}$ 没有自轭延伸，这会被排除，因为 $\hat{H}$ 应当是一个实算子，而每个实对称算子都有自轭延伸。第二种可能性是 $\hat{H}$ 有很多自轭的延伸，那么就可能是"全错"。因为不同自轭延伸的求幂给出了物理上不同的时间演化。大体上讲，不同的自轭延伸对应于位形空间边界点上的不同边界条件。也许某些边界条件会被挑出来并被提升到类似法则的地位，因而提供了唯一的量子力学。但通过这种途径恢复决定论需要对这种假定的挑选和提升给出正当的理由。或者，如果能够表明与不同自轭延伸相关的量子动力学都有着相同的经典极限，那么哈密顿量的非本质自轭性的影响就会被弱化，参见[Earman，2005b]。

第二种可能性的小例子通过一个在电势 $V(x)$，$x \in \mathbb{R}$ 中的正实线$\mathbb{R}_+$上运动的粒子给出。如果电势具有 C/x^2 的形式，其中 $C>0$，那么牛顿初值问题就有唯一解，并且这个解是对所有时间定义的。直观的解释是，不论经典粒子有多大能量，它都不能越过无限高的电势从而到达原点上的奇点，同时它也不能在有限时间内到达 $x=+\infty$ 处。但是当 $c<\frac{3}{4}$时，这种情况的量子哈密顿算子在 $L_c^2(\mathbb{R}_+, dx)$ 上不是本质自轭的，参见[Reed and Simon，1975，Thm. X.10]。直观的解释是，量子粒子可以贯穿势垒到达奇点，并允许漏掉某种可能性。这种漏掉与幺正变换不相容，幺正变换将会作为一个本质自轭的 $\hat{H}$ 的唯一自轭延伸求幂的结果而得到。

在这个小例子的位形空间中的奇点完全是人为假设的，是通过删除一半实线而创造出来的。但是以裸奇点或类时奇点的形式进行的模拟在广义相对论时空中发生(参见6.4节)。有人会问，一个在负质量史瓦西(Schwarzschild)时空背景下传播的相对论量子粒子是否能够穿越围绕原点 $r=0$ 的有效势垒。霍罗威茨和马若夫(Horowitz and Marolf)[1995]指出，答案是肯定的。

哈密顿量的本质自轭性可能会被提升为一种选择定理，有助于确定什么样的系统是"量子可能的"，从而确保(除态矢量塌缩之外)量子演化是决定论的。那些将决定论看作是一种先验真理的人可能会看好这样的提升，但是其他人却很难看到它的优点。

5.4 量子力学中的混沌和可预言性

量子力学不仅能比经典力学具有更强的决定性，而且还能具有更强的可预言性。如果观测者想要预言的系统的行为敏感地依赖于初始条件，而他们又不能完全精确地确定初始条件，那么经典可预言性就要妥协或者完全被破坏。但是如果量子哈密顿算子是本质自轭的，那么不仅量子态的演化是完全决定论的，它的可预言性也不会因为对初始条件的敏感依赖而妥协。关键在于演化的线性和幺正的本质保持了希尔伯特空间的规则：$\| U(t)\psi_2 - U(t)\psi_1 \| = \|U(t)(\psi_2 - \psi_1)\| = \|\psi_2 - \psi_1\|$。总之，如果两个态在开始时就很接近(正如希尔伯特空间规则所测量的)，那么它们会一直保持接近，也就是说，演化是稳定的。

这种稳定性给所有寻找量子力学自身中的混沌的人带来了麻烦——他们被导向这样的极端：诉诸于开放系统(其演化不是幺正的)或者隐变量(它们的演化是不稳定的)。① 但是在其自身之中，量子演化的稳定性没有为解释量子动力学处于某种适当的经典极限时混沌是如何发生的设置任何先验的壁垒。因为那个计划只要求表明相关经典量的期望值能够足够快地发散开(在某种合适的度规下)，从而能够保证形成在宏观尺度上观测到的混沌行为的更高阶的遍历特性，参见[Belot and Earman(贝洛特与厄尔曼)，1997]。实施这个计划的一种显而易见的方法，就是利用埃伦费斯特定理(Ehrenfest's theorem)来表明，在位置表征中，只要波包的均方根宽度保持足够小，量子波包的质心就会遵循经典的轨迹。然而，对于经典的混沌轨迹，后一个条件在其中成立的区间太短了。例如，[Zurek(朱瑞克)，1998]估计，对于木卫七(木星的一颗卫星)的混沌翻转，它的周期只有 20 年。已经有作者提出说量子消相干可以挽救这个问题，参见[Zurek，1998；2003]，但这个话题超出了本章讨论的范围。很明显，经典混沌对我们关于经典世界是如何从量子物理中突现的的理解提出了挑战。② 经

---

① 见克隆兹(Kronz)[1998]、库欣与鲍曼(Cushing and Bowman)[1999]的讨论。相比之下，研究"量子混沌"的物理学家们并不尝试去发现量子力学自身的混沌，而是要去研究其经典对应显示出混沌的量子系统的区别性特性。为此，麦克·贝里(Michael Berry)建议用"量子混沌学"代替"量子混沌"。不幸的是，这条建议没有生效。

② 关于这个问题的综合审视，参见兰兹曼在本书中的论述。

典—量子相对应的另一方面将在下一节进行探讨。

5.5　态矢量约化、隐变量等

表明一个量子系统的哈密顿算子 $\hat{H}$ 是本质上自轭的还不足以保证这个系统中决定论的命运，原因有两个。第一个是，量子态的决定性演化可能因“态矢量约化”而中断。正如量子力学测量问题的一些处理中假定的那样，通过这种态矢量约化，幺正演化 $\psi(0)\mapsto\psi(t)=\exp(-i\hat{H}t)\psi(0)$ 暂停了，且量子态跃入了被测可观测量的一种本征态。以它最粗糙的形式来看态矢量的约化是一个毫不夸张的奇迹——一种对自然法则的违背——使得它成为一个不适合在这里讨论的话题。但是还有更复杂的态矢量约化的形式，这些形式将约化看作一种动力学的过程。吉拉迪等(Ghirardi *et al.*)[1986]和佩尔(Pearle)[1989]分别研究了约化非连续和连续发生在其中的随机模型。佩尔在1976年研究了一种通过对薛定谔方程附加一个非线性项而得到的约化。如果约化的随机模型是正确的并且如果它们假定的随机机制表示不可约化的随机性，那么很明显决定论就被打破了。相比之下，佩尔[1976]的方案依靠隐变量获得了一个决定性的态矢量约化。① 标准非相对论量子力学的所有这些选择都没有被推广为可行的相对论量子场论，据我所知，它们在弦论或圈量子引力的量子引力主流研究路线上都没有起到任何作用(参见第8节)。因此，目前对量子力学可能修正的决定论含义的思考看上去并不多产，这种修正也可能不会成为某种未来物理学的一部分。然而，这里引入态矢量约化的动机意义重大，因为它导致了关于为什么传统量子态动力学可能不能充分保证量子范围中的决定论的第二组理由。

经典(即非量子的)理论公开展示了它们的解释。② 例如，对一个允许(不受约束的)哈密顿公式的经典理论，可观测量与从相空间 $\Gamma(q, p)$ 到实数 $\mathbb{R}$ 的函数都是一一对应的。预期的解释是，如果 f_O 是对应于可观测量 O 的函数，

① 隐变量是相位角，这种思想为安克思和柯亨(Ax and Kochen)[1999]所接受，见下文。

② 但是从第3节开始回顾，如果决定论是硬性要求，那么一开始表面上的解释就可能不得不通过规范自由度进行修正。

那么 O 在 t 处的值是 o，当且仅当 t 处的态 $(q(t), p(t))$ 使得 $f_O(q(t), p(t)) = o$。为了允许只在某几个态上有确定值的确切可观测量，这个方案可以自由化。对于这样的一个 O，表征函数 f_O 只是一个部分函数。另一种自由化是为了允许 f_O 的范围包括“模糊”值(例如间隔)。为了用类似的方法得到量子力学的一种解释，需要在上述形式上至少附加两个东西：(i)对应于量子可观测量的自轭算子的描述；(ii)一种对于量子可观测量的表达，其形式为确定客观测量在什么条件下取什么值的赋值规则。我将简单地假设解释问题的(i)部分已经被解决了。

提供(ii)部分的最明显的方式可能是模仿经典的赋值规则，用量子态空间代替经典态空间 $\Gamma(q, p)$，从而得到如下形式的赋值规则：量子可观测量 O 在 t 点的值是 o，当且仅当对态矢量 $\psi(t)$，有 $f_O(\psi(t)) = o$。其中 f_O 是量子可观测量 O 的表征函数。如果像这个公式中隐含的假设的那样，量子态空间被看作是系统希尔伯特空间 $\mathcal{H}$ 中的单位球体 $\mathbb{S}\mathcal{H}$，即 $\{\psi \in \mathcal{H}: (\psi, \psi) = 1\}$，那么就标准量子力学而言，规范自由就出现了。因为 $\mathbb{S}\mathcal{H}$ 的任意两个由于相位因子而不同的元素对应于同一个量子态，原因在于可观测量的所有期望值对于两个量子态而言都是相同的。这种规范冗余可以通过将态空间作为投影的希尔伯特空间 $\mathbb{P}\mathcal{H}$ 而去除。$\mathbb{P}\mathcal{H}$ 通过行为 $\psi \to \varsigma\psi$ 定义为 $\mathbb{S}\mathcal{H}$ 的商数，其中 $\varsigma \in \mathbb{C}$ 且 $|\varsigma| = 1$；等价地，$\mathbb{P}\mathcal{H}$ 是射线的空间或 $\mathcal{H}$ 的一维子空间。因此，从传统量子力学的观点来看，当单位矢量 ψ' 和 ψ 属于同一条射线的时候，赋值规则应当遵从 $f_O(\psi) = f_O(\psi')$ 的限制。允许赋值规则依赖于相位就相当于在下面要用的术语中引入“隐变量”。

在任何情况下，如果量子赋值规则采取了讨论中的形式，而且如果5.3节中讨论的问题取消了，那么可以证明，量子力学是一个决定性的理论，并且即便 f_O 是一个部分函数(即对某些量子态没有定义)或者 f_O 能取模糊值，它也是一个决定性的理论。由于没有假定态矢量约化，$t=0$ 时的态 $\psi(0)$ 唯一地决定任何 $t>0$ 时的态 $\psi(t)$。同时假设(ii)部分在讨论中得到了实现，那么 $t=0$ 时的态 $\psi(0)$ 唯一地决定之后任何时间 $t>0$ 时的赋值。一个可观测量在 t 时刻没有被赋值或被赋予一个模糊值并不表示决定论的失败，决定论只要求法则加上决定可观测量现在和将来值的初始态，直到这样一种程度：这些值是完全决定

性的。那么，以现在的观点，把科亨—斯佩克(Kochen-Specker)定理和其他随后的不可行的结果看作是表明了量子力学不允许决定性的解释，这是错误的。在一定程度上，这些不能继续的结果表明，在遇到某些自然约束时，① 有些量子可观测量的子集不能全部被同时赋予明确的值。② 对于被最佳地解释为柯亨—斯佩克的不可行结论向一个甚至更小的客观测量集的扩展的贝尔(Bell)类型的定理，情况也是一样。参见[Fine(法因)，1982a；1982b]。

讨论的这一类赋值规则的一个例子是本征值—本征向量规则。它所说的是，对于一个其量子算子 $\hat{O}$ 有着离散谱的可观测量 O 来讲，O 在 t 处拥有明确的值，当且仅当 $\hat{O}\psi(t)=o\psi(t)$，在这种情况下 $O(t)=o$。但正是这个本征值—本征向量的联系导致了量子力学中臭名昭著的测量问题，表现为理论不能解释为什么测量有明确的结果。并且正是这个问题激发了态矢量约化的思想。本质上，问题的产生是因为人们坚持“测量”不应被看作一个简单的术语，而应当在量子力学自身中被当作客体系统和测量工具之间的物理相互作用来进行分析。但是尽管把标准线性和幺正的动力学应用到综合的主体系统 + 测量仪器系统，从而能够在可观测物体的本征态和测量仪器的“指针可观测量”的本征态之间建立一种一对一的关系，但特征向量—本征值规则在测量综合态函数上的应用产生了一种不可接受的结果：测量工具上的“指针”并不指示任何明确的值，参见[Albert，1992]对细节的解释。薛定谔猫的悖论就是对这个难题的残酷说明，在其中“活猫”和“死猫”相当于“指针的位置”。

因此，如果保持本征值—本征向量的联系，就有吸引人的理由去考虑对标准量子力学动力学进行调整，以确保在测量相互作用中量子态演化成为相关可观测量的一个本征态。但是由于上面做出的决定并不处理这些修正，那么随后

---

① 例如很自然就会要求，如果 O 的量子值分配规则分配给 O 一个确定的值，那么这个值就位于对应于 O 的算符 $\hat{O}$ 的谱上。很自然也会要求对于适当的函数 g，有 $F_{g(O)}=g\circ F_O$。

② 从盖里森(Gleason)定理可知，服从值的分配上的很合理的约束，不是所有的三维希尔伯特空间中的自轭算子都能同时被分配以属于这些算子的谱的确定值。柯亨—斯佩克定理表明，一组有限的量子可观测量可以得到相同的结论。这些不可行结果的说明参见[Redhead，1987]。

的讨论将限制在其他选择上，即一种本征态—本征矢量联系的赋值规则的使用可能还会有补充量子态的“隐变量”的帮助。如果运用了隐变量 X，那么赋值规则就采用了以下形式：量子可观测量 O 在 t 处的值就是 o，当且仅当对总态 $(\psi(t), X(t))$ 有 $f_O(\psi(t), X(t)) = o$，其中 f_O 又一次表示了可观测量 O 的表征函数，但是现在这个函数定义在增加态空间上。如果总态的演化是决定性的，那么和之前一样，如果量子力学为真，量子的领域就是完全确定的。玻姆(Bohm)解释提供了一个例子。在玻姆解释中，$X(t)$ 指定了粒子在 t 时刻的位置。总态的量子组分按照薛定谔动力学进化，并且假设的粒子位置的运动方程确保了对 $t>0$，$(\psi(0), X(0))$ 唯一决定了 $(\psi(t), X(t))$，参见[Holland(霍兰德)，1993]。在玻姆解释中，很多量子可观测量有着决定性特征，只有在充分明确的背景(典型地包括一个带有隐变量的测量工具)中才取确定的值。比如，在斯特恩—盖拉赫(Stern-Gerlach)实验背景中，一个自旋为 1/2 的粒子只在粒子位置处于仪器的合适区域时才会自旋向上(或自旋向下)。因此，玻姆解释解决了测量问题这个断言的有效性，使得人们对是否所有测量都能被约化为位置测量感兴趣。

各种量子力学模态解释也试图解决测量问题，通过打破本征值—本征向量的联系并使它们分开得足够远从而允许测量有确定的结果，但又不至于太远而与柯亨—斯佩克型的不可能性结果产生冲突，参见[Bub(巴布)，1997]和迪克森在本书中的概述。但是与玻姆解释相比较，模态解释对维护决定论没有做任何承诺。概略地说，其思想是，一个与处在态 $\psi(t)$ 的复合系统的子系统相联系的可观测量，只有在子系统的约化密度矩阵是一个与可观测量的本征基相联系的投影的加权总数时，才会具有确定的值。这保证了在一个理想的无干扰测量相互作用中，也就是在其中测量工具的指针位置与被测客体系统可观测量的可能值之间的关系非常清楚的相互作用中，指针可观测量和客体系统可观测量都有着明确的值。①

① 更一般地，一个系统与其环境的相互作用将意味着对系统的“测量”一直都在进行。因此退相干通过提供解释的可应用性条件帮助了模态解释。在其他方向上，退相干要求某些类似于模态解释的东西，否则就与它的支持者的断言相反，它不能解决测量问题。更多关于退相干的内容见兰兹曼在本书中的论述。

假定可观测量有适当的条件可以具有确定值，那么大部分形式的模态解释为可观测量提供了有着特定值的可能性。但是在谈及真实值是什么的时候，它们就沉默了。不过量子可观测量实际拥有的值能够被用来扮演隐变量 X 的角色，人们也可以询问是否能赋予总态(ψ，X)一个决定性的动力学。对于哲学文献中讨论的模态解释来说，答案是否定的，因为这些论点不能为决定论提供足够的隐变量。例如，在 $t>0$ 时，当一个理想的测量相互作用得以完成，并且指针位置的本征态与客体可观测量的本征态有着清晰关联时，标准模态解释会断言客体可观测量和指针可观测量都有确定值。在实验的不同时段中，这些相关联的可观测量有着不同的值。但是在所有时段中，初始量子态 $\psi(0)$ 是相同的，实验情况也可以设定为能使模态解释可以断言初始拥有的值 $X(0)$ 是相同的。决定论的这种失败对于模态理论家来说无关紧要。他们的目标是要解决测量问题。为达到这个目标，表明对于拥有的值来说存在一个与量子力学统计预言相容的随机动力学就足够了。事实上，存在大量这样的动力学，参见[Dickson，1997]和[Bacciagaluppi and Dickson(巴奇伽鲁皮和迪克森)，1998]。

安克思和柯亨[1999]提出的不同版本的模态解释采用了上面论述的观点，把射线$\mathbb{P}\mathcal{H}$的标准量子态空间延伸到单位矢量$\mathbb{S}\mathcal{H}$。前者的元素被假定描述了系统的统计效果，而后一种元素则描述了个体系统。这种延伸允许模态解释去确定当一个可观测量处于具有确定值的环境中时，它具有什么样的值，同时也为扩展量子态提供了决定性的演化。人们假定对应于射线$\varsigma\psi$，$|\varsigma|=1$ 的总体，是由有着相位因子ς的个体系统构成的，这些个体系统有着初始一致的随机分布，这就为决定论的明显失败提供了说明。

玻姆解释和各种模态解释面对相对性考虑时都会遇到困难。前者没有对量子场论进行任何自然概括，至少它没有严格地采用在量子场论领域中基本实体和粒子是场行为的副现象这一教义。后者确实拥有对量子场论的自然概括，但它的结果令人遗憾：在标准讨论中，子系统的可观测量都没有明确的值，参见[Earman and Ruetsche(厄尔曼和鲁切)，2005]及其中的参考文献。

量子力学多世界解释可以做出一个字面和形象的解读，参见[Barrett(巴雷特)，1999]中的概述。按照字面解读，因为时空分裂为许多分支，确实存在很多世界。这些分支从分开的时刻起，就不再与对方有任何拓扑的联系，参见

[McCall(麦考尔), 1995]。[1] 这种形式的多世界解释可以通过以下途径被描述为一个隐变量解释：把隐变量 X 看作是描述了时空分支，把表征函数 f_O 看作是从总态 $(\psi(t), X(t))$ 到一个矢量的映射，这个矢量可能会有无数个标记为 α 的组分，它们提供了在 α 分支中 t 时刻的 O 值。这样，决定论的命运就取决于分支在什么时候如何发生和是否会使得总态 $(\psi(t), X(t))$ 的演化成为决定论的。在“多世界”的形象解读中，字面上只有一个世界，但存在许多思想、背景、希望等，而且不存在诸如一个无限制的可观测量 O 这样的东西，而是存在一个语境 α 中的可观测量 O，用 O_α 表示。如果表征函数 $f_{O\alpha}$ 只是量子态的函数，那么决定论看起来就得到了保证。但是，在掩饰对任何给定时间都可能发生的对背景的说明的需要时，我们的符号系统是有缺陷的。如果存在“基础的民主”，也就是，$\psi(t)$ 的任何“分支”都被表示为任何定义一个情况的系统的希尔伯特空间正交基的矢量线性组合时，这种对背景的说明只决定于量子态 $\psi(t)$。这样一种激进的民主看起来与经验并不相容。比如说，在薛定谔猫的实验中，我们或者会看到一个活猫或者会看到一个死猫，但我们永远都不会看到一个活猫和死猫的叠加。[2] 为了克服这个困难，多世界理论家提出运用一组优先基。这样，决定论的问题就转移给对优先基的说明是否是决定论的这一问题。即使多世界解释——不管是字面还是形象的版本——保证了本体论的决定论，其代价似乎是带来了一种激进的认识论非充分决定性：我如何能够知道我处于分裂世界的哪一个分支或者非分裂世界的哪一种情况中？告知不存在“我”而只存在“分支 α 中的我”，或者“情境 α 中的我”，这在我——不管是哪个我——必须对一个测量做出预言的时候是没有帮助的。在这里，我全部所能做的就是退回到量子力学的统计算法中去。多世界解释看起来能够确保即便世界是本体论决定论的，就人们所能判断的来说，它表现得就仿佛存在一个不可约化的随机性。

尽管关于量子测量问题及其结果的讨论非常粗略，但我相信这已经足以表

---

① 如何在传统时空理论的背景中描述分支时空是一件棘手的事情。也许最有希望的行动就是坚持时空是一个微分流形这一假设而放弃它是一个豪斯道夫(Hausdorff)流形的假设。然而，非豪斯道夫流形能显示出多种威胁决定论的反常特性，例如测地线能够分叉，见6.1节。

② 但是我们如何确信？也许短暂的精神混乱是一种叠加现象。

明为什么想要得到量子力学的真理是否会需要决定论为假这一问题的简单明了的答案是徒劳的。为了得到这个问题的答案，需要辨别各种各样的竞争中的量子力学解释，这是一个绝不简单的任务，尤其是判断不可避免地要受到对决定论的态度的影响。

6. 经典广义相对论中的决定论

6.1　爱因斯坦的革命

爱因斯坦的广义相对论在很多方面都是革命性的，但是就我们当前的目的来说，首先最重要的革新在于广义相对论扫除了——所有前广义相对论理论共有的——粒子和场的演化都发生于其上的固定时空背景的概念。在广义相对论中，时空度规是一个动力学场，它的演化受爱因斯坦引力场方程的支配。在讨论此演化是否是决定性的之前，需要注意两点。

首先，广义相对论物理学家代表性地假设了相对论时空$\mathcal{M}$，g_{ab}的流形$\mathcal{M}$是豪斯道夫的。[①] 没有这项约定，决定论将会陷入困境。例如，非豪斯道夫时空能够允许一个分离的测地线；就是说，比如存在从$[0, 1]$到M的光滑映射γ_1和γ_2，使得影像曲线$\gamma_1[0, 1]$和$\gamma_2[0, 1]$是测地线，并满足在$[0, b)$上$0<b<1$，但是它们却有着不同的终点$\gamma_1(1)$和$\gamma_2(1)$。根据广义相对论，大量不受除引力之外的力影响的测试粒子的世界线是类时测地线。但是这样的粒子怎样知道该走测地线的哪条分支？另外，如果允许非豪斯道夫附加条件，那么在下面6.3节中讨论的爱因斯坦场方程初值问题(局域)解的唯一性将会失效。

其次，读者会想到，我们的关注点被限制在时间定向的相对论时空M，g_{ab}上，并且假设其中一个方向被挑选出来作为时间的方向。但即使在空间上附加这种限制，某些以爱因斯坦场方程的解体现的时空仍不利于2.1节给出的全域拉普拉斯决定论的明确表达。例如，这样的时空可能不允许全域时间函数。事

① $\mathcal{M}$是豪斯道夫的，当且仅当对任意的p，$q\in\mathcal{M}$(其中$p\neq q$)，有邻域$M(p)$和$N(q)$使得$N(p)\cap N(q)=\varnothing$。当然，一个流形是局部欧几里得的，因此是局部豪斯道夫的。

实上，戈登宇宙哲学模型的时空不仅不允许全域时间函数，甚至连单个全域时间片段(类空的无边超曲面)也不允许，以至于人们不能有意义地论及给定瞬间的宇宙。①

有一种回应将会缩小广义相对论的物理可接受模型的类别范围，它要求除满足爱因斯坦方程之外，这些模型也必须符合时空的全域因果结构，而这种结构会排除那些畸变的模型如哥德尔模型和其他包含封闭类时曲线的模型。这种运动的动机独立于“时空旅行佯谬”。② 但是为了确定广义相对论中拉普拉斯决定论的全域观点，需要强得多的因果关系条件。

从一开始，为了使拉普拉斯决定论的概念可以延续到广义相对论时空的背景中，就要求时空允许一个全域时间函数。这个函数并不是由于缺乏封闭的类时曲线而发生的。但是即使是对全域时间函数的要求，也不足够有力，因为它并不能保证任何这种函数的位势面都会拥有柯西特性。为了切题，拉普拉斯决定论要求全域双曲的时空$\mathcal{M}$和 g_{ab}是两个条件的结合：第一，$\mathcal{M}$和 g_{ab}必须是强因果性的，因为对于任何 $p\in\mathcal{M}$以及 p 的邻域，存在一个子邻域，使得一旦一个指向未来的因果曲线不在了，那么它就永远不会再出现(直观上说，不存在几乎封闭的因果曲线)。第二，对于每个 p，$q\in\mathcal{M}$，因果方块 $J^+(p)\cap J^-(q)$ 是紧致的。③ 全域双曲性确保$\mathcal{M}$，g_{ab}能被柯西表面分成叶片状，并且$\mathcal{M}$是微分同胚地$\Sigma\times\mathbb{R}$，其中如果 $\dim(\mathcal{M})=n$，Σ是一个 $n-1$ 维的流形。但是只约定全域双曲性有投机取巧之嫌，所以，让我们看看诚实劳作会获得什么吧！

6.2 广义相对论中的决定论和规范自由

对于前相对论理论，一个永恒的主题就是创造一种对决定论友好的环境，

---

① 这是戈登时空三个特点的结果：它是时间上定向的(即它允许一个连续不消失的类时矢量场)，它是单连通的并且每个时空点有一条闭合的未来指向的类时曲线通过。关于戈登解的描述，见[Hawking and Ellis, 1973, pp. 168—170]和[Malament, 1984]。

② 但是见厄尔曼、思蒙克(Smeenk)和乌斯雷彻(Wüthrich)[Earman *et al.*, 2005]，它认为所谓时间旅行佯谬并没有表明时间旅行是概念上或物理上不可能的。

③ $J^+(p)$表示 p 的因果未来，相对地 $J^-(p)$表示 p 的因果过去，即所有点 q 的集合使得存在一个从 p 到 q(相对地，从 q 到 p)的未来指向的因果曲线。

这要求自发地去补充背景时空的结构，或者去发现吸收表面非决定论(回想 3.3 节)时起作用的规范自由。但是在广义相对论中不存在固定的背景结构，因而人们会认为广义相对论要么产生了非决定论，要么存在一个非平凡的规范对称在其中起作用。这种观点没有令人失望。

要明白其缘由，有必要更加详细地对待如下的爱因斯坦方程：

$$R_{ab}-\frac{1}{2}Rg_{ab}+\Lambda g_{ab}=\kappa T_{ab} \tag{16}$$

其中 R_{ab}和 $R:=R_c^c$ 分别是李奇张量(用 g_{ab}及其微商的术语定义)和李奇标量，Λ 是宇宙学常数，T_{ab}是应力—能量张量。就本文的目标来说宇宙学常数可以忽略，但是它是当前宇宙学中人们的兴趣所在，因为一个正的 Λ 是推动宇宙加速膨胀的“暗能量”的候选者之一，参见埃利斯在本书中的论述。

理论的一种潜在模型是一个三元组$\langle\mathcal{M},\ g_{ab},\ T_{ab}\rangle$，其中 g_{ab}和 T_{ab}在$\mathcal{M}$的所有点上都满足(17)。建立这样一种模型看似太简单了：从任意的广义相对论时空$\mathcal{M}$，g_{ab}开始，计算爱因斯坦张量 $G_{ab}:=R_{ab}-\frac{1}{2}Rg_{ab}$，通过 $T_{ab}:=\kappa G_{ab}$定义应力—能量张量。因此，人们必然把 T_{ab}理解为源自一个已知的物质场。并且为了使环境尽可能对决定论友好，假设式(17)右边插入的 T_{ab}会实现具有支配性的能量条件(参见 4.1 节)，这个条件与局域守恒定律 $\nabla^a T_{ab}=0$ 一起，确保了物质—能量不会比光传播得快。

即使具备了这些可能的规定，乍一看似乎决定论并没有受到牵动，至少如果对理论模型给出一个朴素实在论的解释的话，情况就是如此。通过重复 3.2 节中给出的构造的一种变化，这种困难会变得明显起来。令$\langle\mathcal{M},\ g_{ab},\ T_{ab}\rangle$为一种满足上述所有规定的模型，并且假定时空$\mathcal{M}$，$g_{ab}$满足所有你喜欢的因果性条件，比如，它是全域双曲的。由于不存在固定的背景结构，除了$\mathcal{M}$的拓扑和可微结构，人们可以自由地选择微分同胚 d：$\mathcal{M}\rightarrow\mathcal{M}$，使得 d 是$\mathcal{M}$，g_{ab}的某些柯西面Σ上向过去的恒等映射。那么$\langle\mathcal{M},\ d^*g_{ab},\ d^*T_{ab}\rangle$(其中 d^* 表示通过 d 的拖拽)也将是一个满足$\langle\mathcal{M},\ g_{ab},\ T_{ab}\rangle$所有规定的模型。通过构造模型，对于所有在$\Sigma$之上或其过去之上的 p，都有 $d^*g_{ab}(p)=g_{ab}(p)$和 $d^*T_{ab}(p)=T_{ab}(p)$。但是对有些Σ的未来的点 p，有 $d^*g_{ab}(p)\neq g_{ab}(p)$，$d^*T_{ab}(p)\neq T_{ab}(p)$

(除非我们已经不经意地选择了一个d是g_{ab}和T_{ab}的对称，否则这就总是可以避免的)。这种对过去所有时间都相同但在未来却不同的模型的存在，即便对拉普拉斯决定论最弱的近亲理论来说，都是一种违背。至少如果从表面上理解理论的表层结构(即流形上的张量场)的话，情况就是这样的。

当这种类型的构造会对前广义相对论环境中的决定论产生威胁时，有两种选择可以支撑决定论：对背景时空增加更多的结构或者抛弃时空的容器观点。第一种选择只有在时空结构的附加元素是非动力学的对象时才有效，但这代表着从爱因斯坦革命的一种关键特征的后退。在当前情境中，抛弃时空容器观点的选择，其表现形式是抛弃朴素实在论，这种实在论把理论解读为描述了流形上的张量场。

第二种选择是基于一个原理化的动机，它的出现不是为了挽救广义相对论中的决定论，而是遵循了前广义相对论理论中规范对称性的处理。广义相对论场方程(17)是源自一种作用原理的欧拉—拉格朗日方程。这种作用原理允许微分同胚群作为一种变分对称群。这样，诺特第二定理得以应用，并标志着我们拥有一个非充分决定性的例子——比起独立的场方程来说，还有更多“未知”的东西——并且时空变量的任意函数都会在场方程的解中显示出来。

从拉格朗日公式转向哈密顿量公式就会发现，正如预期的一样，广义相对论是一个约束哈密顿理论。第一类约束有两种，动量约束和哈密顿约束。① 不幸的是这些约束的泊松括号代数不是李代数，② 因此不能直接与微分同胚群产生联系。微分同胚群是作用于时空的，存在一个通过在约束代数中找到微分同胚群的李代数的自然同态而作用于哈密顿相空间的群。这个小问题被艾莎姆和库查尔(Isham and Kuchař)[1986a, 1986b]克服了，他们表明，如果用适当的嵌入变量及其共轭动量来扩大相位空间，那么扩大了的约束代数就是一种李代数，并且存在一个时空微分同胚群李代数在新的约束代数中的同态。因此，可以运用处理规范对称性的标准工具，并产生这样的结果：广义相对论的微分同

---

① 使用复数是因为对空间每一点都有动量约束和哈密顿约束。

② 一对约束的括号不总是“结构常数”相乘得到的约束的线性组合。约束代数不能形成李代数意味着广义相对论不是杨—米尔斯意义上的规范理论。但是它肯定不是说广义相对论不包含非平凡的规范自由度。

胚不变性被解释为一种规范对称性。按照这种解释，上述构造并没有证明广义相对论是非决定论的，而是人为地打破了决定论，它在表层结构模型和它们所表述的内禀物理情形之间的多对一关系的意义上利用了表层结构理论的冗余。特别地，上述构造中的模型$\langle \mathcal{M}, d^* g_{ab}, d^* T_{ab}\rangle$和$\langle \mathcal{M}, g_{ab}, T_{ab}\rangle$并不能威胁到决定论，因为它们可以被解释为相同物理情形的不同描述。当然现在讨论的工具已经使之成为对决定论的保证，不能因为它在广义相对论上的应用部分证明了广义相对论的正确性就认为它是一种决定性理论。这里唯一可断言的是，这种对广义相对论的决定论的挽救的行动不是独立的，而是规范对称性系统方法的一部分，被用来产生前广义相对论理论的“正确”结果。①

现在如此清楚的事情，爱因斯坦却花费了多年的努力才理解。他极少受到关注的“洞问题”可以看作是非充分决定性问题的发现。② 使问题复杂化的是广义协变性的两种含义。正规广义协变性要求运动/场方程要被写作一种形式，在这种形式下，它们在任意坐标变换下是协变的。术语“正规”的选择是带有预谋性的。因为正规广义协变性的要求更多的是在理论的形式而不是内容方面。例如，牛顿理论和狭义相对论可以重新规范而不用做内容上的改变就能满足这种要求。事实上，牛顿理论和相对论能够以完全独立于坐标的方式得到规范这一事实应该可以说明坐标无关紧要。③ 实体的广义协变性要求微分同胚不变性，例如对于$\mathcal{M}$的任意微分同胚，如果$\langle \mathcal{M}, g_{ab}, T_{ab}\rangle$是理论的一种模型，那么$\langle \mathcal{M}, d^* g_{ab}, d^* T_{ab}\rangle$也是。并且这种微分同胚不变性是一种规范不变性。同样，术语“实体”的选择也带有预谋性，由于实体的广义协变性并不能不改变内

---

① 因为在哲学文献中这些事情都被如此多的争论围绕着，我想要尽可能地强调，我不是要提出看待广义相对论的新方法，而只是想阐明在广义相对论物理学家们眼中什么是标准的观点，比如见[Wald, 1984]。

② 对于在爱因斯坦引力场方程的研究中如何用图描述“洞问题”的说明参见[Norton, 1984]和[Stachel(斯塔彻), 1986]。罗韦利(Rovelli)在本书第十二章描述了思考“洞问题”的教训如何影响了他对经典广义相对论的理解和他解决量子引力的方法。

③ 在以上陈述中我已经有意地使用“抽象标志”标记法。因此，g_{ab}代表一个对称协变张量场，以无坐标的形式定义为一对正切矢量向$\mathbb{R}$的双线映射。这个客体可以用它在坐标系$\{x^i\}$中的坐标分量g_{jk}来表示。这样两种表述之间的变换是规范变换，尽管是数学上最简单的那种。

容就自动获得，而对于正规广义协变的牛顿理论和狭义相对论，在处理这里谈及的规范对称性的时候至少不需要通过工具，参见[Earman，2006]。

是如下事实导致混淆的出现，即时空坐标变换可以或者被用来指示时空点的重新标记，或者被用来指示(局域的)微分同胚。在第一种情况下，这些变换都是简单的规范变换：因为它们，在不同坐标系上采用不依赖于坐标的客体g_{ab}和T_{ab}的组分而得来的这些内在客体的各种坐标表示联系了起来。但是在这里不存在关于广义相对论的任何新东西，因为对于前广义相对论理论不依赖于坐标的表示，同样的故事也成立。然而，在第二种情况下，这些变换可能是也可能不是规范变换——这取决于理论的内容。

当通过“洞问题”首先发现了非充分决定性问题时，爱因斯坦将它解释成一个真实而又不能接受的非充分决定性形式。为了避免这种情况，他认为不得不抛弃引力场方程组必需的正规广义协变性。在非协变方程组的荒原中徘徊了将近3年之后，他才重新拾起了广义协变性。为了使这个想法重新有效，爱因斯坦不再使用规范对称性的语言，所以他并没有说广义相对论的规范解释降低了决定论的门槛，因为它只需要规范不变量演化的唯一性。但他定义了什么才是相同的事物；或者他为了广义相对论规范不变量的子类而这样定义——什么被叫作“点一致性”，即类似于光线的交叉的东西。①

很多哲学家都已经以不同的方法跟随了爱因斯坦的研究之路。然而几乎没有人能说出这条路指向哪里。如果广义相对论中的决定论通过将规范对称性看作微分同胚协变量来处理而得以保留，那么广义相对论中唯一的“可观测量”(即真正的物理量)是规范不变量。这说起来很简单，但是作为表层结构基础的规范不变结构的本质究竟是什么呢？对于那些通过运用某种广义相对论正则量子化的计划来追寻引力量子理论的物理学家来说，这是一个关键问题。因为按照这个计划，正是经典广义相对论的“可观测量”将在量子引力的希尔伯特空间自轭算子的意义上被转化为量子可观测量。对于如何最好地描述经典广义相对论的可观测量，不存在标准答案。但有一件事是确定的：教科书中所用的熟悉的量，甚至是像场方程(17)中出现的里奇标量那样的标量曲率不变量，都不能

① 爱因斯坦对此术语的使用，见[Howard(霍华德)，1999]。

在我们讨论的意义上被当作是可观测量。更加特别地，没有任何局域量——与时空点或有限区域联系的量——是规范不变的。从这个意义上讲，理论规范自由的内容有着非实体论的意味。这个内容是否能用一种也能满足传统相对论者顾虑的方法来描绘仍要拭目以待。

第二种紧密相关的含义涉及时间和变化的本质，它将广义相对论的微分同胚当作规范对称处理了。在哈密顿表征下，这种含义具有"冻结的动力学"的形式。应用到广义相对论的哈密顿约束，第一类约束条件促成的规范变换学说直接导致的结论是，在广义相对论中运动是纯粹规范的。采用另一种方法，这个理论的哈密顿构想中的瞬时态包含了多余的结构，并且如果在两种态中有一种是通过解哈密顿形式的爱因斯坦方程而从另一种得到的，那么任意两个这样的态是同一内在规范不变情形的等价描述。①

对于那些不接受这些暗示的人来说，3.3 节中提到的非正统运动可能具有吸引力。但是据我了解，将这类应用到广义相对论上的非正统说法还没有被物理学阵营认真对待。

6.3 广义相对论中的初值问题

为了简单起见，考虑无源或是真空的爱因斯坦方程的初值问题。无源或真空的意思是指在式(17)中 $T_{ab}\equiv 0$。由于这些方程在时间中是二阶的，可以推出适当的初始数据由某些给定时刻的时空度规及其一阶时间导数的值构成。这种思想的技术公式采用一组初始数据来构成一个三元组(Σ，h_{ab}，k_{ab})，它具有以下特征和解释意图：Σ是一个三维流形，它作为时空$\mathcal{M}$，g_{ab}的类空超曲面嵌入。h_{ab}是Σ上的一个平滑黎曼(Riem)度规，当Σ作为$\mathcal{M}$，g_{ab}的类空超曲面嵌入时，它与时空度规 g_{ab}在Σ上引起的度规一致。k_{ab}是Σ上的一个平滑对称张量场，当Σ作为$\mathcal{M}$，g_{ab}的类空超曲面嵌入时，它与 h_{ab}的法向导数一致。一个拥有了所有这些角色的时空$\mathcal{M}$，g_{ab}被称为初始数据集(Σ，h_{ab}，k_{ab})的发展。如果初始数据集(Σ，h_{ab}，k_{ab})的发展$\mathcal{M}$，g_{ab}满足无源爱因斯坦方程，那么h_{ab}

---

① 更多关于广义相对论和量子引力中时间问题的讨论，见[Belot and Earman，1999]及贝洛特、罗韦利在本书中的论述。

和 k_{ab}就不能被任意指定，而是必须满足一组约束方程。无源爱因斯坦方程的存在和唯一性会导致以下形式的无源爱因斯坦方程①：令(Σ , h_{ab} , k_{ab})为一个满足约束方程组的初始值集，那么就存在初始数据的唯一发展$\mathcal{M}$, g_{ab}——取决于微分同胚——满足无源场方程的最大柯西发展。而且，g_{ab}连续地依赖于初始数据，参见[Howking and Ellis, 1973]相关拓扑的细节。

正如均匀性麦克斯韦方程初值问题适定性的证明探索了电磁势中的规范自由(见4.2节)，爱因斯坦方程的存在性和唯一性也同样探索了微分同胚不变性是广义相对论的一种规范对称性的思想。当度规势 g_{ij}(即度规 g_{ab}的坐标组分)遇到规范条件(称为和谐坐标条件)时，爱因斯坦方程采用了准线性对角二阶双曲偏微分方程的形式，人们已经知道这种形式有着局域适定的初值问题。

给定初始数据的发展$\mathcal{M}$, g_{ab}是一个柯西发展，这意味着Σ是$\mathcal{M}$, g_{ab}的一个柯西面(因此，这个时空是全域双曲的)。它是最大的柯西发展意味着不存在$\mathcal{M}$, g_{ab}的适当延伸，这样的延伸是无源爱因斯坦方程的解，并且对此Σ是一个柯西面。从上一小节对规范自由的讨论开始，我们就期望满足取决于微分同胚的唯一合格者，相应地这个合格者表明这里给出的启发式讨论可以有精确的内容。这里，合格者的意思是，如果$\mathcal{M}'$, g'_{ab}是满足无源爱因斯坦方程的任何其他最大发展，那么存在微分同胚 d: $\mathcal{M}\rightarrow\mathcal{M}'$使得 $d^{*}g_{ab}=g'_{ab}$。

居里原理(参见上文2.3节)使人相信，真空爱因斯坦方程的初值集(Σ, h_{ab}, k_{ab})的对称性应当通过相应的解延续下来。事实确实如此。令 φ: $\Sigma\rightarrow\Sigma$为一个微分同胚，它在 $\varphi^{*}h_{ab}=h_{ab}$和 $\varphi^{*}k_{ab}=k_{ab}$的意义上是初始数据的一个对称。那么就像弗里德里克和伦德尔(Friedrich and Rendall)[2000, pp. 216—217]所表明的，如果ψ是Σ在由(Σ, h_{ab}, k_{ab})决定的最大柯西发展中的嵌入，那么存在一个这个发展自身上的等距$\bar{\psi}$，使得$\bar{\psi}\circ\varphi=\varphi\circ\psi$。即，存在最大柯西发展的一个等距，它对 $\varphi(\Sigma)$的约束为ψ。而且，初始数据对称性的这种延伸是唯一的。

有源爱因斯坦方程的初值问题不仅涉及应力—能量张量 T_{ab}，而且涉及产生了 T_{ab}的物质场的运动方程。特别地，还涉及这些场与引力或者它们相互之间的耦合。耦合爱因斯坦—物质方程是否允许初值公式或如果允许的话初值问

① 这个公式来自瓦尔德[1984, Thm. 10.2.2]。

题是否是适定的，这就成为不得不逐例研究的问题了。对于什么看起来是耦合的适当选择，组合的爱因斯坦—克莱因—戈登方程和爱因斯坦—麦克斯韦方程的初值问题有着类似于无源爱因斯坦方程的存在性和唯一性的结果。对其他例子，结果就没有这么好了。①

以上提到的结果证明了实体广义协变性(在微分同胚不变性是规范对称的意义上)是与存在适定初值问题相容的。但是两者之间有着明显的张力，因此人们可以对这两个要求的联合对可能原理的约束有多牢固感到惊讶。②

6.4　宇宙监督和时序保护

之前章节中所提到的正面结果很难穷尽广义相对论中的决定论问题。主要的疑虑在于，当初始数据(Σ，h_{ab}，k_{ab})满足约束方程的最大柯西发展$\mathcal{M}$，g_{ab}不是绝对最大的时候，即，当$\mathcal{M}$，g_{ab}可以作为一个满足无源爱因斯坦方程的更大时空$\mathcal{M}'$，g'_{ab}的适当子集而被嵌入的时候，将会发生什么。对于$T_{ab}\neq 0$的情况，也可以提出类似问题。更大时空中非独立的未来区域$D^+(\Sigma)$(的影像)的未来边界$H^+(\Sigma)$被称为Σ的未来柯西视界，类似地可以定义过去柯西视界$H^-(\Sigma)$。③ 直

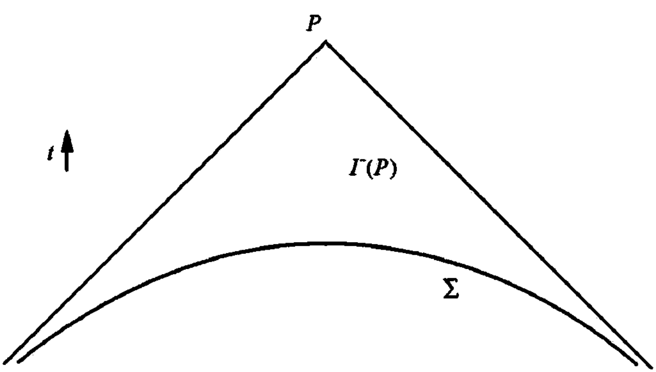


图3　初值超曲面的较差选择

① 对已知的东西的综合性综述，见[Friedrich and Rendall，2000]与[Rendall，2002]。

② 与这里所讲的不同的规范对称分析在[Geroch，2004]中给出。他只给出了原理的两个例子，它们有初值公式，有着像对称不变性一样的微分同胚不变性。

③ 更加精确地，$H^+(\Sigma):=\overline{D+(\Sigma)}-I^-(D^+(\Sigma))$，且对于$H-(\Sigma)$也类似。

观地说，在Σ的柯西视界之外，存在着时空区域，在其之上事物的态不由给定的Σ上初始数据唯一确定。因为总体来讲，如果初始条件的最大柯西发展$\mathcal{M}$, g_{ab}不是绝对最大的，那么Σ对其来说不是柯西面的更大延伸不是唯一的。

至于为什么最大柯西发展可能不是绝对最大的，一个相对来讲不怎么令人感兴趣的原因是，Σ是初值超曲面的一个较差的选择。举一个很小的但是有用的例子：选择Σ作为如图3中所示的闵可夫斯基时空的类空双曲面。这里$H^+(\Sigma)$是点p的过去零锥。

这个例子的一些特征得到了推广，特别地，$H^+(\Sigma)$应该总是零测地线产生的零曲面。比如通过要求Σ是紧致的或渐近地平直的，就可以排除不适当的例子。当然，这些条件也排除了很多物理上适当的例子，但是为了便于讨论，让我们先放下它们。即便这样，最大柯西发展也可能不会由于更加令人感兴趣和更加令人不安的原因而成为绝对最大的。

一种原因是，时空可以从很好的因果特性开始，但在其演化中特性都丢失了。米斯纳(Misner)的(1+1)维时空就阐明了这一点。这种时空抓住了塔布(Taub)-NUT时空的某些因果特征，而塔布-NUT时空是无源爱因斯坦方程的一个解。在图4中，Σ是时空的因果行为良好的塔布部分中的紧致类空片断，并且它的未来柯西视界$H^+(\Sigma)$是一个封闭零锥。穿过这个视界，就将进入时空的一个区域，在这里存在封闭的类时曲线。

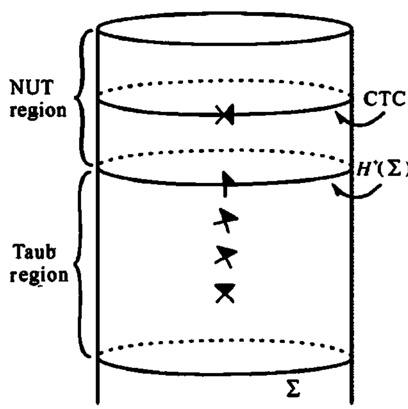


图4 米斯纳时空

图5阐明了最大柯西发展可能不会成为绝对地最大的另一个原因。其中展示了一个非紧致渐近平直的类空片断Σ，在其之上一个球对称物质开始经历引力塌缩。在有限的时间之后，塌缩物质的密度变得无穷大，产生了一个曲率奇点，这个奇点在图中被画成一条类时曲线。然而严格地说，将奇点叫作类时曲

线毫无意义，因为奇点不是时空的一部分。① 但是这并不影响这里的相关重点，即一条结束于 $H^{+}(\Sigma)$的未来上一点并延伸到其过去的因果曲线，可能不会到达Σ，并不是因为它有着过去的终点或者因为它陷入了 $H^{+}(\Sigma)$(如图4的时空中发生的)，而是因为它“进入了一个奇点”，或者更好的说法是(由于奇点不是时空的一部分)，因为它“离开了时空的边界”。

众所周知，爱因斯坦方程加上上述讨论的 T_{ab}上的各种能量条件还并不足以阻止图4和图5指出的那种反常。但是在所有已知的这种例子中，有些东西值得怀疑：其中涉及的物质场都不是“基本的”，也就是说，即使是引力被禁用，这些物质场也不会表现良好。这是在以下意义上讲的：在闵可夫斯基时空中，它们运动方程的初值问题不允许全域存在性和唯一性的结果(见4.2节)，否则奇点中终结的初始条件就会非常特殊。比如，图5中为产生奇点而要求的物质的初始构造必须是完全球对称的。人们可能会猜测，能支持这些例子的那些东西是普遍成立的：考虑一个可以作为引力源的基本物质场。这样的话，考虑爱因斯坦物质场方程，对于它而言，唯一的(取决于微分同胚)最大柯西发展并非绝对最大。假定排除了初值超曲面的较差选择，这样一个场方程初始数据的子集在此类数据的整个空间上是零测度的。要使这种模糊的主张变成一种精确猜想，就要求详细说明什么物质场被当作是基本的、初始数据空间上的适当测量是什么样的，以及什么被看作初始值超曲面的较差选择。这样一种有目的的猜想被称为彭罗斯(Penrose)的宇宙监督猜想。

关于这种猜想，不太全面的版本可能特别关注图4和图5中阐述的两种反常中的其中一种。霍金时序保护猜想的目的在于表明从一个因果无问题的过去发展出来的封闭类时曲线在爱因斯坦基本物质场方程的解中是极不一般的。弱宇宙监督猜想的目的是要表明在有着渐近平直时空的一般解中，奇点在它对无限远处的观测者可见的意义上来说，并非“裸”的。因为任何奇点的发展(比如在引力塌缩中)都隐藏于黑洞的事件视界之中，而这个视界作为单行的因果膜，

① 有人会尝试附加上奇点作为时空流形的边界点。然而，这样做的现有方法导致了违反直觉的特征。例如，奇点与流形内部的点不需要是豪斯道夫分离的，见[Geroch *et al.*, 1982]。

保护外在观测者不受任何可能在视界中发展的非充分决定性的病态影响。在公式化以及证明时序保护和弱宇宙监督的精确版本方面，已经取得了一些进步，但对于强宇宙监督，还缺乏有效的评判。①

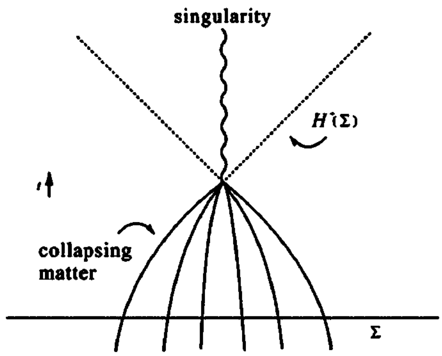


图5　在球面引力塌缩中裸奇点的发展

6.5　广义相对论时空中的可预言性

在4.3节中我们看到，闵可夫斯基时空的结构对于决定论和可预言性具有两面性特征，尽管这个结构使得决定论的清晰例子成为可能，但它同时也使嵌入观测者运用决定论来执行真正的预言成为不可能。这些嵌入观测者必须通过与世界的因果相互作用来收集他们关于初始条件的信息。这一点的形式化是通过定义时空$\mathcal{M}$，g_{ab}上一点$q\in\mathcal{M}$的可预言性，并注意到闵可夫斯基时空对于每一个q都有$p(q)=\varnothing$而得到的。可预言性的非空区域是在闵可夫斯基时空有着如图2中所描绘的紧致空间片段的修正版本中获得的。这个例子的一个特征推广到任意的广义相对论时空之中，也就是说，如果时空$\mathcal{M}$，g_{ab}拥有一个柯西面Σ使得对于某些$q\in\mathcal{M}$有$\Sigma\subset I^{-}(q)$，那么Σ就是紧致的。由于一个有着柯西面Σ的时空是微分同胚地$\Sigma\times\mathbb{R}$的，那么随着对某些$q\in\mathcal{M}$有$\Sigma\subset I^{-}(q)$而来的彻底可预言性只有在空间有限的宇宙中才成为可能。反之则不对，紧致柯西面的存在并不能保证存在一个柯西面Σ，使得对某些$q\in\mathcal{M}$有$\Sigma\subset I^{-}(q)$，德西特(de Sitter)时空提供了一个相关的反例。[Hogarth(霍加斯)，1993]研究

① 宇宙监督进展的概述见[Chruściel(克鲁塞尔)，1992；Isenberg(伊森伯格)，1992]；[Penrose，1998]；[Wald，1998]。时序保护进展的概述见[Earman，2005]。

了广义相对论时空中可预言性的许多有意义的特征。

6.6　广义相对论时空中的决定论和可计算性

在数论中存在一个未解决的猜想，其前述范式为$(\forall_{n_1})(\forall_{n_2})\cdots(\forall_{n_m})F(n_1, n_2, \cdots, n_m)$或$(\exists_{n_1})(\exists_{n_2})(\exists_{n_m})F(n_1, n_2, \cdots, n_m)$，其中$F$是递归的。获取关于这个未决猜想的数学知识所需要的柏拉图(Plato)机制可以概念化为一个在芝诺方式中运转的普通图灵机：列举自然数的m元组，在前1/2分钟检查计算机，察看F对于第一个元素是否成立，在接下来的1/4分钟检查F对第二个元素是否成立，等等。在这一分钟结束的时候，猜想的实际情况就确定下来了。尽管存在各种各样的反对意见，我在这种装置中却并没有看到任何概念上的不一致。但是，狭义相对论对这种装置的物理学实例有所妨碍，因为芝诺加速似乎要求这种策略的某些部分最终运动得要比光速还快。①

广义相对论时空似乎开启了创造不存在芝诺把戏并且不存在超光速但却与柏拉图机制有着等效作用的东西的可能性。考虑一个具有下述特征的时空：首先，存在一条类时半曲线γ_1，它具有过去终点却没有未来终点以及无限固有长度。第二，存在另一条类时半曲线γ_2，它具有过去终点p以及一点$q\in\gamma_2$，使得沿γ_2从p到q的固有时是有限的，同时使得$\gamma_1\in I^-(q)$。这样一个时空被称为马拉蒙特—霍加斯时空。任何递归可公理化理论的定理——比如策梅罗—弗兰克尔(Zermelo-Frankel)集合论——都可以被递归地列举，并且世界线为γ_1的装置可以利用图灵机来有效地检查每一个定理，以找到形式为“0 = 1”的那一个。可以程序化地控制这个装置，向观测者发出一个信号：“找到了！”只有在定理中找到“0 = 1”的时候这个观测者的世界线才是γ_2。假定观测者γ_2知道这种安排，她就获得了关于ZF的一致或不一致的知识：只要在她在到达点q那一刻还没有接收到“找到了！”的信号，她就知道ZF是一致的。

至少在原理上，类似的安排可以被用来“决定”图灵不可确定问题，以及“计算”图灵不可计算函数，参见[Hogarth，1994]。因此它们就对查尔彻—图

① 也许通过芝诺收缩几部分可以避免与狭义相对论的冲突，但是这种策略可能与量子约束发生冲突。

灵(Church-Turing)论的正确性产生了威胁。这个论点粗略地主张任何物理计算装置都能用图灵机来模拟，参见[Etsei and Németi(伊森和内梅提), 2002]，有此论点的详细公式。与属于数理逻辑的原始查尔彻—图灵论不同，物理的查尔彻—图灵论处于数理逻辑和物理学领域的交叉区域，参见[Deutsch *et al.*(道成等), 2000]，对它进行评价要难得多，特别是如果人们认为它要求的装置的物理可实现性来实施分支超级任务的话。这里我将把自己限制在对它的少数几个评价中，对于那些感兴趣的读者，请参见内梅提与大卫[2005]更全面的讨论。

马拉蒙特—霍加斯时空出现在爱因斯坦方程的解之中——比如，瑞斯纳—诺德斯多姆(Reissner-Nordström)空间和反德·西特时空(的普遍覆盖时空)。这些特别的时空在它们允许全域时间函数的意义上不包含因果异常。但是所有的马拉蒙特—霍加斯时空都不是全域双曲的。事实上可以表明，如果$\Sigma\subset\mathcal{M}$，这样的时空就可以是任何使上述定义的γ_1位于$I^+(\Sigma)$之内的类空超曲面，那么任何其过去包含了γ_1的马拉蒙特—霍加斯的时空点q就必须位于$H^+(\Sigma)$或超出其之外，参见[Earman, 1995, 117]中的引理4.3。非决定论影响的可能性，它可能开启γ_1接收了一个错误的“找到了！”信号的可能性，似乎破坏了马拉蒙特—霍加斯时空被用来在决定性意义上获得知识的用途。然而，在这里谁也不应该仓促地得出结论，因为正如在下面小节中将要讨论的，一个在非全域双曲时空上传播的场有可能具有决定论的动力学。同样，似乎问题可以由于它能够被安排以使任何从γ_1发出的信号在马拉蒙特—霍加斯时空的点q之前，因此在$D^+(\Sigma)$之内就到达γ_2而得到避免。但是，由于“找到了”的信号可以到达任意接近q的位置，那么接收者就必须拥有无限精确的识别能力来分辨在q之前就到达的信号和潜在的在q之后到达的错误信号。

6.7 非全域双曲时空中决定论动力学的可能性

对于在广义相对论时空中传播的场，全域双曲性的失效会削弱初值问题。例如，众所周知，在有着封闭类时曲线的一般二维时空中，标量波动方程可能没有平滑解，否则就会允许在一个类空双曲面上确定相同初始数据的多个解。但是很明显，在一些有着封闭类时曲线的四维时空中已经证明了存在性且唯一性结果，参见[Friedman(弗里德曼), 2004]的评论。

对于没有像封闭类时曲线这样明显的因果异常但却不是全域双曲的时空来讲，希尔伯特空间的技术有时可以用来避免存在性和唯一性的崩溃。① 考虑一个静态且拥有全域时间函数的广义相对论时空$\mathcal{M}$，g_{ab}。第一个条件意味着存在一个超曲面正交的类时基灵(Killing)场 V^a。② 第二个条件可以通过选择一个正交于 V^a 的类空超曲面Σ并且要求 V^a 的每一条积分曲线都在一个确定点上相交于Σ来得到保证。这样，每个点$p \in \mathcal{M}$都能用(x, t)标记，其中 $x \in \Sigma$ V^a 的积分曲线穿过 p 与Σ的交点，t 是沿这条积分曲线的基灵参数的值。但这样一个因果表现良好的时空却有可能不是全域双曲的，因为在直观上，它拥有一个裸类时起点。(为了帮助确定直觉，考虑没有类时世界管的闵可夫斯基时空。或者熟悉广义相对论的读者可以考虑爱因斯坦方程的负质量史瓦西解，它是静态的且在 $r=0$ 时包含一个类时赤裸奇点。) 现在考虑一个巨大的 $m \geqslant 0$ 的标量场 ϕ，它根据克莱因—戈登方程(13)在这个背景时空中传播。对我们讨论中的这类时空，这个方程可以写成如下形式：

$$\frac{\partial^2 \phi}{\partial t^2} = -A\phi \tag{17}$$

其中 t 是基灵参数，参见[Wald, 1980a]和[Horowitz and Marolf, 1995]。微分算子 A 可以被看作是活动于 L^2_{R} (Σ、$d\vartheta$)上的希尔伯特空间 $\hat{A}$ 算子，这里 $d\vartheta$ 是Σ除以 $\sqrt{-V^a V_a}$得到的体积元。这最初被当作是 C_0^∞ (Σ)，$\hat{A}$ 是一个正对称算子。这里的提议是，用常微分方程代替偏微分方程(18)：

$$\frac{d^2 \phi}{dt^2} = -\hat{A}\phi \tag{18}$$

其中式(19)中的时间微分是一个希尔伯特空间微分。由于希尔伯特空间算子 $\hat{A}$ 是自轭延伸的，并且由于 $\hat{A}$ 为正，因此一个自轭延伸 $\hat{A}_e$ 的正平方根可以

---

① 用希尔伯特空间技巧研究经典物理问题的第一人是库普曼(Koopman) [1931]。然而，库普曼的方法假定了经典层面的决定论，然后表明了如何把这种决定论表示为希尔伯特空间上的幺正演化。

② 关键的条件是$\nabla(_c g_{ab})=0$，其中∇_a 是与 g_{ab}相容的协变微商算子。静态性保证了局部有一个局域坐标系(x^α，t)可以被选择而使得线源的形式为 $ds^2 = g_{\alpha\beta}(x^\gamma)\,dx^\alpha dx^\beta - g_{44}(x^\gamma)\,dt^2$，见马拉蒙特在本书中的相关论述。

被提取出来，给出：

$$\phi(t) := \cos(\sqrt{\hat{A}_e}t)\phi(0) + \sin(\sqrt{\hat{A}_e}t)\dot{\phi}(0) \tag{19}$$

此式对所有的 t 和所有的 $\phi(0)$，$\dot{\phi}(0) \in \mathcal{H}$都有效。因为 $\phi(t)$ 是一个穿过克莱因—戈登方程(13)的时空的解，并且由于它对于 Σ 上给定的初始数据 $\phi(0)$ 而言 $\dot{\phi}(0)$ 是唯一的解，因此它为克莱因—戈登场的动力学提供了(相对于自轭延伸的选择)一种决定论的解决方式。还存在其他获得 ϕ 的动力学的可能方法，但是伊氏巴氏和瓦尔德(Ishibashi and Wald)[2003]已经表明，刚才回顾的那个是唯一满足下面这组约束的：它与(18)在局域上一致；它允许适当守恒的能量；它因果地传播着场；并且它遵循时间平移和时间反演不变性。如果希尔伯特空间算子 $\hat{A}$ 是本征自轭的，那么唯一的自轭延伸 $\hat{A}_e$ 就提供了满足所述约束的场 ϕ 的动力学。并且这个动力学是完全决定性的，尽管场在其上运动的背景时空不是全域双曲的。然而，毫不意外地，对于很多静态却非全域双曲时空的例子，$\hat{A}$ 不是本质自轭的，除非同时添加更强的约束去挑选出某种特定的自轭延伸，通过以上过程无法指定明确的动力学。但是很明显霍罗威茨和马若夫[1995]已经提供了静态非全域双曲时空的例子，其中 $\hat{A}$ 是本征自轭的。并且在这些例子中，以上的解决方法产生了一个 ϕ 场的动力学，它尽管有着裸奇点但却是完全决定性的。

7. 相对论量子场论中的决定论

通常的量子力学是从粒子系统的经典机械描述——特别指一种哈密顿描述——开始的，并且试图产生量子化的版本。类似地，量子场论从场的经典相对论描述开始，并试图产生量子化的版本。然而，某些经典场并不参与物理学家们认为可接受的量子场论。比如将非线性波动方程(13)看作描述波色子—波色子相互作用的一个候选者。一个具有启发性的量子化过程产生了一个结论，即不存在最低能态。这就使得系统对辐射塌缩非常敏感。在这个基础上，量子场论理论家已经将假设的相互作用归类为“不是物理上可实现”的，参见[Baym(贝姆)，1960]。如果人们意识到因为常规的初始数据可以挑选出在有限时间

内发展奇点的解，因此讨论的场在经典水平上表现得并不好，那么量子场论中遇到这些困难也许就并不令人惊讶。经典相对论水平上的决定论行为可以作为一个选择原理，来决定什么场是适合量子化的。

决定论在量子场论中还扮演了一个更具有建设性的角色。在通常的量子力学中，量子化包含了对正则对易关系$[\hat{x}_j, \hat{p}_k] = i\delta_{jk}$(CCR)的适当表示。由于包含了无界算子，这种形式的正则对易关系只在算子的范围得到确定时才有意义。这种担忧可以通过运用正则对易关系的外尔或者指数形式来避免，而这种形式又使得斯通—冯·诺伊曼(Stone-von Neumann)理论成为可能：对于有限数目的自由度，外尔正则对易关系的强烈不可约化的连续表示都是整体等价的——事实上，它们都和熟悉的薛定谔表示等价。就像量子场论的情况一样，当存在无限自由度的时候，这个定理不再有效。这是量子场论的一个特征，它引出了很多有趣的解释问题，在这里是不相关的。这里有意义的是以下事实：比如闵可夫斯基时空的克莱因—戈登场的正则对易关系代数的构造，根本上利用了这个场的决定论传播，参见[Wald, 1994]。这种构造能够被推广到在任意全域双曲的广义相对论背景时空中传播的克莱因—戈登场，因为传播的决定论本质在更加普遍的环境中得到了延续。

对于一个非全域双曲时空$\mathcal{M}$，g_{ab}来说，仍然是这种情况：对于任何$p \in \mathcal{M}$,存在一个邻域$\mathcal{N}(p)$使得$\mathcal{N}$，$g_{ab} \mid_{\mathcal{N}}$自身被看作时空并且是全域双曲的，因此这种迷你时空的场代数$\mathcal{A}(\mathcal{M})$可以用通常的办法进行构造。那么人们就可以询问，这些局域代数是否能够组合起来形成一个全域代数$\mathcal{A}(\mathcal{M})$，而这种代数有着自然的网络特性呢？比如，每一个这样的$\mathcal{A}(\mathcal{N})$都是$\mathcal{A}(\mathcal{M})$的一个子代数，并且如果$\mathcal{N}_1 \subset \mathcal{N}_2$，那么$\mathcal{A}(\mathcal{N}_1)$是$\mathcal{A}(\mathcal{N}_2)$的子代数。凯(Kay)(1992年)运用了一种量子相容的时空，对它来说答案是肯定的。其思想是，非量子相容的时空不是量子场论的适当舞台。各种各样的非全域双曲时空并非量子相容的，比如，通过沿时间轴进行确定而从二维闵可夫斯基时空得到的二维圆柱时空。但是，很明显，有些因果性时空已经被表明是量子相容的，参见[Fewster and Higuchi(菲斯特和海古彻)，1996]与[Fewster，1999]。

8. 决定论和量子引力

可以论证，当今理论物理学的最大挑战在于将广义相对论和量子场论的见解联合起来以产生一种引力的量子理论，参见罗韦利在本书中的论述。由此产生的关于理论背后在追寻什么的一些想法，也许能从半经典量子引力中得到，它是广义相对论和量子场论的一种闪电结合。半经典的意思是，并不尝试将时空度规量子化，而只把广义相对论时空当作一个固定的背景，量子场在其之上传播(就如在前一节中一样)。这种尝试被用来计算度规上的逆反应，通过在爱因斯坦方程(17)右半边的经典应力—能量张量中插入(重整的)应力—能量期望值而进行。尽管在这样一种过程中不存在一致性理论，但好的理论物理学家知道如何从中提取有用的信息。可能最吸引人的就是霍金的结论了，黑洞并不是黑的，而是完全像一个在与黑洞的表面引力成正比的温度下的黑体一样发出辐射。这种霍金效应被看作是下面论点的一种证明：独立推导的黑洞熵①公式并不只是一种形式上的表达，它表明黑洞熵是黑洞的普通热力学熵，参见[Wald, 1994]。②不同流派的理论物理学家一致同意这是一个稳定的结果，必然会被一种充分的引力量子理论所容纳。但这一点之后的分歧也变得很尖锐。

霍金效应意味着，当考虑量子效应时黑洞就不是稳定的物体了，因为霍金辐射必然伴随着黑洞质量的减少。可以推测，随着这个过程在量子领域越来越深入，半经典的计算最终会崩溃。但是，如果计算的延续是可信的，那么在足够的时间之后，黑洞会完全蒸发。(估计太阳质量的黑洞的蒸发时间数量级为大约 10^{67} 年，比宇宙的年龄要大得多。但是这一点，在有着无限未来的宇宙中是不成问题的，最新的宇宙测量表明我们的宇宙很可能就是这样的情况。)并

---

① $S_{bh}=\frac{kc^3}{4G\hbar}A$，其中 A 是黑洞表面积。

② 霍金效应与盎鲁(Unruh)效应是相关的但又有所区别。其区别在于，后者是按照埃姆什在本书中所讨论的量子统计力学的语言分析的。在闵可夫斯基时空中，盎鲁效应的要素是，观测者在匀加速经过闵可夫斯基真空时的经历由一个 KMS 态描述。盎鲁效应已经被推广到广义相对论时空了，可参见[Kay and Wald, 1991]。

且，如果蒸发的结果可以用经典广义相对论时空描述，那么这个结果就是一个暂时的裸奇点，以及全域双曲率的崩溃，就如图 6 中所表明的那样。① 所以即便某些形式的宇宙监督对经典广义相对论来说成立，但量子效应却似乎不支持它。

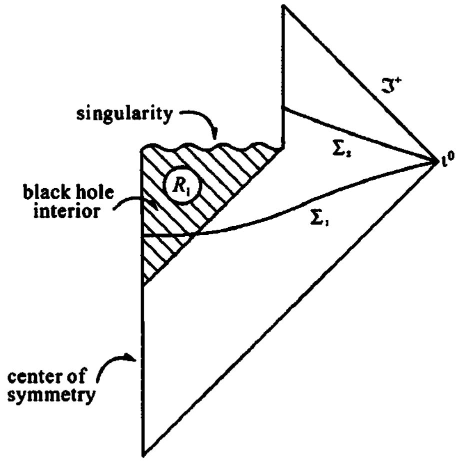


图 6　黑洞蒸发的共形表示

已经走了这么远，并不难建立一种理论：如果在先于黑洞蒸发的时间 Σ_1 时量子场处于纯态，同时如果霍金辐射已经在比如像在黑洞内部区域 R_1（见图 6）那样的相对类空区域和组成关于后蒸发时间 Σ_2 的“三明治”的区域之间建立了相互关联，那么量子场的态将会在后蒸发时间 Σ_2 成为混合态。② 由于从纯态到混合态的转变必然是非幺正的，最终结果就是幺正性的丢失。③

这种“信息丢失悖论”，就像在物理学和大众文学中所经常提到的一样，已经引起了令人吃惊的众多反应，参见[Belot *et al.*, 1999]。最值得注意的是那些拼命想要避免这个结论的人的反应。他们要展开“黑洞互补性”④，因此要抛

---

① 按共形图的惯例，参见[Hawking and Ellis, 1973]，$\mathfrak{I}^+$ 表示未来零无限（向外的零射线的终点），ι^0 表示空间无限。

② 这可以用量子场论的代数公式严格地建立起来，见[Earman, 2002]。

③ 并且顺便地，还存在对时间反演不变性的违背，见[Wald, 1986]和[Earman, 2002]。

④ 考虑狭义相对论中两个观测者 O 和 O' 的情况，他们分别用电场和磁场（**E**, **B**）和（**E**′, **B**′）描述了一个被包围的电磁场。在这两种将 **E**′和 **B**′作为 **E**, **B** 的函数的描述之间存在一个变换，同时 O 和 O' 的相对速度发生了交换，反之亦然。这样一种变换的存在遵从了一个事实：存在一个内在的与观测者无关的实在——在这种情况下，电磁场就用麦克斯韦张量场表征。这个张量场独立于坐标系、参考系、框架和观测者。将其与不同的速度场连接起来，表示不同观测者的运动会导致不同的用电场和磁场的语言进行的描述。不管它还意味什么，“黑洞互补性”和“互补性”的意思是，两个观测者给出的一个蒸发黑洞的不同描述，一种是落入黑洞视界，一种是保持在视界之外，都不会以和 O 和 O' 相关的方式相联系。

弃对相对论的主流解读，也就是说，理论教给我们的是存在一种内在的、独立于观测的实在——与认为任何事物都是相对于观测者的平凡关系论完全相反的观点。

但是从这种争论中退一步，人们会看到根本没必要采取这种孤注一掷的方式。位于“悖论”核心的从纯态到混合态的演化并不需要被看作是量子理论的崩溃。① 也无须因为结果由于全域双曲性的丢失而被打上了“悖论”的标记就感到惊奇。需要询问的是，这种全域双曲性的丢失是否是量子引力的合理预期。半经典量子引力预计了这样的丢失，但这种把广义相对论和量子场论结合在一起的方式至多也就是向量子引力完满理论过渡的一个跳板。而且，就像一般量子力学显示了其消除经典力学的奇点的能力那样，量子引力的正确理论可能也会显示其消除经典广义相对论奇点的能力。

一些正面的表示来自于弦理论学家的工作，他们能够提出消除经典广义相对论模型中奇点的机制。比如，约翰逊等人[2000]表明膜斥力可以消除一类被称为斥力的裸奇点。弦理论家还可以给出排除某些类型的经典奇点的后门论证：假定或者证明M理论之后的探寻给出了一个稳定的基态，然后注意这排除了负质量史瓦西解及其类似的可能性。

其他鼓舞人心的结果来自于圈量子引力，它致力于通过将正则量子化方法运用于广义相对论从而产生一种引力的量子理论。正则量子化方法的基础是一组由安米塔巴·森(Amitaba Sen)引入而由阿拜·阿西提卡(Abay Ashtekar)开拓的新变量。② 在经典广义相对论的弗里德曼—罗伯森—沃克(Friedmann-Robertson-Walker)大爆炸模型中，随着越来越接近大爆炸奇点，时空的尺度因子 a 变为零，并且曲率在 $1/a^2$ 的大小上突然出现。③ 由于没有物理方法可以经过奇点扩展这样一个解，因此关于大爆炸“之前”发生了什么的问题就属于神学或者科幻，而不是科学的问题了。在圈量子引力中，情况惊人地不同。对应于经

① 见[Wald，1994，181—182]和[Belot *et al.*，1999]。

② 圈量子引力的研究，见[Rovelli，2004]和本书相关内容。

③ 弗里德曼—罗伯森—沃克模型的线元可以写成 $ds^2=a(t)d\sigma^2-dt^2$ 形式，其中 $d\sigma^2$ 是空间线元。

典量 $1/a$，存在一个自轭算子作用于空间均匀的希尔伯特空间和各向同性的量子运动学态上，它的谱有上界，这时经典奇点已经被消除了，参见[Bojowald(波究瓦德)，2001]的第一个迹象。这种彻底消除的证据会要求量子动力学给出一个经过经典奇点的明确演化。在圈量子引力中，“动力学”是通过解哈密顿约束方程而得到的，这就限制了物理学所允许的态。对于讨论的情况，这种约束方程的形式是差分方程而不是微分方程。如果尺度因子 a 被看作是一个“时钟变量”，那么约束方程就提供了一种量子态经过时钟变量的离散阶段的“时间演化”。关键在于，这个演化方程确实决定了一个经过经典奇点的唯一延续。① 但在经典奇点处发生了什么是不确定的，因为对应于这个时期的系数与演化方程的其他系数互不相干，参见[Ashtekar and Bojowald，2003]。

从决定论的观点来看，这个最后的结果意味着情况多少有些讽刺。决定论在经典广义相对论中不受弗里德曼—罗伯森—沃克模型的初始大爆炸奇点的威胁，因为这些模型是全域双曲的并且缺乏基于物理机制在奇点中延伸的方法。在圈量子引力中，在曲率仍然有界以及存在一种经过经典奇点的合理延伸方式的意义上，初始奇点被排除了。但其代价是在经典奇点处圈量子引力丧失了决定性，这可以被看作是经典奇点的柴郡猫笑容(意指一个性质分离)。

近来，圈量子引力已经被用来解决黑洞奇点的问题，由此引出一种关于霍金信息缺失悖论的新观点，在此观点中图 6 不是一个黑洞蒸发的有效描述，参看[Ashtekar *et al.*，2005]。它认为，类似于弗里德曼—罗伯森—沃克的例子，量子演化连续地经过经典奇点。② 新的图像不是一幅全域双曲性被恢复的图像；实际上那个概念是没有意义的，因为替换了经典奇点的是一个即使近似地用经典广义相对论的时空几何也不能描述的区域。不过，有人认为在量子演化中，纯态依旧会是纯态，在这个意义上，信息并没有丢失。在它的当前形式中，这

① 但是[Green and Unruh(格林和盎鲁)，2004]表明，在一个空间闭合的弗里德曼—罗伯森—沃克模型中，将尺度因子作为“时钟变量”来用是存在问题的。在非均匀宇宙学中情况要微妙复杂得多，见[Brunnemann and Thiemann(布鲁尼曼和塞曼)，2006a；2006b]。

② 就像在弗里德曼—罗伯森—沃克情况中一样，哈密顿约束方程变成了一个差分方程。“量子演化”是来自于此方程的，是通过选择一个合适的“时钟变量”从而在时钟变量的离散步骤中遵循量子态的实现。

个争论具有启发式的特征，同时也需要详细的计算使它趋于严密化。

9. 总结

世界是决定论的吗？没有了形而上学的帮助，我们只有通过检验科学理论化的成果来处理这个问题。因此我们给自己设置了去检查现代物理理论的任务并对其中每一个发出询问：如果世界就是按照理论成立的那样存在，那么它是决定论的吗？我们的发现之一就是这个任务距离明确还遥不可及，因为理论解释的方法由于我们对于决定论的态度而染上了颜色。例如，不愿看到决定论在牛顿万有引力理论一开始就失效，这影响了人们赞同将万有引力看作是粒子之间的相互作用，而不同意给牛顿的万有引力场赋予独立的自由度。同时不愿看到决定论在广义相对论一开始就失效，这导致了人们拒绝教科书版的时空理论的单纯现实性描述，而接受微分同胚映射不变性作为其规范对称性——需要在教科书展示中使用的变量没有一个可算作真正的(即规范对称的)物理量级。

决定论的命运太过复杂，无法进行短小而精确的总结，但是粗略地讲，对于经典的(即非量子理论的)来讲情况就是这样。在牛顿理论中，如果没有各种辅助假设的帮助，决定论就很难成立。对于狭义相对论，决定论的出现很安全，使得它被用作对“基本场”的选择定则。在合适的测量解释之下，广义相对论在时间上是局域决定论的；但是它在时间中是否是决定论的就发展成了宇宙监督制度和时序保护这些尚未解决的问题了。

量子物理学的情况是最奇怪并且是最困难的。普通量子力学在某些观点上比牛顿机械论更确定，比如量子力学可以解决一些由解的非唯一性或解的崩溃导致的牛顿决定论的失效。但是量子力学中决定论的命运最终依赖尚未解决的解释问题。这些问题背后的驱动力是解释量子力学如何能说明准确的实验结果或者更加一般地，经典世界表面上的确定性——这很讽刺，因为量子力学是最精确的物理理论。一些对这种解释的挑战性的现有回应可能会埋葬决定论，而其他回应则给了它新生。

引力的初步量子理论给出了检验决定论勇气的一个新的竞技场。一些基本迹象表明，就如普通的量子力学能消除牛顿机械论的奇点一样，量子引力效应

可能也能消除经典广义相对论中的奇点。如果这种消除的能力足够广泛，那么它会缓和量子引力中存在的类似于经典广义相对论中与宇宙监督制度的失败相关联的决定论的崩溃的担忧。量子引力可能会对测量问题和普通量子力学中产生的有关解释问题摆出一副新的面孔，但这张新面孔是否会冲着决定论微笑则并不好说。

致谢

在此很高兴感谢从戈登·贝洛特、迈克·迪克森、卡拉斯·兰兹曼、卡罗尔·罗韦利和劳拉·鲁切那里得到的帮助和建议。特别感谢杰里米·巴特菲尔德提供的多次改进。但是毋庸置疑，以上提到的所有人都同意这里表达的观点。

参考文献

[Albert, 1992] D. Albert. *Quantum Mechanics and Experience*. Cambridge, MA: Harvard University Press, 1992.

[Albert, 2000] D. Albert. *Time and Chance*. Cambridge, MA: Harvard University Press, 2000.

[Anderson and Thompson, 1992] I. Anderson and G. Thompson. *The inverse problem of the calculus of variations for ordinary differential equations. Memoirs of the American Mathematical Society*, No. 473. Providence, RI: American Mathematical Society, 1992.

[Ashtekar and Bojowald, 2005] A. Ashtekar and M. Bojowald. Black hole evaporation: a paradigm. 2005. *gr-qc*/0504029.

[Ashtekar *et al.*, 2003] A. Ashtekar, M. Bojowald, and J. Lewandowski. Mathematical structure of loop quantum cosmology. *Advances in Theoretical and Mathematical Physics*, 7: 233-268, 2003. *gr-qc*/0304074.

[Atmanspacher and Bishop, 2002] H. Atmanspacher and R. Bishop (eds.). *Between Chance and Choice: Interdisciplinary Perspectives on Determinism*. Thorveton: Imprint Academic, 2002.

[Ax and Kochen, 1999] J. Ax and S. Kochen. Extension of quantum mechanics to individual systems. 1999. *quant-ph/*9905077.

[Bacciagaluppi and Dickson, 1998] G. Bacciagaluppi and M. Dickson. Dynamics for modal interpretations. *Foundations of Physics*, 29: 1165-1201, 1998.

[Barbour, 1974] J. B. Barbour. Relative-distance Machian theories. *Nature*, 249: 328-329, 1974.

[Barbour, 1999] J. B. Barbour. The development of Machian themes in the twentieth century. In J. Butterfield (ed.), *The Arguments of Time*, pages 83-110. Oxford: Oxford University Press, 1999.

[Barbour and Bertotti, 1977] J. B. Barbour and B. Bertotti. Gravity and inertia in a Machian framework. *Nuovo Cimento*, 38B: 1-27, 1977.

[Bargmann, 1954] V. Bargmann. On unitary ray representations of continuous groups. *Annals of Mathematics*, 59: 1-46, 1954.

[Barrett, 1999] J. A. Barrett. *The Quantum Mechanics of Minds and Worlds.* Oxford: Oxford University Press, 1999.

[Baym, 1960] G. Baym. Inconsistency of cubic boson-boson interactions. *Physical Review*, 117: 886-888, 1960.

[Beig, 2005] R. Beig. Concepts of hyperbolicity and relativistic continuum mechanics. 2005. *gr-qc/* 0411092.

[Belot and Earman, 1997] G. Belot and J. Earman. Chaos out of order: quantum mechanics, the correspondence principle, and chaos. *Studies in History and Philosophy of Modern Physics*, 28B: 147-182, 1997.

[Belot and Earman, 1999] G. Belot and J. Earman. From metaphysics to physics. In J. Butterfield and C. Pagonis (eds.), *From Physics to Philosophy*, pages 166-186. Cambridge: Cambridge University Press, 1999.

[Belot *et al.*, 1999] G. Belot, J. Earman, and L. Ruetsche. The Hawking information loss paradox: The anatomy of a controversy. *British Journal for the Philosophy of Science*, 50: 189-229, 1999.

[Bojowald, 2001] M. Bojowald. Absence of a singularity in loop quantum cosmology. *Physica Review Letters*, 86: 5227-5230, 2001.

[Brading and Brown, 2003] C. Brading and H. Brown. Symmetries and Noether's theorems. In

C. Brading and H. Brown (eds.), *Symmetries in Physics: Philosophical Reflections*, pages 89-109. Cambridge: Cambridge University Press, 2003.

[Brown, 1999] H. R. Brown. Aspects of objectivity in quantum mechanics. In J. Butterfield and C. Pagonis (eds.), *From Physics to Philosophy*, pages 45-70. Cambridge: Cambridge University Press, 1999.

[Brunneman and Thieman, 2006a] J. Brunneman and T. Thiemann. On (cosmological) singularity avoidance in loop quantum gravity, *Classical and Quantum Gravity*, 23: 1395-1427, 2006.

[Brunneman and Thieman, 2006b] J. Brunneman and T. Thiemann. Unboundedness of triad-like operators in loop quantum gravity, *Classical and Quantum Gravity*, 23: 1429-1483, 2006.

[Bub, 1997] J. Bub. *Interpreting the Quantum World.* Cambridge: Cambridge University Press, 1997.

[Butterfield, 1998] J. Butterfield. Determinism and indeterminism. *Routledge Encyclopedia of Philosophy*, 3: 33-39. New York: Routledge, 1998.

[Callender, 2005] C. Callender. Time is the simplest (and strongest) thing. Pre-print, 2005.

[Carroll, 2004] J. Carroll (ed.). *Readings on Laws of Nature.* Pittsburgh, PA: University of Pittsburgh Press, 2004.

[Cassirer, 1956] E. Cassirer. *Determinism and Indeterminism in Modern Physics.* New Haven: Yale University Press, 1956.

[Chalmers, 1970] A. F. Chalmers. Curie's principle. *British Journal for the Philosophy of Science*, 21: 133-148, 1970.

[Chruściel, 1992] P. Chruściel. On uniqueness in the large of solutions of Einstein's equations ('strong cosmic censorship'). In M. Gotay, J. E. Marsden, and V. Moncrief (eds.), *Mathematical Aspects of Classical Field Theory*, pages 235-273. Providence, RI: American Mathematical Society, 1992.

[Courant and Hilbert, 1962] R. Courant and D. Hilbert. *Methods of Mathematical Physics*, 2 vols. New York: Interscience Publishers, 1962.

[Curie, 1894] P. Curie. Sur la symmétrie des phénoménes physiques: symmétrie d'un champélectrique et d'un champ magnétique. *Journal de Physique*, 3: 393-415, 1894.

[Cushing, 1994] J. Cushing. *Quantum Mechanics: Historical Contingency and the Copenhage Hegemony.* Chicago: University of Chicago Press, 1994.

[Cushing and Bowman, 1999] J. Cushing and G. Bowman. Bohmian mechanics and chaos. In J. Butterfield and C. Pagonis (eds.), *From Physics to Philosophy*, pages 90-107. Cambridge: Cambridge University Press, 1999.

[Deutsch *et al.*, 2000] D Deutsch, A. Ekert, and R. Lupacchini. Machines, logic, and quantum physics. *Bulletin of Symbolic Logic*, 6: 265-283, 2000.

[Dickson, 1997] W. M. Dickson. On the plurality of dynamics: transition probabilities and modal interpretations. In R. A. Healey and G. Hellman (eds.), *Quantum Measurement: Beyond Paradox. Minnesota Studies in the Philosophy of Science*, Vol. 17, pages 160-182. Minneapolis, MN: University of Minnesota Press, 1997.

[Dickson, 1998] W. M. Dickson. *Quantum Chance and Non-Locality: Probability and Non-Locality in the Interpretations of Quantum Mechanics.* Cambridge: Cambridge University Press, 1998.

[Driver, 1977] R. D. Driver. *Ordinary and Delay Differential Equations.* New York: Springer-Verlag, 1977.

[Driver, 1979] R. D. Driver. Can the future influence the present? *Physical Review D*, 19: 1098-1107, 1979.

[Earman, 1989] J. Earman. *World Enough and Space-Time: Absolute vs. Relational Theories of Space and Time.* Cambridge, MA: MIT Press/Bradford Books, 1989.

[Earman, 1995] J. Earman. *Bangs, Crunches, Whimpers, and Shrieks: Singularities and Acausalities in Relativistic Spacetimes.* Oxford: Oxford University Press, 1995.

[Earman, 2002] J. Earman. What time reversal invariance is and why it matters. *International Journal for the Philosophy of Science*, 16: 245-264, 2002.

[Earman, 2003] J. Earman. Tracking down gauge: an ode to the constrained Hamiltonian formalism. In K. Brading and E. Castellani (eds.), *Symmetries in Physics: Philosophical Reflections*, pages 140-162. Cambridge: Cambridge university Press, 2003.

[Earman, 2004a] J. Earman. Determinism: what we have learned and what we still don't know. In J. K. Campbell, M. O'Rourke, and D. Shier (eds.), *Determinism, Freedom, and Agency*, pages 21-46. Cambridge, MA: MIT Press, 2004a.

[Earman, 2004b] J. Earman. Curie's principle and spontaneous symmetry breaking. *International Studies in Philosophy of Science*, 18: 173-199, 2004b.

[Earman, 2005] J. Earman. Essential self-adjointness: Implications for determinism and the classical-quantum correspondence. Pre-print, 2005b.

[Earman, 2006] J. Earman. Two challenges to the substantive requirement of general covaraince. *Synthese*, 148: 443-468, 2006.

[Earman and Roberts, 2006] J. Earman and J. Roberts. Contact with the nomic: a challenge to the deniers of humean supervenience. *Philosophy and Phenomenological Research*, in press, 2006.

[Earman *et al.*, 2005] J. Earman, C Smeenk, and C. Wuthrich. Do the laws of phyics forbid the operation of a time machine? Pre-print, 2005.

[Earman and Ruetsche, 2005] J. Earman and L. Ruetsche. Relativistic invariance and modal interpretations. *Philosophy of Science*, forthcoming, 2005.

[Etesi and Németi, 2002] G. Etesi and I. Németi. Turing computability and Malament-Hogarth spacetimes. *International Journal of Theoretical Physics*, 41: 342-370, 2002.

[Fefferman, 2001] C. Fefferman. Existence and smoothness of the Navier-Stokes equation. Clay Mathematics Institute, Millennium Problems, 2001. *http://www.claymath.org/*

[Fewster, 1999] C. J. Fewster. Bisolutions to the Klein-Gordon equation and quantum field theory on 2-dimensional cylinder spacetimes. *Classical and Quantum Gravity*, 16: 769-788, 1999.

[Fewster and Higuchi, 1996] C. J. Fewster and A. Higuchi. Quantum field theory on certain non-globally hyperbolic spacetimes. *Classical and Quantum Gravity*, 13: 51-62, 1996.

[Fine, 1982a] A. Fine. Joint distributions, quantum correlations, and commuting observables. *Journal of Mathematical Physics*, 23: 1306-1310, 1982.

[Fine, 1982b] A. Fine. Hidden variables, joint probability, and the Bell inequalities. *Physical Review Letters*, 48: 291-295, 1982.

[Foias *et al.*, 2001] C. Foias, O. Manley, R. Rosa, and R. Teman. *Navier-Stokes Equations and Turbulence*. Cambridge: Cambridge University Press, 2001.

[Friedrich and Rendall, 2000] H. Friedrich and A. Rendall. The Cauchy problem for the Einstein equations. In B. G. Schmidt (ed.), *Einstein's Field Equations and their Physical Impli-*

cations: *Selected Essays in Honor of Jürgen Ehlers*, pages 127-223. New York: Springer-Verlag, 2000.

[Friedman, 2004] J. L. Friedman. The Cauchy problem on spacetimes that are not globally hyperbolic. In P. T. Chruściel and H. Friedrich (eds.), *The Einstein Equations and the Large Scale Behavior of Gravitational Fields*, pages 331-346. Basel: Birkhäuser, 2004.

[Geroch, 1977] R. P. Geroch. Prediction in general relativity. In J. Earman, C. Glymour, and J. Stachel (eds.), *Foundations of Spacetime Theories. Minnesota Studies in the Philosophy of Science*, Vol. 8, pages 81-93. Minneapolis, MN: University of Minnesota Press, 1977.

[Geroch, 1995] R. P. Geroch. Relativistic theories of dissipative fluids. *Journal of Mathematical Physics*, 36: 4226-4241, 1995.

[Geroch, 2004] R. P. Geroch. Gauge, diffeomorphisms, initial-value formulation, etc. In P. T. Chruściel and H. Friedrich (eds.), *The Einstein Equations and the Large Scale Behavior of Gravitational Fields*, pages 441-477. Basel: Birkhäuser, 2004.

[Geroch *et al.*, 1982] R. P. Geroch, C. B. Liang, and R. M. Wald. Singular boundaries of spacetimes. *Journal of Mathematical Physics*, 23: 432-435, 1982.

[Ghirardi *et al.*, 1986] G. C. Ghirardi, A. Rimini, and T. Weber. Unified dynamics for microscopic and macroscopic systems. *Physical Review D*, 34: 470-491, 1986.

[Green and Unruh, 2004] D. Green and W. H. Unruh. Difficulties with recollapsing models in closed isotropic loop quantum cosmology. *Physical Review D*, 103502-1-7, 2004.

[Hadamard, 1923] J. Hadamard. *Lectures on Cauchy's Problem in Linear Partial Differential Equations*, 1923. Reprinted by Dover, New York, 1952.

[Hawking and Ellis, 1973] S. W. Hawking and G. F. R. Ellis. *The Large Scale Structure of Space-Time*. Cambridge: Cambridge University Press, 1973.

[Henneaux and Teitelboim, 1992] M. Henneaux and C. Teitelboim. *Quantization of Gauge Systems*. Princeton, NJ: Princeton University Press, 1992.

[Hogarth, 1993] M. Hogarth. Predicting the future in general relativistic spacetimes. *Studies in History and Philosophy of Science*, 24B: 721-739, 1993.

[Hoefer, 2003] C. Hoefer. Causal Determinism. *Stanford Encyclopedia of Philosophy*. *http://plato.stanford.edu/archives/sum2005/entries/determinism-causal/*

[Hogarth, 1994] M. Hogarth. Non-Turing computers and non-Turing computability. In

D. Hull, M. Forbes, and R. M. Burian (eds.), *PSA* 1994, Vol. 1, pages 126-138, 1994.

[Holland, 1993] P. R. Holland. *The Quantum Theory of Motion.* Cambridge: Cambridge University Press, 1993.

[Horowitz and Marolf, 1995] G. Horowitz and D. Marolf. Quantum probes of spacetimes singularities. *Physical Review D*, 52: 5670-5675, 1995.

[Howard, 1999] D. Howard. Point coincidences and pointer coincidences: Einstein and the invariant content of space-time theories. In H. Goenner, J. Renn, J. Ritter, and T. Sauer (eds.), *The Expanding Worlds of General Relativity*, *Einstein Studies*, Vol. 7, pages 463-500. Boston: Birkhäuser, 1999.

[Isenberg, 1992] J. Isenberg. Progress on strong cosmic censorship. In M. Gotay, J. E. Marsden, and V. Moncrief (eds.), *Mathematical Aspects of Classical Field Theory*, pages 403-418. Providence, RI: American Mathematical Society, 1992.

[Isham and Kuchař, 1986a] C. J. Isham and K. Kuchař. Representations of spacetime diffeomorphisms. I. Canonical parametrized field theories. *Annals of Physics*, 164: 316-333, 1986.

[Isham and Kuchař, 1986b] C. J. Isham and K. Kuchař. Representations of spacetime diffeomorphisms. II. Canonical geometrodynmamics. *Annals of Physics*, 164: 288-315, 1986.

[Ismael, 1997] J. Ismael. Curie's principle. *Synthese*, 110: 167-190, 1997.

[Ishibashi and Wald, 2003] A. Ishibashi and R. M. Wald. Dynamics in non-globally hyperbolic static spacetimes: II. General analysis of prescriptions for dynamics. *Classical and Quantum Gravity*, 20: 3815-3826, 2003.

[John, 1982] F. John. *Partial Differential Equations*, 4th ed. New York: Springer-Verlag, 1982.

[Johnson *et al.*, 2000] C. V. Johnson, A. W. Peet, and J. Polchinski. Gauge theory and the excision of repulsion singularities. *Physical Review D*, 61: 08600-1-11, 2000.

[Kato, 1995] T. Kato. *Perturbation Theory for Linear Operators.* Berlin: Springer-Verlag, 1995.

[Kay, 1992] B. S. Kay. The principle of locality and quantum field theory on (non-globally hyperbolic) curved spacetimes. *Reviews of Mathematical Physics*, Special Issue, 167-195, 1992.

[Kay and Wald, 1991] B. Kay and R. M. Wald. Theorems on the uniqueness and thermal properties of stationary, quasifree states on spacetimes with a bifurcate Killing horizon. *Physics Reports*, 207: 49-136, 1991.

[Koopman, 1931] B. O. Koopman. Hamiltonian systems and transformations in Hilbert space. *Proceedings of the National Academy of Sciences*, 17: 315-318, 1931.

[Kronz, 1998] F. M. Kronz. Nonseparability and quantum chaos. *Philosophy of Science*, 65: 5-75, 1998.

[Keller, 1957] J. B. Keller. On solutions of nonlinear wave equations. *Communications on Pure and Applied Mathematics*, 10: 523-530, 1957.

[Lanford, 1974] O. E. Lanford. Time evolution of large classical systems. In J. Moser (ed.), *Dynamical Systems, Theory and Applications. Lecture Notes in Physics*, 38: 1-207. New York: Springer-Verlag, 1974.

[Laplace, 1820] P. S. Laplace. *Théorie analytique des probabilitiés.* Paris: V. Courcier, 1820.

[Lewis, 1973] D. Lewis. *Counterfactuals.* Cambridge, MA: Harvard University Press, 1973.

[Lichtenberg and Lieberman, 1991] A. J. Lichtenberg and M. A. Lieberman. *Regular and Chaotic Dynamics*, 2nd ed. New York: Springer-Verlag, 1991.

[Malament, 1984] D. Malament. Time travel in the Ğodel universe. P. D. Asquith and P. Kitcher (eds.), *PSA* 1984, Vol. 2: 91-100, 1984.

[Malament, 2004] D. Malament. On the time reversal invariance of classical electromagnetic theory. *Studies in History and Philosophy of Modern Physics*, 35B: 295-315, 2004.

[McCall, 1995] S. McCall. Time flow, non-locality, and measurement in quantum mechanics. In S. Savitt (ed.), *Time's Arrow Today*, pages 155-172. Cambridge: Cambridge University Press, 1995.

[Moore, 1990] C. Moore. Unpredictability and undecidability in dynamical systems. *Physica Review Letters*, 64: 2354-2357, 1990.

[Moore, 1991] C. Moore. Generalized shifts: Unpredictability and undecidability in dynamical systems. *Nonlinearity*, 4: 199-230, 1991.

[Nagel, 1961] E. Nagel. *The Structure of Science.* New York: Harcourt, Brace, and World, 1961.

[Németi and David, 2005] I. Németi and D. David. Relativistic computers and the Turing barrier. *Journal of Applied. Mathematics and Computation*, forthcoming, 2005.

[Norton, 1984] J. D. Norton. How Einstein found his field equations: 1912-1915. In D. Howard and J. Stachel (eds.), *Einstein and the History of General Relativity. Einstein Studies*, Vol 1: 101-159. Boston: Birkhauser, 1984.

[Norton, 1999] J. D. Norton. A quantum mechanical supertask. *Foundations of Physics*, 29: 1265-1302, 1999.

[Olver, 1993] P. J. Olver. *Applications of Lie Groups to Differential Equations*, 2nd Ed.. New York: Springer-Verlag, 1993.

[Payne, 1975] L. E. Payne. *Improperly Posed Problems in Partial Differential Equations.* Philadelphia, PA: SIAM, 1975.

[Pearle, 1976] P. Pearle. Reduction of the state vector by a nonlinear Schrödinger equation. *Physical Review D*, 13: 857-868, 1976.

[Pearle, 1989] P. Pearle. Combining stochastic dynamical state-vector reduction with spontaneous localization. *Physical Review A*, 39: 2277-2289, 1989.

[Penrose, 1998] R. Penrose. The question of cosmic censorship. In R. Wald (ed.), *Black Holes and Relativistic Stars*, pages 103-122. Chicago, IL: University of Chicago Press, 1998.

[Pérez Laraudogoitia, 2001] J. Pérez Laraudogoitia. Supertasks. *The Stanford Encyclopedia of Philosophy*, 2001. *http://plato.stanford.edu/archives/win2001/entries/spacetime-supertasks/*

[Pérez Laraudogoitia, 2005] J. Pérez Laraudogoitia. Avoiding infinite masses. Preprint, 2005.

[Pitowsky, 1996] I. Pitowsky. Laplace's demon consults an oracle: computational complexity of prediction. *Studies in History and Philosophy of Modern Physics*, 27: 161-180, 1996.

[Planck, 1932] M. Planck. *Der Kausalbegriff in der Physik.* Leipzig: Johann Ambrosius Barth, 1932.

[Poincaré, 1952] H. Poincaré. *Science and Method.* New York: Dover Publications, 1952.

[Popper, 1982] K. R. Popper. *The Open Universe.* Totowa, NJ: Roman and Littlefield, 1982.

[Pour-el and Richards, 1981] M. B. Pour-el and J. I. Richards. The wave equation with computable initial data such that its unique solution is not computable. *Advances in Mathematics*, 39: 215-239, 1981.

[Pour-el and Richards, 1989] M. B. Pour-el and J. I. Richards. *Computability in analysis and* physics. New York: Springer-Verlag, 1989.

[Prince, 2000] G. E. Prince. The inverse problem in the calculus of variations and its ramifications. In P. J. Vassiliou and I. G. Lisle (eds.), *Geometric Approaches of Differential Equations*, pages 171-200. Cambridge: Cambridge University Press, 2000.

[Radicati, 1987] L. A. Radicati. Remarks on the early developments of the notion of symmetry breaking. In M. Doncel, A. Hermann, L. Michel, and A. Pais (eds), *Symmetries in physics (1600-1980)*, pages 195-207. Barcelona: Bellatera, 1987.

[Raju, 1994] C. J. Raju. *Time: Towards a Consistent Theory.* Dordrecht: Kluwer Academic, 1994.

[Redhead, 1987] M. Redhead. *Incompleteness, Nonlocality, and Realism.* Oxford: Clarendon Press, 1987.

[Reed and Simon, 1975] M. Reed and B. Simon. *Methods of Modern Mathematical Physics. II: Fourier Analysis, Self-Adjointness.* New York: Academic Press, 1975.

[Reichl, 1992] L. E. Reichl. *The Transition to Chaos in Conservative Classical Systems: Quantum Manifestations.* New York: Springer-Verlag, 1992.

[Reichl, 2004] L. E. Reichl. *The Transition to Chaos in Conservative Classical Systems: Quantum Manifestations*, 2nd ed. New York: Springer-Verlag, 2004,

[Rendall, 2002] A. Rendall. Theorems on existence and global dynamics for the Einstein equations. *Living Reviews in Relativity*, 2002. *http://www.livingreviews.org/Articles/Volume5/2002-6rendall*

[Rice, 1953] H. G. Rice. Classes of recursively enumerable sets and their decision problems. *Transactions of the American Mathematical Society*, 74: 358-366, 1953.

[Rovelli, 2004] C. Rovelli. *Quantum Gravity.* Cambridge: Cambridge University Press, 2004.

[Saari and Xia, 1995] D. Saari and J. Xia. Off to infinity in finite time. *Notices of the American Mathematical Society*, 42: 538-546, 1995.

[Scheffer, 1993] V. Scheffer. An inivisid flow with compact support in spacetime. *Journal of Geometry and Analysis*, 3: 341-401, 1993.

[Schnirelman, 1997] A. Schnirelman. On the nonuniquenes of weak solution of the Euler equation. *Communicatins on Pure and Applied Mathematics*, 50: 1261-1268, 1997.

[Skow, 2005] B. Skow. What makes time different from space. *Nôus*, forthcoming, 2005.

[Stachel, 1986] J. Stachel. Einstein's search for general covariance: 1912-1915. In D. Howard and J. Stachel (eds.), *Einstein and the History of General Relativity. Einstein Studies*, Vol. 1: 63 – 100. Boston: Birkhauser, 1986.

[Stöltzner, 2003] M. Stöltzner. Vienna indeterminism: from Exner to Frank and von Mises. In P. Parrini, W. Salmon, and M. Salmon (eds.), *Logical empiricism: Historical and contemporary perspectives*, pages 194-229. Pittsburgh, PA: University of Pittsburgh Press, 2003.

[Suppes, 1991] P. Suppes. Determinism or instability. Does it matter? In G. G. Brittan (ed.), *Causality, Method, and Modalities*, pages 5-22. Dordrecht: Kluwer Academic, 1991.

[Suppes, 1993] P. Suppes. The transcendental character of determinism. In P. A. French, T. E. Uehling, and H. K. Wettstein (eds.), *Midwest Studies in Philosophy*, 18:, 242-257, 1993.

[van Fraassen, 1989] B. C. van Fraassen. *Laws and Symmetry.* Oxford: Oxford University Press, 1989.

[Wald, 1980a] R. M. Wald. Dynamics in non-globally hyperbolic, static spacetimes. *Journal of Mathematical Physics*, 21: 2802-2805, 1980a.

[Wald, 1980b] R. M. Wald. Quantum gravity and time reversibility. *Physical Review D*, 21: 2742-2755, 1980b.

[Wald, 1984] R. M. Wald. *General Relativity.* Chicago, IL: University of Chicago Press, 1984.

[Wald, 1994] R. M. Wald. *Quantum Field Theory in Curved Spacetime and Black Hole Thermodynamics.* Chicago, IL: University of Chicago Press, 1994.

[Wald, 1998] R. M. Wald. Gravitational collapse and cosmic censorship. In C. V. Vishveshwara, B. R. Iyer, and B. Bhawal (eds.), *Black Holes, Gravitational Radiation, and the Universe.* Dordrecht: Kluwer Academic Publishers, 1998.

[Weyl, 1932] H. Weyl. *The Open World.* New Haven, CN: Yale University Press, 1932.

[Widder, 1975] D. V. Widder. *The Heat Equation.* New York: Academic Press, 1975.

[Xia, 1992] X. Xia. Existence of noncollision singularities in the N-body problem. *Annals of Mathematics*, 135: 411-468, 1992.

[Zeipel, 1908] H. von Zeipel. Sur les singularités du problème des *n* corps. *Arkiv för Matima-*

tik, Astronomie och Fysik, 4: 1-4, 1908.

[Zurek, 1998] W. H. Zurek. Decoherence, chaos, quantum-classical correspondence, and the algorithmic arrow of time. *Physica Scripta*, T76: 186-198, 1998.

[Zurek, 2003] W. H. Zurek. Decoherence, einselection, and the quantum origins of the classical. *Reviews of Modern Physics*, 75: 715-775, 2003.

索　引