



TROY — W  
PHOTOGRAPHY  
— LAKE BAIKAL  
Озеро Байкал

# 光子真实存在吗？

——结构实在论的一个实例

作者：Minakami Yuki

时间：2023 年 2 月 12 日

---

## 目录

---

<b>0 序：前置与导言</b>	<b>1</b>
<b>1 光的量子性</b>	<b>2</b>
1.1 光电效应 . . . . .	3
1.2 康普顿散射 . . . . .	4
1.3 光子的引入 . . . . .	5
1.3.1 A 光电效应 . . . . .	5
1.3.2 B 康普顿散射 . . . . .	5
1.4 波粒二象性 . . . . .	6
<b>2 光子必要吗？</b>	<b>8</b>
2.1 重回光电效应 . . . . .	8
2.2 康普顿散射新解 . . . . .	9
2.3 结论 . . . . .	9
<b>3 光子存在的实验证据</b>	<b>10</b>
3.1 亚泊松光 . . . . .	10
3.2 光子反相关 . . . . .	12
3.3 纠缠光子与不可分性 . . . . .	12
<b>4 量子场论下的光子</b>	<b>16</b>
4.1 $U(1)$ 规范对称性 . . . . .	17
4.2 对称性自发破缺 . . . . .	19
4.3 电弱对称性破缺 . . . . .	20
<b>5 科学实在论导引</b>	<b>22</b>
5.1 无奇迹论证 . . . . .	23
5.2 悲观元归纳 . . . . .	24
5.3 三个维度 . . . . .	26
<b>6 光子与结构实在论</b>	<b>28</b>
6.1 光子作为基本实体 . . . . .	29
6.2 光子的结构进路 . . . . .	30
6.3 规范场论的建构 . . . . .	31
<b>7 附录 A 标准模型</b>	<b>34</b>
<b>8 附录 B 拓扑量子场论</b>	<b>36</b>
8.1 公理化拓扑场论 . . . . .	37
8.2 瞬子几何学 . . . . .	38



# Chapter 0

## 序：前置与导言

在开始整个 notes 之前，我想要说明一下这个 notes 究竟是什么。首先，这个 notes 并不是一个面向物理的“综述”，它试图讨论的是一个科学哲学问题——不可观察的理论实体，诸如光子和夸克的实在性问题。它们是否真的像客观实体那样，不依赖于人的意志，存在于物理世界？抑或仅仅是作为用于组织我们的经验并预言未来事件的一种人类构造物而存在？如果情况是前者，那么相关的问题是：我们能否有关于它们的真知识？如何拥有？

这些理论科学中假设的不可观察实体是高度依赖于理论物理和实验观测的。因此在 notes 中才会花费如此大篇幅来介绍光的量子性以及光子在现代理论中的地位。宏观上讲，这个 notes 的动机有两个层面：

- 对于一般读者而言，这个 notes 的目标是帮助他们理解光子这一概念在物理学中的发展路径，通过对光子实在性的具体讨论，了解科学哲学中的结构实在论是如何为类似的理论建构作辩护的，而不为一般性的技术细节所烦扰。
- 对于出身物理系的读者而言，这个 notes 希望为他们提供一部好的科学观念史。目前科学史和科学哲学研究者们已追索出一些十分有趣的东西，为科学知识的客观性和进步作出某些断言提供了一个坚实的基础，也为当前关于科学知识的性质及其历史变革的核心论争话题提供了一个坚实的基础。

有关科学与哲学，许多人抱着这样一种观点——形而上学一旦依附于科学，实质上等于取消了形而上学的独立地位和合法性。隐含在这个断言背后的是这样一种认识，即科学可以独立于形而上学。而笔者则同意 Quine 的主张，科学与哲学是连续的；任何一种形而上学，都应当接受方法论自然主义的指引，应当与科学理论中最坚实的部分相容。如果说逻辑经验主义的兴起从反面揭示出科学中含有形而上学命题，后来的迪昂—蒯因论题、不可通约性理论和认识论的结构实在论则从正面揭示出科学中包含形而上学假设。

此外，这个 notes 的目标是讨论**光子与结构实在论**。我想在开始讨论之前有必要明确一下所需的前置内容。在个人看来，notes 的讨论内容主要涉及三个部分：“物理学史”、“科学哲学”和“场论”。这分别对应了 notes 的三部分前置内容：

- “物理学史”：了解 20 世纪物理学有关量子理论的重大进展和重要事件（参考新概念高中物理读本第三册）。
- “科学哲学”：了解科学哲学中实在论与反实在论的论争（参考牛津通识读本《科学哲学》）。
- “场论思想”：了解现代数学物理中的场论思想，具备相应数理基础更佳。

受上述种种目标约束，文艺批评家所称的“灵气”便不免在文字中隐没，代之的是体系和严格论证所不时带来的滞重感。与滞重相对的这种灵气是哲学之于人心最初的，兴许也是最终的叩鸣。然而依笔者之见，轻盈灵动，单刀直入的写法或见诸理论初创时期，或是依托完备的参考书籍为后援，而这些依托须嵌入原生的语言文化，为一切读者共享。笔者选择负重，但盼望随着本 notes 的面世，未来的作者们能享有轻装上阵的自由。

序曲将终，敬无穷的远方，与无尽的人们。

# Chapter 1

## 光的量子性

经典物理学始于十七世纪的牛顿力学定律，十九世纪由麦克斯韦综合电、磁、光而完成。在这两个世纪中，牛顿力学在解释范围广泛的地表现象和天文观测方面极为成功。麦克斯韦电动力学的重要预言也被赫兹等人的实验所证实，新的应用至今不断涌现。当与热力学定律中的一般统计原理相结合时，经典物理学似乎为未来对物理世界的完备解释提供了一个坚实的基础。

在 20 世纪之交，这种乐观的观点被新的实验发现打破，随之而来的经典物理学危机只有通过创建量子理论才能解决。1909 年卢瑟福发现原子核的存在之后，原子模型引起了人们的兴趣。按照经典的电磁理论，具有加速度的电子会不断辐射电磁波而损失能量，因此原子的行星模型不可能稳定存在。同时对原子光谱的测量结果也令人费解，一种原子似乎只能吸收或发射特定频率的光，如果原子的能量是可以连续变化的，应该可以发出任意频率的光才对。为此，1913 年波尔提出 Bohr 模型来解决原子光谱的问题。Bohr 模型认为原子核外电子的轨道不是任意的，电子只能在满足特定量子化条件的轨道上运动。解释原子稳定性的必要性、原子光谱中离散线的存在、电子的衍射以及许多其他实验观察，决定性地说明新生的量子力学（而不是牛顿力学）更好地解释了微观物质粒子。

有关量子力学的一些后续发展也值得一提。1921 年，测量原子磁矩的斯特恩-盖拉赫实验也发现了奇异的结果。实验用一个非均匀的磁场来偏转带有磁矩的银原子，原子带有的磁矩角度与原子被偏转的角度相关。原本人们预期会看到一个宽阔的斑——因为随机运动的原子磁矩应该可以有各种取向，然而实验结果看到的却是两个条带，这意味着原子磁矩只有两个取向，角动量也是量子化的。

为进一步解决原子光谱的问题，1924 年泡利提出电子具有一个内禀的双值的自由度，随后这一思想被发展为形式化的自旋理论。然而电子自旋仍然是一个奇怪的东西，因为电子为达到自旋角动量，其“转动”会超过光速。因此，在量子力学的范畴中，自旋只是被唯象地描述。直到 1928 年狄拉克提出相对论性的狄拉克方程，描述了相对论性电子的运动，此时电子自然地带有内禀的角动量，与自转无关。

光行为的意外特征构成了经典物理学危机中同样重要的部分。黑体光谱、光电效应和原子光谱被证明与经典电动力学的预测结果不一致。爱因斯坦以其典型的大胆方式提出了一个解决这些难题的方法，即提供一个全新的模型——频率为  $\nu$  的光应由能量为  $\epsilon = h\nu$  的离散光量子组成，其中  $h$  是普朗克常数——光量子的数量密度与光的强度成正比这一假设兼容了经典电磁理论的其他部分。

我们将遵循当前的用法将光量子称为光子，尽管这个术语在此使用并不严谨，它导致概念上的模糊——这个名称暗示光子与电子、质子、中子等粒子意义相同。在接下来的章节中，我们将看到“光子”一词的物理含义随着人们对实验和理论的理解而演变。爱因斯坦对光子的介绍是迈向真正的光量子理论的第一步——正如玻尔原子模型是迈向量子力学的第一步一样——但这些平行发展之间存在重要差异。

## § 1.1 光电效应

在本节和接下来的两节中，我们将考虑由具有明确定义的频率辐射引起的影。其中最著名的一种是光电效应：紫外线照射在经过适当清洁的金属表面上会导致电子发射。在光谱学的早期，这种紫外光源通常是在汞弧中激发的尖锐汞线（波长为 253.6 nm）。

为了简化对这种效应的经典分析，我们将用电子被困在势阱中的模型代替更复杂的实际金属。根据麦克斯韦理论，入射光是电磁平面波， $|\mathcal{E}| = c|\mathcal{B}|$ ，电子受到洛伦兹力  $\mathbf{F} = e(\mathcal{E} + \mathbf{v}\mathcal{B})$ 。功仅由电子上的电场完成。因此，电子需要时间从场吸收足够的能量以克服与金属的结合能，从而从表面逃逸。随着场强的降低，所需的时间必然会增加。由于发射电子的动能是所做的功与结合能之间的差值，因此它还取决于光的强度。

这导致以下两个预测：

(P1) 在辐射开始和电子首次发射之间会有一个与强度相关的时间间隔。

(P2) 发射电子的能量将取决于强度。

现在假设我们有一个实验装置，它可以测量射出的光电子的动能和光到达与电子首次发射之间的时间延迟（可以通过将集电板放置在离表面很近的地方来实现）。相对于表面，该板保持负电势  $V_{\text{stop}}$ ，并将电势调整到刚好足以停止发射电子的值。可以通过能量守恒方程确定光电子的动能

$$\frac{1}{2}mv^2 = (-e)(-V_{\text{stop}})$$

由捕获光电子引起的电流的开始决定了辐射脉冲到达和光电子发射开始之间的时间延迟。电流的幅度与电子射出的速率成正比。实验结果如下：

(E1) 在第一个电子发射之前没有可测量的时间延迟。

(E2) 射出的光电子的动能与光强无关。

相反，能量的观测值取决于频率。它们非常准确地符合经验关系：

$$\epsilon_e = eV_{\text{stop}} = \frac{1}{2}mv^2 = \hbar\omega - W \quad (1.1)$$

其中  $\omega$  是光的频率。常数  $W$  称为功函数；它是从金属中释放电子所需的能量。 $W$  的值取决于金属，但常数是通用的。实验结论为：

(E3) 电子发射的速率——但不是它们的能量——与场强成正比。

理论预测 (P1) 和 (P2) 与实验结果 (E1)–(E3) 明显的不符对经典物理学提出了一个严峻的挑战。特别地，关系式(1.1)称为爱因斯坦光电方程。

在光电效应的早期实验中，很难确定光电子能量是否更符合光频率的线性或二次相关性。这个困难被米利肯的漂亮实验解决了。在这个实验中，他通过使用碱金属验证了等式(1.1)，这些碱金属是在真空系统中通过真空金属抛光技术制备的，具有干净的面。这些干净的碱金属表面具有足够小的功函数  $W$ ，因此即使是朝向可见光谱红色部分的光（这部分光的波长比较长，频率低）也能够发射光电子。通过这种方式，他能够测量从光谱的红色到紫外线部分的光电效应——比之前观察到的频率范围增加了近三倍。这使得验证光电子射出能量增量与入射光频率的线性相关性成为可能。此外，密立根在他的油滴实验中已经非常准确地测量了电子电荷  $e$  的值。将此与方程式(1.1)中  $V_{\text{stop}}$  与  $\nu$  的斜率  $\frac{h}{e}$  相结合，能够推断出普朗克常数  $h$  的值，该值与最佳现代测量值的偏差在 1% 以内。

## § 1.2 康普顿散射

随着光与物质相互作用的研究扩展到更短的波长，另一个令人费解的结果出现在单色 X 射线的散射实验中，图(1.1)展示了散射角为  $\theta$  时的情形。

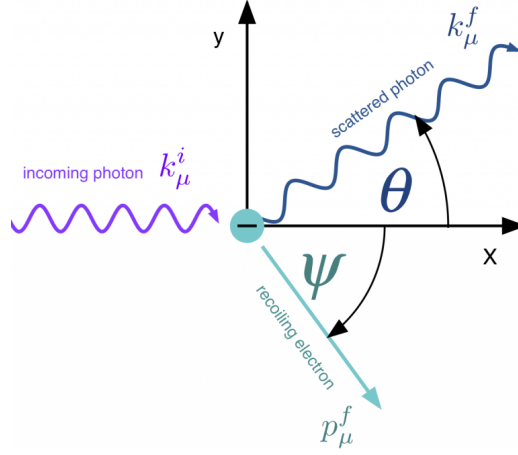


图 1.1: Compton Scattering

散射辐射的波长  $\lambda$  通过布拉格晶体光谱仪使用关系式  $2d \sin \psi = m\lambda$  测量，其中  $\psi$  是布拉格散射角， $d$  是晶体的晶格间距， $m$  是对应于衍射级的整数。康普顿的实验设定  $m = 1$ 。康普顿为他的实验建造的布拉格光谱仪由放置在铅盒内的可倾斜方解石晶体（以布拉格角  $\psi$  定向）组成，用作屏蔽不需要的背景 X-射线。检测器是一个电离室，也在这个盒子内，放置在一系列准直狭缝后面以定义角度  $\theta$  和  $\psi$ 。

该实验的一个简单经典模型由落在原子电子上的频率为  $\omega$  的电磁场组成。根据经典理论，入射场会使电子以频率  $\omega$  振荡，进而产生同频率辐射。这个过程称为汤姆森散射。实际上，入射辐射并不是完全单色的，但光谱确实有一个明确定义的峰值。经典的预测是散射辐射的光谱也应该在相同频率处有一个单峰。图 (1.2) 中  $\theta = 45^\circ, 90^\circ$  和  $135^\circ$  散射角的实验结果确实在入射波长处出现峰值，但每个散射角处，在较长波长处都有一个额外的峰值，这是经典理论无法解释的。

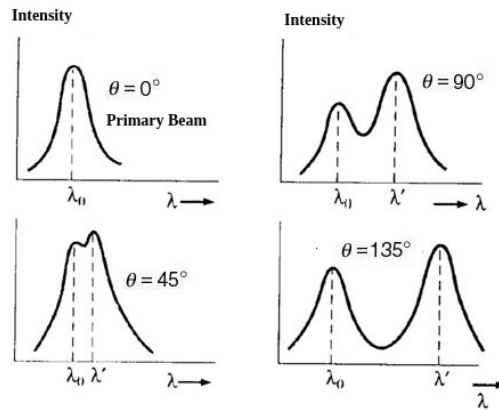


图 1.2: Data from the Compton scattering experiment

### § 1.3 光子的引入

在 1905 年的三篇著名论文中，爱因斯坦提出了一种新的光模型，该模型解释了前面章节中讨论的所有实验结果。

在这个模型中，频率为  $\omega$  的光应该由能量为  $\epsilon = \hbar\omega$  的离散光子凝胶组成。与物质粒子一样，光子携带动量和能量。在第一篇关于相对论的论文中，爱因斯坦已经指出，支配能量和动量的相对论变换定律与支配平面波的频率和波矢的变换定律相同。换句话说，四分量矢量  $(\omega, c\mathbf{k})$  的变换方式与物质粒子的  $(E, c\mathbf{p})$  相同。因此，光量子的能量为  $\hbar\omega$  的假设意味着其动量必须为  $\hbar\mathbf{k}$ ，其中  $|\mathbf{k}| = (\omega/c) = (2\pi/\lambda)$ 。

这是普朗克思想的深远延伸，即能量只能以  $\omega$  为单位在辐射和物质之间传递。新提议将量子化完全归因于电磁场本身，而不是归因于光与物质之间的能量交换机制。将模型的结果分为两组是很有用的：

1. 第一组包括模型的运动学特征，即那些仅依赖于能量和动量守恒定律以及其他对称属性的特征。
2. 第二组包括动态特征，即那些涉及对基本相互作用的明确假设的特征。

接下来我们用光子模型来解释前面提到的两个实验结果。

#### 1.3.1 A 光电效应

光子模型的第一个成功是它对光电效应令人费解的特征的解释。由于光的吸收是通过转移大小恰到好处的离散能量束发生的，因此在第一个电子发射之前没有时间延迟。金属吸收单个光子将其全部能量  $\hbar\omega$  转移到束缚电子上，从而以等式 (1.1) 给出的能量  $\epsilon_e$  从金属中射出它，(1.1) 现在也就表示整体能量守恒。因此，射出电子的能量取决于频率而不是光的强度。由于每次光电子的发射都是由单个光子的吸收引起的，因此每单位时间发射的电子数与光子通量成正比，从而与光强成正比。

光子模型所隐含的光电方程本质上是运动学的，因为它仅取决于能量守恒，并且不假设任何模型用于描述光子与金属中电子间的动态相互作用。

#### 1.3.2 B 康普顿散射

基于动量和能量守恒的运动学论证也预测了康普顿散射中第二个峰的存在。考虑从弱束缚电子散射的 X 射线光子。在这种情况下，考虑一个静止的自由电子并施加能量和动量守恒来确定可能的最终状态就足够了，如图 (1.3) 所示。

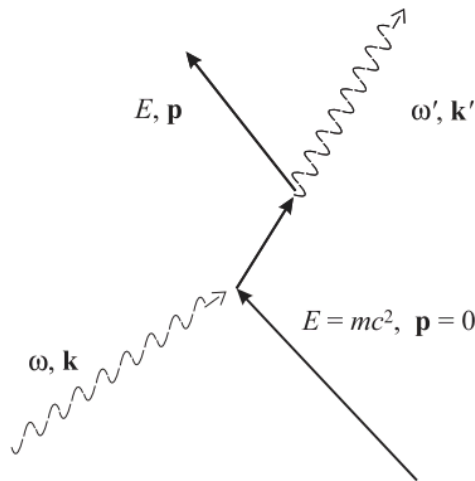


图 1.3: Scattering of an incident X-ray quantum



对于高能 X 射线，电子可能会以与光速相当的速度反冲，因此有必要使用相对论形式进行此计算。能量和动量的相对论守恒定律是

$$mc^2 + \hbar\omega = E + \hbar\omega', \hbar\mathbf{k} = \hbar\mathbf{k}' + \mathbf{p} \quad (1.2)$$

其中  $\mathbf{p}$  和  $E = \sqrt{m^2c^4 + c^2p^2}$  分别是最终的电子动量和能量， $|\mathbf{k}| = \omega/c, |\mathbf{k}'| = \omega'/c$ . 由于散射电子的反冲动能 ( $K = E - mc^2$ ) 应当为正，方程 (1.2) 便解释了为什么散射量子必须具有比入射量子更低的频率（更长的波长）。结合这两个守恒定律产生康普顿位移

$$\Delta\lambda \equiv \lambda' - \lambda = \lambda_C(1 - \cos\theta) \quad (1.3)$$

波长  $\lambda_C$  是散射角  $\theta$  ( $\mathbf{k}$  和  $\mathbf{k}'$  之间的角度) 的函数，其中电子的康普顿波长为

$$\lambda_C = \frac{h}{mc} = 0.0048\text{nm} \quad (1.4)$$

这个简单的论证与图 (1.2) 中的数据以及使用各种入射波长的其他实验非常准确地吻合。康普顿散射的分数波长偏移受  $\Delta\lambda/\lambda < 2\lambda_C/\lambda$  的限制。这表明  $\Delta\lambda/\lambda$  对于光波长  $\lambda \sim 10^3\text{nm}$  可以忽略不计；这解释了为什么需要 X 射线观察康普顿位移。

引出方程式 (1.3) 的方式似乎留下了一些问题——它没有为入射波长处的峰值留下空间，这一点在数据中也很明显。这是假设电子是弱束缚的后果。在对强束缚电子进行相同的运动学分析时，方程 (1.4) 中的电子质量  $m$  必须用原子质量  $M$  代替。由于  $M \gg m$ ，由此产生的偏移即使在 X 射线波长下也可以忽略不计，并且可以恢复入射波长处的峰值。

## § 1.4 波粒二象性

在先前的光子模型中，我们似乎默认光子是经典意义下的粒子，尽管这种想法在解释光电效应和康普顿散射时不存在问题，但光子本质上是一种量子实体，具有波粒二象性。它不能够单纯地由经典物理意义的粒子或波来描述。

可以从德布罗意物质波的角度来说明量子实体具有类波函数性。考虑一个粒子，静质量为  $m_0$  相对于粒子静止的参考系为  $S'$ ，参考系  $S'$  相对于参考系  $S$  正以速度  $v$  向  $x$  轴正向移动。假设有一波函数  $\varphi = A \sin(\omega_0 t_0)$  与这一粒子有关，满足  $E = \hbar\omega_0$ ，其中  $\omega_0 = \frac{m_0 c^2}{\hbar}$ ， $t_0$  为粒子固有时。

根据洛伦兹变换，参考系  $S'$  的时间  $t_0$  与参考系  $S$  的坐标关系为  $t_0 = \frac{t - vx/c^2}{\sqrt{1 - \beta^2}}$  ( $\beta = v/c$ )。

代入波函数中有：

$$\varphi = A \sin(\omega t - kx), \quad \omega = \frac{\omega_0}{\sqrt{1 - \beta^2}}, k = \frac{\omega v}{c^2}$$

波的相速度为  $v_p = \frac{\omega}{k} = \frac{c^2}{v}$ ，通过质能关系

$$E^2 = m^2 c^4 + p^2 c^2, p = \hbar v \frac{v^2}{c} = \frac{h}{\lambda}$$

可以得到  $v^2 = (\frac{mc^2}{\hbar})^2 + c^2 k^2, k = \frac{1}{\lambda}$  是波矢。实际的传播速度应当为群速度  $v_g$ ：

$$v_g = \frac{\partial \omega}{\partial k} = \frac{\partial \omega}{\partial \beta} / \frac{\partial \beta}{\partial k} = \beta c = v, \quad k = \frac{\omega v}{c^2} = \frac{\beta \omega_0}{c \sqrt{1 - \beta^2}}$$

与实际中粒子的速度  $v$  相符。综上可以得到对于实物粒子，

$$E = \frac{E_0}{\sqrt{1 - \beta^2}} = \frac{\hbar \omega_0}{\sqrt{1 - \beta^2}} = \hbar \omega$$



$$\lambda = \frac{v_p}{\nu} = \frac{c^2}{\nu v} = \frac{2\pi c^2}{v\omega} = \frac{h\sqrt{1-\beta^2}}{m_0 v} = \frac{h}{p}$$

这说明微观粒子同时也具有波函数性质，也就是所谓的德布罗意关系。

特别地，在量子场论中，上述过程实际上是电磁场间的相互作用，光子是电磁场的激发态。而各种量子场(或导出的产生湮灭算符)被视为是基本物理实体。

这里给出一个平凡的例子，只考虑**原子和单模光场**的相互作用，可以使用初等量子力学的语言来描述：

所谓单模光场，即光场只有一个模式。例如在完美的单模光纤或者单模谐振腔中，原子只与一个光场模式发生相互作用。在量子力学中，我们学过对谐振子的哈密顿量进行量子化，此时产生湮灭算符(或者叫做阶梯算符)和谐振子位移算符之间存在如下关系：

$$\hat{x} = x_{\text{zpf}}(\hat{a} + \hat{a}^\dagger)$$

其中  $x_{\text{zpf}}$  为位移的基态涨落(即零点能要求基态时谐振子不可能具有确定的位移)， $\hat{a}, \hat{a}^\dagger$  为声子的湮灭算符和产生算符。在量子光学中，也有类似的结论：

$$\hat{E} = E_{\text{zpf}}(\hat{a} + \hat{a}^\dagger)$$

其中  $E_{\text{zpf}}$  为电场的基态涨落， $\hat{a}, \hat{a}^\dagger$  为光子的湮灭算符和产生算符。从这一步我们就可以看出光子与电场的联系，只要我们知道光子态，就可以通过上述算符算出电场。

从光子态通过电场算符得到电场后，如果我们假设电场的空间变化率远小于原子尺度的倒数，那么在原子范围内我们可以认为电场为匀强场(实际上这一近似是非常合理的，只有对于 X 射线级别的高频电磁场才不能忽略其场强在原子尺度的变化)，因此光场-原子系统的总哈密顿量为

$$\hat{H} = \hat{H}_0 + \hbar\omega(\hat{a}^\dagger\hat{a} + 1/2) + \mathbf{E}_{\text{zpf}}(\hat{a} + \hat{a}^\dagger) \cdot \hat{\mathbf{p}}$$

其中第一项为孤立原子的哈密顿量，第二项为光场的哈密顿量(实际上和谐振子哈密顿量一模一样)，第三项为原子-光子相互作用项， $\hat{\mathbf{p}}$  为原子偶极矩。在这一模型下我们可以得到一些很有意思的结论。

例如考虑一个处于激发态的二能级原子，其能级间隔为  $\Delta E$ ，那么当其与一个频率为  $\Delta E/\hbar$  的基态单模光场相互作用时，原子与光场会自发的在激发态与基态之间震荡。这一现象比较反物理直觉，因为这实际上与我们认知中的原子自发辐射模型非常相似，而自发辐射是不可逆的。其实这二者的区别就在于前者为单模光场，而后者为多模，甚至是模式稠密的光场相互作用，因此后者会引入信息丢失和量子退相干的概念，导致了后者的不可逆。

## Chapter 2

### 光子必要吗？

现在我们已经确定光子模型足以解释第一章中描述的两个实验，但引入光子这一做法是否必要呢？是否存在其他方式能够解释这些现象？

从经典电磁理论到光子模型的转变比从经典力学到量子力学的相应变更为激进。如果把经典力学看作象棋这样的游戏，棋子是点粒子，规则是牛顿运动方程。量子力学的棋子相同，规则不同。经典电动力学的情况就大不相同了。这个游戏的棋子是连续的电场和磁场  $\mathcal{E}(\mathbf{r}, t)$  和  $\mathcal{B}(\mathbf{r}, t)$ ，规则由麦克斯韦方程组提供，爱因斯坦的光子无处可寻；因此，游戏的量子版本需要新的棋子和新的规则。应谨慎对待如此巨大的概念变化。换句话说，光子模型足以解释前面所提到的实验，但问题是是否有必要为此目的引入光子。要弄清这个问题的唯一方法是寻找一个替代模型——而目前唯一可用的候选模型是半经典电动力学。在这种方法中，构成原子的带电粒子由量子力学描述，但电磁场仍然采用经典方法处理。

因此，在第二章我们将尝试用半经典理论解释每个实验。在这方面，必须记住，对半经典理论或光子模型的最低阶近似的修正在早期实验中是检测不到的。正如我们将看到的，这些尝试有不同的成功程度；有人可能会问：为什么要考虑半经典方法呢？答案是，对给定实验结果存在半经典解释意味着该实验对光子的**不可分性**——这是爱因斯坦模型的基本假设一不敏感。

### §2.1 重回光电效应

爱因斯坦对光电效应的解释非常直接地依赖于光子概念。然而，事实证明，使用半经典理论的替代描述要简单得多。对于此计算，金属中的电子由量子力学描述，光被描述为外部经典场。因此，电子的总哈密顿量是  $H = H_0 + H_{\text{int}}$ ， $H_0$  是在没有任何外部电磁场的情况下电子的哈密顿量， $H_{\text{int}}$  是相互作用项。对于弱外场中的单个电子，标准量子力学结果为：

$$H_{\text{int}} = -\frac{e}{m}\mathcal{A}(\hat{\mathbf{r}}, t) \cdot \hat{\mathbf{p}}$$

$\hat{\mathbf{r}}$  和  $\hat{\mathbf{p}}$  分别是位置和动量的量子算符。

在通常的位置空间表示中，算符的作用是  $\hat{\mathbf{r}}\psi(\mathbf{r}) = \mathbf{r}\psi(\mathbf{r})$  和  $\hat{\mathbf{p}}\psi(\mathbf{r}) = -i\hbar\nabla\psi(\mathbf{r})$ 。常量函数  $\mathcal{A}(\hat{\mathbf{r}}, t)$  是经典矢量——可以选择来满足辐射规范条件  $\nabla \cdot \mathcal{A} = 0$ ——它通过以下方式确定辐射场：

$$\mathcal{E} = -\frac{\partial \mathcal{A}}{\partial t}, \quad \mathcal{B} = \nabla \times \mathcal{A}$$

对于频率为  $\omega$  的单色场，矢量势为

$$\mathcal{A}(\mathbf{r}, t) = \frac{1}{\omega}\mathcal{E}_0\mathbf{e}\exp(i\mathbf{k} \cdot \mathbf{r} - \omega t) + \mathcal{C}$$

其中  $\mathbf{e}$  是单位极化矢量， $\mathcal{E}_0$  是电场幅度， $|\mathbf{k}| = \omega/c$  且有  $\mathbf{e} \cdot \mathbf{k} = 0$ ， $\mathcal{C}$  是常数项。

应用费米黄金定则可以得到速率：

$$W_{fi} = \frac{2\pi}{\hbar} |\langle f | H_{\text{int}} | i \rangle|^2 \delta(\epsilon_f - \epsilon_i - \hbar\omega) \quad (2.1)$$

对于从初始结合能级  $\epsilon_i$  到自由能级  $\epsilon_f$  的转变。该结果对观察时间  $t \gg 1/\omega$  有效。对于  $\omega \sim 10^{15}\text{s}^{-1}$  的光场，方程(2.1) 预测电子的发射没有明显的延迟。此外， $\delta$  函数保证了射出电子的能量满足光电方程。最后，基元素  $\langle f | H_{\text{int}} | i \rangle$  与  $\mathcal{E}_0$  成正比，因此电子发射速率与场强成正比。

综上，这个简单的半经典理论解释了光电效应所有令人费解的方面，而没有引入光子的概念。方程(2.1)中的能量守恒  $\delta$  函数再现了运动学关系(1.1)，但它只出现在详细的动力学计算的末尾。

大多数检测光子的技术都采用光电效应，因此不需要光子存在的光电效应的解释有点令人沮丧。此外，其他种类的探测器（如感光乳剂、固态光电倍增管等）的响应归根结底也是基于光电效应。因此，它们也可以完全用半经典理论来描述。

## § 2.2

### 康普顿散射新解

上一章中给出的康普顿位移的运动学解释通常作为光子存在的确凿证据，但著名的 Klein-Nishina 公式 [下面的式子(2.2)] 关于康普顿散射的微分截面的推导，是在半经典近似的一个稍微扩展的形式中进行的。

$$\frac{d\sigma}{d\Omega} = \frac{\alpha^2}{2m^2} \left( \frac{\omega'}{\omega} \right)^2 \left[ \frac{\omega'}{\omega} + \frac{\omega}{\omega'} - \sin^2 \theta \right] \quad (2.2)$$

这种分析比半经典的光电效应分析更为复杂，原因有二：

1. 首先，电子的运动可能会变成相对论性的，因此非相对论性的薛定谔方程必须被相对论性的狄拉克方程所取代。
2. 第二个复杂的问题是受激电子发出的辐射不能被忽略，因为观察这种辐射是实验的重点。

因此，康普顿散射是一个两步过程，其中电子首先被入射辐射激发，随后产生的电流产生散射辐射。在 Klein 和 Nishina 的原始论文中，电子在入射平面波作用下的狄拉克方程是使用一阶含时微扰理论求解的。在经典的麦克斯韦方程组中，电流密度算符在扰动状态下的期望值被用作源项。用这种方法产生的辐射场自动满足运动学关系(1.2)，因此它也可以得到公式(1.3)中给出的康普顿位移。

此外，使用半经典 Klein-Nishina 模型计算的康普顿截面与量子电动力学<sup>1</sup>中得到的结果完全一致。我们再一次看到爱因斯坦的光子模型对实验的运动学方面提供了一个美丽而简单的解释，但更复杂的半经典处理达到了同样的目的，同时还提供了正确的横截面动力学计算。在这个计算中引入光子的概念是非必要的。

## § 2.3

### 结论

第一章中讨论的实验通常作为光子存在的证据。这种说法背后的原因是经典物理学与实验结果不一致，而爱因斯坦的光子模型以非常简单的方式描述了所有实验结果。然而，我们看到，电动力学的半经典版本可以解释同一组实验，而无需求助于光子的概念。这就引起了对一些声称证明光子存在的实验的解释的严重质疑。

一个早期的例子是 1909 年 Taylor 的单光子双缝干涉实验，其中使用非常低强度的光——在任意给定时间设备中存在的平均能量最多是  $\hbar\omega$ 。结果是感光底板上的斑点缓慢堆积。经过足够长的曝光时间，光斑显示出预期的双缝干涉图案。这被认为是光子存在的证据。显然，这也是狄拉克断言每个光子只与自身相干的基础。这种解释显然依赖于这样的假设，即板上的每一个光点都代表着单个光子的吸收。光电效应的半经典解释表明，该结果同样可以很好地解释为来自两个狭缝的经典电磁波的干涉，与感光板中电子激发的半经典量子理论结合。在这种观点下，光子的概念就没有必要了，因此对电磁场的量子化也就没有必要了。

那么，光子存在的实验证据在哪里呢？在下一章中，我们将描述与这个问题有关的实验。

<sup>1</sup>其中电磁场由量子理论处理。

# Chapter 3

## 光子存在的实验证据

经典的电磁理论用电场和磁场描述光，而光的光子模型则认为电磁能集中在离散的量子中，不能进一步细分。特别是，一个经典的电磁波必须在分束器上连续可分，而一个不可分的光子必须作为一个整体完全透射或完全反射。经典波的连续可分和光子的不连续反射或透射选择是相互排斥的；因此，涉及单个量子的光入射到分束器时的表现，光的量子理论和经典理论给出了完全不同的实验预测。

光子的不可分性是爱因斯坦光子模型最核心的一个假设，因为即使是最复杂的半经典理论也是用不断变化的经典场来描述光的。正如我们在第一章和第二章中所见：虽然能量量子化的光粒子型激发图景，给出了光电效应简单而迷人的解释，爱因斯坦还因此获得了 1921 年的诺贝尔奖。但它也可用半经典理论解释，其中光被描述成经典的电磁波，同量子化的能级系统相互作用。所以决定性的实验必须依赖于单个光子的不可分性。

借助光子的不可分性和光子统计学，在后来的实验中人们得到了不借助光子概念<sup>1</sup>就无法解释的结果。

本节在初等水平讨论其中一些实验，领略一下过去几十年来光学研究相当活跃的一个领域。这些实验背后的理论通常用量子算符处理，但因为我們只讨论物理，将避免使用它们。毕竟，它们只是数学工具。

### §3.1 亚泊松光

即使是非常稳定的激光，所出射的光子分布也是不均匀的。根据光场的统计性质，可以把光场分为三类：泊松，超泊松，亚泊松。

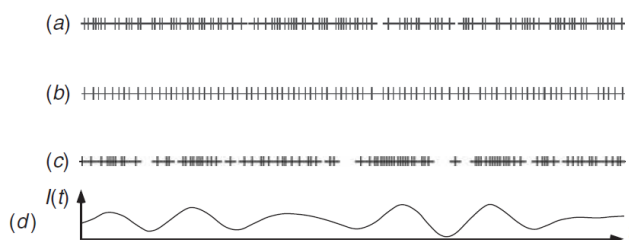


图 3.1: Light Intensity Fluctuation

图(3.1)(a) 是泊松光模拟，(b) 是亚泊松光而 (d) 是超泊松事件沿时间轴的模拟。(c) 对应 (d) 中显示的涨落光强，对非相干光是典型的。混沌光符合泊松统计，单色激光束是它的一个很好的实例。这意味着，光强的方均根涨落正比于光强的平方根。此外，Brown 和 Twiss (1956) 的实验显示，相邻频率成拍引起准单色光中增强的涨落。这些称为“超泊松”涨落。这两个现象都能按半经典理论，用经典光波和光探测器间相互作用的统计解释。在这种方法下，给定强度的电磁波扰动光探测器中的电子，它们随后统计地激发到将被探测到的更高能级。然而，这种半经典模型无法解释“亚泊松”光——涨落小于泊松统计预言的光。

<sup>1</sup>严格意义上说，光子的定义是具有零质量，能量  $\hbar\omega$  和自旋 1，在真空中以速度  $c$  运动的非局域实体。



在 (a) 中，我们看到以速率  $r$  发生的一系列非相关光子事件。在 (b) 中，我们取序列 (a)，速率加倍到  $2r$ ，并在每次事件后引入平均值  $1/r$  的探测死区时间，这段时间中任何发生的事件都被擦除，因此平均速率仍为  $r$ 。不难明白，我们得到了更稳定的亚泊松光子流，等价于上述第二种方法。在 (c) 中，我们先生成经典混沌波（显示在 (d) 中），再根据泊松统计生成合适的超泊松光子序列。

光场最常用到的态是所谓的**相干态**，这种状态可以用来近似描述高于激光阈值的单频激光器的输出。光场的统计性质可以看作是对光相干态基于统计物理学的描述。相干态是最接近经典的态，其在 Fock 态表象上满足的是泊松分布：

$$P(n) = \frac{|\alpha|^{2n}}{n!} e^{-|\alpha|^2}, \quad \sqrt{\bar{n}} = \Delta n = |\alpha|$$

事实上，在这一过程中我们假设了激光的“光子分布”在任何时刻都是等概率的以及  $\delta P = \omega \delta Q$  ( $P, Q$  是线性腔中光平面波模式量子化后的相空间坐标)。

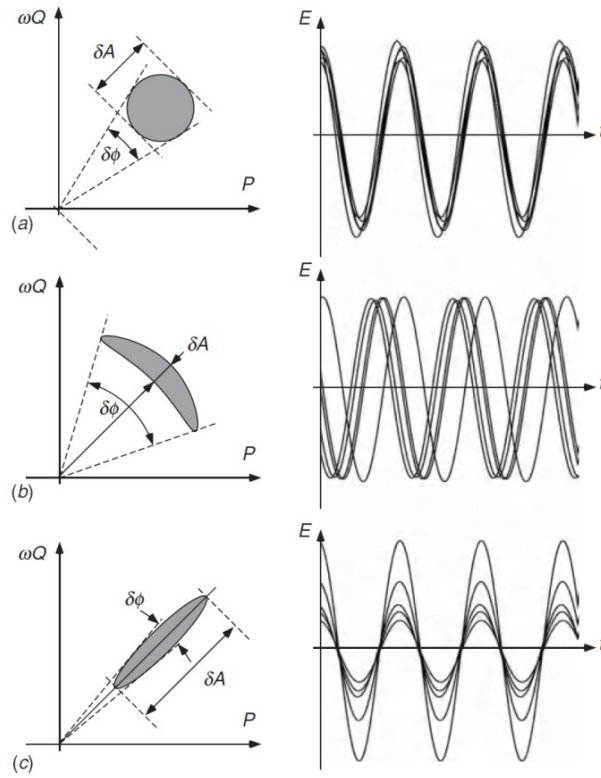


图 3.2: Wigner Graph

图(3.2)(b) 显示了不确定区域形状被扭曲，以相位涨落增强的代价，振幅涨落减弱的例子。这只能在光阴极中的电子发射统计受正则化光子束调整时实现。

目前已经发明了好几种实现方式，如：

- 以稳定的恒流（抽运）源运行半导体激光器或发光二极管。这些电子是费米子，趋于反成团，形成更有序的电子流，由于结中电子空穴对寿命很短，它正则化了光子输出。
- 从单个原子或离子发射共振荧光，每次发射后都必须再激发。再激发造成每次发射后有“死区时间”，通过创造每个光子对先前光子的依赖，光滑化了光子流。

仅当光以一个一个量子发射出来时，这些方法才能实现，因此亚泊松光无法经典描述。

### §3.2 光子反相关

Brown 和 Twiss (1956) 的实验调查了准单色束经分束器产生的两束光间的光强关联。我们看到过，这关联可由二阶相干函数  $\gamma^{(2)}(\tau)$  描述，它是探测时延  $\tau$  的函数；对经典波模型， $\gamma^{(2)}(\tau)$  同时间相干函数  $\gamma(\tau)$  有关系：

$$\gamma^{(2)}(\tau) = 1 + |\gamma(\tau)|^2$$

因此，经典地， $\gamma^{(2)}(\tau)$  的值不能小于 1，那么  $\gamma^{(2)}(\tau) < 1$  的实验结果就表明需要量子描述。测量技术的改进将 Brown 和 Twiss 的光强关联方法换成光电子符合计数。这时考虑在  $t = 0$  时接收到一个光电子后，测量在  $t = \tau$  和  $t = \tau + \delta\tau$  的窗口中探测器  $D_2$  再接收到一个光电子的概率。正确归一化后，就得到  $\gamma^{(2)}(\tau)$  的分布图像。对上述类型的源，实验已经确凿显示可能出现  $\gamma^{(2)}(\tau) < 1$ ；这基本上确立了光子作为量子实体的实在性。

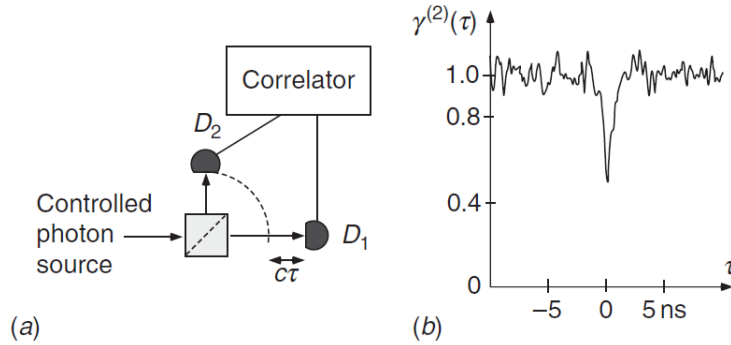


图 3.3: Photon Correlation

不难解释为何会出现这样的结果：

- 考虑泊松统计的光束（图(3.1)(a)），光子在任意时间区间  $\delta\tau$  内到达的概率恒定为  $p\delta\tau$ 。如果光束一分为二，如图(3.3)(a)，光子随机前往  $D_1$  或  $D_2$ ，到达各探测器的概率现在是  $p\delta\tau/2$ ，但  $D_1$  处事件的分布不依赖于  $D_2$  处的，因此对所有  $\tau$  有  $\gamma^{(2)}(\tau) = 1$ 。
- 如果光束是经典超泊松的（如图(3.1)(c)）， $\tau < \tau_c$  [ $\tau_c$  是相干时标]，那么给定时间窗口中  $D_1$  处光子到达的概率高于涨落峰值处， $D_2$  处同样；因此  $\gamma^{(2)}(|\tau| < \tau_c) > 1$ 。
- 然而，如果光子是亚泊松的，因此更有序，那么一个光子到达  $D_1$  就意味着，另一个光子在短时间内到达  $D_2$  的可能性更小。对上述第二种源，在死区时间内就有  $\gamma^{(2)}(\tau) = 0$ 。

1977 年首次证实了这种反相关（即  $\gamma^{(2)}(\tau)$ ），可以说是光子概念第一次得到确凿的支持。之后是来自单个小原子簇共振荧光的异常清晰的结果，最近还有来自量子点的（图(3.3)）。后者是微小的离子阱，可以说是人造的单个原子，不仅可以产生“急需的光子”，还有成组的相关或“纠缠”光子，这是我们的下一个主题。

### §3.3 纠缠光子与不可分性

1935 年，EPR 思想实验被爱因斯坦等三人提出，它给出的反直觉结果引出量子力学能否用可测量参数完备描述系统的问题。在量子系统中，一个参数仅当被实际测量时才获得特定的值。例如，Taylor 干涉实验中的弱光可能将在特定的时间区间中，在胶片的某一点上引发一事件，但不可能预先确定是哪个点。EPR 描述的实验中，两粒子由于自发衰变从源释放。因为质心保持静止，两粒子的动量和位置之和必须始终保持为零 ( $p_2 = -p_1, x_2 = -x_1$ )。于是，通过测量  $x_1$  和  $p_2$ ，可以同时确定  $x_2$  和  $p_1$ ；那么两粒子的位置和动量就都同时知道了。海森堡不确定性原理不容许这点！此外，这些参数是在两粒子相隔很远时测量的，所以它们无法彼此影响。这称为 EPR 悖论。

在光学中，利用光子对进行了这样的实验，它们由单次事件生成，因此偏振相关。测量结果显示——测量出一个光子的偏振，确实就能正确预言另一个光子的偏振，即使它们相隔很远。这有两种可能的解释：

1. 第一种可能，存在“隐变量”，光子在发射时被注入了内禀的偏振，但进行测量后我们才知道是什么。
2. 第二种可能，一个光子在测量时获得偏振，同时另一个光子由于相关也获得了。

考虑到测量时光子相隔很远无法通信，它是如何获知第一次测量结果的？这样的问题在量子力学基础的研究中扮演至关重要的角色，实验的设计和验证者也因此获得了2022年的诺贝尔奖。第一个解释假设存在隐变量，我们直到第一次测量才知道它的值。第二个解释要求两光子共享一波函数，它不能分解成两独立函数之积，参数在对一个粒子测量后才获得值；但结果对两个粒子都适用！我们称这样的粒子存在量子纠缠。

这项实验共分为三个部分，第一、二部分同量子纠缠，量子预测与 Bell 不等式息息相关且分析流程颇为复杂。我们建议读者参考原始论文来了解这项开创性的工作，在这里简要描述第三个实验，该实验为单个光子的不可分性提供了最清晰、最直接的证据，从而证明了电磁场中单个量子的存在。

实验装置(图(3.4))在位于 S 的钙原子中采用了三能级级联(图(3.5))。两个连续的相关光脉冲（以频率  $\nu_1$  和  $\nu_2$  为中心）在与源相反的方向。

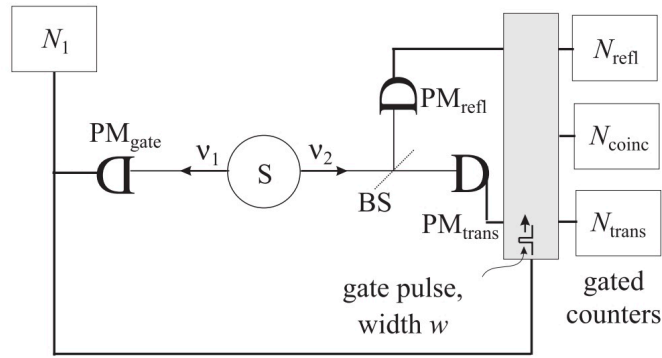


图 3.4: The photon-indivisibility experiment

在论证的这一点上，我们保留了光被经典电磁波描述而不是光子描述的可能性，并假设探测事件完全可以被半经典光电效应理论描述。

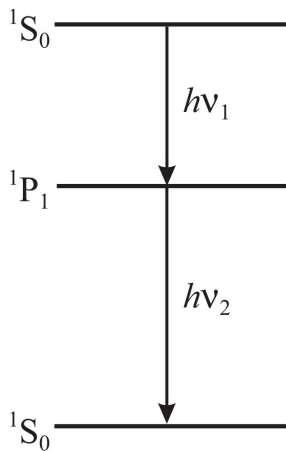


图 3.5: Three-level cascade in a calcium atom

通过原子束传输的原子在到达源区 S 前不久被激发到最高能级。光电倍增管  $\text{PM}_{\text{gate}}$  配备了一个滤光片，可以滤除第二次跃迁频率为  $\nu_2$  的辐射，同时允许第一次跃迁频率为  $\nu_1$  的辐射通过。 $\text{PM}_{\text{gate}}$  的输出监视频率为  $\nu_1$  的辐射脉冲，由计数器  $N_{\text{gate}}$  记录，并且还用于激活（触发）称为门发生器的设备，该设备在指定时间内产生标准化的矩形门脉冲区间， $T_{\text{gate}} = w$ ，称为门宽。光电倍增管  $\text{PM}_{\text{refl}}$  和  $\text{PM}_{\text{trans}}$  的输出监测频率为  $\nu_2$  的辐射脉冲，仅在门宽  $w$  指定的时间间隔内由门控计数器  $N_{\text{refl}}$  和  $N_{\text{trans}}$  记录。

如果检测到  $\nu_1$  频率的辐射突发，则来自第二跃迁的频率为  $\nu_2$  的辐射突发必然指向分束器 BS，分束器 BS 部分反射并部分透射落在其上的光。以这种方式产生的两束光束指向两个光电倍增管  $\text{PM}_{\text{refl}}$  和  $\text{PM}_{\text{trans}}$ 。

$\text{PM}_{\text{refl}}$  和  $\text{PM}_{\text{trans}}$  的输出用于驱动门控计数器  $N_{\text{refl}}$  和  $N_{\text{trans}}$ ，它们记录来自两个光电倍增管的每个脉冲，还用于驱动符合计数器  $N_{\text{coinc}}$ ，它仅在这两个光电倍增管同时产生电流脉冲时响应指定开门时间间隔  $w$ 。

因此，单个计数器触发的概率（单次概率）由  $p_{\text{refl}} = \dot{N}_{\text{refl}}/\dot{N}_{\text{gate}}$  和  $p_{\text{trans}} = \dot{N}_{\text{trans}}/\dot{N}_{\text{gate}}$  给出，其中  $\dot{N}_{\text{gate}} \equiv \dot{N}_1$  是门打开的速率——光电倍增管  $\text{PM}_{\text{gate}}$  的计数速率—— $\dot{N}_{\text{refl}}$  和  $\dot{N}_{\text{trans}}$  分别是  $\text{PM}_{\text{refl}}$  和  $\text{PM}_{\text{trans}}$  的计数率。重合率  $\dot{N}_{\text{coinc}}$  是在开门间隔  $w$  期间同时触发两个检测器  $\text{PM}_{\text{refl}}$  和  $\text{PM}_{\text{trans}}$  的速率；因此，巧合概率为  $p_{\text{coinc}} = \dot{N}_{\text{coinc}}/\dot{N}_{\text{gate}}$ 。实验包括测量单打计数率  $\dot{N}_{\text{gate}}, \dot{N}_{\text{refl}}, \dot{N}_{\text{trans}}$  和重合率  $\dot{N}_{\text{coinc}}$ 。

根据爱因斯坦的光子模型，每个原子跃迁都会产生一个不可细分的光量子。能量为  $h\nu_2$  的不可分量子从分束器中散射出来，只能被检测到一次。因此它必须去到  $\text{PM}_{\text{refl}}$  或  $\text{PM}_{\text{trans}}$ ；它不能同时去到两个地方。在没有复杂因素的情况下，光子模型可以预测符合概率  $p_{\text{coinc}}$  恰好为零。但由于这是一个真实的实验，复杂的因素也不在少数。在开门区间内，源区 S 中的两个不同原子可能发射出两个量子  $h\nu_2$ ，从而产生一个假符合计数。这种困难可以通过选择门间隔  $w \ll \tau'$  来最小化，其中  $\tau'$  是级联中间能级的寿命，但它不能从这个实验安排中完全排除。

分析这个实验只需要半经典理论的三个一般特征：(1) 原子是由量子力学描述的；(2) 每个原子跃迁都会产生经典场描述的辐射爆发；(3) 光电倍增管电流与入射辐射强度成正比。

前两个特征是半经典理论定义的一部分，第三个特征是光电效应的半经典分析所隐含的。分束器将来自原子的经典辐射转换成两束，一束指向  $\text{PM}_{\text{refl}}$ ，另一束指向  $\text{PM}_{\text{trans}}$ 。因此，根据半经典理论，重合概率不可能为零——即使没有上面讨论的错误计数——因为经典电磁波必须在分束器处平滑地分开。半经典理论预测了一个最小重合率，它与反射强度和透射强度的乘积成正比。落在  $\text{PM}_{\text{refl}}$  和  $\text{PM}_{\text{trans}}$  上的瞬时强度与落在分束器上的原始强度成正比，门控测量有效地平均了开门间隔。因此，在第  $n$  个栅极间隔中产生的光电流与分束器处的时间平均强度成正比：

$$I_n = \frac{1}{w} \int_{t_n}^{t_n+w} dt I(t)$$

其中门在区间  $(t_n, t_n + w)$  中打开。原子跃迁由量子力学描述，因此它们在门间隔内随机发生。这意味着强度  $I_n$  表现出从一个门间隔到另一个门间隔的随机变化。为了最小化这些波动的影响，对一系列门打开的计数数据进行平均。因此，单打概率是根据平均强度确定的

$$\langle I \rangle = \frac{1}{M_{\text{gate}}} \sum_{n=1}^{M_{\text{gate}}} I_n$$

其中  $M_{\text{gate}}$  是门开口的总数。单打概率由  $p_{\text{refl}} = \eta_{\text{refl}} w \langle I \rangle$  和  $p_{\text{trans}} = \eta_{\text{trans}} w \langle I \rangle$  给出。 $\eta_{\text{refl}}$  是检测器效率与原始强度分数的乘积，指向  $\text{PM}_{\text{refl}}$  和  $\eta_{\text{trans}}$  与  $\text{PM}_{\text{trans}}$  的量相同。由于单门中的重合率与  $\text{PM}_{\text{refl}}$  和  $\text{PM}_{\text{trans}}$  的瞬时光电流的乘积成正比，因此重合概率与强度平方的平均值成正比： $p_{\text{coinc}} = \eta_{\text{refl}} \eta_{\text{trans}} w^2 \langle I^2 \rangle$ 。

其中  $\langle I^2 \rangle = \frac{1}{M_{\text{gate}}} \sum_{n=1}^{M_{\text{gate}}} I_n^2$ 。通过使用恒等式  $\langle (I - \langle I \rangle)^2 \rangle \geq 0$  很容易证明  $\langle I^2 \rangle \geq \langle I \rangle^2$ ，结合概率表达式有：

$$p_{\text{coinc}} \geq p_{\text{refl}} p_{\text{trans}} \quad (3.1)$$

这种半经典预测可以通过定义参数方便地表达：

$$\alpha \equiv \frac{p_{\text{coinc}}}{p_{\text{refl}} p_{\text{trans}}} = \frac{\dot{N}_{\text{coinc}} \dot{N}_{\text{gate}}}{\dot{N}_{\text{refl}} \dot{N}_{\text{trans}}} \geq 1 \quad (3.2)$$

其中后一个不等式来自等式(3.1)。将门间隔设置为  $w = 9\text{ns}$ ，并调整原子束电流以产生门速率  $\dot{N}_{\text{gate}} = 8800\text{c/s}$ ，发现  $\alpha$  的测量值为  $\alpha = 0.18 \pm 0.06$ 。



这违反了半经典不等式 (3.2) 13 个标准差；因此，该实验果断地拒绝了任何基于发射的半经典处理的理论。这些数据表明，当被触发电倍增管  $\text{PM}_{\text{gate}}$  的放电门控时，光电倍增管  $\text{PM}_{\text{refl}}$  和  $\text{PM}_{\text{trans}}$  的放电之间存在很强的反相关性。单个光子  $h\nu_2$  在离开分束器后，可能导致光电倍增管  $\text{PM}_{\text{refl}}$  或  $\text{PM}_{\text{trans}}$  中的任何一个发射，但这两种可能的结果是相互排斥的。这个实验令人信服地证明了爱因斯坦光子的不可分性。

细心的读者会发现，这一实验实际上可以看作是上一节光子反相关现象的一个详细描述。

最近，光子不可分性实验中使用的关联光子对的级联发射已被自发下转换所取代。在这种更方便、更紧凑的光源中，原子束——需要大量使用不方便的真空技术——被单个非线性晶体所取代。紫外激光束进入晶体，并将其原子相干激发到虚激发态。随后快速衰减成光子对  $\gamma_1$  和  $\gamma_2$ ，如图 (3.6) 所示。这个过程似乎违反了光子的不可分性，因此我们强调入射的 UV 光子作为一个整体被吸收，另外两个光子也作为一个整体被发射。这些光子中的每一个都能通过先前讨论的实验的不可分性测试。

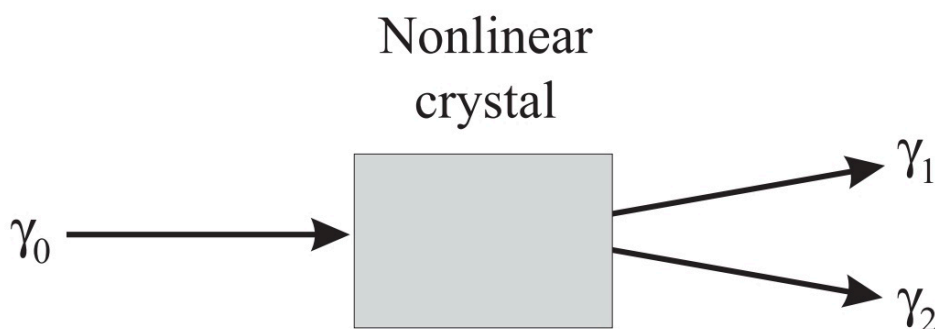


图 3.6: The process of spontaneous down-conversion

正如在一个激发的母核放射性衰变成两个子核的类似过程中，能量和动量在自发下转换中是守恒的。由于非线性晶体的色散和双折射的结合，结果是以多种颜色的彩虹形式高度定向发射光。这条彩虹的独特量子特征是在紫外激光束的相对两侧发射的光子对彼此强烈相关。例如，图 (3.6) 中放置在针孔 1 后面的盖革计数器检测到光子  $\gamma_1$  总是伴随着放置在针孔 2 后面的盖革计数器检测到光子  $\gamma_2$ 。这种光的高方向性源使得相关光子对的收集及其特性的测量比原子束光源的情况简单得多。

除了改进光源外，固态技术还改进了光子探测器。利用基于真空管技术的光电倍增管的光子探测器现已被许多基于半导体晶体中的光伏效应的更简单的固态探测器。进入晶体的光子产生电子-空穴对，然后在强内部电场的作用下被拉开。这个场足够大，以至于光子产生的初始带电粒子对的加速导致晶体内部的雪崩击穿，这可以被认为是由第一对带电粒子引发的多个碰撞电离事件分支组成的链式反应。半导体光电二极管的这种工作模式称为盖革模式，因为它与初始电离粒子通过盖革计数器导致的气体雪崩电离击穿非常相似。

每个雪崩击穿事件都会产生一个大的标准化电脉冲（我们将其称为光子计数器的咔哒声），对应于单个光子的检测。例如，许多当代量子光学实验使用硅雪崩光电二极管，这是一种单光子计数器，在近红外区的量子效率约为 70%。这远高于相同波长区域中光电倍增管的量子效率。固态探测器还具有更短的响应时间（在纳秒范围内），因此可以通过传统电子设备直接实现标准化脉冲的快速符合检测。固态单光子探测器的另一个重要的实际优势是它们需要比光电倍增管低得多的电源电压。

在前三章的内容中，我们已经看到爱因斯坦的光子模型正确地描述了光电效应、康普顿散射和自发辐射，但我们也看到了可以使用半经典电动力学方程的扩展形式来构建对这些现象的合理解释。然而，没有半经典的解释可以解释实验中展示的光子的不可分性；因此，包含不可分性的理论必须基于经典电磁学中未发现的新物理原理。幸运的是，量子理论也必须满足**对应原理**；也就是说，对于经典电动力学正确描述的一大类现象，它必须与经典理论一致。这对构建量子理论是一个无价的帮助。最后，只能通过将量子理论的预测与实验结果进行比较来判断新原理的有效性。第四章将从形式理论（主要是量子场论）的角度来考察光子和电磁场的性质，它将向我们展示光子是如何在标准模型中出现的。

## Chapter 4

### 量子场论下的光子

关于量子力学的公理形式，1920 年代先后有薛定谔的波动力学与海森堡的矩阵力学，后来 1926 年薛定谔证明了两种表述的等价性，于是（非相对论性）量子力学的公理形式便完成了。1948 年，费曼提出路径积分表述，作为第三种构建量子力学的方式，并明确了量子力学与经典力学的关系。至此提出的量子力学是不包含相对论的。关于相对论性量子力学有诸多尝试，大多数都是基于薛定谔方程进行改写。

我们知道薛定谔方程用了类似经典力学的哈密顿量表达式，如果我们使用相对论性哈密顿量：

$$\hat{H} = \sqrt{p^2 c^2 + m^2 c^4}$$

薛定谔方程会变成：

$$i\hbar \frac{\partial}{\partial t} |\phi\rangle = \sqrt{-\hbar^2 c^2 \nabla^2 + m^2 c^4} |\phi\rangle \quad (4.1)$$

我们对(4.1)两边同时乘方：

$$-\hbar^2 \frac{\partial^2}{\partial t^2} |\phi\rangle = (-\hbar^2 c^2 \nabla^2 + m^2 c^4) |\phi\rangle \quad (4.2)$$

这个看起来好像不错的方程(4.2)叫做 Klein-Gordon 方程，但遗憾的是，这个方程存在极其严重的问题——它不能保证粒子出现的概率是正的，波函数不能再做概率解释，构造出的概率流是不守恒的。为了解出这个方程还需要知道态矢随时间的一阶导数，也不现实。此外，方程还允许所谓的负能量解，这是人们不想看到的。

狄拉克认为问题的根源是方程必须对时间求一阶导数。于是他试图想办法把(4.1) 等式右侧的内容做变换，假设有一线性算符的平方满足： $(i\hbar c \sum_{i=1}^3 A_i \partial_i + cB)^2 = -\hbar^2 c^2 \nabla^2 + m^2 c^4$ 。展开后比较系数可得：

$$A_i^2 = 1 \quad A_i A_j + A_j A_i = 0 \quad (i \neq j) \quad A_i B + B A_i = 0 \quad B^2 = m^2 c^2$$

显然，没有任何复数可以满足这四条关系。所以，这四个系数必须是矩阵。而且，要满足条件至少需要四阶矩阵。这里泡利矩阵正好可以用来满足这个条件，所以我们构造：

$$A_i = \begin{pmatrix} 0 & \sigma_i \\ \sigma_i & 0 \end{pmatrix} \quad B = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix} mc$$

于是，薛定谔方程变成了：

$$i\hbar \frac{\partial}{\partial t} |\phi\rangle = (i\hbar c \sum_{i=1}^3 A_i \partial_i + cB) |\phi\rangle \quad (4.3)$$

方程(4.3)叫做狄拉克方程。通常，我们借助  $\gamma$  矩阵，把这个方程写成更为协变的形式，以方便体现洛伦茨变换的对称性：

$$(i\hbar \gamma^\mu \partial_\mu + mc) |\phi\rangle = 0 \quad (4.4)$$

这里  $\gamma^0 = B/mc, \gamma^i = \gamma^0 A_i$ . 在自然单位制中, 方程更为简洁:

$$(i\gamma^\mu \partial_\mu + m)|\phi\rangle = 0 \quad (4.5)$$

在狄拉克方程中, 态矢  $|\phi\rangle$  不再是一个函数, 而是四个函数! 这个方程解决了概率不守恒的问题, 但依然有负能量的问题。狄拉克将负能量解释成正电子。狄拉克认为负能级是存在的而且全部填充了电子, 如果某个负能级的电子被光子激发到正的能级, 会在负能海中形成空穴, 而空穴的运动方程对应于负能量解, 它看起来正像一个拥有正电荷的电子。狄拉克方程成功预言了正电子的存在——4 年后的 1932 年美国物理学家卡尔·戴维·安德森在宇宙射线中发现了正电子。

狄拉克方程是 20 世纪上半叶相对论性量子力学最成功的尝试, 但它仍是一个描述单粒子的相对论性量子力学, 而相对论的引入意味着描述单粒子的理论框架不再准确, 这是因为在高能的情况下, 相对论效应将不能保证粒子数不变。正式的相对论性量子理论是量子场论, 它不再描述单粒子的运动, 而通过正则量子化或路径积分描述量子化的场的运动。经过场量子化后, 粒子和场也将用统一的语言进行描述。

### § 4.1 U(1) 规范对称性

一般的, 场论系统的任何对称性都限制着它的作用量, 这种限制通过要求作用量在对称变换之下保持不变来完成。在现代物理中有一条构造作用量的基本原则, 即首先猜测出场论系统的对称性, 然后再写出这些对称性所允许的作用量的一般形式。对称性和守恒定律密切相关。每一种连续对称性都对应着一条守恒定律, 这就是所谓的诺特定理。

其中, 保持时空坐标不变的对称性被称为内部对称性。一种常见的内部对称性是所谓的相位不变性, 也称作 U(1) 不变性。具体来说, 假设我们考察的场论系统由一个复标量场组成, 并具有如下作用量:

$$S[\phi] = - \int d^4x [\partial_\mu \bar{\phi} \partial^\mu \phi + \mathcal{U}|\phi|^2] \quad (4.6)$$

显然, 这个作用量在如下复标量场的相位变换 (也称作 U(1) 变换) 下保持不变:

$$\phi(x) \rightarrow e^{i\theta} \phi(x), \quad \bar{\phi}(x) \rightarrow e^{-i\theta} \bar{\phi}(x)$$

取  $\theta$  为无穷小量  $\theta = \epsilon$ , 则 U(1) 变换前后场的改变量为  $\delta\phi = i\epsilon\phi$ ,  $\delta\bar{\phi} = -i\epsilon\bar{\phi}$ . 现在, 将上面无穷小参数  $\epsilon$  变成依赖于时空坐标的  $\epsilon(x)$ , 则变换前后作用量的改变量为

$$\begin{aligned} \delta S &= - \int d^4x [\partial_\mu \bar{\phi} \partial^\mu (\delta\phi) + \partial_\mu (\delta\bar{\phi}) \partial^\mu \phi] \\ &= -i \int d^4x [\partial^\mu \bar{\phi} \phi - \bar{\phi} \partial^\mu \phi] \partial_\mu \epsilon \end{aligned}$$

相应的守恒流  $J^\mu = i[\bar{\phi} \partial^\mu \phi - \partial^\mu \bar{\phi} \phi]$ . 特别的, 现实世界的电荷守恒就对应于这样一个 U(1) 守恒流。事实上一个具有局域性<sup>1</sup>的理论中天然就具有 U(1) 对称性, 为了看到这一点, 考虑一个平移变换:

$$S[\phi] = \int d^4x \mathcal{L}(x) \rightarrow \int d^4x \mathcal{L}(x + \delta x) = \int d^4(x + \delta x) \mathcal{L}(x + \delta x) = S[\phi]$$

从而, 一个具有局域性的理论中天然就具有 U(1) 对称性。

我们一直在强调对称性, 但是到此为止我们讨论的对称性全都是所谓的整体对称性, 现在我们要开始讨论局域对称性, 也称之为**规范对称性**。对于整体对称性来说, 它在时空整体上的变换都完全一样, 但是, 对于规范对称性来说, 不同时空点的对称变换可以相互独立。为了将一个整体对称性局域化成一个规范对称性, 我们往往需要引入规范场。最简单的规范场就是电磁场, 它所对应的规范对称性就是所谓的 U(1) 规范对称性。

<sup>1</sup>局域性是指拉氏量满足  $\mathcal{L}(x) = \mathcal{L}$ .

现代的观点认为，电磁场是 U(1) 规范场，是我们要求物质场满足定域规范不变性而引入的。为了说明这一点，我们先来考虑物质场是旋量场的情形。对于一个自由旋量场，其拉格朗日量是

$$\mathcal{L} = \bar{\Psi}(i \not{\partial} - m)\Psi \quad (4.7)$$

其中我们使用了费曼符号，也即  $\not{p} := \gamma^\mu p_\mu$ 。容易验证，这个拉格朗日量对应的经典场方程就是 Dirac 方程：

$$\Psi(i \not{\partial} - m) = 0, \quad \bar{\Psi}(i \overleftarrow{\not{\partial}} + m) = 0 \quad (4.8)$$

这个拉格朗日量具有整体的 U(1) 变换，也就是  $\Psi \rightarrow \Psi e^{i\alpha}$  下的不变性。如果要求这一拉氏量具有定域的 U(1) 规范不变性，也就是在以下变换下不变：

$$\Psi(x) \rightarrow \exp[iq\Gamma(x)]\Psi(x), \quad \bar{\Psi}(x) \rightarrow \bar{\Psi}(x) \exp[iq\Gamma(x)] \quad (4.9)$$

由于  $\Gamma(x)$  是  $x$  的函数，对  $\exp[iq\Gamma(x)]\Psi(x)$  求偏导时会多出来一项  $-iq(\partial\Gamma(x)) \exp[iq\Gamma(x)]\Psi(x)$ 。如果我们想要令这个拉氏量在定域 U(1) 规范变换下不变，那么我们可以引入一个额外的项，它在规范变换下的变化正好抵消这个额外项。这相当于引入一个协变导数算子：

$$D_\mu(x) := \partial_\mu - iqA_\mu(x) \quad (4.10)$$

这个  $A_\mu(x)$  就是我们引入的 U(1) 规范场。引入这个协变导数算子后的拉氏量为：

$$\mathcal{L} = \bar{\Psi}(x)(i \not{D} - m)\Psi(x) \quad (4.11)$$

如果我们令  $A_\mu(x)$  的规范变换是  $A_\mu(x) \rightarrow A_\mu(x) - \partial_\mu\Gamma(x)$ 。那么我们发现， $D_\mu$  的后一项在定域 U(1) 规范变换下的变换刚好和求偏导数得出的项  $-iq(\partial\Gamma(x)) \exp[iq\Gamma(x)]\Psi(x)$  抵消。

在添加了规范场以后，我们的理论中就有了两组广义坐标：物质场  $\Psi$  和规范场  $A_\mu$ 。既然我们多了一个场，当然就要在我们的 Lagrangian 中加入自由规范场的拉氏量。于是至此，我们就从要求旋量场满足定域 U(1) 规范不变出发，得到了以下的拉氏量：

$$\mathcal{L} = -\frac{1}{4}F^{\mu\nu}F_{\mu\nu} + \bar{\Psi}(x)(i \not{D} - m)\Psi(x) \quad (4.12)$$

这个拉氏量生成的量子场论就被称为**旋量量子电动力学**。此外，在经典电磁理论中，麦克斯韦方程描述的电磁场同样具有 U(1) 规范对称性。这说明，电磁场起源于 U(1) 规范对称性，电场和磁场是一个整体，它们共同构成二阶反对称张量场  $F_{\mu\nu}$ ，因此电场和磁场在洛伦兹变换下会相互转变。而 U(1) 对称性守恒流中的  $J_A^0 = \rho_A$  正是电荷密度， $\mathbf{J}_A$  则是电流密度，流守恒方程  $\partial_\mu J_A^\mu = 0$  正是电荷守恒方程<sup>2</sup>。

对于真空中的自由电磁场，即电流四矢量  $J_A^\mu$  等于零时的电磁场，满足如下动力学方程：

$$-\partial_\mu \partial^\mu (A^\nu) = 0 \iff (\partial_t^2 - \nabla^2)A^\nu = 0$$

显然，对于  $A^\mu$  场的每一个分量，这个方程描述的都是一种以  $c=1$  的速度传播的波动，也就是电磁波。由于在我们的单位制中光速正是  $c=1$ ，这和电磁波的传播速度一样，因此，麦克斯韦提出**光就是一种电磁波**。麦克斯韦的这一设想是对历史上人们长期迷惑不解的光的本性的一次深刻洞察。而从量子场论的角度上，光子实际上是电磁场的激发态，光子是传递电磁相互作用的基本粒子。

到此为止，我们就成功地将 U(1) 整体对称性局域化成了一个局域 U(1) 规范对称性。代价是引入了一个额外的矢量场  $A_\mu$ ，通常称之为规范场。

<sup>2</sup>需要注明的是，真实世界中产生电磁场的电流四矢量  $J_A^\mu$  通常并不是由某个复标量场产生的，实际上，除了超导情形中由库珀对所形成的场，宏观世界中并没有带电的复标量场。宏观世界中的电荷密度电流密度其实是由运动的带电粒子（而不是场）提供的。



## 对称性自发破缺

在基本规律的层次上，我们的世界也许有很高的对称性，但是我们所生活的低能世界却没有那么高的对称性，它反而展现出千奇百怪的多样性。如何从基本层次的对称性到现实的多样性呢？回答是通过对称性破缺，尤其是通过对称性自发破缺。

自然界中常常发生的是，基本物理规律具有某种对称性，但是它描述物理系统的某个特定解却没有那么高的对称性，这就是对称性自发破缺。特别的，对称性自发破缺尤其指场论系统的能量最低解（即真空解）没有那么高的对称性，尽管作用量的对称性比较高。在对称性自发破缺前后，也就是选择真空态前后，量子作用量都具有自发破缺的对称性，于是自发破缺的对称性始终是量子体系的对称性，并没有真正的破缺。对称性自发破缺的过程所“破缺”的只不过是显式的对称性，而不是对称性本身。

我们先阐述自发对称性破缺一个十分重要的定理——Goldstone 定理。

**【定理 4.1】(Goldstone 定理)** 在一个对称性自发破缺的理论中，对于自发破缺的、连续的量子对称性<sup>3</sup>的每个生成元而言，自发破缺后的理论中都会出现一个无质量、无自旋的粒子。

这些无质量、无自旋的粒子被称为 **Goldstone 玻色子**。这个定理的证明涉及到许多节外生枝的内容，也并非本章的重点，此处按下不表。

需要关注的是 Goldstone 定理的几何含义——正如 Goldstone 定理所预料的，在对称性自发破缺后，的确出现了一个无质量的玻色场  $\pi$ ——在势能曲面上，如果我们选取了一个真空态  $|\Omega\rangle$ ，那么对于任何自发破缺对称性的生成元，它都给出对称群的一个单参子群。 $|\Omega\rangle$  在这个单参子群下的轨道中所有的态都是真空态，这些真空态具有同样的势能。从而在这个轨道对应的方向上，势能曲线是平的，从而这个方向上的激发就是无质量的。

此外，定理中的一个遗留问题是，我们如何判断一个理论是否存在对称性自发破缺呢？一种可行的方法是通过考察理论中的标量算符的真空期待值来判断理论中是否存在对称性自发破缺。我们有如下结论：

**【命题 4.2】** 如果一个在对称群下处于非平凡表示的算符  $\mathcal{O}$  获得了非零的真空期待值，那么该对称性是自发破缺的。

**证明.** 利用反证法来证明这一结论。首先令这个对称群在 Hilbert 空间上的生成元是  $Q$ ，那么如果这一对称性不是自发破缺的，那么对称变换总是保持真空不变的，从而有  $Q|\Omega\rangle = 0$ 。

我们只需要证明：如果对称性不是自发破缺的，那么一个在对称群下处于非平凡表示的算符  $\mathcal{O}$  一定没有真空期待值，就证明了这一结论。

为此，考虑一系列处于表示  $R$  的算符  $\mathcal{O}_n$ ，它们在对称群下的变换关系是

$$\delta\mathcal{O}_n = i\epsilon[Q, \mathcal{O}_n] = i\epsilon T_{nm}\mathcal{O}_m$$

其中  $T_{nm}$  是表示  $R$  中对称群的生成元。如果真空不破缺对称性，那么就有

$$0 = \langle\Omega|[Q, \mathcal{O}_n]|\Omega\rangle = T_{nm}\langle\Omega|\mathcal{O}_m|\Omega\rangle$$

如果  $R$  并不是平凡表示，那么  $T$  就不是 0 矩阵。由于这一关系对任何的  $n$  都成立，于是也就蕴含着  $\langle\Omega|\mathcal{O}_m|\Omega\rangle = 0$ 。从而如果一个处于非平凡表示的算符满足  $\langle\Omega|\mathcal{O}|\Omega\rangle$ ，那么一定有  $Q|\Omega\rangle \neq 0$ ——理论中存在对称性自发破缺。□

因此，如果理论中包含一个处于对称群的非平庸表示，并且具有非零真空期待值的算符，那么理论中一定具有一个质量为 0 的、非真空的粒子态——它就对应 Goldstone 粒子。对于连续对称性的情况，如果一个系统的对称性从  $G$  破缺到  $H$ ，那么生成的 Goldstone 粒子的数量就为  $\dim(G/H)$ 。

根据上面的结论，我们可以将 Goldstone 定理以一种更形式化的方式表述：

**【定理 4.3】(Goldstone 定理)** 在相对论协变的定域场论中，如果  $\partial_\mu J^\mu = 0$  且存在算符  $\mathcal{O}_a$  使得

$$\langle\Omega|[Q, \mathcal{O}_a]|\Omega\rangle \neq 0$$

<sup>3</sup>注意我在这里强调了量子对称性。对于一个反常的对称性而言，如果它被自发破缺了，在 Lagrangian 层面上的确会出现一个没有质量项的场，但是这个场会被量子修正所赋予物理质量，从而它也就不再是 Goldstone 场。

则存在零质量粒子态  $|G\rangle$  (Goldstone 玻色子) 使得

$$\langle G|J^\mu|\Omega\rangle \neq 0, \quad \langle \Omega|\mathcal{O}_a|G\rangle \neq 0$$

其中  $\partial_\mu J^\mu = 0$  是算符满足的**流守恒条件**。这里要尤其提醒一下，流守恒关系不一定总是满足的！比如在讨论规范理论的手征流时，这个关系就是不成立了，此时被称为**手征流反常**。

现在来看一个对称性自发破缺的例子。考虑一个具有  $O(N)$  对称性的例子：

$$\mathcal{L} = \frac{1}{2}\partial^\mu\phi\partial_\mu\phi - \frac{1}{2}\mu^2\phi^2 - \frac{1}{4}\lambda\phi^4$$

势能函数  $V(\phi) = \frac{1}{2}\mu^2\phi^2 + \frac{1}{4}\lambda\phi^4$ ，我们很容易解出来势能的最小点为

$$\langle \Omega|\phi|\Omega\rangle = \sqrt{\frac{-\mu^2}{\lambda}} = v$$

因此我们在  $|\phi_0| = v$  附近得到的是一个稳定真空，也就可以在这个真空附近寻找破缺之后的对称性。为了不失一般性，我们假设

$$\langle \phi \rangle = \phi_0 = (0, \dots, 0, v)^T$$

因此我们发现对于这个真空而言，只能发生局部激发，因此  $O(N)$  对称性就自发破缺了，并且破缺到了  $O(N-1)$ 。为了显示这个破缺，注意到我们做出变换

$$\phi' = \phi_N - v$$

于是拉格朗日密度可以写成

$$\mathcal{L} = \frac{1}{2}\partial^\mu\hat{\phi}\partial_\mu\hat{\phi} + \frac{1}{2}\partial^\mu\phi'\partial_\mu\phi' + \mu^2\phi'^2 - \lambda v\phi'(\phi'^2 + \hat{\phi}^2) - \frac{1}{4}\lambda(\phi'^4 + \hat{\phi}^4) + \mathcal{C}$$

显然这个拉氏量有  $O(N-1)$  对称性，并且生成了  $N-1$  个 Goldstone 玻色子。

### § 4.3 电弱对称性破缺

在构建理论时，对称性自发破缺可以在要求理论具有某个对称性的基础上帮助我们在拉氏量中引入显式上被对称性禁戒的项。这个用法的一个典型例子就是希格斯机制：在电弱理论中，实验表明传递弱作用的 W 玻色子和 Z 玻色子具有质量。但是电弱理论是一个规范理论，W 和 Z 玻色子是规范场的激发，从而它们的质量项会破坏规范对称性。这个问题的解决方案就是在理论中引入规范对称性的自发破缺，从而可以在拉氏量中引入规范场的质量项。

我们从一个复标量场和 U(1) 规范场的耦合开始，如果要求一个标量场  $\mathcal{L} = -\partial_\mu\phi^\dagger\partial^\mu\phi/2 - V(\phi)$  具有局域的 U(1) 规范对称性，那么我们可以引入一个规范场，耦合后的 Lagrangian 是

$$\mathcal{L} = -(D^\mu\phi)^\dagger D_\mu\phi - V(\phi) - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}$$

考虑一个模型：如果  $V(\phi)$  有形式  $V(\phi) = \mu\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2$ ，其中  $\lambda \geq 0$ ，因为如果它为负，那么势能就没有下界，在量子理论中也会变得极不稳定。但我们无法确定  $\mu$  的大小，若  $\mu \geq 0$ ，这个场的基态是  $\phi_0 = 0$ 。

若  $\mu < 0$ ，从经典场论来看，这个场的基态是  $\phi(x) = \sqrt{\frac{-\mu}{\lambda}}e^{i\zeta(x)} \equiv ve^{i\zeta(x)}$ ，于是在量子场论中，我们要求理论有一系列基态  $|\zeta\rangle$ ，这些基态下场的期待值是  $\langle \zeta|\phi|\zeta\rangle = ve^{i\zeta(x)/v}$ 。

我们对场做重新定义，使得有  $\langle \Omega|\phi(x)|\Omega\rangle = 0$ ，从而 LSZ 约化公式<sup>4</sup>成立，我们可以利用微扰量子场论处理这个体系。为此定义： $\phi(x) = (v + \rho(x))e^{i\zeta(x)/v}$ 。其中  $\zeta, \rho$  都是实标量场。

<sup>4</sup>它可以说是量子场论中最重要的公式之一。它告诉我们，如果要求出 S 矩阵（一个多粒子态从初态跃迁到末态的概率幅矩阵元），就只需要求出编时关联函数即可。

直观上说可以定义一个新的场  $h = \phi - \phi(x)$  然后在  $h = 0$  处展开。因为势能在  $\phi = 0$  处是超光速的，按此处微扰展开会出现不收敛的情况。对于规范场论而言，我们可以做一个局域的规范变换<sup>5</sup>使得考虑的基态  $|\xi\rangle$  下场的期待值变为  $v$ 。也就是说，我们可以通过确定规范来将场  $\xi$  设置为 0。

那么 Lagrangian 可以写为：

$$-\frac{1}{2}\partial_\mu\rho\partial^\mu\rho - e^2(v+\rho)^2A^\mu A_\mu - \lambda v^2\rho^2 - \mu v\rho - \frac{1}{4}F_{\mu\nu}F^{\mu\nu}$$

可以看到，对于场  $\rho$  而言，体系不再具有在  $\rho \rightarrow e^{ie\Gamma}\rho$  下的对称性，也就是说 U(1) 整体规范对称性自发破缺了。并且我们看到， $\xi$  是一个无质量的场，它的激发就是 Goldstone 玻色子。在对称性出现自发破缺的时候，规范场会自动获得一个质量项。我们考虑一个 U(1) 规范理论的动能项：

$$\begin{aligned} -(D^\mu\phi)^\dagger D_\mu\phi &= -|\partial_\mu h - ieA_\mu\phi|^2 \\ &= \dots + e^2\phi^2A^2 \\ &= \dots + e^2(h + \phi(x))^2A^2 \\ &= \dots + e^2v^2A^2 \end{aligned}$$

通过对比确认规范场的质量  $m = ev$ 。U(1) 对称性自发破缺的情况在标准模型里并没有出现，但是它在凝聚态物理中，尤其是超导中起到了关键性的作用。

规范场在对称性自发破缺后获得质量这件事可以用来解释超导体中的 Meissner 效应。在理论上，超导体中的库珀对可以用一个标量场  $\phi$  来有效地描述，而如果我们再在理论中加入电磁场，那么这些库珀对的动力学就可以用上面讨论的这个 Lagrangian 所描述。倘若由于标量势的形状，这些库珀对获得了一个真空期望值  $\langle\phi\rangle$ ，那么根据上面的讨论，由于库珀对带有电荷  $2e$ ，那么电磁场将获得一个质量  $m_\gamma = 2ev$ 。从而在这个真空附近，恒定电磁场满足的 Maxwell 方程将从  $\nabla^2\mathbf{B} = 0$  变为以下的 Proca 方程：

$$\nabla^2\mathbf{B} = m_\gamma^2\mathbf{B}$$

于是如果超导体外有一个沿着  $z$  方向的恒定磁场  $\mathbf{B} = \mathbf{B}_0$ ，那么在超导体内部的磁场就是：

$$\mathbf{B}(z) = \mathbf{B}_0 e^{-m_\gamma z}$$

也就是说，超导体内部的电磁场是指数衰减的，磁通量无法穿透进超导体的深处，这便是 Meissner 效应。而在超导体的表面，磁通量可以有效地穿透  $L = 1/m_\gamma$  的距离，这个距离就被称为穿透深度。

通俗地说，(一部分)微观超导机理本质上是 U(1) 对称性自发破缺产生了质量项，即一个具有质量的 U(1) 规范场会导致外电磁场无法穿透，但是电子却可以正常穿越，于是我们有了不需要外场驱动电流。

而在标准模型中，自发破缺的是  $U(1) \times SU(2) \equiv U(2)$ 。这个时候基态可以写为  $v$  乘上一个 2 维复空间的一个矢量。不丧失一般性地，我们取

$$\phi_0 = \frac{v}{\sqrt{2}} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

此时的协变导数是  $D_\mu\phi = \partial_\mu\phi - i[g_2A_\mu^aT^a + g_1B_\mu Y]\phi$ 。其中  $T^a$  是泡利矩阵 (乘上常数  $\frac{1}{2}$ )， $Y$  是单位阵 (乘上常数  $-\frac{1}{2}$ )。从群论的角度看，U(2) 的 4 个生成元分别是单位阵和 3 个泡利矩阵，我们发现对于  $\phi_0$  这个矢量而言， $(I - \sigma^3)\phi_0 = 0$ 。

也就是说，基态  $\phi_0$  在绕轴  $I - \sigma^3$  的转动下是不变的。因此我们发现这个 U(2) 群里，有一个 U(1) 的子群并没有出现对称性自发破缺。因此，对应的规范场也就不会获得质量。这一机制对所有非阿贝尔规范理论都适用，在某些方向上基态仍具有对称性，从而对应的规范场依然是没有质量的。

因此，在电弱规范理论中，一定存在一个质量为 0，不带电的矢量粒子，这个粒子就被认为是光子。

<sup>5</sup>这个规范被称为么正规范。

## 科学实在论导引

**科学实在论**是对我们最好的理论和模型的内容的一种积极的认识论态度，建议同时相信科学所描述的世界的可观察和不可观察的两个方面。这种认识论态度具有重要的形而上学和语义学维度。关于科学实在论的辩论几乎与科学哲学中的其他一切密切相关，因为它们涉及科学知识的本质。这些不同的承诺受到了一些对立的科学认识论<sup>1</sup>的质疑。

反实在论的许多动机都来源于如下信念：我们实际上不能获得关于实在的不可观察部分的知识——这种知识超出了人类的知识视域。按照这一观点，科学知识的界限是由我们的观察能力设定的。所以，科学能够给予我们关于化石、树木、冰糖的知识，但是不能给予我们关于原子、电子和夸克的知识——后者都是不可观察的。这一观点并非完全不合情理。没有人会真正怀疑化石和树木的存在，但是却可能去怀疑原子和电子的存在。事实上，19世纪后期的许多重要科学家都怀疑原子的存在。如果科学知识限于可被观察的范围内，那么所有接受这一观点的人，显然都必须解释为什么科学家会提出关于不可观察的实体的理论。反实在论者给出的解释是，这些理论都是顺手虚构的，是为了预测事物在可观察世界中的表现而提出的。

实在论者不认为科学知识受限於我们的观察能力。相反，他们认为我们已经切实了解不可观察的实在，因为存在着种种理由认为我们最好的科学理论都是正确的，而最好的科学理论都在探讨不可观察的实体。例如，关于物质的原子论，这一理论声称所有的物质都是由原子构成的。原子论能够解释有关世界的大量事实。按照实在论者的说法，这是表明这一理论正确的很好的证据，也就是说，物质的确是由如这一理论所描述的以某种方式表现的原子构成的。尽管存在着支持该理论的明显证据，该理论当然仍可能是错误的，但所有的理论都可能是这样。仅仅因为原子不可观察，还没有理由不将原子论解释成一种对实在的尝试性描述——多半是一种成功的描述。

在许多特殊的条件和立场的变体之下，科学实在论有一个共同的核心思想，典型的是对科学研究的结果持一种认识论上的积极态度，涉及世界上可见和不可观察的方面。在这里，可观察和不可观察的区别反映了人类的感官能力：可观察的是在有利条件下，可以用无辅助感官感知的东西（如行星和鸭嘴兽）；不可观察的是不能用这种方式检测的东西（如蛋白质和质子）。这仅仅是为了术语上的方便而赋予视觉以特权，与科学上的可观察性概念不同，后者一般扩展到可使用仪器检测的事物。

大多数人是从小科学理论或理论的某些方面的真理或近似真理的角度来定义科学实在论的。有些人则从理论术语成功地论及到世界上的事物（包括可观察的和不可观察的事物）的角度来定义它。另一些人则不是从真理或参照物的角度来定义科学实在论，而是从对科学理论本体论的信念来定义。所有这些方法的共同点是致力于我们最好的理论具有一定的认识论地位：它们产生了关于世界各个方面的知识，包括不可观察的方面。

在这里还需要做一个一般性的免责声明：实在论者一般都是可错主义者，认为实在论对于我们最好的理论是合适的，即使这些理论很可能无法得到绝对肯定的证明；我们的一些最好的理论可能会被想象成是重大错误的，但实在论者认为，即便考虑到这种可能性，实在论者还是有理由的。将适合实在论者承诺的理论领域限制在那些足够成熟和非临时性的理论上，就可以支持这些理由。

---

<sup>1</sup>它们在形式上被统称为科学反实在论。



## § 5.1 无奇迹论证

在气体动力学理论中，通常认为任何体积的气体都包含了大量处于运动中的微小实体。这些实体（分子）是不可观察的。通过动力学理论，我们能够推知关于气体可观察行为的各种结论。例如，如果气压保持不变，加热气体就会导致气体膨胀，这一点可以通过实验证实。

许多类似的假设不可观察实体存在的理论在经验上是成功的——它们对可观察世界中物体的表现作出了准确的预测。上面提到的气体动力学理论是一个例子，还有许多这样的例子。此外，这类理论通常还具有重要的技术用途。例如，激光技术所依据理论的基础，就与一个原子中的电子从高能状态转变为低能状态过程中发生的情况相关。激光技术不断发挥作用——使我们能够矫正视力、用导弹袭击敌人，以及做其他很多事。因此奠定激光技术之基础的理论在经验上是高度成功的。



图 5.1: No Miracle Argument

假设不可观察实体存在的那些理论在经验上的成功，是支持科学实在论的一种最有力的论点，被称为“无奇迹”论证<sup>2</sup>。按照这种说法，如果一个探讨电子和原子的理论精确预测了可观察的世界，那么，除非实际上电子和原子真的存在，否则这一理论就将是一种罕见的巧合。如果不存在原子和电子，怎么解释理论与观察数据的紧密相符？类似地，除非假设相关理论是正确的，我们又该怎样去解释理论推动了技术的进步？如果正如反实在论者所言，原子和电子仅仅是“顺手虚构的”，为什么激光能有效地发挥作用？

按照这一观点，成为一个反实在论者就等于相信奇迹。既然在存在非奇迹的替代情形时显然最好不要相信奇迹，我们就应该成为实在论者而不是反实在论者。需要格外注意的一点是，NMA 并不是想证明实在论是正确的而反实在论是错误的。确切地说，它是一个合理性论点——一种**最佳解释推论**。待解释的现象是这样一种事实，即许多假设不可观察实体存在的理论在经验上高度成功。“无奇迹”论证的倡导者们声称，对于这一事实的最佳解释就是这些理论是正确的——相关的实体的确存在，正如理论所说的那样以某种方式表现出来。除非接受这种解释，否则我们的理论在经验上的成功就仍是一个无法解释的谜团。

最佳解释推论是一种所谓的溯因推理方法，它明确地诉诸于科学理论的经验成功这一现实来为科学实在论辩护。因为如果存在这样的东西（如电子、DNA、弯曲的时空，等等），那么对这些理论的成功的一个自然的解释就是理论部分真实地说明了这些东西是怎样运行的。并且，对科学理论相互超越途径的一个自然的说明——比如说，爱因斯坦相对论超越牛顿万有引力——是对一个理论客体部分正确或不正确的说明，如引力场或时空结构、或二者被一个对同一客体更好的说明所取代。但如果这些物体根本就不存在，那么这将是一个奇迹：一个理论描述了远处的重力行为并成功预测了现象。总之，在面对当代科学不断取得成功的经验事实时，科学实在论的解释是最具说服力的，否则，难道我们真的要把科学的成功简单归于一连串的巧合吗？这是 NMA 向我们提出的最现实、最重要的疑问，在这方面实在论的解释还是最符合我们直观感觉的。

<sup>2</sup>我们将其简称为 NMA。

§ 5.2  
悲观元归纳

面对 NMA 这一看似合理的“最佳解释推论”，反实在论求助于历史案例分析，将批判矛头指向“非奇迹论证”的关键环节——“真理是导致理论成功的真正原因”，它直接从科学历史事实出发提出了 NMA 的一系列反例。反实在论试图来一个“事实胜于雄辩”：实在论论证并不符合科学史事实，所以，其论证是没有说服力的。

NMA 为科学实在论的主张提供了最有力的辩护，但这种论证本身还并不完善。NMA 在逻辑上存在的最大问题在于它缺乏严格性的逻辑依据，其解释不具有必然性，这也是导致反实在论批判的关键环节。NMA 仅仅将对科学实在论的论证建立在直觉之上是很难令反实在论者满意的。

所谓“悲观元归纳”<sup>3</sup>是作为经验前提的论证。如果考虑到任何一个学科的科学理论的历史，通常会发现随着科学知识的发展，旧的理论会有规律地更替，而新的理论则会出现。从现在的角度看，过去的大多数理论一定被认为是错误的；事实上，从大多数时代的角度看也会如此。因此，通过列举式的归纳（即从这些情况中归纳），任何一个时期的理论最终都会被取代，从未来的某个角度看也会被认为是假的。因此，当前的理论也是错误的。

## The Pessimistic Induction

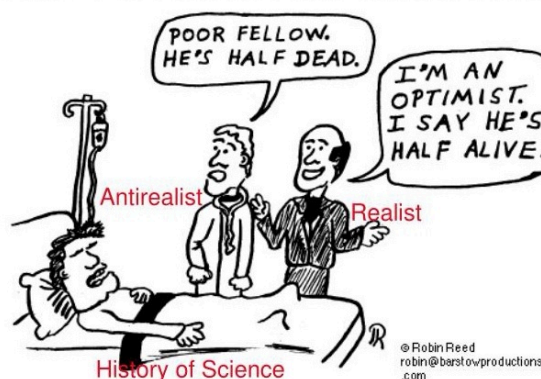


图 5.2: The Pessimistic Induction

在历史上有许多理论，我们现在认为那些理论是错误的而在当时它们在经验上却相当成功。科学哲学家劳丹从不同科学门类和科学时期中抽取、列举了不下 30 种这类理论。燃烧的燃素说就是一个例子。这一理论直到 18 世纪末还广泛流行，它主张任何物体燃烧时都会释放一种叫做“燃素”的物质进入空气中。现代化学告诉我们这一理论是错误的：不存在类似燃素的物质。相反，燃烧发生于物体在空气中与氧气发生化学反应时。尽管燃素是不存在的，燃素说在经验上却相当成功：它非常恰当地符合当时能获取的观察数据。这里也就可以看出，NMA 的溯因推理方法是缺乏严密逻辑依据的，其解释也不是必然为真的，即科学的“成功”与理论“真理性”之间不存在必然性的因果逻辑关系。

而且，如果我们把科学实在论看作一个经验假说，实在论者认为通过最佳说明推理得出科学实在论是对当前科学成功状况的最佳说明。但科学实在论并没有像成熟的自然科学理论一样做出什么成功的科学预言或应用，至多只是解释了一些科学取得成功的现象而已，这不符合科学实在论辩护中“科学成功”的条件。

一些人认为，NMA 实际上是一种被称为基本比率谬误的谬误推理实例。考虑下面的例证。有一种疾病的测试，其假阴性率（疾病存在的情况下的阴性结果）为零，假阳性率（疾病不存在的情况下的阳性结果）为十分之一（也就是说，没有疾病的个体有 10% 的时间测试阳性）。如果一个人的检测结果为阳性，那么他得病的几率有多大？如果根据假阳性率得出的结论是 90% 的概率，那就错了，因为实际的概率还取决于另外一些关键的信息：疾病在人群中的基本比率（患病人群的比例）。总体而言，疾病的发病率越低，阳性结果表明疾病存在的概率就越低。以此类推，将一个科学理论的成功作为其近似真理的指标（假设假阳性率很低——在这种情况下，与真理相去甚远的理论仍然是成功的），同样可以说是基本比率谬误的一个例子。一个理论的成功本身并不表明它很可能是近似真的，而且由于没有独立的方法可以知道近似真的理论的基本比率，所以无法评估它近似真的机会。

<sup>3</sup>因为它涉及到产生科学理论和法则陈述的“底层”归纳推论，我们将其简称为 PMI。

一些实在论者通过稍微修正 NMA 来加以回应。按照改进后的说法，一种理论的经验性成功证明的是：理论就不可观察的世界所述的内容是近似为真，而不是准确地为真。这一弱化的陈述更容易抵御来自科学史上反例的攻击。同时它也更为温和：它允许实在论者承认今天的理论在某个细枝末节上可能不正确，但仍坚称今天的理论大致是正确的。另一种修正 NMA 的方法是改进经验成功这一概念。一些实在论者认为，经验成功不仅仅是与已知的观察数据相符的问题，它还能使我们去预测新的尚属未知的观察数据。与经验成功的这一更严格的标准相比较，找出在经验上成功、后来却被证实为错误的史例就更不容易了。

尚不确定这些改进能否真的挽救 NMA。它们固然减少了历史上的反例数量，却没有完全消除。仍旧存在的一个反例是 1690 年首次由惠更斯提出的光的波动理论。按照这一理论，光是由“以太”这种不可见的介质中的波状振动构成的，而以太被认为充满整个宇宙。直到法国物理学家菲涅耳在 1815 年用公式表示了光的波动理论的数学形式，并将其用以预测一些令人惊奇的新光学现象，这一理论才被广为接受。光学实验证实了菲涅耳的预测，并且使许多 19 世纪的科学家相信光的波动理论是正确的。然而，现代物理学又告诉我们这一理论并不正确：并不存在类似以太这样的物质，所以光并不是由以太中的振动组成。我们再次碰到一个错误的但在经验上成功的理论。

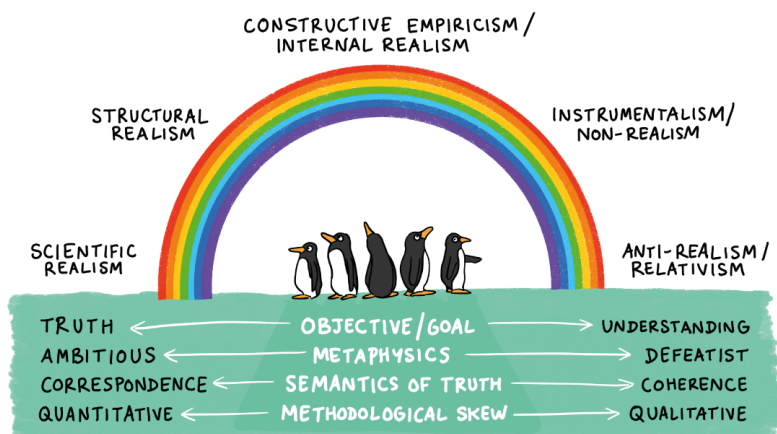


图 5.3: Scientific Philosophy

这一例子的重要特征是，即使是改进后的 NMA，同样可以被推翻。菲涅耳的理论的确作出了新颖的预测，所以即使就经验成功的更为严格的标准来说，它也有资格被认为是经验上成功的。而且很难理解，既然菲涅耳的理论建基于并不存在的以太概念之上，它如何能被称做“近似为真”。无论声称一个理论近似为真的确切所指是什么，一个必然条件一定是这一理论所谈论的实体的确存在著。简言之，即使按照一种严格的对经验成功概念的理解，菲涅耳的理论在经验上也是成功的，但却不是近似为真的。反实在论者说，这一例子的寓意是我们不应该仅仅因为现代科学理论在经验上如此成功就假设它们大致正确。

PMI 并没有直接否认当前成功的科学理论（可能碰巧）是真实的。为了达到否定“当前成功的科学理论是趋近真理的”这个主张，PMI 比较了一系列过去历史上存在的理论和当前的理论主张。通过比较指出，“如果当前成功的科学理论是真实的，那么过去的理论就不可能也同样是真实的”。历史上的科学理论不能被视为真实的，是因为“它们所假设的实体不再被认为存在，或者是因为他们假定的规律和科学机制不是我们当前对世界的理论描述”。但这些具有错误特征的旧理论，在历史上却曾经是极为成功的。因此，PMI 指出，科学在经验方面的成功与其真实性并不相关，也就是说理论的真实性不能因果性地解释其成功性，实在论的 NMA 是失败的。这种思路正如劳丹所表述的那样：“因为（许多过去的理论）建立在我们现在视为基本错误的理论模型和结构之上，实在论不能够指望用真理性来解释这些理论在经验上取得的成功。”劳丹指出，“这些具有错误特征的理论，然而，在经验上是成功的”，因此，科学在经验上的成功不能保证理论的真实性，即真实性不能因果性地解释科学所取得的成功。



### §5.3 三个维度

将科学实在论描述为对理论的一种积极的认识论态度，包括假定涉及不可观察的部分，是一种更精确的承诺的简略表达。传统上，实在论更普遍地与任何赞同相信某物的真实性的立场联系在一起。因此，一个人可能会对自己对桌椅的感知是一个实在论者（感觉材料实在论），或者对桌椅本身是一个实在论者（外部世界实在论），或者对数学实体如数字和集合是一个实在论者（数学实在论），等等。科学实在论是关于我们最好的科学理论所描述的任何事物的实在论。——从这一点来说，这里的“实在论”表示科学实在论。但是，更确切地说，那是什么呢？为了明确科学范畴内的实在论相当于什么，并将其与一些重要的反实在论替代方案区分开来，从三个维度来理解它是有用的：形而上学（或本体论）维度、语义维度和认识论维度。

**形而上学维度：**在形而上学维度上，实在论主张科学所研究的独立于心灵的世界的存在。这一观点最好与否认这一观点的立场进行对比说明。例如，任何属于传统的“唯心主义”标题下的立场，包括某些形式的现象学，都否认这种观点，根据这种观点，不存在外在于心灵的世界，因而不存在独立于心灵的世界。不过，这种唯心主义虽然在历史上很重要，但在当代科学哲学中却很少遇到。更常见的对独立于心灵这一说法的拒绝来自新康德主义对科学知识本质的看法，这些看法否认我们经验的世界是心灵独立的，即使（在某些情况下）这些立场接受世界本身并不依赖于心灵的存在。这里的论点是，科学所研究的世界——不同于“世界本身”（假设这是一个连贯的区别）——在某种意义上取决于一个人带入科学研究的观念，这可能包括，例如，理论假设和感知训练。在这方面必须注意到，科学分类学中的人类约定俗成与独立于心灵是相容的。例如，尽管 Psillos 将实在论与世界的“独立于心灵的自然类结构”联系在一起，但 Chakravartty 认为，独立于心灵的属性往往被约定俗成地归为不同的种类。

**语义学维度：**在语义学维度上，实在论致力于对世界的科学主张进行字面解释。通俗地说，实在论以“表面价值”来看待理论陈述。根据实在论，关于科学对象、事件、过程、属性和关系（我以后将用“科学实体”作为这类事物的总称）的说法，无论它们是可观察的还是不可观察的，都应按字面解释为具有真值域，无论真假。这种语义上的承诺主要与某些“工具主义科学认识论”的承诺形成了鲜明的对比，后者把对不可观测物的描述简单地解释为预测可观测现象的工具，或者把观测报告系统化。

**认识论维度：**在认识论维度上，实在论致力于这样一种观点，即理论主张（从字面上解释为描述一个独立于心灵的现实）构成对世界的认识。这与怀疑论的立场形成了鲜明的对比，后者即使认可实在论的形而上学和语义学维度，也怀疑科学研究在认识论上足以产生这样的知识，或者像一些反实在论的立场一样，坚持认为科学研究只足以产生关于可观察物的知识。实在论的认识论维度虽然为实在论者普遍认同，但有时也会以相反的方式进行更具体的描述。例如，虽然许多实在论者赞同用真理符合论的某个版本来理解理论的真理（或近似真理），但有些人更喜欢使真者的说法或真之紧缩论的说法。尽管大多数实在论者将其立场与理论术语的成功指涉（包括那些不可观察的实体的术语）相结合。但一些人否认这是一个要求。然而，在这些分歧中，实在论的一般秘诀是广泛共享的：我们最好的科学理论对一个独立于心灵的世界的可观察和不可观察的方面给出了真实或近似真实的描述。

刚才描述的实在论的一般配方就其内容而言是准确的，但仍未达到大多数实在论者所提供的精确程度。到目前为止，不精确的两个主要来源是在一般配方本身，它提到了“我们最好的科学理论”和“近似真理”的概念。这些限定的动机也许是明确的。如果人们要对科学理论采取积极的认识论态度进行辩护，那么，这样做大概不单纯是针对任何理论（特别是当我们考虑到，在迄今为止的漫长的科学历史中，有些理论并不成功或不是特别成功的时候），而是针对初步看来值得这样辩护的理论（或理论的某些方面，我们稍后将看到），即我们最好的理论（或其某些方面），这样做是明智的。而且人们普遍认为，至少是实在论者认为，严格地说，即使我们的许多最好的科学理论可能是假的，也可能“接近”真理（即近似于真）这一概念的重要性。然而，如何使这些限定条件更加精确是一个重大的挑战，并引起了许多讨论。

对上述实在论者的一般配方的一个特别重要的限定是是以若干变体的形式出现的。其中最具有影响力的是结构实在论——它试图更具体地确定科学理论中最值得认识论承诺的组成部分。



作为最大限度地提高实在论的可信度的一般战略，特别是在科学的不可观测因素方面。这一策略的采用部分是为了使实在论与广泛接受的观点相一致，即严格地说，即使是我们最好的理论，也有大部分是错误的。如果，尽管如此，这些理论中还是有一些方面是真实的（或接近真实），而且人们能够识别这些方面，那么人们就可以合理地将自己的实在论投向对这些理论中最值得认识论承诺的方面采取认识论上的积极态度。

结构实在论是一种提倡选择性的观点，但在这种情况下，怀疑的是不可观察的实体的性质，实在论承诺的是不可观察领域的结构——我们最好的理论所描述的某些关系所代表的那些结构。这一立场的众多版本都属于两个阵营中的一个：第一个阵营强调结构概念和自然概念之间的认识论区别；第二个阵营强调本体论的论点。认识论观点认为，我们最好的理论很可能不能正确地描述不可观察的实体的性质，但却成功地描述了它们之间的某些关系。本体论观点认为，实在论者之所以应该只向往结构的知识，是因为传统的站在关系中的实体概念在形而上学上是有问题的——事实上，并不存在这样的东西，或者说，如果有这样的东西，它们在某种意义上是由它们的关系产生的，或者说是依赖于它们的关系。

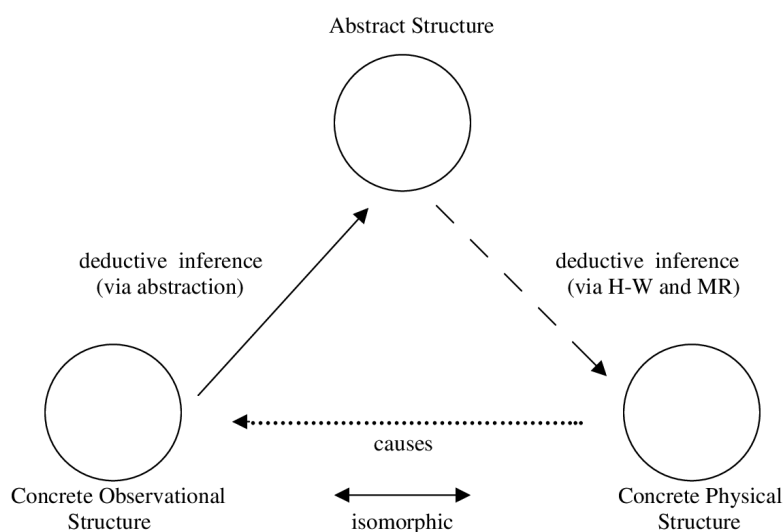


图 5.4: Structural Realism

根据科学理论的繁荣与衰亡的周期报告，科学展示出一个理论接着一个理论走向毁灭的图景。但是不是这些理论的建构是徒然无益的呢？没有那个理论像菲涅尔理论那样曾经建立于巩固的以太运动的基础上，而现在却为麦克斯韦理论代替了，是不是菲涅尔的工作是徒劳无益呢？

正如庞加莱所说：“不是的，菲涅尔的目标并不是想要知道，以太是否真实存在着，如果它存在，是否由某种原子组成，而这些原子又以什么方式运动的。他的目标是要预言光学现象。菲涅尔理论即使在今天也能做到这一点，而事实上它在麦克斯韦以前就已经能做到了，但微分方程总是正确的，它们总是能用同样的方法积分出来，而积分的结果总是保持它们的数值……这并不是说将物理学还原为一种实用的处方。这些（从菲涅尔转到麦克斯韦的）方程表达了一种关系，而如果方程是真的，那是因为这些关系保持它的真实性，它告诉我们，在这样那样的事物中存在着这样那样的真实的关系。只不过是这样那样的事物原来叫作（以太的）运动，而现在叫作电流。它们是我们想出来替代真实客体的名词，而这些真实客体永远隐蔽于我们视野之外，我们唯一能达到的实在就只是这些真实客体之间的关系。”

菲涅尔方程组完整无缺地保留于麦克斯韦的电磁场理论中，并且从麦克斯韦方程中可以将菲涅尔方程推出，新的观点对于旧的论证来说只是重新诠释了它的变量。菲涅尔将振动强度说成是以太粒子之间的振动最大距离，而麦克斯韦说这是电磁波的振幅的平方。从麦克斯韦观点看，菲涅尔对实体解释完全错了，根本没有什么以太粒子这东西，至于麦克斯韦的经典电磁场，后来又为量子力学的光子所代替。尽管科学对实体的认识像走马灯式的不停地转换，但他们的方程式及其表现出来的结构则是正确的、真的。它在理论的转折中和革命中持续下来，坚持下去，表现出科学发展的连续性，和科学本体的某种实在性，即结构实在性。

结构实在论提供了一种“综合”，将主要的先前的实在论论证，即无奇迹论证与反实在论的主要论证，即“悲观元归纳”论证综合起来。

# Chapter 6

## 光子与结构实在论

无论是从实在论的视角看，还是从假设-演绎的方法论看，理论科学中潜在的基本实体的重要性是显然的，并且是可以理解的；在大多数情形中，这些潜在的基本实体是假设性的、不可观察的，甚或本质上是猜测性的，诸如在电弱规范理论中的光子和 Z 玻色子。这些潜在实体的重要性已经深深地扎根于人类渴求说明的夙愿中。

当科学实在论与反实在论的论争相持不下的时候，John Worrall 于 1989 年以他的一篇论文“结构实在论：两个世界的最优选择？”打破了僵局。一方面他不承认标准的科学实在论这样的论点，即认为成熟的科学或经验上实践上成功的理论都能正确地描述了不可观察的实体，他根本就怀疑这一点。他径直地放弃了描述和理解不可观察的实体的企图。另一方面他极力反对反实在论关于理论变革时，从一个旧理论转向一个新理论时没有任何持续性、没有任何东西继承下来的主张，他认为从一个旧理论向一个新理论的转变不仅有成功的经验内容从旧理论转移到新理论，而且有理论的形式从旧理论转移到新理论，这就是物理世界的形式的和数学的结构。不同理论之间，特别是它的不可观察的实体可以大不相同，但却可以具有共同的结构，因此科学实在论没有被推翻，它只是改变了形式，成为一种结构实在论。世界和事物的结构是实在的，是我们认识的；我们的认识，在认识结构上是可以有持续、积累和进步的。

根据这种立场，一个科学理论的经验内容在于所研究领域的理论和现象在结构层面的全域对应，这些经验内容是靠数学结构而兑现的，丝毫不涉及现象的性质，无论是就它们的内禀性质而言，还是就潜在的不可观察实体而言。

库恩曾对科学实在论提出挑战：在理论更替中**本体论的不连续性**显然已经推翻了实在论的主张：科学给予了我们世界上存在什么和发生什么的真知识。当这些结构实在论者对库恩的科学实在论挑战提出异议时，他们通常停留在结构层面，并求助于数学结构的连续性，而没有在本体论层面和解释层面去直面库恩的论证。对于沃勒尔来说，世界的终极性质和内容是不可知的；对于弗伦奇和莱德曼来说，除了结构，再没有什么是世界的终极性质和内容。也就是说，对于他们来说，不存在本体论激进变换的科学革命。因此，他们与库恩学派的对抗，通过拒斥科学史上真正存在革命，而退化成为一种伪对抗。

在我看来，采取这样一种回避问题的策略于事无补，为实在论作有效辩护的唯一途径是在本体论层面直面库恩的论证，对表观的本体论不连续性给出一个可接受的说明，同时为科学发展中更为深层的连续性意义作辩护。下面采用的这样一个推理线路大大有赖于本体论和本体论综合的概念。然而，这些概念根植于潜在实体的结构进路上，因此，在以基本实体观念为基础的科学的还原分析方法和结构主义的整体论立场之间的方法论张力将会减弱。

在这一章，我将对(与光子有关的)规范理论的建构给出一个建构论版本的结构实在论说明，表明规范理论的建构是本体论综合的结果。在这之前首先要阐明科学理论中基本实体的地位和作用，并对本体论范畴(实体、性质、关系、过程等等)之间的关系给予一些一般性的讨论，结论是赞成把某些个性化的因素构成的基本实体看作是基本物理理论的本体论的候选范畴，诸如电弱统一理论，还要详尽阐述在理论科学中基本实体的结构进路。

按照结构实在论的认识论版本，只有结构关系是可知的，潜在的实体是永远隐藏着的；而结构实在论的本体论版本则否定物理实体的真正存在，理由是实体溶解进了数学结构中，至少在量子场论中是如此；不同于结构实在论的认识论版本和本体论版本，结构实在论的建构论版本基于如下的形而上学理解：为高阶抽象结构提供基本的具体结构是科学的首要兴趣，这些具体结构和它们的要素在本体论上不是可分的，享有同等的本体论地位；以此为基础，结构实在论的建构版本完善了传统实在论，把理论实体看作是对存在于世界中的真实实体的指称，并且这些真实实体通过我们对于世界的结构知识是可以认知的。

但是，建构结构实在论与传统实在论有一个重要的区别。传统实在论者通常把物理世界设想为由固定的自然类组成，但是，建构结构实在论者不这样认为。建构结构实在论者以一种结构化的方式来构想物理实体的个性（或本性），这为把科学革命期间激进的本体论转变包容进来打开了一个广阔的概念空间，同时保留了我们对于世界上存在什么和发生什么的一种实在论意义上的认知连续性。

在理论科学史中，一个不可否认的事实是，基本实体的概念在理论科学的概念结构中占据着一个中心位置。科学理论的本体论承诺规定这些基本实体的概念。虽然本体论概念通常指世界上存在什么，但是一个本体论承诺是承诺存在某些特殊实体（或其他的本体论范畴），这些特殊实体在所研究领域的实在的逻辑建构中充当不可化约的概念要素。不同于现象，也与仅仅是理论中启发式的和约定的设置相反，基本实体被假定表示深层实在，因此享有巨大的说明能力。也就是说，能从基本实体中把所有现象作为它的行为结果推导出来。尽管一个基本实体的存在可能与其他基本存在相互关联，诸如光子和电子相互关联<sup>1</sup>，但光子的存在不能从理论论述中的其他任何东西推导出来。

本体论承诺的概念对于我们理解科学革命也是至关重要的。一场科学革命（或者说科学中的一场概念革命）是以科学家在看待相同的现象领域时视角上发生了变革为特征的，这种变革总是伴随着一个本体论承诺的改变，甚至是源于本体论承诺的改变。例如，在看待物理相互作用上，从力学的视角转变为场论的视角，这一转变是伴随着或说是源于从力作为一个独立实体的承诺（如在牛顿力学的场合）转变为力只是作为基本过程的一个附带现象，场才是传递的基本实体的承诺。如果我们采用几何学的观点，如惠勒所假定的观点，把力看作只是时空几何学的表现，那么这个观点传达了看待物理相互作用的另一种概念革命的信号。当然，相互作用由相互作用的实体通过激发普遍的真空信号背景，被概念化为量子的发射和吸收过程这样一种观点，仍然是另一种概念革命。在电磁相互作用的场合，从经典视角到量子电动力学视角的转变，源于一个不带电，零质量和自旋 1 的非局域实体的本体论承诺取代了一个某些方面上看同样基本的电磁波的本体论承诺。

本体论观念也有助于将科学进步概念化。人们可能会论证说，如果我们接受概念革命是本体论承诺的改变，进而认为革命前后的理论是不可通约的，那么我们如何能够主张在革命前后科学中的连续性和进步？库恩声称，虽然在科学（发展过程）中具有不连续性，但是仍然能够承认科学（发展过程）中有进步，尽管只是在我们解决疑难问题的能力提高这种工具意义上，而不是关于世界上存在什么的客观知识的累积的意义上。在否定了第二种意义上的科学进步后，库恩坚定地声称，“我不能看到任何前后一贯的本体论发展方向”。但是，正如我们将在第 6.2 节中阐明的，借助于本体论综合的概念，能在库恩所否定的意义上主张科学（发展过程）中有进步。

用本体论承诺来认定应该把什么样的假想实体看作是基本的和应该研究的，这也为说明所研究领域中的现象提供了资源，规定了理论本身的概念结构，并为进一步的理论和实验研究提供了激励和指导原则，从而迫使相关发展进入一个特定的方向。结果，理论的演化和分支构成一个显著首尾一贯的研究纲领，而其内核就是本体论承诺。

在量子场论中，为了确定起伏真空及其局域激发是需要研究的基本实体，要求进一步研究庞加莱群的么正表象、虚量子的含义，以及为避免真空局域激发不一致的后果所做的重整化。在一种非常强的意义上，物理学家指定的量子场论的理论结构由一个强烈起伏的真空的基本本体论承诺所规定，它只受制于守恒律和测不准原理。各种企图重整化物理相互作用（电磁相互作用除外）的模型，都描绘了一个清晰定义的研究纲领，其硬核是承诺了一个局域起伏的真空。

在电弱统一理论中，情况更为清楚。关于散射振幅和希格斯机制的研究直接由光子和 W、Z 玻色子存在的本体论承诺所规定；量子光学、反常磁矩和中微子物理都源起于光子和 W、Z 玻色子的基本假说，因此属于一个独特的研究纲领。

<sup>1</sup>例如光电效应和量子跃迁。



结构及其要素具有本体论上的不可分离性，这一重要的形而上学假设有一个好处：给出了发现一个深层结构（“深”意指有统一力）将有助于发现嵌入在深层结构中的深层实体这一论断的某种辩护。人们可以论证这种好处，声称如果没有在量子色动力学<sup>2</sup>中发现 SU(3) 规范对称，物理学家将不可能发现夸克和胶子<sup>3</sup>。但问题是如何阐明这个形而上学假设发挥作用的方式，或者说，如何阐明我们从结构转移到基本实体的方式。

我们在建构一个数学物理理论，例如量子电动力学时，我们不得不使用数学概念和数学构造，诸如希尔伯特空间、局域场、对称性，甚至鬼粒子，因为我们只有通过这些概念和构造才可以在某种程度上接近深层实在。由于我们关于深层实在的知识典型地采用一组关于基本的虽然是假设的不可观测实体及其结构的结构陈述的形式（绝大多数是数学陈述），因此，对于这种主张（即我们关于世界的结构关系的客观知识保证了这些实体的实在性及其客观知识）至关重要，注意的焦点应该从数学结构本身移向已被赋予因果有效性和完全不同的质性的潜在实体的结构知识。

但是，我们如何能够确保我们真的通过结构知识到达不可观测的实体和它们的内禀属性，而不是停留在这种结构知识的范围之内呢？

- 对于属性来说，更容易论证。由于具体结构作为它们的关系物的特定关系，由作为它们的关系物的因果有效的属性所决定，因此我们能从关于一个结构的知识合法地推论出关于一个有特定因果力的特定属性的知识，这个特定因果力负责特定关系或结构的形成。
- 对于实体来说，情形更为复杂得多。为了说明任何类型的实体都能通过它们的认识上可获取的内部或外部结构性质和关系来接近，我们可以从表达关系的定律开始，依照这些定律来构成实体，或者从不变性质开始，（从对称群导出的）不变性质是重要的结构性质，广泛用来表征量子领域中的实体。

但另一方面，本体结构论者会论证说，既然实体可以通过定律和不变量来决定，而定律和不变量是本体论优先于实体的，那么实体实际上只享有一个隐喻的地位，它们在本体论上已被融解进了表达定律和对称性的数学关系中。诚然，我们能论证说诸如电荷和质量这些不变性质不足以确定某类实体。例如，正电子有和电子相同的质量，有和质子相同的电荷。因此，需要更多的东西。也就是说，必须找到一种确定的关系，这种关系指明了与一类特殊实体相联系的全部不变性质，因而对于一类特殊实体来说是特定的，这样才能表明某类实体是客观构成的并存在于物理世界，而不仅仅是隐喻。

这里的一个关键概念是“约束条件”，这个概念为世界的（相对稳定和客观的）结构知识从（纯主观的和总是飘忽不定的）感觉中涌现提供了一个主要的辩护；根据逻辑经验主义阵营的感觉资料的基础主义，感觉是我们唯一可直接获取的东西。有了约束条件这个概念后，我们能把这个概念应用于新的概念场景，并论证说当我们的结构知识满足某些约束条件时，我们就可以对世界上真正的不可观测实体<sup>4</sup>拥有客观知识。然后，关键的事情是需要阐明这些约束条件对于不可观测实体的客观知识从它们的结构陈述中涌现是必要的，对于辩护把这些实体引进论述是必要的，最终也对于辩护实体的实在性从结构知识中建构是必要的。当我们有了一组经验上适当、性质上不同的结构陈述（所有这些结构陈述涉及一个不可观测的实体并描述它的某些特征）时，就能对上面提到的对于涌现和辩护必要给定集合组织上的约束条件说明。

在结构进路中，实体诸如光子、胶子和希格斯玻色子都由展示它们的内禀和本质（因果有效的）特征的结构关系构成。在此，客观性概念不是一个与人类参与相分离的概念，那是一个幻想。确切地说，客观性是用自然界反抗任何随意的人为建构来定义的。以量子电动力学的情形为例，物理学家以各种各样的方式来构想电磁场的组成部分。沿着一条概念线路，物理学家最初通过一组结构知识把电磁场的组成部分构想为微粒，然后又把微粒重新构想为光子。随着规范对称性概念提出的结构约束进入“电磁相互作用”图像，物理学家再次把光子重新构想为某种规范玻色子，光子概念获得了自然界的认可，也为科学家共同体所接受。沿着另一条思想线路，物理学家构想电磁场的组成部分是电磁波，这没有得到自然界的认可，因而也为科学家共同体所遗忘。所有这些都是人类用结构关系建构的结果，但是，某些观念和构造物的客观性，而不是其他一些观念和构造物的客观性，是以自然界的认可和不认可反应作为根据的。由于这个理由，不能把这里讨论的建构和再建构当作纯粹的主观行为而不予考虑。

<sup>2</sup>量子色动力学（简称 QCD）是一个描述夸克、胶子之间强相互作用的标准动力学理论。

<sup>3</sup>胶子是负责在两个夸克之间传递强作用力的基本粒子。



基本实体的结构进路为基本物理学中的理论建构提供了一个实在论的说明。正如我们在 6.1 节中所指出的，建构一个基本理论的中心任务是规定它的概念基础，阐明它的基本实体是什么。除了从现存的几乎经验上适当、概念上合乎逻辑的理论和可靠的观察资料提出的约束中获得一些暗示外，物理学家没有接近假想的基本实体的任何认识途径，在电弱规范理论的案例中，这些指的是可获得的关于黑体辐射、光子反相关、散射截面、弱中性流和量子反常的经验材料；以及重整化、电弱相变、规范对称性、对称性自发破缺和希格斯机制的理论概念。如何从这些暗示行进到对一个像电弱统一理论这样的新基本理论的概念基础的公式化表述，这是一个令人畏惧的挑战。然而，在规范场论的案例中，一群物理学家采纳了一种与上面讨论的本体论综合概念相一致的方法，令人满意地回应了这一挑战。这一节致力于表明，对于规范场论的建构至关重要的是一种包含了构成各种对称性自发破缺、希格斯机制、规范对称性和可重整性的结构约束的综合。

对称性学派的指导思想是从破缺的对称性中提取有用的物理信息，即关于规范粒子的结构知识。出于这个目的，Higgs 从唯象的观测结果提示的一个对称性破缺等级推导出规范玻色子质量的产生机制，然后把它们呈送给实验检验<sup>4</sup>，尽管人们也不知道自发对称性破缺会不会破坏理论的可重整性。从结构主义的视角看，这里涉及规范粒子的外部结构特征，这些外部结构特征刻画规范粒子之间的外部关系而不是它们的内部结构组成和约束。非阿贝尔规范理论的可重整性的证明是非常困难的，一个首先面临的困难就是非阿贝尔规范场的量子化，因为规范粒子具有非物理的极化自由度，这会对量子化带来很大的困难，传统的正则量子化将很难进行。't Hooft 的行动富有成果，他利用 Faddeev-Popov 量子化把对正则量子化的应用扩展到非阿贝尔规范理论，并证明了电弱统一理论的可重整性。

对称性破缺的数学结构表明了电弱相互作用的深刻统一性并启发了强子层面规范对称性的探索。在此之前，根据重子和介子的八重分类法，盖尔曼试图描述强子的电-弱流属性，而奈曼受汤川秀树的激励，尝试得出强相互作用的一个精确的 SU(3) 规范对称性。然而，在探索的那个阶段，基本费米子和玻色子，包括奈曼的规范玻色子，都是重子和矢量介子八重态。通过事后的觉悟，可以理解，与泡利十年前的失败相似，奈曼在强子层次探索规范对称性将一无所获。

而盖尔曼受到电弱统一理论的影响，利用流把对 SU(3) 么正对称性的应用从强子光谱扩展到强子的弱相互作用和电磁相互作用。根据 Noether 定理——如果有一个流守恒，那么就有一个对称性。U(1) 是电磁对称性；SU(2) 是电弱同位旋矢量流对称性（由于守恒矢量流假说）；那么，QCD 中假想的 SU(3) 对称性就是假想的么正流对称性。因此，流作为电弱过程中能量交换的媒介，变成了 SU(3) 表示，其结果就是流代数的产生。

流代数描述的是一个局域场理论的抽象结构特征，人们期望它在未来的一个真实理论中是有效的。流代数的整个框架只是流之间的一个数学结构。由于流表示一些未知的物理实体的物理过程的结构，而不是表示存在的物理实体本身，因此可以把流代数看作是结构的数学描述。

在某种意义上，流代数是结构主义方法论的一个范例应用。普遍的结构特征从一个场论模型的假想对称性中推导出来，而不考虑模型的细节。然而，一旦假定流而不是强子对强子过程负责，这一假定就在本体论承诺中发出了从强子到流的一个微妙转变的信号，从而进一步研究流的（外部和内部的）结构属性的大门就敞开了。因此，尽管这个进路初看起来是唯象的，似乎有一个不可捉摸的性质（即回避特定的实体的本体论和动力学问题），但是，正如我们将要看到的，由结构知识提供的这一楔子为探索更深层次开辟了道路。

尽管不清楚这些流实际上是什么，但是物理学家已从与各种强子过程相关的流代数中提取了许多结构属性（“代数关系”）。流可以从各种模型，例如  $\sigma$  模型和夸克模型中推导出来。一般来说，物理学家不把这些模型看作是实在的东西，而是作为启发性手段。然而，尽管最初物理学家也不把夸克看作是真实的实体，但是夸克模型确实为我们探查强子的组成部分提供了一个机会。

盖尔曼最初想从夸克概念得到的东西是：在场论模型中使用它们来提取强子过程中可能有效的代数关系，也就是说，把夸克作为流中的占位符，流则作为结构，满足流代数提出的所有结构约束。然而，从概念上来讲，夸克的提出使得从寻找结构知识转变为寻找作为结构基础的实体的结构知识成为可能。然而，为了把这种可能性转变为现实，必须取得实质性的进展。夸克不是作为独自存在的强子的组成部分而引入的，而是假定它纠缠在关系的网络之中。沿着与盖尔曼的结构探索相反的线路，一些物理学家假定夸克是强子的真实组分，从夸克中能推导出强子光谱。

<sup>4</sup>基于 Higgs 机制对  $W^\pm$  和  $Z^0$  粒子的高精度的质量预言给 Higgs 机制提供了很强的间接支持。

沿着由盖尔曼本人开创的另一条线路，夸克由流（作为关系结构）定义，夸克是流的占位关系物，它们与强子的关系最初完全不清楚。只是随着从流代数推导出来的现象的结构特征的累积，夸克与强子的关联才逐渐显露。由于这个理由，这一研究线路堪称是体现从结构进路到实体再到理论创造的典范。

从探查强子的外部结构特征移向探查强子的内部结构特征，这最初的一步由盖尔曼提议，阿德勒在他的局域流代数求和定则的推导中采取了这一提议。阿德勒求和定则表示对用轻子来探查强子的结构特征这一过程给予的一般结构约束。正如比约肯所做的正确承认，阿德勒求和定则是逐步建立一些重要进展的概念基础。然而，阿德勒求和定则本身提出的结构约束的性质，不论是内部的还是外部的，都仍然不清楚。这主要是因为求和定则中，中间态、强子态或亚强子态的性质还不清楚。关键的一步是比约肯通过“历史类比”，在他的标度无关性假说中迈出的，这一步清楚地把阿德勒约束转变为内部结构约束：强子由类点的、渐近自由的部分组成。这是从结构进路到实体的关键一步，这一步把物理学研究的焦点从寻找结构知识转向寻找作为结构基础的潜在实体的结构知识。

这一步在 1969 年为高能物理界采纳之后，潜在实体的结构知识迅速在斯坦福直线加速器中心和世界上许多其他高能物理实验室以及许多理论研究机构积累起来。在这个范畴中最为重要的成就与标度无关性密切相关，这些成就导致了部分子概念和部分子的一组结构特征。除了类点和渐近自由外，物理学家也坚定地确立了在强子中一半部分子有夸克量子数，另一半没有。这一了解揭示了组成强子的组成部分的结构特征，因此对于组分概念和量子色动力学的建构都是至关重要的。

在量子色动力学中，夸克和胶子不是由组成部分子的结构特征组成的。实际上，它们是由组成部分子的因子和组成其他实体的因子联合组成。更为特殊的是，**色概念**携带了强子及其组分之间结构约束的大量信息，对在量子色动力学中建构出的夸克和胶子的组成起着十分重要的作用，色概念起源于 SU(6) 组分夸克模型，但在量子色动力学建立之前用来组成流夸克和部分子的一个要素。（色）是从组分夸克模型输运到流夸克模型的，这种输运不是机械的而是有改造能力的：（色）不再限制在 SU(6) 模型的语境中，而是能应用于其他语境的情形中，诸如电子-正电子过程，这些过程与 SU(6) 对称性无关。但是，从组分夸克模型输入的色以及所有与色相关的结构约束，并不真的是构成量子色动力学中夸克和胶子的色。量子色动力学中夸克和胶子的色是另一个本体论综合的结果。也就是说，它是组分夸克模型的色与组成**靴祥强子**的一个额外要素（即强子的任何部分都不能存在于渐近自由态的结构约束）形成的一个组合。在设想夸克是带色的语境中，色单态被限制为物理世界的渐近自由态，相当于在世界的渐近自由态的色守恒。运用这样一个色守恒的构成性约束，弗里奇和盖尔曼在建构夸克和胶子上迈出了决定性的一步，即把胶子看作形成一个色八重态，把胶子八重态看作精确的规范色 SU(3) 对称的规范玻色子。一旦迈出了这一步，量子色动力学的概念基础的建构就完成了。

本体论综合的概念有助于我们理解科学发展（在常规科学时期和科学革命时期）的丰富结构。除了简单地抛弃和保留理论改变过程中的基本实体外，这一概念提供了一个能容纳明显的本体论不连续性（新的基本实体取代旧的基本实体）的机制，同时，就激进的理论变革前后存在组成新的自然类实体的要素而言，揭示了我们对于世界上存在什么的一个深刻的认识连续性。然而，这一项目更为有趣的地方是它有助于我们理解一个新的基本理论是如何创建的。

除了规范场论的创建外，夸克的组成要素的约束和对称性自发破缺的约束一起导致了另一组以二元论概念为中心的复杂约束，二元论概念甚至在量子色动力学建构之前就成为了弦的构成要素。在一个深层的意义上，结构实在论现在似乎正在经历一个相变，从主要关心理论变革的旧阶段向主要关心新的基本理论的建构的新阶段转变，这一转变的方法论导向和概念历史过程能按照本体论综合给予清楚的分析。

显然，老的建构经验主义在可观察和不可观测之间所做的区分，完全无力表达当前的科学尝试，特别是在分析像量子色动力学、弦论和量子引力这样一些基本理论的建构上，完全无能为力。由于这个原因，这种经验主义的区分在人类是否可接近所谓的“不可观测之物”的讨论中完全没有分量。实际上，建构结构实在论的总体思想起源于认识到结构进路在我们接近“不可观测的”物理实在上的特殊特征。

有两种获取源。首先，前驱理论（它们的领域全部或部分与打算研究的领域相重叠）提出的基本约束已对我们在理论建构中打算研究和概念化的不可观测实在提供了一个主要的认识途径，因为这些约束已编码了人类在过去已获得的具体化在前驱理论中的全部可靠的结构知识。本体论综合的思想非常关键地建基于对这种获取源的认识。其次，正如盖尔曼反复提醒物理学共同体的，一些结构约束，他称之为“代数结构”，原则上是可知的。结构主义者假定结构知识的可达性，并在科学经验的基础上广泛接受结构知识的可达性。

凭借这两个进入窗口，以及其他的认知手段，诸如实验装备，概念的、理论的和数学工具，微观实体诸如电弱规范玻色子、夸克和胶子都是认知上可接近的，足以让我们声称结构地建构的实体的实在性。



当然，一些哲学家已论证说，在现代物理学中微观实体的一些形而上学性质能用同一性和个体性这样一些基本范畴来阐释，但是，这些形而上学性质是完全不可认知的。他们进一步论证说，既然按照传统的形而上学，这些形而上学性质是一个微观实体的完整组成中必要的组成部分，没有它们就不能把理论实体看作是一个真实实体，因此，我们的确不能声称我们有充分接近微观实体的认知通道。然而，如果这些性质原则上是不可认知的，那么不管它们在传统形而上学中是多么基本，它们对于现代形而上学和现代物理学来说都是不相关的。形而上学必须随着物理学的进步而修正。因此，同一性和个体性这些范畴必须从当代形而上学中清除出去。就构成一个实体物理上（也是形而上学上）相关的要素而言，原则上，没有一个是不可认知的。

最后，回到我们一直企图解答的问题——**光子和夸克这样的微观粒子真实存在吗？**

在物理学界建构和接受规范场论的过程中，成为一个真粒子的标准正处在激进的修正之中。目前足以指出，对于盖尔曼来说，在他建构夸克和胶子时，所谓一个真粒子意味着一个能渐近自由地存在、因而是在实验室中可探测的粒子。一旦修订了这一标准，物理学家就能完全确信夸克的物理实在性。例如，布赫霍尔兹通过把重整化群变换应用于局域可观测量，有说服力地论证说，夸克、胶子和色的物理实在性能够从可观测量的代数（描述假想实体的结构特征的一组经验上可认知的关系）中，揭示为以可观测量的代数为基础的基础理论的内禀特征。

从认识论上讲，这意味着我们能从关于可观测量的结构关系（可观测量的代数）的知识继续向前，到达微观实体（夸克和胶子）的物理实在性，这可以形而上地构想为经验内容（“可观测量”）的结构，与 Ramsey 语句<sup>5</sup>的指称形而上地一致。

既然理论实体的证据地位被编码在理论中，因此光子的实在性不能与量子电动力学的有效性分开，因为它们正是从量子电动力学中建构出来的。只要量子电动力学得以确证，光子就是真的，是在量子电动力学框架中的真。如果一个更为基本的理论取代了量子电动力学，情形又将如何？这仍然不会剥夺光子的实在性，仅仅是改变了它的本体论地位，从首要的（基本实在）变为派生物（副现象）。这意味着我们关于实在的客观知识是连续累积的，即便这一连续性随着基本实体的更替一次又一次地中断。

上面的讨论还产生了另一个哲学问题：**什么是真实？**

如果一个物理世界中可以测量的只有线性响应，那在这个世界线性响应就代表着所有的事实，物理理论就将是线性响应的理论，在我们的世界里，我们可以测量的事物当然不只是线性响应，但是似乎可测量的也不比线性相应多多少，我们可以测量的其实只是关联函数，这使我们不禁很想用关联函数来定义世界上的物理理论，关联函数可能就代表着我们世界的真实。

这一点之所以重要是因为物理理论中的很多概念并不代表真实，一旦新的理论发展起来，这些概念就可能改变。例如，具有位置和速度的点粒子是牛顿经典理论的基本概念，其理论大厦的建筑砖块，我们过去相信所有事物都由粒子组成，粒子是真实的基础，我们现在相信粒子不代表真实，相反，线性希尔伯特空间中的量子态代表真实。但是，量子态不是我们可以直接测量的事物，有朝一日超越量子理论的新理论发展出来，谁也不能保证量子态概念不会有粒子概念相同的命运。

物理学是一门测量的科学，不幸的是物理理论常常包含许多不代表真实的内容（例如牛顿经典力学中粒子的概念）。同一真实可能用完全不同的理论去描述，甚至可能基于根本不同的概念。因此透过理论中的形式和非真实概念的烟幕，密切注视真实（如关联函数和可观测量）是非常重要的。把这一思路推向极端，我们可以认为量子态和量子算符也不代表真实。它们只不过是用于计算可以为实验所测量的关联函数的数学工具。

正文末了，感谢 Jakob Schwichtenberg 的《基于对称性的现代物理学》一书（和超理汉化组的翻译），本 notes 中绝大部分物理内容都源自个人与此书的不期而遇。此外，也要感谢曹天予所著《20 世纪场论的概念发展》，正是在该书中，Cao 发展出了本体论综合和建构结构实在论，这也是本 notes 后阶段论证的核心。

<sup>5</sup> $\text{Ramsey}(T) := \exists X_1 \dots \exists X_n \hat{T}[X/R]$ ，对于一个物理理论  $T$  来说， $\text{Ramsey}(T)$  表达了  $T$  的所有真实的内容。

# Chapter 7

## 附录 A 标准模型

到目前为止，我们对规范理论有了一个大致地了解。本节的目的是，利用规范理论来试图理解这个世界究竟是如何运作的。根据**还原论**的观点，只要知道这个世界的基本组分以及基本组分之间的相互作用，就明白了世界的运作方式。在本文的结尾，我们会给出目前人类对微观世界已经被实验验证的最深刻理解——粒子物理标准模型。

根据人类现在的认知，世界由夸克、轻子、规范玻色子和 Higgs 粒子组成。其中夸克、轻子有三代，是费米子。这些费米子构成我们常见的一切物质：夸克构成质子和中子，质子和中子组合成原子核，它和电子一起构成原子。从而激发这些粒子的量子场也被称为物质场。规范玻色子负责传递物质之间的相互作用，而人类目前认知的基本相互作用一共有四种——电磁相互作用、弱相互作用、强相互作用和引力。这四种相互作用除引力之外<sup>1</sup>的三种都可以用下面讨论的标准模型所描述。

标准模型是一个规范理论，并且由两部分相对独立的理论构成：描述强相互作用的理论就是量子色动力学，它是一个 SU(3) 的规范理论。描述电磁相互作用和弱相互作用的理论是电弱理论，它是规范群为 SU(2) × U(1) 的规范理论，有时候也称这两个群为电弱 SU(2) 和超荷 U(1)。不过正如前面所见的，如果没有对称性自发破缺的话，由于规范玻色子的质量项会破坏规范不变性，于是规范玻色子总是无质量的。但由于实验要求传递弱相互作用的玻色子是有质量的，于是需要引入 Higgs 机制来赋予传递弱相互作用的玻色子以质量。

最终，我们得到标准模型的 Lagrangian<sup>2</sup>：

$$\begin{aligned}\mathcal{L}_{\text{SM}} = & -\frac{1}{4}G_{\mu\nu}^A G^{A\mu\nu} - \frac{1}{4}W_{\mu\nu}^a W^{a\mu\nu} - \frac{1}{4}B^{\mu\nu} B^{\mu\nu} - (D_\mu\phi)^\dagger (D^\mu\phi) - V(\phi) \\ & + \bar{q}_{I,L,i,\alpha}(i \not{D})_{ij,\alpha\beta} q_{I,L,j,\beta} + \bar{u}_{I,R,\alpha}(i \not{D})_{\alpha\beta} d_{I,R,\beta} + \bar{l}_{I,L,i}(i \not{D})_{ij} l_{I,L,j} + \bar{e}_{I,R}(i \not{D}) e_{I,R} \\ & - y_{e,IJ} \bar{e}_{I,R} \epsilon^{ij} \phi_i l_{j,L} - y_{d,IJ} \bar{d}_{I,R,\alpha} \epsilon^{ij} \phi_i q_{j,L,j,\alpha} - y_{u,IJ} \bar{u}_{I,R,\alpha} \phi^{ti} q_{j,L,i,\alpha} + \text{h.c.}\end{aligned}\quad (7.1)$$

其中，协变导数就是<sup>3</sup>：

$$D_\mu = \partial_\mu - ig_s G_\mu^A T_R^A - ig_2 W_\mu^a I_R^a - ig_1 B_\mu Y \quad (7.2)$$

这个 Lagrangian 就蕴含着人类现在对微观世界的、已经被实验检验的最深刻的认识。

在标准模型中，存在着两个连续的全局 U(1) 对称性——重子数对称性 U(1)<sub>B</sub> 和轻子数对称性 U(1)<sub>L</sub>，其荷被称为重子数和轻子数。在标准模型的粒子谱中，所有夸克都具有重子数 1/3 和轻子数 0，而所有轻子都具有重子数 0 和轻子数 1。

但理论计算显示，这两个全局 U(1) 对称性都是**反常的**，这意味着在宇宙的演化过程中总会有一个过程中出现重子数、轻子数不守恒。

<sup>1</sup>至于为什么引力无法用量子场论描述——引力是不可重整的。

<sup>2</sup>这里面指标实在是有点多，需要梳理一下：首先  $\alpha = 1, 2, 3$  是 SU(3) 基础表示的指标（色指标）， $A = 1, \dots, 8$  是 SU(3) 伴随表示的指标。 $i = 1, 2$  是 SU(2) 基础表示的指标， $a = 1, 2, 3$  是 SU(2) 伴随表示的指标。大写字母  $I = 1, 2, 3$  是代指标。 $\mu = 0, 1, 2, 3$  是 Lorentz 指标。

<sup>3</sup>为了区分 SU(3) 和 SU(2) 生成元，我们将 SU(2) 生成元的记号改为了  $I$ 。



这些重子数、轻子数不守恒的过程对早期宇宙中发生的**重子数产生**而言是十分重要的。在宇宙大爆炸时，宇宙中应该不存在任何重子数，但是我们今天看到的宇宙充满重子数为正的物质，而重子数为负的反物质不见踪影。从而在宇宙演化到今天的过程中，一定发生了一些重子数不守恒的过程。

除了重子数不守恒之外，Sakharov 给出了重子数产生的两个其他条件：CP 不守恒和宇宙演化远离平衡态。前者是为了保证重子和反重子的产生过程具有不同的振幅；后者是因为如果宇宙演化一直处于平衡态，那么由于细致平衡原理，重子数产生速率及其反过程速率一定相同，从而宏观上不可能具有净的重子数产生。但根据计算，标准模型的重子数破坏、CP 破坏和非平衡程度事实上都还不够产生当前我们看到这么多的重子数，这也是很多超出标准模型的新物理模型的动机所在。

直至现在，所有超出标准模型的新物理模型都没有相应的实验验证。这很大一部分因素是我们没有看到除中微子振荡之外的超出标准模型的实验现象<sup>4</sup>。于是验证这些新物理模型，我们就要从实验上寻找超出标准模型的蛛丝马迹。

新物理的能标应该是大于现有对撞机所能达到的能标的，否则我们就应该已经从对撞机中将这新粒子撞了出来。既然新物理能标大于标准模型能标，那么就可以在路径积分中将所有的 BSM 粒子积分掉，得到一个截断在新物理能标  $\Lambda_{\text{BSM}}$  的有效理论——这被称为 SMEFT。在 SMEFT 中可以含有很多高量纲的算符，并且所有这些高量纲的算符的来源都应该是新物理。于是搜寻超出标准模型的新物理可以通过搜寻这些高量纲的算符的方式进行。在找到这些高量纲算符之后，就可以自下而上地利用有效场论找到新物理可能的形式。

我们要搜寻什么样的高量纲算符呢？首先，由于  $\text{SU}(3) \times \text{SU}(2) \times \text{U}(1)$  的规范对称性对世界的描述是十分成功的，于是我们不是很希望新物理会破坏这个规范对称性。如果我们要搜寻的算符不破坏标准模型中的任何偶然对称性，那么标准模型也会贡献量子作用量  $\Gamma[\phi]$  中这一算符对应的项。由于我们的可观测量可以完全由量子作用量给出，于是这一项耦合对任何可观测量的贡献将会被标准模型冲淡。除非我们能够精确测量相应的可观测量，否则搜寻不破坏任何对称性的高量纲算符的过程将是困难的。

于是一个更好的选择是搜寻破坏标准模型中偶然对称性的高量纲算符。标准模型中最重要的两个偶然对称性就是**轻子数守恒**和**重子数守恒**，它们分别对应两个不同的全局 U(1) 变换。对于轻子数变换，可以选定一代轻子，这一代中所有轻子场的变换都是  $e^{i\theta}$ ，反轻子场的变换都是  $e^{-i\theta}$ ，而这一代轻子之外的所有其他场都不变。对于重子数变换，所有夸克的变换就是  $e^{i\theta/3}$ ，反夸克的变换就是  $e^{-i\theta/3}$ ，其他场不变。破坏这两个对称性的过程一旦出现，就相当于找到了超出标准模型高量纲算符。

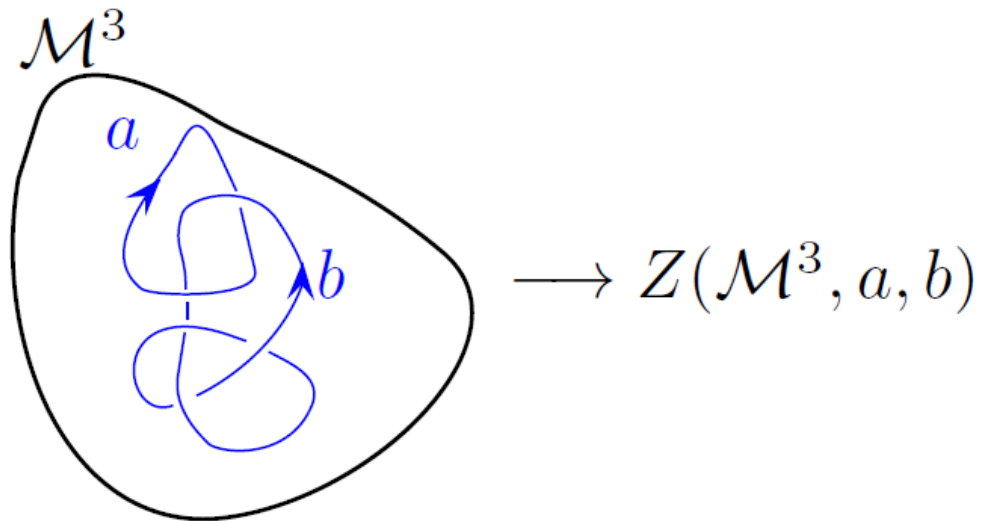


图 7.1: Chern-Simons Theory

<sup>4</sup>暂且不论星系旋转曲线这些观测结果，至少在地球上的实验没有看到超过  $5\sigma$  的实验现象。也许  $\mu$  子反常磁矩会是下一个？

## 附录 B 拓扑量子场论

拓扑量子场论（简称 TQFT）是一类计算拓扑不变量的量子场论，其共同特征是某些相关函数不依赖于背景时空流形的度量。严格地说，TQFT 关注量子场论中一类特殊的构造——**拓扑场构型**。

拓扑场构型是一些满足运动方程的经典场，从而也是作用量的鞍点。讨论拓扑场构型的第一个理由是：在量子场论中，由于我们的路径积分要对所有可能的给定的经典场构型求和，于是一些拓扑场构型（例如**瞬子**）也自然会包括在路径积分中。在平时使用的微扰论中，我们仅仅在真空场构型附近对路径积分做鞍点近似，忽略了其他鞍点对路径积分的贡献。而通过考虑这些拓扑场构型，我们可以将其他的鞍点的贡献考虑在内。

另外一个理由是：在一些不同于真空的边界条件下，一些被称为**孤子**的拓扑场构型可能具有局域的能量，其行为类似一个粒子。并且由于拓扑上的原因，这些场构型是稳定的。于是在量子理论中，我们也相信这些东西对应一个理论中可能出现的、并非场的微扰激发的稳定粒子。于是为了考虑理论中可能出现的所有粒子，也需要将这些孤子考虑在内。更加有趣的是：孤子实际上可以在规范理论中给出**磁单极子**。如果我们在实验中找到了磁单极子，那么它很有可能是一个孤子。

场论的一个重要任务是给出配分函数，大致来说就是场论在每一个（闭的）时空流形会给出一个数。配分函数可以用路径积分来计算，而路径积分最重要的性质应当就是它可以被“分解”来计算：把时空分解成很多小块，如果我们知道了在每个小块上怎么做计算，那么整个时空上的配分函数就是这些小块上的结果“拼”起来。这种可以被分解的性质基本上就是函子性。所以我们或许可以把场论定义成某种几何范畴到某种代数范畴的函子。所谓“拓扑”的意思是，配分函数只与时空流形的拓扑有关，而跟其它的结构（比如，度规）无关。

一个朴素的想法是，我们可以形式化一个只取决于拓扑性质的量子理论，即理论不依赖于时空度规的微小变形：

$$\frac{\delta}{\delta g_{\mu\nu}} \langle * | (\cdots) | * \rangle = 0$$

其中  $\langle * | (\cdots) | * \rangle$  代表这个理论的任意关联函数。这样定义 TQFT，完全合理，但不太有用。

另一种定义（2+1 维）TQFT 的方法是，将它看成一套规则，输入是一个带标记的嵌入在三维流形中的链环，输出是一个复数，在某种意义上，它在光滑形变下不变。这个定义与我们对纽结不变量的定义很像，但有两个关键的不同：

1. 我们容许这条链环带“粒子类型”标记（求出最终结果的规则取决于粒子类型）。
2. 这条链环可以嵌入在任意复杂的三维流形中。

这种映射（如图(7.1)）就像 Chern-Simons 配分函数  $Z(\mathcal{M}, \text{链环})$ 。用这种语言的优势在于，Chern-Simons 泛函积分严格来说往往在数学上不能良定义。我们这里就避开了 Chern-Simons 场论，仅仅将 TQFT 定义为从带链环的流形到某些输出的映射。

一个与此紧密相关但更形式化的 TQFT 定义，由 Atiyah 在 1983 年的一篇文章中以一套公理给出，它包含更丰富的信息。

在文章中，拓扑量子场论 (TQFT) 指的是 2- 态射范畴  $\text{Fct}(\text{Cob}, \mathcal{V}\text{-Alg})$ ，其中  $\mathcal{V}$  是张量范畴，或写作 2- 态射  $\text{Cob} \rightarrow \mathcal{V}\text{-Alg}$ ，关于其的 Atiyah 公理，兹列如下：

设  $\mathcal{M} \in \text{Mor}_0[\text{Fold}_{D+1}]$ ，其的  $D$  维可定向截面记作  $\Sigma^D$ ，那么，

- 公理 1: 存在态射  $\text{Fold}_D \rightarrow \text{Vect}(\mathbb{C})$  使得  $\Sigma^D \mapsto \mathcal{H}(\Sigma^D)$ ，其中  $\mathcal{H}(\Sigma^D) \in \text{Vect}(\mathbb{C})$  是一个复 Hilbert 空间，态射只取决于  $\mathcal{M}$  的拓扑；
- 公理 2:  $\text{Fold}_D \sqcup \text{Fold}_D \rightarrow \text{Vect}(\mathbb{C}) \otimes \text{Vect}(\mathbb{C}), \emptyset \mapsto \mathbb{C}$ ；
- 公理 3: 若  $\Sigma^D = \partial \mathcal{M}$ ，则存在一特殊元素  $Z(\mathcal{M}) \in \mathcal{H}(\partial \mathcal{M})$ ，选取取决于  $\mathcal{M}$  的拓扑；
- 公理 4:  $\mathcal{H}(\Sigma^D)^* = \mathcal{H}^*(\Sigma^D)$ ，右上角标  $*$  代表取反定向；
- 公理 5:  $\mathcal{M}$  兼容配边，即  $Z(\mathcal{M}) \in \mathcal{H}(\Sigma^*) \otimes \mathcal{H}(\Sigma')^*$  且给出  $\mathcal{M}$  的演化。

值得注意的是公理 3，这个公理其实是说，从容许的向量空间选出的系统的特定波函数  $Z(\mathcal{M})$ ，取决于系统的时空历史。我们已经多次遇到这个原理。例如，我们知道如果粒子-反粒子对沿洞运动，我们得到的波函数会变化，这个过程可能是时空历史的一部分。

公理 3 意味着，对闭的  $\mathcal{M}$ ， $\partial \mathcal{M} = \emptyset$ ，有

$$Z(\mathcal{M}) \in \mathbb{C}$$

也就是说，TQFT 赋予流形的拓扑不变量一定是个复数。这就是 Chern-Simons 理论中发现的。

我们已经强调过，拓扑场论可在一般的代数框架下研究，不需要具体的作用量。关联函数和配分函数被看成给定的量，量子场论提供它们间的关系。理论最重要的代数性质是因子化。

考虑  $\mathcal{V} = (\mathcal{V}, \otimes)$  是张量范畴，那么自然可以引出  $\mathcal{V}$  上的圆盘代数  $D : (\text{Disk}, \sqcup) \rightarrow (\mathcal{V}, \otimes)$ ，接下来要做的是进行推广（是否可以对一般的流形  $\text{Fold}$  做类似的事情）。

利用上文定义，即 Mayer-Vietoris 谱序列与同伦极限得到  $H_n(X) = H_n(X^\bullet, d_X^\bullet) \in \text{Mor}_0[\text{CSh}(X)]$ ，考虑一个  $\mathcal{U}$ -小范畴  $\text{Disk}_{\mathcal{M}}$  为嵌入到  $\mathcal{M} \in \text{Mor}_0(\text{Fold})$  中的圆盘构成的范畴，定义圆盘代数  $D$  的因子化同调为其的同伦归纳极限：

$$\mathcal{MH}_n(D) := \text{holim}_{\text{Disk}_{\mathcal{M}}} A$$

其给出张量 2- 态射  $H_n : (\text{Fold}, \sqcup) \rightarrow (\mathcal{V}, \mathcal{V}\text{-Alg})$ ，这其实就是前文所述的拓扑量子场论作为一个 2- 态射  $\text{Cob} \rightarrow \mathcal{V}\text{-Alg}$  的类似版本。于是，我们就得到了如下的关系：

$$\begin{array}{ccc} \text{gp Disk}^{\text{sm}}(\mathcal{V}) & \longrightarrow & \text{gp } \mathcal{V}\text{-Alg} \\ \downarrow f_{\text{disk}} & & \downarrow Z \\ \text{gp}_* H_n[\text{Fold} \rightarrow \mathcal{V}\text{-Alg}] & \longrightarrow & \text{Fct}[\text{Cob}, \mathcal{V}\text{-Alg}] \end{array}$$

其中  $Z$  就是 TQFT 的配分函数，左下角的那部分是所有  $\text{Fold}$  上的因子化同调，右下角那部分是（拓展的）拓扑量子场论。

以下是用到的部分范畴的记号：

$\text{Alg}$  : 代数范畴；

$\text{Ring}, R\text{-Mod}, \text{Mod-}R, (R, R)\text{-Mod}$  : (含么的) 环范畴，环上的左、右模范畴与双模范畴；

$\text{Fct}(\mathcal{A}, \mathcal{B})$  : 2- 态射  $\mathcal{A} \rightarrow \mathcal{B}$  的范畴；

$\text{InvSys}(\mathcal{A}) = \text{Fct}(\mathbb{Z}_{\geq 0}^{\text{op}}, \mathcal{A})$  : 逆向系范畴，其对偶记作  $\text{DirSys}(\mathcal{A})$ ，即正向系范畴；

$\text{Fold}$  : 流形范畴；

$\text{Cob}$  : 配边范畴；

$\text{Disk}$  : 圆盘代数范畴，即对象为圆盘之间的无交并，态射为嵌入；

$\text{PSh}(X), \text{Sh}(X), \text{CSh}(X)$  :  $X$  上的预层，层，余层范畴；

右上角标  $\text{sm}$  表示其光滑，左上角标  $\text{gp}$  表其所括范畴的  $\infty$ - 群胚。

## §8.2 瞬子几何学

现在，我们来看拓扑场论中的一个具体构造——瞬子。简单来说，瞬子是经典场方程的一个含时解，从而也是路径积分中的鞍点。不过瞬子的特殊性在于：瞬子在  $t = \pm\infty$  时趋于两个不同的“真空”场构型<sup>1</sup>，并且具有有限的经典作用量。于是它就会贡献两个不同“真空”之间的跃迁振幅。如果“真空”之间存在这样的跃迁振幅，那么我们必须要对不同的“真空”做一个对角化来使其具有确定的能量，并且我们也只能在这个对角化后的、真正的真空上做微扰论。我们将会看到，这个真空的对角化会带来一些重要的物理效应，而这些物理效应都要归功于瞬子的贡献。

**瞬子**是一个来自量子场论的概念。在经典层面上，它对应 Yang-Mills 方程的一类特殊解。

考虑 4 维可定向紧 Riemann 流形  $M$ ，给定紧半单 Lie 群  $G$ ，以  $E$  记  $M$  上的  $G$  向量丛。联络形式 (规范势)  $A \in \Omega^1(M, \mathfrak{g})$ ，曲率形式 (规范场)  $F \in \Omega^2(M, \mathfrak{g})$ 。4 维的特殊之处在于作为 2 形式空间自同构的 Hodge 星算子满足  $** = 1$ ，从而给出特征子空间分解  $\Omega^2 = \Omega^+ + \Omega^-$ 。这个基本分解也给出

1.  $d_A^\pm : \Omega^1 \rightarrow \Omega^\pm$ ;
2.  $b_2 = b_+ + b_-$ .

$\Omega^2$  可视为作用在  $\Omega^1$  上的斜共轭变换。从表示论的观点看，基本分解  $\Omega^2 = \Omega^+ + \Omega^-$  的存在可以用  $\mathfrak{so}(n)$  中仅有  $\mathfrak{so}(4)$  不是单 Lie 代数来解释：

$$\mathfrak{so}(4) = \mathfrak{so}(3) \oplus \mathfrak{so}(3)$$

在 Lie 群层面上，例外同构  $\text{Spin}(3) = \text{SU}(2)$  给出  $\text{Spin}(4) = \text{SU}(2) \times \text{SU}(2)$ 。SU(2) 丛和 SO(3) 丛是我们的主要例子，在单连通的  $M$  上二者没有本质区别。

满足  $F = F^+$  和  $F = F^-$  的曲率形式 (瞬子) 分别称为自对偶的 (SD) 和反自对偶的 (ASD)：它们将自动满足 Yang-Mills 方程  $d_A^* F = 0$ 。注意到反转定向后，SD 解与 ASD 解互换，故只需集中精力研究其中一类。我们将视情况方便自由转换：通常，在实几何中考虑前者，在复几何中考虑后者，以保持下面提到的拓扑量子数  $k > 0$ 。

在  $\mathfrak{g}$  值 2 形式空间上定义内积  $(A, B) = -\int_M \text{tr}(A \wedge *B)$ 。由 Chern-Weil 理论，

$$(F, *F) = \|F_+\|^2 - \|F_-\|^2 = 8\pi^2 k$$

量子数  $k$  是  $M$  的拓扑不变量： $G = \text{SU}(2)$  时， $k = -c_2[E]G$ ； $G = \text{SO}(3)$  时， $k = p_1[E]/4$ 。另一方面，规范场的 Yang-Mills 泛函 (能量)

$$\text{YM}(A) = \frac{\|F\|^2}{2} = \frac{(\|F_+\|^2 + \|F_-\|^2)}{2}$$

当且仅当  $*F = \text{sgn}(k)F$  时取到最小值  $4\pi^2|k|$ ，换言之，瞬子对应规范场的“基态”。

这也说明了离散量子数产生的一般机制：它们来自紧 Lie 群的表示。

$M$  上的 Yang-Mills 联络构成一个无穷维仿射空间  $\mathcal{A}$ ，其上有规范群  $\mathcal{G} = \Gamma(\text{Aut}(E))$  的作用。 $\mathcal{A}/\mathcal{G}$  通常是一个无穷维参模空间。然而，注意到在规范变换下  $F \mapsto s^{-1}Fs$ ，作为 Yang-Mills 泛函极值点的瞬子是规范不变的。模去群作用后，瞬子的参模空间  $\mathcal{M}$  通常是有限维的，从中可以提取出重要的几何信息。

对于一个规范群为  $G$  的规范理论，其“真空”态一一对应于同伦群  $\pi_3(G)$  中的同伦类。如果  $G$  是紧致、联通且单纯的，那么  $\pi_3(G) = \mathbb{Z}$ ，理论中就会有无穷多个简并的“真空”态。一个瞬子就是场方程的一个具有有限作用量的、联通两个不同“真空”的经典解。瞬子的存在会导致不同的“真空”之间出现跃迁，从而不是量子理论中真正的真空。真正的真空是将这些“真空”的能量对角化后得到的  $\theta$ -真空。在其附近做微扰论时，需要在 Lagrangian 中加上额外的  $\theta$ -项。

<sup>1</sup>这里的“真空”场构型意味着对应经典势的最小值的场构型。我之所以在这里在“真空”一词上加了引号，是因为瞬子贡献了这些“真空”之间的跃迁，于是这里的“真空”并不是量子意义上的真空。在下面的讨论中，我们将沿用这个说法——“真空”对应势的最小值，而不加引号的真空对应量子理论中真正的真空。