

第一章 绪论

第一节 课程简介

量子物理，顾名思义，是以量子力学为基础的物理学，它也是一种精确描述微观粒子运动方式的学科。我们生活中的很多现象，其发生的原因都与量子力学有着深刻的关系。例如我们的世界充满了光，而光是由于原子的跃迁而产生的。如果我们用光谱仪去分析日光灯的光谱，我们会发现它的谱是连续的，但是在这个连续谱的基础上，又会出现几个分立的峰，如图 1-1 所示。

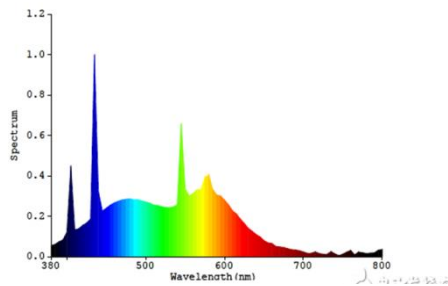


图 1-1 日光灯的光谱（来源于互联网）

为了解释清楚这种光谱特性，就必须利用量子力学的理论。同样的，大家都知道原子是由电子和原子核组成，如果按照经典物理的理解，电子绕着原子核运动，一定会辐射出电磁波，从而会逐渐损失能量并且最终与原子核碰到一起，然而事实上组成我们这个世界的原子大多是稳定的，对这一事实的解释，同样需要借助量子力学。

从技术发展的角度，微电子技术是第三次工业革命的基础，集成电路的发展导致了当前的移动互联网和数据时代，而微电子技术也是完全以量子物理为基础发展起来的。例如固态二极管，是由德国科学家 Karl Braun 于 1874 年提出的。Karl Braun 提出的二极管是一种点接触式的二极管，它是把平板状的自然晶体如 PbS 放置在导电底座上，并且由铜质的尖针作为另外一个电极，与 PbS 的上表面接触，如图 1-2 所示。Karl Braun 发现，如果尖针接触到 PbS 表面的合适位置，当电流的流向不同时，这个结构的电阻是不同的，这个现象就是就是二极管的单向导电性。二极管的发明标志了半导体学科的建立，它也直接导致了无线电接收器的广泛应用，并且在随后的一战中发挥了重要的作用。

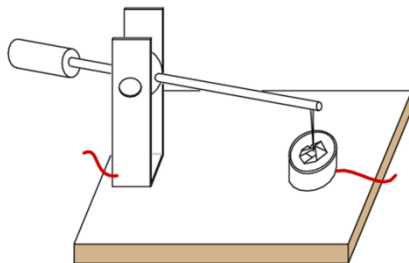


图 1-2 点接触式二极管（来源于互联网）

然而当时人们并不能解释清楚点接触式二极管具有单向导电性的原因，直到一战之后的 1920 年代，当量子力学的理论建立起来之后，A. Sommerfeld、F. Bloch 等人提出了固体中的

电子能带理论，W. H. Schottky、N. F. Mott、H. Bethe 等人进一步基于能带理论，阐释了金属和半导体接触界面的肖特基势垒是这一单向导电性的根本原因。

以点接触式二极管运行机制的理解为标志，人们基于量子力学的原理，对固体材料，尤其是半导体的认识和利用逐步深入。到了二战之后的 1940 年代，贝尔实验室的 J. Bardeen, W. Shockley 和 W. Brattain 首次利用半导体材料，制备了可以实现电压调控电阻的固态器件，学名点接触晶体管，如图 1-3 所示。

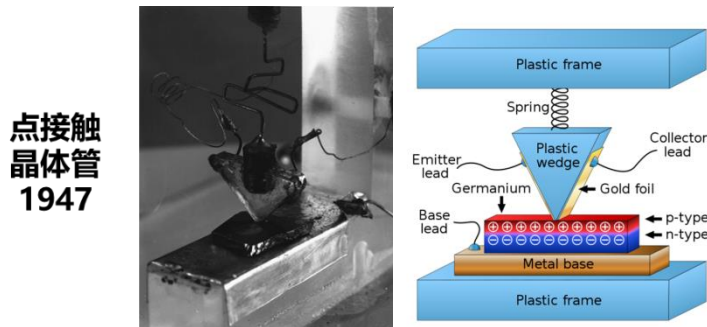


图 1-3 点接触式晶体管及其原理图（来源于互联网）

与点接触式二极管先有现象后有理论解释不同，点接触式晶体管是先有理论构想后有器件实现，这也体现了量子力学理论建立对科技发展的巨大推动作用：人们从被动的认识世界转变为了在理论指导下主动地改造世界。点接触式晶体管的结构与点接触式二极管有很大的相似之处：它是将一块锗单晶放置在导电底座上，并且由一个三角形的楔块与这块锗单晶的上表面接触。与点接触式二极管不同的是，这里的楔块不再是整个一块导体，而是在不导电的塑料侧面镀上了导电的金箔。在与锗接触的地方，金箔被划开一道很细的狭缝，从而整个结构相当于有了三个电极：锗底部的电极，狭缝左侧的电极，狭缝右侧的电极。由于锗的半导体特性，当在其底部的电极上加上不同的电压时，锗上表面的导电特性可以发生很明显的变化，而这一变化体现为狭缝左侧的电极和狭缝右侧的电极之间的电阻变化。这就是最早的晶体管。现代集成电路的基本单元——场效应晶体管(MOSFET)，可以看作是升级版本的点接触式晶体管。

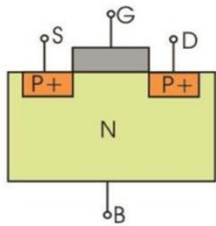


图 1-4 场效应晶体管(MOSFET)（来源于互联网）

如图 1-4 所示，MOSFET 也是由三个电极组成：栅极 G、源极 S 和漏极 D（有时也会有第四个电极 B）。其中栅极 G 就相当于点接触式晶体管中锗底部的电极，而源极 S 和漏极 D 分别相当于点接触式晶体管中狭缝左侧和右侧的金箔电极。当在栅极 G 上加入不同的电压时，源极 S 和漏极 D 之间的导电能力会发生巨大的变化。

固态晶体管的发明带来了我们的信息时代。如图 1-5 所示，最早的电子计算机 ENIAC 是由电子管等非固态器件组成的，它的占地达到了 167 平方米，重达 30 英吨，耗电 150 千瓦。

而这样一个庞然大物，每秒仅可以完成 5000 次简单的加减操作（当然比人力是快多了）。

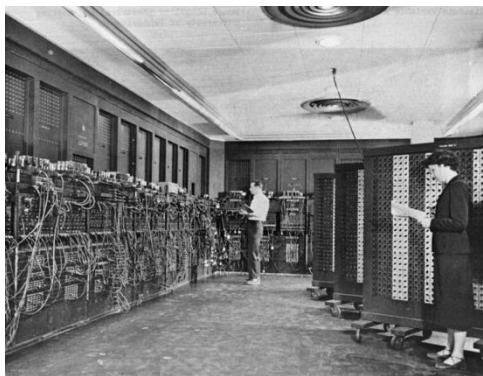


图 1-5 世界首台电子计算机 ENIAC（来源于互联网）

而随着固态晶体管的广泛采用与集成，我们现代的计算机做得越来越小，功耗越来越低，而计算性能越来越高。图 1-6 是 Intel 2020 年发布的一块 i5 10400 CPU 芯片，这块小小的芯片却具有每秒 26 万亿次的浮点运算能力，而与此同时它的功耗仅有 65 瓦。



图 1-6 Intel i5 10400 芯片（来源于互联网）

如前所述，这一巨大进步的最根本原因，是因为人们掌握了量子力学的理论体系，并且进一步基于这一理论去理解固体的中电子状态，从而能够充分利用半导体、金属、绝缘体等材料的电子特性，去设计、制造纳米尺度的微电子器件，并且将这些器件集成起来实现复杂而强大的功能。所以微电子技术的物理基础是量子物理，它包含了量子力学、固体物理、半导体物理和半导体器件物理。

另一方面，当前微电子技术的发展突飞猛进，TSMC 和 Intel 在 2020 年已经分别到达了 5nm 和 10nm 的工艺节点，而中国的 SMIC 也宣布实现了 14nm 的工艺节点。随着集成电路中微电子器件尺寸的持续微缩，短沟道效应、栅极漏电流、亚阈值电流等问题也对摩尔定律的延续构成愈发严峻的挑战。这些问题的解决，需要微电子器件研究和开发者对于势垒、量子限域、量子隧穿、界面散射等量子物理现象具有深刻的理解。所以对于有志于在微电子器件方面从事研究、研发的同学来说，量子物理和固体物理一方面是进一步学习半导体物理和半导体器件物理所必须要打牢的基础，另一方面也是针对微电子科技的进一步发展的瓶颈，将来做出创造性工作甚至解决卡脖子问题所必备的理论储备。

基于以上考虑，本课程旨在面向微电子专业的同学，阐述量子力学的基本概念，使同学们建立量子物理的直观图像，掌握典型系统薛定谔方程的求解，并结合实际体系学习理解固体能带、粒子统计分布、非平衡体系、开放体系与跃迁等相关知识，为后续课程的学习打下必备的基础。

第二节 牛顿力学与经典光学

量子物理是与经典物理相对应的。为了弄清楚什么是量子物理，我们首先需要回顾什么是经典物理。我们主要回顾一下经典力学（牛顿力学）和经典光学。

一、牛顿力学

1. 牛顿第二定律

大家中学时期都学过牛顿三大定律，其中核心的是牛顿第二定律，它说明了物体的运动状态是如何与外界环境关联的。如果我们只考虑一个质点，那么它的运动可以写作 $\mathbf{r}(t)$ ，即位置随时间的依赖关系。我们假设这个质点只能限制在一维运动，那么这个关系可以简化为 $x(t)$ 。

如果在某一个时刻 t_0 ，我们观察到质点位于 x_0 点，那么是否可以预测下一时刻 $t_0 + \delta t$ ，质点将处在什么位置呢？答案是肯定的，前提是我们得知道 t_0 时刻质点的速度 \dot{x}_0 ，这样的话当 δt 无限小的时候，我们有

$$x(t_0 + \delta t) = x_0 + \dot{x}_0 \delta t \quad (1)$$

但是我们能否进一步预测下下时刻 $t_0 + 2\delta t$ ，质点的位置呢？因为(1)式是不依赖于 t_0 的具体数值的，所以我们有 $x(t_0 + 2\delta t) = x(t_0 + \delta t) + \dot{x}(t_0 + \delta t) \delta t$ ，所以这等价于要求我们预测 $t_0 + \delta t$ 时刻质点的速度 $\dot{x}(t_0 + \delta t)$ 。这要求我们知道 t_0 时刻质点的加速度 \ddot{x}_0 ，从而利用

$$\dot{x}(t_0 + \delta t) = \dot{x}_0 + \ddot{x}_0 \delta t \quad (2)$$

得到 $\dot{x}(t_0 + \delta t)$ 。

类似的，如果我们希望进一步预测 $t_0 + 3\delta t$ 时刻质点的位置，我们需要知道 t_0 时刻质点的加速度 \ddot{x}_0 ；如果我们希望进一步预测 $t_0 + 4\delta t$ 时刻质点的位置，我们需要知道 t_0 时刻质点的加加速度；…… 从而无限循环下去。

牛顿第二定律的意义在于告诉我们，质点在任意时刻的加速度等于它收到的力除以它的质量，即

$$\ddot{x}(t) = F[x(t), \dot{x}(t)]/m \quad (3)$$

注意到这里的力一般是质点位置和速度的函数，例如带摩擦力的弹簧体系，滑块作为质点受到的力为 $F = -kx - \eta \dot{x}$ 。

由于这一关系，上述无限循环被打破了：我们只需要知道 t_0 时刻质点的位置 x_0 和速度 \dot{x}_0 ，并且知道了质点所受力的具体形式（即 $F(x, \dot{x})$ 的具体函数关系），就能利用牛顿第二定律预测质点在任意时刻的运动状态：由 x_0, \dot{x}_0 利用(3)得到 \ddot{x}_0 ，从而得到 $x(t_0 + \delta t), \dot{x}(t_0 + \delta t)$ ，继续利用(3)得到 $\ddot{x}(t_0 + \delta t)$ ，从而得到 $x(t_0 + 2\delta t), \dot{x}(t_0 + 2\delta t)$ ，以此类推。这告诉我们，对于单质点体系，任意时刻质点的位置和速度，完整描述了质点此前和此后的行为。我们称 $\{x, \dot{x}\}$ 为质点的运动状态。

上述关系对于多个质点组成的体系同样是成立的。假设我们知道了这些质点两两之间的相互作用力

$$\mathbf{F}_{ij}[\mathbf{r}_i(t), \dot{\mathbf{r}}_i(t); \mathbf{r}_j(t), \dot{\mathbf{r}}_j(t)]$$

以及质点受到外界的作用力

$$\mathbf{F}_i^{(0)}[\mathbf{r}_i(t), \dot{\mathbf{r}}_i(t)]$$

那么对于任意粒子 i ，我们有

$$\ddot{\mathbf{x}}_i(t) = \frac{1}{m_i} \{ \sum_{j \neq i} \mathbf{F}_{ij}[\mathbf{r}_i(t), \dot{\mathbf{r}}_i(t); \mathbf{r}_j(t), \dot{\mathbf{r}}_j(t)] + \mathbf{F}_i^{(0)}[\mathbf{r}_i(t), \dot{\mathbf{r}}_i(t)] \} \quad (4)$$

从而我们同样可以通过初始时刻 t_0 所有质点的运动状态集合 $\{\mathbf{r}_i(t_0), \dot{\mathbf{r}}_i(t_0)\}$ 得到后续任意时刻质点体系的运动状态 $\{\mathbf{r}_i(t), \dot{\mathbf{r}}_i(t)\}$ 。

这导致了牛顿力学完全确定性的世界观：由于世界总可以看作是由大量质点组成的，如果我们通过某种方式测量到了某一时刻世界上所有质点的运动状态，那么我们就可以利用牛顿第二定律算出来后续任意时刻世界上全部质点的运动状态，也可以反推出此前任意时刻世界上全部质点的运动状态。而世界上全部质点在任意时刻的运动状态，决定了这一时刻世界上所有的事件，包括股票的价格，每个人的心情和健康状态。所以原则上只要人类最终掌握了这种能够测量世界上所有质点运动状态的仪器，并且有一台足够强大的电脑去计算公式(4)，那么我们就可以预测未来，洞察历史。

2. 运动曲线

我们可以从图像的角度对上述内容进一步理解。以时间 t 作为横轴，位置 x 作为纵轴，质点的运动 $x(t)$ 可以表示为一条曲线，我们可以称之为运动曲线。对于给定的系统（即给定的 $F(x, \dot{x})$ 函数关系），如果我们知道了运动曲线上任意一点的坐标和斜率，就可以通过牛顿第二定律，逐点推演，画出整条曲线。

例如对于自由质点，它的 $F(x, \dot{x})$ 恒等于 0，所以加速度为 0，我们如果知道了初始的位置和速度，就可以逐渐推演出来它的运动曲线是一条直线；如果我们选取不同的初始位置和加速度，我们发现得到的曲线仍然是直线，只不过直线的斜率和截距依赖于初值的选择。再例如一个受到恒定力作用的质点，它的 $F(x, \dot{x})$ 为常数，加速度为常数，我们如果知道了初始的位置和速度，就可以逐渐推演出来它的运动曲线是一条抛物线；如果我们选取不同的初始位置和加速度，我们发现得到的曲线仍然是抛物线，只不过抛物线的位置依赖于初值的选择。作业 1 要求大家画一下一个受到摩擦力的弹簧滑块系统的运动曲线。

对于多质点体系，上述运动曲线的观点仍然成立，只不过这时纵轴不再是单一的 x 轴，而是包含了每个质点的位置集合 $\{\mathbf{r}_i\}$ 的一个高维空间的坐标轴。但是质点体系的运动依然对应于 $\{\mathbf{r}_i, t\}$ 空间中的一条曲线，并且对于给定的系统，通过运动曲线上任意一点的坐标和斜率 $\{\mathbf{r}_i(t_0), \dot{\mathbf{r}}_i(t_0)\}$ ，我们同样可以通过牛顿第二定律，逐点推演，画出整条曲线。

3. 最小作用原理

从微分方程的角度，牛顿第二定律给出的是质点系统的二阶常微分方程，其通解给出了系统运动曲线所属的曲线集合（例如自由质点的运动曲线属于直线，受到恒定力作用的质点的运动曲线属于抛物线）。对于某个质点，到底其运动曲线是这个集合中的哪条曲线，取决于其初始条件。换句话说，二阶常微分方程的解需要两个定解条件。初始位置和初始速度就是这

两个定解条件。

现在假设我们知道的并不是质点初始位置和初始速度，而是两个不同时刻（我们定义成起始和终止时刻）质点的位置，我们还可以得到系统的运动曲线吗？答案是肯定的，因为起始和终止时刻质点的位置同样可以作为两个定解条件，从而确定出具体的运动曲线。

从算法的角度，当定解条件为初始位置和初始速度时，我们可以利用前面阐述的方法，从运动曲线的初始点初步推演，得到整条曲线。但是当定解条件为起始和终止时刻质点的位置时，我们应该用什么方法去得到运动曲线呢？最小作用原理给出了一个美妙的回答。当给定起始和终止时刻质点的位置时，相当于固定了运动曲线的起点和终点。连接这两个点，我们可以作任意一条假想的运动曲线 C 。如果我们定义拉格朗日量为

$$L[x(t), \dot{x}(t)] = T - V = \frac{1}{2}m\dot{x}^2(t) - V[x(t)] \quad (5)$$

那么对于任意一条假想的运动曲线 C ， L 都是 t 的某个函数。我们把这个函数对 t 进行积分，或者说沿着这条假想的运动曲线对拉格朗日量积分，将得到一个量 S ，我们称为作用量：

$$S[C] = \int_{t_1}^{t_2} L[x(t), \dot{x}(t)] dt \quad (6)$$

注意到不同的 C 一般会给出不同的 S ，所以我们把 S 写作 C 的函数 $S[C]$ （由于 C 本身也是一个函数，而 S 又是 C 的函数，这种函数的函数叫做泛函）。

最小作用原理的内容是：当给定了起始和终止时刻（即 t_1 和 t_2 ）质点的位置时，我们可以作无数条假想的运动曲线 C ，在这无数条假想的 C 中，质点实际的运动曲线一定是使得 S 取稳定值（即极小值、极大值或者非极值驻点）的那一条（或者多条），用数学语言表达就是

$$\delta S = 0 \quad (7)$$

大多数情况下， S 是取极小值的，所以这个原理称为最小作用原理。

最小作用原理与牛顿第二定律是等价的。我们首先来计算任意曲线 C 处的 δS 值。假想对曲线 C 【表达式为 $x_c(t)$ 】进行一个小小的扰动 $\delta x(t)$ ，从而得到另外一条曲线 C' 【表达式为 $x_c(t) + \delta x(t)$ 】，那么这一变化导致的 S 变化为

$$\begin{aligned} \delta S &= S[x_c(t) + \delta x(t)] - S[x_c(t)] \\ &= \int_{t_1}^{t_2} \{L[x_c(t) + \delta x(t), \dot{x}_c(t) + \delta \dot{x}(t)] - L[x_c(t), \dot{x}_c(t)]\} dt \\ &= \int_{t_1}^{t_2} \left[\frac{\partial L}{\partial x} \delta x(t) + \frac{\partial L}{\partial \dot{x}} \delta \dot{x}(t) \right] dt = \int_{t_1}^{t_2} \left[\frac{\partial L}{\partial x} \delta x(t) - \frac{d}{dt} \frac{\partial L}{\partial \dot{x}} \delta x(t) \right] dt \\ &= \int_{t_1}^{t_2} \left(\frac{\partial L}{\partial x} - \frac{d}{dt} \frac{\partial L}{\partial \dot{x}} \right) \delta x(t) dt \end{aligned} \quad (8)$$

对于实际的运动曲线，最小作用原理要求任意的小扰动 $\delta x(t)$ 导致的 δS 都为零，这要求积分号里面 $\frac{\partial L}{\partial x} - \frac{d}{dt} \frac{\partial L}{\partial \dot{x}}$ 这一项为零，代入(5)式，我们得到

$$\frac{\partial L}{\partial x} - \frac{d}{dt} \frac{\partial L}{\partial \dot{x}} = -\frac{dV}{dx} - \frac{d}{dt} (m\dot{x}) = F - ma = 0 \quad (9)$$

这就是牛顿第二定律。

可以证明，对于多质点体系，我们同样可以通过定义适当的拉格朗日量，得到适用于多质点

体系的最小作用原理，并且它与多质点体系的牛顿第二定律形式等价。

牛顿第二定律与最小作用原理的关系类似于微分与积分的关系：牛顿第二定律是一种局部理论，它告诉我们如何从运动曲线上的一点逐步扩展到紧邻的其它点；最小作用原理是一种整体原理，它告诉我们对于给定的起点和终点，用哪个标准去选择实际的运动曲线。从牛顿第二定律，我们可以认为大自然是一台精密的机器，它不断地从一个时刻推算下一个时刻；而从最小作用原理，我们可以认为大自然是一个聪明的决策者，它总是要求质点选择“最经济”的运动方式，即作用量取极值的运动曲线。

二、经典光学

1. 几何光学：

高中光学的知识告诉我们，当光线经过两个折射率不同的介质的界面时（图 1-7），会发生发射和折射，并且反射角和折射角满足下述关系

$$\theta_{\text{反射}} = \theta_{\text{入射}}, n_1 \sin(\theta_{\text{入射}}) = n_2 \sin(\theta_{\text{折射}}) \quad (10)$$

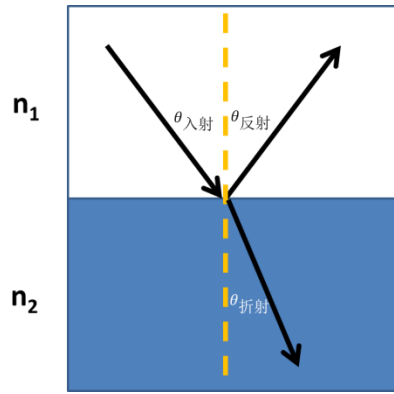


图 1-7 反射角与折射角的定义

一般地，如果光线走过的空间具有不均匀的折射率分布 $n(\mathbf{r})$ ，我们可以利用(10)式，逐点推演光线传播的轨迹。这类似于牛顿第二定律的操作方式。

法国科学家费马 1662 年提出，光线的传播规律还可以用另外一种方式表达：对于给定的起始位置 A 和终止位置 B ，光线总会走使得光程取极值的路径。具体来说，对于连接 AB 之间的任意一条路径 C ，我们定义其光程为

$$S[C] = cT[C] = c \int_{\text{沿着}C} dt = c \int_{\text{沿着}C} dl/v = \int_{\text{沿着}C} n dl \quad (11)$$

那么在这些任意路径 C 中，光线实际走过的路径一定是使得 S 取稳定值（即极小值、极大值或者非极值驻点）的那一条（或者多条），即

$$\delta S = 0 \quad (12)$$

类似于经典力学中最小作用原理与牛顿第二定律的关系，费马原理与反射折射定律也是等价的，或者说费马原理是反射折射定律的整体形式。我们可以较容易地证明这一点，例如对于图 1-8 所示的情况，光线从 A 点传播到 B 点，当我们在满足反射定律的光线基础上做一个

小的扰动，使得光的路径偏离实际的反射光线时，我们发现这一定会导致光程变长，因为 $APB=APB'$ ， $AP'B=AP'B'$ ，而 APB' 是一条直线，具有极短的距离，而反射光线一直在同一种介质中传播，所以路径极短意味着光程极短，从而 APB 是光程取极值的路径，从而对应着光线实际会走的路径。

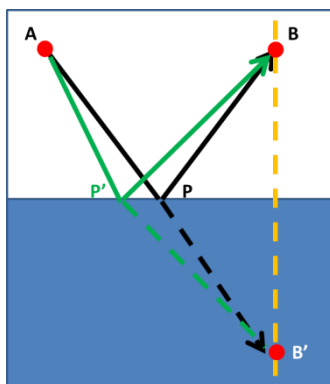


图 1-8 反射光线满足费马原理

2. 波动光学

然而费马原理自从提出来，就遭到了严厉的批评。1662 年（费马原理提出的同一年），著名光学专家、笛卡尔主义者 Clearselier 批评说：“The principle ... is **merely a moral principle** and not a physical one...”“when a ray of light must pass from a point in a rare medium to a point in a dense one, is there not reason for nature to **hesitate** if, by your principle, it must choose the straight line as soon as the bent one, since if the latter proves shorter in time, the former is shorter and simpler in length? **Who will decide and who will pronounce?**”

这一批评的核心元素在于，根据费马原理，光线会自动选择光程取极值的路径；但是如果光不事先走过所有的可能路径，它怎么会提前知道那一条路径是使得光程取极值的呢？

1678 年，惠更斯提出的波动光学原理解决了这一矛盾。根据惠更斯原理，光实际上是一种波动，光从 A 点传播到 B 点的过程，实际上是波的传播过程：例如如图 1-9，首先是 A 点（波源）的振动（当时还不知道这实际上是电磁场的振动）带动了其周围空间的振动，例如某一时刻这些振动传播到球面 α 上（振动同时传播到达的这些点，它们组成的面称为波前面，球面 α 就是一个波前面）， α 面这些被带动的点又成为新的次级波源，带动其周围空间的振动（如绿色圆圈所示），这一过程一直持续下去，这些新的次级波源叠加在一起，它们的包络构成了新的波前面 β 。

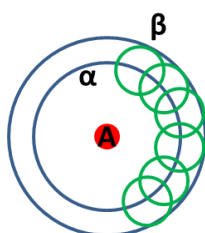


图 1-9 惠更斯原理：球面波的传播

从惠更斯原理也可以反射和折射定律：如图 1-10 所示，假设一个平面光源发射出一束平面波，当图最右侧的光线刚刚到达反射界面的时候，左边的三条光线早已到达界面并且已经在界面处激发出次级球面波。根据它们到达界面时间的早晚，次级球面波的半径线性变化，从而根据简单的几何关系，可知这些次级球面波的包络（图中土黄色虚线）与界面所成的角度（即 $\theta_{\text{出射}}$ ）与原平面波的包络（图中绿色虚线）与界面所成的角度（即 $\theta_{\text{入射}}$ ）相等，这就是反射定律。如图 1-11 所示，当界面处的次级波进入界面下方的介质时，由于不同介质中的光速不同，相应的平面波波前相比于入射波也要发生倾斜，使得折射角和入射角之间折射定律(10)。

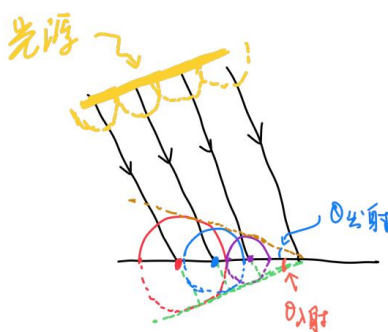


图 1-10 由惠更斯原理导出反射定律

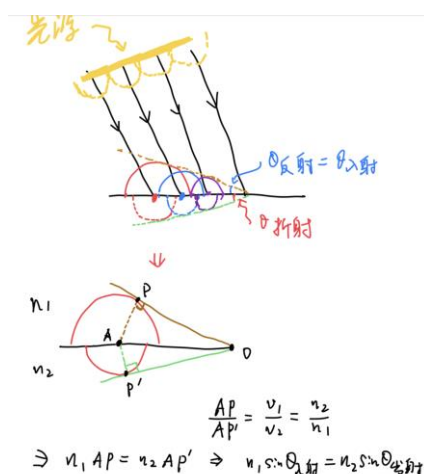


图 1-11 由惠更斯原理导出折射定律

惠更斯原理解决了费马原理逻辑上的问题：因为光是一种波动，所以光传播的过程实际上是波动传播的过程；波源的振动原则上是可以传播到空间所有位置的，而根据惠更斯原理，波前面上的每一点都可以看作一个次级波源，而这个次级波源的振动可以也传播到传播到空间所有位置。所以波动的传播，实际上是遍历了空间中所有可能的路径。光实际上**确实**走遍了所有可能的路径，并且选择了光程取极值的那一条（或者多条）！

从数学上来说，我们知道一束波矢为 k 的平面波可以写成 $e^{i(kx - \omega t)}$ 的形式，其中 $e^{-i\omega t}$ 项表示空间所有点都在以 ω 为原频率进行简谐运动，而 e^{ikx} 项表示如果一只蚂蚁沿着波的传播方向爬行，那么它每爬行 x 的距离，会发现该处的振动相位增加了 kx ，所以平面波在空间中的传播过程可以看作相位的积累过程。注意到在介质中， $k = nk_0$ （其中 k_0 为真空波矢），所以

假如平面波传播了距离 l ，那么它积累的相位为 $\Delta\varphi = nk_0 l$ 。现在假设光沿着任意一条起点为 A ，终点为 B 的曲线 C 传播，由于我们总可以把这条曲线的每一小段近似为直线，同时我们近似认为光在这些小段都是以平面波的形式传播的，那么当它从 A 点传播到 B 点的时候，总共积累的相位是每一小段积累相位的叠加，即 $\varphi = k_0 \int_{\text{沿着} C} n dl = k_0 S[C]$ ，其中 $S[C]$ 是(11)式中定义的光程。假设 A 点处，波的复振幅为 1，那么它沿着 C 传播到 B 点，将具有复振幅 $e^{i\varphi} = e^{ik_0 S[C]}$ 。根据惠更斯原理，由于光的传播是波动的传播，实际上光会走过连接 AB 所有可能的路径，每一条路径都贡献一个大小为 $e^{ik_0 S[C]}$ 的复振幅。由于波的叠加原理，这些不同路径的复振幅应该叠加起来，才是 B 点振动真正的复振幅，即

$$u_B \sim \sum_C e^{ik_0 S[C]}, \quad (13)$$

注意上式需要对所有可能的路径 C 求和。

我们注意到 $k_0 = \frac{2\pi}{\lambda_0}$ ，而可见光的波长一般在 500nm 左右。当光线传播的尺度远大于 500nm 时，一般来说，只要路径稍微变化一点点， $k_0 S[C]$ 就会发生非常巨大的变化。由于(13)是对 $e^{ik_0 S[C]}$ 的求和（把这一项理解成复平面上单位圆内的矢量），这导致一般来说，不同路径的矢量互相抵消（相消干涉），最终导致求和的结果为零。但是如果存在一条特殊的路径，它满足(12)式，即对这一路径稍微做一个小的扰动， S 依然保持不变，那么这条路径和它周围的路径都参与到(13)式的求和中，并且产生相长干涉，从而主导了(13)式求和的结果（这个原理称为驻定相位原理）。这条特殊的路径正是费马原理所要求的路径。通过把光的传播理解成一种波动，惠更斯原理给出了费马原理的一个更深层次的解释：**光确实遍历了所有可能的路径，但是由于波的叠加（干涉）特性，只有满足费马原理的光线才具有非零的贡献。**

通过观察(13)式，我们也可以得出在哪些情况下几何光学（费马原理）是不成立的。例如当我们通过某种方式限定了光传播的路径，使得(13)式的求和只有少数项时，驻定相位原理不再成立。一个极端的例子就是双缝干涉。再例如当光运动的尺度和波长在同一量级时，驻定相位原理也不再成立。

三、波粒二象性

事实上，与费马原理类似，经典力学中的最小作用原理也同样会遇到这样的逻辑问题：如果质点不事先走过所有的可能路径，它怎么会提前知道那一条路径是使得作用量取极值的呢？一个诱人的想法是，我们能不能完全类比惠更斯原理，认为质点实际上也是一种波，从而通过波的叠加解决这一问题呢？实际上，这一观点正是量子力学的核心思想之一，即所有的物质（粒子及粒子体系）都是以一种波的形式存在，物质的运动也是波的传播过程，即“**波粒二象性**”。这一思想最早由德布罗意提出，并且经过了薛定谔的完善，成为量子力学的标准形式之一。

从形式上，我们可以完全类比惠更斯原理，认为质点的运动对应的波由一个复振幅描述，由于波的传播特性，质点从运动曲线上的 A 点运动到 B 点的过程也就是这个波从 A 点传播到 B 点的过程；由于波的传播特性，每一点的“振动”都可以引起周围所有点的“振动”（惠更斯原理），这意味着在波从 A 点传播到 B 点的过程中，质点其实是走过了所有可能的路径的，只是由于波的叠加特性，对于宏观质点来说，只有满足最小作用原理的路径才具有非零的贡献。用数学去表述，就是

$$u_B \sim \sum_C e^{\frac{i}{\hbar} S[C]}, \quad (14)$$

(14)式称为费曼路径积分，即假设 A 点的复振幅为 1，那么 B 点的复振幅是所有路径贡献的相位因子的求和，这与(13)式的思想是完全一致的。唯一的不同在于，费马原理的光程 S 具有长度的量纲（参见式(11)），所以它需要除以一个量纲为长度的常量才能作为无量纲的相位，这个常量对应于光的波长；而最小作用原理的作用量具有能量*时间的量纲（参见式(6)），所以它需要除以一个量纲为能量*时间的常量才能作为无量纲的相位，这个常量实际上是有确定的数值的

$$\hbar = 1.054571817... \times 10^{-34} \text{J} \cdot \text{s}, \quad (15)$$

并且称为约化普朗克常数。量子物理中另外一个常用的常数叫做普朗克常数，记作 h ，它和约化普朗克常数之间具有如下关系

$$h = 2\pi\hbar. \quad (16)$$

为了进一步把费曼路径积分(14)与惠更斯原理(13)做比较，我们考虑一个自由质点的运动，它感受到的势能应该为零，所以拉格朗日量 $L = T - V = T = \frac{p^2}{2m}$ ，其中 p 是动量。根据牛顿第二定律，自由质点的运动是匀速直线运动，所以如果它沿着这条直线运动了时间 t ，对应的作用量为 $S = Lt = \frac{p^2 t}{2m} = p \frac{l}{2}$ ，其中 $l = vt = \frac{pt}{m}$ 是质点运动的距离。带入(14)式，我们得知这条经典路径将贡献一个形如 $e^{i\frac{pl}{\hbar 2}}$ 的相位因子，这与惠更斯原理在量纲上就一致了：一个依赖路径的、量纲为长度的量 $\frac{l}{2}$ 除以一个同样量纲为长度的量 $\frac{\hbar}{p}$ ，得到对应的相位。而这个量纲为长度的量 $\frac{\hbar}{p}$ ，应该与惠更斯原理类似，对应于（至少正比于）波长。

实际上，后面我们会看到，一个自由粒子的物质波是平面波，具有 $e^{i(kx - \omega t)}$ 的形式，并且平面波的波矢 k 与这个粒子的动量 p 之间具有关系

$$p = \hbar k \quad (17)$$

或者等价地，物质波的波长 λ 与粒子的动量之间具有关系

$$\lambda = \frac{2\pi\hbar}{p} = \frac{h}{p} \quad (18)$$

这与我们刚才的定性结论是一致的。

与此同时，自由粒子物质波的圆频率 ω 与粒子的能量之间满足关系

$$E = \hbar\omega \quad (19)$$

(17)和(19)式称为德布罗意关系。

我们去计算一下不同物质的物质波波长（也称为德布罗意波长）。例如我们假设有一个质量 1 克的物体，以 1 毫米每秒的速度运动，代入(18)式，可以算出它的物质波波长为 6.63×10^{-28} 米，我们知道原子的半径大约为 10^{-10} 米的量级，所以这个宏观物体的物质波波长小到不可思议的程度。这意味着(14)式的求和计算可以很好地使用驻定相位原理，即只有满足最小作用原理的路径才具有非零的贡献。这就是为什么宏观物体的运动规律一般符合经典力学，而不会体现出量子效应。

然而,对于微观粒子,情况将有很大的不同。例如我们考虑一个以 1 米每秒速度运动的质子,由于质子质量为 1.67×10^{-27} 千克,如果我们按照(18)式计算,它的物质波波长大约为 400nm。再例如考虑一个被 1V 电压加速的电子,其能量为 1eV,由于电子质量为 9.11×10^{-31} 千克,代入(18)式计算,它的物质波波长大约为 1.23nm。当前集成电路工艺节点下,晶体管的栅长大约为 10nm 量级,栅介质的厚度大约为 1nm 量级,而 Si 中电子的能量大约为几个 eV 量级。对于这些器件来说,电子的运动范围与物质波的波长大致在同一个量级,所以(14)式的求和计算不再可以用驻定相位原理,换句话说,我们应该考虑电子的波动特性,即量子效应。

通过电磁学的课程,我们知道光的波动其实对应于电磁场的振动,那么物质波到底对应于什么振动呢?或者说,物质波的复振幅到底对应什么物理量呢?下一节我们会详细介绍,这一复振幅对应(注意,不是等于)的是粒子出现在该点处的几率。

第三节 几率幅与量子物理的基本假设

一、经典几率与统计物理

按照牛顿第二定律,如果我们认为世间万物都是由质点组成的,那么一旦我们可以发明某种庞大的测量机器,测量出某一瞬间宇宙中所有质点的位置和速度,并且把两两质点之间的相互作用力带入到牛顿第二定律中,那么我们原则上就可以算出宇宙中所有质点在后续及之前任意时刻的位置(和速度)。这意味着我们可以完全预测宇宙的未来,并且知晓宇宙的过去。

然而这一观点被证明是不可行的。第一个阻碍在于,即使质点之间的相互作用形式非常简单,例如两两之间通过万有引力产生作用力,如果当我们考虑的质点数目大于 2 时,牛顿第二定律计算的难度也会超过人类的能力。作为一个极端的例子,即使只有区区 3 个质点,除了一些特殊情况以外,都会产生所谓的混沌现象: 3 个质点的初始位置和速度只要稍微改变一点点,都会在它们后续的运动中都会导致巨大的变化。由于对质点初始位置和速度的测量总会有误差,同时计算机的数值精度也一定是有限的,这导致我们实际上是无法通过牛顿第二定律计算这 3 个质点后续的运动状态的。

仅仅 3 个质点已经无法实际操作定量计算了,对于大量质点组成的宏观系统,我们更加无法精确预测每个质点的运动演化过程。但是,如果我们放弃精确预测每个质点的执念,而是只考虑一些宏观的物理量,那么我们可以引入几率的观点去处理问题,得到很多相当精确的结果,这就是统计物理的方法。我们假设一个宏观物体是由 N 个质点组成的(N 非常大),如果精确描述这些质点的运动状态,我们需要 $6N$ 个参量,即每个质点的位置及每个粒子的动量(速度): $\{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N; \mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_N\}$, 而质点体系的运动可以看作是这些参量随时间的演化, 即 $\{\mathbf{r}_1(t), \mathbf{r}_2(t), \dots, \mathbf{r}_N(t); \mathbf{p}_1(t), \mathbf{p}_2(t), \dots, \mathbf{p}_N(t)\}$ (注意这等价于运动曲线 $\{\mathbf{r}_1(t), \mathbf{r}_2(t), \dots, \mathbf{r}_N(t)\}$, 只不过把动量随时间的依赖关系也加进去了)。由于质点数如此之多,并且质点与质点通过两两相互作用导致的运动是如此复杂而迅速,我们假设平衡状态下,质点体系在短时间内遍历了所有可能的 $\{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N; \mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_N\}$ 取值组合, 只不过在不同的 $\{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N; \mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_N\}$ 取值处,质点体系停留的时间(或者说经历的次数)不同,所以体系在不同的 $\{\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N; \mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_N\}$ 处具有不同的几率密度,我们用

$$w(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N; \mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_N) \quad (20)$$

表示。我们在平时的测量中,只会与这个宏观物体的宏观状态例如体积、能量等等打交道,