

# 高二物理竞赛

陈博



# 目录

<b>第一章 运动学</b>	<b>7</b>
1.1 时空与物质 . . . . .	7
1.1.1 时空观,坐标系 . . . . .	7
1.1.2 物质 . . . . .	9
1.1.3 参考系,物理规律与其不变性 . . . . .	11
1.2 运动的描述 . . . . .	12
1.2.1 质点的运动 . . . . .	12
1.3 参考系变换 . . . . .	13
1.4 运动的牵连 . . . . .	13
<b>第二章 动力学</b>	<b>15</b>
2.1 牛顿定律 . . . . .	15
2.2 动量定律 . . . . .	15
2.3 能量定律 . . . . .	15
2.4 角动量定律 . . . . .	15
2.5 位力定律 *	15
2.6 动力学问题 . . . . .	15
2.7 碰撞问题 . . . . .	15
<b>第三章 静力学</b>	<b>17</b>
3.1 约束 . . . . .	17
3.2 力系化简 . . . . .	17
3.3 平衡问题 . . . . .	17
3.4 虚功原理 . . . . .	17
3.5 平衡态稳定性 . . . . .	17
<b>第四章 振动与波</b>	<b>19</b>
4.1 方程与谐振 . . . . .	19
4.1.1 简谐振动的定义 . . . . .	19
4.1.2 简谐振动的性质 . . . . .	19
4.2 阻尼振动与受迫振动 . . . . .	20
4.3 多自由度小振动 *	20
4.4 非线性摄动 . . . . .	20
4.5 格波 . . . . .	20
4.6 波动方程 . . . . .	20
4.7 波的色散 . . . . .	20

<b>第五章 万有引力</b>	<b>21</b>
5.1 有心力下运动 . . . . .	21
5.2 万有引力下运动 . . . . .	21
5.3 二体与潮汐 . . . . .	21
<b>第六章 刚体</b>	<b>23</b>
6.1 刚体的物理描述 . . . . .	23
6.2 平面平行运动 . . . . .	25
6.3 空间刚体运动 * . . . . .	25
<b>第七章 流体</b>	<b>27</b>
7.1 流体的物理描述 . . . . .	27
7.2 定常流动动力学 . . . . .	27
7.3 黏滞流体动力学 . . . . .	27
<b>第八章 弹性体</b>	<b>29</b>
8.1 弹性体的物理描述 . . . . .	29
8.2 弹性棒, 弹性膜与弹性体 . . . . .	29
8.3 弹性波 . . . . .	29
<b>第九章 相对论</b>	<b>31</b>
9.1 相对论运动学 . . . . .	31
9.2 相对论动力学 . . . . .	31
9.3 相对论连续物质 . . . . .	31
9.4 相对论电磁场 . . . . .	31
<b>第十章 热力学第一定律</b>	<b>33</b>
10.1 热力学第零定律 . . . . .	33
10.2 热力学第一定律 . . . . .	34
10.2.1 准静态与非准静态过程 . . . . .	34
10.2.2 内能 . . . . .	35
10.2.3 功 . . . . .	36
10.2.4 热量 . . . . .	36
10.2.5 耗散 . . . . .	37
10.3 理想气体 . . . . .	38
10.3.1 理想气体的定义 . . . . .	38
10.3.2 理想气体物态方程 . . . . .	39
10.3.3 混合理想气体 . . . . .	41
10.3.4 理想气体的过程 . . . . .	42
10.4 开放系统的理想气体 . . . . .	46
10.4.1 静态平衡问题——重力场中的大气 . . . . .	46
10.4.2 能量守恒——伯努利方程 . . . . .	47
10.4.3 动量守恒——欧拉方程 . . . . .	48

第十一章 热力学第二定律	51
11.1 循环过程	51
11.1.1 热机与热泵	51
11.1.2 热机循环	54
11.2 理想气体的熵	56
11.3 热力学第二定律	58
11.4 熵的计算	61
11.4.1 理想气体的熵	61
11.4.2 固定热容固体熵	62
11.4.3 大热容恒温热库	62
11.4.4 传热熵	62
11.4.5 混合熵	62
11.5 热力学函数与其特性	64
11.5.1 四个热力学函数与四个状态参量	64
11.5.2 若干定理的证明	65
11.5.3 自由能的含义	70
11.5.4 化学势	71
11.6 近平衡态热力学 *	72
11.6.1 线性输运现象	72
11.6.2 传热的熵产生	75
11.6.3 普遍理论与昂萨格倒易关系	75
11.6.4 推证热电耦合的普遍规律	76
第十二章 液体与固体的性质摘要	79
12.1 固体晶格论	79
12.2 固体电子论 *	80
12.3 液体的彻体性质	81
12.3.1 液体性质综述与其微观成因	81
12.3.2 热容	81
12.4 液体的表面性质	82
12.5 极端条件下的其他物态	82
第十三章 相与相变摘要	85
13.1 相平衡	85
13.2 气液相变	85
13.3 连续相变	86
13.4 拓扑相变	86
第十四章 统计物理基础摘要	87
14.1 数学基础	87
14.1.1 概率与独立性	87
14.1.2 随机变量及其数字特征	88
14.1.3 信息熵	89
14.2 统计假设	93
14.3 麦克斯韦分布律	94

14.4 能均分定理 . . . . .	94
14.5 功, 热, 熵 . . . . .	94
14.6 量子与相对论 . . . . .	94
<b>第十五章 光波与光线</b>	<b>95</b>
15.1 界面上的反射与折射 . . . . .	95
15.1.1 光波与光线 . . . . .	95
15.1.2 菲涅尔公式 . . . . .	97
15.2 光线方程 . . . . .	101
15.2.1 光线方程与折射定律 . . . . .	101
15.2.2 光力类比 . . . . .	102
15.3 费马原理 . . . . .	103
<b>第十六章 光学成像</b>	<b>105</b>
16.1 傍轴光成像 . . . . .	105
16.1.1 物与像 . . . . .	105
16.1.2 球对称成像系统与符号法则 . . . . .	106
16.1.3 光具组成像 . . . . .	108
16.2 理想成像系统 . . . . .	109
16.2.1 作图法 . . . . .	110
16.2.2 基点基面性质 . . . . .	111
16.2.3 实例与望远系统 . . . . .	112
16.3 更多讨论 *	114
16.3.1 理想成像本质 . . . . .	114
16.4 非傍轴成像 . . . . .	116
<b>第十七章 光学仪器知识摘要</b>	<b>117</b>
17.1 光度学 . . . . .	117
17.1.1 色视觉 . . . . .	117
17.2 光阑与光瞳 . . . . .	118
17.3 眼睛 . . . . .	118
17.4 显微镜 . . . . .	118
17.5 望远镜 . . . . .	118
17.6 照相机 . . . . .	118
<b>第十八章 静电学</b>	<b>119</b>
18.1 电荷与电场 . . . . .	119
18.1.1 电磁相互作用与电荷 . . . . .	119
18.1.2 库仑定律 . . . . .	120
18.1.3 电场 . . . . .	120
18.2 两个定律与电势 . . . . .	123
18.2.1 电场的高斯定律 . . . . .	124
18.2.2 电势与电场的环路定律 . . . . .	125
18.2.3 总结 . . . . .	126
18.3 静电能 . . . . .	127
18.3.1 静电势能 . . . . .	127

18.3.2 自能与相互作用能 . . . . .	128
18.3.3 电场能 . . . . .	130
18.4 电荷体系 . . . . .	131
18.4.1 电偶极子 . . . . .	131
18.4.2 电荷密度 . . . . .	134
18.4.3 极化强度 . . . . .	135
<b>第十九章 导体与介质</b>	<b>137</b>
19.1 导体与静电平衡 . . . . .	137
19.1.1 导体特点 . . . . .	137
19.2 电像法 . . . . .	137
19.3 电介质 . . . . .	137
19.4 再议静电能 . . . . .	137
<b>第二十章 稳恒电流</b>	<b>139</b>
20.1 稳恒电流描述与形成 . . . . .	139
20.1.1 德鲁德模型 . . . . .	139
20.1.2 费米气观点 * . . . . .	140
20.1.3 能带论 * . . . . .	142
20.1.4 惯性, 阻尼与回复力 . . . . .	144
20.1.5 稳恒电流与形成条件 . . . . .	145
20.2 电路与电路方程 . . . . .	147
20.3 电路分析基础 . . . . .	147
20.4 电路分析方法 . . . . .	147
20.5 半导体 . . . . .	147
<b>第二十一章 静磁场</b>	<b>149</b>
21.1 电流与磁场 . . . . .	149
21.2 两个定理与矢势 . . . . .	149
21.3 磁偶极子 . . . . .	149
21.4 磁介质与磁能 . . . . .	149
21.5 超导简介 . . . . .	149
<b>第二十二章 电磁感应</b>	<b>151</b>
22.1 动生电动势 . . . . .	151
22.2 感生电动势 . . . . .	151
22.3 自感与互感 . . . . .	151
22.4 法拉第电磁感应定律 . . . . .	151
<b>第二十三章 麦克斯韦方程组</b>	<b>153</b>
23.1 麦克斯韦方程组 . . . . .	153
23.2 平面电磁波 . . . . .	153
23.3 电磁场能量与动量 . . . . .	153
23.4 电磁波辐射 . . . . .	153

第二十四章 交流电路	155
24.1 代数与几何表示	155
24.2 常见电路	155
24.3 变压器	155
24.4 电能传输	155
第二十五章 光的干涉	157
25.1 标量波理论	157
25.2 分波面干涉	161
25.2.1 杨氏双缝干涉仪	161
25.3 分振幅干涉	161
25.4 偏振干涉	161
25.5 相干性	161
25.6 多光束干涉	161
第二十六章 光的衍射	163
26.1 光栅与波带片	163
26.2 布拉格衍射	163
26.3 衍射积分公式	163
26.4 波前分析法	163
第二十七章 物理光学	165
27.1 经典色散理论	165
27.1.1 复波矢与复折射率	165
27.1.2 经典电子论的解释	168
27.2 色散, 散射与吸收	170
27.3 群速与展宽	170
27.4 光量子	170
第二十八章 量子论	171
28.1 黑体辐射	171
28.2 光粒子性	171
28.3 玻尔原子	171
28.4 电子波动性	171
28.5 物质波与波函数	171
第二十九章 物理学尺度	173
29.1 宇观	173
29.2 宏观	173
29.3 介观	173
29.4 微观原子	173
29.5 微观亚原子	173

# 第一章 运动学

## 1.1 时空与物质

物理学,从刚开始成为实验性的科学的伽利略时期开始,到近半个世纪年来理论物理学家对额外维度的探讨,都给予了时空(spacetime)最核心的地位.牛顿的理论,分析力学,经典场论,相对论这些理论最基本的图像都是时空与物质(matter)的分立性.时空是装备了一个能体现出物理物理意义的度量(metric)的3+1维对象<sup>1</sup>.而物质是在每个时空点处的某种结构.下面会分别介绍不同物质体系的描述方法.在这之前我们来看看不同的时空观:

### 1.1.1 时空观,坐标系

牛顿力学理论体系基于伽利略时空观(Galilean spacetime picture),也就是绝对时空观(absolute spacetime picture).在这里空间是三维的平直空间.设想一个人站在该空间的某个空间点,他的胸前,头顶和右手平举的三个方向就是相互垂直的方向.如果以自己的臂长为标准长度,他能够定义空间中每两个点之间的空间间隔.事实上,这个人能够以一种正确的方式为每个空间点定义一个坐标,那么所有三维空间点的集合与任意两个点之间的距离为:

$$A(x, y, z) \in \mathbb{R}^3 ; \quad x, y, z \in \mathbb{R}$$

$$A(x_1, y_1, z_1), B(x_2, y_2, z_2) : l^2 = \overline{AB}^2 = (x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2$$

因为空间不同于时间,我们为以上数字赋予特殊的物理含义,也就是加上量纲(dimension).如果两点坐标为(0,0,0)和(1,1,1),那么 $l = \sqrt{3}\text{m}$ ,而不是 $l = \sqrt{3}$ .其单位m一方面表示了这个物理量的属性,另一方面指定了某个实际物理体系确定下来的固有长度大小.现行的(2018年1月1日,下文同)国际单位制对1m的定义如下<sup>2</sup>:

1m是光在1/299792458 s内在真空中行进的距离.

这就是我们的三维平直空间(3-dimensional flat space).注意空间点具有物理实际意义,它可以脱离坐标系而单独存在.事实上坐标系的原点可以建立在空间中的任意点处,朝向也可以是任意方向,两个空间点之间的距离 $l$ 不会依赖于坐标系的选取,但两个点的坐标会因坐标系不同而改变.如果我们选取的坐标系总是下述使得两个相距很近的点之间的微元距离公式成立:

$$dl^2 = dx^2 + dy^2 + dz^2$$

那么建立的坐标系就是一个笛卡尔坐标系(Cartesian coordinate system),即空间直角坐标系(3D orthogonal coordinate system).但是以空间直角坐标系为基础,我们又经常建立球坐标(spherical coordinate)和柱坐标(cylindrical coordinate)系统.通常取 $x$ 轴为幅轴(azimuth axis),点在 $x-y$ 平面上的投影与原点的连线相对 $x$ 轴转过的角度为 $\varphi$ ,即幅角(azimuth angle).而 $z$ 轴为极轴(polar axis),而点与原点连线与极轴的夹角 $\theta$ 为极角(polar angle).天文观测用球坐标就很方便,它是以描述的空间点到原点之间的距离 $r$ ,也

<sup>1</sup>数学上有一套严格的说法,把这种连续,光滑的四维对象称为伪黎曼流形(pseudo-Riemannian manifold).

<sup>2</sup>一方面,它依赖于狭义相对论的正确性,目前极少理论物理工作者会质疑它.另一方面,应该要先定义下文的1s,再来定义1m.

称矢径(radius), 和两个描述角位置的极角幅角来构成三个坐标  $(r, \theta, \varphi)$  的. 而理论物理里也常用到的柱坐标是以  $(\rho, \varphi, z)$  为描述空间点的坐标,  $\rho$  是空间点到  $z$  轴的距离.

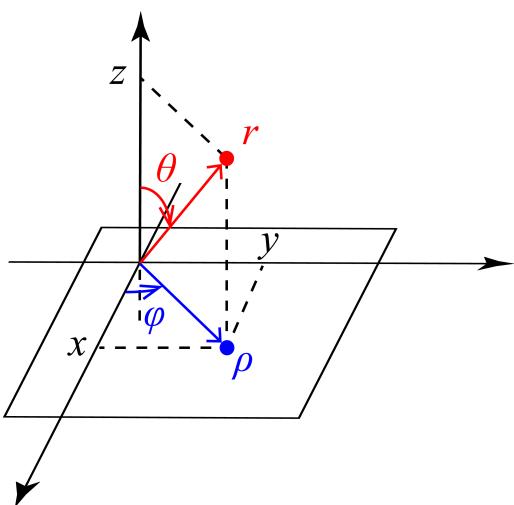


图 1.1: 三种坐标

需要注意, 我们对以上  $\theta, \varphi$  角度定义比较微妙,  $\theta$  一般限制其在  $[0, \pi]$  范围内, 在端点  $\theta = 0, \pi$  处  $\varphi$  的不同取值代表同一个空间点. 但  $\varphi$  的取值我们不加以任何限制, 也就是原则上  $\varphi \in \mathbb{R}$ . 而任意  $\varphi$  相差  $2n\pi$  的坐标实际上代表同一个空间点. 这么做是因为我们最好不要用静态的几何观点去看待角度这一概念. 而是用运动, 变换的观点去看待角度. 作为连续的运动的点的幅角变化也应是连续的, 为了使点绕  $z$  轴一圈后幅角不至于突然改变, 就必须默认每一点的幅角可以有多种取值, 在具体的运动中应该灵活选择其具体取值大小.

之后经常会说到各种对称性, 在本系列教材中我们采取如下说法: **球对称**(spherical symmetric) 仅仅代表某个函数  $f(r, \theta, \varphi)$  与  $\varphi$  无关. 而**柱对称**(cylindrical symmetric) 代表的是  $f(\rho, \varphi, z)$  与  $z$  无关. 与  $\varphi$  和  $\theta$  都无关的

$f(r, \theta, \varphi) = f(r, \forall\theta, \forall\varphi)$  被称为**各向同性**(isotropic). 最后**中心对称**(centrosymmetric) 是一个很弱的对称性, 它仅仅代表函数在**中心反演**(space inversion) 下的对称性:

$$f(x, y, z) = f(-x, -y, -z)$$

绝对时空观中的时间则是一种完全与空间独立的属性. 实际上, 每一个空间点处的人都能感受到时间的流逝. 十分抽象的把这些不同的时刻画在一根时间轴上, 便是一维的时间坐标:

$$A(t) \in \mathbb{R} ; t \in \mathbb{R}$$

而时空又是两个独立的概念. 任意一个坐标点处都有时间轴, 而任意一个时刻都有一个三维空间切片. 事实上, 我们的时空是一个 3+1 维的结构, 合理地选取坐标后, 实际上可以把时空结构写成四维坐标:

$$A(x, y, z, t) \in \mathbb{R}^4 ; x, y, z, t \in \mathbb{R}$$

而对于任意两个时空点, 存在绝对的时间, 也就是可以找到两个事件的绝对时间差:

$$\tau = |t_2 - t_1|$$

但空间距离却具有相对性. 我们要求在同一时刻的同一空间切片上定义空间的度量:

$$t_1 = t_2 : l = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$$

不同时刻的空间, 被线性地组织在一起, 保证每一刻都平直的空间既不会随时间膨胀收缩, 也不会旋转加速. 这种时空结构被叫做**牛顿-嘉当几何**(Newton-Cartan geometry)<sup>3</sup>.

时间是一个新的量纲, 其国际单位制对 1s 的定义为:

1s 是铯-133 原子基态的两个超精细能级之间跃迁所对应的辐射周期时长的 9192631770 倍.

牛顿-嘉当几何的显著特点是物理量的定义没有所谓的**协变性**(covariance). 固然, 在三维空间坐标框架的旋转下, 任何物理过程中的事件所发生的位置坐标发生旋转变换, 而几乎所有标量(比如质量, 能量)都不

<sup>3</sup>用更为严格的语言来说, 我们需要一个装备了一阶形式“钟” $c_\nu$  和二阶逆半正定对称张量类空间度规  $s^{\mu\nu}$  的四维微分流形  $\mathcal{M}$ :

$$\mathcal{M} : s^{\mu\nu} c_\nu = 0$$

变,几乎所有矢量(比如动量,角动量)都按照类似坐标变换的法则发生变换.但是,对于更普遍的一列伽利略变换(Galilean transformation),也就是相对匀速运动的参考系之间的变换,时空坐标的形式却不能简单地推广到能量动量角动量上.

不同于绝对时空观中时间与空间成为相互独立的量纲的特点,狭义相对论改变了对基本物理量的看法.狭义相对论把时空看成为可以相互转化的不可分割的新的3+1维整体,同时性已然破缺.两个时空点之间无法定义绝对的时间间隔.在相对论时空观(relativistic spacetime picture)下,时间和空间可以用同一把尺子去丈量:这是由于光速的不变性:

$$c = 299792458 \text{ m/s}$$

而量出来的长度叫做时空间隔(spacetime interval):

$$ds^2 = c^2 dt^2 - dx^2 - dy^2 - dz^2$$

这赋予时空以截然不同的结构,最关键的一点是,时间的绝对性被取消了,不同事件发生时间先后的比较不总是可行,相对论的有限速度因果律在这里取代了经典观点的时序因果律.我们将在本书最后介绍相对论理论.

### 1.1.2 物质

时空是物理过程发生的舞台.那么将物质引入后舞台上所演出的便是事件(event).事件这个物理概念由这样的数学工具来描述:它的第一个要素自然是时空坐标( $r, t$ ),是事件在时空上的外化.第二个要素是事件内禀的属性.它取决于所引入的物质种类,还取决于我们所关心问题的层次.一般用标量,矢量,乃至张量这样的数学工具来描述它.举例,电磁场物质在经典物理中用电场磁场来描述,但在量子力学(quantum mechanics, QM)中这不够了,需要用矢势和标势来描述才是完整的.在更深的量子场论(quantum field theory)中甚至这也是不够的.还需要量子化为光子才合适.又比如,电子这个很有意思的现象,最简单的模型是质点模型,内禀的属性是质量,动量与能量<sup>4</sup>.然而与电磁场的经典相互作用强度告诉我们还有一项内禀属性叫做电荷量.近代人们惊奇地发现原来电子还固有磁矩,也就是电子的自旋.最后狄拉克等人发展出量子电动力学(quantum electrodynamics, QED),统一地用一个四分量的旋量波函数和它的方程中的若干参数来完整地描述所有发现的电子内禀属性.

经典物理学<sup>5</sup>(classical physics)中涉及到的物质主要有:

#### 1. 质点 (point mass, particle)

不得不承认质点是牛顿力学的根基,一切可观的结论的出发点.质点所对应的事件集合为时空中的一条世界线(world line).每一个时间仅仅有可能只有一个事件发生.实际上质点的运动用运动学方程(kinematic equation)来描述:

$$\mathbf{r} = \mathbf{r}(t) = (x(t), y(t), z(t))$$

而质量(mass)是质点必要的内禀属性.它将作为参数出现在下一节介绍的动力学方程中.它反应物质受到同样大小的相互作用运动状态改变的难易程度.国际单位制对1kg的定义如下:

1kg是保存在法国巴黎布勒特伊宫的国际计量局实验室的约47立方厘米立式铂铱合金小圆柱的质量.当然,出于实用考虑,也是很多它的官方复制体的质量.

这个定义从1889年至今已经有一百多年了,历史远长于其他六个国际基本单位.2018年末将有望修改这一古老的规定方式,新的定义<sup>6</sup>为:

<sup>4</sup>不要认为动量能量不是内禀的而是由时空运动所决定的,读者可以思考在水中一个气泡的上升,它为体系带来的动量方向如何?

<sup>5</sup>本书的经典物理学指牛顿力学,分析力学,经典电磁学,经典统计力学与狭义相对论的范畴.

<sup>6</sup>一方面,它依赖于狭义相对论和量子力学的正确性,目前极少理论物理工作者会质疑它.另一方面,应该要先定义上文的1s和1m,再来定义1kg.

1kg 被这样定义: 取普朗克常数的固定数值在单位  $\text{kg} \cdot \text{m}^2 \cdot \text{s}^{-1}$  下为  $6.62607015 \times 10^{-34}$ .

长度、时间和质量为三大力学量纲, 量纲是物理量的属性, 物理量的表示方法为数值加单位, 每个量纲都有自己独特的一套单位, 不同单位间差一个纯数的倍率. 在物理量的加减时量纲必须相同而且结果保持量纲不变. 但不同量纲物理量可以进行乘除而生成新的量纲. 除了简单的加减乘除的以上规则以外, 其他特殊函数必须只能作用在无量纲的纯数上. 这叫做**量纲法则**(dimensional rules).

## 2. 质点系 (particle system)

质点系是对质点的一种自然扩展. 讨论质点的运动时质点受到的相互作用来自何方? 在牛顿力学的框架下必然来自施力物体. 故把两个质点作为整个体系讨论, 把相互作用区分为内力与外力而推广之前的结论, 在这个过程中我们能体会到哪些结论是一脉相承的普遍规律而哪些需要重新审视与修改. 从原理上质点系不过是多个质点同时存在的情况而已. 而经典统计力学是这种观点的提炼与延伸, 它着眼于一个宏观大分子数的体系的长时间平均下的行为, 理论上常常结合概率论的做法, 对等概率的体系代表点构成的系统做平均. 从中提取统计力学下不平凡的物理量.

## 3. 连续介质 (continuous media)

一方面, 连续介质是从纯粹的理论过渡到实际问题的至关重要一步. 另一方面, 在处理连续介质的问题中形成了场论, 它作为一种与质点截然不同的物理图像形成了自己的一整套理论. 更重要的, 量子场论认为由于不确定原理, 质点与世界线的观点实际上是作为场的传播的一种近似. 也就是说, 场论相比质点的力学更具有兼容性和普适性. 对于简单的连续介质, 内禀的属性由每一个点处的质量密度  $\rho$  和速度  $v$  描述, 这实际上形成了一个标量场和矢量场:

$$\rho = \rho(\mathbf{r}, t) ; \quad \mathbf{v} = \mathbf{v}(\mathbf{r}, t)$$

而动力学方程的列法又强烈依赖于介质内的相互作用. 于是这个大的问题又分出很有代表性的弹性力学和流体力学, 还有介于两者之间的粘弹性塑性模型等等. 还有从热力学平衡态出发考虑的近平衡态统计力学方法.

## 4. 场 (field)

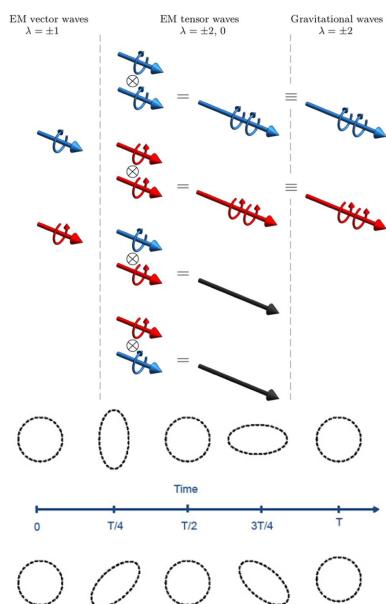


图 1.2: 引力波与电磁波的区别

关于场的理解历史上经历了两个过程. 第一个阶段是正如上面的连续介质的做法, 场是作为某种事先已经被研究的足够清楚的粒子体系与它们之间的相互作用的描述方法与数学工具. 而从牛顿时代开始人们就默认了超距作用的存在, 而场数学工具的第二个作用就是描述电荷之间, 质量之间的超距作用. 但是逐渐人们意识到相互作用传播速度的有限性, 于是场开始变成某种与物质其他耦合的物质而单独存在, 这就是第二阶段, 人们意识到了场本身就是一种物质, 电子与电子之间不能直接相互作用, 电子改变了真空, 稳定地与电磁场相互作用而产生库仑场, 然后在对第二个电子产生相互作用力. 将场作为某种先验地存在于时空中的物质形式进行描述与研究的学科就叫做**场论**(field theory). 有哪些可能的场形式? 场的物理量如何书写? 场与粒子, 场与场之间如何相互作用便是场论的研究内容. 比如说, 动力学所关心的内容, 场的质量本身没有定义, 可以定义的一般是场的定域动量与波矢, 能量与角频率, 平面波的色散关系等等. 的确, 场论最基本的做法是傅里叶分解, 它认为平面波是不受任何相互作用时的场的惯性运动方式, 而任何场的运动, 每一时刻都可以分解成一系列平面波的叠加. 就好像任何复杂体系的运动本质上都是质点系的运动, 而质点的运动每时每刻都可以看出匀速运动那样.

色散关系等等. 的确, 场论最基本的做法是傅里叶分解, 它认为平面波是不受任何相互作用时的场的惯性运动方式, 而任何场的运动, 每一时刻都可以分解成一系列平面波的叠加. 就好像任何复杂体系的运动本质上都是质点系的运动, 而质点的运动每时每刻都可以看出匀速运动那样.

场论根据其描述方法分为标量场论,矢量场论,张量场论和旋量场论等. 电磁场论是矢量场论,它用四势矢量  $(\varphi, \mathbf{A})$  来描述. 引力场论在广义相对论弱场近似下是张量场论,但传播的引力波和电磁波类似只能有两种偏振,其模式为 + 型和  $\times$  型,这又是不同于电磁波的. 在更现代的观点中,场是真空的某种激发,十分类似于一张二维的拉紧的薄膜,没有任何振动时它已经固有能量了,场物质反应为膜的各种振动形式. 而一个石子受到外力而压弯了膜,就可以理解为粒子与场之间的相互作用.

### 1.1.3 参考系,物理规律与其不变性

**坐标系**(coordinate system) 是一套用来描述时空点位置的坐标系统,而**参考系**(reference system) 则是下文将引申的一类坐标系的统称.

在伽利略时空观<sup>7</sup>下参考系的定义相对简单. 我们只需要任意指定一个观察者,观察者可以在  $3+1$  维的时空中作任意运动,每一个时刻对应一个测量的坐标系统: 观察者分别以右手边,胸前和头顶三个方向建立坐标系,坐标系的刻度就是其代表的位置到原点观察者处的空间距离. 这样便可以表示任意事件的时空坐标. 而且构成一个三维空间直角坐标系. 叫做观察者的**固有参考系**(proper reference system).

物理规律的本质是什么? 在一个特定的参考系中,我们进行一些物理量的测量,测量得到的值之间在特定的实验条件下会满足特定的关系而不依赖于某些具体的参数,一般都是等式. 这样的普遍成立的等式就可以反映某些物理规律. 从**动理学**(kinetics) 的角度理解,体系的演化是**状态**(state) 随时间的演化,而体系的状态由一组完备的**态参量**(state variable) 描述,态参量可以包含一些广义坐标  $q_\alpha$  和他们对时间的导数即广义速度  $v_\alpha = \dot{q}_\alpha$ . 而以下**决定论**(determinism) 则对物理规律的形式做出了规定:

动理学系统:

已知某时刻  $t=0$  的初态  $\{q_\alpha(0), \dot{q}_\alpha(0)\}$  后,应能预言任意将来的状态  $\{q_\alpha(t), \dot{q}_\alpha(t)\}$ .

事实上,不光是将来的状态,而且在一般经典物理的框架下,过去的状态也是可以被“预言”的. 这是因为我们进一步要求动理学系统的规律是以下形式的二阶方程:

$$f_i(\ddot{q}_\alpha, \dot{q}_\alpha, q_\alpha, t) = 0$$

而且广义坐标的个数应该和方程的数目一致,方程是至多二阶的. 微分方程的描述在数学上完全导致了体系彻底的决定性,只要知道初始条件便可以往前往后推理所有发生的物理现象<sup>8</sup>.

这一物理学科研究范围内的原理很强大,强大到历史上很多物理学家和哲学家都质疑过它的真实性和适用性. 最著名的莫过于拉普拉斯提出的**拉普拉斯妖**(Laplace's demon) 的论点: 如果存在一个全知的拉普拉斯妖,能够知道在某一时刻宇宙中所有原子的精确位置与动量,那么任意时刻的状态也能被预言. 这种决定论观点总是被拿来与**自由意志论**(free will) 去归谬,如果过去与未来都已经被谱写,那么人类所有的行为的意义和动机也就不能被解释. 这直到今天也是科学与哲学未能获得满意的回答的问题.

但是科学的一个目的就是在于能够对未发生的事情进行预言,所以本着对决定论的认可与尊重,我们才能继续我们的科学的研究.

下一个伽利略相对性原理.

<sup>7</sup>本章仅限讨论非相对论,相对论见后

<sup>8</sup>更有甚者,经典物理也难以解释耗散现象,在忽略耗散时,方程应该还具有平移对称性(不显含  $t$ ) 和时间反演对称性(方程对广义速度是偶函数). 此时对将来和对过去的预言在广义速度取相反数时是完全对称的,而且预言只依赖于时间差不依赖于初始时刻.

## 1.2 运动的描述

### 1.2.1 质点的运动

质点的运动是复杂的连续体的运动的基础. 所谓运动(motion) 就是每一个时间都有一个位置. 就是说位置是时间的函数:

$$\mathbf{r} = \mathbf{r}(t)$$

我们使用实数上的矢量空间和矢量来刻画真实的物理学上的二维或三维空间和其中的点. 也即:

$$\mathbf{r} \in \mathbb{R}^2 \text{ or } \mathbb{R}^3$$

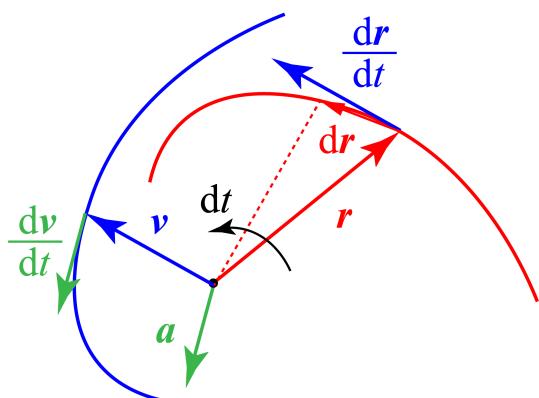


图 1.3: 质点的位置,速度与加速度

前者对应二维运动,后者对应三维运动. 我们从中可以发现, 函数是描述运动的数学工具, 对矢量与函数的分析学知识便是运动学和物理学的基础. 而运动学本身是可以看成是数学的一个分支的. 对质点的位置矢量求导数便得到速度矢量, 再求导数或是求二阶导数便是加速度矢量:

$$\mathbf{v} = \frac{d\mathbf{r}}{dt} = \lim_{\Delta t \rightarrow 0} \frac{\mathbf{r}(t + \Delta t) - \mathbf{r}(t)}{\Delta t}$$

$$\mathbf{a} = \frac{d\mathbf{v}}{dt} = \lim_{\Delta t \rightarrow 0} \frac{\mathbf{v}(t + \Delta t) - \mathbf{v}(t)}{\Delta t}$$

$$\mathbf{a} = \frac{d^2\mathbf{r}}{dt^2} = \lim_{\Delta t \rightarrow 0} \frac{[\mathbf{r}(t + \Delta t) - \mathbf{r}(t)] - [\mathbf{r}(t) - \mathbf{r}(t - \Delta t)]}{\Delta t^2}$$

正如上图??所示, 我们可以把每一个瞬时的速度矢量共起点画出来, 并把端点连做曲线, 此时背景的空间往往叫

做速度空间(velocity space), 它的更普遍的对应是动量空间, 这在近代物理的理论值具有重要地位. 引入速度空间后质点的运动便可以等价地在速度空间中进行研究. 比如在均匀重力场中的抛体运动在速度空间中就是简单的竖直向下的匀速直线运动. 而平方反比力下的天体运动(开普勒问题)实际上在速度空间中可以证明将变为变速偏心圆周运动. 这使得通过速度求加速度就好像通过位置求速度那样方便. 具体求法接下来还得进一步展开.

加速度的导数偶尔也会进入物理研究的视野, 它被叫做急动度(jerk). 即:

$$\mathbf{j} = \frac{d\mathbf{a}}{dt} = \frac{d^3\mathbf{r}}{dt^3}$$

矢量函数概念就是实数到矢量的映射, 但是具体写出函数形式的例子在物理中却不是那么常见. 比如抛体运动的运动对应的函数, 即运动方程(equation of motion)<sup>9</sup>为:

$$\mathbf{r} = \mathbf{r}_0 + \mathbf{v}_0 t + \frac{1}{2} \mathbf{g} t^2$$

它是牛顿定律下以下微分方程结论的解:

$$\mathbf{a} = \frac{d^2\mathbf{r}}{dt^2} = \mathbf{g}$$

但是在一般情况下, 要表示出一个既有大小又有方向的矢量函数, 我们得采用分解的思想, 研究能够完全确定这个矢量值的参数. 通常有以下做法:

#### 1. 直角坐标系下的分解

在一个欧几里得空间里的一个参考系中研究问题, 很常见的一种做法是建立直角坐标系来表示一个点的位置. 以更普遍的三维运动为例. 既然物体的位置由坐标  $(x, y, z)$  来表示. 我们就能认为表示运动的任务

<sup>9</sup>指运动学运动方程, 动力学运动方程就是反映运动的规律的, 能够最终解出来运动学运动方程的微分方程.

由三个标量函数  $x(t), y(t), z(t)$  来承担. 其实这与上面介绍的用矢量来表示位置是非常一致的, 因为实际上点的坐标通常也被理解为原点指向这个点的矢量在三个基矢的坐标:

$$\mathbf{r} = (x, y, z) = xi + yj + zk$$

出于哈密顿(Hamilton) 爵士关于把三维空间三个方向基矢视作四元数代数运算的旋转等效的奇妙观点, 我们分别把  $x, y, z$  方向的基矢按照惯例记做  $i, j, k$ . 也常常因为它们是单位矢量(长度为 1), 而记做  $e_x, e_y, e_z$ . 这样便能把  $\mathbf{r}$  分解出  $x, y, z$  来.

分解以后, 重新合成便能得到需要的物理量. 这其中的意思是: 比如我们为了求一个曲线运动速度或者加速度的  $x$  分量. 也就是求下列表达式:

$$\mathbf{v} = v_x i + v_y j + v_z k = \frac{d\mathbf{r}}{dt}$$

$$\mathbf{a} = a_x i + a_y j + a_z k = \frac{d^2\mathbf{r}}{dt^2}$$

中的分量  $v_x, a_x$ . 其求法非常直接, 便是:

$$v_x = \frac{d}{dt}x(t) \quad , \quad a_x = \frac{d^2}{dt^2}x(t)$$

这相当于是说, 求导数和向某固定方向投影是可以交换前后次序的:

$$i \cdot \frac{d}{dt}\mathbf{r} = \frac{d}{dt}(i \cdot \mathbf{r}) \quad , \quad i \cdot \frac{d}{dt}\mathbf{v} = \frac{d}{dt}(i \cdot \mathbf{v})$$

这是当然的, 因为矢量点乘的导数的法则也符合与标量乘积求导数类似的莱布尼茨法则, 而常矢量  $i$  不随着运动改变. 所以导数只有对  $\mathbf{r}, \mathbf{v}$  求导的项.

在数学上, 曲线段也经常被描述为是与某实数上区间同胚(homeomorphic) 的几何图形, 这其实也就是说, 存在自变量  $t$ , 不失一般性地, 我们让其取值范围为  $(-\infty, +\infty)$ . 而存在一个从这个自变量到三维空间点  $(x, y, z)$  的光滑映射:

$$x = x(t) \quad , \quad y = y(t) \quad , \quad z = z(t)$$

在数学上看, 这其实就是曲线方程: 其参数方程的唯一给法. 这样的式子就给出了几何学上的光滑曲线的定义. 而不同于以往我们对曲线是  $y = f(x)$  的图像的刻板认识. 但在这里, 我们恰巧发现如果把参数  $t$  理解为时间的话, 这恰好就对应到了

## 1.3 参考系变换

## 1.4 运动的牵连



## 第二章 动力学

### 2.1 牛顿定律

中世纪到近代以来物理学的发展,可谓沧海桑田,旧的观点被否定,几十年以后突然又以全新的面貌出现在物理学理论中.其中,实验中观察到各种令人惊诧的结果,无论物理学所依托的数学形式一变再变,要说亘古不变的主题,恐怕只有自然界的神奇与深邃,和我等凡人难以参透其中奥妙却仍为之孜孜不倦研究的决心.物理是一门研究自然界存在的运动并提出解释与理解方法的学科.运动形式多种多样,变化万千,但很多都被我们或多或少地解释,其实总结来看,从古至今我们也仅仅拥有过两类解释:

牛顿开创的相互作用观:物质间存在相互作用,相互作用是改变物质运动状态的原因.这种观点植根于伽利略等先人对古希腊经验学说的批判.牛顿很成功地发展出系统的理论来,并很好地把天体的运动归结到互相之间产生的相互作用上.并表述为牛顿第二定律的形式.后人建立了场的理论以后,超距的粒子间相互作用就不再被理论学家们采纳了,而要理解为场与粒子的相互作用,再晚一点,场和粒子都被赋予了波粒二象性,粒子也要被理解为场,但所有场都要被分解为平面波,平面波又要被量子化.如果没有相互作用,那么这些平面波的运动状态就不会发生改变,只有相互作用才会改变其运动,或传播方向发生偏折(发生散射),或强度发生改变(粒子的产生与湮灭).

爱因斯坦开创的几何观:

### 2.2 动量定律

### 2.3 能量定律

### 2.4 角动量定律

### 2.5 位力定律 \*

### 2.6 动力学问题

### 2.7 碰撞问题



# 第三章 静力学

3.1 约束

3.2 力系化简

3.3 平衡问题

3.4 虚功原理

3.5 平衡态稳定性



# 第四章 振动与波

## 4.1 方程与谐振

### 4.1.1 简谐振动的定义

振动是最常见的物理现象. 而振动中的最简单 (simple) 而和谐 (harmonic) 者谓之简谐振动 (simple harmonic oscillation). 对谐振子 (harmonic oscillator) 的学习与研究是会贯彻整个物理理论不同层次内容的始终的. 现在是经典力学, 以后会上升到场论, 量子力学与量子场论的高度.

简谐振动是指一个物理量  $Q$  随时间围绕其平衡位置做上下的波动. 其形式符合:

$$Q = Q_0 + \Delta Q \cos(\omega t + \varphi)$$

我们经常会有用复数表示振动的习惯, 其做法是在三角函数  $\cos$  与其宗量 (argument)  $\phi = \omega t + \varphi$  构成的项后添加一个虚的  $i \sin \phi$  项, 于是新的写法变成:

$$\tilde{Q} = Q_0 + \Delta Q e^{i(\omega t + \varphi)} ; Q = \Re(\tilde{Q})$$

又或者:

$$\tilde{Q} = Q_0 + \Delta \tilde{Q} e^{i\omega t} ; \Delta \tilde{Q} = \Delta Q e^{i\varphi}$$

以上各个常量中,  $Q_0$  是平衡位置,  $\Delta Q$  叫振幅 (amplitude), 宗量  $\omega t + \varphi$  叫做相位 (phase),  $\omega$  叫角频率 (angular frequency),  $\varphi$  叫初相位.  $\tilde{Q}$  为复化的复物理量, 而  $\Delta \tilde{Q}$  叫做复振幅 (complex amplitude).

复数表示最大的一个好处就在于很方便计算物理量的线性组合与导数. 事实上:

$$\dot{Q} = -\omega \Delta Q \sin \phi ; \dot{\tilde{Q}} = i\omega \tilde{Q} \Rightarrow \dot{Q} = \Re(\dot{\tilde{Q}})$$

$$\ddot{Q} = -\omega^2 \Delta Q \cos \phi ; \ddot{\tilde{Q}} = -\omega^2 \tilde{Q} \Rightarrow \ddot{Q} = \Re(\ddot{\tilde{Q}})$$

### 4.1.2 简谐振动的性质

以一个质点水平坐标  $x$  围绕  $x = 0$  左右做谐振为例. 其运动方程写作<sup>1</sup>:

$$x = A e^{i(\omega t + \varphi)}$$

那么其速度与加速度为:

$$v = i\omega A e^{i(\omega t + \varphi)}, a = -\omega^2 A e^{i(\omega t + \varphi)}$$

我们发现在运动过程中的任意一个时刻, 加速度都是指向平衡位置的, 与偏离平衡位置的位移是成正比的:

$$a = -\omega^2 x$$

<sup>1</sup> 以后我们对物理量和物理量的复化只在必要的时候加以区分, 看到复数只需要认为省写了取实部这一例常操作罢了.

而任意一个时刻既然  $x$  是余弦函数,  $v$  是正弦函数, 它们就满足其绝对值大小的“此消彼长”关系<sup>2</sup>:

$$v^2 + \omega^2 x^2 = \omega^2 A^2$$

相位是很重要的物理概念, 相(phase), 状态也, 相位则是一个可以用来表示状态的数. 反过来, 我们可以根据物体在一个时刻的状态反过来确定这个时刻的相位. 如果位移是  $x$  而振幅为  $A$ , 则:

$$\phi = \text{Arcsin} \frac{x}{A}, \text{Arcsin} \frac{x}{A} \in \left\{ \arcsin \frac{x}{A} + 2n\pi \mid n \in \mathbb{Z} \right\} \cup \left\{ \pi - \arcsin \frac{x}{A} + 2n\pi \mid n \in \mathbb{Z} \right\}$$

同理如果已知速度  $v$  和速度振幅(velocity amplitude)  $\omega A$ , 那么相位为:

$$\phi = \text{Arccos} \frac{v}{\omega A}, \text{Arccos} \frac{v}{\omega A} \in \left\{ \arccos \frac{v}{\omega A} + 2n\pi \mid n \in \mathbb{Z} \right\} \cup \left\{ -\arccos \frac{v}{\omega A} + 2n\pi \mid n \in \mathbb{Z} \right\}$$

相位的取法上具有多值性. 但是相位随时间的变化我们约定必须是连续的, 事实上它随时间线性增加. 也就是说如果前一个状态下相位为  $\phi_1$ , 后一个状态下相位为  $\phi_2$ , 那么这两个状态间历时:

$$\Delta t = \frac{\phi_2 - \phi_1}{\omega}$$

## 4.2 阻尼振动与受迫振动

## 4.3 多自由度小振动 \*

## 4.4 非线性摄动

## 4.5 格波

## 4.6 波动方程

## 4.7 波的色散

---

<sup>2</sup>注意到这里要避免使用复数, 因为一个复数的实部平方不会等于其平方的实部  $v^2 \neq \Re(\vec{v}^2)$

# 第五章 万有引力

5.1 有心力下运动

5.2 万有引力下运动

5.3 二体与潮汐



# 第六章 刚体

## 6.1 刚体的物理描述

近代以前人们意识到了物质世界的连续性(continuum), 同时针锋相对地也提出了原子论(atomism). 原子最简单的模型就是质点, 而调和物质连续性与原子学说的中间模型就是刚体(rigid body) 模型. 刚体是不允许形变发生的系统. 由牛顿力学对质点的讨论推广到质点系的讨论, 使我们也很容易将相关结论进一步推广到刚体.

由于刚体上一点受力, 则整体同时运动起来, 这个模型与相对论力学体系是不兼容的. 具体来说, 相互作用必须以有限的速度传播, 否则就违背了因果律. 刚体不符合因果律这一时空的固有结构. 在很多相对论情境下将招致矛盾的结果.

刚体的物理学量是哪一些呢? 必要的内禀的属性是其质量的分布. 某一默认时刻  $t_0$  刚体占据了空间区域  $\Omega_0$ , 由大量体积微元  $dV$ (记做  $d^3\mathbf{R}_0$ ) 组成, 则刚体的总质量为:

$$m = \int_{\Omega_0} \rho(\mathbf{R}_0) d^3\mathbf{R}_0 = \int dm$$

随着刚体的运动, 原来在  $\mathbf{R}_0$  处的体积元现在在  $t$  时刻位于  $\mathbf{r}$  处, 整个刚体的运动由一个多元映射来定义:

$$f : \mathbf{R}_0, t \longrightarrow \mathbf{r}(\mathbf{R}_0, t)$$

刚体的刚性的要求, 使得这个映射必须保持体积元的不变性:

$$f : d^3\mathbf{R}_0 \in \Omega_0 \longrightarrow d^3\mathbf{r} \in \Omega ; \quad d^3\mathbf{R}_0 = d^3\mathbf{r} = dV$$

而且所有这个体积元内的所有内禀属性, 这里包括密度都不能变. 所以质量元  $dm$  也是不变的. 从而刚体具有不变的总质量. 马上就会发现, 质量几何(mass geometry) 对刚体的动力学来说也十分重要. 质量几何研究质量对特定原点  $O$  的各级矩(moment). 其中零级矩即为质量:

$$M^0 = m = \int_{\Omega} dm$$

一级矩是个矢量, 它定义了刚体的质心(center of mass) 的位置:

$$M_i^1 = mr_{Ci} = \int_{\Omega} r_i dm$$

$$\mathbf{r}_C = \frac{\int \mathbf{r} dm}{\int dm}$$

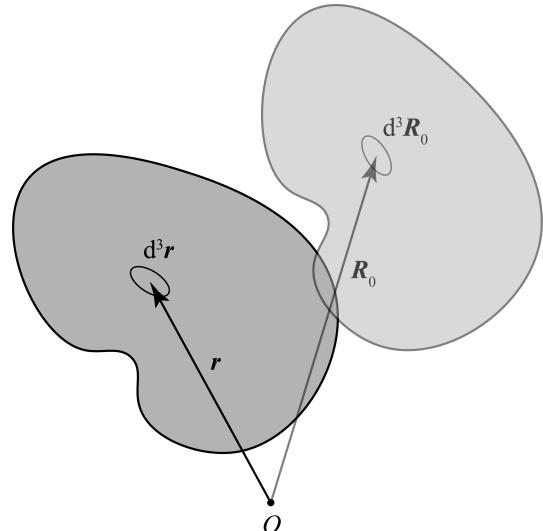


图 6.1: 刚体的描述

二级矩则是一个张量,它的九个分量代表**惯量积**(product of inertia):

$$M_{ij}^2 = \int_{\Omega} r_i r_j dm$$

这些矩和原点的选取有关,随着刚体的运动也会不断变化,这三阶矩的信息对刚体动力学来说就是充分的了,通过后面的动力学可以发现,刚体的运动完全依赖于外力和这三阶矩.如果要研究广义相对论里的引力波辐射问题,更高阶的矩才变得重要起来.

刚体的运动可以被我们更精确地描述,在  $t_0$  时刻建立固定在刚体上,沿  $x, y, z$  三方向的单位矢量  $e_1, e_2, e_3$ ,那么刚体的运动同时也把三个矢量旋转到新的三个方向:

$$f: e_i \longrightarrow \epsilon_i$$

这三个矢量仍然要互相垂直,且长度为一.这在数学上导致了可以通过这三个矢量的导数定义**角速度**(angular velocity) 矢量的结果:

$$\dot{\epsilon}_i = \omega \times \epsilon_i$$

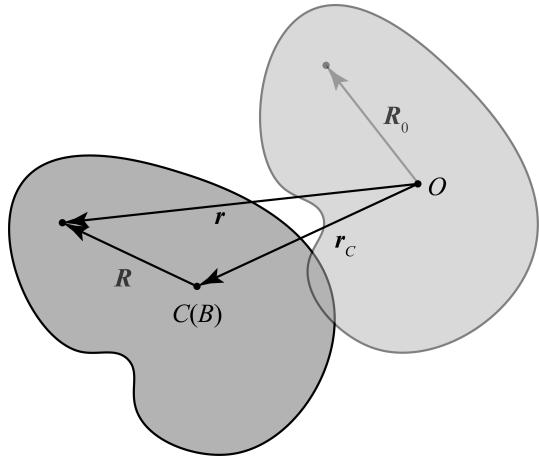


图 6.2: 基点法

十分类似于旋转参考系的变换的运动学,刚体的运动实际上就是有一个唯一的旋转参考系固连在刚体上,以后称作**刚体系**(reference system of rigid body),要研究的刚体上各个点速度实际上就是刚体系中定点的运动.于是习惯上我们采用**基点法**(method of base point) 来计算刚体上任意点的运动学量.定义**基点**(base point) 为  $t_0$  时刻  $\mathbf{R}_0 = \mathbf{0}$  的点  $B$ ,之后的位矢为  $\mathbf{r}_B$ ,对应的基点速度加速度为  $\mathbf{v}_B, \mathbf{a}_B$ ,而刚提上待研究的点相对基点的位矢为  $\mathbf{r} - \mathbf{r}_B = \mathbf{R}$ ,那么该点的速度加速度即为:

$$\mathbf{v} = \mathbf{v}_B + \omega \times \mathbf{R}$$

$$\mathbf{a} = \mathbf{a}_B + \omega \times (\omega \times \mathbf{R}) + \dot{\omega} \times \mathbf{R}$$

出于力学的考虑,基点  $B$  一般就取做质心  $C$ .作用在刚体上  $\mathbf{r}$  处的力  $\mathbf{F}$  固然对原点会有力矩  $\mathbf{M}$ ,但是为了研究使刚体自身转动的效应,考虑到这个力同时也会使得质心运动起来,故我们重视这个力相对质心的力矩  $\mathbf{M}_C$ :

$$\mathbf{M} = \mathbf{r} \times \mathbf{F}$$

$$\mathbf{M}_C = \mathbf{R} \times \mathbf{F}$$

刚体受到一个力系  $\mathbf{F}_i$  的作用,那么以下六个定理则来自于之前的动力学理论:

$$\begin{aligned} \sum_i \mathbf{F}_i &= \frac{d\mathbf{p}}{dt} ; \quad \sum_i \mathbf{F}_i = m\mathbf{a}_C \\ \sum_i \mathbf{v}_i \cdot \mathbf{F}_i &= \frac{d}{dt} \left( \frac{1}{2} m \mathbf{v}_C^2 + E_{kr} \right) ; \quad \sum_i (\mathbf{v}_i - \mathbf{v}_C) \cdot \mathbf{F}_i = \frac{dE_{kr}}{dt} \\ \sum_i \mathbf{r}_i \times \mathbf{F}_i &= \frac{d}{dt} (\mathbf{r}_C \times m\mathbf{v}_C + \mathbf{L}_r) ; \quad \sum_i \mathbf{R}_i \times \mathbf{F}_i = \frac{d\mathbf{L}_r}{dt} \end{aligned}$$

前两个式子实际上几乎没有区别,因为刚体的总动量其实就是质心动量  $\mathbf{p} = m\mathbf{v}_C$ .后面几个式子则涉及到相对(平动)质心系的能量与角动量.它们与质量二级矩有着密切的联系,我们在第三讲6.3阐明.现在我们写出相对质心系能量角动量的定义:

$$E_{kr} = \int_{\Omega} \frac{1}{2} (\omega \times \mathbf{R})^2 dm$$

$$\mathbf{L}_r = \int_{\Omega} \mathbf{R} \times (\omega \times \mathbf{R}) dm$$

## 6.2 平面平行运动

一个底面磨平的物体贴在平坦的地面上运动给人以平面平行运动(plane-parallel motion)的概念. 其他的一些物体运动特征也相似于它: 黑板刷在黑板上的运动, 车轮在直行时的滚动... 它们的特征是: 刚体任何一个体积元都在一个特定的平面上运动, 而这些平面又彼此平行. 对于这样的运动基点法只需要给出质心  $C$  的位矢  $r_C$  和对  $t_0$  时刻刚体转过的角度  $\theta$  即可. 角速度与角加速度即可用角度的导数来给出  $\omega = \dot{\theta}$ ,  $\beta = \ddot{\theta}$ . 一般来说, 在垂直于这些平面方向要么由于动力学对称性不需要也不存在任何作用力. 要么这些力作为约束平面对物体的约束力而不在我考虑范围内. 所以我们对一些物理对象做如下修改:

- 所有参考点改成垂直平面过该点的参考轴. 并约定其正方向

## 6.3 空间刚体运动 \*



# 第七章 流体

7.1 流体的物理描述

7.2 定常流动动力学

7.3 黏滞流体动力学



# 第八章 弹性体

## 8.1 弹性体的物理描述

所谓弹性体就是完全弹性(elasticity)的物体. 弹性描述的是使物体发生形变的力撤除以后物体可以回到静息状态的属性. 弹性力学研究的对象与范围就是弹性体的力学性质. 一般来说, 固体主要具有弹性而液体主要具有黏性, 若是研究中间的状态, 非牛顿流体(non-newtonian fluid) 和塑性固体(plastic solid), 那就是黏弹性力学(rheology) 要研究的对象了. 典型的黏弹性过程受力不是简单地正比于位移而是与速度, 与历史相关. 因此而可以发生永久的不可恢复的变形.

正因为如此, 完整描述弹性体的运动学时, 不得不额外留心所有点的实际位移. 在流体时也许速度更需要注意. 所以我们写出一个初始  $t = 0$  位置矢量为  $\mathbf{R}$  的点, 经过  $t$  时间到达位置为  $\mathbf{r}$  处, 也就是我们要定义一个  $3D \times 1D$  到  $3D$  的映射:

$$\mathbf{r} = \mathbf{r}(\mathbf{R}, t)$$

不失普遍性地, 我们考虑如何刻画在  $\mathbf{R} = \mathbf{0}$  的形变. 我们需要研究在  $\mathbf{R} = \mathbf{0}$  的附近  $d\mathbf{R} = dX\mathbf{e}_x + dY\mathbf{e}_y + dZ\mathbf{e}_z$  处的位移与中心的位移去比较. 数学上有以下泰勒展式:

$$\mathbf{r}(d\mathbf{R}, t) = \mathbf{r}(\mathbf{0}, t) + d\mathbf{R} \cdot \nabla \mathbf{r}$$

上式中  $\nabla \mathbf{r}$  是一个有九个分量的张量, 它的作用是可以与之前的矢量点乘把它线性地映射为另一个矢量:

$$\begin{aligned}\nabla \mathbf{r} &= \sum_{i,j} \frac{\partial x_j}{\partial X_i} \mathbf{e}_i \mathbf{e}_j \\ \nabla \mathbf{r} : \sum_i dX_i \mathbf{e}_i &\rightarrow \sum_j dx_j \mathbf{e}_j = \sum_j \left( \sum_i \frac{\partial x_j}{\partial X_i} dX_i \right) \mathbf{e}_j\end{aligned}$$

不难发现第二个式子是不证自明的. 所以实际上刻画形变的包含于  $\nabla \mathbf{r}$  这个张量. 但是并不是完全取决于它, 考虑像刚体这样的不能变形的物体, 由于

## 8.2 弹性棒, 弹性膜与弹性体

## 8.3 弹性波



# 第九章 相对论

**9.1 相对论运动学**

**9.2 相对论动力学**

**9.3 相对论连续物质**

**9.4 相对论电磁场**



# 第十章 热力学第一定律

## 10.1 热力学第零定律

热力学系统的状态可以分为平衡态(equilibrium state) 和非平衡态(non-equilibrium state). 平衡态这样一种状态, 首先它是一种宏观性质都不随着时间演化的状态---稳定态<sup>1</sup>, 但稳定态不一定是平衡态, 如两端温度保持不变后稳定导热的杆, 如稳定喷气的火箭, 如不断气化液化工作循环制冷的空调工作物质. 原因是它们没有符合以下三个平衡条件:

热学平衡 系统各部分温度严格相等.

力学平衡 系统内力(如压强)与外力达到平衡, 如重力场下的等温大气模型.

化学平衡 系统各部分热运动扩散的趋势相互平衡, 如饱和糖水与糖块共存.

可以发现, 既使系统状态不均匀, 如重力场等温大气模型的各处压强和饱和糖-糖水模型中的两相糖分子数密度, 系统也是处于平衡态的. 而如果这些平衡条件没有达成, 系统内部将发生各种宏观可见的输运过程(transport phenomena). 如温度不均导致的热传导, 速度与压强不均导致的黏滞, 及浓度不均导致的扩散. 不平衡的力产生了不平衡的流, 相应的研究是非平衡态物理研究的对象. 这种情况下既使系统随时间稳定, 也不是平衡态.

上文中已多次提及温度(temperature)这一物理概念. 今天的我们完全可以用统计力学的思考方式来理解它: 诚然, 温度是分子热运动剧烈程度的量化. 如理想气体的温度定义为:

$$T = \frac{2\bar{\varepsilon}_k}{3k}$$

其中  $\bar{\varepsilon}_k$  表示分子平均平动能, 而  $k$  表示玻尔兹曼常数. 也就不难理解, 为什么温度不均将导致传热, 因为分子平均动能大的子系统 A 与分子平均动能小的子系统 B 发生热接触时, 平均效应是 A 中的分子通过相互作用要把能量给 B. 这叫热传递(heat transfer).

然而, 在统计力学还没有诞生前, 甚至在很长的一段时间内人们对热与温度的能量本质并没有明确的认识, 热质说(caloric theory)大行其道, 人们把热量导致的温度升高视为与功导致的能量升高为完全独立的两个维度. 但正是在这样的环境内, 诞生了基于经验规律的物态方程, 发明了测温, 量热的各种实用仪器, 乃至甚至产生了公理化的热力学体系. 热力学第零定律即给出了温度可定义的依据.

热力学第零定律 任意两个各自达到热平衡的系统若发生热接触可以直接形成热平衡, 则说这两个系统具有一致的温度. 等温具有反身性, 交换性与传递性.

热力学第零定律的补充 如果两个各自达到热平衡的系统若发生热接触一者传递给另一者热量, 则说前者比后者温度高.

<sup>1</sup>注意, 一是从微观上看, 组成系统的微观粒子仍然处于无休无止的运动中, 其微观坐标与速度在发生几急剧的变化, 但平衡态只对宏观性质定义. 二是统计力学可以证明, 处于平衡态的系统总是由于内在随机性而产生涨落, 有时甚至不可忽略, 如临界乳光现象. 这说明我们的平衡态模型只能是一种近似, 实际情况往往还有很多需要考虑的因素.

这一定律确定了温度作为系统的**态函数**(state function), 它能像一个数那样可以进行大小的比较. 有两点值得注意:

一是热力学第零定律定义了温度, 但它没有量化一个**温标**(temperature scale). 如何量化一个温标呢? 从各种系统平衡态的性质出发可以以某一具有特点的状态量作为温度评价的标度. 如理想气体温标, 黑体辐射谱温标和噪声温标等等. 而我们发现, 温度高低的比较似乎与其传热的难易程度有关, 是否可以利用这一点自然而然地定义普适温标呢? 热力学第二定律回答了这个问题, 答案是肯定的.

二是热力学第二定律与热力学第零定律的内在联系. 作为热力学第零定律的补充, 其实它正是热力学第二定律的一部分. 热力学第二定律中得到了很好的量化定义了热力学温标以完善温度这一概念. 其实, 这两个定律的背后是简单的统计规律, 若我们采取一种统计力学的方法, 我们的理论基础便是统计公设, 而热力学定理反而成为我们需要去证明和解释的对象了.

## 10.2 热力学第一定律

热力学第一定律的研究对象是热力学过程, 是既定系统连续发生的这样或那样的真实状态改变. 而热力学第一定律的微观本质是普遍的能量转化与守恒原理. 它写为:

$$dU = dW + dQ$$

$U$  为系统内能, 表示系统内部能量的总和. 等式左边为内部能量增量, 右边为外界输入系统内部的能量, 其中  $dW$  为外界对系统内部的做功, 而  $dQ$  为外界通过热接触, 以微观碰撞交换能量的形式, 产生的平均宏观能量输送的大小. 注意,  $U$  前面的  $d$ , 表示无穷小变化量, 但  $W$  与  $Q$  前面的  $d$ , 仅仅表示无穷小量(微元), 不是某个热力学态函数的微分. 这也是两个微分符号记法不相同的原因. 这两种不同的热力学量可以称为**状态量**(state quantity) 与**过程量**(process quantity).

在具体考察上述三个量的具体范围前, 明确一下过程发生的具体形式是十分有必要的.

### 10.2.1 准静态与非准静态过程

**准静态过程**(quasi-static process): 过程进行地速度无限缓慢, 以至于每一时刻系统都几乎处于平衡态. 如等温传热的过程; 保持外界压强始终等于气缸压强地压缩气体的过程; 亦或是缓慢地冷却热溶液, 使得溶质析出的过程, 都是准静态过程. 相反, 高温物体向低温物体放热的过程, 气体向真空自由膨胀的过程, 把糖丢进纯水里溶解的过程或两种不同气体混合的过程, 就都是非准静态过程.

过程的准静态与非准静态有以下两个特征:

一是准静态由于每时每刻都是平衡态, 所以可以由平衡态的均匀状态参量去描写, 从而得出各状态参量随时间的变化关系, 消去时间可得**状态方程**(process equation). 更特别地, 在过程中所发生的内能变化, 做功, 吸热全都可以用状态方程去确定, 比如以  $p, V$  作为状态参量的理想气体:

$$dU = \frac{\partial U}{\partial p} dp + \frac{\partial U}{\partial V} dV ; \quad dW = -pdV ; \quad dQ = dU - dW$$

注意在上式中最后一式不是说  $dQ$  是  $dU$  与  $dW$  的结果, 在热一的公式中, 等式右边的做正负功和吸放热永远是原因, 等式左边内能的增减永远是结果. 不过这是一个很常用的计算式, 图像也还算直观: 气体吸收的热量用来增加内能和对外做功.

而非准静态过程则不能由状态方程描写, 也没有以上三个方程的成立, 我们在更晚些时候展开讨论.

二是准静态过程由于每时每刻都达到了平衡, 在平衡的状态下发生热量, 功或者粒子的交换, 所以是可逆过程<sup>2</sup>. 而由于非准静态过程没有达到各种平衡, 故自发产生的各种过程都是不可逆的.

我们接下来仔细考虑内能, 功与热量三个物理概念.

<sup>2</sup>严格地说, 还要求过程没有耗散, 耗散是一类功变热的行为, 它使得有序的功对应的能量转化为无序的热对应的能量.

### 10.2.2 内能

**内能**(internal energy), 就是系统内部的能量和. 它由许多可能的成分构成, 如:

分子平动能 三个自由度;

分子转动动能 线性分子只有两个自由度<sup>3</sup>; 非线性分子则有三个;

分子振动动能<sup>4</sup> 分子间键长, 键角偏离平衡位置的集体振动模式, 其自由度数与上两者之和为  $3n$ ,  $n$  为分子含原子数;

分子振动势能 由于分子的振动, 其势能也会比其平衡位置更大, 近似地与分子振动动能相等;

键能 即电子运动的动能势能之和 (全部能量);

静能 这里指以上运动的动能, 势能全部去除后, 由于更加深层次内部相对运动造成的静能量, 如核能, 基本粒子的自能等等.

有两点值得注意:

一是在内能具体的计算中, 对于我们不关心的, 过程中不发生改变的部分我们将其置零. 如双原子分子理想气体的问题中研究的内能不会去包括两原子间化学键的能量, 除非发生化学反应. 而在常温下其振动自由度往往难以被激发, 故其等效自由度约为 5. 自然界对于尺度的放缩并不具有对称性, 所以很多能量实际上在更微小的层次内封装为不变的整体, 这使得我们只需要研究我们关心的尺度上的问题, 即**有效理论**(effective theory) 的思想.

二是有一些形式的能量是否属于内能的一部分视问题的不同而可以做不同的处理. 这体现为功能原理上:

$$dW_{\text{conservative}} = -dE_{\text{potential}}$$

即, 在热力学第一定律的计算中, 我们不去计算保守力部分做的功. 而是把它对应的势能视为系统内能的一部分:

$$dU = dW_{\text{others}} + dW_{\text{conservative}} + dQ \implies dU_{\text{total}} = dU + dE_{\text{potential}} = dW_{\text{others}} + dQ$$

最经典的例子是那些热力学系统中每个粒子都具有外场中的相互作用势能的情况. 如气体的重力势能与电介质的静电能, 就常常当做系统内能的一部分, 然而有些情况下又使用不含这些能量的系统内能更加方便. 再比如在恒定的外压强下的气体系统, 外界对气体做功可以用体积功来表示:

$$W = \int_V^{V+\Delta V} (-p_0 dV) = -p_0 \Delta V$$

从中可以看出, 气体具有体积能  $E_{\text{volume}} = p_0 V$ , 可以定义新的总内能为:

$$H_0 = U + p_0 V$$

如果气体与外界压强平衡<sup>5</sup>, 则上式中的压强就是气体压强  $p$ , 以上量称为气体的**焓**(enthalpy):

$$H = U + pV$$

<sup>3</sup>分子对自转轴的转动不会带来系统的改变, 这与全同粒子的量子本性有关系.

<sup>5</sup>如系统在大气压下发生一个化学反应等等.

### 10.2.3 功

**功**(work) 有两个特点:一, 功是一种能量转移, 考虑功能原理, 对一个系统做正功使得系统动能增加, 而这是系统内能的一部分, 从而能量传入系统. 二, 功总是与力对应. 热力学中为了方便我们引入所谓**广义力**(generalized forces) 与**广义坐标**(generalized coordinates) 来方便表示我们的系统. 如理想气体广义坐标如果取做体积  $X = V$ , 则广义力就是外压强的相反数  $Y = -p_0$ , 它们满足外界对研究的系统做功为:

$$dW = Y dX = -p_0 dV$$

这被叫做**体积功**(volume work). 与其他形式的功对应的广义力与广义坐标并在一张表中供参考:

功形式	广义力	广义坐标
体积功	$-p$	$V$
表面张力功	$\sigma$	$S$
电路功	$U$	$Q$
极化功	$E$	$P$
磁化功	$H$	$M$

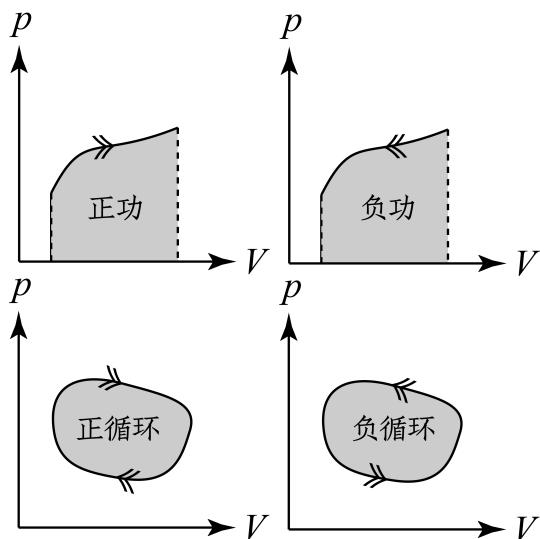


图 10.1: 准静态过程的做功

对于体积功, 在准静态过程中可以由  $p - V$  图下的面积计算. 在循环过程中则等于过程曲线所包含的面积. 注意其正负号与过程方向之间的关系. 但对于非准静态过程. 做功一定要由外界参数 (活塞加速度, 活塞外气压等) 计算而不是用内部气体的过程方程. 这是因为即使形式地写出内部气体所发生的过程的过程方程. 由于气体发生膨胀处的压强不等于过程方程中的平均压强, 对做功的计算是没有意义的.

### 10.2.4 热量

做功与**传热**(heat transfer) 是截然不同的两种向系统注入能量的方式. 做功依赖于广义坐标这一参数的改变. 本质上与每一个微观粒子的能量对于广义坐标的依赖有关. 然而热量则完全不依赖于外在的参数. 它与系统粒子关于不同能量的状态的占据数**配分**(partition) 有关. 也就是与温度息息相关. 温度高的物体, 占据在高能量状态上的粒子较多, 低能量状态的粒子较少, 与低温物体发生热接触时, 单位时间内高能粒子碰撞低能粒子传递能量的可能性更高. 其统计平均效应就是高温物体向低温物体放热. 准静态的**等温传热**又是怎样一个过程呢? 从以上分析可知, 产生温度梯度是有热传导<sup>6</sup>的必要条件, 没有了温度差热量的交换也就无从谈起了. 从这个意义上说, 实际发生的过程其实都是非准静态过程, 准静态过程仅仅是一种实际过程的一种近似. 或是我们虚设的一种模型. 如在等温传热模型中, 我们设想只让一个物体的高能粒子与另一个物体的低能粒子发生热接触而实现热传递, 这无疑会破坏热力学的统计规律<sup>7</sup>, 但有助于我们理解这个过程的发生机理.

<sup>6</sup>注意, 广义的热量传递还包括热对流与热辐射.  
<sup>7</sup>或是借助于所谓的麦克斯韦妖, 引入信息熵.

### 10.2.5 耗散

还有一类向系统输送能量的方式,表面上看是功,效果上却是热。这被我们称为**耗散**(dissipation)。俗称**功变热**。例如常见的摩擦生热过程,非弹性碰撞的过程,激光泵浦的过程。都是典型的功变热的例子。我们看一个十分浅显的例子。涡轮叶片搅动气缸中的空气使其温度升高。在这个过程中,涡轮无疑对气体做功,但这个做功并不与系统广义坐标的变化相关。它的特点是仅仅作用在系统的一个子系统上。再由系统倾向于热平衡的趋势把得来的能量平分到系统各个自由度上。从而造成了与吸热完全类似的结果:广义坐标没有变化,但系统从外界获得了能量。

究其原因,本质上是因为系统具有了所谓的**耗散结构**(dissipative structure)。它不把系统看成一个平衡的整体,而是有着复杂内部非线性相互作用的对象。故输入的功不一定改变体系的广义坐标,但像加热一个系统一般,增加了系统的内能。

耗散结构的特点之一是外力驱动系统远离平衡态。在非弹性的碰撞过程中固体内部的低频弹性波的成分被突然增强,而激光发射的过程中大量粒子被输送到亚稳态而实现布居反转,这些都破坏了原来热平衡下的布局。需要结合动力学与统计方法来讨论这些现象。

特点之二是开放条件下的自组织现象。以普利高津(Proggine)为首的研究者把生命现象,连同研究过程中的很多奇特现象看成是一种耗散结构的结果。这些过程伴随着系统与外界连续不断的能量、粒子与信息的交换,使得微小的涨落可以在具有正反馈的功能结构中被放大,使得体系的对称性被破坏,形成有组织的自发序。

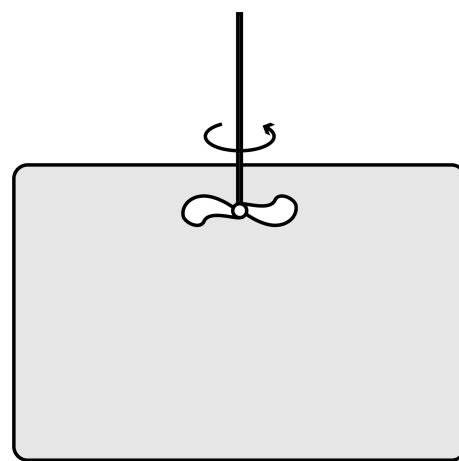


图 10.2: 搅动空气发热



图 10.3: 自组织现象:Bénard 对流与斑马身上的条纹

## 10.3 理想气体

### 10.3.1 理想气体的定义

满足以下三个条件的气体为理想气体:

1. 分子间作用力可以忽略;
2. 分子体积可以忽略;
3. 分子按自由度平分能量.

我们讨论一下这三个条件. 首先第一个条件的意思, 旨在把理想气体间的作用力当成短程力(short-range force) 处理. 仅仅当两个分子间距离进入力程范围才可以当做可以发生相互作用(碰撞或散射). 在这个意义上, 分子体积实际上也可以处理为半径为力程大小的钢球模型. 我们忽略这一个体积, 也就是说所有分子可以自由活动的体积就等于容器体积  $V$ . 且分子间相互作用力为零.

然而这两点合在一起将导致一个严重的问题: 那便是热平衡不可以达到. 若每个分子不可以通过相互作用或刚性的碰撞与别的分子交换能量与动量. 那从根本上就不可以达到热平衡, 统计方法也就无济于事. 故分子体积足够小, 小到可以忽略其对热力学性质的影响, 但又足够大, 大到在我们关心的时间空间尺度内体系能很快地达到热平衡. 这才是对这两点的正确理解方式.

我们设分子的速率分布函数为  $f(v)$ , 意思是每取  $N$  个分子, 速率介于  $v$  到  $v + dv$  的分子数平均为:

$$dN = Nf(v)dv$$

从而分子的平均速率与方均速率为:

$$\bar{v} = \int_0^\infty v f(v)dv ; \quad \bar{v^2} = \int_0^\infty v^2 f(v)dv$$

分子质量为  $m$ , 直径为  $D$ , 数密度为  $n$ . 从中我们得出分子间平均距离  $d$ :

$$d \sim \frac{1}{\sqrt[3]{n}}$$

可以想见, 分子间平均距离  $d \sim 10\text{nm}$ (大气压下) 越大, 分子直径  $D \sim 1\text{nm}$  越小, 气体越能够近似为理想气体. 同时其相隔两次碰撞的平均时间间隔  $\tau$  与空间间隔  $\lambda$  也会越大. 如何计算这两个量呢? 由于分子的速度符合麦克斯韦分布律(参见第五章), 故两分子间的平均相对速率为  $\sqrt{2}\bar{v}$ . 将待研究的粒子处理为钢球, 速率恒定为  $\sqrt{2}\bar{v}$ , 其半径为原分子直径  $D$ . 而其他分子处理为静止的质点, 半径为零. 这样钢球碰到质点, 即相当于原来的两个分子发生碰撞. 而时间  $t$  内钢球扫过的体积为:

$$V(t) = \pi D^2 \cdot \sqrt{2}\bar{v}t$$

发生碰撞次数  $c$ :

$$c(t) = nV(t)$$

当  $c = 1$  时, 对应的时间即为平均自由时间  $\tau$ , 又称为驰豫时间(relaxation time), 而这一段时间内粒子走过的距离即为平均自由程  $\lambda$ (mean free path):

$$\tau = \frac{1}{\sqrt{2}\pi n D^2 \bar{v}} ; \quad \lambda = \frac{1}{\sqrt{2}\pi n D^2}$$

即有:

$$\lambda \sim \frac{d^3}{D^2} \sim 100d \sim 1\mu\text{m}$$

以上数据仅仅在大气压下有效,注意平均自由程与分子平均速度无关,仅仅是一个几何相关量.而它与  $n$  成反比,也就是与  $p$  成反比,如果压强减小到约 1Pa(而温度不变),则平均自由程  $\lambda$  增大 5 个量级,也就是分米量级.

现在我们做一个总结:我们取了一个分子间相互作用力与分子的大小可以忽略的模型来研究理想气体,它对应的条件是:

$$D \ll d ; \quad \lambda \sim \frac{d^3}{D^2} \ll L$$

第一个条件是保证分子体积可以忽略,第二个条件是在研究的问题尺度内保证分子有足够的空间来实现动量能量交换以达到热平衡.

现在让我们专心研究第三个条件.能量按自由度均分是麦克斯韦分布的一个独特结果.详情也可以参见第五章.现在我们指出,由于系统中的各个自由度,如  $x, y, z$  方向的平动,分子的转动,振动的能量具有相似的形式<sup>8</sup>,而且可以通过相互作用(碰撞)交换能量,总能量守恒.在这样的条件下,统计规律告诉我们每一个分子每一个自由度上都分得相同的平均能量值.这个平均能量值<sup>9</sup>即代表了体系的温度:

$$\bar{\varepsilon} = \frac{1}{2}kT$$

体系自由度越多,分子数越多,在一定温度下储能就越多:

$$U = N \cdot \frac{f}{2}kT$$

其中  $f$  是系统能量自由度数<sup>10</sup>.它被视为与温度无关的常量<sup>11</sup>.对于上式,我们再引入摩尔(mole)这一化学计量单位来表示巨大的原子数  $N$ :

$$\frac{N}{N_A} = \nu ; \quad N_A = 6.022140857 \times 10^{23}/\text{mol}$$

同时,普适气体常量被定义为:

$$R = N_A k = 8.314 \text{J/mol} \cdot \text{K}$$

这样气体的内能即为:

$$U = \frac{f}{2}\nu RT$$

### 10.3.2 理想气体物态方程

我们发现,  $dp_1 = f(v)dv$  表示的含义为任意一个分子的速率落在  $v \sim v + dv$  区间内的概率.然而对于这样的分子,它速度方向在空间中的取向是任意的,各向同性的.所以建立极轴后,速度与极轴夹角在  $\theta \sim \theta + d\theta$  范围内的概率又为:

$$dp_2 = \frac{d\Omega}{4\pi} = \frac{2\pi \sin \theta d\theta}{4\pi} = \frac{\sin \theta d\theta}{2}$$

对于这样的分子,它在容器中的任何位置都是有可能且等概率的,所以出现在体积微元  $dV$  中的概率还有  $dp_3$ ,联合概率为  $dp = dp_1 dp_2 dp_3$ .

<sup>8</sup>指谐振子形式,详见第五章.

<sup>9</sup>及其对应的按能量高低的配分.

<sup>10</sup>注意其与动力学自由度的区别,振动的势能也对应能量自由度,但其动力学自由度与振动动能重合.

<sup>11</sup>然而实验发现这些不是常量,而是随着温度升高越来越大,存在一些基本不变的平台,思考这意味着什么.

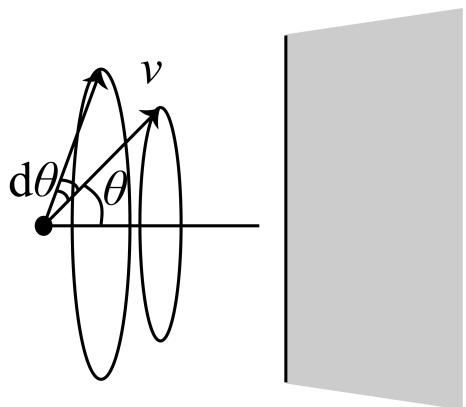


图 10.5: 粒子速度方向分布

我们接下来考虑气体的压强.在这之前我们了解两个概念是十分有帮助的:处于平衡态的系统可以看成互相达到平衡的各部分子系统(subsystem)直接组成.这些部分的所有强度量(intensive property)都是完全一致的,如理想气体的压强,温度,粒子数密度,摩尔体积,摩尔内能等等.而子系统的广延量(extensive property)大小之和即为总系统的对应量的大小,如分子数,摩尔数,体积,内能,焓等等.我们可以这么理解:强度量描述体系的内禀属性,而广延量仅仅反应全同子系统的简单堆积.如一杯盐水,确定了我们关心的强度量——温度和浓度之后,我们就可以回答“怎样的盐水”这一个问题.而接下来只需要再回答“有多少”这一个简单的问题,就能算出所有的广延量来.基于这一点理解,我们可以发现以下两个性质是成立的:

保持所有强度量不变,一个广延量变化  $\lambda$  倍,所有广延量随之变化  $\lambda$  倍.

两个广延量之比为强度量,典型的如质量  $M$  比体积  $V$  为密度  $\rho$ ,体积  $V$  比摩尔数  $\nu$  为摩尔体积  $v$  等.

理想气体作为极其简单的系统,它的独立强度量仅仅只有两个.一是系统的温度  $T$ .它决定了分子的平均热运动剧烈程度,也就是平均能量大小.另一个就是分子数密度  $n$ ,它反映单位体积里有多少个全同单元,是一种“浓度”.其他所有强度量都是这两个量的导出.比如压强  $p$ ,用分子动理论的方法:

$$\begin{aligned} p &= \int_0^\infty \int_0^{\frac{\pi}{2}} n dS v \cos \theta dt \cdot f(v) dv \cdot \frac{\sin \theta d\theta}{2} \cdot 2mv \cos \theta / dS dt \\ &= nm \cdot \int_0^\infty v^2 f(v) dv \cdot \int_0^{\frac{\pi}{2}} \sin \theta \cos^2 \theta d\theta \\ &= \frac{1}{3} nm v^2 \\ &= \frac{2}{3} n \bar{\varepsilon}_k \\ &= nkT \end{aligned}$$

可见对容器壁的压强正比于温度和分子数密度,这一点的物理图像还是很直观的.这个压强被我们称之为外压强(outer pressure).它是由于分子碰撞容器壁引起.还有所谓的内压强(inner pressure)的概念,它由于分子间的碰撞引起.理想气体的内压强与外压强相等<sup>12</sup>.但如果气体足够稀薄,以至于容器尺度与分子平均自由程可以比拟.此时内压强的概念也失去了意义,这意味着气体完全不可以用动力学方法处理,压强的梯度不会导致气体团的加速度,而是以下面所介绍的泻流的方式进行.

上文我们推出来的关系式被称作(强度量之间的)物态方程(equation of state),它有很多变式,如:

$$pV = NkT ; \quad pV = \nu RT ; \quad pv = RT ; \quad p = \frac{\rho RT}{\mu}$$

上面最后一式中  $\mu$  表示摩尔质量,稍作变形还可以得到密度公式:

$$\rho = \frac{\mu p}{RT}$$

用分子动理论方法还可以得到另一个十分重要的量——泻流数(effusion number).如果在容器壁上开

<sup>12</sup>然而范德瓦尔斯气体则不然,详见第四章.

一个小孔. 则速率在在  $v$  到  $v + dv$  区间内的分子, 单位面积单位时间通过小孔的分子数为:

$$\begin{aligned} d\Gamma &= \int_0^{\frac{\pi}{2}} n dS v \cos \theta dt \cdot f(v) dv \cdot \frac{\sin \theta d\theta}{2} / dS dt \\ &= \frac{n}{2} \cdot \int_0^{\frac{\pi}{2}} \sin \theta \cos \theta d\theta \\ &= \frac{1}{4} nv f(v) dv \\ &= \frac{1}{4} dn v \end{aligned}$$

可见速率为  $v$  的那一部分分子, 泻流数就是  $d\Gamma = \frac{1}{4} dn v$ . 那么对于所有分子, 总泻流数为:

$$\Gamma = \frac{1}{4} n \bar{v}$$

泻流的条件是开的孔的大小足够小, 以至于分子在穿过孔时不会受到别的分子的碰撞. 也就是其平均自由程  $\lambda$  显著地大于孔的直径  $L$ . 相反地, 如果  $\lambda \ll L$ , 那么就变成了一种压强差驱动气体集团运动的动力学模型, 我们在下一节阐述.

最后, 理想气体的内能则可以根据之前关于温度定义的讨论, 直接写出:

$$U = \nu C_{mV} T$$

其中  $C_{mV}$  被称为定容摩尔热容(molar heat capacity at constant volume). 其意义之后将了解. 现在只需知道  $C_{mV} = \frac{f}{2} R$ .

### 10.3.3 混合理想气体

考虑几种理想气体的混合, 我们也需要忽略几种组分分子间的相互作用, 尤其是应该排除几种分子间发生化学反应的情况<sup>13</sup>. 此时, 几种气体由各自的数密度  $n_i$  和热平衡下的公共温度  $T$  描述. 另一种描述方法是给出总分子数密度  $n$  与各组分分子所占的比例(强度量) $x_i$ :

$$n = \sum n_i ; \quad x_i = \frac{n_i}{n}$$

称为摩尔分数(mole fraction), 注意  $x_i$  间要满足的归一关系:

$$\sum x_i = 1$$

不同组分间分子纵然有碰撞, 但这对于这一部分分子碰撞容器壁产生的压强大小没有影响. 也就是说, 压强等于各组分以  $n_i$  单独存在时对容器壁产生的分压  $p_i$ . 这叫做道尔顿分压定律(Dalton's law of partial pressures):

$$p = \sum p_i ; \quad p_i = n_i kT = x_i p ; \quad p = nkT$$

值得注意, 每个分子的质量  $m_i$  或能量自由度数  $i_i$  并不影响压强的计算. 压强单方面地决定于分子数密度与温度(分子平均平动动能. 分子质量改变的是整个体系的摩尔质量  $\mu$ , 取一定体积的气体:

$$M = \nu \mu = \sum M_i = \sum \nu_i \mu_i \Rightarrow \mu = \sum x_i \mu_i$$

其中出现的摩尔数  $\nu_i$  之比即为分子数密度  $n_i$  之比, 而摩尔质量  $\mu_i$  与分子质量的换算关系也很基础, 请读者一定要熟练使用:

$$\nu_i : \nu_j = \frac{n_i V}{N_A} : \frac{n_j V}{N_A} = n_i : n_j ; \quad \mu_i = N_A m_i$$

能量自由度数影响的是摩尔内能  $u$ :

$$u = \nu \overline{C_{mV}} T = \sum u_i = \sum \nu_i C_{mVi} T \Rightarrow \overline{C_{mV}} = \sum x_i C_{mVi} = \sum \frac{x_i f_i}{2} R$$

上面我们定义了平均摩尔定容热容  $\overline{C_{mV}}$ . 平均的含义一定是按分子数  $n_i$ (摩尔数  $\nu_i$ , 比例  $x_i$ ) 平均.

<sup>13</sup>如混合NH<sub>3</sub>与Cl<sub>2</sub>.

### 10.3.4 理想气体的过程

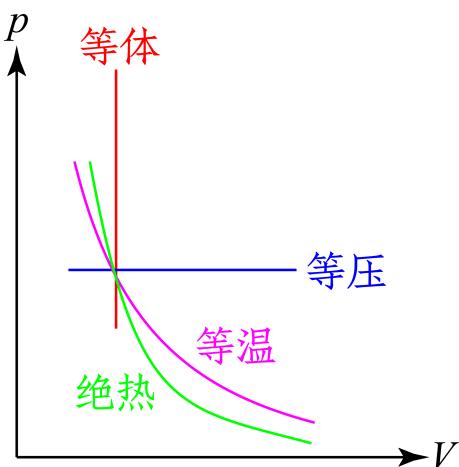


图 10.6: 四大准静态过程

所谓**过程**(process) 是状态的集合. 我们考虑的范围内, 过程的初态和末态都需要是均匀的平衡态. 中间态如果也可以视为平衡态, 就叫做**准静态过程**(quasi-static process), 若中间态蕴含着非平衡, 则称为**非准静态过程**(non-quasi-static process). 下面列举常见的过程:

#### 等体过程 (isochoric process)

作为等体过程体积功为零  $W = 0$ , 只能以从外界吸放热  $Q$ (或是如图10.2所示的功变热形式.). 系统吸热导致温度  $T$  升高, 定义系统**热容**(heat capacity) 为:

$$C = \frac{dQ}{dT}$$

而压强  $p$  将升高, 定义**压强系数**(pressure coefficient) 为:

$$\beta := \frac{1}{p} \frac{dp}{dT}$$

根据理想气体的物态方程与内能公式我们算出<sup>14</sup>:

$$C_V = \left( \frac{dQ}{dT} \right)_V = \nu C_{mV} ; \quad \beta_V = \frac{1}{p} \left( \frac{\partial p}{\partial T} \right)_V = \frac{1}{T}$$

#### 准静态的等压过程 (isobaric process)

若是控制压强  $p$  而不是体积  $V$  不变, 区别在于气体从外界吸收的热量中一部分要用来对外做功:

$$dU = dQ - pdV$$

热容按同样的方法定义, 分别称为等体热容和等压热容<sup>15</sup>. 而**热膨胀系数**(thermal expansion coefficient) 定义为:

$$\alpha := \frac{1}{V} \frac{dV}{dT}$$

注意这实际上是**体膨胀系数**(volumetric thermal expansion), 由近似方法可知, 近似为线膨胀系数  $\alpha_L$  的三倍. 对于理想气体等压过程, 同理可得:

$$C_p = \left( \frac{dQ}{dT} \right)_p = \nu(C_{mV} + R) = \nu C_{mp} ; \quad \alpha = \frac{1}{V} \left( \frac{\partial V}{\partial T} \right)_p = \frac{1}{T}$$

#### 准静态的等温过程 (isothermal process)

系统与恒温大热库接触, 吸热膨胀, 压强减小; 放热收缩, 压强增大. 这时  $pV$  为过程不变量. 我们定义**压缩系数**(compression coefficient):

$$\kappa := -\frac{1}{V} \frac{dV}{dp}$$

注意若是在固体情形, 以上量被称为**体模量**(bulk modulus). 对于理想气体, 等温压缩系数为:

$$\kappa_T = -\frac{1}{V} \left( \frac{\partial V}{\partial p} \right)_T = \frac{1}{p}$$

<sup>14</sup>热力学的角标表示这个量不变, 而这里热容的表达式下一章学了熵的概念后可以写:

$$C = T \frac{\partial S}{\partial T}$$

<sup>15</sup>中文太灵活, “等体” 也有人说成“定体”, “等容”, “定容”. 统一术语十分困难.

## 准静态的绝热过程 (adiabatic process)

绝热过程中  $dQ = 0$ , 对于理想气体:

$$\begin{aligned} dU + pdV &= 0 \\ \nu C_{mV} dT + \frac{\nu RT}{V} dV &= 0 \\ C_{mV} \frac{dT}{T} + R \frac{dV}{V} &= 0 \end{aligned}$$

惯例上定义绝热指数(adiabatic index) $\gamma$  为:

$$\gamma := \frac{C_{mp}}{C_{mV}} = 1 + \frac{R}{C_{mV}} = \frac{f+2}{f}$$

则可以由  $\gamma$  和  $R$  表示两个摩尔热容:

$$C_{mV} = \frac{1}{\gamma-1} R ; \quad C_{mp} = \frac{\gamma}{\gamma-1} R$$

从而我们从绝热过程的热力学第一定律方程积分得:

$$TV^{\gamma-1} = \text{Const.}$$

也可以写为:

$$pV^\gamma = \text{Const.} ; \quad \frac{p^{\gamma-1}}{T^\gamma} = \text{Const.}$$

前一式可以用于在  $p-V$  图 10.6 上确定过程曲线. 后一式则注重关心强度量之间的变化关系而不关心广延性质. 在绝热过程中 (用角标  $S$  表示绝热,  $S$  其实表示熵, 具体原因见下一章.), 热膨胀系数, 压强系数与压缩系数分别为:

$$\alpha_S = \frac{1}{V} \left( \frac{\partial V}{\partial T} \right)_S = -\frac{1}{\gamma-1} \frac{1}{T} ; \quad \beta_S = \frac{1}{p} \left( \frac{\partial p}{\partial T} \right)_S = \frac{\gamma}{\gamma-1} \frac{1}{T} ; \quad \kappa_S = -\frac{1}{V} \left( \frac{\partial V}{\partial p} \right)_S = \frac{1}{\gamma} \frac{1}{p}$$

## 准静态的多方过程 (polytropic process)

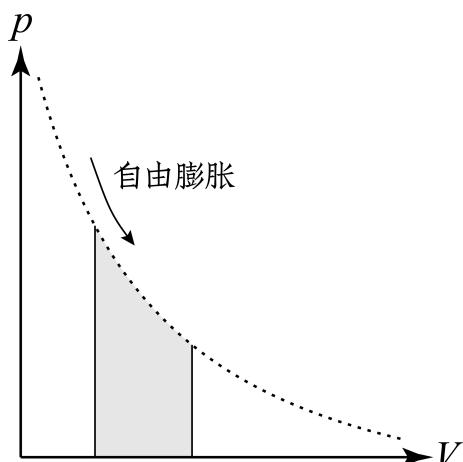
多方过程可以由过程方程  $pV^n = \text{Const.}$  描述, 多方指数  $1 < n < \gamma$ . 实际过程若发生缓慢, 可以近似处理为某种形式的多段多方过程. 多方过程的热容为:

$$C_{mn} = \left( \frac{1}{\gamma-1} - \frac{1}{n-1} \right) R$$

而其他响应函数 (指上面定义的  $\alpha, \beta$  与  $\kappa$ ) 的形式与绝热过程算出来的类似, 只需要把  $\gamma$  改为  $n$ . 多方过程在特殊的参数条件下即变成以上过程:

绝热指数 $n$	摩尔热容 $C_m$	描述
$n < 0$	$\left( \frac{1}{\gamma-1} + \frac{1}{1-n} \right) R$	吸收大量热量, 压强体积温度同时增加
$n = 0$	$C_{mp}$	等压过程
$0 < n < 1$	$\left( \frac{1}{\gamma-1} + \frac{1}{1-n} \right) R$	相比吸热等温膨胀, 温度有升高
$n = 1$	$\infty$	等温过程
$1 < n < \gamma$	$-\left( \frac{1}{n-1} - \frac{1}{\gamma-1} \right) R$	膨胀降温时要从外界吸热
$n = \gamma$	0	绝热过程
$n > \gamma$	$\left( \frac{1}{\gamma-1} - \frac{1}{n-1} \right) R$	压缩气体同时给气体热量的过程
$n = \infty$	$C_{mV}$	等温过程

## 自由膨胀过程

图 10.7:  $p - V$  图上画自由膨胀

理想气体向真空的自由膨胀过程(free expansion process)是一类典型的非准静态过程. 在这样一个过程中, 由于外压为零, 外界对气体做功为零. 而系统绝热, 从而从外界吸热也为零. 故由于热力学第一定律, 内能不变. 再根据理想气体内能直接依赖于温度的性质, 我们发现: 理想气体的自由膨胀是一个非准静态的等温膨胀.

从而在  $p - V$  图上, 我们画一根等温线, 它是否代表向真空的自由膨胀过程? 我们需要画成虚线, 线上每一个点都表示平衡态, 而从一个点到另一个相邻的点中发生的过程则没有达到力学平衡, 从而不能视为准静态过程. 这一根虚线能够代表气体在自由膨胀过程中的状态, 因为即使不均匀, 整体的内能和可以确定出平均温度, 整体的体积和平均温度又可以确定出平均压强. 如果在某个时刻自由膨胀结束, 体系达到平衡态后其状态就落在了这根等温线上. 只不过, 这根线下面积此时也没有了实际意义, 因为气体对外做功应该是零.

## 向固定压强膨胀

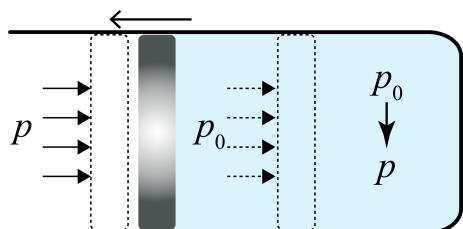


图 10.8: 向固定压强膨胀

考虑原平衡态压强为  $p_0$  的气体, 突然令外压变小为  $p$ , 气体将发生某种膨胀, 如果膨胀稍过度, 其压强小于外界压强  $p$  时, 又会被压缩, 这样来回振动, 直到达新的平衡态. 如果这个过程系统绝热, 则系统对外界做功应该写为:

$$W = p(V - V_0)$$

这一个模型可以这样理解: 它保证了外界压强  $p$  的严格不变性. 内部气体由于发生了急剧的体积变化, 最靠近活塞的部分气体压强由于急剧的膨胀实际上就会等于外界气压  $p$ . 这是忽略活塞质量(轻活塞)的情形. 若是考虑活塞的质量, 这个结论也还是成立, 因为内外气体对活塞做功之和转化为活塞动能. 而前后平衡态活塞实际上动能都是零. 故内部气体对外做功可以由外界固定压强部分气体的体积功来计算. 这就是上式. 那么同样, 由热力学第一定律和理想气体的性质, 我们有:

$$\frac{V}{V_0} = 1 + \frac{p_0 - p}{\gamma p} ; \quad \frac{T}{T_0} = \frac{1}{\gamma} + \left(1 - \frac{1}{\gamma}\right) \frac{p}{p_0}$$

如果我们在初态压强  $p_0$  与末态压强  $p$  之间搭设很多的压强平台  $p_i$ , 使得外压是缓慢地从  $p_0$  逐级减小到  $p$ , 那么用极限的方法容易证明(请读者自己完成), 过程恰好变为准静态的绝热过程:

$$pV^\gamma = p_0V_0^\gamma ; \quad \frac{p^{\gamma-1}}{T^\gamma} = \frac{p_0^{\gamma-1}}{T_0^\gamma}$$

此时屡次的非平衡态将无限趋于平衡态, 而非准静态过程也就无限趋于准静态过程了.

### 节流过程 (throttling)

节流过程亦是对很多复杂系统过程的抽象,如冷凝机内部工作物质所发生的循环的定常流动. 我们从中提取出一个模型,如下图,气体在内壁光滑的管道内做定常流动,通过一多孔塞后压强减小,气体膨胀. 整个过程绝热. 我们取一个固定的气团(图中红加蓝)作为研究对象. 经过一定时间  $t$  后左侧与右侧体积变化分别为  $V_1$  与  $V_2$ ,那么气体的流量被定义为:

$$Q = \frac{m_i}{t} = \frac{\rho_i V_i}{t} = \rho_i v_i A$$

由于是定常流动,左右两侧流量必须相等,这是质量守恒的要求,而体积流量  $v_i A$  可以不相等,因为气体可以被压缩.

下面我们考虑气体的热力学第一定律,由于  $Q$  小,我们忽略气体的整体动能(这一点在下一节我们将做详细讨论).

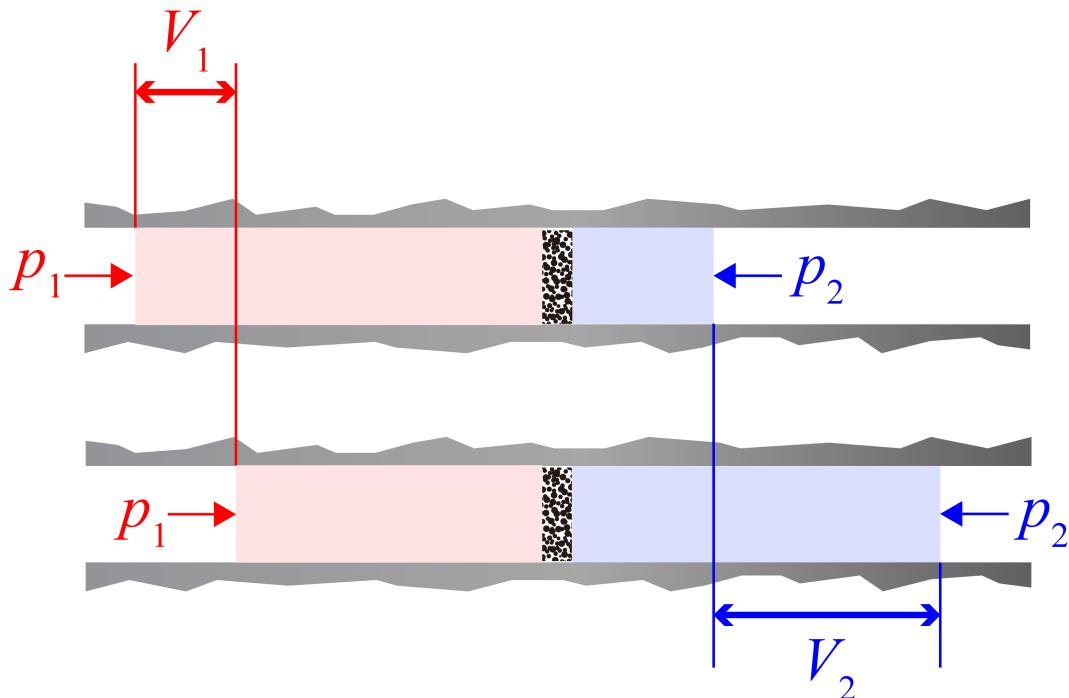


图 10.9: 多孔塞制节流阀

对于红 + 蓝的整体,热一定律表述为:

$$p_1 V_1 - p_2 V_2 = U_2 - U_1$$

即:

$$U_1 + p_1 V_1 = U_2 + p_2 V_2 ; \quad H_1 = H_2$$

也就是说,节流过程前后焓不变的过程. 若流过的气体是理想气体,那么

$$H = \nu C_{mV} T + pV = \nu(C_{mV} + R)T = \nu C_{mp} T$$

可见这也是一个温度不变的过程. 但是吸热,做功都为零,它并不是一个达到平衡的准静态过程.

## 10.4 开放系统的理想气体

开放的气体系统一般有两个特点,一是其性质一般被描述为:在某处  $r$  的气体性质怎样.而描述该气体性质的只能是各种强度量.如压强  $p(r)$ ,温度  $T(r)$  等等.二是气体往往处于动态的物质,能量的交换与流动中.这就意味着我们分析其物理规律时必须单独取出一个气团(parcel of air) 作为研究对象.研究周围气体对它可能造成的微观扩散,黏滞,传热过程,研究周围气体对气团压力造成的体积功与加速过程.下面我们由简入难地分析这些问题.

### 10.4.1 静态平衡问题——重力场中的大气

对于静态平衡的热力学体系,气团首先要符合的是静力学平衡条件:

$$\nabla p + f = 0$$

其中  $f$  代表体积力,我们考虑重力场系统,则为:

$$\nabla p + \rho g = 0$$

而密度反过来又同时和压强,温度有关:

$$\rho = \frac{\mu p}{RT}$$

于是,压强的分布纯粹由温度分布决定,而温度分布则看似成为了一个纯粹的热学问题,而且在一定范围内粗略地研究我们可以简单地假设温度随高度均匀地变化  $T(z) = T(0) - \Gamma z$ .然而这个问题并不简单,大体上如果讨论限于地球表面对流层的大气,其基本图像可以视为由太阳照射地表,地表升温而加热大气导致的.而对于不同的气候,地理环境,时间与天气条件这个温度随高度递减的方式都十分地不相同,我们也还是定义环境温度递减率(environmental lapse rate) 来线性表示温度的降低:

$$\Gamma = -\frac{dT}{dz} ; \quad T(z) = T(0) - \Gamma z$$

了解这一点后便可计算出其压强分布:

$$p(z) = p(0) \cdot [1 - \frac{\Gamma z}{T(0)}]^{\frac{\mu g}{R}} = p(0) \cdot [\frac{T(z)}{T(0)}]^{\frac{\mu g}{R}}$$

当环境递减率取某些不同的值时大气的性质是不同的,我们考虑其对流性质.考虑一个干的气团,让它在垂直方向上发生微扰,认为力学平衡总是很快地达到——气团压强始终等于外大气压.但热学平衡总是来不及达到——既使气团温度与外大气不相等,它与外界的热量交换可以忽略,从而可以发现,气团做绝热碰撞:

$$p(z) = p(0) \cdot [\frac{T'(z)}{T(0)}]^{\frac{\gamma}{\gamma-1}}$$

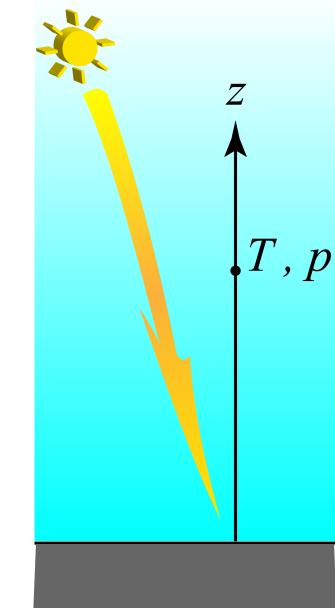


图 10.10: 地球大气

比对气团温度  $T'(z)$  与外大气温度  $T(z)$ ,我们发现有一临界的环境温度递减率(dry adiabatic lapse rate),它表征干燥的气团发生微扰后温度,密度恰好与周围大气相等的情况:

$$\Gamma_d = \frac{\gamma - 1}{\gamma} \frac{\mu g}{R} = \frac{\mu g}{C_{mp}} = 9.8^{\circ}\text{C}/\text{km}$$

若  $\Gamma > \Gamma_d$ ,则环境递减率过高,向上微扰的气团温度高于环境温度,所受浮力偏大,将偏离平衡位置,产生热对流,成云致雨.反之, $\Gamma < \Gamma_d$  时浮力偏小,气团受到回复力而稳定在平衡位置处,对应着稳定无对流的大气.对于地球大気体系,有对流与无对流的情况都有时发生,其平均递减率就在临界绝热递减率附近.

对于临界大气为何温度随高度线性递减,而递减率又与定压热容有直接的关联,有一种较为直接的证明方法.其思想为虚功原理结合下面经常使用的流管分析法.我们让从  $z = 0$  到  $z = h$  的截面积为  $A$  的气柱在顶部与底部的压力作用下向上发生虚位移,由于是临界平衡,故位移以后恰巧每一小块气柱  $Adz$  都与新的位置的环境大气性质全同,也就是与之前在这儿的那块气柱性质一样.整体上看,相当于把  $z = 0$  处的一块质量为  $\delta m$  的气团位移到了  $z = h$  处.写出热力学第一定律,要注意其本质为能量守恒,这里还要包括重力势能:

$$p(0)\delta V(0) - p(h)\delta V(h) = \delta U(h) - \delta U(0) + \delta mgh$$

而由焓的定义  $\delta H = \delta U + p\delta V$ , 以及理想气体的焓有公式  $\delta H = \delta\nu C_{mp}T$ , 我们得出:

$$\delta\nu C_{mp}[T(0) - T(h)] = \delta\nu\mu gh$$

即:

$$T(h) = T(0) - \frac{\mu g}{C_{mp}}h$$

对于实际地球对流层大气来说,还需要考虑空气中水蒸气受冷造成的液化放热现象,其临界递减率将会取决于气体含水蒸气的多少.而地球下层大气还分为对流层(troposphere) 与平流层(stratosphere),平流层含大量臭氧,太阳的紫外辐射将在传播的过程中被平流层逐渐吸收.所以平流层的上部具有较强的辐射和较高的平衡温度,而温度随高度递增,提供了一个稳定的大气环境,极难发生对流.

### 10.4.2 能量守恒——伯努利方程

我们把流体力学里讨论的伯努利方程(Bernoulli Equation) 推广到热力学中. 它适用条件为定常流动(steady flow) 中的同一条流线(streamline) 上. 它也是推广的热力学第一定律, 它代表能量的守恒.

不可压缩流体的伯努利方程的图像其实很清晰, 我们若不考虑流体势能的变化, 那么伯努利原理实际是在说, 两侧的压强差导致了流体的加速:

$$p_1 - p_2 = \frac{1}{2}\rho v_2^2 - \frac{1}{2}\rho v_1^2$$

从而沿流线, 以下量守恒:

$$p + \frac{1}{2}\rho v^2 = \text{const.}$$

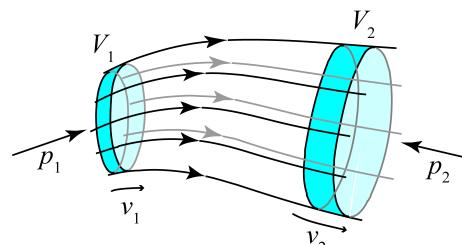


图 10.11: 气体沿一条流管流动

但若考虑可以压缩的气体, 以及考虑保守外力的做功, 是否有以上守恒量呢? 我们假设保守外力为质量力: 单位质量的流体的势能为  $\Phi$ , 那么对于一定时间间隔  $t$  下的流管发生的流动  $m = Qt$ , 有以下能量守恒:

$$p_1 V_1 - p_2 V_2 = (U_2 + \frac{1}{2}mv_2^2 + m\Phi_2) - (U_1 + \frac{1}{2}mv_1^2 + m\Phi_1)$$

两边同时除以质量, 得守恒律:

$$u + \frac{p}{\rho} + \frac{v^2}{2} + \Phi = \text{const.}$$

其中  $u + p/\rho = u + pv = h$  即为比焓, 而  $u$  为比内能,  $v$  为比体积, “比” 表示单位质量的某物理量, 它是一种强度量. 从而以上量我们还可以写为:

$$h + \frac{v^2}{2} + \Phi = \text{const.}$$

我们还定义所谓的阻滞焓(stagnation enthalpy) 的概念. 它表示让运动的气体发生某种虚构的定常流动绝热地停下来后气体应该有的焓. 那么比阻滞焓为:

$$h' = h + \frac{v^2}{2}$$

最后这一广义的伯努利方程,就被我们写为一个广义的两项的能量守恒形式:

$$h' + \Phi = \text{const.}$$

那么对于理想气体,我们还可以定义阻滞温度:

$$T' = T + \frac{\mu v^2}{2C_{mp}} ; \quad h' = \frac{1}{\mu} C_{mp} T'$$

而对于普通不可压缩流体,我们一般不关心其热学性质,故焓可以扣除内能项  $h = p/\rho + v^2/2$ ,而  $\rho$  为常数.这样我们又可以定义阻滞压强:

$$p' = p + \frac{1}{2}\rho v^2 ; \quad h' = \frac{p'}{\rho}$$

### 10.4.3 动量守恒——欧拉方程

让我们考虑管道内的理想气体流动,管道的截面积在不断减小的情况  $A = A(x)$ .可以想见,这会导致流速的增加,进一步导致压强的减小,对应方程为[连续性方程](#)(continuity equation)与伯努利方程,分别对应质量与能量守恒:

$$\rho v A = \text{const.}$$

$$\frac{1}{\mu} C_{mp} T + \frac{v^2}{2} = \text{const.}$$

以上方程还应联系物态方程理解:

$$\rho = \frac{\mu p}{RT}$$

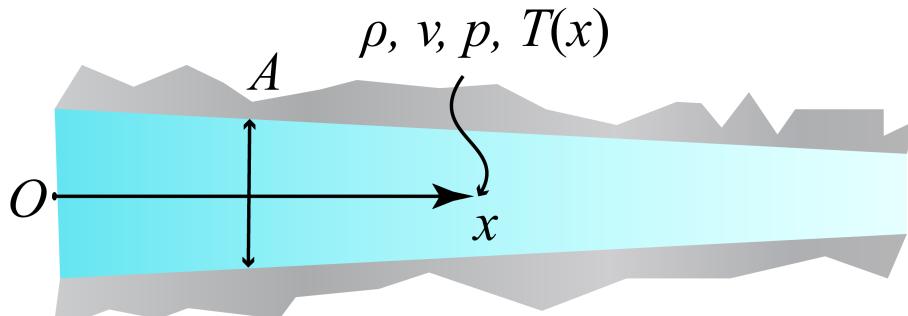


图 10.12: 截面积减小的管道内的气流

注意到我们把  $A$  的变化当做已知量以后,却仍设了待求解的四个物理量  $v(x), p(x), T(x), \rho(x)$ ,但目前却只有三个方程.那么我们漏了怎样一个方程呢?这个方程就是流体的[欧拉方程](#)(Euler equation),它是一个纯粹的动力学方程,不包含任何热效应.一般写作:

$$\frac{D\mathbf{v}}{Dt} = -\frac{\nabla p}{\rho} - \nabla \Phi$$

其中  $D$  代表[随体导数](#)(material derivative):

$$\frac{D}{Dt} = \frac{d}{dt} + \mathbf{v} \cdot \nabla$$

上式代表了压强梯度力与保守外力对流体的加速效应,在上述情形中没有外力所以等式右边只有一项.实际上,它代表了整个体系动量守恒.而对于上述情形,气体做定常流动,上式写为:

$$v \frac{dv}{dx} = -\frac{1}{\rho} \frac{dp}{dx}$$

这一个式子也可以由上一节的类似方法——取一定长度的气团来导出. 但切不可忽略管壁对气体在过程中作用的冲量:

$$p_1 A_1 \mathbf{e}_x \cdot t - p_2 A_2 \mathbf{e}_x \cdot t + \int -pd\mathbf{A} \cdot t = m\mathbf{v}_2 - m\mathbf{v}_1$$

左边恰巧是在整个表面的压强的面积分, 可以由高斯定理化为体积内的体积分:

$$\int_{x_1}^{x_2} -\frac{dp}{dx} \cdot Adx = Q(v_2 - v_1)$$

又考虑到以上积分式无法直接积出结果, 故对积分上限求导, 再考虑到  $Q = \rho v A$ , 化简后即得欧拉方程:

$$\frac{dp}{dx} = -\rho v \frac{dv}{dx}$$

对于管道截面积相对于长度变化缓慢的情况<sup>16</sup>, 联立以上四个方程, 解得:

$$\begin{aligned}\frac{dv}{v} &= -\frac{dA}{A} \left(1 - \frac{\mu v^2}{\gamma RT}\right)^{-1} \\ dp &= \rho v^2 \cdot \frac{dA}{A} \left(1 - \frac{\mu v^2}{\gamma RT}\right)^{-1} \\ dT &= \frac{\mu v^2}{C_{mp}} \cdot \frac{dA}{A} \left(1 - \frac{\mu v^2}{\gamma RT}\right)^{-1}\end{aligned}$$

可见, 随着截面积  $A$  的减小, 流速加快, 压强和温度都要减小, 利用这个效应可以制作纯粹依靠风力制冷的“土空调”. 注意到以上结论还依赖于一个条件:

$$\frac{\mu v^2}{\gamma RT} < 1$$

这意味着分子代表的气团集体运动能量要远小于分母代表的微观热运动能量, 这在一般情况下是很容易满足的, 因为  $v \sim 10\text{m/s}$  而热运动平均速率  $\bar{v} \simeq 500\text{m/s}$ , 若对于高速运动的气体, 其发生的过程中气团速度的改变与热运动平均速度可以比拟, 实际上把气团处理为平衡的理想气体也是不妥当的.

<sup>16</sup>即, 在所关心的范围内截面积的相对变化很小:

$$\frac{\Delta A}{A} \ll 1$$



# 第十一章 热力学第二定律

热力学第一定律固然强大,它根本上杜绝了所谓的**第一类永动机**(perpetual motion machine of the first kind) 的可能性,也排除了一大类不现实的热力学过程的发生.然而由这样一个只关心整体性质,不深入到微观本质的热力学规律又能告诉我们多少呢? **炼金术**(alchemy) 不违背热力学第一定律,它企图实现物质与物质之间的随意转化, **第二类永动机**(perpetual motion machine of the second kind) 亦不违背热力学第一定律,它基于能量守恒,主张同一份能源的无限再生利用.对于前者,微观机理彻底地否定了它,不同的原子有不同的构成,原子核的结构十分稳定,至今无法操控.而对于后者,统计规律也彻底地进行了否定,它告诉我们,一个体系的过程总是有确定的方向性,在一定条件下,过程被我们分类为**自发过程**(spontaneous process) 与**不自发过程**(non-spontaneous process),具有特定方向性的过程一旦发生,其时间反演的过程就将不现实.但十分神奇的一点是,不同体系的不可逆特性居然是紧密相关的,若将一个过程的不可逆性与另一个不可逆过程的不可逆性建立联系,我们就能从中同时发现统一的热力学温标与态函数熵这两个概念,从而量化过程的不可逆性,这一套宏观的规律称为热力学第二定律.它道出了热力学过程的方向性,本质是微观统计规律,限制了另一大类不现实的热力学过程的发生.

## 11.1 循环过程

在历史上,热力学基本规律的确立与发展和热机的使用与改进是紧密相关的.直到今天卡诺循环的引入仍然不失为介绍热力学第二定律理论体系的强有力工具.本节将介绍热机,热泵的基本模型与著名的卡诺循环.为之后引入热力学第二定律相关概念打下基础.

### 11.1.1 热机与热泵

上一章介绍过循环过程的概念.我们介绍了  $p - V$  系统的正负循环.现在我们把这样的  $p - V$  体系看成很多实际生活中一些机器的工作过程的抽象.其实际承担了状态循环变化的体系称为**工作物质**(working body).而其联动的**机械装置**(mechanical device) 可以与外界交换功.最后,这样的体系还需要可以与其交换热量的对象,凡是工作物质从中吸热的被称为**热源**(heat source),而工作物质向之放热的被称为**冷库**(cold sink).这样我们就抽象出了一个理论模型,根据热力学第一定律,如果做正循环,工作物质一个循环将从高温热源吸收更多的热量,向低温冷库放更少的热量,而通过机械装置输出功,故被称为**动力循环**(power cycle),对应机器为**热机**(heat engine).而如果考虑相反的过程<sup>1</sup>,工作物质一个循环向高温热源放热,从低温冷库吸热,机械装置要对它做功,则是一个**泵热循环**(heat pump cycle).对应机器被称为**热泵**(heat pump).

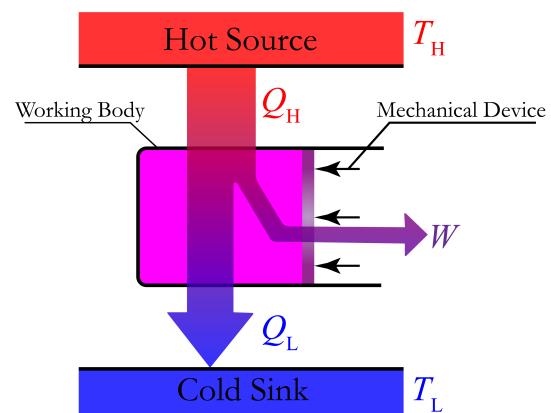


图 11.1: 热机模型

<sup>1</sup>这样的过程不一定可逆,这里指的是那些定性现象相反的循环.



图 11.2: 俄罗斯巴拉科沃 (Balakovo) 核电站汽轮机

根据以上原理,热机主要分为以下两大类:

**蒸汽机 (steam engine)** 蒸汽机是一类外燃机(external combustion engines),工作物质是在气相与液相混合的水(水蒸气).用外界燃烧的热量,或者来自太阳能,核能,地热等的非燃烧热量加热锅炉(boiler)中的水,由此产生高温高压蒸汽,用蒸汽带动机械装置运转向外输出能量.经过这样的装置后(这中间一般视为节流过程),水蒸气一般变为过饱和的冷气体或者还夹杂着水滴,进一步导入冷凝器而液化变为液态水回收进入锅炉.完成一个循环.18世纪末与大半个19世纪可谓“蒸汽时代”,蒸汽机在大型机械生产工厂,蒸汽机车,蒸汽机轮船等交通工具中大放异彩.虽然石油开采为小型汽车生产提供了可能,大型交通工具和生产业动力来源之后也慢慢地几乎被内燃机与电力所取代.但直到今天,我们的生产生活有90%以上的电量都来自于汽轮机(steam turbine)形式的热电厂.值得一提的是其中化石能源(煤电,气电,以及较为小规模的油电)仍然占据75%的高比重(中国与美国的情况大致相同),新型能源的比重不大,但在稳步上升中.

**内燃机 (internal combustion machine)** 蒸汽机的锅炉设计让燃烧能量给了流动的工作物质,实现了工作物质不含燃料,不与燃料混合或被燃料污染.这有效地提高了燃料能量利用的效率.然而却需要为工作物质准备额外的空间,不利于把热机做小.与之相反,内燃机利用了燃烧室(combustion chamber)的设计,工作物质直接是高能量密度的燃料和氧化剂(一般就是空气).利用高温高压下点燃化学反应释放的巨大化学能来产生热,导致气体反向对机械装置——活塞(piston)或涡轮(rotor)做功.与外燃机工作物质发生的真循环过程不同,内燃机发生的过程是一种开放循环.以常规的四冲程活塞式内燃机为例,气体经历四个冲程:吸入(intake),压缩(compression),做功(power),排出(exhaust).整个循环从常态燃料中开始,最后以排出气体在大气中冷却结束.而燃气涡轮发动机(gas turbine engine)更是针对四个阶段分别设计功能区,使得四个阶段同时进行,提供了极大的功率输出.按照其机械装置的不同分为四类:军用战斗机往往采用涡喷发动机(turbojet engine),而民航客机往往是涡扇发动机(turbofan engine);运输机,巡逻机经常采用涡桨发动机(turboprop engine);而直升机和其他陆上,海上交通工具则通过涡杆发动机(turboshaft engine)将动力用传动轴导出.在这样的设计中,燃料产生的能量用来

驱动涡轮带动整个装置大量吸入外界空气<sup>2</sup>并有效地压缩气体从而产生动力. 涡轮转动的能量也正是很多民航飞机上电力的来源.

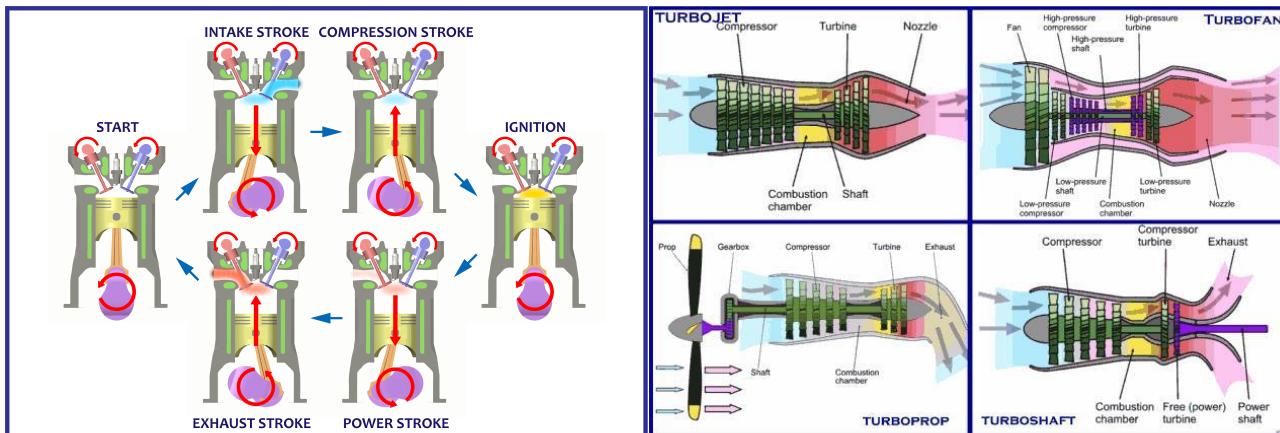


图 11.3: 左: 四冲程活塞内燃机; 右: 四类涡轮机

我们发现, 将实际的热机抽象为前图11.1那样的模型往往是过分地简单了. 实际问题中的工作物质往往成分在发生改变, 有物质的输入输出, 甚至无法视为一个完整的循环过程. 但理论可以帮我们分析与把握这个过程的一些主要信息. 在实际问题中热机的能效(energy efficiency) 是主要的关心因素之一, 它定义为输出的有用功  $W$  与燃料提供的化学能  $Q_H$  之比:

$$\eta = \frac{W}{Q_H}$$

而热泵也根据其设计功能分为制冷机(cooling machine), 制热机(heating machine) 与可逆机(reversible machine). 可逆的热泵同时具备制冷制热两种功能, 采用换向阀来控制气流的流向从而改变效果. 当采用制热模式时, 其效率往往比使用单独的电阻加热高 3 到 4 倍.

无论何种类型的热泵, 其构成与原理可以被抽象为典型的四步. 高温热源处更高温的气体在冷凝器(condenser) 处变为液态而放热; 膨胀阀(expansion valve) 处的节流过程使液态流体温度与压强大降而成为气液混合的状态; 低温冷库处稀薄的半液态流体在蒸发器处蒸发吸热; 最后通过压缩机(compressor) 泵入冷凝管, 压强与温度显著升高. 这样的循环称为斯特林循环(Stirling cycle).

而对于热泵的性能系数(coefficient of performance), 需要对两种不同的模式区分:

对于制冷模式, 取低温冷库处吸热比输入功率:

$$\eta = \frac{Q_L}{W}$$

对于制热模式, 取高温热源处放热比输入功率:

$$\eta = \frac{Q_H}{W}$$

两者都可能大于一.

<sup>2</sup>火箭除外, 它需要自备燃料与氧化剂.

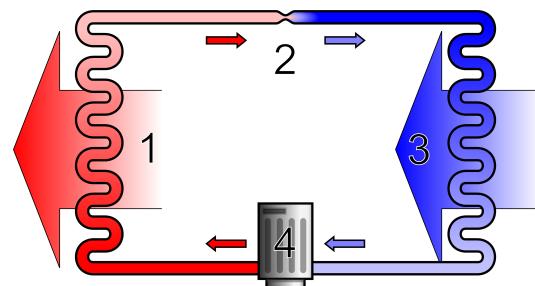


图 11.4: 热泵的普遍构成

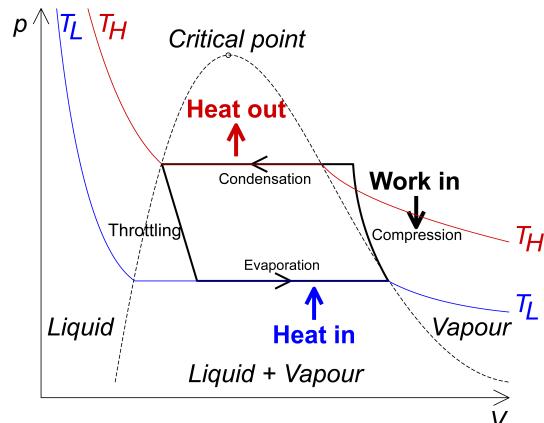


图 11.5: 热泵的工作循环

### 11.1.2 热机循环

从瓦特 (Watt) 改良蒸汽机开始,一代代的工程师开始致力于提高热机的效率.这其中便产生了大量实用的热机模型.直到 1824 年法国物理学家与工程师卡诺 (Carnot) 给出了惊人的卡洛热机模型与关于热机工作效率上限的论断.这实际上意味着热力学(thermodynamics) 的诞生.之后热机优化建立在严格的热力学计算的基础上.发展到今天理论早已成熟,只待技术的发展与革新.

简单的热机循环计算往往将工作物质视为理想气体在  $p-V$  图上的准静态过程构成的循环.下将一一列举:

#### 勒鲁瓦循环 (Lenoir cycle)

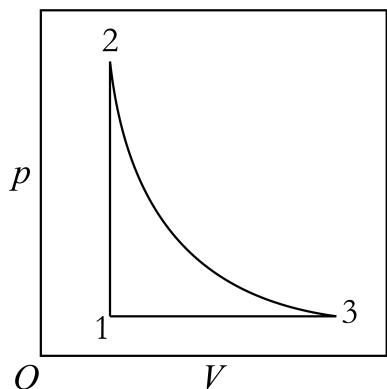


图 11.6: Lenoir cycle

1860 年法国工程师勒鲁瓦发明的勒鲁瓦机是第一个为商业设计的内燃机.其对应循环常常用来描述脉冲喷气发动机(pulse jet engine) 的模型.它区别于其他热机的特点是无需压缩气体.一个循环内,理想气经历:

- i 1-2: 定容爆炸吸热
- ii 2-3: 绝热膨胀
- iii 3-1: 定压放热

定义爆炸比(explosion ratio) $r_p$ :

$$r_p = \frac{p_2}{p_1}$$

则可以计算出其效率为:

$$\eta = 1 - \gamma \frac{T_3 - T_1}{T_2 - T_1} = 1 - \gamma \frac{r_p^{\frac{1}{\gamma}} - 1}{r_p - 1}$$

#### 奥托循环 (Otto cycle)

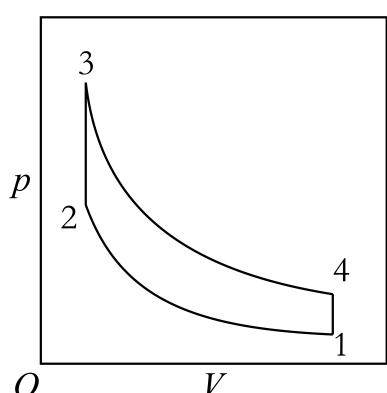


图 11.7: Otto cycle

这是最典型的四冲程型汽油机(gasoline engine, gas engine) 的工作循环.特点是通过压缩气体产生高温高压易燃汽油空气混合体,再通过火花塞点燃混合体,过程可以视为瞬间发生的,从而体积不变,压强突然增大.一个循环内,理想气经历:

- i 1-2: 绝热压缩
- ii 2-3: 定容爆炸吸热
- iii 3-4: 绝热膨胀
- iv 4-1: 定容放热

定义压缩比(compression ratio) $r$ :

$$r = \frac{V_1}{V_2}$$

则可以计算出其效率为:

$$\eta = 1 - \frac{T_1}{T_2} = 1 - \frac{T_4}{T_3} = 1 - r^{-(\gamma-1)}$$

对于空气与燃油的混合气  $\gamma \approx 1.3$ ,由于爆炸燃烧机制,目前的压缩比能做到  $r \approx 10$ ,这样看其理论效率上限  $\sim 50\%$ ,目前一般的汽油机效率为  $25\% \sim 35\%$ .最高能到  $40\%$ ,离理论上限还尚有一段提升空间.

### 迪塞尔循环 (Diesel cycle)

迪塞尔机又称为**柴油机**(Diesel fuel engine). 其与汽油机设计上存在区别. 为了提高汽油机的效率, 柴油机企图增大压缩比, 故采用粗重的活塞. 并取消汽油机中点火爆炸的过程, 而让压缩冲程的末温直接使得燃料能够燃烧, 而缓慢注入燃料以使得燃烧充分. 这样, 典型的迪塞尔循环可以认为是这样四个过程:

i 1-2: 绝热压缩

ii 2-3: 定压燃烧吸热

iii 3-4: 绝热膨胀

iv 4-1: 定容放热

定义压缩比  $r$  与**停气比**(cut off ratio) $\alpha$ :

$$r = \frac{V_1}{V_2} ; \quad \alpha = \frac{V_3}{V_2}$$

则可以计算出其效率为:

$$\eta = 1 - \frac{1}{\gamma} \cdot \frac{T_4 - T_1}{T_3 - T_2} = 1 - \frac{\alpha^\gamma - 1}{\gamma r^{\gamma-1} (\alpha - 1)}$$

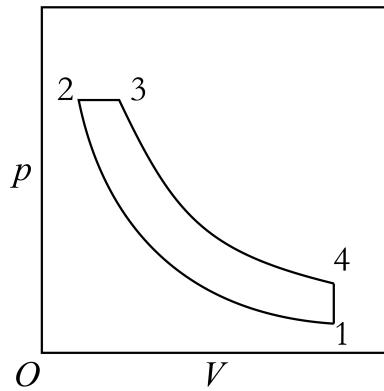


图 11.8: Diesel cycle

### 布莱顿循环 (Brayton cycle)

涡轮机等开放式热力学循环的一般抽象是带有两个等压过程的布莱顿循环. 可以认为是这样四个过程:

i 1-2: 绝热压缩

ii 2-3: 定压燃烧吸热

iii 3-4: 绝热膨胀

iv 4-1: 定压放热

定义**压强比**(pressure ratio) $r_p$ :

$$r_p = \frac{p_2}{p_1}$$

则可以计算出其效率为:

$$\eta = 1 - \frac{T_1}{T_2} = 1 - r_p^{-\frac{\gamma-1}{\gamma}}$$

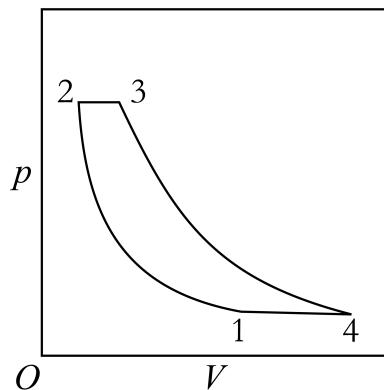


图 11.9: Brayton cycle

### 卡诺循环 (Carnot cycle)

卡诺循环是一种具有极其重要的理论意义的理想循环。它不能对应实际应用与生活与生产中的实际热机。但是却像直线与平面之于几何理论那样，对卡洛循环与热机的很多讨论都将与热力学理论体系紧密相关。对于理想气体，卡诺循环意味着其吸放热都发生在与两个恒温热源做等温接触的过程中。

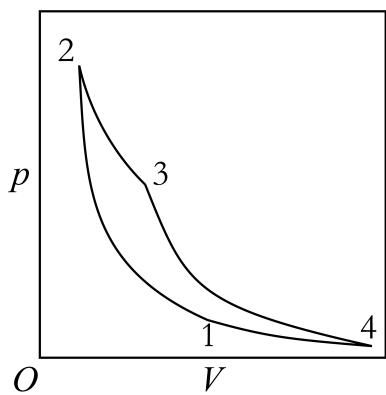


图 11.10: Carnot cycle

整个循环由两条绝热线与两条等温线围成。过程如下：

- i 1-2: 绝热压缩
- ii 2-3: 等温吸热
- iii 3-4: 绝热膨胀
- iv 4-1: 等温放热

高温热源的温度为  $T_H$ ，低温冷库的温度为  $T_L$ 。那么可以计算出卡诺热机的效率为：

$$\eta = 1 - \frac{T_L}{T_H}$$

## 11.2 理想气体的熵

理想气体的准静态过程是这一章内我们关心的问题。如何判断一个微元准静态过程中理想气体吸热还是放热？具体来说，我们关心在  $(T, V)$  点沿着  $(dT, dV)$  变化方向改变状态所需要吸收的热量微分  $dQ$ 。由热一，它应为：

$$dQ = dU + pdV = \nu C_{mV} dT + \frac{\nu RT}{V} dV$$

这可以进行如下的化简：

$$dQ = \nu C_{mV} T \cdot \left[ \frac{dT}{T} + (\gamma - 1) \frac{dV}{V} \right] = \nu C_{mV} T \cdot d \ln(TV^{\gamma-1})$$

我们定义态函数——**熵**(entropy)：

$$S_1 = \nu C_{mV} \ln(TV^{\gamma-1})$$

那么自然得出：

$$dQ = T dS_1$$

这给出了气体吸放热与状态变化之间的联系。一个直接的推论就是，准静态绝热过程中  $TV^{\gamma-1}$  是不变的（绝热方程）。或者说，准静态绝热过程可以被称为**等熵过程**(isentropic process)。而熵作为一个态函数，等熵线就是绝热线。

熵在定义时可以相差一个与状态无关的常数（但与摩尔数  $\nu$  有关，目前它不随着状态变化而变化）。我们为了方便暂时不写这个常数。但如果一开始以  $(p, V)$  或者  $(p, T)$  作为状态参量。那么自然而然写出来的符合  $dQ = T dS$  的熵为：

$$S_2 = \nu C_{mV} \ln(pV^\gamma) \quad ; \quad S_3 = \nu C_{mV} \ln \frac{T^\gamma}{p^{\gamma-1}}$$

它们都相差一个过程不变的常数：

$$S = S_3 \quad ; \quad S_1 = S + \nu R \ln(\nu R) \quad ; \quad S_2 = S + \nu C_{mp} \ln(\nu R)$$

我们把  $S_3$  取为标准的熵。原因在于，如果我们要讨论摩尔数  $\nu$  变化的情况时。在对数里放两个强度量可以使表达式成为一个广延量——当摩尔数增大到  $\lambda$  倍而所有强度量不变时，熵  $S$  也增大到  $\lambda$  倍。而其他两个  $S$  附加的项  $\nu R \ln(\nu R)$  与  $\nu C_{mp} \ln(\nu R)$  使得定义失去了这种广延性。

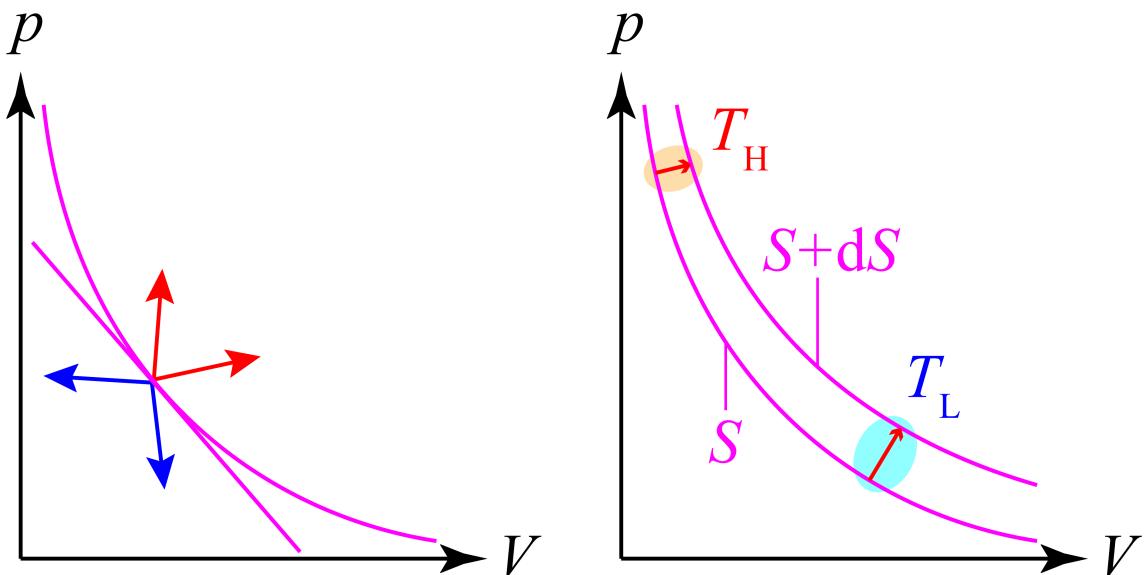


图 11.11: 左: 吸放热的元过程方向; 右: 两根等熵线间的“热量差”

我们由此得出这样的结论: 在一个准静态过程中. 沿着等熵线以下的方向熵降低 (如图11.11中蓝色箭头方向), 从而要放热. 而往上走则熵增加, 要吸热. 但注意到既使沿着切线方向走熵还是会降低 (无论往上还是往下), 只不过这时候其降低是一个二阶小量而已. 也就是说如果气体历经  $p - V$  图上的负斜率直线过程, 一定是先吸热, 后放热. 在中间有一个熵的极值点符合:

$$k = \frac{dp}{dV} = -\gamma \frac{p}{V}$$

我们画出两条相隔很近的等熵线 (差  $dS$ ). 注意到两条线之间的“热量差”这个概念本身不够严谨, 因为热量是过程量不是状态量. 但由于两条等熵线相隔很近, 所以从一点出发到另一条绝热线上的点还在这一点附近 (微元过程, 取  $\Delta S \rightarrow 0$  的极限). 所以可以认为温度近似不变. 尽管这样, 吸热仍然与过程发生的平均温度  $T$  有关:

$$Q_H = T_H dS \quad ; \quad Q_L = T_L dS$$

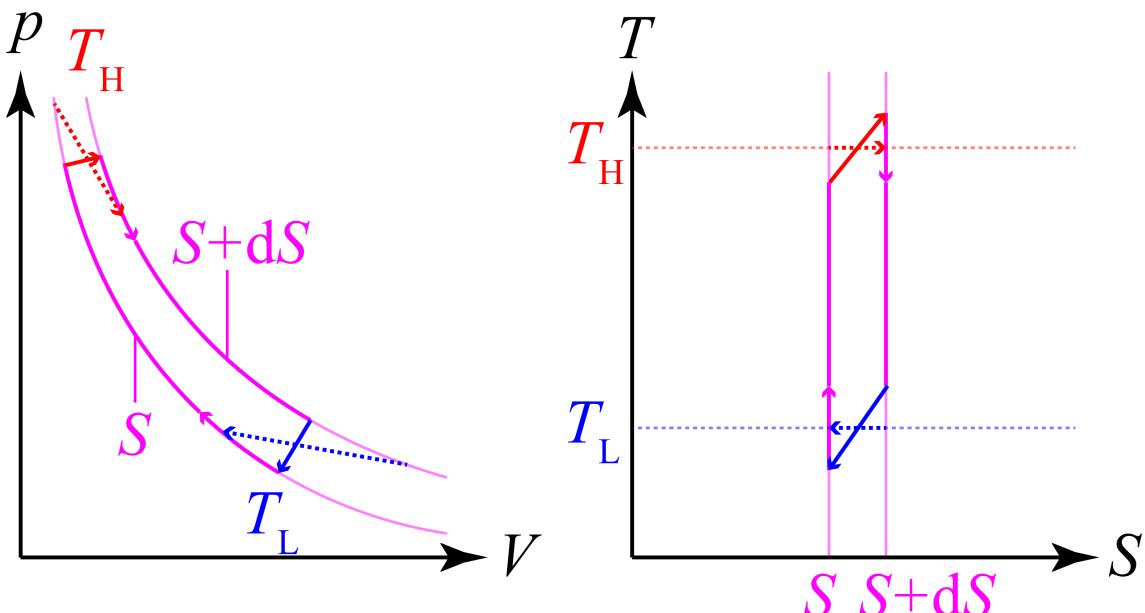


图 11.12: 示功图与示热图

由于  $S$  是态函数, 我们也可以选取  $(T, S)$  来作为体系的状态参量. 而做  $T - S$  图表示状态与过程. 此时等温线实际上就是水平线, 而绝热线就是竖直线. 而其包围的面积实际上就是整个过程所吸收的热量:

$$Q = \oint T dS$$

而由于热一, 它一定等于  $p - V$  图中的面积, 它表示对外做功:

$$W = \oint p dV$$

正是由于这个直观的关系, 热力学理论分析时我们经常对照两个图像. 分别叫做示功图 (work indicator diagram) 与示热图 (heat indicator diagram).

还有两点值得讨论:

第一, 以上讨论仅仅限于理想气体的准静态过程. 我们发现对于理想气体, 如果采用理想气体温标 (即分子内能恰正比于温度), 恰好可以引入独特的态函数熵来计算其吸热. 然而, 对于非准静态过程初末态之间的联系我们还一无所知, 仅仅知道热一总是成立的.

第二, 对于其他体系是否有  $dQ = T dS$ ? 这看上去似乎不显然. 首先我们知道热力学第零定律定义了温度是某种热平衡的系统共同具有的属性, 但我们一直使用的温度似乎是根据理想气体的分子平均动能定义出来的 (理想气体温标):

这样, 直接把这个温度用于其他体系似乎不太可能会得出相似的结论, 而另外定义新的温标又会造成物理规律的不统一性. 温度也就失去了作为互为热平衡态共同参数的性质. 然而, 需要知道上式不是无来由的, 它来源于麦克斯韦分布律——微观体系的经典动力学与统计规律的共同结果, 在这一点上其实温度的定义是普适的, 它都反映了各个自由度上热运动的剧烈程度. 所以实际上存在普适的热力学温标, 而同样的也将存在对应的普适的态函数熵. 它们都由大量微观粒子的统计规律给出. 然而十分神奇的一点是, 普适的温度与熵的存在性在历史上和逻辑上的确是可以由简单的公理得出的热力学第二定律的体系. 这正是下一节我们要说明的. 而要深刻了解温度与熵的微观本质需要深入了解统计力学方法, 我们在下一章将给出相关的论述.

$$T = \frac{2\bar{\varepsilon}_k}{3k}$$

### 11.3 热力学第二定律

费曼说过, 所有自然现象的最自然特征莫过于它们明显的不可逆性. 大量微观粒子构成的宏观热力学体系在种种现象上体现出不可逆的特性, 即热力学时间箭头 (time arrow of thermodynamics). 如果令时间反演, 系统的各状态逆序发生, 外界对其做功变为对外做功, 向外界放热变为从外界吸热. 那么很多现象都给出了从未被观察到的结果: 混合均匀地两种气体自发地分开, 低温冷库向高温热源自发放热, 热能单一地转化为机械能等等. 无耗散的准静态过程是理想的可逆过程 (reversible process), 仅有系统内部处处达到细致平衡, 在与外界平衡的状态下统一地改变其状态才能将其过程反演而符合物理规律. 而广泛存在的不可逆过程 (irreversible process), 一方面在于过程反演的不实际性, 还在于如果我们把系统和外界各部分视为一个孤立的体系, 那么发生的不可逆过程无法通过任何可能的过程回到初始状态. 在发生不可逆过程的时候, 体系的某些特征沿着一个方向被永久而不可恢复的改变了, 衡量这个特征的物理量就是熵, 而这种断言就是热力学第二定律. 它回答了热一所不能回答的问题, 给出了热力学过程发生的方向性与限度. 对于这些问题的讨论是热二的特色.

热力学第二定律在逻辑体系上承接热力学第零, 第一定律, 我们现在取消前面已经有良好定义的理想气体与针对它的温标的概念. 而热机模型与其效率只取决于吸放热仍然有良好的定义:

$$\eta = 1 - \frac{Q_L}{Q_H}$$

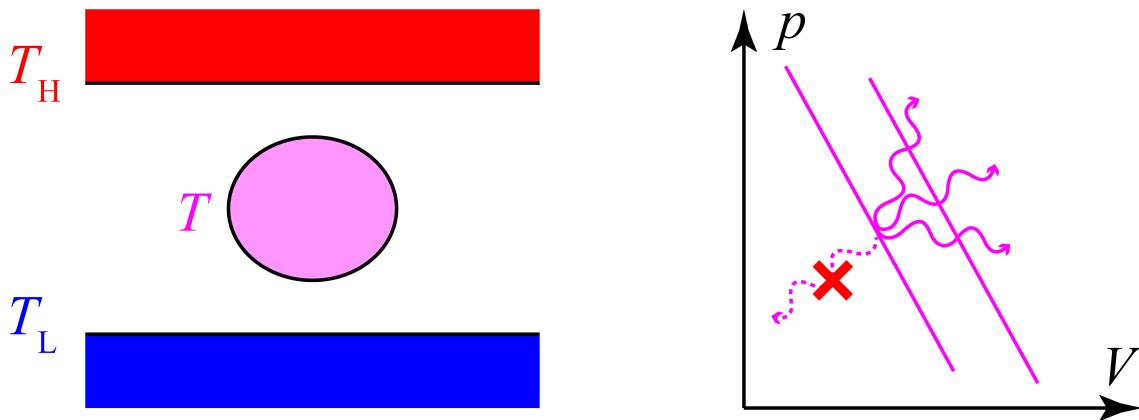


图 11.13: 熵增原理

我们讨论一个  $p - V$  系统作为工作物质在两个恒温热源 (冷库在此被称为低温热源) 之间的工作, 各个温度暂时只是一个记号, 其值暂不确定. 为了统一起见, 我们研究系统与外界的功热交换时规定  $dU = dW + dQ$ , 即以从外界吸热和外界对其做功为正. 首先, 我们孤立那个  $p - V$  系统, 让它做绝热过程而只能与外界有功的交换. 那么准静态的绝热过程<sup>3</sup>将给出系列绝热线, 它们彼此绝不可以相交, 否则在一点处将存在多个绝热的方向. 图中为了简化而近似画成了平行直线 (这在微元的循环过程中——微小吸放热, 微小温差——是合理的处理方法). 现在考虑一个非准静态的绝热过程. 如果是绝热压缩, 那么外界给予气体的压强应该大于系统的平均压强, 这样造成的不平衡才可以使得系统自发收缩. 如果是绝热膨胀, 同理, 外界给予气体的压强应该小于系统的平均压强. 不管怎样, 都有:

$$dW \geq -pdV$$

这样造成的结果是系统的状态只能往绝热线的一侧走, 类似于热力学第零定律, 对于这个特定体系, 我们暂时可以把这些可以通过准静态绝热过程相互连接的状态称为具有共同的熵. 而更高的绝热线代表更高的熵, 以上观点给出了**熵增原理**(principle of entropy increase):

**熵增原理: 孤立系统的熵只增不减,  $\Delta S \geq 0$**

这一点的意义不仅仅在于末态到达另一条绝热线上以后仍然孤立地自发回到初始状态, 还意味着既使借助外界——两个恒温热源, 可能出现的其他辅助工作物质的帮助, 也无法在不改变其他物体的状态 (两个热源净吸放热为零, 其他工作物质初末状态复原) 的条件下回到初始状态, 原来的功也被补偿. 这构成热力学第二定律的一种表述, 它在热力学体系里是公理, 代表了自然界的某种规律.

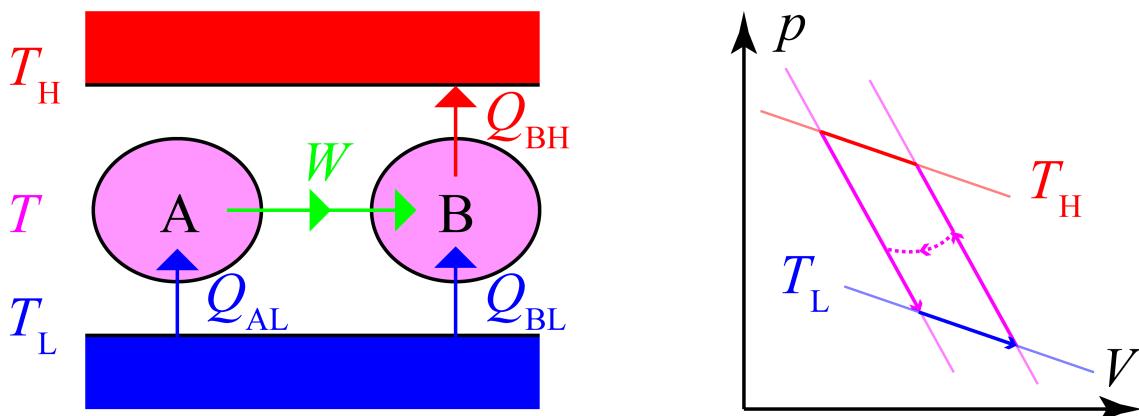


图 11.14: 三种不可逆过程的联系

<sup>3</sup>它具有极其独特的含义, 参见之后第五章的讨论.

质疑这一点会引发什么结果? 如果在某个任意中间温度  $T$ ,  $T_L < T < T_H$  下可以通过某个过程让较高绝热线上的状态回到相距很近的较低绝热线上的状态. 那么我们将它与两条绝热线和在低温热源上吸热膨胀的过程组成一个循环, 其合效果是从低温热源吸收了热, 并完全转变为对外界的做功. 这也是被热力学第二定律禁止的, 或者说其反过程不可逆:

开尔文表述: 功变热不可逆,  $W = -Q \geq 0$

表述为文字, 则为热机不可能在一个循环过程中从单一热源吸热转变为对外做功而不产生其他影响.

继续质疑这一点. 再构造一个热泵 B, 它接受来自热机 A 的功而从低温热源向高温热源泵热, 其总结果是, 低温热源向高温热源传了热, 其他体系状态历经一个循环回到初态. 同样的, 它被热力学第二定律禁止:

克劳修斯表述: 温差传热不可逆,  $Q_H = -Q_L \geq 0$

可见热力学第二定律的不同表述之间都是相互联系的. 实际上如果把两个热源和工作物质视为整体, 那么其实后两个表述就是整个体系的熵增原理, 此时, 整个循环过程来看, 如果过程中时时, 处处都是无耗散的准静态过程, 那么熵应该是不变的, 过程应该是可逆的, 这就是卡诺热机模型: 是高温热源与气体等温时放热  $Q_H > 0$ , 气体与低温热源等温时再放热  $|Q_L| < 0$ . 而气体对外做功为  $|W = -Q_H - Q_L| < 0$ . 或者是这个过程的逆过程——卡诺热泵循环. 对于所有热机, 热力学第二定律断言:

$$\text{卡诺定理: 卡诺热机效率最高, } \eta = 1 - \frac{-Q'_L}{Q'_H} \leq 1 - \frac{-Q_L}{Q_H}$$

证明思路是, 如果有热机效率高于卡诺热机效率, 意味着输出相同的功时其在两个热源的吸放热可以同时减到比卡诺热机小, 那么用输出的功给一个对应的卡诺热泵, 总的结果是低温热源向高温热源放热了, 由于热力学第二定律, 这不可以成立.

从而, 卡诺定理实际上也是热力学第二定律的等价表述. 这些表述都有一个特点, 可逆过程对应取等号, 而不等号发生在不可逆的过程中, 代表过程的方向性. 在等号情况下, 工作于两个恒温热源间的可逆卡诺热机的吸放热比是恒定的, 这给出了热力学温标的定义:

$$\frac{-Q_L}{Q_H} = \frac{T_L}{T_H}$$

只需确定一种特定系统的温度, 便可以由可逆循环的吸放热比测量任意其他平衡态的温度. 目前, 这种特定的系统被取做水的三相共存系统<sup>4</sup>, 温度定义为 273.16K. 这就是热力学温标. 而这也给出了与之共轭的熵的概念:

$$dS = \frac{dQ}{T}$$

上式对可逆过程适用. 对于卡诺循环来说, 如果着眼点在整个系统, 那么一个循环下来工作物质的熵是复原的. 而两个恒温热源的熵变根据上面温度的定义:

$$\Delta S = -\frac{Q_H}{T_H} - \frac{Q_L}{T_L} = 0$$

可见大的系统初末态在一条等熵线 (实际上整个循环中每一个状态之间都彼此等熵, 因为整个系统做绝热可逆过程) 上. 而如果我们着眼点在气体上, 而且让气体做一个任意的可逆循环过程, 一定有:

$$\Delta S = \oint \frac{dQ}{T} = 0$$

以上式子为克劳修斯等式 (Clausius equality), 同理对于不可逆过程将给出不等式, 普遍的, 我们对一个系统, 有:

$$\Delta S = S_{\text{final}} - S_{\text{initial}} \geq \int_{\text{initial}}^{\text{final}} \frac{dQ}{T}$$

<sup>4</sup>2018 年预通过的新国际单位制提案指出将可能把联系微观统计规律与宏观测量的常数  $k_B$  作为基本常数. 这样确定了温度与能量换算的线性关系, 从极低温到极高温完全一致.

上式也是热力学第二定律的表述,它十分强大,其他表述都可以简单地从这个表述验证. 它甚至有十分清晰的图像: 右边是从外界流向系统的熵流, 左边的区别在于系统内部不平衡条件下的各种流造成的熵产生, 详见下一节和本章最后一节的相关讨论. 对于上式, 如果不等号左边为零, 可以理解为功变热的不可逆性: 右边积分里的  $dQ$  小于零将使得不等式成立, 说明在过程中体系更自发的行为是接受外界对其做功, 而以放热的形式形成能量守恒, 当然条件是初末态等熵——一般考虑循环过程. 如果不等式右边为零, 给出了熵增原理——绝热体系的熵只增不减.

## 11.4 熵的计算

### 11.4.1 理想气体的熵

理想气体温标与热力学温标为何统一? 一者, 用理想气体温标定义的温度自动符合卡诺定理: 卡诺循环效率为热力学定义值  $\eta = 1 - \frac{T_L}{T_H}$ , 且关键是它能使得下式成为全微分:

$$dS = \frac{dQ}{T} = \frac{dU + pdV}{T} = f(T, V)dT + g(T, V)dV$$

如果理想气体温标是  $T' = h(T) \cdot T$ , 那么  $\frac{dQ}{T'}$  不可能再是一个全微分, 而与热力学第二定律定义的温标不一致.

二者, 热力学第二定律的体系看似复杂而充满精巧的构造, 实则可以在更深的层次: 微观动力学规律与统计规律的基础上进行自然的推导. 此时温度, 熵, 热量不仅可以普适而简单地定义, 而且具有清晰的物理图像, 我们在第五章统计物理基础中展开说明.

所以, 理想气体的熵我们可以直接使用第二节的推导, 不过我们这里写作:

$$S(T, V) = \nu C_{mV} \ln T + \nu R \ln V$$

$$S(p, V) = \nu C_{mV} \ln p + \nu C_{mp} \ln V$$

$$S(p, T) = \nu C_{mp} \ln T - \nu R \ln p$$

以上三个定义不太一样, 差一些取决于摩尔数  $\nu$  的常数. 而且只有最下面一个是广延量. 一种常见的处理方法是把体积换为摩尔体积:

$$V_m = \frac{V}{\nu}$$

摩尔体积实际上与比体积(单位质量的体积) $v$ , 密度  $\rho$ , 粒子数密度  $n$  有如下正反比关系:

$$\rho = \frac{1}{v} = \frac{\mu}{V_m} = nm$$

这样三个式子都变成了广延量. 它们还差一些与  $\nu$  成正比, 对数里面是  $R$  的一些广延量. 这种定义的非唯一性很正常, 实际过程中产生实际效果的是熵的变化量, 仅有在极低温下, 讨论量子体系时才需要讨论熵的绝对数值(热力学第三定律). 还有一点, 为了使得对数里的物理量不带单位, 我们规定  $p = p_0$ ,  $v = v_0$ ,  $T = T_0$  时的熵为零, 最后得到统一的熵的公式:

$$S = \nu C_{mV} \ln \frac{T}{T_0} + \nu R \ln \frac{v}{v_0} = \nu C_{mV} \ln \frac{p}{p_0} + \nu C_{mp} \ln \frac{v}{v_0} = \nu C_{mp} \ln \frac{T}{T_0} - \nu R \ln \frac{p}{p_0}$$

### 11.4.2 固定热容固体熵

理想情况下这种固体是以温度为单参数热力学系统, 其内能为

$$U = CT$$

而吸热就是  $Q = C\Delta T$ , 从而熵:

$$dS = C \frac{dT}{T}$$

从而:

$$S = C \ln \frac{T}{T_0}$$

### 11.4.3 大热容恒温热库

无疑, 这种情况是上一种情形的大热容极限近似, 如果温度变化  $\Delta T$  相对于  $T$  很小:

$$\Delta S = C \ln \frac{T + \Delta T}{T} \simeq \frac{C\Delta T}{T} = \frac{\Delta U}{T} \simeq \Delta \frac{U}{T}$$

即此时等价的定义不会造成区别:

$$S = \frac{U}{T}$$

### 11.4.4 传热熵

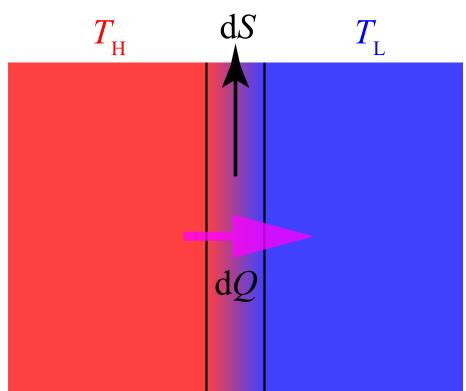


图 11.15: 传热熵产生

如果高温物体温度  $T_H$ , 低温物体温度  $T_L$ , 两者发生一个热接触而有热流  $dQ$  从高温物体流向低温物体. 这是一个非准静态过程, 它不可逆, 如何分析其熵的呢? 考虑高温物体放热  $dQ$ , 实际上如果它是与一个温度同为  $T_H$  的恒温热库接触而放出这个热量, 其状态的改变是完全相同的, 不过这是一个准静态过程, 可以计算其熵变:

$$dS_H = -\frac{dQ}{T_H}$$

即, 有  $|dS_H|$  的熵从高温物体中流了出来. 同理, 流向低温物体的熵为:

$$dS_L = \frac{dQ}{T_L}$$

整个体系熵要增加:

$$dS = dQ \left( \frac{1}{T_L} - \frac{1}{T_H} \right) > 0$$

上式被称为传热过程中的熵产生 (entropy production). 它其实也代表某种微观形式的耗散, 它发生于高温与低温物体接触时平均起来高温的动能较大的粒子动能被拉低而低温物体动能被拔高的过程中.

### 11.4.5 混合熵

讨论理想气体的混合, 而且混合后各种气体分子间亦无任何形式的相互作用. 命气体的第  $i$  组分的摩尔分数为  $x_i$ , 混合前不同气体处于温度为  $T$ , 压强为  $p$  的平衡, 不同组分占体积比为:

$$V_1 : V_2 : \cdots : V_n = x_1 : x_2 : \cdots : x_n ; \quad V_i = x_i V$$

混合后温度仍然是  $T$ , 体积倒是共用一个  $V$  了. 而不同组分去对  $p$  进行分压:

$$p_1 : p_2 : \cdots : p_n = x_1 : x_2 : \cdots : x_n ; \quad p_i = x_i p$$

熵是广延量, 混合理想气体的熵满足吉布斯定理(Gibbs' theorem): 混合气体的熵等于所有组分的熵的和, 而每一组分在与其他组分共存的过程中, 其他组分对这个组分可能的微观状态形式并无影响(分子间无相互作用). 所以这个组分的熵就等于它单独存在时的熵. 那么过程前后的熵变其实也就是气体向真空自由膨胀的熵变(因为  $T$  恰好不变, 与自由膨胀特征相同), 这是一个典型的不可逆过程:

$$\Delta S_i = \nu_i R \ln \frac{V}{V_i} = -\nu R x_i \ln x_i$$

故算出混合熵(entropy of mixing) 为:

$$\Delta S = -\nu R \sum x_i \ln x_i > 0$$

我们写出末态混合均匀后熵的表达式, 注意为了保证其广延性, 我们采用  $p, T, x_i$  这些强度量作为其可以放到  $\ln$  后面的自变量:

$$\begin{aligned} S &= \sum S_i \\ &= \sum \nu_i C_{mp} \ln T - \nu_i R \ln p_i \\ &= \nu C_{mp} \ln T - \nu R \ln p - \nu R \sum x_i \ln x_i \\ &= S_0 + \Delta S \end{aligned}$$

从中混合熵的概念就十分明显了: 它代表混合气体与同样条件下的单组分气体的熵  $S_0$  的区别. 这很自然地引入了所谓的吉布斯佯谬(Gibbs' paradox): 一是为什么吉布斯定理不适用于单组分气体, 如果把一缸气体体积分成两份, 那么熵可以直接相加(广延性), 但不能把这缸气体的压强(粒子数)分成两份而占据相同的体积, 除非分成的两份分子有本质的不同. 二就是多组分与单组分的区别在哪里, 会造成物理量的差别.

这其实不是奇怪的事. 我们考虑一缸氢气, 它的熵应根据  $S_0$  计算, 用它计算也符合测量结果. 而某一天科学家们突然得知, 原来氢分子  $H_2$  分为两种: 核自旋平行的正氢(orthohydrogen) 和核自旋反向的仲氢(parahydrogen), 两者占比为 3 : 1. 两种分子仅仅是深入到原子核的性质略有不同, 对常温下的分子物理化学性质几乎没有影响<sup>5</sup>. 于是似乎要在熵公式后附加一项混合熵  $-\nu R (\frac{1}{4} \ln \frac{1}{4} + \frac{3}{4} \ln \frac{3}{4}) \approx 0.56\nu R$ . 这一项如何解释? 通过第五章的学习我们将知道, 熵实际上对应着信息的多少. 在混合氢情况下我们知道了很多的信息: 每一个氢分子究竟是正氢还是仲氢. 所以熵增加了. 但加这一项有必要吗? 其实是没有必要的, 因为它作为常数, 在公式  $dQ = TdS$  中求变化量时不出现, 从而不会出现在实际可以测量的系统发生的过程中. 然而一涉及到微观构成, 它就是必要的了, 如果两种氢之间可以相互转化, 那么如果整缸氢气都是正氢或是仲氢, 在一定时间后它将在两种氢之比为 3 : 1 处平衡, 这个过程是个不可逆过程. 这是因为正氢又分为三种可能的自旋态, 那么把这三个态与仲氢的摩尔分数记做  $x_1, x_2, x_3, x_4$ , 有:

$$x_1 + x_2 + x_3 + x_4 = 1$$

$$\Delta S = -\nu R \ln \sqrt[4]{x_1 x_2 x_3 x_4}$$

由算术几何平均不等式, 在  $x_1 = x_2 = x_3 = x_4 = \frac{1}{4}$  处熵增取极大值. 在这个状态往任何一个其他状态熵必然减少而不被热力学允许. 它就是我们要求的平衡态. 所以两种氢之比为 3 : 1, 同时我们还可以看出之前我们算的熵变还不够严格, 考虑四种氢时混合熵应为

$$\Delta S = \nu R \ln 4 = 1.4\nu R$$

值得一提, 我们上面反过来通过预先在非平衡态都有良好定义的熵的极值来求平衡状态的参数条件的方法称为最大熵原理(principle of maximum entropy), 这种目前看似难以理解的方法亦会在第五章给出详细论述.

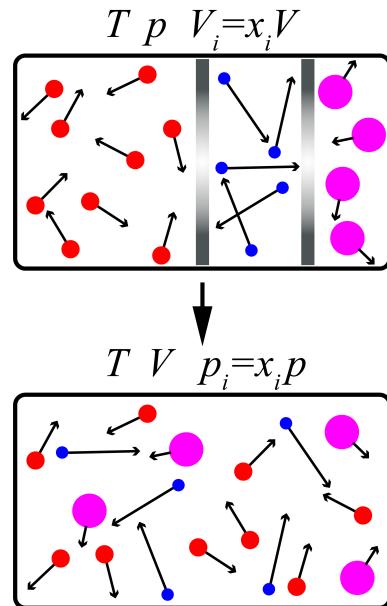


图 11.16: 混合熵

<sup>5</sup>然而低温下影响巨大, 这与量子效应有关

## 11.5 热力学函数与其特性

### 11.5.1 四个热力学函数与四个状态参量

在本节中我们将把态函数熵  $S$  视为与  $p, V, T$  并列的状态参量. 由热力学第一定律:

$$dU = dQ + dW$$

无论是在准静态或非准静态过程中上式都使用, 而只有在准静态过程中:

$$dQ = TdS \quad ; \quad dW = -pdV$$

于是:

$$dU = TdS - pdV$$

这个式子又是对准静态和非准静态过程都适用的了, 因为等式两边都是状态量, 它们不关心过程的性质. 这是对符合热力学规律的正确定义的各个物理量之间提出的必然要求, 被称作**热力学基本方程**(master equation). 这里, 可以理解为如果把内能写作熵  $S$  和体积  $T$  的函数, 它必须满足:

$$\left(\frac{\partial U}{\partial S}\right)_V = T; \quad \left(\frac{\partial U}{\partial V}\right)_S = -p$$

由混合偏导数与求导次序无关的性质, 可以得到:

$$\left(\frac{\partial T}{\partial V}\right)_S = -\left(\frac{\partial p}{\partial S}\right)_V$$

而我们之前定义了焓  $H = U + pV$ , 那么相应的有:

$$dH = TdS + Vdp$$

$$\left(\frac{\partial H}{\partial S}\right)_p = T; \quad \left(\frac{\partial H}{\partial p}\right)_S = V$$

$$\left(\frac{\partial T}{\partial p}\right)_S = \left(\frac{\partial V}{\partial S}\right)_p$$

我们再定义**亥姆霍兹自由能**(Helmholtz's free energy), 简称**自由能**(free energy) 为  $F = U - TS$ , 从表达式上看, 它反映了一种能量最低与熵最高的“抗衡”. 它的相应结论为:

$$dF = -SdT - pdV$$

$$\left(\frac{\partial F}{\partial T}\right)_V = -S; \quad \left(\frac{\partial F}{\partial V}\right)_T = -p$$

$$\left(\frac{\partial S}{\partial V}\right)_T = \left(\frac{\partial p}{\partial T}\right)_V$$

最后还可以定义**自由焓**(free enthalpy), 它又被叫做**吉布斯自由能**(Gibbs' free energy), 定义为  $G = H - TS = U + pV - TS$ , 相关性质为:

$$dG = -SdT + Vdp$$

$$\left(\frac{\partial G}{\partial T}\right)_p = -S; \quad \left(\frac{\partial G}{\partial p}\right)_T = V$$

$$\left(\frac{\partial S}{\partial p}\right)_T = -\left(\frac{\partial V}{\partial T}\right)_p$$

以上四个热力学基本函数的定义自然而然的给出了四个基本关系式,称为麦克斯韦关系 (Maxwell relations):

$$\left(\frac{\partial T}{\partial V}\right)_S = - \left(\frac{\partial p}{\partial S}\right)_V$$

$$\left(\frac{\partial T}{\partial p}\right)_S = \left(\frac{\partial V}{\partial S}\right)_p$$

$$\left(\frac{\partial S}{\partial V}\right)_T = \left(\frac{\partial p}{\partial T}\right)_V$$

$$\left(\frac{\partial S}{\partial p}\right)_T = - \left(\frac{\partial V}{\partial T}\right)_p$$

如何理解这些关系?它们实际上都来自热力学规律对从  $p, V$  确定  $T, S$  的映射的性质.如果我们画出  $p - V$  图与  $T - S$  图,则两个图形上的点一一对应. $p - V$  图上的一个闭区域的面积必须和对应的  $T - S$  图上的面积相等.这是因为其边界对应的循环过程的吸热与做功就是两个面积,由热一两者必然相等.这种性质在数学上表述为雅可比行列式 (Jacobian determinant) 必须等于一:

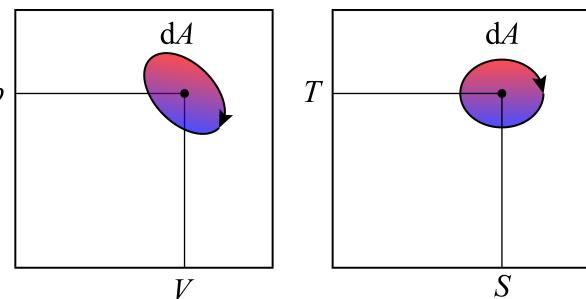


图 11.17: 循环过程的做功与吸热相等

$$\frac{\partial(T, S)}{\partial(p, V)} = \begin{vmatrix} \frac{\partial T}{\partial p} & \frac{\partial S}{\partial p} \\ \frac{\partial T}{\partial V} & \frac{\partial S}{\partial V} \end{vmatrix} = \left(\frac{\partial T}{\partial p}\right)_V \left(\frac{\partial S}{\partial V}\right)_p - \left(\frac{\partial S}{\partial p}\right)_V \left(\frac{\partial T}{\partial V}\right)_p = 1$$

而且是 +1, 这代表循环的顺时针与逆时针在映射下保持不变.对于符号  $\partial(p, V)$ , 我们理解为点  $(p, V)$  周围的一小块邻域的面积大小.这样这个式子就是一个直观的微商式,而上式两个面积比之所以可以用行列式来表示,是因为在  $p - V$  图上  $(p, V)$  点移动  $dp$ , 在  $T - S$  图上移动矢量  $(\frac{\partial T}{\partial p}, \frac{\partial S}{\partial p})dp$ ,若移动  $dV$ ,则在  $T - S$  图应移动  $(\frac{\partial T}{\partial V}, \frac{\partial S}{\partial V})dV$ ,那么在  $p - V$  图上由两个移动构成的矩形面积为  $dpdV$ ,而  $T_S$  图上面积应为  $\left|(\frac{\partial T}{\partial V}, \frac{\partial S}{\partial V}) \times (\frac{\partial T}{\partial V}, \frac{\partial S}{\partial V})\right| dpdV$ ,根据代数知识这个叉乘就等于以上行列式.

也是基于这个认识,不难写出:

$$\left(\frac{\partial T}{\partial V}\right)_S = \frac{\partial(T, S)}{\partial(V, S)} = \frac{\partial(p, V)}{\partial(V, S)} = -\frac{\partial(V, p)}{\partial(V, S)} = -\left(\frac{\partial p}{\partial S}\right)_V$$

其他三个麦克斯韦关系也可以依据这个方法进行推导.在  $p - V$  图和  $T - S$  图上应该如何直观理解这个式子还请读者自己思考.

### 11.5.2 若干定理的证明

#### 能态定理

对于  $p - V$  体系,若给出状态方程,则能量关于状态参量的偏导数有以下条件必须满足:

$$\left(\frac{\partial U}{\partial V}\right)_T = T \left(\frac{\partial p}{\partial T}\right)_V - p$$

证法一：

利用  $U(S, V)$  是特性函数(characteristic function), 用它可以表示出任意状态参量与其偏导数:

$$\begin{aligned} p &= -U_V, \quad T = U_S \\ \left(\frac{\partial p}{\partial T}\right)_V &= \left(\frac{\partial S}{\partial T}\right)_V \cdot \left(\frac{\partial p}{\partial S}\right)_V = -\left(\frac{\partial S}{\partial T}\right)_V \cdot U_{SV} \\ dT = dU_S &= U_{SS}dS + U_{SV}dV \Rightarrow \left(\frac{\partial S}{\partial T}\right)_V = \frac{1}{U_{SS}}, \left(\frac{\partial p}{\partial T}\right)_V = -\frac{U_{SV}}{U_{SS}} \end{aligned}$$

从而等式右边为:

$$T \left(\frac{\partial p}{\partial T}\right)_V - p = U_V - \frac{U_S U_{SV}}{U_{SS}}$$

而等式左边为:

$$\begin{aligned} \left(\frac{\partial U}{\partial V}\right)_T &= \left(\frac{\partial U}{\partial V}\right)_S + \left(\frac{\partial U}{\partial S}\right)_V \cdot \left(\frac{\partial S}{\partial V}\right)_T \\ dT = dU_S &= U_{SS}dS + U_{SV}dV \Rightarrow \left(\frac{\partial S}{\partial V}\right)_T = -\frac{U_{SV}}{U_{SS}} \end{aligned}$$

代入, 即有:

$$\left(\frac{\partial U}{\partial V}\right)_T = U_V - \frac{U_S U_{SV}}{U_{SS}}$$

从而两式相等.

证法二:

方程以  $T, V$  作为自变量, 故考虑自由能  $F = U - TS$ , 它恰好以  $T, V$  为自变量时为特性函数, 有:

$$\left(\frac{\partial U}{\partial V}\right)_T = \left(\frac{\partial F}{\partial V}\right)_T - T \left(\frac{\partial S}{\partial V}\right)_T = -p + T \left(\frac{\partial p}{\partial T}\right)_V$$

后一步推导用到了麦克斯韦关系.

证法三:

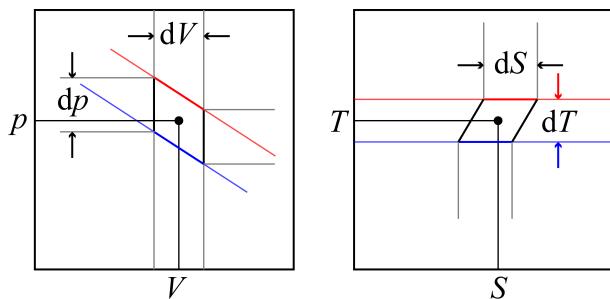


图 11.18: 构造循环过程证明能态定理

构造在两根相差  $dV$  的等体线和相差  $dT$  的等温线之间的循环过程. 这不是一个卡诺循环, 但依然近似有两个等温过程中的吸放热定义的效率为:

$$\eta = 1 - \frac{dQ_L}{dQ_H} = \frac{\delta dQ}{dQ} = \frac{dT}{T}$$

这一点从这个循环对应的  $T - S$  图中可以看出来, 小段红线下方的面积与小段蓝线下方的面积, 由于微元过程近似为平行四边形, 两个面积的底  $dS$  相等, 面积差即为平行四边形的面积  $dTdS$ , 而吸放热约为  $TdS$ , 它们的比就是上式.

现在我们把注意力集中在左图中. 在等体过程中增加的压强为  $\left(\frac{\partial p}{\partial T}\right)_V dT$ , 从而过程做功为:

$$dW = \left(\frac{\partial p}{\partial T}\right)_V dTdV$$

而两个等温过程中的吸热, 由热力学第一定律:

$$dQ = pdV + \left(\frac{\partial U}{\partial V}\right)_T dV$$

把两式与循环效率比对:

$$\frac{dW}{dQ} = \frac{dT}{T}$$

即得:

$$\left(\frac{\partial U}{\partial V}\right)_T = T \left(\frac{\partial p}{\partial T}\right)_V - p$$

把这个式子用于理想气体的物态方程:

$$pV = \nu RT \Rightarrow \left(\frac{\partial U}{\partial V}\right)_T = 0$$

可见理想气体的内能只和温度有关是由物态方程决定的. 而满足该物态方程的其他气体模型也需要符合这一点, 所以内能对温度的偏导数仍然符合等体热容的定义, 但热容一般是温度的函数:

$$U = \nu \int C_{mV}(T) dT$$

实际气体, 一方面在高密度, 高温度情况下其状态方程会偏离理想物态方程的形式, 另一方面, 多原子分子气体其内能-温度关系在常温-低温区一般体现为温度的单调递增函数的形式, 而在某些温度区间达到一些近似不变的平台. 这反应了气体内部自由度的量子本质. 在一种近似的模型处理中, 我们可以让  $T < T_c$  时摩尔等体热容为  $C_{mV} = \frac{i}{2}R$ , 而  $T > T_c$  时摩尔等体热容为  $C_{mV} = \frac{f}{2}R$ , 而在  $T = T_c$  附近的微小区间内不仅热容迅速升高, 内能也突然从  $\frac{i}{2}\nu RT_c$  上升到  $\frac{f}{2}\nu RT_c$ <sup>6</sup>. 这实际上是一种相变-期间两种热容的分子可以视为共存, 详见第四章相与相变的内容.

这种情况也很容易推广到一个类似体系中: 液膜的热力学系统. 在此时热力学基本方程写作:

$$dU = TdS + \sigma dA$$

可见在这儿表面张力系数  $\sigma$  是广义力, 广义坐标是表面积  $A$ . 故我们可以写出:

$$\left(\frac{\partial U}{\partial A}\right)_T = \sigma - T \left(\frac{\partial \sigma}{\partial T}\right)_A$$

一般  $\sigma$  是温度  $T$  的函数, 只不过与理想气体情况不太相同的是,  $\sigma$  依赖于  $T$  但一般不依赖于液面表面积  $A$ , 液膜的拉伸改变的是膜的厚度, 而表面层永远是膜的两个表面上极薄的一层, 它的性质不会改变, 从而不会改变表面张力系数 (详见下一章液体与固体的性质). 事实上, 此时我们注意到自由能的性质为:

$$dF = -SdT + \sigma dA$$

而  $\frac{\partial \sigma}{\partial A} = 0$  意味着:

$$\left(\frac{\partial F}{\partial A}\right)_T = \sigma(T) ; \quad \left(\frac{\partial^2 F}{\partial A^2}\right)_T = 0$$

这意味着  $F = \sigma(T)A + f(T)$ , 对于  $f(T)$ , 显然与表面现象没有关系, 因为当面积  $A = 0$  时这一项仍然存在, 说明不是因为表面现象引起的, 实验时与理论计算都应该扣除. 从而我们给出:

$$F = \sigma A$$

也就是说平时说的表面能在热力学下实际上是自由能. 它恰恰是表面积  $A$  的广延量. 而熵应该为:

$$S = -\left(\frac{\partial F}{\partial T}\right)_A = -\sigma' A$$

<sup>6</sup>事实上由于需要吸热而温度不变, 这个过程热容是发散的.

最后再代入到  $U = F + TS$ , 得:

$$U = (\sigma - T\sigma')A$$

内能与面积的比例系数不是  $\sigma$ , 但由  $\sigma$  决定, 意料之外, 情理之中. 这意味着等温地拉伸液膜是需要吸放热的, 从微观机理上来看, 熵必须大于零, 从而  $A$  增加熵应该也要增加. 从以上熵的表达式来看必然有:

$$\sigma' < 0$$

从而我们得出了温度升高表面张力系数减小的结论. 微观上看, 温度升高导致分子从液相通过表面层到气相更简单, 表面层变薄, 所以表面能减小. 从而拉伸液膜是要吸热的. 而如果面积不变, 而温度升高, 我们得到等面积热容为

$$C_A = -T\sigma''A$$

对于一般的有机液体有经验公式:

$$\sigma = \sigma_0 \left(1 - \frac{T}{T_C}\right)^n$$

一般  $n \simeq \frac{11}{9}$ , 从而我们发现  $C_A < 0$ <sup>7</sup>. 这没什么奇怪的, 要知道液膜的表面层是依附于内部的液体而存在的. 所以既使增加  $A$  对内部液体没有影响, 升温绝对是对两侧的液, 气相产生影响的. 这个热容将被包含在正的较大的液体气体热容中. 它的减小可以理解为温度升高时表面层变薄而直接减少内能的缘故. 另外, 在  $T = T_C$  处  $\sigma$  减小到零, 内能与自由能也都减小到了零. 这也是预料之中的. 这一点就是液相与气相共存的临界点. 第四章相与相变将仔细研究这个现象.

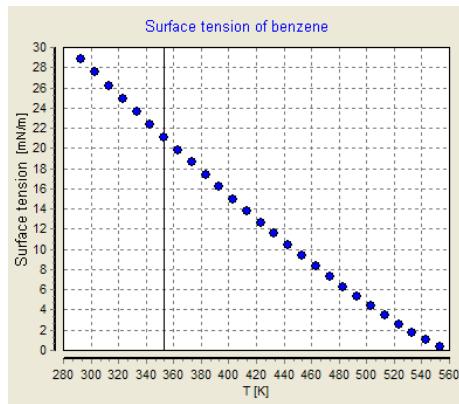


图 11.19: 苯的  $\sigma \sim T$  关系

### 物性函数间关系

除了两个热容  $C_{mp}, C_{mV}$ , 实验中容易测得的量一般只能是  $p, V, T$ . 从而连带着可以测的还有在上一章理想气体的过程这一节中介绍的各种**物性函数**(material property function): 膨胀系数  $\alpha$ , 压强系数  $\beta$  和压缩系数  $\kappa$ . 不同的过程中这些量间有关系:

$$\alpha_X = -p\beta_X\kappa_X$$

$$\alpha_p = p\beta_V\kappa_T$$

$$\gamma = \frac{C_p}{C_V} = \frac{\kappa_T}{\kappa_S}$$

$$C_p - C_V = pVT\alpha_p\beta_V$$

其中下标  $X$  表示一个任意的过程.

证明:

二自由变量的  $p-V$  系统引入一个过程方程后变为一自由变量, 所有变化的微元互相之间都成正比, 即:

$$dp = \lambda_p dx, dV = \lambda_V dx, dT = \lambda_T dx$$

其中  $x$  作为过程参数反应过程发生的程度. 从而三个  $X$  过程的响应函数分别为:

$$\alpha_X = \frac{1}{V} \left( \frac{\partial V}{\partial T} \right)_X = \frac{\lambda_V}{V\lambda_T}$$

<sup>7</sup>热力学预言简单的  $p-V$  系统如果热容小于零将是不稳定的——传热将使得局部高温更高, 低温更低.

$$\beta_X = \frac{1}{p} \left( \frac{\partial p}{\partial T} \right)_X = \frac{\lambda_p}{p \lambda_T}$$

$$\kappa_X = -\frac{1}{V} \left( \frac{\partial V}{\partial p} \right)_X = -\frac{\lambda_V}{V \lambda_p}$$

从而得到待证第一式. 而第二个式子的区别在于它不依赖于某个过程, 而是直接从物态方程中得出:

$$F(p, V, T) = 0 \Rightarrow F_p dp + F_V dV + F_T dT = 0$$

$$\Rightarrow \alpha_p = -\frac{F_T}{VF_V}, \beta_V = -\frac{F_T}{pF_p}, \kappa_T = \frac{F_p}{VF_V}$$

即可证明第二式. 这两个过程都仅仅是数学恒等式, 不依赖于任何体系的物理性质. 而第三个式子则需要热二的相关推论来推导. 首先注意到一下关于两个热容的定义是与之前说法等价的:

$$C_V = \left( \frac{\partial U}{\partial T} \right)_V = T \left( \frac{\partial S}{\partial T} \right)_V; \quad C_p = \left( \frac{\partial H}{\partial T} \right)_p = T \left( \frac{\partial S}{\partial T} \right)_p$$

从而利用雅可比行列式的方法:

$$\frac{C_p}{C_V} = \frac{\left( \frac{\partial S}{\partial T} \right)_p}{\left( \frac{\partial S}{\partial T} \right)_V} = \frac{\partial(S, p)}{\partial(T, p)} \cdot \frac{\partial(T, V)}{\partial(S, V)} = \frac{\partial(T, V)}{\partial(T, p)} \cdot \frac{\partial(S, p)}{\partial(S, V)} = \frac{\left( \frac{\partial V}{\partial p} \right)_T}{\left( \frac{\partial V}{\partial p} \right)_S} = \frac{\kappa_T}{\kappa_S}$$

这就是第三式, 而最后一式则考虑  $(V, T)$  为自变量:

$$C_p - C_V = T \left( \frac{\partial S}{\partial T} \right)_p - T \left( \frac{\partial S}{\partial T} \right)_V = T \left( \frac{\partial S}{\partial V} \right)_T \left( \frac{\partial V}{\partial T} \right)_p = T \left( \frac{\partial p}{\partial T} \right)_V \left( \frac{\partial V}{\partial T} \right)_p = pVT\alpha_p\beta_V$$

期间用了一次麦克斯韦关系. 至此四个关系式都得证.

### 吉布斯-亥姆霍兹方程

吉布斯-亥姆霍兹方程 (Gibbs-Helmholtz equation) 用来确定化学反应中的吸放热 (实际上是焓变) 与吉布斯自由能的变化间的关系, 而吉布斯函数又和化学反应发生的方向, 化学平衡的位置紧密相关. 这个关系写作:

$$\left( \frac{\partial}{\partial T} \right)_p \frac{G}{T} = -\frac{H}{T^2}$$

证明:

实际上:

$$H = G + TS = G - T \left( \frac{\partial G}{\partial T} \right)_p = -T^2 [G \left( \frac{\partial}{\partial T} \right) \frac{1}{T} + \frac{1}{T} \left( \frac{\partial}{\partial T} \right) G] = -T^2 \left( \frac{\partial}{\partial T} \right) \frac{G}{T}$$

这就是上式. 相似的关系我们还可以写出:

$$\left( \frac{\partial}{\partial T} \right)_V \frac{F}{T} = -\frac{U}{T^2}$$

$$\left( \frac{\partial}{\partial V} \right)_T \frac{F}{V} = -\frac{G}{V^2}$$

$$\left( \frac{\partial}{\partial p} \right)_T \frac{G}{p} = -\frac{F}{p^2}$$

### 11.5.3 自由能的含义

我们先讨论亥姆霍兹自由能,对于双参数描述的简单系统,此时热力学基本关系牢固地成立,不管过程是否准静态:

$$dU = TdS - pdV$$

但对于非双参数的,比如我们考虑两种气体的相互扩散,初态与末态似乎其表现出来的性质是类似的.比如标准状态下1mol正氢与1mol仲氢混合均匀.如果视为一个 $p-V$ 系,整个过程前后等温,等压,等体,但熵增.这类问题上式就不再是一个等式.究其本质,在与内部发生的各种过程的各种不平衡与耗散的因素造成的自发熵产生.体系总是受迫地远离平衡态,而有自发地从非平衡态回到平衡态.为了简化,我们假设系统与外界的热,功交换仍然是准静态且无耗散的<sup>8</sup>,熵产生仅仅发生在体系内部.此时上式应该写为:

$$\Delta U \leq T\Delta S - p\Delta V$$

这是因为外界对系统做功仍然为 $\Delta W = -p\Delta V$ ,但根据克劳修斯不等式系统从外界吸热满足 $\Delta Q \leq T\Delta S$ .而最终 $\Delta U = \Delta Q + \Delta W$ 的缘故.注意到克劳修斯不等式的本质就是系统与外界一起看成一个整体的熵增原理.它是普适的,对任意复杂体系都适用的统计规律.上式也通常写成

$$\Delta S \geq \frac{\Delta U}{T} + \frac{p\Delta V}{T}$$

上式也与熵增原理兼容,等体,绝热的体系自然可以认为是孤立的,它内能不变,从而上式给出了 $\Delta S > 0$ .但上式实际上比熵增原理更强.因为即使如果对于 $V$ 增大的过程,即使不可逆,上式对理想气体也应该严格成立,但对于一般多内部参量的 $p-V$ 系统上式则可以取不等号.上式还给出了平衡态的条件:我们可以在不改变体系总内能和总体积的情况下分配各部分所占的内能和体积,如果可以通过某种分配方式让熵增加,则体系会自发地往这个方向演化.直至熵达到它的极大值,对应热力学平衡态.这叫做**最大熵原理**(principle of maximum entropy).

从内能的相似表达式中可以得到完全一致的结论,不过此时我们要保持熵 $S$ 与体积 $V$ 不变.而结论是对应的**热力学势**(thermodynamic potential)——内能取最小.对应的原理为**最小能原理**(principle of minimum energy).从内能出发的这个结论有些不足之处:熵理论上计算公式复杂,实际操作上又因为无法直接控制其不变而没有意义,且难以用微观统计观点诠释的熵的概念解释.所以仅有较为好的热力学理论价值.

亥姆霍兹自由能的优势就体现在这儿了,由于 $F = U - TS$ ,在等温的条件下我们写出:

$$\Delta F \leq -S\Delta T - p\Delta V$$

从而这个关系式对应的平衡判据为——在等温等体情况下自由能必须极小.这个式子实际上经常意味着纯粹的机械平衡.例如我们还记得液膜的自由能就是 $F = \sigma A$ ,而后一项 $-p\Delta V$ 理解为 $\Delta W$ ,这个原理就是:在一定温度下固定边界(外界不做功)的液膜的面积要收缩到最小.利用**变分法**(calculus of variation)我们可以把这个最小曲面确定下来.我们将在下一章介绍液体表面张力时介绍相关概念.

如果放宽条件,允许系统在不变的温度下对外做功 $\Delta W' = -\Delta W = p\Delta V$ ,那么两边同时积分给出:

$$W' \leq F_{\text{initial}} - F_{\text{final}}$$

等号成立时,这就是个机械能守恒,自由能 $F$ 具有势能的含义.而不等号的存在意味着热力学系统的不可逆性.上式又被称为**最大功原理**(principle of maximum work).虽然名称与上面两个原理相似,但这个原理不是用作一个平衡判据,而是用于实际过程,工程上往往把整个系统与外界和合自由能称为**火用**<sup>9</sup>(exergy).

<sup>8</sup>为了表述的简化我们让体系的每处都处于共同温度 $T$ 而每个与外界交换的体积功都发生在共同压强 $p$ 处,只要我们愿意对于较复杂的系统我们随时都可以把下面式子写为:

$$\Delta U \leq \sum_i T_i \Delta S_i - \sum_j p_j \Delta V_j$$

<sup>9</sup>笔者电脑字库里没有这个字 orz.

接着我们讨论吉布斯自由能  $G$ . 同样, 吉布斯自由能的能量最小判据在等温等压的条件下被写作:

$$\Delta G \leq -S\Delta T + V\Delta p$$

也就是说, 在等温, 等压的过程中吉布斯自由能  $G$  总是取到它的极小值. 而等温等压的过程又常常是各种化学反应发生的初末态要符合的条件. 此式在判断化学反应的发生方向上极为有用. 在各种燃烧的模型中, 我们发现气体发生化学反应体积膨胀对外做的功不全由体积功  $p_0\Delta V$  表示, 这是因为把外界压强(初末态平衡压强)与过程中气体压强划等号的不合理. 此时还会有其他形式的机械功输出, 要么转化为气体的动能, 要么利用起来推动活塞对外做功. 这一部分功称为非体积功  $W''$ , 由于  $W' = W'' + p\Delta V$ , 我们通过上面的亥姆霍兹自由能的最大功原理写出等温等压下的吉布斯自由能的最大功原理:

$$W'' \leq G_{\text{initial}} - G_{\text{final}}$$

#### 11.5.4 化学势

最后我们再考虑一下粒子数可变的系统. 对于理想气体类似的仅由  $p - V$  决定的系统在考虑粒子数变化后将有三个独立变化的状态参量. 我们注意到粒子数  $N$  不变时:

$$dG = -SdT + Vdp$$

而显然  $G, S, V$  恰好都是广延量, 而作为自变量的  $p, T$  是两个强度量. 这就是说, 如果我们引入“每个粒子携带的吉布斯自由能” $\mu = \frac{G}{N}$ , 同样地, 我们不假思索的把所有广延量平分到每个粒子上, 用它们的小写表示, 如  $s, v, u, h, f$  等等. 研究吉布斯自由能便可以发现:

$$d\mu = -sdT + vdp$$

$$dG = d(N\mu) = -SdT + Vdp + \mu dN$$

而如果研究内能便可以发现:

$$du = Tds - pdv$$

$$dU = d(Nu) = NTds - Npdv + udN = TdNs - pd(Nv) + (u - Ts + pv)dN = TdS - pdV + \mu dN$$

另外两个热力学势是类似的:

$$dF = -SdT - pdV + \mu dN$$

$$dH = TdS + Vdp + \mu dN$$

从中可以看出  $\mu$  具有某种特殊的含义, 我们把这个量叫做**化学势**(chemical potential). 它是在准静态的等温等压过程中向体系添加一个粒子外界需要做的非体积功. 它与相之间的平衡与转化问题息息相关. 如一个在等温等压条件下自发进行的化学反应过程  $A \rightarrow B$ , 我们如果不对体系做非体积功, 也不去吸收非体积功, 那么根据最小功原理吉布斯自由能必须减小. 也就是  $A$  的化学势必须比  $B$  的化学势高可以利用化学势的差对外输出机械功, 这便是内燃机的原理. 达到平衡时有  $\mu_A = \mu_B$ , 此时两相之间的粒子交换已经是准静态的过程. 不会造成过程的不可逆性.

在引入化学势后对于共存的, 或多相混合的, 甚至非均匀的  $p - V$  系统我们统一地写出热力学的基本关系式:

$$dU = TdS - pdV + \sum_{\alpha} \mu_{\alpha} dN_{\alpha}$$

上式是一个恒等式, 不再像之前的讨论那样由于缺少内部的参数而导致成为不等式. 这一点容易理解, 因为考虑到化学势实际上说明我们对不同的组分进行了区分, 考虑到了实际  $p - V$  系统的内部构成而不是仅仅研究其宏观彻体性质. 在这一点上, 上式中各个量都是明确的微观含义, 对于平衡态和下一节要介绍的近平衡态都是可以使用的. 不会造成不相等的情形.

## 11.6 近平衡态热力学 \*

从古至今,物理学对**非平衡态**(non-equilibrium state) 的研究兴趣丝毫不亚于对平衡态的研究.著名的**开尔文勋爵**(Lord Kelvin) 就曾试图处理不可逆过程.直到 20 世纪三四十年代才有**昂萨格**(Onsager) 等人建立了**近平衡态**(near-equilibrium) 的热力学与统计理论.而在六七十年代又由**普利高津**(Prigogine) 等人对**远离平衡态**(far from equilibrium) 的研究而大放异彩.远离平衡态的性质将在本书的讨论范围外,我们在这儿对近平衡态中的线性特性做一个简介.

### 11.6.1 线性输运现象

当系统处于非平衡态时,系统内部一般存在压强,温度,浓度,电势梯度.分别将引起质量<sup>10</sup>,热量,粒子与电荷的迁移,称为**输运现象**(transport phenomenon).形象地看这种因果关系,我们把各种非平衡的梯度称为**力**(force),而导致的输运称为**流**(flow).在力较小时流近似与力成正比,这个线性的区域正是近平衡态热力学所研究的区域.历史上人们相继发现了如下经验规律:

**傅里叶 (Fourier) 热传导定律**

$$\mathbf{J}_q = -\kappa \nabla T$$

其中  $\mathbf{J}_q$  表示热流密度.热交换是指通过非粒子交换,也非物质团集体运动而通过热接触,分子碰撞而交换的能量部分.而系数  $\kappa$  称为**热导率**(thermal conductivity).

**菲克 (Fick) 扩散定律**

$$\mathbf{J}_n = -D_0 \nabla n$$

其中  $\mathbf{J}_n$  表示粒子流密度.但为了热力学的统一处理我们又经常把上式写为:

$$\mathbf{J}_n = -D \nabla \mu$$

注意到  $\mu = \mu(n, T)$ ,一般一定温度下化学势都是与浓度正相关的,两种定义因此而能够互相换算  $D_0 = D \left( \frac{\partial \mu}{\partial n} \right)_T$ ,仅仅是**扩散率**(diffusivity) 的取法不同.然而如果不等温的情形如何处理?这不会带来问题,因为温度梯度也能造成粒子流,见下.

**欧姆 (Ohm) 定律**

$$\mathbf{J}_e = -\sigma \nabla \varphi$$

其中  $\mathbf{J}_e$  表示电流密度.而  $\sigma$  为导体的**电导率**(electrical conductivity).

**索瑞 (Soret) 效应**

温差传热与浓度差传粒子的现象存在耦合,温度差(而不是浓度差)也将导致粒子的扩散.这个效应称为**热迁移**(thermophoresis) 或**热扩散**(thermodiffusion).类似的可以由以下经验公式表示:

$$\mathbf{J}_n = -D_T \nabla T$$

---

<sup>10</sup>这是黏滞现象,本书将不讨论

其中  $D_T$  称为热扩散率(thermodiffusivity). 从而在傅里叶热传导实验中, 如图11.20左, 在温度梯度下实际上存在着热流与粒子流两种同时存在的现象.

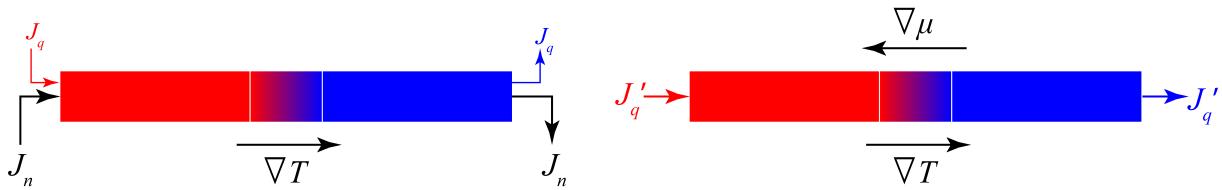


图 11.20: Soret 效应与 Dufuer 效应

### 杜福尔 (Dufuer) 效应

如图11.20右所示, 在傅里叶热传导实验中阻断两侧的粒子流, 那么由于热扩散现象的存在将造成两侧粒子的浓度差. 两个效应由此产生的粒子流也因此而达到平衡而抵消. 在此时对热流进行测量, 其值将不同于有粒子流时的值, 这可以被解释为由于粒子浓度梯度所造成的. 经验规律可以写为:

$$\mathbf{J}_q - \mathbf{J}'_q = \kappa_\mu \nabla \mu$$

### 伽伐尼 (Galvani) 接触电势差

金属内的电子体系不是一个简单的体系, 在量子理论建立后才有能带理论(energy band theory) 能给出电子性质的正确描述. 一般来讲, 不同金属内的电子可以认为具有不同的化学势  $\mu_i$ . 此时如果让两种金属 1 与 2 相互接触, 化学势高的金属, 设为 1, 其电子将自发地向化学势低的金属 2 上运动. 这直接改变的不是化学势, 而是更直接地导致了电磁学的结果—2 上电势降低而 1 上电势升高. 我们可以引入电化学势(electrochemical potential) 来表示两种驱动电子运动的效果的和, 达到平衡时应有:

$$\bar{\mu}_1 = \bar{\mu}_2 ; \quad \bar{\mu}_i = \mu_i - e\varphi_i$$

这一点也很好理解,  $\mu$  的含义为从无穷远移动一个电荷进入导体中外界所需要的功, 现在只需要把整体电荷分布所带来的电势能改变所需要的功也计及在内即可. 上式给出了两种金属的接触电势差(contact potential difference):

$$\varphi_1 - \varphi_2 = e(\mu_1 - \mu_2)$$

然而它既不能产生回路电流, 甚至也不能被直接测量: 如果把第三种金属制作的电极加在两种金属两端, 由于三个界面上都有接触电势差而使得第三种金属的两个电极上电势差为零.

### 塞贝克 (Seebeck) 效应

1821 年物理学家塞贝克发现由两种材料的导线构成的闭合回路如果把两个连接点置于不同的温度环境下则会使得小磁针偏转. 他把这个效应称为“热磁效应”, 这是因为当时人们对电磁现象之间的联系缺乏认识, 毕竟恰好在几乎同时奥斯特 (*Ørsted*) 才发现电流能够产生磁场. 之后奥斯特就意识到这个现象是典型的热电耦合现象, 展开了人们对热电耦合现象研究的序幕. 类似于温差传热与浓度差传粒子的耦合, 这个现象实际上是有以下经验规律:

$$\mathbf{J}_e = -\sigma \eta \nabla T$$

其中  $\eta$  被称为塞贝克系数(Seebeck coefficient), 而实际温差电现象有以下特点: 温差电动势不依赖于导线的长度, 不依赖于导线上温度的分布, 既使中间有连接第三种材料的导线, 只要连接点等温, 也不影响温差电动势的大小. 这很容易理解. 首先意识到接触电势差不会带来电动势, 其次在第  $i$  段导线上的非静电力为:

$$\mathbf{J}_e = \sigma \mathbf{K} \Rightarrow \mathbf{K} = -\eta_i \nabla T$$

从而两端的总电动势为:

$$\mathcal{E}_i = \int_{H_i}^{L_i} -\eta_i \nabla T \cdot d\mathbf{l} = \eta_i (T_{H_i} - T_{L_i})$$

上式以高温端向低温端为正,两个温度分别用  $T_H$  和  $T_L$  表示.通过上式发现的确这个积分与路径无关,而两端等温的导线段将不产生电动势.故如果把导线 1 和 2 连成回路,以从 1 上高温到低温,2 上低温到高温为回路正方向的回路电动势为:

$$\mathcal{E} = (\eta_1 - \eta_2)(T_H - T_L) = \eta_{12}\Delta T$$

而  $\eta_{12}$  则是实用的温度系数(temperature sensitivity).塞贝克系数是可正可负的,类似于热扩散现象,如果载流子为正电荷,如 p 型半导体,那么热扩散导致电荷从高温端向低温端迁移,系数为正.如果载流子为负电荷,如 n 型半导体,相反的系数就为负.金属的塞贝克系数就都是负的.由此现象制作的热电偶(thermocouple)是很好的宽温度范围测温仪器.下图 11.21 左演示了其半导体型热电偶的工作原理.

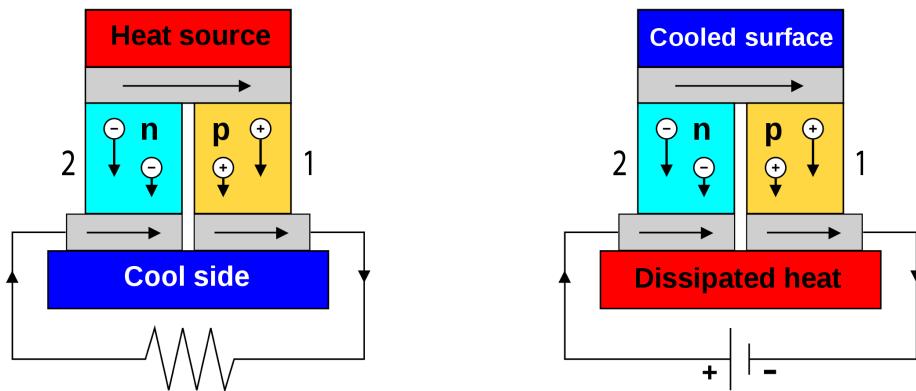


图 11.21: Seebeck 效应与 Peltier 效应

### 佩尔捷 (Peltier) 效应

类似于索瑞效应与杜福尔效应的关系,温差可以耦合电流,那么这个效应也会反过来对热传导造成影响.第一个效应是佩尔捷效应,当电流  $I$  以其正向流过以上连接点处时,连接点会产生一部分独立于热传导外的吸放热.这个热量被称为佩尔捷热,以电流从 1 到 2 为正,热量产生为正,则经验公式为:

$$\dot{Q} = (\pi_1 - \pi_2)I$$

其中  $\pi_i$  为材料  $i$  的佩尔捷系数.

### 汤姆孙 (Thomson) 效应

汤姆孙也就是开尔文爵士,他在 1851 年前后系统地研究了热电耦合的两个效应,并在理论上预言和实验上观测了连续版本的佩尔捷效应---即汤姆孙效应.这个效应只需要把上式中的热产生率换成热产生率密度,而佩尔捷系数之差理解为某种和温度有关的性质的梯度,电流改为电流密度,得到:

$$\dot{q} = -k \mathbf{J} \cdot \nabla T$$

这样一些线性现象可以代表在统计力学方法出现前人们对近平衡态研究成果.

### 11.6.2 传热的熵产生

我们先建立简单的热传导过程的热力学理论。在热传导的过程中每一处的熵增都是由于左右两侧的传热造成的，根据热力学第一定律，单位体积内的内能增加为：

$$\frac{\partial u}{\partial t} = -\nabla \cdot \mathbf{J}_q$$

而凡是内能增加即体现为熵增：

$$\delta u = \delta q = T \delta s$$

那么熵的等式可以化为：

$$\frac{\partial s}{\partial t} = -\frac{1}{T} \nabla \cdot \mathbf{J}_q$$

至此已足够明显，如果定义**熵流**(entropy flow)，它由其他流共同体现，在热传导时热流即也携带熵流：

$$\mathbf{J}_s = \frac{\mathbf{J}_q}{T}$$

那么熵流本身不能完全反应每一点处熵的增加，还要引入**熵产生**(entropy production) 的概念，比如在此种情形下为：

$$\theta = \mathbf{J}_q \cdot \nabla \frac{1}{T}$$

那么有以下熵连续性方程：

$$\frac{\partial s}{\partial t} = -\nabla \cdot \mathbf{J}_s + \theta$$

### 11.6.3 普遍理论与昂萨格倒易关系

一般来讲，熵密度的改变往往与体积内的某种“荷” $q_\alpha$  与“势” $f_\alpha$  有关，一般可以表示为：

$$\delta s = \sum_{\alpha} f_{\alpha} \delta q_{\alpha}$$

而每一种“荷”都有其连续性方程：

$$\frac{\partial q_{\alpha}}{\partial t} = -\nabla \cdot \mathbf{J}_{\alpha}$$

从而一样地我们写出：

$$\frac{\partial s}{\partial t} = -\nabla \cdot \mathbf{J}_s + \theta$$

$$\mathbf{J}_s = \sum_{\alpha} f_{\alpha} \mathbf{J}_{\alpha} ; \quad \theta = \sum_{\alpha} \mathbf{J}_{\alpha} \cdot \nabla f_{\alpha}$$

从而流就是  $\mathbf{J}_{\alpha}$  而力就是  $F_{\alpha} = \nabla f_{\alpha}$ ，那么近平衡区的线性输运特性可以表示为：

$$\mathbf{J}_{\alpha} = \sum_{\beta} L_{\alpha\beta} \nabla f_{\beta}$$

中间的经验系数构成的矩阵就是**昂萨格矩阵**(Onsager matrix)。通过对微观过程可逆性的研究，昂萨格断言：

$$L_{\alpha\beta} = L_{\beta\alpha}$$

即这个矩阵是一个对称的(半正定)矩阵，任意两种流与力之间的交叉耦合强度实际上应该相等。这就是**昂萨格倒易关系**(Onsager reciprocal relation)，由于其普适性，它经常被当成热力学第四定律<sup>11</sup>。从而整体熵产生被表示为一个二次型：

$$\theta = \sum_{\alpha\beta} L_{\alpha\beta} \nabla f_{\alpha} \cdot \nabla f_{\beta}$$

---

<sup>11</sup>本书将不涉及热力学第三定律

### 11.6.4 推证热电耦合的普遍规律

热电耦合问题中存在两种流,一是热流  $\mathbf{J}_q$ ,定义为单位面积单位时间内通过的热量.二是电流  $\mathbf{J}_p$ ,我们令载流子为正电荷,电量为  $p$ ,而把化学势约化为单位电荷的: $\zeta = \mu/p$  则  $\mu n$  项变为  $\zeta \rho \cdot \rho$  为电荷密度.在导体导电问题中有一些不同于普遍理论之处:一是有外电场的存在,它会对流做功而直接导致输运,二是化学势的变化一般可以忽略,而仅在两种导体接触处才需要考虑.我们写出此时的内能连续方程:

$$\frac{\partial u}{\partial t} = -\nabla \cdot \mathbf{J}_q - \nabla \cdot (\zeta \mathbf{J}_p) - \mathbf{J}_p \cdot \nabla \varphi$$

第一项是传热输入的能量,第二项是传入粒子所带来的能量,第三项还有外电场  $\mathbf{E} = -\nabla \varphi$  对电流所做的功.而电荷连续方程写为:

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot \mathbf{J}_p$$

熵的大小为:

$$\delta s = \frac{1}{T} \delta u - \frac{\zeta}{T} \delta \rho$$

代入化简得:

$$\begin{aligned} \frac{\partial s}{\partial t} &= -\frac{1}{T} \nabla \cdot \mathbf{J}_q - \frac{1}{T} \mathbf{J}_p \cdot \nabla (\zeta + \varphi) \\ &= -\nabla \cdot \left( \frac{1}{T} \mathbf{J}_q \right) + [\mathbf{J}_q + (\zeta + \varphi) \mathbf{J}_p] \cdot \nabla \frac{1}{T} - \mathbf{J}_p \cdot \nabla \frac{\zeta + \varphi}{T} \\ &= -\nabla \cdot (f_1 \mathbf{J}_1 + f_2 \mathbf{J}_2) + \mathbf{J}_1 \cdot \nabla f_1 + \mathbf{J}_2 \cdot \nabla f_2 \\ &= -\nabla \cdot \mathbf{J}_s + \theta \end{aligned}$$

其中第一个流是能量量纲的  $\mathbf{J}_1 = \mathbf{J}_q + (\zeta + \varphi) \mathbf{J}_p$ ,对应的“势”为  $f_1 = 1/T$ ;第二个流是电荷量纲的  $\mathbf{J}_1 = \mathbf{J}_p$ ,对应的“势”为温度约化的电化学势  $f_2 = -(\zeta + \varphi)/T$ .这样我们写出昂萨格线性关系:

$$\begin{aligned} \mathbf{J}_q + (\zeta + \varphi) \mathbf{J}_p &= L_1 \nabla \frac{1}{T} - M \nabla \frac{\zeta + \varphi}{T} \\ \mathbf{J}_p &= M \nabla \frac{1}{T} - L_2 \nabla \frac{\zeta + \varphi}{T} \end{aligned}$$

对应欧姆定律,命上式温度梯度为零,在一种导体中  $\zeta$  始终认为是一个常数,有:

$$L_2 = \sigma T$$

而对应温差电效应,我们消除外电场,让  $\varphi$  是一个常数,用塞贝克系数可以确定  $M$ :

$$M = \sigma T (\eta T + \zeta + \varphi)$$

最后再由傅里叶定律,让电流  $\mathbf{J}_p$  为零,得到:

$$L_1 = \kappa T^2 + \sigma T (\eta T + \zeta + \varphi)^2$$

最后整理得:

$$\begin{aligned} \mathbf{J}_q &= -(\kappa + \sigma \eta^2 T) \nabla T - \sigma \eta T \nabla (\zeta + \varphi) \\ \mathbf{J}_p &= -\sigma \eta \nabla T - \sigma \nabla (\zeta + \varphi) \end{aligned}$$

首先考虑伽伐尼接触电势差,显然当且仅当上式中两个梯度为零时所有流都将消失.从而相互接触的两种介质有

$$\zeta_1 + \varphi_1 = \zeta_2 + \varphi_2 \implies \varphi_1 - \varphi_2 = -\frac{1}{p} (\mu_1 - \mu_2)$$

如果是金属  $p = -e$  我们便得到了之前熟知的结果. 再考虑佩尔捷效应, 我们依然认为两介质满足以上式子, 这是因为此时由  $\nabla\varphi$  驱动的电流应该认为是发生在两种介质中, 而分界面上仍是这个固定的接触电势差. 那么对于电流从介质 1 流向 2 的情况, 再命上式中温度均匀不变, 有:

$$\mathbf{J}_{q,i} = \eta_i T \mathbf{J}_{p,i}$$

也就是说电流恰好携带着一个热流. 由内能连续性:

$$\dot{Q} = \frac{dU}{dt} = \eta_1 T I - \eta_2 T I + (\zeta_1 + \varphi_1) I - (\zeta_2 + \varphi_2) I$$

后两项表示输入输出的电流粒子带来的能量与电场力的做功, 这两项恰好消掉了. 于是得到了佩尔捷效应的公式, 并发现佩尔捷系数实际上就是:

$$\pi = \eta T$$

这是开尔文爵士当年得出的两个关系之一, 称之为第二汤姆孙关系(the second Thomson relation). 我们再研究连续情形的内能连续性, 再次使用之前的表达式:

$$\frac{\partial u}{\partial t} = -\nabla \cdot \mathbf{J}_q - \nabla \cdot (\zeta \mathbf{J}_p) - \mathbf{J}_p \cdot \nabla \varphi$$

首先根据昂萨格关系, 注意到任意情况下都有:

$$\mathbf{J}_q = \eta T \mathbf{J}_p - \kappa \nabla T$$

再注意到在我们要研究的问题中电流仍连续:

$$\nabla \cdot \mathbf{J}_p = 0$$

代入上式, 通过昂萨格关系化简便得:

$$\dot{q} = \nabla \cdot (\kappa \nabla T) + \frac{\mathbf{J}_p^2}{\sigma} - T \mathbf{J}_p \cdot \nabla \eta$$

上式第一项表示热传导积累的热, 第二项是焦耳热, 第三项就是汤姆孙热了. 可见它是由于塞贝克系数  $\eta$  不均匀而导致的, 一般就是其与温度有关<sup>12</sup>, 温度梯度也造成了一个塞贝克系数的梯度  $\nabla\eta = \frac{d\eta}{dT} \nabla T$ , 从而我们也得到了开尔文系数:

$$k = T \frac{d\eta}{dT} = \frac{d\pi}{dT} - \frac{\pi}{T}$$

上式即被称作第一汤姆孙关系(the first Thomson relation).

这就是近平衡态线性热力学理论对热电耦合现象的解释大意.

<sup>12</sup>既然  $\eta$  可以随温度改变, 那么  $\kappa$  随温度的改变其实也应考虑在内. 当然这一项是纯导热现象而与电流没有耦合.



# 第十二章 液体与固体的性质摘要

在讲解热力学基本规律的过程中,我们已经对气体,尤其是理想气体这样的热力学体系有了一定的了解.但是对于与我们的生活息息相关的液体(liquid)与固体(solid)两种凝聚态(condensed matter)我们还知之甚少.在此稍作一步展开介绍.

## 12.1 固体晶格论

固体在此处实际上特指晶体(crystal).晶体应具有长程序(long-ranged order),也就是说组成的晶体的原子,分子,或者离子应体现出大范围的空间周期性.如图12.1所示,一整块物质如果完全由严格的周期性结构组成,那么一般其表面也是规则的几何形状,称作单晶体(single crystal).但是大量的晶体内部是充满着缺陷(defect)的,无论是原子的缺失(vacancy)还是填隙(interstitial),还是更复杂的面缺陷与体缺陷,都会使得晶体的不同区域变得孤立起来,常见的金属,冰,陶土,岩石都是由大量可视作晶体的晶粒拼接而成,在物理性质上丧失了单晶体所具有的各向异性而成为多晶体(precrystal).最后还有丧失了长程序的无定形体(amorphous solid),比如玻璃,蜡,塑料等,它们都不是晶体,但一般仍认为是固体.

固体的结合方式在极大程度上决定了其热学性质,而固体的结合又在很大程度上依赖于其外层的电子在量子统计规律下的诸多行为.根据最后的结果我们可以粗浅地分类:

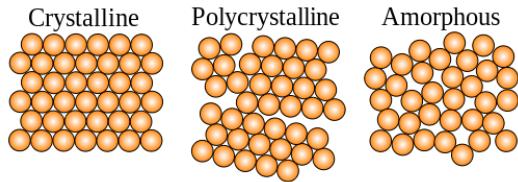


图 12.1: 单晶体, 多晶体与无定形体

晶体类型	结合方式
原子晶体	局域在原子间的共用价电子对与原子实之间的相互吸引
离子晶体	一种原子的电子几乎完全被另一种原子剥夺形成离子, 两种离子相互吸引
金属晶体	非局域的共享电子们与原子实们之间的相互吸引
分子晶体	分子与分子之间十分微弱的氢键, 范氏相互吸引, 还包含电偶极矩磁偶极矩相互作用

在某些较为简单的情形下,我们会认为固体可以拆解为两个体系,一是由原子实(atomic core)构成的点阵(lattice),一是参与原子间相互作用的所有价电子(valence electron)构成的电子气(electron gas).后者是那么地轻盈,但是却带有与原子实等量反号的电荷,这一点造成了如下的局面:体系的动能几乎由具有绝大多数质量的点阵承担<sup>1</sup>;而势能也由点阵的位置与变形来计算,对于电子的分布,总是遵循泡利不相容原理和最小势能原理,故可由点阵位置计算.这种平均场论(mean field theory)的思想使我们写出体系的总能量来:

$$E = T(\mathbf{v}_{i,g}) + V(\mathbf{r}_{i,g}) \quad , \quad i = 1, 2 \cdots n ; g \in G$$

<sup>1</sup>其实这也是个量子效应,详见下一章.

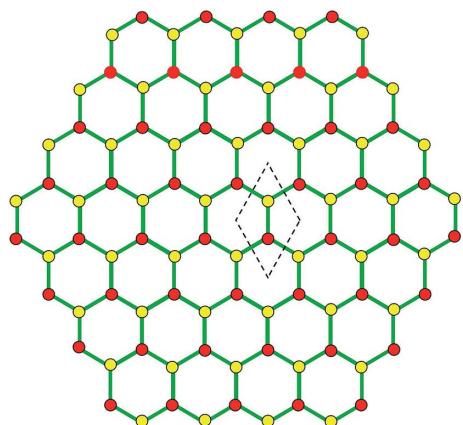


图 12.2: 石墨烯的结构

式中  $G$  表示所有原胞(unit cell) 的集合, 将一个原胞按照  $G$  的方式平移便得到了整个晶体, 称作布拉伐格子(Bravais lattice). 而每一个原胞都是全同的, 内含  $1, 2 \dots n$  号原子. 以石墨烯(graphene) 为例, 石墨(graphite) 实际上即由一层一层的石墨烯通过范德瓦尔斯力吸引而形成. 每一层可以认为是一种平面型的晶体. 其原胞由两个并不具有平移等价性的原子构成. 原胞形状为菱形, 故可以在平面上按密铺,

- 晶体分为原子晶体, 离子晶体, 金属晶体, 分子晶体...
- 晶格为晶体带来摩尔热容  $C = 3R$ , 此即杜隆-裴替定律(Dulong-Petit law)
- 晶格为晶体带来热膨胀.  $l = l_0(1 + \alpha t)$ .  $\alpha$  为线膨胀系数.  $\beta = 3\alpha$  为提膨胀系数.

• 晶格振动为集团模式, 可约化为一个个彼此独立的谐振子, 量子化即成为准粒子—声子(photon). 声子具有动量和能量, 符合  $E = \hbar\omega$ ,  $\mathbf{p} = \hbar\mathbf{k}$ . 以及色散关系  $\mathbf{v} \cdot d\mathbf{p} = dE$ .

- 可以把声子体系视作理想气体, 小动量下类似于光子气, 所有声子的速度都等于声速. 大的动量的声子本质上不应该存在, 因为其波长不能小于晶格常数. 对声子计算可以分别正确给出固体的热容和热导率. 前者在低温下正比于温度的三次方, 后者可以看做类似理想气体的输运过程类似的结果.

## 12.2 固体电子论 \*

- 固体中的电子分三类: 参与导电的自由电子, 在所有原子间自由游走<sup>2</sup>, 低温下到了固体边缘被势垒反弹回来; 参与成键的价电子, 仅在个别原子间来回; 以及与单个原子绑定的内电子. 后两者都是束缚电子.

- 经典理论为德鲁特模型(Drude model), 把电子视作类似理想气体, 在原子实间做匀速直线运动, 平均速率为  $v$ , 频繁发生碰撞, 碰撞自由程  $\lambda$ , 自由时间  $\tau$ , 经典理论给出其输运的两个特性: 电导率与热导率:

$$\sigma = \frac{ne^2\tau}{m} ; \quad \kappa = \frac{1}{3}nc\lambda v$$

- 经典理论对平均速率, 自由程与自由时间的估计都严重偏离真实值好几个量级, 电子质量也应该修正为有效质量才准确. 但最后对电导率, 热导率的估计仅差几倍, 但仍然没有解释金属的热容为什么没有因为电子热运动而增加的问题.

- 考虑到电子的波动性, 不确定原理告诉我们电子的波函数的弥散范围至少能扩展到原子的尺寸. 把电子处理为波得出了单电子的布洛赫波理论(Bloch wave theory), 而且得出了电子运动动量与能量之间的独特色散关系, 也就是电子的能带结构(energy and structure), 能量既不是像真空中电子那样连续也不是像原子内电子那样有能级, 而是介于两者之间. 价电子将一个能带填满处于满带, 故不导电; 而导电的电子处在未填满的能带.

<sup>2</sup>不一定所有方向, 比如石墨导电只发生在层内.

- 同时,与原子中的电子一样,电子要符合泡利不相容原理(Pauli exclusion principle),能带中的电子往往是从最低能量的静止态往上填充,填充到其动能约为几个电子伏的费米能级上.即使在常温下,电子在各个能级上的自带能量也会远大于热力学特征能量  $kT$ ,所以它不是符合麦克斯韦-玻尔兹曼分布,而是符合由于电子作为费米子(fermion)其微观特性带来的费米-狄拉克分布(Fermi-Dirac distribution).这样的电子体系是强简并气体(strongly degenerate gas).
- 部分绝缘体带隙(band gap)较窄,部分处于满带的电子可以因为热运动激发到空带.这就成为了本征半导体(intrinsic semiconductor).激发的电子数密度若为  $n$ ,那么载流子就为数密度  $n$  的电子和数密度  $p = n$  的空穴:

$$p \cdot n \propto e^{-\frac{V_g}{kT}} ; \quad \sigma = p e \mu_p + n e \mu_n$$

其中  $\mu = e\tau/m$  为迁移率,为速度正比于外力的系数.

- 掺杂半导体分为 p 型和 n 型,主要载流子为空穴 p for positive 和电子 n for negative.用它们制作的pn 结(pn junction)进一步就是各半导体晶体管的基元了.在 pn 结内电子的输运有两大类:扩散电流(diffusive current)和漂移电流(drift current).

## 12.3 液体的彻体性质

### 12.3.1 液体性质综述与其微观成因

液体在很多方面与气体相似,而与固体区分开来.具体来说,液体没有固定的体积易于流动;没有弹性只有黏性;黏度与压缩系数比气体倒是大不少,这得益于其接近固体的高密度;热容,热导率等更细节的内容我们后几节详细阐述;微观细节则在以后章节中加以说明.最后,我们指出特殊条件下液体也具有某些与固体相似的性质,尤其是当变形率十分的急剧时.这种现象是非牛顿流体(non-newtonian fluid)的一种典型,比如图12.3中用玉米面粉与水和成的欧不裂(oobleck)就是一种典型的一旦作用力变化十分快,就会主要表现出弹性固体的性质而不是液体的流动性的材料,所以人才可以在液面上快步走动而不会陷进去.但使我们把液体固体彻底区分开来的主要原因是固体具有序(order),而且是长程(long range)的序<sup>3</sup>.



图 12.3: 在有“弹性”的液体上行走

### 12.3.2 热容

- 微观结构:局部有序,有“类晶区”概念,径向分布函数(radial distribution function)与固体相似.但无长程序(long range order).原子分子时停时动,也区别于固体,但不像气体那样仅当碰撞时才短暂停驻片刻.
- 热容:一般来说,按原子的摩尔数计算,才有等压等体热容约为  $3R$
- 热膨胀:部分液体与晶体反常膨胀,结晶过程也不是收缩反而是膨胀,一种解释是加热使得一些固体到液体成键倾斜以后更有利在其他方向节省空间.

<sup>3</sup>凡事皆有例外,液晶(liquid crystal)就有长程序的液体,非晶体(non-crystalline)便是无长程序的固体.更准确的液体固体区分方式应该要看结合的强度.

- 黏度: 液体一般温度越高黏度越小. 这是因为不同于气体黏度, 液体黏度更像固体那样, 靠的是相互作用形成的“激发”. 而热运动加剧时液体分子间相互作用更难形成结果, 比如两个原子很难形成一种亚稳态的“集团”.

## 12.4 液体的表面性质

- 液体与气体交界面为表面(surface), 与固体交界面为界面(interface). 表面一般就若干个分子层那么厚, 表面的分子数密度由于过渡到气体中所以明显会降低. 界面则可能降低可能升高.
- 分子与分子间相互吸引形成内聚力, 它的计算一般包括在体的内能中. 但表面与界面的形成会改变能量, 叫做表面能与界面能. 它正比于其面积  $S$ . 习惯上:

$$W_s = +\sigma_s S = +\sigma_{lg} S$$

$$W_i = -\sigma_i S = -(\sigma_{ls} - \sigma_{gs}) S$$

前者  $S$  为液体表面积. 后者  $S$  为液体固体接触的界面面积, 就是说要考虑由于液体与固体的接触使得固体与气体少接触了相同的面积可能导致的  $\sigma_{gs} S$  能量, 尽管这一项几乎总是可以忽略.  $\sigma_{ls}$  是可正可负可为零的.  $\sigma_{lg}$  一般大于零.

- 表面能对应表面张力(surface tension):

$$f_s = \sigma_s L \quad ; \quad f_i = \sigma_i L$$

前者使表面尽可能小, 后者使界面尽可能大.

- 附加压强的杨-拉普拉斯方程(Young-Laplace equation):

$$\Delta p = \sigma \left( \frac{1}{\rho_x} + \frac{1}{\rho_y} \right)$$

其中  $\rho_x$  和  $\rho_y$  是可以任意选取的两个相互垂直的方向上曲面上法曲线的曲率.

- 液体固体接触的接触角(contact angle) 的杨-杜普雷方程(Young-Dupré) 方程:

$$\cos \theta = \frac{\sigma_i}{\sigma_s}$$

若  $\sigma_i > 0$  称为浸润(wetting). 若  $\sigma_i > \sigma_s$  上式无解称为完全浸润(perfect wetting), 液体会在固体表面尽可能铺展. 若  $\sigma_i \leq 0$  称为不浸润(non-wetting), 若  $\sigma_i < -\sigma_s$  称为完全不浸润(perfect non-wetting)

- 表面能实际上是自由能而不是内能, 等温等压拉伸液膜要伴随着吸放热. 详情参考之前章节的计算.

## 12.5 极端条件下的其他物态

- 超冷 ( $\sim \mu\text{K}$ ): 玻色-爱因斯坦凝聚(Bose-Einstein condensate): 宏观数目的原子的动量  $p = \hbar k$  凝聚到了基态. 其波函数可以被干涉实验探测到.
- 低温 ( $1 \sim 10\text{K}$ ): 超流(superfluid) 现象: 如 He-4 液体黏度消失. He-3 由于是费米子其超流直到 1970 年代才被发现, 需要  $\sim \text{mK}$  的低温.
- 低温物理中的高温 ( $10 \sim 100\text{K}$ ): 超导(superconductor): 常规金属的电阻率会消失, 伴随着迈斯纳效应(Meissner effect): 完全抗磁性. 电子结合成库珀对(Cooper pair), 可以像 BEC 那样通过晶格作用而凝聚成整个波函数而传播.

- 高温低密 ( $10^3 \sim 10^4$ K): 等离子体(plasma), 所有电子都与原子核不再形成束缚态, 而是构成独立的气体体系, 可以不考虑相对论效应也不考虑电子的量子简并来处理.
- 超高温, 高密度 ( $10^{11} \sim 10^{12}$ K,  $10^{17}$ kg/m<sup>3</sup>): 引力场下坍缩形成的中子星. 电子数不再守恒而是在中子质子互变的反应中产生与消灭.
- 最高温 ( $> 10^{12}$ K): 各大型对撞机上已经实现的夸克-胶子等离子体(quark-gluon plasma): 质子数和中子数不再守恒, 内部的构成的夸克与胶子形成强相互作用的等离子体.
- 每一种粒子数不再守恒的极端相对论性的气体, 其内能密度与温度的关系都是: 对于每自由度的玻色子与费米子:

$$u_b = \frac{2\sigma}{c} T^4 \quad ; \quad u_f = \frac{7}{8} u_b$$

例如对于光子气, 两个偏振自由度再结合  $J = uc/4$  便得到黑体辐射公式. 对于不包含第二代基本粒子的 QGP, 夸克有 2 味 3 色 2 自旋 2 电荷一共 24 自由度, 胶子有 8 色 2 偏振(螺度)共 16 自由度. 代入得:

$$u = \frac{74\sigma T^4}{c}$$



# 第十三章 相与相变摘要

## 13.1 相平衡

- 相平衡条件: 等温等压, 化学势相等:

$$\mu_1(T, p) = \mu_2(T, p)$$

即三种平衡同时被满足: 热学, 力学与化学平衡.

- 若  $\mu_1 > \mu_2$ , 粒子从势高的向势低的自发相变. 如某温度下未饱和的水蒸气内化学势就会低于水中的化学势.
- 除了这三个量相等, 摩尔熵  $s$ , 摆尔体积  $v$ , 数密度  $n$  两相平衡时都可以不相等.
- 相变潜热(latent heat) 指等温等压条件下的吸热, 所以是摩尔焓的增量:

$$Q = \nu l = h_2 - h_1 = u_2 - u_1 + p(v_2 - v_1)$$

- 0°C 下若把水的比热容记做  $1\text{cal/g}\cdot\text{K}$ , 那么水的溶解热与汽化热为  $80\text{cal/g}$  与  $600\text{cal/g}$ , 100°C 下汽化热降低到只有  $540\text{cal/g}$ . 冰和水蒸气的比热容几乎刚好只有水的一半  $0.5\text{cal/g}\cdot\text{K}$

## 13.2 气液相变

首先范德瓦尔斯方程(van der Waals equation):

$$p = \frac{\nu RT}{V - \nu b} - \frac{\nu^2 a}{V^2}$$

- 用范德瓦尔斯方程同时可以描述气体和液体, 低于临界温度  $T_c = 8a/27Rb$  时  $p - V$  曲线不再单调. 提供了等温等压等化学势相平衡的可能性. 事实上在麦克斯韦等面积法则(Maxwell's equal area rule) 决定的两点处能够相平衡. 这定义了气液两相的分歧背后的数学原因.

- 汽化的熵变其实简单地决定了潜热:

$$l = T(s_g - s_l)$$

- 两相共存时压强与温度的导数关系, 也就是克劳修斯-克拉伯龙关系(Clausius-Clapeyron relation):

$$\frac{dp}{dT} = \frac{s_g - s_l}{v_g - v_l} = \frac{l}{T(v_g - v_l)}$$

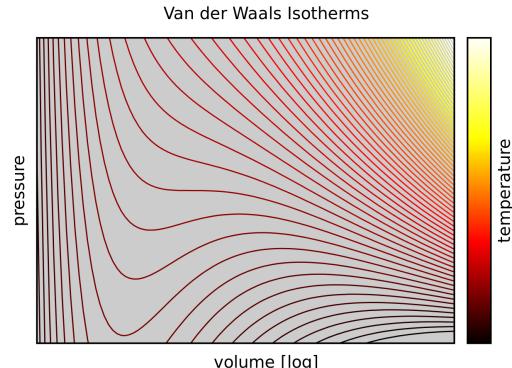


图 13.1: 范氏等温线

- **蒸发**(evaporation): 由于大气的存在,水蒸气在空气中仅是分压,不等于水的压强,但是大气压对水的压强使水压缩程度很小对水的化学势几乎可以忽略.故,可以认为某温度平衡时空气中的水汽分压就是该温度下水蒸气单独与水在该温度共存时的压强,称作**饱和蒸气压**(saturate vapour pressure). 0°C 时为 0.6kPa, 室温下变成 2 – 4kPa, 最后 100°C 变成 1atm. 蒸发条件是外界大气环境水蒸气分压不能达到该温度下的饱和蒸气压. 这样化学不平衡会让水变成蒸汽而增大水面上的总压强,但是力学不平衡又会让蒸汽转移向上方的大气形成蒸发流. 易挥发的气体会尤其明显. 高压锅也是利用这个特点,取消了力学平衡,用化学平衡来使压强单方面增加. 日常生活中的空气一般都是不饱和的,降雨的天气空气湿度可达 100% 而形成饱和的空气. 如果空气水蒸气分压较大而经历降温至**露点**(dew point) 以下,水蒸气就会析出而结露.
- **沸腾**(boiling): 当温度达到甚至高于某压强下两相共存对应的温度时. 剧烈的沸腾现象就发生了. 比如 1atm 下水在 100°C 沸腾, 此时水蒸气不断产生希望在更高的压强下实现化学平衡,但是外界大气压的力学平衡又不允许. 所以就发生了从内到外的剧烈汽化现象.

### 13.3 连续相变

### 13.4 拓扑相变

# 第十四章 统计物理基础摘要

## 14.1 数学基础

### 14.1.1 概率与独立性

现实生活中有很多偶然事件,偶然事件的成因是多种多样的,它们集中表现在相似的条件下进行试验,而能够得到完全不同的结果.数学上用抽象的集合来表示所有可能的结果:

$$x \in X, \quad X = \{x\}$$

其中每一个元素代表一种可能的结果.这些结果应该有以下特点:

**忠实性**:每一个不同的实验结果如实地反应为集合中的不同元素.也就是不能用一个元素代替一类实验结果.

**互斥性**:当一个结果发生时,另一个结果就必须排除,也就是不能有多个元素对应同样的实验结果.

**完备性**:所有的实验结果必须都有集合中的元素对应.

这样就能把集合  $X$  称为**样本空间**(sample space),而每一个元素  $x$  也称为一个**元事件**(elementary event).我们常说的**事件**(event),其实一般指样本空间的一些定义良好的子集  $A \subset X$ .只要实验结果在这个子集中,就说这个事件发生了.所谓定义良好我们可以做如下理解:

在**古典概型**(classic models) 情形下,样本空间是一个有限的集合,此时任意子集都可以视为某种事件.如投一颗骰子,样本空间为  $X = \{1, 2, 3, 4, 5, 6\}$ ,则投出偶数是一个事件  $A = \{2, 4, 6\}$ .

但在**几何概型**(geometric models) 情形下,样本空间一般具有与  $\mathbb{R}^n$  类似的结构,一般都是无限的集合,有一些“事件”的提法应当给予摒弃否则会引起矛盾.例如在闭区间  $[0, 1]$  间任取一个点,这个点恰好是有理数这样的“事件”可能就不是那么定义良好<sup>1</sup>.一般常见的事件如点落在区间  $(a, b)$  内等等.

设想同时做好几个实验,这几个实验互不干扰它们的结果是完全独立的,那么联合到一起就构成了一个大的实验,其结果应表示为一个数组:

$$\mathbf{x} = (x_i) \quad ; \quad x_i \in X_i$$

而新的样本空间称为原来那些样本空间的**独立直积**(independent product).记做:

$$\bigotimes_i X_i = \{\mathbf{x} = (x_{ij}) | x_{ij} \in X_i\}$$

互斥,独立这样的一些概念如何用数学严格表述?事实上它们恰恰是用概率去定义的.概率是一种我们关于实验结果可能性的**先验假设**(a priori presumption),它是**随机变量**(random variable) 的正实数值函数  $P$ ,而随机变量既可以取为样本空间中的单个结果,也可以取为结果的集合:事件:

$$P(A) = P(\bigcup_{x_i \in A} x_i) = \sum_{x_i \in A} P(x_i)$$

<sup>1</sup>事实上有理数集是定义良好---可测的,但有些集合不可测从而不能讨论它们的概率.参考[https://en.wikipedia.org/wiki/Non-measurable\\_set](https://en.wikipedia.org/wiki/Non-measurable_set)

这其中已用到了  $P$  的属性: **互斥事件**(mutual-exclusive events, disjoint events) 的加法原理:

$$A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$$

再加上:

$$P(A) \geq 0 ; P(X) = 1$$

就构成了一个合适的概率定义. 而对于随机变量可取连续样本空间, 比如区间  $[a, b]$  的情形, 引入**概率密度函数**(probability density function, pdf):

$$P([x, x + dx]) = p(x)dx ; p(x) > 0, \int_a^b p(x)dx = 1$$

便是一个合适的概率定义.

现在来讨论事件的独立性. 两个非互斥的事件可能同时发生, 同时发生这一个新的事件即被定义为  $A \cap B \neq \emptyset$ , 此时可以定义事件  $B$  在事件  $A$  下的**条件概率**(conditional probability):

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

显然, 若条件概率为零, 那么实际上两个事件互斥. 而实际上如果  $B$  在  $A$  下的条件概率等于  $B$  的概率, 那么这种情况称为两个事件相互独立:

$$P(B|A) = P(B)$$

注意到上式实际上也可以写为:

$$P(X)P(A \cap B) = P(A)P(B)$$

所以两个事件相互独立的确是一个相互的关系, 此时  $A$  的条件概率也有  $P(A|B) = P(A)$ . 两个事件相互独立是一种很微妙的关系, 这意味着一个事件的发生既不会阻碍另一个事件, 也不会促成另一个事件. 而是完全没有影响. 物理上看, 很有可能两个事件没有因果关系.

我们之前构造的独立直积, 用概率表示即为:

$$P((x_{ij})) = \prod_i P(x_{ij})$$

而对于连续概率分布, 两个样本空间的独立直积给出的新概率应该是有以下概率密度:

$$p(x, y) = p_1(x)p_2(y)$$

### 14.1.2 随机变量及其数字特征

物理实验中的随机现象, 常常体现为实验测量中测量结果数据的上下波动. 此时一般认为这个测量数据  $x$  就是一个随机变量, 而具有先验的某种概率密度函数  $p(x)$ . 有时候我们关心测量数据的某种函数  $y = f(x)$ , 那么不同的  $y$  其出现概率分布应该修改:

$$p_y(y) = \frac{dp}{dy} = \frac{p(x)dx}{dy} = \frac{1}{f'}p(x)$$

例如, 如果考虑在  $x$  方向的粒子速度  $v$  具有分布律:

$$p(v) = \frac{1}{\sqrt{2\pi}}e^{-\frac{v^2}{2}}$$

可以验证这个概率归一, 那么其能量  $e = \frac{v^2}{2}$  的分布为

$$p_{e+} = \frac{p(v)}{v} = \frac{1}{2\sqrt{\pi e}}e^{-e}(v > 0) ; p_{e-} = \frac{p(v)}{-v} = \frac{1}{2\sqrt{\pi e}}e^{-e}(v < 0)$$

注意到同一个能量对应着两种可能的运动方向,故把两个  $de$  对应概率分开算,最后合到一起:

$$p_e(e) = \frac{1}{\sqrt{\pi e}} e^{-e}$$

一个在统计中关心的数学处理,是在  $n$  次全同独立测量中获得的测量结果的平均数(average):

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$$

而著名的**大数定律**(law of large numbers) 指出:如果概率分布的一个重要数字特征:**期望**(expectation) 存在<sup>2</sup>:

$$\mu = \langle x \rangle = \int x dp$$

那么如果随着测量次数趋于无穷,平均数将会无限趋近于期望值:

$$\lim_{n \rightarrow \infty} P(|\bar{x} - \langle x \rangle| > \varepsilon) = 0, \forall \varepsilon > 0$$

如何理解这一点? 我们再定义另一个随机变量的数字特征:**方差**(variance), 物理上称为**涨落**(fluctuation), 数学上还称**中心二极矩**(second central moment):

$$\sigma^2 = \langle (x - \mu)^2 \rangle = \int (x - \int x dp)^2 dp$$

物理上还常常把涨落除以期望的平方称为相对涨落. 那么可以证明:

$$\langle \bar{x} \rangle = \langle x \rangle ; \quad \langle (\bar{x} - \mu)^2 \rangle = \frac{\langle (x - \mu)^2 \rangle}{n}$$

故随着实验次数增加,平均数作为新的随机变量它的期望不会变,而方差在不断减小,故有中心极限定理的说法.

以上讨论适用于任何  $y = f(x)$  型的随机变量,其中  $x$  可以是一次实验的实数结果,也可以是多次实验的数组结果,也可以是不同独立甚至非独立测量结果的复合,此时  $y$  即被理解为事件,代表某些可能发生的元事件的集合. 期望与方差被定义为:

$$\begin{aligned} \mu &= \int y dp \\ \sigma^2 &= \int (y - \mu)^2 dp \end{aligned}$$

其中  $dp$  表示  $y$  落在该点附近范围内的概率.

### 14.1.3 信息熵

为了刻画样本空间与其概率的多样性,我们引入**信息熵**(information entropy) 的概念. 对于一个样本空间与随机分布,它具有的熵被定义为:

$$S = \sum_{x_i \in X} -P(x_i) \ln P(x_i)$$

而每一种可能性  $x_i$  带来的信息量被定义为  $I_i = -\ln P(x_i) > 0$ . 而总熵是事件信息量的期望. 为什么如此定义? 比如我们设想投一枚骰子,那么投出 1 这个数的信息量应该是投出奇数的信息量加上在已知投出奇数后再投出 1 的信息量,投出奇数称作  $A$  事件那么这个性质写作:

$$I(x) = I(A) + I(x|A)$$

<sup>2</sup>值得指出,有些数学上诡异的概率分布是不存在期望的,比如典型的柯西分布:

$$p(x) = \frac{\gamma/\pi}{x^2 + \gamma^2}$$

当然这样的理想概率分布物理上几乎不存在,因为物理量总是有限的,在一定尺度下适用.

而应该有一种概率与信息量之间的普适对应,以上三个信息都可以理解为:

$$I(x) = f(P(x)), I(A) = f(P(A)), I(x|A) = f(P(x|A))$$

但三个概率本来就有:

$$P(x) = P(A)P(x|A)$$

这暗示着  $f(xy) = f(x) + f(y)$ , 在假定  $f$  性质足够好 (可以求二阶导数) 情况下:

$$\begin{aligned} & \xrightarrow{\partial_y} xf'(xy) = f'(y) \\ & \xrightarrow{\partial_x} f''(xy) + xyf'(xy) = 0 \\ & \xrightarrow{xy=u} \frac{d}{du}(u \frac{df}{du}) = 0 \\ & \Rightarrow f = C \ln u + C' \end{aligned}$$

重新代入  $f$  性质, 得  $C' = 0$ . 剩下的  $C$  可以任取, 一般理解为  $-1$ , 这样给出了正的信息量. 这个推理过程还给出了事件的信息量  $I(A)$  与条件信息量  $I(A|B)$  两个概念的定义. 根据定义, 可见  $A$  取整个样本空间的信息量就是零.

现在我们理解到信息量表示稀有程度, 具体事件发生概率越小则信息量越大. 那熵又表示什么含义呢. 它表示多样性. 我们设想如果某一种元事件  $x_i$  的可能性被发现还可以细分为  $x_{i1}, x_{i2}, \dots, x_{im}$  种子可能性. 那么根据此式所引入的熵就会因此而增加, 增量为:

$$\begin{aligned} \Delta S &= \sum_{j=1}^m -P(x_{ij}) \ln P(x_{ij}) - (-P(x_i) \ln P(x_i)) \\ &= \sum_{j=1}^m -P(x_{ij}) \ln P(x_{ij}) - \sum_{j=1}^m -P(x_{ij}) \ln P(x_i) \\ &= \sum_{j=1}^m -P(x_{ij}) \ln \frac{P(x_{ij})}{P(x_i)} \\ &= \sum_{j=1}^m P(x_{ij}) I(x_{ij}|x_i) \\ &= P(x_i) \cdot \sum_{j=1}^m -\frac{P(x_{ij})}{P(x_i)} \ln \frac{P(x_{ij})}{P(x_i)} \\ &= P(x_i) \cdot \sum_{j=1}^m -P(x_{ij}|x_i) \ln P(x_{ij}|x_i) \\ &= P(x_i) S(x_i) \end{aligned}$$

一方面, 等于细化这一元事件所增加的信息量的概率平均. 另一方面, 也等于事件熵与概率的乘积. 此处也定义了事件细分熵:

$$S(A) = \sum_{x_i \in A} -P(x_i|A) \ln P(x_i|A)$$

可见单个元事件的细分熵是零. 而如果把样本空间可以分为一系列不相交的事件:

$$X = \bigcup_i A_i ; \quad A_i \cap A_j = \emptyset (i \neq j)$$

那么有:

$$S(A) = \sum_i P(A_i)[I(A_i) + S(A_i)] = S(\{A_i\}) + \sum_i P(A_i)S(A_i)$$

可见把不相交的事件并在一起时, 总熵就是把  $A_i$  视作元事件的熵加上各个子事件的熵的加权和.

最后让我们来看看作为以上熵的定义所符合的一个十分重要的性质.为了方便起见我们把样本空间记做  $\Omega$ .并在样本空间的每个元事件引入两个随机变量  $x_i \in X, y_j \in Y$ .还要求两个随机变量独立:

$$P(x_i, y_j) = P(x_i)P(y_j), \forall i \forall j$$

那么对于整个样本空间按照随机变量  $X$  来划分的熵与  $Y$  的熵:

$$S_X = \sum_i -P(x_i) \ln P(x_i) ; S_Y = \sum_j -P(y_j) \ln P(y_j)$$

而总熵:

$$S = \sum_{ij} -P(x_i, y_j) \ln P(x_i, y_j)$$

计算其关系,容易发现:

$$\begin{aligned} S &= \sum_{ij} -P(x_i, y_j) \ln P(x_i)P(y_j) \\ &= \sum_{ij} -P(x_i, y_j) \ln P(x_i) + \sum_{ij} -P(x_i, y_j) \ln P(y_j) \\ &= \sum_i \sum_j -P(x_i, y_j) \ln P(x_i) + \sum_j \sum_i -P(x_i, y_j) \ln P(y_j) \\ &= \sum_i -P(x_i) \ln P(x_i) + \sum_i -P(y_j) \ln P(y_j) \\ &= S_X + S_Y \end{aligned}$$

即,独立随机变量的熵是可以直接相加的.而我们再来分析一下若是非独立随机变量的叠加导致的熵:

$$\begin{aligned} S_X + S_Y - S &= \sum_{ij} -P(x_i, y_j) \ln P(x_i) + \sum_{ij} -P(x_i, y_j) \ln P(y_j) - \sum_{ij} -P(x_i, y_j) \ln P(x_i, y_j) \\ &= \sum_{ij} -P(x_i, y_j) \ln \frac{P(x_i)P(y_j)}{P(x_i, y_j)} \\ &\geq -\ln \sum_{ij} P(x_i, y_j) \cdot \frac{P(x_i)P(y_j)}{P(x_i, y_j)} \\ &= -\ln \sum_{ij} P(x_i)P(y_j) \\ &= -\ln 1 \\ &= 0 \end{aligned}$$

其中用到了  $\ln$  是上凸函数,所以自变量的加权平均处的函数值总是大于函数值的加权平均:

$$\ln(\sum_i p_i x_i) \geq \sum_i p_i \ln x_i ; \sum_i p_i = 1$$

以上结论告诉我们非独立的两个随机变量所带来的熵有一部分是重合的,从而导致总的熵小于两个随机变量熵的和.小的这一部分叫做两个变量之间的互信息(mutual information):

$$M = S_X + S_Y - S$$

互信息代表着已知一个随机变量的结果以后对另一个随机变量条件概率分布的影响.例如考虑自然光在某个方向的分振幅,以  $x$  方向的振幅作为第一个随机变量,如果另一个方向也是  $x$ ,那么互信息恰好等于  $x$  方向振幅概率分布的熵;如果另一个方向为  $y$  方向,那么两个随机变量独立,从而互信息为 0;如果选取一个与  $x$  方向夹锐角的方向,则  $x$  方向的振幅在一定程度上也会影响该方向的振幅的概率分布,从而两个随机变量之间具有一定的互信息量,并随着夹角增大到 90 度而消失.

- 概率模型: **经典概型**(classic models) 中原子事件构成的集合是有限的, 一般为等概率分布. 也可以想象不是等概率的样本空间. 事件是部分原子事件构成的子集. 一般写作:

$$x \in \Omega, \quad \Omega = \{\omega\}$$

$$P(A) = P(\bigcup_{\omega_i \in A} \omega_i) = \sum_{\omega_i \in A} P(\omega_i)$$

- 概率模型: **几何概型**(geometric models) 中的样本空间是与  $\mathbb{R}^n$  具有类似的结构的连续空间, 一般认为参数化. 比如原子事件抽象为区间  $[a, b]$  内的一个数, 或是二维球面  $S^2 = \{(\theta, \varphi) | \theta \in (0, \pi), \varphi \in [0, 2\pi]\} \cup \{\theta = 0, \theta = \pi\}$  上的一个点. 而概率则是一个归一化的**概率分布函数**(probability distribution function):

$$\int_a^b dp = \int_a^b p(x)dx = 1$$

$$\oint_{S^2} dp = \oint_{S^2} p(\theta, \varphi)dS = 1$$

- 独立事件**(independent events): 对于经典概型, 事件  $A$  与事件  $B$  独立:

$$P(A \cap B) = P(A)P(B)$$

如果把事件的交写作直接的乘积  $A \cap B = AB$ , 事件的补  $\Omega \setminus A$  写作  $\bar{A}$ , 上式等价于以下四式:

$$P(AB) = P(A)P(B) \quad ; \quad P(A\bar{B}) = P(A)P(\bar{B})$$

$$P(\bar{A}B) = P(\bar{A})P(B) \quad ; \quad P(\bar{A}\bar{B}) = P(\bar{A})P(\bar{B})$$

在样本空间定义一个函数变量  $X : \omega_i \mapsto x_i = X(\omega_i)$ , 那么可以根据不同的函数结果对样本空间进行分类, 每一个结果就是一个事件, 概率为  $P(x_i)$ . 此时变量  $X$  与变量  $Y$  独立即定义为:

$$\forall x_i \forall y_j : P(x_i y_j) = P(x_i)P(y_j)$$

对于定义在  $(x, y) \in \mathbb{R}^2$  上的几何概型, 概率分布关于两个坐标独立的条件为可以找到两个坐标各自的概率分布函数:

$$\forall x \forall y : p(x, y) = p_x(x)p_y(y)$$

- 两种概型下变量  $X$  的**期望**(expectation)  $E = \langle X \rangle$  与**方差**(variance)  $\sigma^2 = \langle (X - E)^2 \rangle$  的定义:

$$E = \sum_{\omega_i \in \Omega} P(\omega_i)X(\omega_i) = \int_{\Omega} X(x)dp$$

$$\sigma^2 = \sum_{\omega_i \in \Omega} P(\omega_i)(X(\omega_i) - E)^2 = \int_{\Omega} (X(x) - E)^2 dp$$

- 两种概型下**统计熵**(statistic entropy) 或**信息熵**(information entropy) 的定义:

$$S = - \sum_{\omega_i \in \Omega} P(\omega_i) \ln P(\omega_i) = - \int_{\Omega} \ln P(x)dp$$

- 在  $\mathbb{R}$  上的分布若要求  $\sigma^2$  取固定值, 那么**正态分布**(normal distribution) 可以使得熵最大:

$$p(x) = N(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

如果还要求期望为零, 那么上式  $\mu = 0$ . 注意这等价于说, 如果一个一维粒子的平均动量为 0 而平均动能为  $\langle mv^2/2 \rangle = kT/2$ , 熵最大的分布为:

$$f(v) = \sqrt{\frac{m}{2\pi kT}} e^{-\frac{mv^2}{2kT}} \quad ; \quad S = \ln \sqrt{\frac{m}{2\pi kT}} - \frac{1}{2}$$

- 如果已知  $\mathbb{R}^2$  上概率分布  $p(x, y)$  的边缘分布(marginal distribution)  $p_x, p_y$ :

$$p_x(x) = \int_{-\infty}^{\infty} p(x, y) dy$$

$$p_y(y) = \int_{-\infty}^{\infty} p(x, y) dx$$

那么在  $\mathbb{R}$  分别以  $p_x, p_y$  作为分布函数的分布的熵与二维分布的熵满足:

$$S \leq S_x + S_y$$

成立条件是  $X$  与  $Y$  随机变量独立:

$$p(x, y) = p_x(x)p_y(y)$$

这就是说,两个随机分布合成时,独立合成能够使得熵最大.

## 14.2 统计假设

原则上来说,经典的物理体系应当是确定性(certainty)的,用经典力学语言可以描述(质点系的位置与动量)的,不包含随机性(randomness)的,不应引入概率描述的系统.然而一方面由于体系在大粒子数自由度数下所体现出来的统计复杂性,即使初态是一个高度特殊的状态:比如一个长方体气缸内理想气体初始时刻所有粒子排列为规则的点阵,具有完全相同的速度矢量.在历经一段时间之后也会体现出完全的随机性:体系在很长一段时间内看来各个状态都没有明显的序(order),显得杂乱无章.而稍微有序的状态几乎不可能再次出现.这期间的过程与演化,原则上说理应是确定性的过程,如果我们在某一时刻令所有粒子的速度方向反向,这个体系将会重新回到最初的状态.

但是我们将研究随机过程(stochastic process),在一个随机过程中,即使上一刻只有唯一的一种状态,下一刻的状态也不是决定性的,而是体现出不确定性或是随机性.从而对一个时刻系统完整的描述就要包含在各种可能状态下的概率,这实际上是一个系综(ensemble).用系综而不是单个的系统描述,从而引入概率的原因有以下几个:

- 外在随机性:多自由系统是混沌系统,初态微小的改变将会随时间以指数形式迅速放大.热力学体系总是处在外界环境中,外界总是对体系有各种各样的微小扰动.即使在之前的情况下在某个时间让所有粒子速度反向,也会因为外界对体系的影响而无法观察到体系回到初始状态.
- 内在随机性:以衰变为例,如果认为衰变发生在一个瞬间,粒子只可能处于衰变前或衰变后两个不同状态,那么衰变就是一个随机的过程.每一小段时间粒子均以均等的概率衰变,那么大量全同未衰变的粒子构成的体系的未衰变数期望随时间将指数衰减,但是每一个时刻未衰变粒子数仍然具有某种概率分布,粒子数对于其期望仍具有某种方差.事实上原子与原子的碰撞涉及到量子力学行为,同样的初始条件是会导致不同的结果的.
- 等概率原理(principle of equal a-priori probability):平衡态分布在相空间等能量曲面上所有点处的概率分布函数都应该相等.例如大量摆角固定的单摆构成的热力学体系,显然摆球出现在最高点的概率要大于最低点的概率(最高点速度慢停的久).但是如果把其摆动的  $p - x$  空间画出来合适的比例下相轨迹为圆,而且一定做匀速圆周运动,所以肯定单位圆弧上出现摆球的概率相等.这个结论是理论力学里通过刘维尔定理(Liouville's theorem)的一种合情推理,但一般作为平衡态统计力学的最初假设.
- 通过等概率原理可以证明,如果理想气体也符合这个原理的话,平衡态分子速度作为随机变量分布为麦克斯韦分布律.这个推导参考热力学统计物理书籍的系综理论.

- 通过等概率原理 + 求最可几分布(most probable distribution) 也可以得到实际的分布为麦克斯韦分布律.这个推导也参考热力学统计物理书籍的最可几分布.
- 玻尔兹曼 H 定理(Boltzmann H theorem): 通过分子混沌假说(molecular chaos hypothesis),认为从非平衡态分子通过碰撞向平衡态转变的过程中,每一时刻都存在的概率分布,即,真实世界的情况只是万千可能情况中的一种.而每一次分子的碰撞都是把单一的初态的可能性变成了更多的末态的可能性.数学上很容易证明,这样的概率演化过程熵只增不减.从而末态的熵一定达到了极值.再根据上一章的论述,粒子在三个方向的速度分布一定是正态分布,且彼此独立,平均动能一致时熵最大.也给出了麦克斯韦分布律.

### 14.3 麦克斯韦分布律

三维麦克斯韦速度分布律是指:

$$f(v) = \left(\frac{m}{2\pi kT}\right)^{\frac{3}{2}} e^{-\frac{mv^2}{2kT}} = \prod_1^3 \sqrt{\frac{m}{2\pi kT}}$$

### 14.4 能均分定理

### 14.5 功, 热, 熵

### 14.6 量子与相对论

# 第十五章 光波与光线

## 15.1 界面上的反射与折射

### 15.1.1 光波与光线

众所周知,光是电磁频谱的一段. 我们所说的光一般是指可见光(visible light), 学术上代表波长为 400 – 700nm 的电磁辐射. 它大致代表了人对光的色视觉的界限. 而一般的光学研究还包括了波长大至 1mm 左右的红外光(infrared light, IR) 与短至 10nm 的紫外光(ultraviolet light, UV).

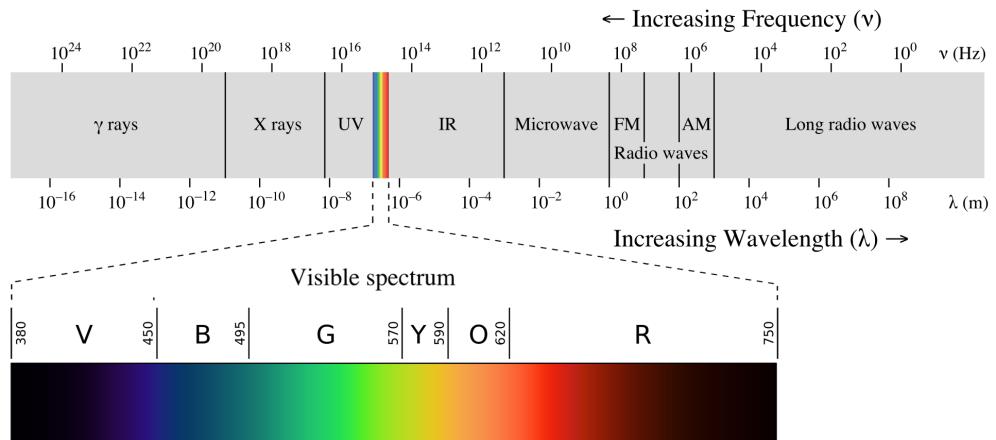


图 15.1: 光学研究范围

光的波动理论建立在众多干涉衍射实验上, 由 19 世纪后半页麦克斯韦 (Maxwell), 韦伯 (Weber) 与赫兹 (Hertz) 等人的理论与实验工作进行总结与升华. 最后通过迈克尔孙 (Michelson) 的干涉实验和随之由爱因斯坦 (Einstein), 洛伦兹 (Lorentz) 等人建立的狭义相对论来完善与归入到更加普适的经典电磁场论下. 然而整个 19 世纪实际上都处于光的粒子学说与波动学说的无休止争论中. 这是因为此时发现的光的几乎所有现象都可以同时用两种学说进行解释. 今天我们知道光实际上也有波粒二象性 (particle-wave duality), 最准确的描述光的语言应该是量子电动力学 (quantum electrodynamics, QED). 光是一份份的量子场, 被称为光子 (photon). 几何光学问题, 用经典的电磁场模型给出的描述最为精确, 而在特殊场合或者精度要求不高的情况下, 粒子光学的处理也是十分合适的.

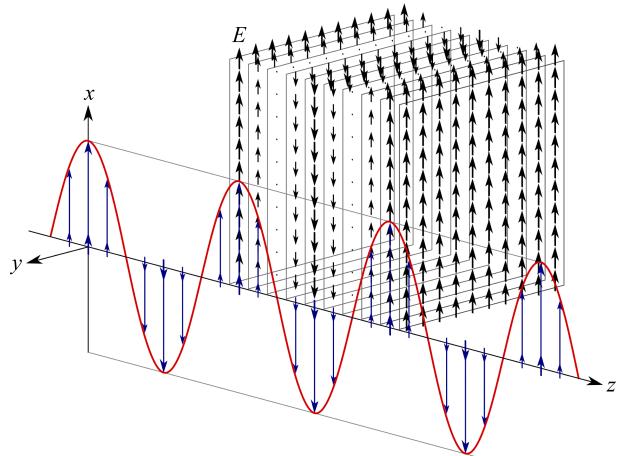


图 15.2: 平面偏振电磁波

真空中的平面波(plane wave) 提供了一种最为简单的情形, 如图, 此时  $\mathbf{E} = E\mathbf{e}_x \cos(\omega t - kz)$ . 按照光学惯例, 我们把它写作一个复数, 模就是在每一点处场强的振幅, 而幅角我们取为  $\cos$  里的相位宗量的相反数, 而它的实部表示物理上的场强:

$$A(z, t) = E e^{i(kz - \omega t)}$$

有几点值得注意: 首先, 之所以取相反数, 是因为在频率甚高的光频波段我们不太用关心每一个时空点处场随时间  $t$  的变化, 而是不同时空点  $z$  对振幅与相位的影响, 故把  $kz$  项提前. 而对于  $\omega t$  项由于每一点的振动都含这个随时间振动的因子, 而且角频率都是  $\omega$ (单色波), 所以我们往往又省去这个因子写成  $A = E e^{ikz}$ , 相当于仅考虑  $t = 0$  时的初相位与振幅. 这种操作意味着指数上的宗量随时间减小, 这与我们平时对相位随时间增加的理解是相反的. 其次, 我们取消了其矢量性, 我们知道沿特定方向  $z$  传播的平面单色电磁波可以分解为  $x$  方向的偏振与  $y$  方向的偏振, 两者的同相反相组合可以造成不同方向的线偏振模式(linear polarized mode), 而相差  $\frac{\pi}{2}$  的组合则可以造成不同的圆偏振模式(circular polarized mode). 但这些只有在特殊的光学器件下进行偏振光干涉实验才能发现区别, 故在一般的几何光学条件下我们只关心光的传播方向, 相位与振幅信息, 而不关心其偏振时用标量  $A$  来代表任意模式下约化的光的振幅大小,  $A$  具有电场强度的量纲<sup>1</sup>, 相同的  $A$  代表相同的光强与能流.  $A$  被称为标量振幅(scalar amplitude).

平面波是其他一大类波在局部可以采取的近似, 我们把等相位的面在局部近似为平面, 则波传播方向为垂直于平面的法线方向, 用波矢(wave vector) $\mathbf{k}$  代表这个方向, 它的大小就是上述平面波的  $k$ , 代表单位长度所改变的相位大小, 正如角频率(angular frequency) 代表单位时间所改变的相位大小. 两者具有如下关系:

$$\frac{\omega}{k} = \frac{\nu}{\sigma} = \frac{\lambda}{T} = \frac{c}{n}$$

其中  $\nu$  表示频率(frequency), 单位时间所经历的振动周期数, 由于一个周期相位改变  $2\pi$ , 故  $\omega = 2\pi\nu = \frac{2\pi}{T}$ , 同理,  $\sigma$  表示波数(wave number), 有  $k = 2\pi\sigma = \frac{2\pi}{\lambda}$ ,  $\lambda$  表示波长(wave length) 而  $T$  表示周期(period). 这里考虑到介质中的传播, 波速小于真空中的波速,  $n$  为介质的折射率(index of refraction).

而我们所要研究的更加普遍的一类光波一般写为:

$$A(\mathbf{r}) = E(\mathbf{r}) e^{i\varphi(\mathbf{r})}$$

这类光波的共性是仍然以统一地角频率  $\omega$  振动, 也就是所谓的单色波(monochromatic wave)<sup>2</sup>. 所以我们省去包含时间的因子  $e^{-i\omega t}$ , 直接写出只依赖于位矢  $\mathbf{r}$  的波动形式. 一般来说  $E$  是  $\mathbf{r}$  的缓变函数, 而  $\varphi$  则是  $\mathbf{r}$  的快速变化函数, 且其变化要与波矢大小  $k = \omega/c$  相吻合:

$$\nabla\varphi = \mathbf{k}$$

例如, 波源在  $\mathbf{r} = 0$  的球面波(spherical wave) 有  $E = \frac{u}{r}$ ,  $\varphi = kr$ , 而波源在  $\rho = 0$  的柱面波(cylindrical wave) 有  $E = \frac{u}{\sqrt{\rho}}$ ,  $\varphi = k\rho$ .

<sup>1</sup>这是因为与物质发生相互作用而感光时电场的作用占据主导地位.

<sup>2</sup>此时波的波动方程变为亥姆霍兹方程:

$$\left[ \nabla^2 - \frac{1}{(c/n)^2} \frac{\partial^2}{\partial t^2} \right] A e^{-i\omega t} = 0 \Rightarrow (\nabla^2 + k^2) A = 0$$

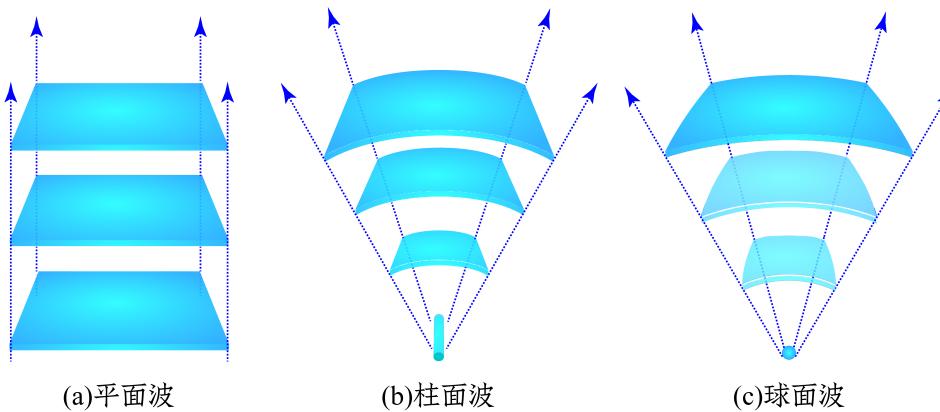


图 15.3: 三种常用波形式

显然这些理想的波不能看成是生活中的一般光源的波特性,一言以蔽之,日常光波具有很差的相干性(coherency).来自不同位置的点光源的不同偏振,不同频率,不同持续时间的波列非相关叠加,使得需要在十分仔细的实验条件下才可以观察到干涉衍射等现象.但这些对一般的几何光学性质没有影响,几何光学关心光的传播与光的强度,于是抽象出所谓的光线(ray)的概念,光线  $L$  始终沿着波传播的方向,与等相位面垂直:

$$L : \quad \mathrm{d}l // \mathbf{k} = \nabla \varphi$$

对于以上理想的波光线的方向是显而易见的(图中虚线方向),而强度(intensity),可以定义为  $I = A^2$  平面波不变,柱面波与半径成反比,球面波则与半径的平方成反比.而日常生活中的线状点状光源其性质在几何光学意义下是完全相同的.我们也可以把光的传播视为光线的集合,如果遇到界面则会发生折射与反射,折射率的不均匀导致光线的偏折,而波前可以被部分的障碍物遮挡,没被遮挡的部分仍然以光线的方式向前传播.

也就是说,在几何光学研究中,光线是表,其实作为电磁场的光波才是里.光在两种介质表面的反射(reflection)与折射(refraction)就是波动理论应用的绝佳例证.

### 15.1.2 菲涅尔公式

如图,在一个典型的界面上反射与折射的过程中,反射光,折射光的强度实际上是与入射光的偏振状态有关的.这一点在接下来的计算中也可以表明.我们按照惯例,取两个独立的线偏振为:

- **横电波**(TE wave),即垂直偏振波(<sup>3</sup>s-wave, s-polarized wave),它的电场方向垂直于主平面<sup>4</sup>.而磁场方向则在主平面内.
- **横磁波**(TM wave),即平行偏振波(<sup>5</sup>p-wave, p-polarized wave),它的电场方向平行与主平面.而磁场方向垂直主平面.

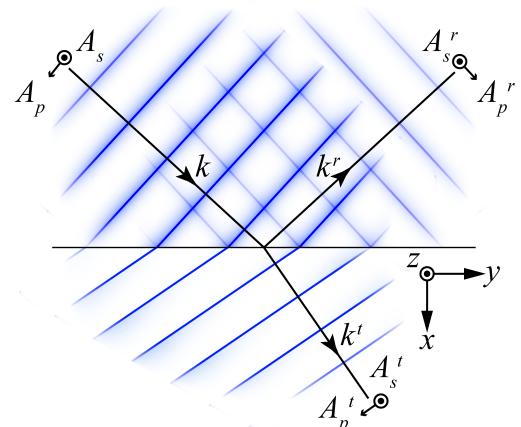


图 15.4: 界面与两种模式

<sup>3</sup>来自德语词 senkrecht: 垂直.

<sup>4</sup>由入射线与界面法线确定的平面,显然,出射波与折射波光线方向也在这个平面内.

<sup>5</sup>来自德语(英语)词 parallel: 平行.

它们的正方向如图所示, 满足  $\mathbf{A}_s \times \mathbf{A}_p // \mathbf{k}$ . 而反射折射系数则定义为在  $(0, 0, 0)$  点处的界面两侧各光振幅比:

$$A_{s,p} = E_{s,p} e^{i(\mathbf{k} \cdot \mathbf{r} - \omega t)}, \quad A_{s,p}^r = E_{s,p}^r e^{i(\mathbf{k}^r \cdot \mathbf{r} - \omega^r t)}, \quad A_{s,p}^t = E_{s,p}^t e^{i(\mathbf{k}^t \cdot \mathbf{r} - \omega^t t)}$$

$$r_{s,p} = \frac{E_{s,p}^r}{E_{s,p}}, \quad t_{s,p} = \frac{E_{s,p}^t}{E_{s,p}}$$

对于  $s$ -光, 电磁场在界面两侧切向电场连续, 切向磁场强度连续<sup>6</sup>. 而对于  $p$ -光, 切向也有电场连续, 法向可以用电位移连续. 注意到边界条件对任意  $t$  成立, 这给出了:

$$\omega = \omega^r = \omega^t$$

也就是折射反射不能改变光的频率, 也就是颜色. 从光量子论的角度理解这代表弹性(elastic) 作用, 光子的能量将不受改变, 介质不会以某种共振的方式吸收入射光的能量. 上式还意味着  $k = \frac{\omega}{c/n} = n \frac{\omega}{c} = nk_0 \propto n$ , 介质内的波矢比同样颜色的真空光要大折射率  $n$  倍.

而这个边界条件对任意界面  $y$  坐标也是恒成立的, 这给出了:

$$k_y = k \sin \theta = k^r \sin \theta^r = k^t \sin \theta^t$$

即折射时平行于界面方向的波矢是不变的. 光量子论下可以理解为这个方向动量不变, 没有受到沿平行界面方向的相互作用. 但注意到  $k$  与  $n$  的正比关系, 将折射方折射率记做  $n'$ , 折射角记做  $\theta'$  我们得到了著名的斯涅耳定律(Snell's Law):

$$\theta^r = \theta; \quad n \sin \theta = n' \sin \theta'$$

再注意到在光学问题中我们一般只讨论透明介质, 它们磁导率十分接近真空  $\mu \approx \mu_0$ , 在此时实际上  $\frac{c}{n} = \frac{1}{\sqrt{\epsilon \mu}}$  给出了  $n = \sqrt{\frac{\epsilon}{\epsilon_0}} = \sqrt{\epsilon_r}$ . 而此时实际上电位移振幅就是  $D = \epsilon E = n^2 \epsilon_0 E \propto n^2 E$  磁感应强度振幅就是  $B = \sqrt{\epsilon \mu} E = \frac{n}{c} E \propto n E$ , 磁场强度  $H$  与它只差一个真空磁导率  $\mu_0$ . 依此我们写出  $s$ -光与  $p$ -光的边界条件:

$$E_s + E_s^r = E_s^t; \quad n E_s \cos \theta - n E_s^r \cos \theta = n' E_s^t \cos \theta'$$

$$E_p \cos \theta - E_p^r \cos \theta = E_p^t \cos \theta'; \quad n^2 E_p \sin \theta + n^2 E_p^r \sin \theta = n'^2 E_p^t \sin \theta'$$

解之, 我们得到的则是著名的菲涅尔公式(Fresnel's equations):

$$r_s = \frac{n \cos \theta - n' \cos \theta'}{n \cos \theta + n' \cos \theta'}; \quad t_s = \frac{2n \cos \theta}{n \cos \theta + n' \cos \theta'}; \quad t_s - r_s = 1$$

$$r_p = \frac{n' \cos \theta - n \cos \theta'}{n' \cos \theta + n \cos \theta'}; \quad t_p = \frac{2n \cos \theta}{n' \cos \theta + n \cos \theta'}; \quad \frac{n'}{n} t_p - r_p = 1$$

都满足:

$$n(1 - |r_{s,p}|^2) \cos \theta = n' |t_{s,p}|^2 \cos \theta'$$

其中反射率(reflectance) 与透射率(transmittance) 分别被定义为:

$$R_{s,p} = |r_{s,p}|^2; \quad T_{s,p} = \frac{n' \cos \theta'}{n \cos \theta} |t_{s,p}|^2$$

透射率与透射振幅系数  $t_{s,p}$  关系前多出来的系数, 一是由于在介质中能流密度正比于折射率  $I \propto |\mathbf{E} \times \mathbf{H}| \propto n E^2$ , 二是折射后方向改变, 两侧对应光束部分的横截面积也会发生改变. 下图展示了空气 ( $n = 1$ ) 与玻璃 ( $n = 1.5$ ) 界面上的外入射(external incidence) 与内入射(internal incidence) 时的情形.

<sup>6</sup>法向电位移为零.

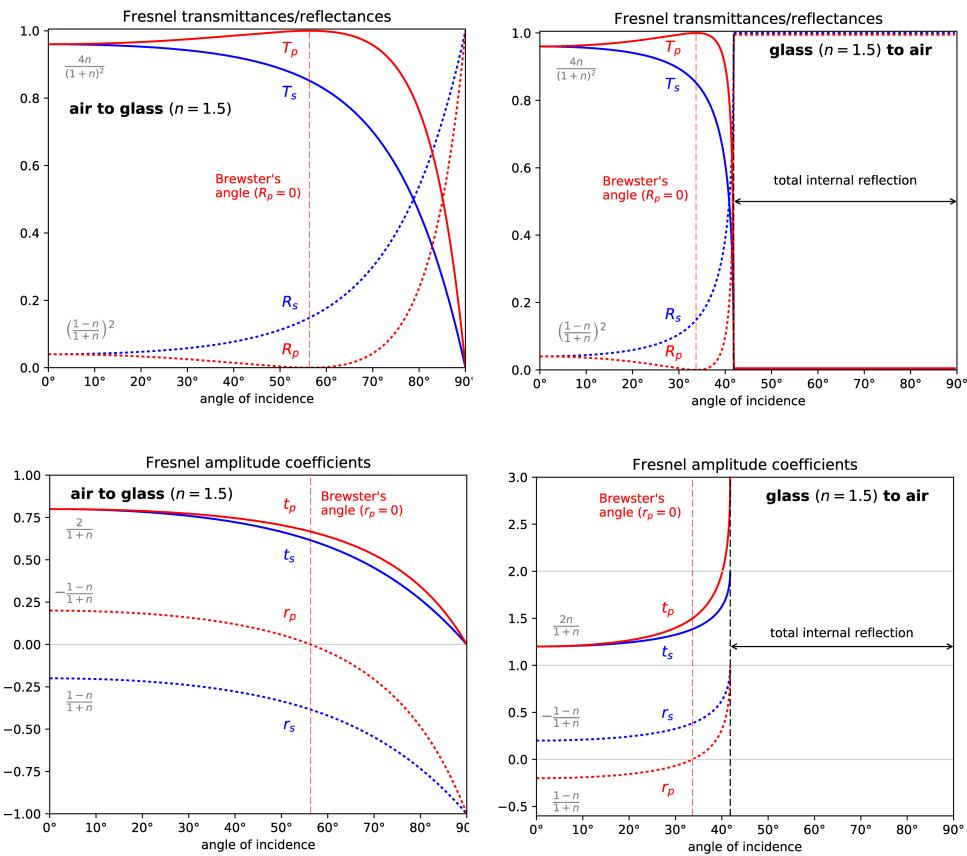


图 15.5: 界面反射与折射

有几点值得注意:

### 半波损失发生条件

当正入射时,反射光可以与入射光发生叠加. 我们关心在  $x = 0$  处,也就是界面上的叠加结果. 当  $n' > n$  时,由上图可见  $r_p > 0$  而  $r_s < 0$ . 但注意到在此处我们预设的入射反射电场正方向对  $p$ -光本来也相反而  $s$ -光相同. 故事实上都是相干相消. 发生了半波损失(half-wave loss). 而若  $n' < n$  时情况恰好相反,两种光都是相干相长,没有半波损失.

而对于入射角为  $90^\circ$  的掠入射情形,观察可知  $r_{s,p} = -1$ , 即反射光完全与入射光交叠区域几乎完全相干相消. 这个结论对两种偏振模式与两种入射方式均给出了半波损失的结果.

注意透射光几乎在所有情形下都与入射光同相. 唯一的例外发生在全内反射情况下. 之后将仔细讨论之.

### 布儒斯特角

注意到:

$$r_p = \frac{n' \cos \theta - n \cos \theta'}{n' \cos \theta + n \cos \theta'} = \frac{\sin \theta \cos \theta - \sin \theta' \cos \theta'}{\sin \theta \cos \theta + \sin \theta' \cos \theta'} = \frac{\tan(\theta - \theta')}{\tan(\theta + \theta')}$$

可以发现在  $\theta + \theta' = \frac{\pi}{2}$  时  $r_p = 0$ , 横电的  $p$ -光将发生消光的情形. 这种特殊的情况在内入射与外入射时都可以发生. 对应的入射角称为布儒斯特角(Brewster's angle). 其原因可以这样理解: 将入射波变为折射波与反射波实际上是介质电偶极子在电场

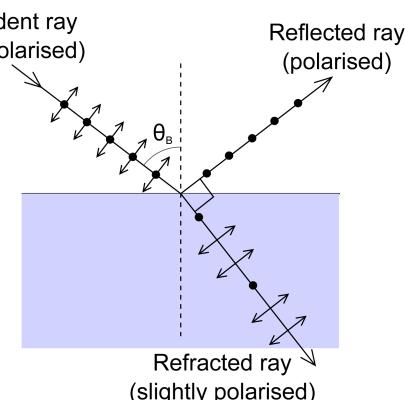


图 15.6: 布儒斯特角

驱动下振动的集体结果. 而对于  $p$ -光, 反射方向恰恰是介质内电场激发的偶极振动的振动方向. 按照电磁波的辐射理论, 这个方向上恰好是辐射强度为零的. 所以恰好反射波消失.

依照这一种特性制作出来的布儒斯特窗(Brewster window) 常见于各种气体激光器的内置谐振腔或发射口. 它将一块透明介质倾斜至布儒斯特角的位置, 它有效地将  $s$ -光进行反射, 而  $p$ -光没有反射的直接出射, 使得出射激光一般都是沿  $p$ -方向高度偏振的.

### 全内反射与隐失波

在内入射状态下如果入射角足够大, 根据斯涅耳定律将无法确定折射角  $\sin \theta' > 1$ . 这种情形为全内反射(total internal reflection). 因为接下来的计算可以表明反射系数  $R_{s,p} = 1$ . 而实际上此时并不是完全没有折射光场, 而是折射光场被局限在有光照射的表面区域无法传播开来.

首先, 折射光场由于边界条件在边界上每个时空点都成立, 它依然符合:

$$\omega' = \omega, \quad k'_y = k_y = k \sin \theta$$

但此时由于  $x, y$  方向波矢仍然满足  $k_x'^2 + k_y'^2 = k'^2$ , 而由于入射角足够大使得  $k_y > k'$ , 所以实际上  $k'_x$  只能成为一个虚数:

$$k'_x = \pm i\kappa; \quad \kappa = \sqrt{k^2 \sin^2 \theta - k'^2} = \frac{\omega}{c} \sqrt{n^2 \sin^2 \theta - n'^2}$$

取正号还是取负号呢? 这实际上取决于我们想研究的问题. 在此处我们关心  $k_x = +i\kappa$  的解. 这是因为这使得相位因子  $e^{ik_x x}$  成为衰减因子  $e^{-\kappa x}$  而不是正的指数爆炸因子<sup>7</sup>. 这样, 再将菲涅尔公式中的各个  $\cos \theta$  理解为  $\frac{k_x}{k}$ , 我们得到:

$$r_s = e^{-2i\delta_s}, \quad t_s = \frac{2n \cos \theta}{\sqrt{n^2 - n'^2}} e^{-i\delta_s}, \quad \delta_s = \arctan \frac{\sqrt{n^2 \sin^2 \theta - n'^2}}{n \cos \theta}$$

$$r_p = e^{-2i\delta_p}, \quad t_p = \frac{2nn'}{\sqrt{(n^2 - n'^2)(n^2 \tan^2 \theta - n'^2)}} e^{-i\delta_p}, \quad \delta_p = \arctan \frac{n \sqrt{n^2 \sin^2 \theta - n'^2}}{n'^2 \cos \theta}$$

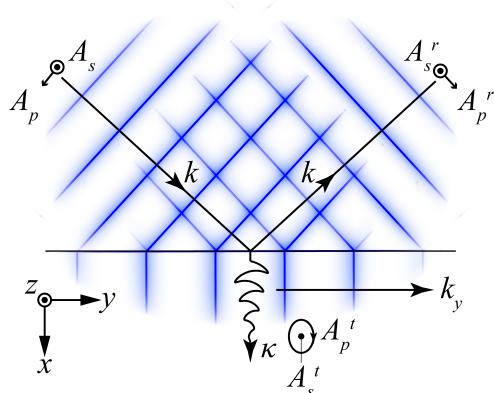


图 15.7: 全内反射与隐失波

可见, 此时最显著的特点除了反射振幅系数的模为一, 就是反射波将会与入射波拉开一个非  $0$  非  $\pi$  的相位差  $2\delta_s$ , 它随着从临界全内反射到掠入射从  $0$  不断增加至  $\pi$ . 而折射波相位则恰好介于两者之间, 其振幅在临界全内反射时有极大值, 此后随着入射角增加而不断减小至零.

更细致的分析可以发现折射光场的  $p$ -光模式实际上是  $x-y$  平面上的顺时针椭圆偏振光.  $s$ -光模式则显然还沿垂直平面的  $z$  方向. 两者都要在  $x$  方向上做衰减, 指数衰减可以表示为:

$$A(x) = A(0)e^{-\frac{x}{H}}$$

其中  $H = \frac{1}{\kappa} = \frac{\lambda_0}{2\pi\sqrt{n^2 \sin^2 \theta - n'^2}}$  称为衰减深度(attenuation depth) 或趋肤深度(skin depth),  $\lambda_0$  是真空中的该色光波长. 由表达式可见其一般  $H$  就光的波长的数量级. 正是因为这个原因这种折射波被称为隐失波(evanescent field). 它一定会产生但是由于存在厚度很薄所以难以被探测到. 如果在隐失波未衰减到零处又重新进入下一层折射角正常的介质, 那么将会有有限光强的光出射出来. 这种现象叫做隐失波耦合(evanescent wave coupling). 在光量子论的角度看这实际上是光子的量子隧穿(quantum tunnelling) 行为.

<sup>7</sup>这意味着下方的场才是需要关心的而界面上的场仅仅是弱地可以忽略的边缘效应.

值得指出,如果入射光并不是无限宽的平面波,而是可以确定其入射点位置的窄光束,而且也不是时刻连续的光,而是一个光脉冲,那么它的行为体现出更加有意思的结果.首先是著名的古斯-汉森效应(Goos-Hänchen effect),光脉冲似乎能进入下层介质,并在其中传播一段时间,随后在另一处离开下层介质出射.如图所示,这造成的前后光束中心位置的横向移位称为古斯-汉森位移(Goos-Hänschen shift).这个位移的计算需要把入射光束看成有一定角度展宽的单色平面波的叠加,而在入射点处恰好各个平面波相干相长,而光束出射点也是各个出射平面波相干相长的位置,不难说明:

$$\Delta_{s,p} = 2 \frac{d\delta_{s,p}}{dk_y} = \frac{\lambda \cos^2 \delta_{s,p}}{\pi \cos \theta} \cdot \frac{d \tan \delta_{s,p}}{d\theta}$$

计算可得:

$$\Delta_s = \frac{2H \tan \theta}{n} ; \quad \Delta_p = \frac{\Delta_s}{(1 + \frac{n^2}{n'^2}) \sin^2 \theta - 1}$$

另一种效应是英伯特-费多罗夫效应(Imbert-Fedorov effect).如果入射光是两种圆偏振光,那么可以证明隐失波场将具有垂直于纸面方向的能量流,如果入射光是窄光束,那么反射光束将在垂直于纸面的方向发生位移,这个结果与光的角动量守恒相协调.在这里不加论述,详情参见光学教材.

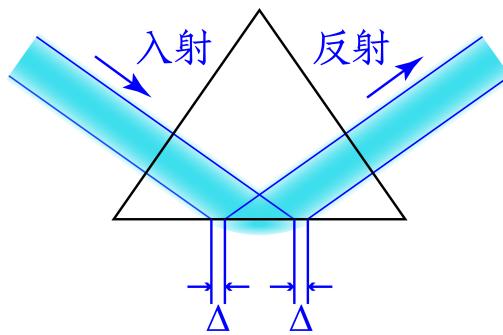


图 15.8: Goos-Hänchen Shift

## 15.2 光线方程

### 15.2.1 光线方程与折射定律

在给定折射率  $n$  随着空间位置连续变化的情况下,光将如何进行传播呢?把光的传播简化为光线,光线垂直于等相位面,那么光线方程(ray equation)给出了这个问题的答案.这个方程写作:

$$\frac{d}{ds}(n\tau) = \nabla n$$

其中  $ds$  表示光线经过的路程.我们来考虑证明这个式子.首先我们先观察上式在法向与切向的两个分量式:

$$(\frac{dn}{dl})_\tau \tau = (\nabla n)_\tau \tau$$

$$n \frac{d\tau}{ds} = n \frac{d\theta}{ds} \mathbf{n} = (\nabla n)_n \mathbf{n}$$

那么第一个式子是一个重言式(恒等式),而第二个式子则给出了由光线传播方向(给出了切向量  $\tau$  与法向量  $\mathbf{n}$ )和局域折射率分布(给出了  $\nabla n$ )决定的光线弯曲程度(有曲率  $k = \frac{d\theta}{ds}$  描述).是名副其实的能够决定光线传播规律的光线方程.事实上,如图,如果把在该点局域处折射率分布理解为沿  $\nabla n$  的方向进行分层,把光线传播方向与局域统一的  $\nabla n$  方向的夹角记为  $\varphi$ ,则当光线走了  $ds$  时折射率变化为:

$$dn = \nabla n \cdot ds \tau = |\nabla n| ds \cos \varphi$$

由折射定律:

$$d(n \sin \varphi) = 0 \Rightarrow d\varphi = -\frac{\sin \varphi dn}{n \cos \varphi} = -\frac{|\nabla n|}{n} ds \sin \varphi$$

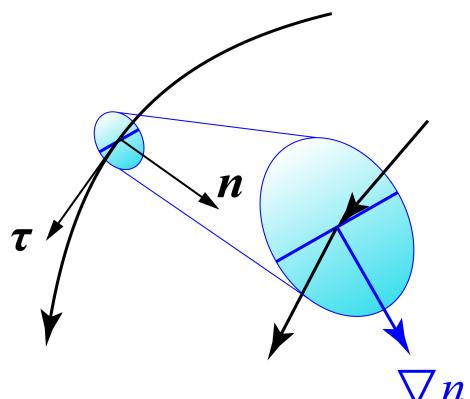


图 15.9: 光线方程推导

这便是:

$$n \frac{d\theta}{ds} = n \frac{|d\varphi|}{ds} = (\nabla n)_n$$

从而给出了光线方程的表达式.

具体的,如果研究在二维平面上,折射率  $n$  仅仅与  $y$  有关而与  $x$  无关的问题,如稳定大气中光线的传播问题,此时原光线方程化为:

$$\frac{d}{ds}(n \sin \theta \mathbf{e}_x + n \cos \theta \mathbf{e}_y) = \frac{dn}{dy} \mathbf{e}_y$$

注意这时候我们一般取  $\theta$  为光线传播方向与界面法线 ( $y$  方向) 间的夹角.上式一方面意味着

$$n \sin \theta = \text{Const.}$$

这可以看成是两层界面上折射定律的推广,它给出了每一个  $y$  处(因而知道  $n$ ) 的光线传播方向,从而本质上余下工作只需要解出具体轨迹即可.只需解方程:

$$\sin \theta = \frac{1}{\sqrt{1+y'^2}} = \frac{(n_0 \sin \theta_0)}{n(y)}$$

其中  $(n_0 \sin \theta_0)$  是根据初始条件确定的不变量.

### 15.2.2 光力类比

牛顿提倡光的粒子说,他的粒子光学理论中光粒子<sup>8</sup>运动状态由位置与速度描述,而运动由轨迹方程与运动方程描述.的确,由光线方程可知,这是一个定义良好的动力学方程,它类似于势场中的牛顿第二定律:

$$\frac{d}{dt}(mv\tau) = -\nabla V$$

由粒子的初位置与初速度便可决定论地的给出之后的运动.这一种类似性为我们进行光力类比提供了良好的依据.然而要注意仍然有一个不同之处:光的速度大小不可控.在一个典型的光线求解的过程中,我们知道初始时刻的光的传播方向即可,而传播速度无需给出,事实上它也不能被改变,因为光速是常数,换一个角度理解,改变速度就是改变动量与能量,那么实际上光子的动量能量改变实际上反应在了光的颜色(频率)上.这是与相对论相吻合的一个结论.

那么如何进行光力类比呢?还是从以上两个公式的相似性出发,我们想象有一个质量为  $m=1$  的粒子在光的真实传播轨迹上运动,规定其速度为  $v=n$ <sup>9</sup>,那么其受力应该为:

$$\mathbf{F} = \frac{d}{dt}(mv\tau) = \frac{d}{ds/v}(n\tau) = v\nabla n = n\nabla n = -\nabla\left(-\frac{n^2}{2}\right)$$

也就是说如果认为粒子是在  $V=-\frac{n^2}{2}$  的势场中运动的,那么其模仿光线的轨迹与用该处折射率来定标自己的运动速度的做法又恰巧是符合动力学规律的.恰巧,这个运动的守恒能量:

$$E = T + V = \frac{v^2}{2} - \frac{n^2}{2} = 0$$

光粒子的运动只能是能量为零的,如果忽略光的色散,那么不同频率的光在同一个折射率场中将具有完全相同的传播轨迹,实际上没有光粒子运动的“快慢”的说法.

我们总结光力类比法求光线传播问题的大意:

1. 明确折射率场分布  $n(\mathbf{r})$ ,明确光线初始入射位置与方向  $\tau_0(\mathbf{r}_0)$ .
2. 光力类比:设置势场  $V(\mathbf{r}) = -\frac{n^2}{2}$ ,设置粒子,质量为 1,初始位置  $\mathbf{r}_0$ ,速度  $\mathbf{v}_0 = n\tau_0$ .

<sup>8</sup>与光子概念有本质区别

<sup>9</sup>注意这不是实际光传播的相速度!  $v \propto n$  而  $v_g = c/n \propto 1/n$

3. 求解粒子的运动, 得出运动轨迹.

4. 这个轨迹也就是光线的轨迹, 再由已知的轨迹求解其他问题 (如传播时间等).

这种方法往往是方便的, 它会使得很多我们在力学问题中有明确结论的结果可以直观地呈示出来. 例如在之前的折射率随  $y$  改变的情形, 它对应着水平方向平移对称的势场, 从而有  $x$  方向的动量守恒, 这就是:

$$n \sin \theta = \text{Const.}$$

同样的道理, 如果光线在一个球对称的折射率场  $n(r)$  中运动时, 其轨迹恰好对应有心力场中粒子运动的轨迹. 例如取:

$$V = -\frac{k}{r} + C \Rightarrow n = n_0 \sqrt{\frac{2a}{r} - 1}$$

那么这个粒子的运动是在环域  $0 < r < 2a$  上的椭圆, 而且计算可以表明改变初始条件, 其半长轴固定为  $a$ , 保证其轨迹范围不会超过这个环域.

而此时应该有角动量守恒, 用光线方向与径向夹角  $\theta$  表示则为:

$$nr \sin \theta = L = \text{Const.}$$

### 15.3 费马原理

类似于力学变分原理, 光线方程的存在也意味着光学的变分原理, 这个原理即为著名的费马原理:

$$\delta \int_A^B n ds = 0, \text{ if } \delta \mathbf{r}_A = 0 \text{ and } \delta \mathbf{r}_B = 0$$

如何理解这个原理呢, 首先这个折射率对路程的积分称为光程(optical path), 在光力类比下它实际上就是满足能量守恒条件下的相积分:

$$S = \int_{A \xrightarrow{\Gamma} B} n ds = \int_{A \xrightarrow{\Gamma} B} \mathbf{p} \cdot d\mathbf{r}$$

而从  $A$  点到  $B$  点的光传播可以由某种参数方程描述:

$$\Gamma : \quad \mathbf{r} = \mathbf{R}(t), \quad \mathbf{R}(t_A) = \mathbf{r}_A, \quad \mathbf{R}(t_B) = \mathbf{r}_B$$

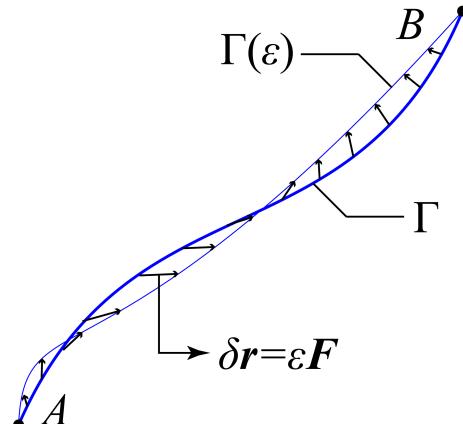


图 15.10: 光程变分

而变分的意思是我们另取一种可能的轨迹, 比如:

$$\Gamma(\epsilon) : \quad (\mathbf{r} + \delta\mathbf{r})(t) = \mathbf{R}(t) + \epsilon\mathbf{F}(t)$$

那么费马原理要求在轨迹端点不变的条件下, 新的光程关于所有可能的微扰方向  $\mathbf{F}(t)$ , 关于其微扰强度的一阶导数都要等于零:

$$S(\epsilon) = \int_{A \xrightarrow{\Gamma(\epsilon)} B} n ds \quad ; \quad \left. \frac{dS(\epsilon)}{d\epsilon} \right|_{\epsilon=0} = 0$$

这个原理是等价于光线方程的, 证明这一点需要用到较为专门的矢量分析技巧, 在此略去. 用文字一般可以表述为:

两点之间光线的实际轨迹, 是光程取平稳值的路径.

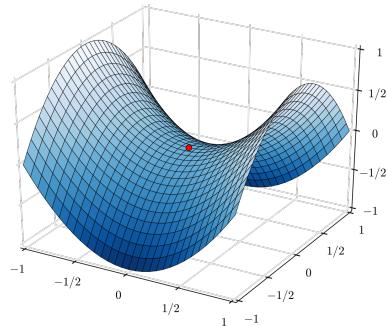


图 15.11: 二维鞍点处势取平稳值

其中的“平稳值”的说法,既包含了极大,也包含了极小,还包含了一些复合的情况,比如如图,我们考虑一个马鞍面上有一个质点在鞍点处于平衡状态,那么在一个方向上势能最高,另一个方向上势能最低,然而该点能平衡,因为势能取得了“平稳值”.二自由度下已经出现了两个自由度方向微扰后极大极小性质出现差异的情形,而光程问题则是在所有可能的轨迹空间中进行无穷多自由度的微扰的问题,普遍来说符合之前我们定义的一阶变分为零即可以符合费马原理的条件,故冠以“平稳值”的说法.

最后,值得指出以上结论都是在波动理论尚未发展到一定程度时的替代性的粒子理论.在有了波动理论以后以上两个结论都将得到极为简单的解释.我们把变折射率问题中的光的等相位面找到,光线即将垂直于等相位面.而相位场<sup>10</sup>  $\varphi(\mathbf{r})$  与光传播方向,折射率间关系即为:

$$\nabla\varphi = \mathbf{k} = k_0 n \boldsymbol{\tau}$$

该式两边同时对传播方向取方向导数:

$$\frac{d}{ds}(\nabla\varphi) = k_0 \frac{d}{ds}(n\boldsymbol{\tau})$$

而此式左边,在沿传播方向求的导数应该理解为与下一个梯度算符独立的,故可以交换顺序:

$$\frac{d}{ds}(\nabla\varphi) = \nabla\left(\frac{d\varphi}{ds}\right) = \nabla k = k_0 \nabla n$$

两式联立便得到光线方程:

$$\frac{d}{ds}(n\boldsymbol{\tau}) = \nabla n$$

而费马定理也非常容易理解:光程积分:

$$\int n ds = \int n \boldsymbol{\tau} \cdot d\mathbf{r} = \frac{1}{k_0} \int \nabla\varphi \cdot d\mathbf{r} = \frac{\Delta\varphi}{k_0}$$

所以在光程微扰下只要起点与终点固定,光程就不会有改变.

<sup>10</sup> 在粒子论里称为程函(eikonal)

# 第十六章 光学成像

## 16.1 傍轴光成像

### 16.1.1 物与像

成像不是一个平凡的过程. 延续上一章我们对变折射率问题的讨论, 现在我们引入所谓物(object)的概念, 它指的是一个点光源  $S$ , 它能够向各种不同的方向发射光线, 每一个既定的方向则可以根据光线方程求解之后的光线传播. 这样它所发出的光场构成了物的光场. 而传播过程中到达的任何一点  $P$ , 满足  $S$  与  $P$  间的真实光线有且只有一条. 它恰好由费马原理确定. 但传播过程中可能发生一种奇异的现象, 就是在空间中另一点  $S'$  处成了像(image). 从光线角度看, 便是:

一定范围内, 物发出的每一条光线都经过像.

而根据费马原理, 真实光线旁边的微扰光线光程对此光线的一阶变分为零, 但此时有一大组真实光线, 彼此之间可以连续的过渡, 我们必然得出:

物与像之间的所有真实光线, 其光程都是相等的.

以上就是物像之间的等光程原理(the principle of equal path length). 它实际上也是作为能否成像, 成像在何处的判据. 这一个结论也可以依靠干涉理论, 或者更普遍的, 光场概率幅的路径积分理论来理解: 如果两点间光程沿某种路径不是取得极值, 那么这种情形将会由于存在很多其他类似的路径, 使得最后在终点处贡献的概率幅相干相消. 仅仅是在光程取得极值的路径将对最后终点处的概率幅有主要贡献. 但在物像点间如果存在大量路径其光程完全不变, 最后像点的概率幅将会非常地大, 这也就是成像的过程了.

以上描述的实际上只是几何光学中的实物与实像的概念. 在一般的光学系统中, 光路都是傍轴的. 即满足傍轴条件(paraxial condition): 有一条中心光线作为主光轴(optical axis) 参考. 光具组(optical system) 把光路分为物方(object space) 与像方(image space). 两个空间一般都是均匀折射率空间, 故一般考虑光锥(pencil)型光束的传播(就是球面波). 注意两个空间有可能重合(反射情况). 这样的光束偏离主光轴的角度满足:

$$\theta \ll 1$$

或者表述为

$$\frac{z}{l} \ll 1$$

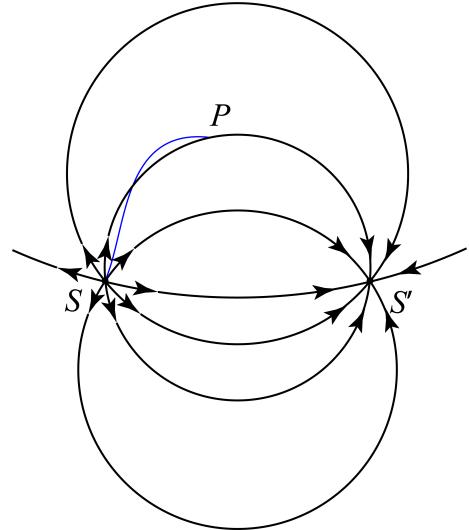


图 16.1: 物与像

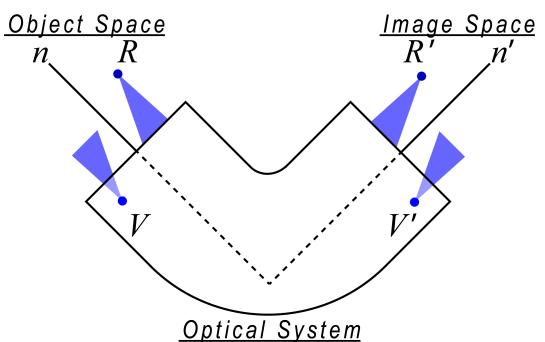


图 16.2: 实与虚

其中  $z$  是光路的特征偏离主光轴的距离. 而  $l$  是光路沿主光轴传播的特征长度. 如果进入光具组时光束呈现发散(diverging)态, 则说明进入光具组前已有物, 尽管这一个物有可能也不在真实空间中(如前一次经过光学组后成虚像作为这次的实物), 也把这样的物称为实物(real object). 它一定位于物方. 而相反的情况是进入光具组时为会聚(converging)光的情况, 那么物就不可能在物方找到了, 它位于光具入射面的另一侧(不一定是像方), 这样的物称为虚物(virtual object). 同理, 对于像, 出射光会聚的像位于像方, 是实像(real image), 而出射光发散, 像位于像方的另一侧, 则是虚像(virtual image).

### 16.1.2 球对称成像系统与符号法则

顾名思义, 球对称成像系统指的是主光轴在一条直线上而整个光具组具有对主光轴的旋转对称性. 当然实际光锥光路不一定要求关于主光轴对称. 两种典型的基础光具是折射球面与反射球面. 如下图所示:

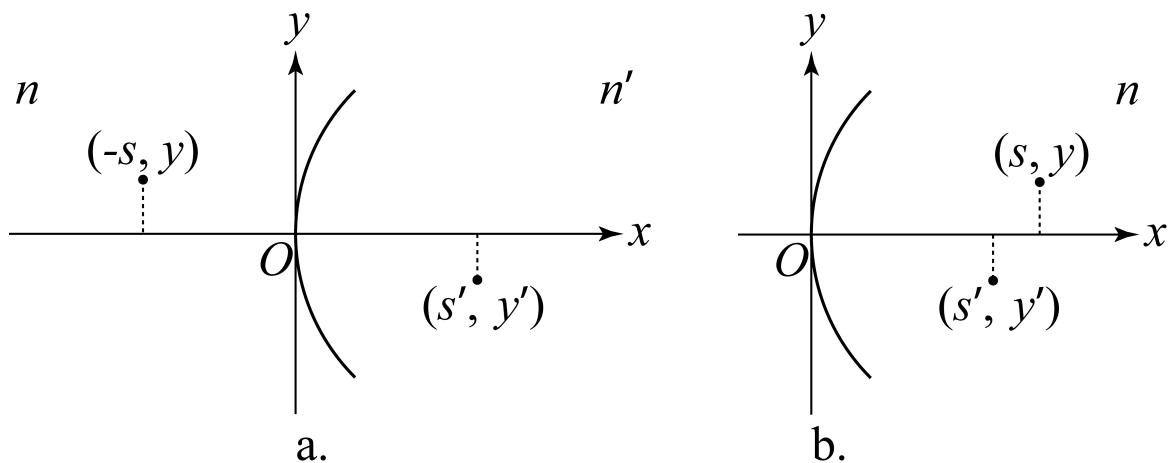


图 16.3: 球面折射与球面反射

在球面折射情况16.3下, 我们先考虑向像方凹的, 曲率半径为  $R$  的近似球面的傍轴折射成像(作左图). 各量的符号设定如图,  $s$  为物距(object distance),  $s'$  为像距(image distance),  $y, y'$  则为物, 像的高(height). 由物像的等光程性, 应有:

$$f(X, Y) = n\sqrt{(X + s)^2 + (Y - y)^2} + n'\sqrt{(X - s')^2 + (Y - y')^2} = \text{Const.}$$

由于傍轴条件, 上式中参量  $Y$  是理解为小量变量即可, 而  $X$  近似为:

$$X = \frac{Y^2}{2R}$$

代入, 小量近似, 表示为  $Y$  的二次多项式:

$$\begin{aligned} f(X, Y) &= n\sqrt{s^2 + y^2} + n'\sqrt{s'^2 + y'^2} \\ &\quad - \left( \frac{ny}{\sqrt{s^2 + y^2}} + \frac{n'y'}{\sqrt{s'^2 + y'^2}} \right) Y \\ &\quad - \left[ \frac{n[(1 + s/R)(s^2 + y^2) - y^2]}{2(s^2 + y^2)^{3/2}} + \frac{n'[(1 - s'/R)(s'^2 + y'^2) - y'^2]}{2(s'^2 + y'^2)^{3/2}} \right] Y^2 \end{aligned}$$

从而一次项与二次项系数都必须为零. 由傍轴条件,  $s, s'$  是  $u, v$  的小量. 从而整理得:

$$\begin{aligned} \beta &= \frac{y'}{y} = -\frac{ns'}{n's} \\ \frac{n}{s} + \frac{n'}{s'} &= \frac{n' - n}{R} \end{aligned}$$

上式给出了成像的横向放大率(transverse magnification)与物像距关系(object-image distance relation), 抽象地看, 在傍轴加二阶近似意义下物方空间每一点都能成像. 而给定任意物距处的像距与横向放大率, 就决定了一个光学系统傍轴成像的所有特性. 以上成像的两个公式的特点是: 物像距都从球面的顶点开始计算, 这个点称为光心(optical centre). 这两个公式称为高斯型公式(Gaussian formula). 一般把第二个物像距公式右端的结果定义为光焦度(focusing power):

$$\Phi = \frac{n' - n}{R}$$

其单位是屈光度(diopter), 中文简称“度”.  $1D = 1\text{m}^{-1}$ .

当物距和像距取特殊值时会出现另一个趋于正负无穷, 而对应的物像高也趋于正负无穷的情形, 此时意味着入射, 出射光成为了平行光(collimated light). 这些特殊的物像距称为焦距(focal length), 而对应的主光轴上的点称为焦点(focus). 物方焦距与像方焦距就分别为:

$$f = \frac{n}{\Phi} = \frac{nR}{n' - n} ; f' = \frac{n'}{\Phi} = \frac{n'R}{n' - n}$$

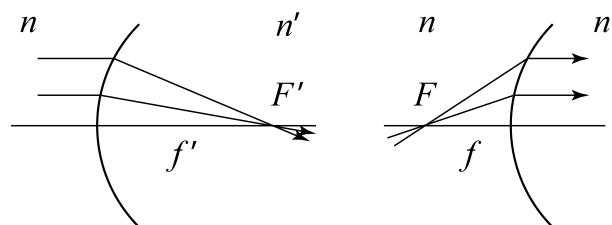


图 16.4: 焦点

自然地有:

$$f : f' = n : n'$$

通过引入两个焦距我们可以把成像公式写作:

$$\frac{f}{s} + \frac{f'}{s'} = 1 ; \beta = -\frac{f s'}{f' s}$$

实际上,  $n, n'$  都不是通过这个折射成像系统可以测量的量, 而两个长度量纲的  $f, f'$  代表了这个系统成像的所有特征.

为了更直观地反应物像距指尖关系的特性, 一般把顺光传播方向, 物到物方焦点的距离记为另一种物距  $x = s - f$ , 而像方焦点到像的距离记为另一种像距  $x' = s' - f'$ . 那么代入高斯型公式化简可得:

$$x x' = f f' \\ \beta = -\frac{f}{x} = -\frac{x'}{f'}$$

以上两个公式更为精简, 称为牛顿型公式(Newtonian formula). 图像也更加清晰: 物像距恰好是一个反比关系. 而横向放大率取决于物离物方焦点有多近, 也等效于像离像方焦点有多远. 运用起来往往更加方便. 仔细观察还会发现, 仅需  $ff'$  作为整体的大小实际上已经确定了物像距的关系. 但如果要求的是横向放大率,  $f, f'$  各自取多少就是必要的了. 也就是说不仅需要知道两个焦点的位置, 还得已知光心离各自有多远.

对于16.3右图的球面反射镜, 只需要把两个焦距定为

$$f = f' = \frac{R}{2}$$

那么高斯型公式与牛顿型公式都以完全一样的形式适用, 只需注意此时的主光轴正方向在反射时发生了反转, 物像距的定义会产生相应的改变, 参照图理解即可.

当我们考虑不同的球面凹凸性, 不同的折射率分布, 不同的物像实虚性时显然不希望每次都重新进行推导, 而是希望我们之前推出的结论具有普适性. 那么实际上之前我们涉及的每个量都可以有正负(折射率除外). 约定不同量的正负性的方法称为符号法则(sign convention). 本书介绍一下符号法则:

1. 光具球面曲率半径: 出射方看, 凹正凸负. 出射方也就是单次成像的像方, 如果曲率中心位于这一侧, 那么曲率半径计为正. 如果在另一侧, 计为负.

2. 物像距, 焦距: 实正虚负. 如果是物点发出的发散光沿主光轴正向传播到物距的零点处则为实物, 物距计正 (实物). 如果是会聚光传播到物距零点处还未到达物点, 则物距计负 (虚物). 同理, 像距零点处出射会聚光可以沿主光轴传播形成像点, 像距计正 (实像). 若像距零点处出射发散光, 像点可以认为应在之前形成, 像距计负 (虚像). 物方焦距是使像距趋于无穷的特殊物距. 像方焦距是使物距趋于无穷的特殊像距.
3. 物像高: 上正下负. 不管任何情形, 都规定主光轴的一侧高度大于零, 一侧小于零. 也就是把与这个方向垂直的轴上坐标定义为高度. 如果把主光轴理解为水平方向, 一般上半空间物, 像高为正, 下半空间物像高为负.
4. 光线的倾角: 增正减负. 有时候我们不仅关心物像点间的对应关系, 也关心一根光线与另一根光线的对应关系. 而实际上相互对应的光线不仅是物方入射光线与像方出射光线的关系, 由于每一点某能成像, 两条线上的点也是一一对应的物像点. 此时光线的角度被定义为沿传播方向单位长度所增加的物或像高. 如果光线从下往上传播, 角度为正, 从上往下则角度为负.

我们之前推出来的, 针对于球面折射, 反射的公式自然是在以上符号法则下普适的. 同样的道理也适用于之后所有情况下提出的公式. 而且不失一般性地, 高斯型公式与牛顿型公式其实是对任意球对称成像系统傍轴适用的. 推广到非傍轴情形便是下一节讨论的理想成像系统.

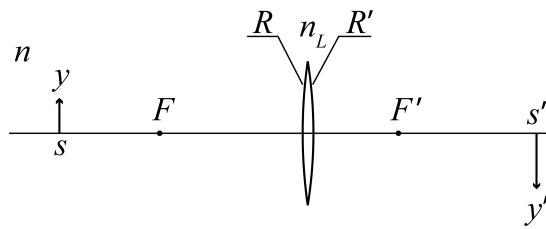


图 16.5: 薄透镜

我们接着讨论**薄透镜**(thin lens) 模型. 这是一个两次成像的过程, 第一次成像的像方零像距点, 第二次成像的物方零物距点恰好重合在透镜的光心位置. 从而有:

$$\frac{n}{s} + \frac{n_L}{s'_1} = \frac{n_L - n}{R} ; \quad \beta_1 = -\frac{ns'_1}{n_L s}$$

$$\frac{n_L}{s_2} + \frac{n'}{s'} = \frac{n' - n_L}{-R'} ; \quad \beta_2 = -\frac{n_L s'}{n' s_2}$$

注意在第二个公式中  $R'$  是在右表面凸是取正的曲率半径 (以双凸透镜为正), 它与我们既定的符号法则相反, 故在照抄公式时在分母加以符号以示区别. 再有上讨论的两个物像距参考关系:

$$s'_1 + s_2 = 0$$

消去中间的物像距, 得到:

$$\frac{n}{s} + \frac{n'}{s'} = \frac{n_L - n}{R} + \frac{n_L - n'}{R'} ; \quad \beta = -\frac{ns'}{n's}$$

这说明把两次成像的密接的结果是光焦度的直接相加:

$$\Phi = \Phi_1 + \Phi_2$$

而除此以外余下公式都没有变化:

$$f = \frac{n}{\Phi} ; \quad f' = \frac{n'}{\Phi}$$

$$\frac{f}{s} + \frac{f'}{s'} = 1 ; \quad \beta = -\frac{fs'}{f's}$$

这几个规律也适用于密接的透镜组. 任何发散和会聚光的能力直接用光焦度评价, 密接时这些规律都直接叠加.

### 16.1.3 光具组成像

如图,如果考虑**厚透镜**(thick lens) 模型,或者更普遍地讨论两个薄透镜相隔一定距离放置的情况.那么一般把前透镜像方焦点到后透镜物方焦点的以传播方向为正的距离即为**光学距离**(optical distance).即  $\Delta = \overline{F'_1 F_2}$ ,两透镜间距离:

$$d = f'_1 + f_2 + \Delta$$

上图的  $\Delta$  实际上取负,事实上当  $\Delta = -f'_1 - f_2$  时变为密接的情形,而  $\Delta$  增大也代表着  $d$  增大.我们写出此时的成像物像距公式:

$$\begin{aligned} \frac{n}{s} + \frac{n''}{s'_1} &= \Phi_1 \\ \frac{n''}{s_2} + \frac{n'}{s'} &= \Phi_2 \\ s'_1 + s_2 = d &= \frac{n''}{\Phi_1} + \frac{n'}{\Phi_2} + \Delta \end{aligned}$$

得到:

$$(s - f_1 - \frac{f_1 f'_1}{\Delta})(s' - f'_2 - \frac{f_2 f'_2}{\Delta}) = \frac{f_1 f'_1 f_2 f'_2}{\Delta^2}$$

上式即成像公式的牛顿形式.而由牛顿型的放大率公式:

$$\beta = \left( -\frac{f_1}{s - f_1} \right) \left( -\frac{s' - f'_2}{f'_2} \right)$$

代入上式可得:

$$\beta = -\frac{-\frac{f_1 f_2}{\Delta}}{s - f_1 - \frac{f_1 f'_1}{\Delta}} = -\frac{s' - f'_2 - \frac{f_2 f'_2}{\Delta}}{-\frac{f'_1 f'_2}{\Delta}}$$

从中可以发现,这个成像公式恰恰等价于一个物像距的零点不位于两个透镜光心.而两侧的焦点也不与原来的焦点重合的成像系统,仍然有其焦距为:

$$f = \overline{FH} = -\frac{f_1 f_2}{\Delta}, \quad f' = \overline{H'F'} = -\frac{f'_1 f'_2}{\Delta}; \quad f : f' = n : n'$$

而各个特征点位移为:

$$\begin{aligned} \delta_F &= \overline{F_1 F} = -\frac{f_1 f'_1}{\Delta}; \quad \delta'_F = \overline{F' F_2} = -\frac{f_2 f'_2}{\Delta} \\ \delta_H &= \overline{O_1 H} = -\frac{f_1 d}{\Delta}; \quad \delta'_H = \overline{H' O_2} = -\frac{f'_2 d}{\Delta} \end{aligned}$$

其中  $H$  被称作等效光学系统的物方主点,  $H'$  被称作等效光学系统的像方主点.以上公式的理解中如果  $\Delta < 0$ ,那么各个量就都是正的.图画的正是这个情况.此时的成像规律,可以认为是在  $H$  处具有光焦度为  $\Phi = -\frac{n'' \Delta}{f'_1 f'_2}$  的透镜,再把其直接成像的结果进行一个平移,平移距离即为两个主点间距  $\overline{HH'}$ .正是因为这个原因,考虑多次成像系统的复合时可以再把两个光具等效的光学系统再与第三个光具按照以上公式再进行等效,直到化为最基础的成像公式的情形.

## 16.2 理想成像系统

把傍轴成像系统的特性进行数学上的抽象与大胆的推广,我们得到了理想光具组成像系统.它是一个全空间中点到全空间中点的可逆映射:

$$\mathcal{J}: S \rightarrow S', \quad S \in \mathbb{R}^3 \cup \{\infty\}$$

在这个映射中无穷远点也是有意义的.一般正负无穷远处的点其实没有区别(发出平行光),不同的无穷远点根据其发出的光与主光轴的夹角定义.而理想光具组需要有以下特征:

1. 主光轴必须被映射到主光轴.
2. 映射关于主光轴对称. 可以认为是一个平面上的映射绕主光轴旋转而生成的.
3. 保直线, 平面的形状. 直线点集与平面点集映射到直线与平面.

仅这三个条件, 足以把映射的性质确定到仅含极少参数, 也就是我们上一节介绍的成像系统. 这是因为可以根据以上定义证明以下性质 (作为思考题请读者自己思考):

1. 垂直于主光轴的平面一定被映射到垂直于主光轴的平面.
2. 在垂直于光轴的平面之间的映射是简单的放缩变换 (横向放大率为常数).
3. 不同的物距的横向放大率, 要么始终保持为常数, 要么连续, 单次地取遍任意正负值. 前者称为**望远系统**(afocal system).

所以我们把这个映射的性质总结为在一个过主光轴的平面内的问题, 这个平面称为**子午面**(meridinal plane), 只需要物像距公式和放大率公式两个成像公式. 而定义若干的辅助点, 辅助面可以帮我们有效的研究非望远系统的性质. 这些点与面被称作**基点**(cardinal point) 与**基面**(cardinal plane). 除了望远系统, 所有理想光具组成像都有两对关键基点, 它们都是主光轴上的特殊点, 而有基面即过基点的垂直于主光轴的平面, 代表特殊的物像距处物像高的所有可能性. 它们是:

**主点 (principal points):** 横向放大率为 1 的那对共轭点即为物方主点与像方主点.

**焦点 (focal points):** 横向放大率发散 ( $\infty$ ) 时, 物方为物方焦点, 像方为无穷远点. 横向放大率为 0 时, 物方为无穷远, 像方则为像方焦点.

有了以上定义以后, 便可以用作图法确定成像公式.

### 16.2.1 作图法

作图法的基础是以下两个简单结论:

- 凡入射到物方主平面上一点的光线必然在像方主平面上等高处出射.
- 平行光入射, 出射光聚焦于像方焦平面上某点. 物方焦平面上一点入射, 则出射平行光. 更特别地, 经过物方焦点的光线出射光平行于主光轴, 平行于主光轴的光入射后出射光经过像方焦点.

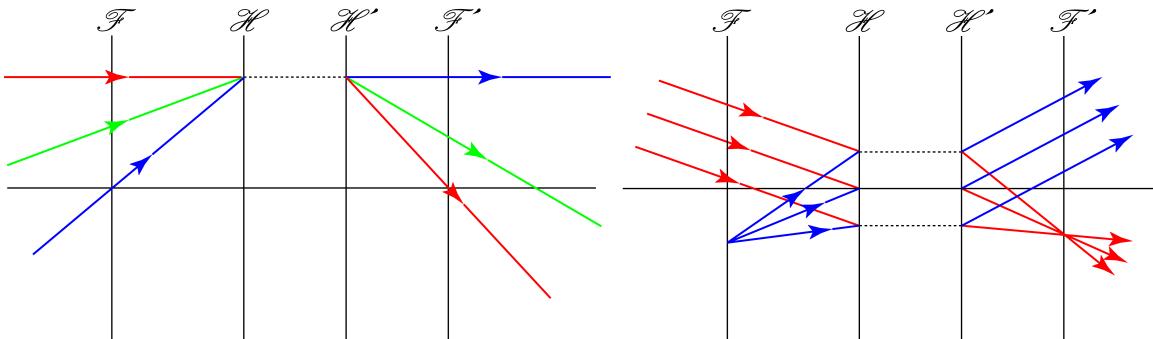


图 16.7: 特殊光线

通过以上特殊光线我们可以确定任意一条入射光线对应的出射光线, 也能进一步明确到每一个物对应的像的位置. 如下图左所示. 给出一条入射光线  $AB$ . 如果它入射到物方主面上的  $B$  点. 那么必然从像方主面上的等高  $C$  点出射. 于是做出两条辅助光线: 一是从入射光线经过的物方焦面上的点作平行与主光轴

的光线(红色),那么我们很容易找到它的出射光线.那么原光线与此特殊光线必然出射平行光而找到出射 $CD$ 的方向.二是也可以做出过物方焦点且与 $AB$ 平行的辅助入射光线.也可以找到它的出射光线从而确定出 $CD$ 来.而如果要找物 $S$ 的像 $S'$ 的位置,也是利用特殊的光线确定的,如下图右.通过几何上三角形1与2,3与4的相似关系:

$$\frac{y}{-y'} = \frac{x}{f} ; \quad \frac{-y'}{y} = \frac{x'}{f'}$$

整理即为牛顿型的成像公式:

$$xx' = ff' ; \quad \beta = -\frac{f}{x} = -\frac{x'}{f'}$$

再利用 $s = f + x, s' = f' + x'$ 可以确定从主面出发计算物像距的对应高斯型成像公式:

$$\frac{f}{s} + \frac{f'}{s'} = 1 ; \quad \beta = -\frac{fs'}{f's}$$

- 凡入射到物方主平面上一点的光线必然在像方主平面上等高处出射.
- 平行光入射,出射光聚焦于像方焦平面上某点.物方焦平面上一点入射,则出射平行光.更特别地,经过物方焦点的光线出射光平行于主光轴,平行于主光轴的光入射后出射光经过像方焦点.

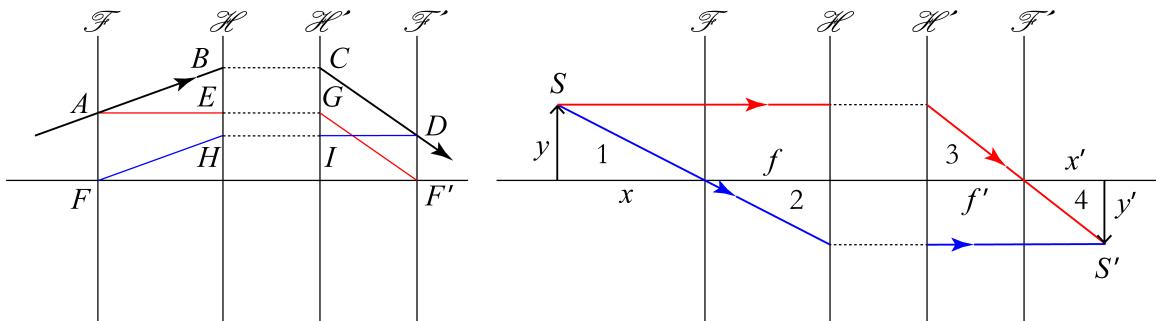


图 16.8: 共轭线与共轭点

### 16.2.2 基点基面性质

在实际傍轴成像过程中还有一个我们比较关系的量.就是在一定物距与对应的像距处的成像的角度放大率(angular magnification).它被定义为物点傍轴发出一个顶角为 $u$ 小角的光锥时<sup>1</sup>,成像处对应的光锥顶角 $u'$ 的放大率.在傍轴的小角近似下,也可以定义为过零物像高处的共轭点的共轭光线的夹角比.后一个定义能产生正负号,我们采用后一种定义(右图中红色的角).

$$B = \frac{u'}{u}$$

而如果考虑的是理想光具组情形,我们可以将上式写为更严格的形式:

$$B = \frac{\tan u'}{\tan u}$$

从图中可以发现,这样定义角度放大率后它与物像距的关系从图中可以写出:

$$s \tan u = s' \tan(-u') \Rightarrow B = -\frac{s'}{s} = \frac{f'}{f} \beta$$

<sup>1</sup>这里考虑平面问题,空间情形可以比较类似地讨论,此时应该考虑立体角.

考虑牛顿公式  $xx' = ff'$ , 除去焦点与焦面, 还有有四种独特的物像距分别具有独特的性质. 它们对应的主光轴上的点与平面都被称作基点与基面. 而各个点性质如下:

**主点 (principal points)** :  $x = -f, x' = -f' \Rightarrow \beta = 1$ . 成等大正立像.

**二倍焦距点** :  $x = f, x' = f' \Rightarrow \beta = -1$ . 成等大倒立像.

**节点 (nodal points)** :  $x = -f', x' = -f \Rightarrow B = 1$ . 角放大率为一.

**物像距等值点** :  $x = f', x' = f \Rightarrow B = -1$ . 角放大率为负一.

其中节点概念在光学测量中有重要的作用. 它的特点是, 入射到物方节点  $N$  光线, 方向不改变地从像方节点  $N'$  出射. 注意到这个角放大率为一的事实也仅仅在物像距为节点时能够取到. 节点也是经常讨论的关键基点

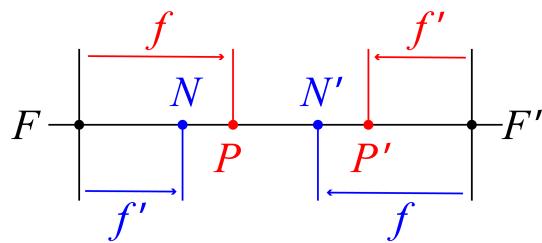


图 16.10: 三对关键基点

我们可以这么理解这三对基点的关系,  $xx' = ff'$  提示我们, 只需要知道  $ff'$  乘积整体的值便可以确定物像距之间的对应关系, 而两者之比  $f : f' = n : n'$ <sup>2</sup> 则告诉我们放大率的信息. 实际上就是分别确定物方主点在焦点后方多远, 像方主点在像方焦点前多远. 而把这两个距离对换, 就变成了对应的节点位置. 换句话说, 如果把两焦点的中点看做对称中心, 则两个主点对这个点中心反演就到了两个节点.

这也就是说, 如果物方像方折射率相等 (这实际上是比较常遇到的情况), 那么也就没有  $f$  与  $f'$  的区别, 那么上面列举的四对基点退化为两对. 节点与主点必然重合, 横向放大率与角度放大率都是 1. 而二倍焦距处的那两个共轭点 (物方  $2F$  点, 像方  $2F'$  点, 成像系统常称  $4F$  系统), 物像互为倒立关系, 横向放大率与角度放大率为  $-1$ .

### 16.2.3 实例与望远系统

球面折射, 即可以傍轴近似为两个主点重合到光心 (与两节点都不重合), 物像方焦距分别为:

$$f = \frac{nR}{n' - n} ; \quad f' = \frac{n'R}{n' - n}$$

如果  $R \rightarrow \infty, \Phi \rightarrow 0$ . 即平面折射, 那么成像系统化为望远系统. 此时物像距为简单线性关系, 横向放大率处处一样:

$$\frac{s'}{s} = -\frac{n'}{n}, \beta = 1$$

横向放大率为 1 是系统沿平面界面平移对称的结果.

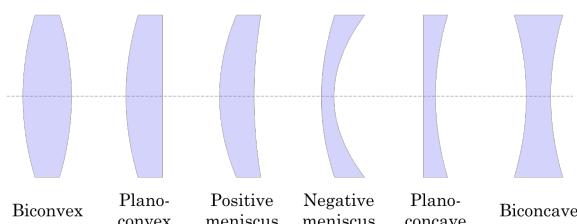


图 16.11: 透镜种类

而对于更复杂而常见的厚透镜系统, 一般我们做如图所示的分类. 即分为**双凸透镜**(biconvex lens), **双凹透镜**(biconcave lens), **平凸透镜**(planoconvex lens), **平凹透镜**(planoconcave lens), **正弯月透镜**(positive meniscus lens) 与**负弯月透镜**(negative meniscus lens). 一般置于折射率为 1 的空气中, 透镜折射率计为  $n$ , 厚度记为  $d$ . 则有以下**磨镜者公式**(lensmaker's equation):

$$\Phi = \frac{1}{f} = (n - 1) \left[ \frac{1}{R_1} + \frac{1}{R_2} - \frac{(n - 1)d}{nR_1 R_2} \right]$$

<sup>2</sup>这个表达式的普遍性参考下一节与下一章.

推导参考之前的证明过程. 注意此处的焦距是要从两个主面开始计算的, 而两个主面的位置一般并不会与左右两折射面与主光轴交点重合. 而是相距:

$$\overline{HH'} = \frac{d(\Delta + f_1 + f'_2)}{\Delta} ; \quad \Delta = d - f'_1 - f_2$$

$$f_1 = \frac{R_1}{n-1} ; \quad f'_1 = \frac{nR_1}{n-1}$$

$$f_2 = \frac{nR_2}{n-1} ; \quad f'_2 = \frac{R_2}{n-1}$$

我们讨论四种特殊的情形. 一是  $R_1 + R_2 = d$  的情形. 此时恰有:

$$\Delta + f_1 + f'_2 = d - (f'_1 - f_1) - (f_2 - f'_2) = d - R_1 - R_2 = 0$$

即两个主点重合. 其实, 两个主点必然重合到左右球面的公共球心. 这是因为如果入射光线过这个点, 那么出射光线方向恰好不改变, 必然也过这个点, 从而这个点是节点, 而对称成像系统(物像方折射率相等, 焦距相等)的主点与节点重合的. 从而主点也在这里. 只需要确定焦距便可以决定的所有特征. 它由下式给出:

$$f = \frac{nR_1R_2}{(n-1)d}$$

第二种情况是  $\Delta = 0$ , 即  $\frac{nR_1}{n-1} + \frac{nR_2}{n-1} = d$ . 此时光焦度变为零, 成像系统成为望远系统. 望远系统由于各种量的发散性, 其物像距关系与放大率都必须重新讨论. 如图16.12, 我们把物距  $x$  仍然计做物到左折射球面左焦点的距离, 而像距  $x'$  仍计做像到右折射球面右焦点的距离, 有:

$$xx'_1 = f_1f'_1 ; \quad \beta_1 = -\frac{f_1}{x}$$

$$x_2x' = f_2f'_2 ; \quad \beta_2 = -\frac{x'}{f'_2}$$

再考虑到两次成像有  $x'_1 = -x_2$ , 得到:

$$\frac{x'}{x} = -\frac{f_2f'_2}{f_1f'_1} = -\frac{R_2^2}{R_1^2}$$

$$\beta = -\frac{f_2}{f'_1} = -\frac{R_2}{R_1}$$

可见物像距间有线性关系, 而横向放大率是一个常数. 我们也定义**纵向放大率**(longitudinal magnification):

$$\Xi = \frac{dx'}{dx} = -\beta^2$$

纵向放大率的大小为横向放大率的平方, 这个结论其实对对称薄透镜也是适用的(请读者证明), 此时物像距关系非线性, 故应采用严格的导数来定义局域放大率.

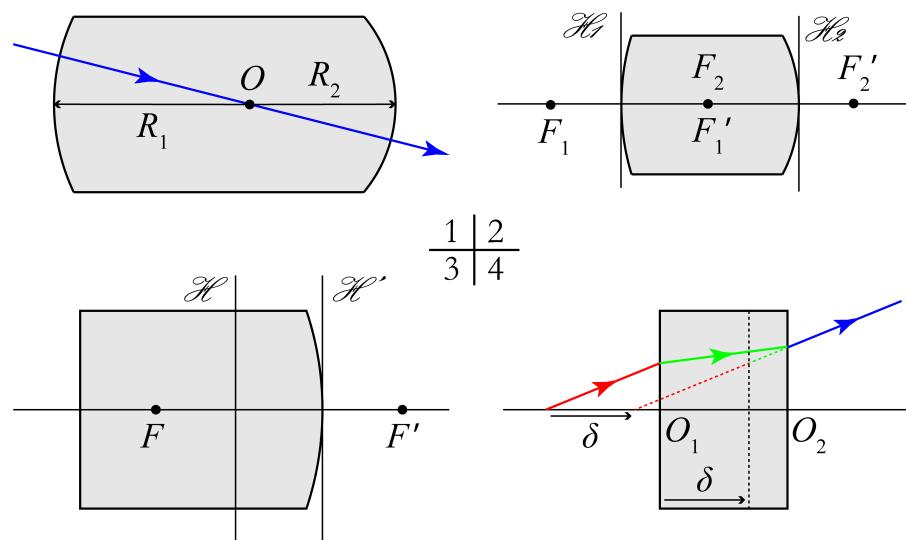


图 16.12: 四种特殊厚透镜

第三种情况是有一个球面退化为平面的情况,此时各公式也因为发散而失效.如  $R_1 \rightarrow \infty$  成为平面.如图16.12,我们可以根据此时成像系统的特征确定等效的各基点.像方焦点显然位置仍然在右折射面像方焦点的位置,这是因为平行光经过平面折射仍然是平行光,就应该会聚在这个焦面上.而像方主面也仍然在右折射面顶点处.这可以从平行于主轴的光线的传播确定出来.从而焦距即为右球面的像方焦距  $f = \frac{R_2}{n-1}$ .物方主面位于像方主面经左平面折射成像的位置,而物方焦点还在其左侧距离  $f$  处,这些都很容易理解,留给读者思考其原因.

最后一种情况是两个球面都退化为平面的情况.这是不仅公式发散需要重新计算,它的结果还是一个望远系统.事实上我们可以考虑傍轴的几何光学,用简单的光线追迹法(ray-tracing method),我们发现如果把玻璃砖的厚度进行压缩到  $d/n$ ,那么所有光线都变成了直线传播.从而物就重合到了像.所以我们发现此时的物像映射实际上就是一个固定距离  $\delta = \frac{n-1}{n}d$  的平移,其横向纵向放大率都是 1.

## 16.3 更多讨论 \*

### 16.3.1 理想成像本质

理想光具组是一种十分强大的工具,但一些普遍的规律仍然悬而未决:为什么根据光线的对应性能把成像确定到如此小的范围.物像方的焦距为什么总是与折射率正相关,同正同负.而各个放大率间的关系又是如何.所以我们提出理想成像的数学本质抽象:射影变换.

首先是数学上有著名的射影几何基本定理(fundamental theorem of projective geometry):

实线性空间上的共线性变换(collineation) 是更高维度内的射影变换(projection).

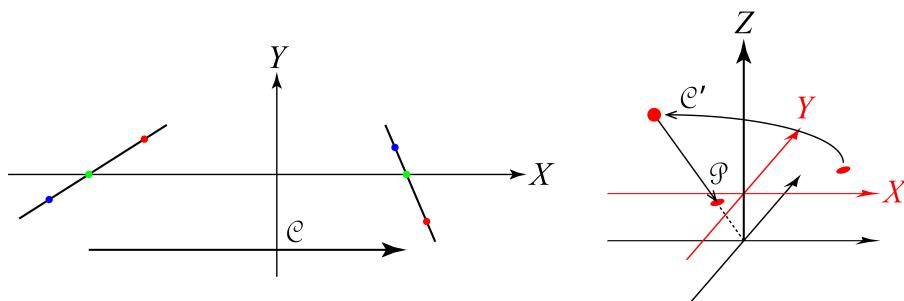


图 16.13: 射影映射

那么成像作为二维平面上的点到点之间的共线性映射:

$$\mathcal{C} : (X, Y)^T \mapsto (X', Y')^T$$

可以加一个第三维度  $Z, (X, Y, Z)$  可以以原点为中心投影到  $Z = 1$  平面, 那么原映射理解为三维内的非退化仿射变换加上投影:

$$\mathcal{P} : \{(X, Y, Z)\} \mapsto (X/Z, Y/Z, 1)$$

$$\mathcal{C}' : \{(X, Y, 1)^T\} \mapsto \{(Z'X', Z'Y', Z')^T\}$$

$$\begin{bmatrix} Z'X' \\ Z'Y' \\ Z' \end{bmatrix} = \begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix}, \quad \begin{vmatrix} a & b & c \\ d & e & f \\ g & h & i \end{vmatrix} \neq 0$$

此即:

$$X' = \frac{aX + bY + c}{gX + hY + i}; \quad Y' = \frac{dX + eY + f}{gX + hY + i}$$

但为了使得  $Y = 0$  的  $X$  轴上的点仍然被保留在  $Y' = 0$  轴, 且  $Y$  改变正负号时  $X'$  不能变, 这给出:

$$b = h = 0; \quad d = f = 0$$

$$X' = \frac{aX + c}{gX + i}; \quad Y' = \frac{eY}{gX + i}$$

为了使变换非退化, 要求  $e \neq 0, gi \neq 0, ac \neq 0$ .

情况一是  $g = 0, i \neq 0$ , 那么:

$$X' = \frac{a}{i}X + \frac{c}{i}; \quad Y' = \frac{e}{i}Y$$

这代表所有的望远系统, 物像距间为线性关系, 横向放大率在整个空间都是常数. 两个共焦放置的薄透镜, 平面镜与平面的折射系统在傍轴近似下都是这样的系统. 三个放大率分别为:

$$\begin{aligned} \beta &= \frac{Y'}{Y} = \frac{e}{i} \\ \Xi &= \frac{dX'}{dX} = \frac{a}{i} \\ B &= \frac{dY'}{dX'} / \frac{dY}{dX} = \frac{\beta}{\Xi} \quad \text{i.e. } \beta = \Xi B \end{aligned}$$

从推导过程可以发现  $\beta = \Xi B$  其实是个恒等式.

情况二是  $g \neq 0$ , 对应所有的非望远系统. 先考虑纵向放大率:

$$\Xi = \frac{dX'}{dX} = \frac{ai - cg}{(gX + i)^2}$$

我们指出, 沿主光轴  $X$  方向传播的光线如果在像方仍然沿同样的  $X$  方向传播, 那么物沿  $X$  正方向移动时必然有像也沿  $X$  正方向移动<sup>3</sup>, 此时  $ai - cg > 0$ , 这被称为**折射式系统**(dioptric system). 相反地, 如果沿主光轴方向传播的光线如果在像方沿  $-X$  方向传播, 此时物若沿  $X$  正方向移动, 像将沿  $-X$  方向移动, 被称为**反射式系统**(catoptric system). 此时我们将  $X'$  变为  $-X'$  再讨论这个问题, 就会发现对应系数也满足  $ai - cg > 0$  了. 而更普遍的一类情况, 物像方的主光轴甚至不会在同一个空间中重合, 被称为**折轴式系统**(off-axis system). 此时  $X, X'$  本质上就位于不同空间. 我们接下来总是默认  $ai - cg > 0$  即  $\Xi > 0$ .

<sup>3</sup>请读者用费马原理证明它.

进行坐标平移:

$$x' = X' - \frac{a}{g} \quad ; \quad x = -\left(-\frac{i}{g} - X\right)$$

把大写  $Y$  改为小写  $y$  便得到:

$$xx' = ff' \quad ; \quad \beta = \frac{y'}{y} = -\frac{f}{x}$$

$$f = \frac{-e}{g} \quad ; \quad f' = \frac{ai - cg}{(-e)g}$$

这便是牛顿式的成像公式. 我们从中可以发现:

$$\frac{f'}{f} = \frac{ai - cg}{e^2} > 0$$

可见两个焦距一定同号. 它与折射率的关系是我们接下来要解决的关键问题.

## 16.4 非傍轴成像

# 第十七章 光学仪器知识摘要

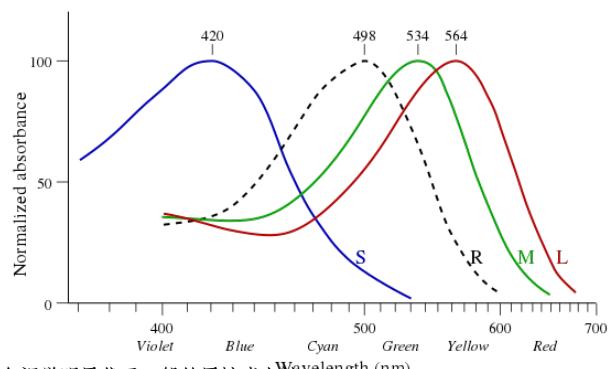
我们暂时不讨论光的干涉与衍射,故暂时不太需要光的波动理论.但实际上到了实用的层面上,各个光学仪器,从原理上说尽管几乎只用到几何光学的内容,似乎也是避不开对光的强度的讨论的.故本章先对在光的波动学说成型之前就已经蓬勃发展起来的光度学(photometry)进行说明.然后从若干方面对实用频率最高的一些光学仪器做未免以偏概全的介绍.

## 17.1 光度学

光源产生的光,在光学系统中可能被光学仪器最终接收,如光电二极管(photodiode),电荷耦合元件(charge-coupled device, CCD)等.也可以直接被人眼所直接观察.很明显,不同仪器作为探测仪器用时,其光谱响应将存在很大的区别.然而,既然是光学,从古到今仪器的设计都是围绕人的视觉展开.即最核心的波段是可见光波段.只是,随着科技的发展,眼睛的作用逐渐被物理仪器所替代,对光的探测也逐渐融入了更大范围的对电磁辐射的探测中,后者即辐射度量学(radiometry).比如当今天体物理中用到的各种探测手段,从波长最长的关于宇宙微波背景的射电探测,到波长最短的中子星 $\gamma$ 射线暴的探测,俨然不再是古人仰观星空可以达到的广度和深度.辐射度量学以能量作为最基本的物理量进行测量,单位就是焦耳,J.但是如果是考虑日常生活中以人眼视觉为核心设计的各类照明,媒体灯光,红外线,紫外线以外的波段几乎就被我们排除在外.在此之前我们需要先对人眼的独特的色视觉做一个了解:

### 17.1.1 色视觉

大部分人类的视网膜内有两种大类的视觉感受细胞:视锥细胞(cone)与视杆细胞(rod).而视锥细胞又根据其所含视蛋白分为感受500–700nm波长的视红素细胞(protan, erythrolabe, L,  $\rho$ );感受450–630nm波长的视绿素细胞(deutan, chlorolabe, M,  $\gamma$ )和感受400–500nm波长的视蓝素细胞(tritan, cyanolabe, S,  $\beta$ ).<sup>1</sup>



<sup>1</sup> 在女性中常见异常X染色体导致个体产生四种色素的视锥细胞从而其色视觉明显优于一般的男性或女性

**17.2 光阑与光瞳****17.3 眼睛****17.4 显微镜****17.5 望远镜****17.6 照相机**

# 第十八章 静电学

## 18.1 电荷与电场

### 18.1.1 电磁相互作用与电荷

物质若携带电荷(charge), 则可以发生电磁相互作用(electromagnetic interaction). 在经典情形下理解为电荷受到一个力. 我们先研究静电学(electrostatics), 它要求受力物体与施力物体<sup>1</sup>都处于静止状态, 或者认为一切电荷都处于静止状态.

电荷分正负, 阴阳激荡.

作为协变电磁场理论的基本要求与实验规律, 电荷是一个参考系不变的标量, 微观地看电荷具有量子化(quantized) 的奇妙特性<sup>2</sup>. 质子的质量与电子的质量没有简单的整数比关系, 但电荷量确实严格相反. 如果设想两者电量有着微小差异, 那么构成自然界的原子也就携带了一个能够传递远程相互作用的库仑力, 从而导致自然界的不稳定<sup>3</sup>. 而质子的电量被定义为基本电荷(elementary charge)<sup>4</sup>:

$$e = 1.602176634 \times 10^{-19} \text{ C}$$

实际上今天我们知道, 电荷量的量子化单位实际上更小, 至少是  $\frac{1}{3}e$ , 夸克的电量就是这个电量的一倍或两倍.

电荷如同静质量, 是基本粒子的基本属性. 它只能随着基本粒子运动, 没有相互作用时不能被创造与消灭. 在存在电磁相互作用时电荷可以被创造与消灭, 典型的过程如电子与正电子湮灭(annihilation) 为两个光子或高能光子在原子核附近产生(production) 电子正电子对. 这些过程电荷的代数和守恒, 其实现要求反物质与相对论性运动的存在. 在经典情形下有严格的电荷守恒定律(charge conservation), 电荷可以流动形成电流, 但流动的过程中电荷的总量不能发生改变. 我们总是计算一定体积内的电荷代数和:

$$dQ = \sum_i n_i q_i dV$$

其中  $n_i$  是第  $i$  种基本构成粒子的数密度,  $q_i$  是这种基本粒子的带电量.  $dV$  则为微元体积. 而上式也被写为:

$$dQ = \rho dV ; \quad \rho = \sum_i n_i q_i = \sum_i \rho_i$$

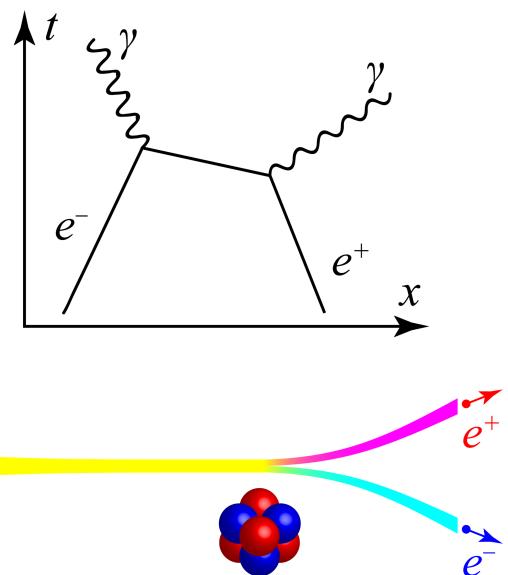


图 18.1: 电子对湮灭与产生

<sup>1</sup>实际上物理上不允许超距的施力物体, 施力物体就是局域的电场.

<sup>2</sup>与磁单极子的存在与量子化有关.

<sup>3</sup>是的, 这种思考方式就是人择原理(anthropic principle).

<sup>4</sup>这个值是 2018 新国际标准单位制提出的规定值.

### 18.1.2 库仑定律

**库仑定律**(Coulomb's law) 用于描述真空中的两个孤立静止点电荷之间的相互作用力:

$$\mathbf{F}_{12} = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{r^2} \mathbf{e}_{12}$$

其中系数  $k_e = \frac{1}{4\pi\epsilon_0} \approx 8.99 \times 10^9 \text{ N} \cdot \text{m}^{-2} \cdot \text{C}^{-2}$  为库仑常数.

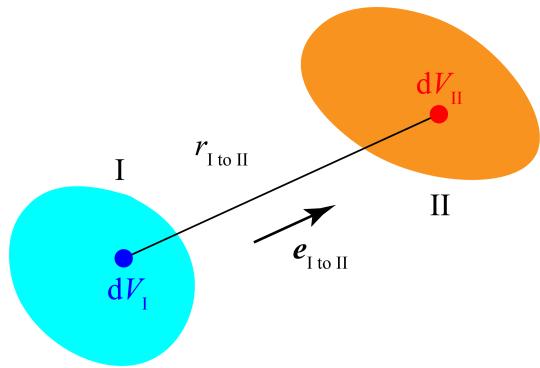


图 18.2: 电荷体系间的相互作用力  
加本性. 一个复杂电荷系统 (包括构成介质的电荷) 可以用连续的电荷密度描述:

$$dQ = \rho dV$$

那么两部分电荷之间的相互作用力为:

$$\mathbf{F}_{I \text{ to } II} = \frac{1}{4\pi\epsilon_0} \int_{V_{II}} \rho(\mathbf{r}_{II}) \int_{V_I} \frac{\rho(\mathbf{r}_I)}{r_{I \text{ to } II}^2} \mathbf{e}_{I \text{ to } II} dV_I dV_{II}$$

相互作用力在定义上就是成对存在的, 所以两部分电荷体系间的相互作用力满足牛顿第三定律: 大小相等, 方向相反, 对整个体系产生的力矩的贡献为零.

### 18.1.3 电场

现代物理认为, 狭义相对论下的时空观是正确的. 一个重要的推论就是**局域性原理**(principle of locality): 物体仅仅被此时此刻周围的物质所影响. 所有相互作用必须被理解为**近距作用**(action upon contact) 而不能是**超距作用**(action at a distance). 这是为了能与更本质的**因果律**(principle of causality) 相适应. 如果物体这一时刻的状态极其改变取决于同一时刻的空间上相隔一定距离的另外一个点 (类空间隔). 那么就可以另一个点的事件称作因, 而物体的状态与状态改变称作果. 但类空间隔在洛伦兹变换下具有可逆性. 也就是原则上可以在果的时空点也可以制造事件去对因的点产生影响. 从而造成逻辑上的紊乱. 所以我们必须规定这两个时空点间不存在因果关系, 互相独立没有影响.

局域性原理的直接结论便是, 在引力与电磁相互作用时, 作为物质的**场**(field) 的概念的引入. 场本是数学概念. 在一个三维的欧几里得空间  $\mathbb{R}^3$  中, 如果每一个点都定义了一个数 (标量), 代表某种物理量, 物理现象可以在不同参考系, 惯性系中观察, 所对应的时空点的位置坐标可能发生改变, 但特定的时空点的概念, 和与之相结合的物理现象与描述它的这个数, 不应该发生改变, 称为**标量**(scalar):

$$f: \mathbf{r} \mapsto f(\mathbf{r})$$

如,压强场,温度场,波动的相位场,都是典型的标量场的例子.数学上定义了标量场的梯度(gradient)来描述在一点极其附近的标量场的行为,它是一个矢量,方向是使得标量上升最快的欧几里得空间中的方向,大小是沿这个方向的标量变化的方向导数.数学上可以证明:

$$\nabla f = \frac{\partial f}{\partial x} e_x + \frac{\partial f}{\partial y} e_y + \frac{\partial f}{\partial z} e_z$$

物理量在某点  $\mathbf{r}$  的局部,偏离一个  $d\mathbf{r}$  后值变为:

$$f(\mathbf{r} + d\mathbf{r}) = f(\mathbf{r}) + \nabla f \cdot d\mathbf{r}$$

这就是微元形式的牛顿-莱布尼茨定理(Newton-Leibniz theorem).而标量场的梯度往往也具有重要的含义.压强场的梯度决定浮力的方向与大小(反向),温度场的梯度决定热传导的方向与强弱(也是反向),波动的相位场的梯度则决定波的传播方向与波长(波矢).

而在三维的欧几里得空间中在每一个空间点也可以引入一个矢量(vector),它不占三维空间的有限体积<sup>5</sup>而是定义在一个点上,可以认为是这个点上扩展出了一个新的与原三维空间坐标平行的  $e_x - e_y - e_z$  坐标系.而这一点定义的矢量就位于这个“浓缩在一点”的坐标系中<sup>6</sup>:

$$\mathfrak{F}: \mathbf{r} \mapsto \mathbf{F}(\mathbf{r})$$

显然,标量场的梯度恰是一个矢量场.矢量不能简单理解为三个标量场的合成.因为在空间旋转(或时空的惯性系变化)下,标量应该是不变的,而矢量将发生与坐标系相似的旋转,从而三个分量之间相互转化,只有矢量的模长是不变的.另一种典型的矢量场是流体的流速场  $\mathbf{v}$ ,或者更普遍的,某种物理属性的流密度场  $\mathbf{j}$ ,前者定义为单位时间通过单位面积的体积,后者则推广为单位时间通过单位面积的某物理量  $Q$ :

$$dQ = \mathbf{j} \cdot dA dt$$

在一点附近的矢量场的行为则比标量场复杂很多:

$$\mathbf{F}(\mathbf{r} + d\mathbf{r}) = \mathbf{F}(\mathbf{r}) + d\mathbf{r} \cdot \nabla \mathbf{F}$$

此处  $\nabla = e_x \frac{\partial}{\partial x} + e_y \frac{\partial}{\partial y} + e_z \frac{\partial}{\partial z}$  可以形式地视为矢量微分算符,它写在上一式的后一项中实际上是在表示:

$$\begin{aligned} d\mathbf{F} &= dF_x e_x + dF_y e_y + dF_z e_z \\ &= (\frac{\partial F_x}{\partial x} dx + \frac{\partial F_x}{\partial y} dy + \frac{\partial F_x}{\partial z} dz) e_x \\ &\quad + (\frac{\partial F_y}{\partial x} dx + \frac{\partial F_y}{\partial y} dy + \frac{\partial F_y}{\partial z} dz) e_y \\ &\quad + (\frac{\partial F_z}{\partial x} dx + \frac{\partial F_z}{\partial y} dy + \frac{\partial F_z}{\partial z} dz) e_z \\ &= (dx \frac{\partial}{\partial x} + dy \frac{\partial}{\partial y} + dz \frac{\partial}{\partial z})(F_x e_x + F_y e_y + F_z e_z) \\ &= d\mathbf{r} \cdot \nabla \mathbf{F} \end{aligned}$$

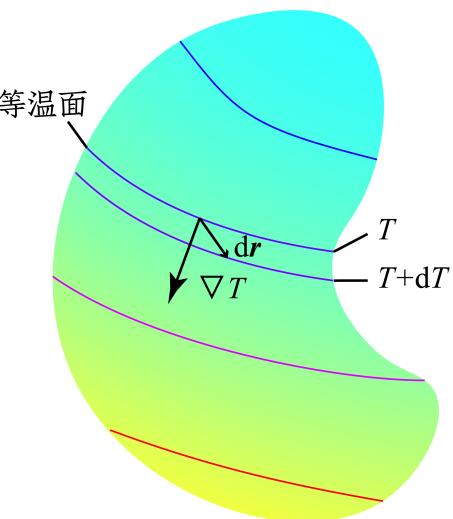


图 18.3: 温度场的梯度

<sup>5</sup>即,不同于位移矢量那样必须与三维空间中的不同点同时建立联系

<sup>6</sup>数学术语把空间叫做微分流形(differentiable manifold),而局部的矢量位于的空间叫做纤维(fiber)或切空间(tangent space),每一点都有一个这样的空间,合称纤维丛(fiber bundle).

而也可以把  $\nabla F$  认为能完整描述这个矢量在该点处的一阶变化行为,也称为矢量的梯度,但它实际上由九个分量组成,分别描述三个分量在三个方向的变化率,一般写在一张数表中,也就是矩阵<sup>7</sup>:

$$\nabla F = \begin{bmatrix} \frac{\partial F_x}{\partial x} & \frac{\partial F_y}{\partial x} & \frac{\partial F_z}{\partial x} \\ \frac{\partial F_x}{\partial y} & \frac{\partial F_y}{\partial y} & \frac{\partial F_z}{\partial y} \\ \frac{\partial F_x}{\partial z} & \frac{\partial F_y}{\partial z} & \frac{\partial F_z}{\partial z} \end{bmatrix}$$

显然用这样的一个矩阵去描述矢量场在一点处的行为是代数的,抽象地,不几何直观的.所以我们引入两个物理上会更实用的概念.

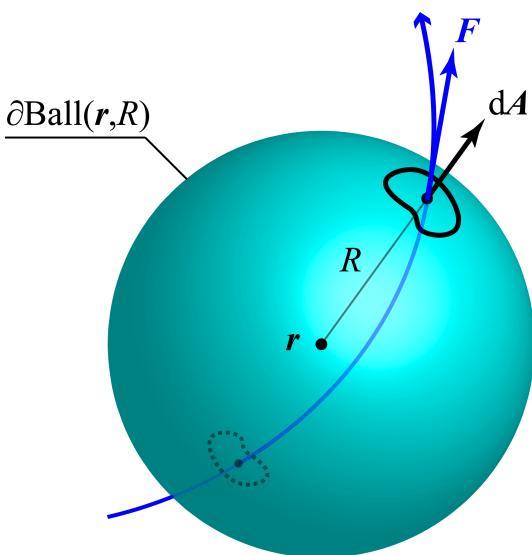


图 18.4: 矢量场的散度

一是矢量场的散度(divergence),它是一个标量,定义为矢量场的源强度:

$$\nabla \cdot F = \lim_{R \rightarrow 0} \left( \oint_{\partial \text{Ball}(r, R)} F \cdot dA \right) / \left( \frac{4}{3} \pi R^3 \right)$$

其中  $\text{Ball}(r, R)$  表示以待研究的点为球心的,半径为  $R$  的球形体,而  $\partial \text{Ball}$  表示它的表面.积分号  $\oint$  用来代表此场合下用来积分的区域是一个闭合的区域<sup>8</sup>.积分的量称为通量(flux),即矢量场与面积的电场,速度场情况下即单位时间通过面  $dA$  的体积.现在对闭合的面进行积分,再除以面包含的体积,代表的是该点单位体积内产生这个矢量场的场线的源的多少,即源强度.数学上可以证明,它与体积以怎样的取法趋于零无关,如果取成与坐标系平行的长方体的形状,能够在坐标系下计算出它的值为:

$$\nabla \cdot F = \frac{\partial F_x}{\partial x} + \frac{\partial F_y}{\partial y} + \frac{\partial F_z}{\partial z}$$

第二个概念便是旋度(curl)了.它实际上代表矢量场的变化矩阵表中反对称的那一部分<sup>9</sup>.它可以被写为一个矢量.与在这一点处的无限小的环积分有关,数学上不难证明:

$$n \cdot \nabla \times F = \lim_{R \rightarrow 0} \left( \oint_{\partial \text{Circ}(r, n, R)} F \cdot dl \right) / (\pi R^2)$$

由于半径为  $R$  的圆的选取还依赖于其朝向.故在上式中我们对积分的方向和垂直于积分圆面的法向单位矢量  $n$  采取了右手定则式的对应.而最后积分出来得到的数恰好是面法向单位矢量与从  $F$  的各个导数中提取的旋度矢量的点乘.环积分通常被称为环量(circulation),旋度则反应矢量场局域的涡旋程度.数学上可以证明旋度可以由行列式表示为:

$$\nabla \times F = \begin{vmatrix} \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ F_x & F_y & F_z \\ e_x & e_y & e_z \end{vmatrix}$$

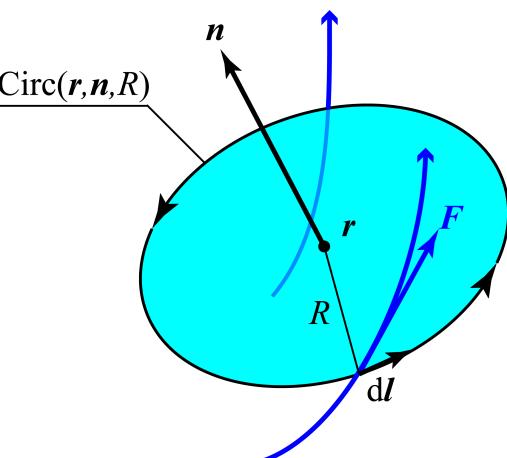


图 18.5: 矢量场的旋度

<sup>7</sup>实际上,形成了一个张量场(tensor filed).张量将不在本教材的讨论范围内.

<sup>8</sup>数学上称为又闭又开(clopen)的集合,表示它没有边界.

<sup>9</sup>任何一个矩阵都可以被唯一地分解为一个对称矩阵与反对称矩阵之和,前者有六个独立分量,包含原来的对角线与对称位置的两元素和的一半,后者仅有三个独立分量,为对称两个位置元素差的一半.

作为与牛顿-莱布尼茨定理对应的散度与旋度版本,根据两个量的定义可知前者发生在体积内部的积分与其闭合表面上,后者发生在曲面上面积分和它的闭合曲线边界上:

$$\int_V \nabla \cdot \mathbf{F} dV = \oint_{\partial V} \mathbf{F} \cdot d\mathbf{A}$$

$$\int_A \nabla \times \mathbf{F} \cdot d\mathbf{A} = \oint_{\partial A} \mathbf{F} \cdot dl$$

前者为[奥斯特诺格拉德斯基-高斯定理](#)(Ostrogradsky-Gauss theorem),后者则为[开尔文-斯托克斯定理](#)(Kelvin-Stokes theorem)<sup>10</sup>.

现在让我们回到库仑定律,现在就必须把它理解为两个公式:

$$\mathbf{E} = \frac{1}{4\pi\epsilon_0} \frac{Q}{r^2} \mathbf{e}_r \quad ; \quad \mathbf{F} = q\mathbf{E}$$

前一半式子说明,即使没有引入任何试探电荷  $q$ ,在空间中距离场源电荷  $Q$  为  $r$  处依然存在某种独立于电荷而存在的“场”.否则也就不可能对引入的电荷瞬间产生力的作用.这个场由电荷激发,以后将知道它有自己的能量与动量,而且都是可以被局域地确定<sup>11</sup>.经典物理中似乎只能由电荷产生场,但高能物理下场也能反过来产生电荷粒子,实际上此时会将电磁场理解为“光子”这样的物质.电磁场的特点是能对带电荷的物质产生作用.

而后一半的式子则给出了电磁场中一点处电场强度矢量  $\mathbf{E}$  的定义.带点粒子如果在电磁场中静止,那么将受到一个与电荷量成正比的力.把力与电荷量的比值定义为电场强度矢量.

显然,带电粒子的引入也会产生电场.这个电场与原来场源电荷产生的电场叠加,形成了空间的真实电场分布.但在计算这个“试探电荷”的电场时,不应该考虑自己在自己这一点处产生的电场.因为自己对自己的力总是零.而总是计算“外场”在试探电荷这一点的值.不过,试探电荷的引入与否可能会对外界的场源电荷部分产生改变.从而改变外场.这又需要另说了,我们下一章的导体静电感应就是在不断讨论这个问题.

电场满足[叠加原理](#)(superposition principle).我们总是一对一对地考虑电荷之间的二体作用,把相互作用力做非相干的矢量叠加.这也就是说,给定一个场源电荷分布,可以叠加计算出空间中任意一点的电场强度:

$$\mathbf{E}(\mathbf{r}) = \int_{V'} \frac{1}{4\pi\epsilon_0} \frac{\rho e_R}{R^2} dV' \quad , \quad \mathbf{R} = \mathbf{r} - \mathbf{r}'$$

$$\mathbf{E}(\mathbf{r}) := \frac{\mathbf{F}_{\text{static}}}{q}$$

也是出于这个原因,所有由电场强度派生出来的量:电势,电通量等等都符合线性的叠加原理.马上就会介绍,这是由于描述电场与电荷关系的基本方程是一个线性的微分方程的缘故.

## 18.2 两个定律与电势

库仑定律是静电学最基本的公理,但我们研究电场作为场的性质,它与电荷的关系.为以后深入了解电磁场整体打好基础.

<sup>10</sup>微分流形上的微分形式有一个[嘉当-斯托克斯定理](#)(Cartan-Stokes theorem) 统一了以上所有形式:

$$\oint_{\partial\Omega} \omega = \int_{\Omega} d\omega$$

<sup>11</sup>即,可以确定每一点的场能量动量密度.

### 18.2.1 电场的高斯定律

首先让我们研究电通量(electric flux), 通过一个曲面(不一定闭合)的电通量定义为曲面上电场强度的积分:

$$\Phi = \int_A \mathbf{E} \cdot d\mathbf{A}$$

我们在这儿必须引入有向曲线, 曲面, 体积的概念. 固然, 曲线, 曲面, 体积都没有方向与正负, 它们仅仅是特定空间中的几何形状. 但定义各种积分时, 曲线积分的线元矢量  $d\mathbf{l}$ , 曲面积分的面元矢量  $d\mathbf{A}$  都有两个可能的方向. 而体积元  $dV$  也不一定恒正, 在某些情况下将取负. 那么我们通常会有一个符号约定:

- 对于可定向(orientable)曲面(不可定向的比如莫比乌斯环), 总是定义一个方向为正, 即为每一个面元指定一个面元矢量, 它取垂直于面的两个方向之一, 而相邻的面元具有相同的取向. 而如果涉及到面边界上闭合曲线的环积分, 则线元矢量取的方向与面元构成右手定则(right-hand rule).
- 对于连通.connected)体积区域, 如果取体积微元为正, 那么涉及到区域边界闭合曲面上的面积分时, 面元取从体积内到体积外的方向. 如果取体积微元为负, 则面上面元取体积外到体积内的方向.

而我们之前写出的奥-高定理和开-斯定理, 都是在这个意义下正确的.

从而以上电通量也就因为我们选择的面的两种可能的正向而造成两个不同的结果, 它们相差一个相反数.

电场的高斯定律(Gauss's law)指出:

$$\oint_{\partial V} \mathbf{E} \cdot d\mathbf{A} = \frac{1}{\varepsilon_0} \int_V \rho dV$$

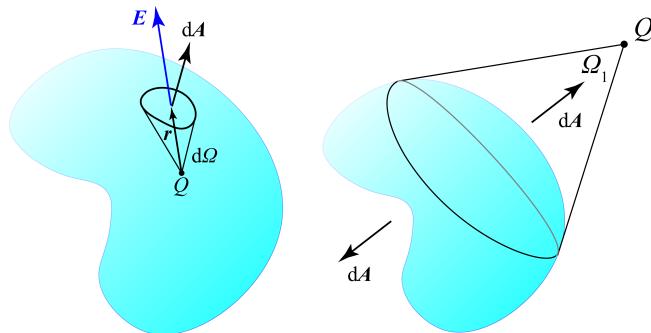


图 18.6: 证明高斯定律

我们来用库仑定律证明高斯定律. 首先考虑一个在闭合曲面内部的点电荷, 我们发现面元上的通量:

$$d\Phi = \mathbf{E} \cdot d\mathbf{A} = \frac{1}{4\pi\varepsilon_0} \frac{Q e_r \cdot d\mathbf{A}}{r^2}$$

而这一个小面元对点电荷张的立体角恰好是:

$$d\Omega = \frac{e_r \cdot d\mathbf{A}}{r^2}$$

故对面积分总立体角给出  $4\pi$  的结果:

$$d\Phi = \frac{Q}{\varepsilon_0} \cdot \frac{d\Omega}{4\pi} \Rightarrow \Phi = \frac{Q}{\varepsilon_0} \cdot \oint_{\partial V} \frac{d\Omega}{4\pi} = \frac{Q}{\varepsilon_0}$$

我们再考虑在面外的点电荷  $Q$  对面产生的电通量, 此时面的正向与点电荷产生电场方向的点乘就有了正负号, 积分以后得到的结果是通过点电荷去看曲面, 把曲面对点电荷张的最大视角处的曲线找到, 将曲面

分为两个部分,立体角都是  $\Omega_1$ ,总积分为:

$$\Phi = \frac{Q}{\varepsilon_0} \cdot \oint_{\partial V} \frac{d\Omega}{4\pi} = \frac{Q}{\varepsilon_0} (\Omega_1 - \Omega_1) = 0$$

从而我们证明了,电荷若在闭合曲面内,则电通量贡献为  $\frac{Q}{\varepsilon_0}$ .而若在闭合曲面外,则电通量贡献为 0.再由叠加原理,得任意情况下闭合曲面的电通量等于内部总电荷量除真空介电常数  $\varepsilon_0$ .考虑电荷连续分布或准连续分布的情况下,有:

$$\oint_{\partial V} \mathbf{E} \cdot d\mathbf{A} = \frac{1}{\varepsilon_0} \int_V \rho dV$$

此即高斯定律,由数学上的奥-高定理,我们可以把上式写为局域的微分形式:

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\varepsilon_0}$$

高斯定律旨在说明:电荷产生电场(电荷是电场的源强度).

### 18.2.2 电势与电场的环路定律

对点电荷产生的电场  $\mathbf{E} = \frac{Q}{4\pi\varepsilon_0 r^2} \mathbf{e}_r$ ,再引入以下标量场是方便的:

$$\varphi = \frac{1}{4\pi\varepsilon_0} \frac{Q}{r}$$

这是因为如果计算球坐标下的梯度:

$$\nabla = \mathbf{e}_r \frac{\partial}{\partial r} + \frac{\mathbf{e}_\theta}{r} \frac{\partial}{\partial \theta} + \frac{\mathbf{e}_\varphi}{r \sin \theta} \frac{\partial}{\partial \varphi}$$

便得到以下关系式:

$$\mathbf{E} = -\nabla \varphi$$

即电场可以表示为我们引入的另一个标量场的负梯度的形式.这个标量场就叫做**电势**(electric potential).而以上关系,结合梯度的含义理解,便是沿电场线方向电势降低.可以通过牛-莱定理写为积分形式(下式与路径无关):

$$\int_A^B \mathbf{E} \cdot d\mathbf{l} = \varphi_A - \varphi_B$$

以上关系式由叠加原理,在任意电荷分部的情况下都成立,此时对应的电势标量场为:

$$\varphi(\mathbf{r}) = \int \frac{1}{4\pi\varepsilon_0} \frac{\rho}{|\mathbf{R}|} dV' , \quad \mathbf{R} = \mathbf{r} - \mathbf{r}'$$

对于电势的引入可能存在一点疑惑:我们都应该知道电场的定义是可以独立于场源电荷的,它为在一点处引入静止的试探电荷时单位电荷的受力.那么电势是否可以独立于场源电荷而定义呢?观点一比较基础,它认为在静电场情况下既然全空间电场已经被定义了,它就应该反映了该点处场的全部实际信息,那么电势作为与电场符合一个梯度关系的梯度场,也就能被相应的确定下来,两个可能的电势还可以差一个全空间都一致的常数.也就是这种观点把电场视为更加基础的场的特征.下面便可以知道电势的含义是引入试探电荷后单位试探电荷在该点处的电势能.而观点二则更加本质,我们今天知道了势比场强实际上包含更多的信息.由于微观粒子的量子特性,力的概念已然失效,从而场强反而不如用势来描述一个场更为直接而实用.电磁场论实际上也就是一个矢量场论,它由矢势和标势组成一个四维矢量来完整地描述,但取法不唯一,多出来的冗余自由度被称为**规范**(gauge),它恰好就在表示电荷与场的相互作用时,与电荷的量子相位协同变化,从而造成实际可以观测的影响.为了实际问题的方便,通常约定势有特定的取法.在静电学的常见情况下(电荷分布区域有限时),常常取无穷远处电势为零,注意上电势积分式也恰恰符合这一点.

我们进一步考察电势的物理含义,如果把以上电场视为外电场,引入试探电荷  $q$ ,并在以上积分形式的场强与势的关系中乘以  $q$ ,便发现  $q\mathbf{E}$  就是电荷的受力,左边就是表示电场力的做功:

$$W_{A \rightarrow B} = q\varphi_A - q\varphi_B$$

很自然地我们发现电场力是保守力,而对应的电势能,基于它应该与电荷量成正比的观点,我们总是直接取为:

$$E_p = q\varphi$$

从而电势的效果,是为试探电荷带来一个电势能.

最后我们再考察电场的一个特性,由于任何路径上的电场积分可以根据牛-莱定理写为端点的电势差的形式,考虑闭合的曲线积分,它可以看成一个可定向曲面的边界,由于两端点重合:

$$\oint_{\partial A} \mathbf{E} \cdot d\mathbf{l} = 0$$

这就是电场的环路定律(circuital law),再根据数学上的开-斯定理,我们写出环路定律的微分形式:

$$\nabla \times \mathbf{E} = \mathbf{0}$$

也就是静电场无旋.值得指出这不过是静电场可以表示为标量电势场的负梯度的直接结果,因为任意一个标量场的梯度都是无旋的<sup>12</sup>:

$$\oint_{\partial A} \nabla f \cdot d\mathbf{l} = 0 \Rightarrow \nabla \times \nabla f = \mathbf{0}$$

### 18.2.3 总结

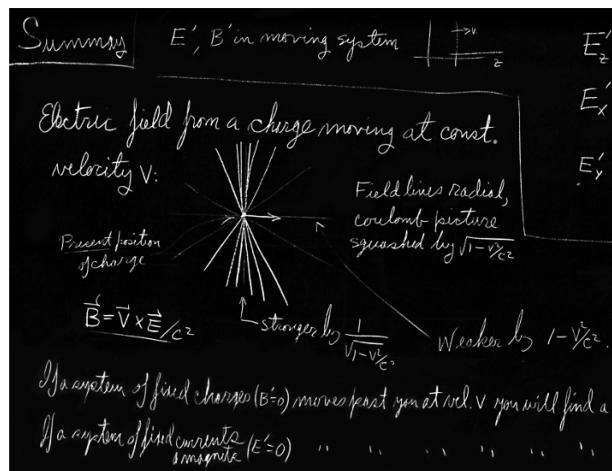


图 18.7: Feynman 讲座一瞥

我们现在有三个不同的层次去看待电荷与电荷之间的相互作用问题<sup>13</sup>:

#### 1. 不引入场:

$$\mathbf{F} = \frac{1}{4\pi\epsilon_0} \frac{Qq}{r^2} \mathbf{e}$$

#### 2. 引入场,用场强描述:

$$\nabla \cdot \mathbf{E} = \frac{\rho}{\epsilon_0} ; \quad \nabla \times \mathbf{E} = \mathbf{0}$$

$$\mathbf{F} = q\mathbf{E}$$

#### 3. 引入场,用势描述:

$$\nabla^2 \varphi = -\frac{\rho}{\epsilon_0}$$

$$E_p = q\varphi$$

<sup>12</sup>这个结论向流形推广就是著名的庞加莱引例(Poincaré lemma)了,它指出任意  $p-$  形式  $\omega$  的二次全微分为零:

$$d\omega = 0$$

一个以后会用得上的推论可以通过考察矢量场的旋度在闭合曲面上的面积分,而把闭合曲面看成是边界曲线缩小到零的有边界曲面而得:

$$\oint_{\partial V} \nabla \times \mathbf{F} \cdot d\mathbf{A} = 0 \Rightarrow \nabla \cdot \nabla \times \mathbf{F} = 0$$

<sup>13</sup>拉普拉斯算子  $\nabla^2 = \nabla \cdot \nabla$ ,如作用在标量上即表示先求梯度后求散度:

$$\nabla^2 f = \nabla \cdot (\nabla f)$$

以上描述一个比一个适用范围更广,第一个仅仅适用于静电场或低速情况,第二个则是对高速运动的电荷与电磁场适用,比如匀速运动的电荷由简单相对论讨论可得:

$$\mathbf{E} = \frac{1}{4\pi\epsilon_0} \frac{Q}{r^2} \frac{1 - \beta^2}{(1 - \beta^2 \sin^2 \theta)^{3/2}} \mathbf{e}_r$$

$$\mathbf{B} = \frac{\mathbf{v} \times \mathbf{E}}{c^2}$$

它要符合第二个式子(旋度需要做相对论修正),而第三个式子更是适用于非经典场论情况.

## 18.3 静电能

我们将进一步深入研究电场相互作用中的能量概念.

### 18.3.1 静电势能

之前已经引入了电势能的概念,我们把外场中引入的试探电荷(点电荷)所具有的势能的大小计做电势能,并得出了电势能的表达式:

$$E_p = q\varphi$$

还指出了,它的作用是用来表示不变的外电场对电荷的做功的多少,基于这样的观察,我们把这个能量称为**静电势能**(electrostatic potential energy),静电的含义,是要求外场不能随着试探电荷的引入与移动而改变.而更常见的一类情况是,随着试探电荷的移动,场源电荷也在随着改变,如之后将会经常讨论的静电感应的情况.此时仍然可以计算外场的场强与电势,并引入试探电荷以后依然可以形式地写出**电势能**(electro potential energy)的值  $E_p = q\varphi$ ,但它不再代表实际移动试探电荷电场对电荷做的功了,它代表什么?电场对电荷做功普遍地应该怎么表示?外力做功又如何和它发生联系,这是我们之后的讨论关心的问题.

首先对于这个问题的第一步考察是,虽然试探电荷的特定位置带来特定的外场,但一旦确定了外场的场源的电荷分布,便可以独立于试探电荷地产生一个分布于全空间的电势与电场,而可以设想固定这个电荷分布不变,而把试探电荷放在空间中的任意一点,讨论其受力与能量.这些改变试探电荷位置而不改变外场的位移称为**虚位移**(virtual displacement).它是一种广泛用来建立理论模型的方法.它的强大之处在于虚位移的任意性,在每一个实际时刻试探电荷可以在任意位置,外场在缓慢随时间发生改变,而试探电荷发生的实际位移是所有虚位移中的一个特例.在这个意义上,电势能的含义便在于,试探电荷发生虚位移时,即场源电荷分布不变时,电场力对试探电荷做的虚功即为电势能的减小,此即静电场下的虚功原理:

$$\delta E_p = -q\mathbf{E} \cdot \delta\mathbf{r}$$

我们可以把这个概念推广,如果在该外场中引入不止一个点试探电荷,我们写出一个电荷体系在外场中的电势能来:

$$E_p = \sum_i q_i \varphi_i$$

这个写法存在一些问题,首先是如果没有引入任何试探电荷,体系是否固有一个能量?用场的角度来理解就会发现不对,场是一种物质,物质必定伴随着能量.或者考虑场源电荷之间的相互作用,它也一定会伴随着能量.再者,引入两个以上的点电荷时,它们之间的相互作用能没有被包含在以上项中.第三,即使是单个点电荷单独存在,也必定会带来其自能.我们需要从更基础的角度来讨论电荷体系的能量.

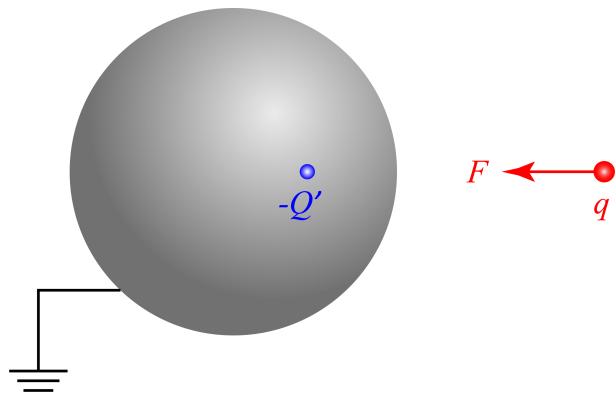


图 18.8: 情形: 场源电荷取决于试探电荷

### 18.3.2 自能与相互作用能

考虑真空中的两个点电荷之间的相互作用. 电荷  $q_1$  在电荷  $q_2$  处产生的电场与电势为  $\mathbf{E}_{12}$  与  $\varphi_{12}$ , 电荷  $q_2$  在电荷  $q_1$  处产生的电场与电势为  $\mathbf{E}_{21}$  与  $\varphi_{21}$  无疑, 相互作用力是相互的, 满足牛顿第三定律  $q_1\mathbf{E}_{21} + q_2\mathbf{E}_{12} = \mathbf{0}$ . 那么电势能是相互的吗? 注意到:

$$q_1 U_{21} = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{r} = q_2 U_{12}$$

从而两个电势能实际上是相等的. 根据电势能的定义, 它等于 1 不动 2 移到无穷远电场力所做的功, 也会等于 2 不动 1 移到无穷远电场力所做的功. 这两种情况平移对称, 从而我们有理由判断: 无论如何移动两个电荷, 将 1 与 2 分开到无限远时电场力做功都会是以上值, 这一点也不难在数学上证明.

从而我们把这个能量称为 12 间相互作用力对应的相互作用能(interaction energy). 它可以用任何一个电荷乘以另一个电荷在这个电荷处产生的势相乘得到, 或者找到这一对相互作用, 直接写出:

$$I_{12} = \frac{1}{4\pi\epsilon_0} \frac{q_1 q_2}{r_{12}}$$

按照这思路, 由叠加原理, 我们可以很轻易地写出点电荷体系与连续分布电荷体系的总相互作用能:

#### 1. 点电荷体系:

$$I = \sum_{\{i,j\}!} \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}}$$

#### 2. 连续分布电荷体系:

$$I = \iint_{\{\mathrm{d}V, \mathrm{d}V'\}!} \frac{1}{4\pi\epsilon_0} \frac{\rho \rho'}{r} \mathrm{d}V \mathrm{d}V'$$

关于以上求和我们做出以下解释:

首先, 点电荷体系的相互作用能表示把点电荷从计算这个能量的状态移动到点电荷彼此之间都相距无穷远时电场力所做的功. 但一定要注意每个点电荷移动到无穷远以后还会因为自身的电磁相互作用而产生一个待定的能量, 叫做自能, 我们紧接着会讨论到这个问题. 而连续分布电荷体系则表示把每一份电荷微元移动到无穷远以后电场力做的功, 移动到无穷远以后电荷无限分散(宏观意义上), 所以每一份微元电荷对应的电磁相互作用能真实地趋于零, 故我们计算的这个能量就是这个体系的总电磁相互作用对应的能量.

第二, 是这个求和表示的含义. 我们用  $(i, j)$  来表示特定集合中两个元素  $i, j$  构成的有序对(ordered pair), 而  $\{i, j\}$  则表示交换  $i, j$  后表示同一代数元素的<sup>14</sup>对称无序对(symmetrized unordered pair), 即  $\{i, j\} = \{j, i\}$ . 求和下加这个对符号表示对所有一对一对的情况进行求和. 每一对只需要要求一次即可. 这之后还跟着一个感叹号, 它表示  $i, j$  必须取不同值(而  $\mathrm{d}V$  与  $\mathrm{d}V'$  可以取相同值, 这些项会在体积很小时求和也很小而可以忽略). 那么由简单的代数变形, 我们发现上两式即:

$$I = \frac{1}{2} \sum_i \sum_{j \neq i} \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}} = \frac{1}{2} \sum_i q_i \phi_i$$

$$I = \frac{1}{2} \int \int_{\mathrm{d}V \mathrm{d}V'} \frac{1}{4\pi\epsilon_0} \frac{\rho \rho'}{r} \mathrm{d}V \mathrm{d}V' = \frac{1}{2} \int \rho \varphi \mathrm{d}V$$

注意到第一式中  $\phi_i$  表示的是别的电荷在  $q_i$  处产生的电势, 不包括自己在自己上产生的, 我们把符号稍微变动一下以示区别. 而在后一个表达式中则不需要考虑这种区别, 这个式子将作为一个普遍的求总电磁相互作用能的表达式而使用.

<sup>14</sup>以后还会用到  $n$  个元素的对称无序类(symmetrized unordered setoid) 与反对称无序类(antisymmetrized unordered setoid):

$$\{i_1 \cdots i_p \cdots i_q \cdots i_n\} = \{i_1 \cdots i_q \cdots i_p \cdots i_n\}$$

$$[i_1 \cdots i_p \cdots i_q \cdots i_n] = -[i_1 \cdots i_q \cdots i_p \cdots i_n]$$

第三,是这个求和的使用方法.实际上,如果电荷连续分布系统从一个状态变化到另一个状态,电荷分布与它产生的电势分布都发生改变,那么前后的能量增加:

$$\Delta I = \frac{1}{2} \int \rho' \varphi' dV - \frac{1}{2} \int \rho \varphi dV$$

自然表示电场对电荷体系所做的负功,或者表示由电荷流向电场的能量,它使得场的能量,也就是  $I$ ,发生了增长<sup>15</sup>.而反过来考察电荷体系,如果外力对电荷体系做功为  $W$ ,那么根据动能定理应该有:

$$W - \Delta I = \Delta E_k$$

在十分缓慢地移动电荷的情况下,亦或是导体在载流子为电子,其动能十分的小可以忽略的情况下,动能的增长可以视为零,从而有:

$$W = \Delta I$$

即,外力做功,通过电荷为媒介,转化为了场的能量的增长.这个观点也是我们经常用到的.

我们最后回到电势能与相互作用能的关系,考虑整个体系由两部分电荷体系构成,那么由叠加原理,整个空间中每一点的电势可以写成两部分贡献的和  $\varphi_1 + \varphi_2$ .我们只需要对全空间的电荷积分,就得到了整个体系的相互作用能:

$$\begin{aligned} I &= \frac{1}{2} \int_{I+II} \rho \varphi dV \\ &= \frac{1}{2} \int_I \rho_1 (\varphi_1 + \varphi_2) dV_1 + \frac{1}{2} \int_{II} \rho_2 (\varphi_1 + \varphi_2) dV_2 \\ &= \frac{1}{2} \int_I \rho_1 \varphi_1 dV_1 + \frac{1}{2} \int_{II} \rho_2 \varphi_2 dV_2 + \frac{1}{2} \left( \int_I \rho_1 \varphi_2 dV_1 + \int_{II} \rho_2 \varphi_1 dV_2 \right) \\ &= I_1 + I_2 + I_{12} \end{aligned}$$

我们于是找到了三个能量,  $I_1, I_2$  是两个电荷分布体系单独存在时各自的自相互作用能, 我们把它简称作**自能**(self-energy), 它一定是大于零的, 这一点下一节就可以发现. 而余下的交叉求和项就是表示两个体系之间的相互作用能. 仔细一看不难发现一项其实就是电荷体系 II 产生的外场下电荷体系 I 的电势能, 而另一项就是电荷体系 I 产生的外场下电荷体系 II 的电势能, 两者仔细研究下发现其实相等:

$$I_{12} = \iint_{I II} \frac{1}{4\pi\epsilon_0} \frac{\rho_1 \rho_2}{r} dV_1 dV_2 = \int_I \rho_1 \varphi_2 dV_1 = \int_{II} \rho_2 \varphi_1 dV_2$$

这一项可以大于零也可以小于零, 表示的是相互作用. 现在我们试探电荷体系在外场中的电势能有了一个新的看法, 它其实表示试探电荷体系与场源电荷体系的相互作用能, 作为相互作用, 它也可以反过来用场源电荷体系在试探电荷体系产生的场中的电势能来计算, 或是对称地, 用两者和的一半来计算.

最后让我们来看一下点电荷的自能. 实际上在之前点电荷体系相互作用能体系中我们已经指出, 其表达式中电荷  $q_i$  上电势  $\phi_i$  与真实电势  $\varphi_i$  的区别就在于没有计及点电荷在自己位置处产生的电势  $\varphi_{ii}$ , 从而漏掉的项其实就是点电荷的自能:

$$I_i = \frac{1}{2} q_i \varphi_{ii}$$

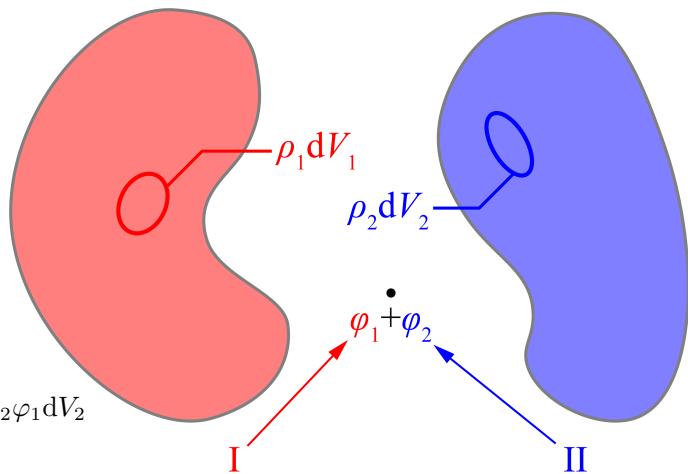


图 18.9: 自能与相互作用能

<sup>15</sup>场只能和带荷的物质发生作用, 这也是唯一的使得场能量变化的方法

这些项不会影响我们的实际计算,因为点电荷总被我们认为是没有内部改变的理想体系,我们经常把微观基本粒子做这样的抽象,因为它们都是小到可以忽略其尺寸.当粒子速度远小于光速,且不涉及到高能粒子反应的情况下,它们也的确不发生任何内部的改变,自能精确地保持不变.所以只需要计算扣除点电荷自能的相互作用能表达式,就能够知道有多少势能被存储在场中,就能够用它来写出电荷分布发生改变时释放出来的能量.

点电荷的自能如何计算呢?这取决于我们采取的模型.如果我们考虑的是一个电荷均匀分布在表面的球体,那么自能为:

$$I = \frac{1}{2}Q \cdot \frac{1}{4\pi\epsilon_0} \frac{Q}{r}$$

对于电子,如果我们认为电子的静质量全部来源于它携带的场的惯性,那么能够写出:

$$I = \frac{e^2}{8\pi\epsilon_0 r} = m_e c^2$$

事实上我们并不能确定电子内部电荷分布,所以我们更常直接定义**经典电子半径**(classical electron radius) 为:

$$\frac{e^2}{4\pi\epsilon_0 r_e} = m_e c^2 \Rightarrow r_e = \frac{1}{4\pi\epsilon_0} \frac{e^2}{m_e c^2} = 2.82 \times 10^{-15} \text{ m}$$

然而这仅仅是一个方便计算的实用常数,并不能代表电子的真实情况!事实上用不带电荷的粒子与电子发生散射实验就会发现,电子的尺寸比这要小得多,上面的计算结果甚至比质子的半径还要大几倍!而实验更是发现电子在多种意义上就是一个点粒子,根本没有内部组成结构,十分稳定不能衰变.那么显然从上式发现这也面临着自能发散的问题.双重矛盾如何解决呢?

解决方案是量子力学的引入,电子不是点粒子,在原子尺度下电子就已经必须要由波函数描述了,电子没有同时确定的位置和速度,在原子外形成“电子云”.狄拉克 (P. Dirac) 于 1928 年提出**狄拉克方程**(Dirac equation),把电子像光子那样处理为一个传播的“场”,从而从根本上摒弃了电子作为点电荷的观点,电子的质量被理解为场的质量或者静能,而电子的电荷量被理解为与电磁场(光子)的耦合常数.从而把电荷作为一种物质和场物质彻底分割开来独立描述.还同时一并给出了电子的自旋与预言了电子的反物质:**正电子**(positron) 的存在.它们都被天然而巧妙的包含在一个统一的代数形式中,极大程度地推动了物理学理论的进展.

但这没有给出问题的全部解释,对于点电荷自能发散的问题这种做法仍然没有给出绝对的答案.经典地看一个匀速运动的经典无穷小点电荷会携带着一个发散的场共同向前运动,这在量子场论语言下,变成了电子与光子相互作用的无限可能性求和,从而也理所当然地引进了各种各样的发散.对于这些发散物理学家们发展出了各种各样的处理方法,但直到今天还仅仅停留在现象解释的层面,背后是否有更加简单的解释方法还是一个谜.

### 18.3.3 电场能

我们已经十分熟悉,电荷之间的相互作用能实际上就是场的能量.我们现在将给出这一说法的定域化描述.记得一个电荷连续分布体系的相互作用能为:

$$I = \frac{1}{2} \int_V \rho \varphi dV$$

而又根据场与电荷之间的关系:

$$\nabla^2 \varphi = -\frac{\rho}{\epsilon_0}$$

我们把以上积分写为:

$$I = -\frac{\epsilon_0}{2} \int_V \varphi \nabla^2 \varphi dV$$

再根据著名的<sup>16</sup>第一格林等式(Green's first identity):

$$\nabla \cdot (\phi \nabla \psi) = \phi \nabla^2 \psi + \nabla \phi \cdot \nabla \psi$$

我们再取  $\phi, \psi$  都为  $\varphi$ , 从而:

$$\varphi \nabla^2 \varphi = \nabla \cdot (\varphi \nabla \varphi) - (\nabla \varphi)^2$$

带入以上积分, 我们得到:

$$\begin{aligned} I &= \frac{\varepsilon_0}{2} \int_V (\nabla \varphi)^2 dV - \frac{\varepsilon_0}{2} \oint_{\partial V} \varphi \nabla \varphi \cdot dA \\ &= \int_V \frac{\varepsilon_0}{2} \mathbf{E}^2 dV + \frac{\varepsilon_0}{2} \oint_{\partial V} \varphi \mathbf{E} \cdot dA \end{aligned}$$

我们总是考虑电荷分布有限的情况. 此时最初的积分体积只要包含所有电荷即可, 那么我们最后得到的这个恒等式就十分的有意思, 它将一个体系的相互作用转化为了一个量的体积分和另一个量在包含所有电荷的面上的面积分的形式. 我们如果取体积为无穷大的全空间, 那么考虑到随着远离电荷体系  $R$  处  $\varphi$  以至多  $\frac{1}{R}$  的方式衰减(若体系总电荷量为零则衰减更加剧烈), 而  $\mathbf{E}$  以至多  $\frac{1}{R^2}$  的方式衰减, 而积分面积以  $R^2$  的方式增加, 故整个积分以至多  $\frac{1}{R}$  的方式衰减, 在无穷远处这个积分的第二项就会趋于零.

从而我们得到了:

$$I = \int \frac{\varepsilon_0}{2} \mathbf{E}^2 dV$$

从而发现相互作用能可以用场来定域地计算. 电场的能量密度(energy density) 为:

$$w_e = \frac{\varepsilon_0}{2} \mathbf{E}^2$$

场的能量不再由电荷造成, 而是由场自己的场强去定义, 以后还会进一步证明, 场在一点处对电荷有力的作用, 将会使场的动量密度发生改变, 而对电荷有做功, 将会影响这一点处的能量密度.

## 18.4 电荷体系

讨论若干电荷体系对于之后问题的理解是大有帮助的. 这些包括:

### 18.4.1 电偶极子

我们都应该知道一个电荷体系如果带电, 那么尽管在离电荷体系很近处电场分布不均匀, 但离开分布电荷足够远处的场强却总是趋于一个点电荷的电场. 常见的情况发生在原子核中, 原子核由质子和中子组成, 质子带电而中子不带电. 原子核内电荷分布总是不均匀的, 但在考虑外界电子受力时总是将其产生的场视为库仑场. 这是因为电子典型半径: 波尔半径与原子核的尺寸比差了 5 个数量级. 在这个距离处电场完全可以看成是库仑场, 它与库仑场的偏离将作为一个可以忽略的小量而不去考虑.

但类似的, 我们将考虑到很多不带净电荷的电荷分布体系, 它们在近处产生一个不均匀的电场, 而在远处产生的电场

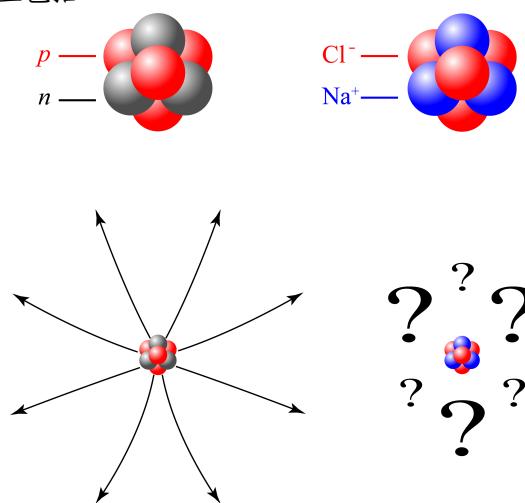


图 18.10: 远方的场

<sup>16</sup>而第二格林等式(Green's second identity) 在格林函数法解边值问题时将十分有用:

$$\nabla \cdot (\phi \nabla \psi - \psi \nabla \phi) = \phi \nabla^2 \psi - \psi \nabla^2 \phi$$

却总是存在着某些规律。这个问题有重要的实用价值。比如我们如果从 NaCl 晶体中取出一小块来，它的总电量总是为零，但由于晶体的变形等因素导致正负电荷产生的电场并不能完全抵消，这个电场将会与晶体的动力学耦合，极大地影响晶体的动力学特征。又比如很多分子实际上属于极性分子(polar molecule)，由于原子的电负性(electronegativity)差异，原子间的键合具有极性，在很多情形下一个原子对公用电子对的吸引能力更强，导致电子集中在一个原子一侧。从而负电荷的中心向这个原子位移。只要分子结构上的对称性不会使得键的极性不会相互抵消，整个分子就会显示出极性来，它是溶液中溶剂溶解机理的关键，也会影响物质的熔点、沸点、极化率等等关键的物理特性。

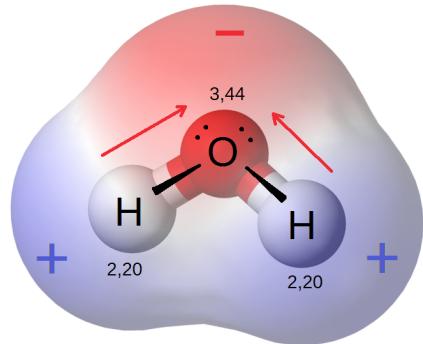


图 18.11: 水分子极性

所以我们关心总电荷量为零的体系在很远处产生的电场，理论上我们可以采取更完整的多极展开(multipole expansion)的做法，在很多情况下电偶极项(dipole term)成为了这种情形下的领头项(leading order)。我们定义小范围分布电荷体系中所有正电荷的“中心”和负电荷的“中心”：

$$\mathbf{r}_+ = \frac{\sum_{q_i > 0} q_i \mathbf{r}_i}{Q}, \quad Q = \sum_{q_i > 0} q_i$$

$$\mathbf{r}_- = \frac{\sum_{q_i < 0} q_i \mathbf{r}_i}{-Q}, \quad -Q = \sum_{q_i < 0} q_i$$

而体系的电偶极矩(dipole moment)就被定义为由负电荷指向正电荷的矢量：

$$\mathbf{d} = Q(\mathbf{r}_+ - \mathbf{r}_-)$$

或者写为：

$$\mathbf{d} = \sum_i q_i \mathbf{r}_i$$

求和遍及所有电荷，无论正负。由于总电荷量是零，不难验证这个矢量与坐标原点的选取无关，为了纪念物理学家德拜(P. Debye)在分子物理领域的相关开创性工作，它一般使用以德拜命名的单位，它是一个高斯制单位：

$$\begin{aligned} 1 \text{ D} &= 10^{-18} \text{ statC} \cdot \text{cm} \\ &= \frac{1}{299792458} \times 10^{-21} \text{ C} \cdot \text{m} \\ &= 3.34 \times 10^{-30} \text{ C} \cdot \text{m} \\ &= 0.208 \text{ e} \cdot \text{\AA} \end{aligned}$$

为何定义这样的一个矢量呢？我们考察在原点附近分布了一个电荷体系  $\{q_i, \mathbf{r}_i\}$ ，那么在很远处产生的电势应该为：

$$\varphi(\mathbf{r}) = \sum_i \frac{1}{4\pi\epsilon_0} \frac{q_i}{|\mathbf{r} - \mathbf{r}_i|}$$

而对于分母中的矢量的模，我们考虑到  $|\mathbf{r}| \gg |\mathbf{r}_i|$ ，可以做如下近似：

$$\begin{aligned} \frac{1}{|\mathbf{r} - \mathbf{r}_i|} &= [(\mathbf{r} - \mathbf{r}_i)^2]^{-\frac{1}{2}} = (\mathbf{r}^2 - 2\mathbf{r}_i \cdot \mathbf{r} + \mathbf{r}_i^2)^{-\frac{1}{2}} \\ &\approx (r^2 - 2\mathbf{r}_i \cdot \mathbf{r})^{-\frac{1}{2}} = \frac{1}{r} (1 - 2\frac{\mathbf{r}_i \cdot \mathbf{r}}{r^2})^{-\frac{1}{2}} \\ &\approx \frac{1}{r} (1 + \frac{\mathbf{r}_i \cdot \mathbf{r}}{r^2}) = \frac{1}{r} + \frac{\mathbf{r}}{r^3} \cdot \mathbf{r}_i \\ &= \frac{1}{r} + \frac{e_r}{r^2} \cdot \mathbf{r}_i \end{aligned}$$

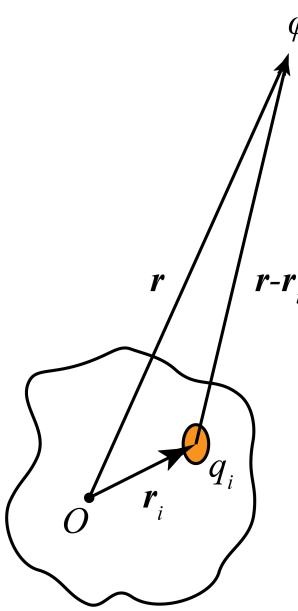


图 18.12: 偶极展开

代入求和表达式, 得:

$$\varphi(\mathbf{r}) \approx \frac{1}{4\pi\epsilon_0} \frac{\sum_i q_i}{r} + \frac{1}{4\pi\epsilon_0} \frac{\mathbf{e}_r \cdot \sum_i q_i \mathbf{r}_i}{r^2}$$

第一项求和由于总电荷量为零放弃了成为领头项的机会, 而第二项凸显了出来, 根据我们之前定义的电偶极矩, 它实际上就是:

$$\varphi(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \frac{\mathbf{e}_r \cdot \mathbf{d}}{r^2} \quad (r \gg d)$$

我们取了等号, 条件是我们考察的点的矢径  $r$  远大于电荷分布的平均尺寸  $d$ . 而理论上常常构造一种十分独特的体系, 我们让一对电量为  $+Q$  与  $-Q$  严格相等的点电荷相离  $\mathbf{r}_+ - \mathbf{r}_- = \mathbf{l}$ , 再令  $Q \rightarrow \infty$  而  $\mathbf{l} \rightarrow \mathbf{0}$ , 而保持  $\mathbf{p} = Q\mathbf{l}$  的值保持不变, 其极限就是抽象的模型: 点电偶极子(point dipole), 它在除原点外, 全空间产生的电势都可以用上式来表示. 这也是我们以后讨论的中心. 如果在点电偶所在空间建立球极坐标系, 而把电偶极矩矢量指向的方向作为极轴, 我们可以写出电势为:

$$\varphi(r, \theta) = \frac{1}{4\pi\epsilon_0} \frac{p \cos \theta}{r^2}$$

这是一个与半径平方反比的电势, 它与角度有关. 利用球坐标下梯度算符的表示:

$$\nabla = \mathbf{e}_r \frac{\partial}{\partial r} + \frac{1}{r} \frac{\partial}{\partial \theta} + \frac{1}{r \sin \theta} \frac{\partial}{\partial \phi}$$

我们计算出点电偶极子的电场来:

$$E_r(r, \theta) = \frac{1}{4\pi\epsilon_0} \frac{2p \cos \theta}{r^3}$$

$$E_\theta(r, \theta) = \frac{1}{4\pi\epsilon_0} \frac{p \sin \theta}{r^3}$$

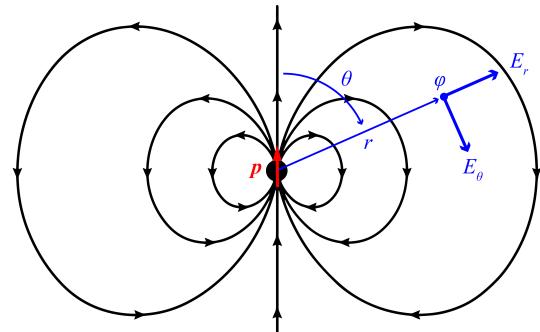


图 18.13: 点电偶极子的场

可见电场倒是三次反比了. 它比点电荷产生的场随距离减少地更快. 理论研究经常会把以上两个分量式组合为一个矢量式, 不难验证<sup>17</sup>

$$\mathbf{E} = \frac{1}{4\pi\epsilon_0} \frac{3(\mathbf{e}_r \cdot \mathbf{p})\mathbf{e}_r - \mathbf{p}}{r^3}$$

如果考虑电偶极子在外场中的情形, 采用两电荷分布极限为点电偶极子的结果, 我们很容易发现它的电势能为:

$$E_p = Q(\varphi_+ - \varphi_-) = Q\mathbf{l} \cdot \nabla \varphi = -\mathbf{p} \cdot \mathbf{E}$$

注意到根据电势能的含义, 这个能量是不包含构成电偶极子的两电荷之间的相互作用能的, 而在电偶极子在电场中发生位移的过程中我们也应该不能让电偶极子发生结构的改变: 即,  $\mathbf{p}$  的大小不能改变, 但  $\mathbf{p}$  的方向, 位置都可以发生改变, 随着这两个剩余自由度发生变化, 电场对电偶极子做的虚功  $-\delta E_p = \delta[\mathbf{p} \cdot \mathbf{E}(\mathbf{r})]$  就被理解为力矩与受力:

$$\delta \mathbf{r} = \mathbf{0}, \delta \mathbf{p} = \delta \boldsymbol{\theta} \times \mathbf{p}: -\delta E_p = \mathbf{M} \cdot \delta \boldsymbol{\theta} \Rightarrow \mathbf{M} = \mathbf{p} \times \mathbf{E}$$

$$\delta \mathbf{r} \neq \mathbf{0}, \delta \mathbf{p} = 0: -\delta E_p = \mathbf{F} \cdot \delta \mathbf{r} \Rightarrow \mathbf{F} = \mathbf{p} \cdot \nabla \mathbf{E}$$

其中表达式  $\mathbf{p} \cdot \nabla \mathbf{E}$  的含义为取沿  $\mathbf{p}$  方向的矢量场  $\mathbf{E}$  的方向导数. 以后会提到导体或电介质颗粒将会在外电场下感应或极化出一个电偶极子来:

$$\mathbf{p} = \alpha \mathbf{E}$$

<sup>17</sup>这个表达式对原点以外的点适用, 如果考虑原点的奇异电场特性, 并引入狄拉克德爾塔函数(Dirac delta function), 则更准确的表达式为:

$$\mathbf{E} = \frac{1}{4\pi\epsilon_0} \frac{3(\mathbf{e}_r \cdot \mathbf{p})\mathbf{e}_r - \mathbf{p}}{r^3} - \frac{\mathbf{p}}{3\epsilon_0} \delta^3(\mathbf{0})$$

把此式代入受力公式, 我们发现:

$$\mathbf{M} = 0 \quad ; \quad \mathbf{F} = \frac{\alpha}{2} \nabla E^2$$

这也就是我们常说的电场有能够吸引轻小物体的本领的解释, 我们发现这些小物体在外电场中的受力方向总是指向电场强度大小增加的方向的, 而与电场强度方向关系不大.

### 18.4.2 电荷密度

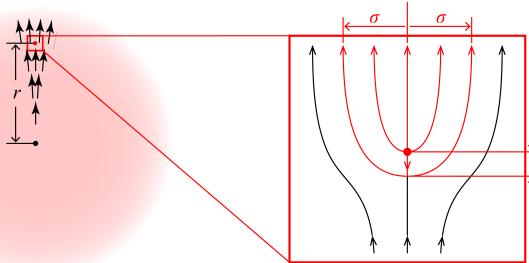


图 18.14: 连续分布点电荷产生平均电场

我们经常会用到连续分布的电荷这样的说法, 把电荷分布写成一个  $\rho(\mathbf{r})$  的形式, 表示每一个体积元  $dV$  中含有的电荷量为  $\rho dV$ . 这种做法在分析理论问题时发挥了很重要的作用, 但是它却并不实际. 实际电荷分布在很小的尺度看来永远是不连续的, 经典情况下一般可以视为由基本电荷量的点粒子产生的. 那么正如前述, 在这样的情况下电荷密度被理解为:

$$\rho = \sum_i n_i q_i$$

更难理解的此时电场强度的定义. 让我们考虑一个均匀带点的球, 此时球外的电场强度自然可以简单的按  $\frac{Q}{4\pi\varepsilon_0 r^2}$  来计算. 若内部的电场若按照完全的连续电荷分布来计算, 给出了:

$$\mathbf{E} = \frac{\rho \mathbf{r}}{3\varepsilon_0}$$

的结果. 但这个结果忽略了电荷的离散性, 我们如果认为这个球是由电荷数密度为  $n$ , 电荷量为  $q$  的大量点电荷构成. 那么设在半径为  $r$  处的电场与单个点电荷在距中心为  $r'$  处的电场相等. 我们可以得到:

$$\frac{nqr}{3\varepsilon_0} = \frac{q}{4\pi\varepsilon_0 r'^2}$$

如果一个电荷的平均占据体积视为球体, 半径为  $R$ , 那么电荷数密度与它的关系为:

$$n \cdot \frac{4}{3}\pi R^3 = 1$$

这样就得到:

$$\frac{r'}{R} = \sqrt{\frac{R}{r}}$$

或者用  $r$  区域内包含的电荷数量  $N = \frac{4}{3}\pi r^3 / \frac{4}{3}\pi R^3$ , 得到:

$$r' = \frac{R}{N^{1/6}}$$

还有一种看法是考察一个电荷发出的电场线在远方形成匀强电场以后的截面:

$$\frac{nqr}{3\varepsilon_0} \cdot \pi \sigma^2 = \frac{q}{\varepsilon_0}$$

我们很容易发现:

$$\sigma = 2r$$

也就是说,离开中心区域越远,包含的总电荷数量越多,该处点电荷产生的电场相对与由内部大量电荷积累而来的大的几乎均匀的电场相比就足够地小.点电荷使得外电场偏离均匀的空间范围就远小于电荷与电荷之间的间距,使得电荷之间的空间就可以视为一个几乎均匀的场.

最后我们总结,在用堆放点电荷的方式来得到一个电荷密度时,尽管微观来看产生的电磁场仍然很不均匀,但我们可以抽象出来每一点的平均场强的概念.它的意义有二:一是下一节即将证明的该点局部微元体内电场的平均值.二是如果保证电荷密度不变电荷无限分散的极限情形下该点处的实际场强.它的算法依然是:

$$\mathbf{E} = \int_{V'} \frac{1}{4\pi\epsilon_0} \frac{\rho e_R}{R^2} dV'$$

而既然电场算法与真实的连续分布的电荷没有区别.电势,局部电荷受力体密度就仍然可以这样计算:

$$\varphi = \int \frac{1}{4\pi\epsilon_0} \frac{\rho}{|\mathbf{R}|} dV' , \quad \mathbf{f} = \rho \mathbf{E}$$

### 18.4.3 极化强度

往往,平常状态下的物质的真实情况是,所有局部都不是由单一的点电荷单元堆积而成.就比如液体的水,它的重复单元是电中性的水分子.尽管电中性,对外表现出偶极子,在原场情况下近似为点电偶极子,便是其电势电场的领头项.故我们需要考虑的实际情况是:单位体积内有  $n(\mathbf{r})$  个偶极矩为  $\mathbf{p}(\mathbf{r})$  的点电偶极子.这里  $\mathbf{p}$  不会是常矢量,首先其方向往往随着空间位置逐渐在改变,而每个电偶极子的电偶极矩大小亦可能因为极化程度的不同而有所区别,下一章我们要介绍的位移极化就是这样的情形.我们将数密度  $n$  与电偶极矩  $\mathbf{p}$  相乘,得到单位体积内的电偶极矩,即**极化强度**(polarization density):

$$\mathbf{P} = n\mathbf{p}$$

对于连续情况,我们发现电势应该用电偶极子的电势进行体积分来得到:

$$\varphi = \int \frac{1}{4\pi\epsilon_0} \frac{\mathbf{P} \cdot \mathbf{e}_R}{R^2} dV' , \quad \mathbf{R} = \mathbf{r} - \mathbf{r}'$$

然而我们发现以下等式是成立的:

$$\begin{aligned} \nabla' &= e_x \frac{\partial}{\partial x'} + e_y \frac{\partial}{\partial y'} + e_z \frac{\partial}{\partial z'} \\ \nabla' \cdot \frac{\mathbf{P}}{|\mathbf{R}|} &= \frac{\nabla' \cdot \mathbf{P}}{|\mathbf{R}|} + \mathbf{P} \cdot \nabla' \frac{1}{|\mathbf{R}|} \\ \nabla \frac{1}{|\mathbf{R}|} &= \nabla \frac{1}{|\mathbf{r} - \mathbf{r}'|} = -\frac{\mathbf{r} - \mathbf{r}'}{|\mathbf{r} - \mathbf{r}'|^3} = -\frac{\mathbf{e}_R}{R^2} , \quad \nabla' \frac{1}{|\mathbf{R}|} = \nabla' \frac{1}{|\mathbf{r} - \mathbf{r}'|} = -\frac{\mathbf{r}' - \mathbf{r}}{|\mathbf{r}' - \mathbf{r}|^3} = \frac{\mathbf{e}_R}{R^2} \\ \Rightarrow \quad \frac{\mathbf{P} \cdot \mathbf{e}_R}{R^2} &= \nabla' \cdot \frac{\mathbf{P}}{|\mathbf{R}|} - \frac{\nabla' \cdot \mathbf{P}}{|\mathbf{R}|} \end{aligned}$$

这样我们就得到在区域  $V'$  内存在极化矢量  $\mathbf{P}$  时在全空间产生的电势的公式:

$$\begin{aligned} \varphi &= \int_{V'} \frac{1}{4\pi\epsilon_0} (\nabla' \cdot \frac{\mathbf{P}}{|\mathbf{R}|} - \frac{\nabla' \cdot \mathbf{P}}{|\mathbf{R}|}) dV' \\ &= \oint_{\partial V'} \frac{1}{4\pi\epsilon_0} \frac{\mathbf{P} \cdot dA'}{|\mathbf{R}|} + \int_{V'} \frac{1}{4\pi\epsilon_0} \frac{-\nabla' \cdot \mathbf{P} dV'}{|\mathbf{R}|} \\ &= \oint_{\partial V'} \frac{1}{4\pi\epsilon_0} \frac{\sigma_P dA'}{|\mathbf{R}|} + \int_{V'} \frac{1}{4\pi\epsilon_0} \frac{\rho_P dV'}{|\mathbf{R}|} \end{aligned}$$

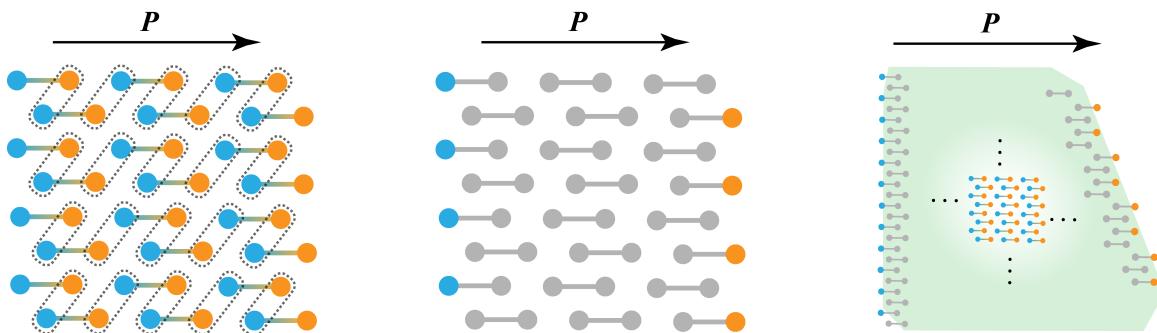


图 18.15: 极化强度导致表面电荷分布

从以上表达式看得出来,最后电势有一种等效的求法,就是认为在体内有一个  $\rho_P = -\nabla' \cdot \mathbf{P}$  的电荷密度分布,而认为在体表面的面元  $dA$  上带有面电荷  $dQ = \mathbf{P} \cdot d\mathbf{A}$ ,或者认为有面电荷密度  $\sigma_P = \mathbf{P} \cdot \mathbf{n}$ ,其中  $\mathbf{n}$  为体积表面的外法向量.与其说这是一种等效,不如说这只是对实际情况的另一种理解.如上图所示,这是一块均匀极化的介质,我们这样理解介质表面上会出现的电荷:对介质内部的电荷进行合适的分组使得一对一对的正负电荷产生的场互相抵消.这样便只剩下了表面的电荷能产生场了.在几何学上不难证明,如果电偶极子长  $l$ ,两电荷为  $q$ ,体密度为  $n$ ,那么在与极化矢量平行的方向看两端未配对电荷的面密度恰好为  $\sigma = nql = np = P$ .再考虑面可能的倾斜便容易发现以上极化电荷面密度  $\sigma_P$  的公式.类似地,体电荷密度  $\rho_P$  对应于沿某方向的  $\mathbf{P}$  分量在变化的情形.具体的物理图像也请读者加以分析与思考.

我们有把握预期空间各处的电场强度的公式,也可以通过上述方式等效,毕竟我们可以通过电势计算电场:

$$\mathbf{E} = \oint_{\partial V'} \frac{1}{4\pi\epsilon_0} \frac{\sigma_P dA'}{|\mathbf{R}|^2} \mathbf{e}_R + \int_{V'} \frac{1}{4\pi\epsilon_0} \frac{\rho_P dV'}{|\mathbf{R}|^2} \mathbf{e}_R$$

然而,若要开始考虑电偶极子在场中受到的外力,那就势必要考虑连续分布的电荷体系产生的电场和微观非均匀分布的电荷体系的局部场之间的区别.我们会惊人地发现以上电场只能作为平均意义上的**宏观场**(macroscopic field) 而区别于电偶极子具体感受到的**局部场**(local field).首先让我们来计算一个概念,一个球体内的平均场强

# 第十九章 导体与介质

## 19.1 导体与静电平衡

### 19.1.1 导体特点

微观地看,物质由原子或分子等组成,其中导电现象通常发生在不同情况下:

1. 真空导电:一般不会说真空具有导电性(conductive),因为真空中是没有载流子(charge carrier)的.的确,在阴极射线管(cathode ray tube)中,加热一个阴极,并辅以合适的偏置电压可以造成真空中的电流.或是用一束能量足够的光子去轰击金属表面造成电子逸出,甚至纯粹由于阴极表面尖处十分强的电场导致电子直接克服逸出功发射出来.三种现象分别称为热发射(thermionic emission),光电效应(photoelectric effect)与场发射(field emission).这种电荷的定向移动现象被统一地称为输运现象(transport phenomenon).由于真空输运的独特性质,比如电子不会受到散射,平均自由程远大于仪器尺寸,与凝聚态物理中的一些概念对应,这被称为弹道输运(ballistic transport).
2. 绝缘体漏电:大多数非金属晶体,或是不含离子的液体与气体,原子核在晶体中都限制在点阵格子的特定位置,在气体,液体中则没有特定位置.电子分为两类,,一类是原子的内层电子,它极为稳定地存在于原子周围,离子晶体中的几乎完全被阴离子夺取的电子也属于这种情况.它们与原子核一起构成原子实(atomic core).而价层电子(valence shell electron),它们用来成键,将原子连接形成分子或者晶体,一般被定域在原子间的特定区域.所有这些电子的特点都是束缚态(bond state).它们不能在介质中自由传导,一个电子的区域到另一个电子的区域间存在势垒(potential barrier),经典物理认为电子的动能不足以穿过这些势垒.但是,量子理论则认为电子的波函数可以通过隧道效应(tunnelling)以小概率在不同区域间转移.这就为有电场的情况下电荷的平均定向移动提供了可能.这种现象称为漏电(leakage).
3. 绝缘体击穿:在十分高的电场下,原子与原子间的典型电压将大于阻碍电子转移的势垒,或是分子中显不同电性的部分之间.此时电子在本质上可以视为在

## 19.2 电像法

## 19.3 电介质

## 19.4 再议静电能



# 第二十章 稳恒电流

## 20.1 稳恒电流描述与形成

### 20.1.1 德鲁德模型

电荷的定向移动形成电流(current). 就好像一缸气体在慢慢挪动那样, 电荷的定向移动并不是纯粹的匀速运动, 而是与无规则的热运动相叠加. 1900 年前后德鲁德 (P. Drude) 和洛伦兹 (H. A. Lorentz) 等人提出德鲁德模型(Drude model) 来解释金属中的导电现象. 主要观点是金属内部自由运动的电子类似于理想气体那样做自由的运动, 称为自由电子气(free electron gas). 我们用电子电量  $-e$ , 质量  $m$ , 电子数密度  $n$ , 和弛豫时间(relaxation time) $\tau$ , 平均速度  $v$  来表示其特征. 弛豫时间就是电子做匀速直线运动, 与原子实两次碰撞之间的平均间隔时间. 与之相关的另一个量还可以是平均自由程(mean free path) $\lambda$ . 容易想像, 典型的情形是, 常温下电子的平均速度是和气体分子的平均热运动速度那样, 一个非常巨大的速度, 而金属原子之间的距离又是那么地短, 导致电子发生十分频繁的碰撞. 而如果在金属中加一个电场  $E$ , 它在两次碰撞内是只可以让电子速度改变一个十分微小的量的:

$$\mathbf{a} = -\frac{e\mathbf{E}}{m}$$

$$\mathbf{v}(0) \rightarrow \mathbf{v}(t) = \mathbf{v}(0) + \mathbf{a}t \rightarrow \mathbf{v}(\tau) = \mathbf{v}(0) + \mathbf{a}\tau$$

我们根据统计力学的思想, 计算在某一时刻电子速度对所有电子取热平衡分布的平均:

$$\mathbf{u} = \langle \mathbf{v} \rangle = \langle \mathbf{v}(0) + \mathbf{a}t \rangle = \mathbf{a}\tau$$

上式  $t$  代表距离上次碰撞每个电子幸存的时间.  $\mathbf{v}(0)$  代表上次碰撞后其速度. 一方面, 认为碰撞使得电子速度完全随机分布, 平均的结果为零. 另一方面, 认为电子的碰撞是一个泊松过程, 其碰后幸存时间的概率分布是一个指数分布:

$$p(t) = e^{-\frac{t}{\tau}}$$

从而平均的  $t$  为

$$\langle t \rangle = - \int_0^\infty t dp = pt|_0^\infty + \int_0^\infty e^{-\frac{t}{\tau}} dt = \tau$$

最后结合电流密度为:

$$\mathbf{j} = -nev$$

我们得到:

$$\mathbf{j} = \frac{ne^2\tau}{m}\mathbf{E}$$

的结果. 称为微观欧姆定律(microscopic Ohm's law), 其中系数被称为电导率(electrical conductivity):

$$\sigma = \frac{ne^2\tau}{m} ; \quad \mathbf{j} = \sigma\mathbf{E}$$

十分值得指出的是,作为德鲁德模型的另一个重要结论,金属的热导率(thermal conductivity)在理论中也可以给出一个估计值.我们都应该知道金属比绝大多数其他固体都拥有好得多的导热性能.这可以用热导率与傅里叶热传导定律(Fourier's law of thermal conduction)来描述:

$$\mathbf{q} = -\kappa \nabla T$$

其中  $\mathbf{q}$  为热流密度,  $\kappa$  即为材料的热导率.水常温下热导率只有  $0.591\text{W}/(\text{m}\cdot\text{K})$ ,但纯的铜却能够达到  $401\text{W}/(\text{m}\cdot\text{K})$ .金属的高热导率全都得益于轻盈的电子气,它能够迅速地把局部的热运动加剧传导开来.类比理想气体的非平衡态统计方法,我们给出:

$$\kappa = \frac{1}{3}nv\lambda c$$

式中  $c$  为每个电子的动能与温度的比.利用  $\lambda = v\tau$ , 我们把热导率和电导率做比,便可以把较难确定的散射弛豫时间消去,得到:

$$\frac{\kappa}{\sigma} = \frac{mv^2c}{3e^2}$$

德鲁德模型认为,作为类似于理想气体的电子气,理应有:

$$c = \frac{3}{2}k \quad ; \quad \frac{1}{2}mv^2 = \frac{3}{2}kT$$

从而得到:

$$\frac{\kappa}{\sigma T} = \frac{3}{2}\left(\frac{k}{e}\right)^2$$

历史上金属热导率和电导率的比值与温度相关的现象很早就被人们发现,一般温度升高时电导率会有十分明显的下降,也就是金属的电阻会上升.小灯泡在未发光时与正常发光时电阻就经常有 2 倍左右的差距.而 1853 年魏德曼 (G. Wiedemann) 和弗朗茨 (R. Franz) 观察到不同金属虽然导电导热性能差距悬殊,室温下两者之比却接近一个常数,称为魏德曼-弗朗茨定律(Wiedemann-Franz's law).而洛伦茨 (L. Lorenz)<sup>1</sup> 在 1872 则把这一经验规律确定到常数与绝对温度成正比的形式.最后德鲁德电子论将这个常数以微观常数的形式确定下来,等式右边现在约为  $1.11 \times 10^8 \text{W}\cdot\Omega/\text{K}^2$ , 这与实验结果数量级是一致的,但却差了约 2 倍.真实的值被称为洛伦兹常数(Lorentz number):

$$\frac{\kappa}{\sigma T} = L = \frac{\pi^2}{3}\left(\frac{k}{e}\right)^2 = 2.44 \times 10^8 \text{W}\cdot\Omega/\text{K}^2$$

这个理论很成功,但最后的结果的不符合让人感到疑惑.还有一个令人感到疑惑的问题.就是对于电子气的电容.既然电子是完全独立于金属原子实的另一个热力学体系.按照经典统计物理.原子实的摩尔热容与离子晶体等类似,大概在  $3R$  附近<sup>2</sup>.而电子按只有平动自由度的理想气体考虑应贡献  $1.5R$  热容.实验却否定了这一点,金属常温下热容仍然是在  $3R$  附近.电子仿佛没有被热激发,然而电子又势必参与导热,因为金属的导热性能明显优于其他物质.那么以上推导过程中哪儿存在不合理之处?几年后到来的量子力学革命让人们认识到这一个经典模型从一开始就做了不止一个的与微观实际情况不符的假定,但又是十分巧合地,最终结果与真实值数量级自动一样了.

### 20.1.2 费米气观点 \*

元素周期律引发了人们对原子核外电子排布规律的研究,人们惊奇地发现电子是费米子(fermion),符合泡利不相容原理(Pauli exclusion principle).这赋予电子独特的量子特性.具体来说,单位体积内如果电子数目越多,那么其最低平均能量就必须越大.因为能量最低的状态一旦被占据,其他电子就必须占据能量更高的状态,即使按照最低能量的方式去堆积(绝对零度时的行为),电子也将具有很高的平均能量.它符合类

<sup>1</sup>注意,丹麦物理学家路德维希·洛伦茨 (Ludwig Lorenz)(1829-1891) 与荷兰物理学家亨德里克·洛伦兹 (Hendrik Lorentz)(1853-1928) 是两个不同的人.有一个方程以它们两人的名字共同命名:Lorenz-Lorentz 关系.

<sup>2</sup>即督龙-裴替定律(Dulong-Petit Law).

似位置-动量不确定性原理的反比率, 电子浓度的减小了单电子占据的位置尺度, 则它的动量就会增加, 从而根据色散关系  $p^2 = 2mE$  其能量也会升高. 这一点使得我们去修改经典的德鲁德模型. 相应的量子气体称为费米气(Fermi gas).

定量计算费米气的特性需要考虑电子的波动本性. 我们暂时取金属为长宽高为  $ABC$  的长方体, 那么如果将  $N$  个电子倒入这个容器, 电子的数密度为:

$$n = \frac{N}{ABC}$$

让我们考虑一下电子对状态的填充, 长方体相当于一个谐振腔, 事实上给出了电子动量状态的量子化:

$$p_x A = ah, p_y B = bh, p_z C = ch$$

$$a, b, c \in \mathbb{N}$$

这是因为根据德布罗意关系  $p = \hbar k$  而电子状态需要满足  $kL = 2m\pi$  的缘故. 从而我们发现电子的三个方向的动量都是量子化的. 量子化的单位为:

$$\Delta p_x = \frac{\hbar}{A}, \Delta p_y = \frac{\hbar}{B}, \Delta p_z = \frac{\hbar}{C}$$

现在把  $N$  个电子倒入动量空间中, 那么电子在低温下将从最低能量态开始填充, 由于电子数目巨大, 最后电子将填充到能量为  $\varepsilon_F$  处. 这样的一个填满电子的动量空间中的球体称为费米海(fermi sea), 最终填充到的能量称为费米能级  $\varepsilon_F$ , 对应的电子动量为费米动量  $p_F$ . 注意到一个动量空间中的一个标志状态的坐标内部实际上有两个独立的状态, 它们表示即使这两个波的波矢  $k$  一样, 它们代表电子的自旋也不一样<sup>3</sup>. 从而我们写出:

$$2 \cdot \frac{4\pi p_F^3}{3} / (\Delta p_x \Delta p_y \Delta p_z) = N$$

从而得到:

$$p_F = \left( \frac{3Nh^3}{8\pi N} \right)^{\frac{1}{3}} = \left( \frac{3nh^3}{8\pi} \right)^{\frac{1}{3}}$$

在以上过程中总电子数和总体积被消去了, 最后电子堆积到的动量值仅仅取决于各处电子的数密度而成为了强度量. 这样一个由泡利不相容原理所造成的动量值对应到速度上, 对一般金属估计约为  $10^6 \text{ m/s}$ . 而如果按非费米气计算, 热运动速度应为  $\sqrt{kT/m} \sim 7 \times 10^4 \text{ m/s}$ , 我们发现经典结果是严重估计少了的. 但电子的热容又估计多了. 量子统计给出:

$$c = \frac{\pi^2}{2} k \cdot \frac{kT}{\varepsilon_F}$$

代入热导率公式, 得:

$$\frac{\kappa}{\sigma} = \frac{\pi^2}{6} \cdot \frac{mv_F^2}{\varepsilon_F} \cdot \left( \frac{k}{e} \right)^2 \cdot T$$

恰好,  $\varepsilon_F = \frac{1}{2}mv_F^2$ , 从而我们得到了正确形式的魏德曼-弗朗茨定律:

$$\frac{\kappa}{\sigma T} = \frac{\pi^2}{3} \left( \frac{k}{e} \right)^2$$

我们最后做一个经典德鲁德模型与量子费米气模型的比较. 两个理论都承认以下基本公式的成立:

$$\sigma = \frac{ne^2\tau}{m}, \quad \kappa = \frac{1}{3}nv^2\tau c$$

<sup>3</sup>电子的微观描述是旋量波, 也就是二分量波函数, 具有两种可能的自旋.

先看电导率, 经典理论与量子理论对各个参数的估计除了  $\tau$  其他都是一致的. 而经典更倾向于对  $\tau$  的估计更小. 因为经典理论一般认为  $\tau = \lambda/v$ , 而  $\lambda$  一般就按照原子实之间的平均距离来估计. 这一点之后就知道是不妥当的, 因为电子实际上可以在严格周期性的势能场中毫无散射的传播下去. 而使得电子能够被散射的其实是晶格的缺陷, 热振动等因素, 从而一般温度升高,  $\tau$  减小, 金属的导电性能就大大减弱. 这给出了金属电阻随温度升高而升高的结论. 这里经典理论对  $\lambda$  的过低估计恰好被对  $v$  的过低估计所抵消掉一部分, 从而最后  $\tau$  的值在常温下差距也不大. 而对于热导率, 经典理论对  $v$  的过低估计又恰好被对  $c$  的过高估计修正, 精确到了仅仅相差一个常数, 除了经典理论说不清楚的  $\tau$ , 经典与量子的在数量级上是基本符合的.

值得一提, 热导率描述导热, 电导率描述导电, 而导电现象的附效应便是热产生. 也就是焦耳热(Joule heating). 微观的功率密度

$$\mathbf{j} = \sigma \mathbf{E} \quad , \quad w = \mathbf{j} \cdot \mathbf{E}$$

也即:

$$w = \sigma E^2 = \rho j^2$$

其中 **电阻率**(resistivity) $\rho$  为电导率的倒数. 上为**微观焦耳-楞次定律**(microscopic Joule-Lenz law).

### 20.1.3 能带论 \*

费米气模型在量子力学迅猛发展的几十年后回头看又是过于 naive 了. 人们发现电子终究是在晶格间传播的波动, 如果原子实与背景的电子给出了严格的周期性的势能, 那么如果电子的总机械能小于最大势能, 那么就成为了束缚态, 限制在一个原子附近运动而成为原子实的一部分. 而如果电子守恒机械能大于最大势能, 就成为不受任何阻碍的传导态. 动能虽在运动中有所波动, 但绝不会越来越小. 描述电子的波函数称为**布洛赫波函数**(Bloch wave function). 而电子如何去填充怎样的能级? 布洛赫(F. Bloch)研究并发现了**能带论**(energy band theory) 来替代旧的费米气理论, 这个理论可以统一地描述导体与绝缘体.

能带论可以看成是电子在单原子外的束缚态特征与传导的自由电子气的特征的一种综合, 一方面电子在这样的一个周期性势场中的运动其守恒的机械能并不是可以取所有值, 而是分为可以连续取值的区间: **能带**(band) 与根本取不到的能量区间: **带隙**(gap) 所构成. 而电子作为波动用其所在的能带与波矢  $\mathbf{k}$  来描述. 其波矢方向的物理意义不再重要, 因为每一个态一般都代表某种驻波与行波的混合, 而每一个态附近的群速度则可以代表这个态代表的真实电子运动速度. 另一方面这些能带的结构又恰好对应到了单原子的束缚态能级. 这使得我们经常是一望某元素的单原子核外电子排布, 便能从其是否具有最外层电子读出这种元素单质是否能导电的信息.

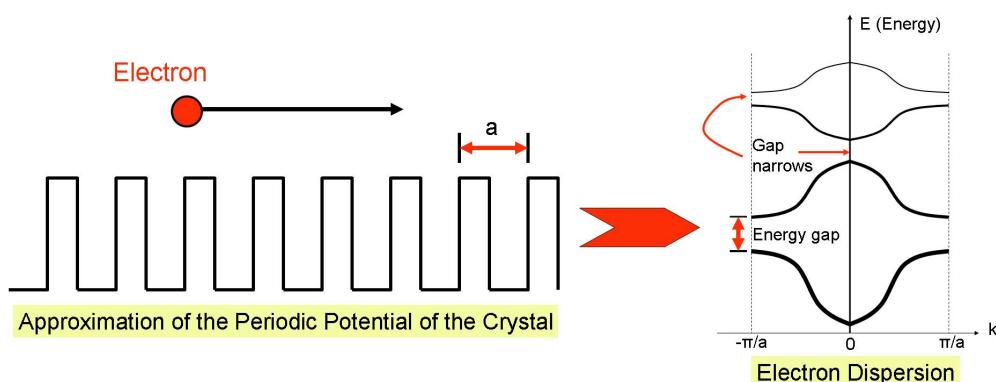


图 20.2: 电子在周期性势场中传播

某种原子形成晶体时, 电子从最低的能带开始填充, 内层电子填满了较低的那些能带, 它们不参与导电, 对应于原子实电子和在晶体共价键键合的价电子. 最外层电子较少的原子, 形成晶体时最后填充的那个能带没有填满, 或者虽然填满但与下一个能带在能量区间上重合而能够导电. 称为**导体**(conductor). 对应

的元素也就显示出明显的金属性<sup>4</sup>. 而若原子外层电子数目多, 则很有可能电子恰好填满一个能带后隔着一个能隙与空的上方能带相望, 则所有态都稳定不变, 材料所有电子都是束缚电子, 最后填充的那个满带称为**价带**(valence band), 代表原子间成共价键的价电子所在的能带. 而上方的那个空带称为**导带**(conduction gap). 意为一旦这个带中存在电子则会大大改善其导电性能. 此时元素体现出非金属性, 一般形成的都是**绝缘体**(insulator). 而在金属元素与非金属元素分界处则存在很多的**类金属**(metalloid), 它们形成的晶体虽然与绝缘体一样, 在最低能量时恰好把价带排满, 而隔着一个带隙与空的导带相望. 但带隙很小 (小于4eV), 导致常温的热激发和掺杂, 抑或是光照激发都将在导带中激发大量载流子而具有可观的导电性能. 这些材料称为**半导体**(semiconductor).

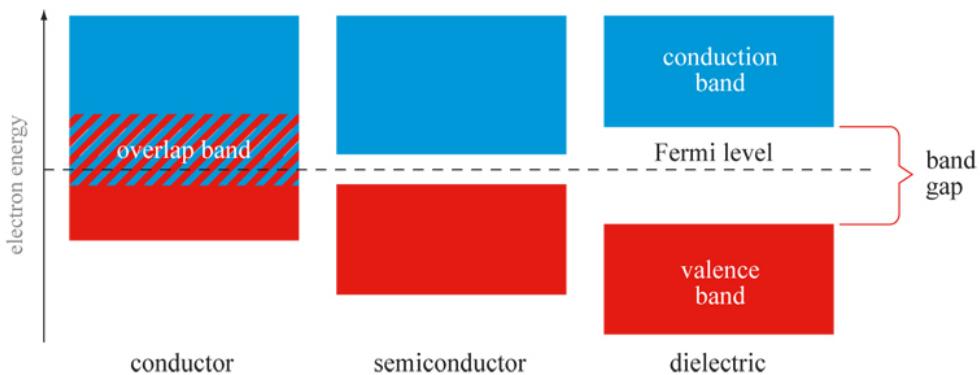


图 20.3: 能带与导电性

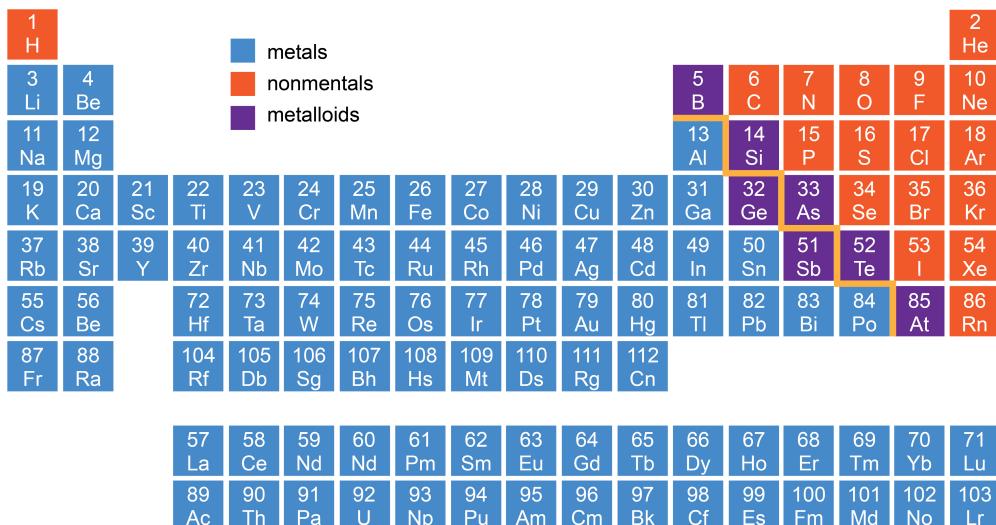


图 20.4: 金属, 类金属与非金属元素

那么能带论如何给出电导率的计算公式? 仍然是:

$$\sigma = \frac{ne^2\tau}{m}$$

与上一个费米气模型区别在于一点, 就是这里的  $m$  不再是电子的**裸质量**(bare mass), 而是由于与周期性势场相互作用, 电子的色散关系(能量与波矢关系)被修正后给出的**有效质量**(effective mass). 即使对于金属, 这个质量与真实电子质量相差好几倍的现象也是十分普遍的.

<sup>4</sup>思考: 氢是否具有金属性? 参考[https://en.wikipedia.org/wiki/Metallic\\_hydrogen](https://en.wikipedia.org/wiki/Metallic_hydrogen).

### 20.1.4 惯性, 阻尼与回复力

在量子力学还未诞生的时期, 为了解释导体的导电, 绝缘体的介电现象, 并适用于任意交变电磁场在介质中的传播, 色散, 吸收与散射. 洛伦兹提出著名的洛伦兹模型(Lorentz model) 作为德鲁德模型的补充. 在这儿电子的惯性被重视, 电子与晶格的碰撞被简化, 可能的原子实对电荷的束缚被简化为线性回复你. 也就是洛伦兹用唯象的谐振子类比来解释电子在外场下的行为:

$$m\ddot{\mathbf{r}} + \gamma\dot{\mathbf{r}} + k\mathbf{r} = -e\mathbf{E}e^{i\omega t}$$

这个模型的解和它对应的各种特性我们将在光学教材里详细讲解. 在这里我们仅仅讨论导体中的自由电子, 故  $k = 0$ . 与导体导电相关的两个因素为:

一是阻尼系数  $\gamma$ . 在直流电场下平衡时稳定的电子速度为:

$$\dot{\mathbf{r}} = -\frac{e}{\gamma}\mathbf{E}$$

对比更加本质的德鲁德等导电模型的电导公式, 容易发现这个唯象系数与电导率和基本参量关系为:

$$\gamma = \frac{ne^2}{\sigma} = \frac{m}{\tau}$$

于是我们可以写出在导体上突然加一个不随时间变化的匀强电场后电子的运动的微分方程:

$$m\ddot{\mathbf{r}} + \frac{m}{\tau}\dot{\mathbf{r}} = -e\mathbf{E}$$

它的解为:

$$\dot{\mathbf{r}} = -\frac{e\mathbf{E}\tau}{m}(1 - e^{-\frac{t}{\tau}})$$

我们十分自然地发现定义电子碰撞的弛豫时间恰好与导体对电场响应的弛豫时间相吻合. 这一点稍微需要一些讨论和修正. 见下.

第二点我们来讨论表示电子惯性的质量  $m$ . 正是它导致了衰减因子  $e^{-\gamma t/m}$ , 从而造成电路对电场响应的弛豫. 但如果讨论电路的弛豫, 或者说电流的惯性, 有一点是被我们过低地估计了, 便是自感现象. 电荷的加速运动形成变化的电流, 从而与之相伴的是变化的磁场. 这个变化的磁场反过来给电流一个反向的作用力. 由于该力正比于加速度, 故等价于增加了载流子的质量. 这里的弛豫时间对应直流暂态电路中的  $\tau' = L/R$ . 如果一定要理解为电子的惯性质量, 那么一般比电子的裸质量要大好几个数量级  $m^* \gg m$ . 从而一般有电路弛豫时间  $\tau = m^*/\gamma$  不等于微观碰撞弛豫时间.

纯粹的导体是否有回复力项? 显然在以上电子运动方程中不含这一项. 但注意到如果有回复力项, 即使外场为零电子也会做振动. 真实情况会发生这样的现象吗? 答案是会, 即等离子体振荡(plasma oscillation). 原因是电子相对于原子实的位移实际上在金属的表面累积了电荷分布, 从而在内部激发了电场. 由于频率很高, 一般要到紫外波段的频率, 我们可以完全忽略阻尼的影响. 当电子有位移  $\mathbf{r}$  时将形成极化强度<sup>5</sup>:

$$\mathbf{P} = -ner$$

我们考虑块状金属板<sup>6</sup>的集体等离子体振荡, 那么由于金属表面的累积电荷在金属内部形成的电场为:

$$\mathbf{E} = \mathbf{E}_{in} = -\frac{\mathbf{P}}{\varepsilon_0} = \frac{ner}{\varepsilon_0}$$

代入原方程:

$$m^*\ddot{\mathbf{r}} = -e\mathbf{E} = -\frac{ne^2\mathbf{r}}{\varepsilon_0}$$

<sup>5</sup>显然此时“极化强度”描述的非极化电荷, 而是自由电荷.

<sup>6</sup>的确, 取不同形状的物体这个频率会有所不同.

从而得到谐振子方程. 振荡频率为:

$$\omega_p = \sqrt{\frac{ne^2}{\varepsilon_0 m^*}} ; \quad \sigma = \varepsilon_0 \omega_p^2 \tau$$

最后我们考虑低频电磁波 (内部实际场强) 下金属的电子行为, 利用光学中讨论过的洛伦兹模型, 我们计算金属的复电容率, 最后得到:

$$\varepsilon_r(\omega) = 1 - \frac{\omega_p^2 \tau^2}{1 + \omega^2 \tau^2} ; \quad \sigma(\omega) = \frac{\sigma}{1 + \omega^2 \tau^2}$$

一般  $\omega_p \tau \gg 1$ , 故低频下金属的介电常数实际上是绝对值很大的负数. 电导率随着频率的增加而减小.

### 20.1.5 稳恒电流与形成条件

作为电荷的流动, 电流密度与电荷密度间满足电荷守恒定律:

$$\nabla \cdot \mathbf{j} + \frac{\partial \rho}{\partial t} = 0$$

而所谓**稳恒电流**(steady current) 指的是一定区域内导电物质形成的电荷, 电场, 电流分布. 其中电流的分布不能随时间变化, 自然的结论是电荷密度分布也不能随时间变化. 否则将产生随时间变化的电场, 电场造成电流的变化. 即上式必有:

$$\frac{\partial \rho}{\partial t} = 0 ; \quad \nabla \cdot \mathbf{j} = 0$$

即在稳恒电流中电流密度是一定没有散度的, 无头无尾的闭合曲线. 稳恒电流必须是**直流**(direct current) 而非**交流**(alternating current).

容易发现, 在稳恒电流中微观欧姆定律写成以下形式是不完整的:

$$\mathbf{j} = \sigma \mathbf{E}$$

在很多情况下, 上式并没有问题.  $\mathbf{E}$  被理解为由电荷分布  $\rho$  形成的静电场<sup>7</sup>, 它是一定可以引入电势的:

$$\mathbf{E} = -\nabla \varphi$$

那么如果承认以上形式的微观欧姆定律就会引起矛盾: 电流线是环状闭合的, 电场线必须与电流线同向平行, 但电场线又不能是闭合的, 否则与静电场环路定理, 电势的可定义性矛盾. 实际上, 微观形式的欧姆定律必须被写成:

$$\mathbf{j} = \sigma(\mathbf{E} + \mathbf{K})$$

上式  $\mathbf{K}$  代表**非静电力场**(non-electrostatic field) 对电子产生的作用. 定义方式与电场量纲一致, 都是单位正电荷的受力. 这么写的不同之处就在于, 把驱动电流的力分成了两部分: 静电场  $\mathbf{E}$  是保守的, 可以引入势的, 在一个回路中不做功的, 不吞吐能量的场 (注意电流本身就会发热, 这不是场造成的而是电荷定向流动碰撞晶格造成的). 但非静电力就是源源不断做功驱动回路中电流流动的, 非保守的, 不能引入势来描述的场. 虽然不能引入电势来描述非保守力, 但可以用**电动势**(electromotive force) 来描述非静电力:

$$\mathcal{E}_\gamma = \int_{A \xrightarrow{\gamma} B} \mathbf{K} \cdot d\mathbf{r} ; \quad \mathcal{E}_\Gamma = \oint \mathbf{K} \cdot d\mathbf{r}$$

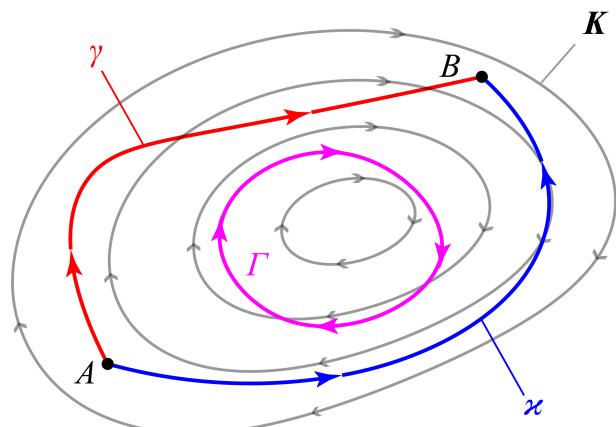


图 20.5: 非静电力场

<sup>7</sup>实际上, 不包括涡旋电场, 因为它被归于之后引入的非静电力. 换句话说, 这里的  $\mathbf{E}$  不再由静止电荷的受力  $\mathbf{F}/q$  去定义, 而是单纯根据静止电荷分布产生平方反比的电场叠加去定义.

不像对静电场的积分那样仅仅取决于端点, 电动势与积分路径一一对应. 回路电动势一般来说不为零, 从而相同端点的两个路径电动势也不一定相等:

$$\mathcal{E}_\Gamma \neq 0 ; \quad \mathcal{E}_\gamma \neq \mathcal{E}_x$$

非静电力既然能够作用在电子上, 其形式一般就是电磁力, 万有引力, 或是惯性力. 磁场力对应**动生电动势**(motional emf), 而由于变化的磁场激发的涡旋电场部分的电场力对应**感生电动势**(transformer emf). 电子会受到引力, 1967 年 Schiff 指出电子会因为引力的原因聚集到金属的底部. 而后 Dessler 提出了严格的考虑晶格在重力场下压缩与不均匀情况下的电子平衡情况. 引力是很弱的保守力, 对金属内部电荷体系的影响虽然很小但绝不是不可观测. 而与引力类似的惯性力则可以人工制造出很大的数值, 而且可以使它非保守. 试想加速旋转一个线圈, 那么在随晶体加速的参考系里电子受到的角向惯性力就和受到一个电场力造成的效果没有本质的区别.

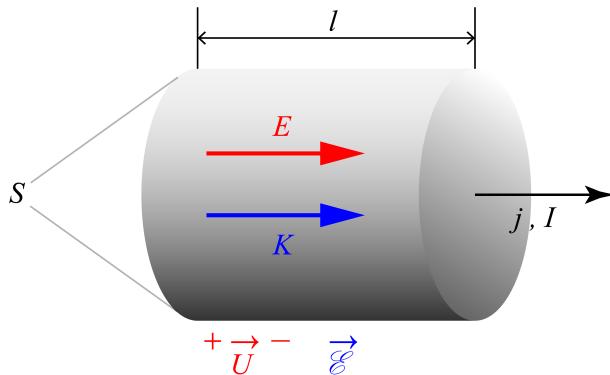


图 20.6: 宏观欧姆定律

还有一些电动势一般无法用非静电力描述. 因为在一个回路中某些部分的确有能量的输入, 但不是以作用在某些具体的电荷上的力的形式. 而是更加广义的一些力作用在一些流上. 例如温差电偶利用两种物质的电子的热扩散性质差异, 从高温端吸取热量而驱动电流. 化学反应更是源源不断地产生新的物质, 利用反应的自发性而驱动物质做定向的移动形成电流. 在这些抽象的场合, 我们采用某段电路中移动单位电荷, 外界向体系注入的能量来定义电动势:

$$\mathcal{E} = \frac{dW}{dQ}$$

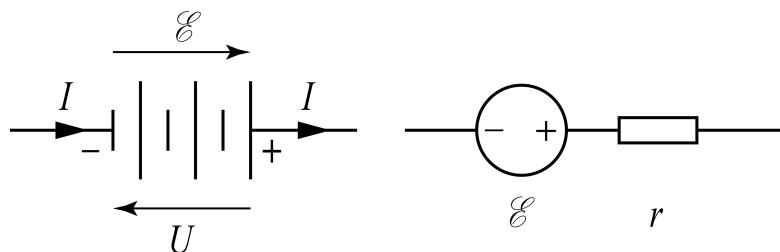


图 20.7: 电池的欧姆定律

最后, 从微观到宏观, 我们考虑一段长为  $l$ , 面积为  $S$ , 电阻率为  $\rho$  的导体上的欧姆定律:

$$E + K = \rho j$$

等式两边同时乘以  $l$ , 把电流密度写成  $I/S$ . 得到:

$$U + \mathcal{E} = IR ; \quad R = \rho \frac{l}{S}$$

一般叫做部分电路的欧姆定律. 而如果考虑**电池**(battery) 的两端, 正常工作时一般正极电压更高, 故约定电动势与电压取相反的方向, 电流则与电动势方向一致, 却与电压方向反向:

$$U = \mathcal{E} - Ir$$

为一段含源电路的欧姆定律. 此时我们把电池等效为理想电压源和一个定值电阻  $r$  的串联. 电阻即为电池的**内阻**(internal resistance).

## 20.2 电路与电路方程

具体到直流电路中, 我们一般提供以下理想的电路元件:

### 20.3 电路分析基础

### 20.4 电路分析方法

### 20.5 半导体



# 第二十一章 静磁场

**21.1 电流与磁场**

**21.2 两个定理与矢势**

**21.3 磁偶极子**

**21.4 磁介质与磁能**

**21.5 超导简介**



## 第二十二章 电磁感应

**22.1 动生电动势**

**22.2 感生电动势**

**22.3 自感与互感**

**22.4 法拉第电磁感应定律**



# 第二十三章 麦克斯韦方程组

**23.1 麦克斯韦方程组**

**23.2 平面电磁波**

**23.3 电磁场能量与动量**

**23.4 电磁波辐射**



## 第二十四章 交流电路

**24.1 代数与几何表示**

**24.2 常见电路**

**24.3 变压器**

**24.4 电能传输**



## 第二十五章 光的干涉

25.1 标量波理论

标量波是指以下标量偏微分波动方程(wave equation) 的解:

$$\left( \nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right) A(\mathbf{r}, t) = 0$$

分别在直角坐标, 柱坐标, 球极坐标下用所谓的分离变量法可以得到一些正确的基础解<sup>1</sup>:

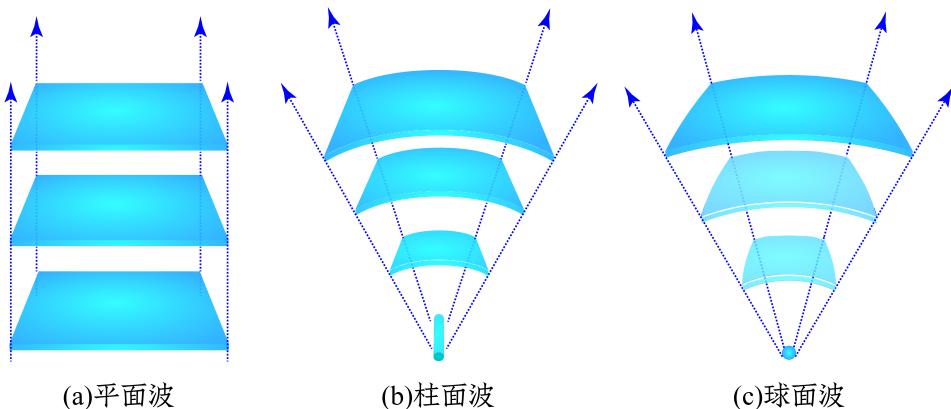


图 25.1: 三种典型的波

$$1. \ A(\mathbf{r}, t) = A \cos(\omega t - \mathbf{k} \cdot \mathbf{r} + \phi)$$

要求  $\frac{\omega}{|\mathbf{k}|} = c$ . 此为传播方向为  $\mathbf{k}$  的平面行波.

$$2. \ A(\mathbf{r}, t) = \frac{I}{\sqrt{\rho}} \cos(\omega t - k\rho + \phi)$$

要求  $\frac{\omega}{|k|} = c$ .  $k > 0$  时为向外传播的柱面行波,  $k$  小于零则向内传播.

$$3. \ A(\mathbf{r}, t) = \frac{I}{r} \cos(\omega t - kr + \phi)$$

要求  $\frac{\omega}{|k|} = c$ .  $k > 0$  时为向外传播的球面行波,  $k$  小于零则向内传播.

容易发现,以上波全都具有这样的形式:

$$A(\mathbf{r}, t) = |A|(\mathbf{r}) \cos(\omega t - \varphi(\mathbf{r}))$$

也就是说,每个点的相位都在以共同的  $\omega$  做等速率的增加,代表一种同步的振动.而振动幅度  $|A| > 0$  是要由于波的传播特点和能量流动的连续性而缓慢变化的.具体传播方向则看相位的梯度:

$$\boldsymbol{k} = +\nabla \varphi$$

<sup>1</sup>以下三式都要求  $\omega > 0$ , 数学上看  $\omega < 0$  也没错, 但是利用三角函数的性质可以将  $t$  系数变号而不带来新的解.

在近似的意義下,  $\frac{\omega}{|\mathbf{k}|} = c$  是沒有問題的. 而  $c$  便是波傳播的速度. 在真空與介質(寫作  $v$ ) 中出現在光的波动方程中的  $\frac{1}{c^2}$  中的  $c$  其實是:

$$c = \frac{1}{\sqrt{\epsilon_0 \mu_0}} , \quad v = \frac{c}{n}, \quad n = \sqrt{\epsilon_r \mu_r} \approx \sqrt{\epsilon_r}$$

約等於  $c$  是因為考慮到了透明光学介質一般磁性質不顯著故  $\mu_r \approx 1$ .

我們不是很推薦用三角函數來計算波动光学. 要寫成複指數的形式, 而實際三角函數理解為它的有效的實部, 虛部是為了方便計算而添加的沒有物理效應的. 而且波动光学統一約定, 利用三角函數的形式, 為相位增添一個負號, 使得光在同一刻沿傳播方向相位是增加的, 改為:

$$A(\mathbf{r}, t) = |A|(\mathbf{r}) e^{i(\varphi(\mathbf{r}) - \omega t)} = |A|(\mathbf{r}) e^{i\varphi(\mathbf{r})} e^{-i\omega t} = A(\mathbf{r}) e^{-i\omega t}$$

在以後情形下我們都將把這種符號約定稱為“光學符號約定”, 而以往的指數上的宗量隨時間增加隨傳播方向減小的符號約定稱為“電磁學符號約定”. 以上討論就引出了複振幅  $A(\mathbf{r})$  的概念. 而之前找到的各點振幅  $|A|$  和  $\varphi$  現在就被整合在了一起, 作為了複振幅的模與幅角:

$$|A|(\mathbf{r}) = |A(\mathbf{r})| , \quad \varphi(\mathbf{r}) = \arg A(\mathbf{r})$$

回過頭來再看, 我們在解波动方程時, 實際上從一开始就不需要認為該標量  $A(\mathbf{r}, t)$  是一個實數, 在複數域內解, 也能够直接得到完全與上述等效一樣的結果. 最後我們發現, 該複數標量場  $A(\mathbf{r}, t)$  如果能夠符合波动方程, 其實部與虛部也是能够分別滿足波动方程的:

$$\left( \nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right) A(\mathbf{r}, t) = 0 \Rightarrow \begin{cases} \left( \nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right) \Re A(\mathbf{r}, t) = 0 \\ \left( \nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right) \Im A(\mathbf{r}, t) = 0 \end{cases}$$

如果再要求這樣的  $A(\mathbf{r}, t)$  具有形式  $A(\mathbf{r}, t) = A(\mathbf{r}) e^{-i\omega t}$ , 數學上也就是對變量  $t$  分離變量, 物理就意味著研究單色光的傳播, 那麼代入便會發現實際上  $A(\mathbf{r})$  需要滿足的對空間的方程為**亥姆霍茲方程**(helmholtz equation):

$$(\nabla^2 + k^2) A(\mathbf{r}) = 0 , \quad k = \frac{\omega}{c}$$

這就是整個波动光学理論的核心方程. 在繼續闡述之前我們先說說  $A$  指什麼. 其實真空與線性介質(不導電)電磁場几乎每個量(標勢, 矢勢, 電場, 磁場)都是符合波动方程的:

$$A : \{\varphi, A_x, A_y, A_z, E_x, E_y, E_z, B_x, B_y, B_z\}$$

$$\left( \nabla^2 - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right) A(\mathbf{r}, t) = 0 , \quad (\nabla^2 + k^2) A(\mathbf{r}) = 0$$

所以作為複數的標量波理論就構成了光干涉, 衍射等波动光学課題的數學基礎. 而且我們率先注意到其中的兩點:

一是, 波動方程, 亥姆霍茲方程是線性方程, 也就是說, 如果兩個光場都符合方程, 那麼兩個場直接逐點做標量加法得到的場也會自動符合方程. 這也就是之後干涉, 衍射操作的重要前提. 如果由於兩個原因同時產生了兩束光波, 那麼只要寫出各自的場, 加在一起便是整個體系現在的光場. 這就是雙光束干涉時我們將會採取的做法.

二是, 平面波與球面波的重要性. 只要是行波, 在任意一點就可以按照平面波來近似. 如果取其  $\mathbf{k}$  方向為  $z$  方向, 那麼局部的光場為:

$$A(\mathbf{r}) = A e^{ikz}$$

但是旁边的点  $(x_1, y_1)$  的最佳平面波的近似却不一定平行于  $z$  轴, 它应该为:

$$A(\mathbf{r}) = A e^{i[k_z z + k_x(x - x_1) + k_y(y - y_1)]}$$

这下便有第一个近似条件了, 我们叫做**傍轴条件**(paraxial condition). 它要求光线传播方向与  $z$  轴夹一个小角度. 此时我们习惯把  $x$  和  $y$  方向方向余弦  $\cos \alpha_x, \cos \alpha_y$  使用的约 90 度的角度  $\alpha_x, \alpha_y$  的余角记做方向角, 也就是说:

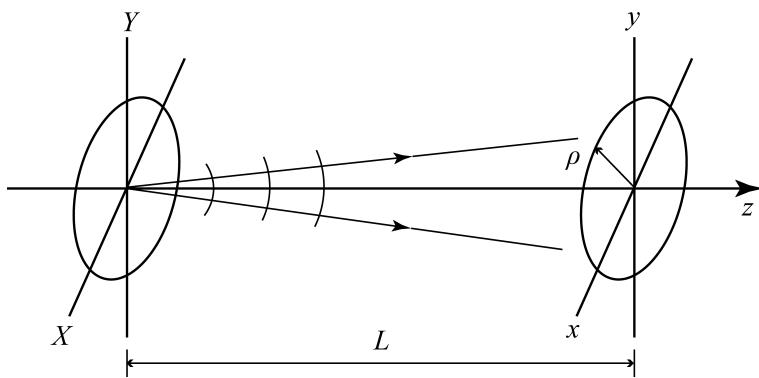


图 25.2: 球面波向平面波近似

$$\sin \theta_x \approx \theta_x = \frac{k_x}{k}, \quad \sin \theta_y \approx \theta_y = \frac{k_y}{k}, \quad k_z \approx \sqrt{1 - \theta_x^2 - \theta_y^2} k \approx k$$

在傍轴近似下, 以上以方向角  $(\theta_x, \theta_y)$  传播的平面波被方便地写为:

$$A(\mathbf{r}) = A e^{ikz} e^{i\phi} e^{ik(\theta_x x + \theta_y y)} = A' e^{ik(\theta_x x + \theta_y y)}$$

经常的, 我们只需要关心在像屏  $z = 0$  上的场, 参考上图. 所以上式我们可以把不变的系数都统一写作  $A'$ .

但是平面波却不总是好的近似, 我们已经发现了, 像上图那样, 如果在物屏  $z = -L$  上原点  $X = 0, Y = 0$  处有一点光源发出球面波, 那么光屏上  $(0, 0)$  处和  $(x_1, y_1)$  处的近似公式就采取了不同的形式. 所以我们会寻求更好的球面波近似. 但且慢, 是不是任意情况下光场中的任意一点附近的场都能近似为一个球面波呢? 答案显然是否定的. 比如柱面波很显然就不能被近似为球面波, 因为其**波前**(wavefront), 即等相位面沿垂直于传播方向的一个方向弯曲而另一个方向却是平直的. 事实上球面波的光源是点, 柱面波的光源是线. 而介于柱面波与球面波中间的某种不对称波前就连光源都没法找到. 但是球面波的地位仍然是重要的. 这体现在两点上. 一是在真实的光路系统中, 点光源是切实存在且常用的. 此时物方到像方的点到点的消除了像差的理想成像是我们努力追求的方向. 故在这样的系统中光波的确一直都适合被近似为球面波, 而平面波也被当做一种光源在无穷远的特殊状态而处理. 二是, 若把波前上各个点当成是新的次波源向前发射出球面波而在前方相干叠加形成新的光场, 后面在衍射这一章我们数学上可以证明这就是计算光的传播的正确方式. 其中用到的球面波作为一种本质上重要的物理对象, 即**传播子**(propagator), 有着重要的物理地位.

故我们研究从物屏  $z = -L$  的  $(X, Y)$  处传播到像屏  $z = 0$  处  $(x, y)$  点的球面波的场:

$$\begin{aligned} A(x, y) &= \frac{I}{\sqrt{L^2 + (x - X)^2 + (y - Y)^2}} e^{ik\sqrt{L^2 + (x - X)^2 + (y - Y)^2}} \\ &= \frac{I}{L} e^{ikL} \left[ 1 + \frac{(x - X)^2}{L^2} + \frac{(y - Y)^2}{L^2} \right]^{-\frac{1}{2}} e^{ikL \left[ \left[ 1 + \frac{(x - X)^2}{L^2} + \frac{(y - Y)^2}{L^2} \right]^{\frac{1}{2}} - 1 \right]} \end{aligned}$$

由于傍轴条件, 我们发现方向角  $(u_x, u_y)$ ,  $u_x = \frac{x}{L}, u_y = \frac{y}{L}$  和方向角  $(U_x, U_y)$ ,  $U_x = \frac{X}{L}, U_y = \frac{Y}{L}$  都是小量, 而传播到光屏上该点的光的方向角( $k$  方向)为  $(u_x - U_x, u_y - U_y)$ . 仅考虑领头项, 把不随  $x, y$  变化的系数记为常数, 我们发现:

$$A(x, y) = A e^{\frac{1}{2}kL[(u_x - U_x)^2 + (u_y - U_y)^2]} = A' e^{ik\frac{x^2 + y^2}{2L}} e^{-ik\frac{Xx + Yy}{L}}$$

也就是说光屏上的波由两个相因子决定. 第一个是平方相关于  $x, y$  的  $\varphi_1 = k \frac{x^2 + y^2}{2L} = \pi \frac{x^2 + y^2}{\lambda L}$ , 第二个恰好是平面波的相因子  $\varphi_2 = -k \frac{Xx + Yy}{L} = -k(U_x x + U_y y) = k(\theta_x x + \theta_y y)$ . 就是说当点光源相对  $(x, y) = (0, 0)$  从  $X < 0, Y < 0$  的方向照过来能够在原点附近获得方向角为正的平面波场.

那么球面波能够被近似为球面波的条件也就不言自明了, 它要求第一个相因子的相位很小不干扰第二个相因子, 也就是  $\varphi_1 \ll \pi$ , 也就是远场条件(farfield condition):

$$\rho^2 \ll \lambda L \Leftrightarrow L \gg \frac{\rho^2}{\lambda}$$

其实换一个角度理解这个条件, 如果远场条件被满足, 但是在这样的  $\rho$  内第二个相因子变化也非常小时, 整个  $A$  几乎就是常数, 这时候也形成不了有价值的光场, 所以我们如果要求  $\varphi_2 \gg \varphi_1$ , 还能发现第三个条件  $X, Y \gg x, y$ . 对于这个条件我们做这样的理解, 相位的绝对大小是没有观测效果的. 我们需要的其实是光束干涉时相位的差值. 故这其实是要求物屏的花样尺寸要远大于像屏上干涉花样的观测范围.

在远场条件不被满足的情况下, 考虑第一个相因子对第二个相因子的影响, 会造成干涉图样的扭曲与形变, 几何光学上会造成最基础的像差: 相散和散焦.

在介绍下两节具体的干涉之前, 对干涉的原理做一个简要的介绍是有必要的. 根据之前的介绍, 空间中如果同时存在两个光场:

$$A_1(\mathbf{r}) = |A_1|(\mathbf{r})e^{i\varphi_1(\mathbf{r})} ; A_2(\mathbf{r}) = |A_2|(\mathbf{r})e^{i\varphi_2(\mathbf{r})}$$

我们可以找到其单独存在时的光强, 一般来说, 光强为位置的缓变函数:

$$I_1(\mathbf{r}) = |A_1|^2 = A_1^* A_1 ; I_2(\mathbf{r}) = |A_2|^2 = A_2^* A_2$$

然而光场的叠加是有两种典型的方式的. 最常见的其实是非相干叠加(incoherent superposition). 此时出于下面要介绍的原因, 光强是可以逐点直接相加的:

$$I(\mathbf{r}) = I_1(\mathbf{r}) + I_2(\mathbf{r})$$

将自然界或人造的两束来源不同的光照在一次是无法观察到干涉花样的. 光的干涉条件比机械波干涉条件来的苛刻的多. 在某些细心制备的实验室条件下才能观察到干涉花样, 此时光场是相干叠加(coherent superposition). 在光学之中相干度是十分值得关注的. 我们将在之后的一节中集中讨论. 读者可能自然地认为相干叠加时为复振幅相加:

$$A(\mathbf{r}) = A_1(\mathbf{r}) + A_2(\mathbf{r})$$

但是其实无论非相干叠加还是相干叠加上式都适用, 那么是什么造成了两者的区别呢? 我们把上式模方求光强:

$$\begin{aligned} I(\mathbf{r}) &= A^* A \\ &= (A_1^* + A_2^*)(A_1 + A_2) \\ &= A_1^* A_1 + A_2^* A_2 + A_1^* A_2 + A_2^* A_1 \\ &= I_1 + I_2 + |A_1||A_2|e^{i(\varphi_2 - \varphi_1)} + |A_1||A_2|e^{i(\varphi_1 - \varphi_2)} \\ &= I_1 + I_2 + 2\sqrt{I_1 I_2} \cos(\varphi_2 - \varphi_1) \end{aligned}$$

上式中出现的交叉项  $2\sqrt{I_1 I_2} \cos(\varphi_2 - \varphi_1)$  就叫干涉项(interference term). 它是关于位置的快变函数.乍一看这也未免太快了, 因为只要当空间位置改变一个波长, 相位  $\varphi$  就会有  $\pi$  的量级的改变, 这样干涉项在这个过程中就会有正负号的改变. 从而产生干涉花样. 在实际的干涉实验中, 合理的仪器设置可以使得这个空间特征长度被放大到人眼或助视仪器可以直接观测的程度. 从而这就是干涉实验的基础.

那么为什么对于非相干叠加情形我们直接去掉了干涉项. 那是因为在之前我们做把光场向标量复振幅简化的过程中去掉了公共的单色光时间演化项  $e^{i\omega t}$ . 这并不是永远合理的做法. 真实光场由于后面要介绍的各种原因, 可以认为相位随时间的演化并不是完全线性增加的而是有随时间偏离期望值的方差越来越大的涨落. 从而两束并没有关联的自然光之间不同时刻同一点  $\mathbf{r}$  的相位差  $\delta = \varphi_2(\mathbf{r}) - \varphi_1(\mathbf{r})$  是一个在随机涨落的数, 只要涨落够大,  $\cos(\varphi_2 - \varphi_1)$  项的随时间平均就会总是趋近于零. 从而干涉现象也就消失了.

在以下的两小节中, 我们先不用关心相干性的问题, 认为不同光场的叠加总是有“完美的相干性”: 完全稳定的相位差. 我们会在不同干涉装置的介绍中去强调为什么总是能保证“完美的相干性”的成立.

## 25.2 分波面干涉

以下几种干涉装置一般统一归为分波面干涉(interference by dividing wave-front). 因为它们实现的核心思想都是让本来沿各个方向独立传播的波前的部分改变传播方式而在空间中产生交叠.

### 25.2.1 杨氏双缝干涉仪

历史上第一个挑战牛顿光的粒子说权威, 构想出更合理的光的波动说的假设并通过实验证明自己的想法的是谁? 那必然要归功于托马斯·杨(Thomas Young)<sup>2</sup>.

## 25.3 分振幅干涉

## 25.4 偏振干涉

## 25.5 相干性

## 25.6 多光束干涉

<sup>2</sup>杨氏为英国博物学家, 也经常作为物理学家而介绍. 除了双缝实验, 也因为杨氏模量, 杨-亥姆霍兹三色视觉论, 杨-拉普拉斯和杨-杜普蕾附加压强与接触角公式而闻名. 他还是杰出的语言学家, 比较了 400 种语言的词汇与语法. 并解密了一部分古埃及象形文字. 在音乐, 医学和神学等上也有贡献.



## 第二十六章 光的衍射

**26.1 光栅与波带片**

**26.2 布拉格衍射**

**26.3 衍射积分公式**

**26.4 波前分析法**



# 第二十七章 物理光学

## 27.1 经典色散理论

蓝天,白云,红太阳.在简单的自然现象中蕴含着光在传播过程中的另外一种典型的现象: **散射**(scattering).而散射却又是一个多么复杂的现象!要完全理解散射,我们得研究光的吸收与发射,因为光与物质中分子的相互作用其实本质上可以看成吸收与发射,这是**分子光学**(molecular optics)的研究范畴.其次,光在介质中的衰减(吸收),不同颜色的光折射率的不同(色散)其实都导致了一类散射现象(瑞利散射).但是还有不同的散射现象则表现出不用的特性,需要从更宏观或更微观的角度去建立新的理论来解释他们.

不同于以往在透明介质中的平面波传播,新的现象需要新的平面波模型.

### 27.1.1 复波矢与复折射率

我们在波动方程的解中引入复数是为了更好地描述振动形式的解.但这却忽视了另一种解的存在性.就是如果波矢  $\mathbf{k}$  本身亦为复数,由于同时又是一个矢量,可以表示为  $\mathbf{k} = \alpha + i\beta$ .代入场的空间部分:

$$A(\mathbf{r}) = A_0 e^{i\mathbf{k} \cdot \mathbf{r}} = A_0 e^{-\beta \cdot \mathbf{r}} e^{i\alpha \cdot \mathbf{r}}$$

这就代表了一个沿  $\alpha$  方向传播,沿  $\beta$  方向衰减的平面波.而  $\alpha$  为波矢的实部,  $\beta$  为波矢的虚部:

$$\alpha = \Re(k) ; \beta = \Im(k)$$

这种波可不可以真空中传播?出人意料的是这居然是可能的.因为如果代入真空中的亥姆霍兹方程:

$$\nabla^2 A + \frac{\omega^2}{c^2} A = 0$$

便会觉得这只需要要求:

$$\mathbf{k}^2 = (\alpha + i\beta) \cdot (\alpha + i\beta) = \frac{\omega^2}{c^2} \Rightarrow \alpha^2 - \beta^2 = \frac{\omega^2}{c^2}, \alpha \cdot \beta = 0$$

这样的波我们并不陌生,早在几何光学我们介绍光的全反射时便研究过另一侧界面中的隐失波场便是以上形式.更典型地,在  $a \times b$  的矩形波导中传播的波满足:

$$(\frac{2\pi}{\lambda_x})^2 + (\frac{2\pi}{\lambda_y})^2 + (\frac{2\pi}{\lambda_z})^2 = \frac{\omega^2}{c^2}$$
$$\lambda_x = \frac{2b}{m}, \lambda_y = \frac{2a}{n}, m, n = 1, 2 \dots$$

此时在  $x, y$  方向的振动是驻波,  $z$  方向传播的波则是行波. $x, y$  方向要形成驻波是因为在边界的导体上要符合相同的边界条件,所以或都是波腹,或都是波节.亦可以用同一个

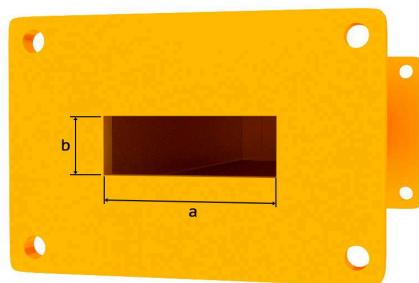


图 27.1: 矩形波导

平面波在侧面上的多次反射的叠加来理解这种兼具驻波和行波特质的场. 我们发现模式为  $(m, n)$  的波, 会有一个截止角频率  $\omega_c$ , 如果角频率低于此值使得波无法向前传播:

$$\omega_c = c \sqrt{\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2}$$

那如果在波导的起点的天线给出信号角频率  $\omega < \omega_c$  会发生什么情况? 此时就是  $z$  方向波矢  $k_z$  变为虚数, 意味着波沿  $z$  方向不是传播而是衰减:

$$\mathbf{k} = \boldsymbol{\alpha} + i\boldsymbol{\beta}$$

$$\boldsymbol{\alpha} = \left( \frac{m\pi}{a}, \frac{n\pi}{b}, 0 \right) , \quad \boldsymbol{\beta} = \left( 0, 0, \sqrt{\left(\frac{m\pi}{a}\right)^2 + \left(\frac{n\pi}{b}\right)^2 - \frac{\omega^2}{c^2}} \right)$$

在真空中尚可以产生衰减的波<sup>1</sup>, 更何况在各种介质中. 事实上, 描述光在介质中随着传播的距离而衰减的现象早有一经验公式, 即**比尔-朗博-布葛律**(Beer-Lambert-Bouguer law):

$$\frac{I(l)}{I(0)} = e^{-(\sum_i n_i \sigma_i)l}$$

其中  $I(0)$  为在传播  $l$  前的光强,  $I(l)$  为在传播  $l$  后的光强.  $n_i$  是构成介质的第  $i$  种分子的数密度,  $\sigma_i$  为描述其吸收本领的**衰减截面**(attenuation cross section). 例如对大气来说主要的衰减因素就来自气溶胶, 干净的海水则能提供 200m 左右的透光带, 如果有的话, 浮游植物含有的光合色素是造成衰减的主要原因. 而对于强衰减性质的介质, 我们可以唯象地认为波在介质中的传播为:

$$|A(l)| = |A(0)|e^{-\beta l} \Rightarrow I(l) = I(0)e^{-2\beta l}$$

这样的选择是有道理的. 下以导电导致的损耗为例: 在不漏电介质中, 麦克斯韦关系本应为:

$$\left\{ \begin{array}{l} \nabla \cdot \mathbf{E} = 0 \\ \nabla \times \mathbf{E} + \frac{\partial \mathbf{H}}{\partial t} = \mathbf{0} \\ \nabla \cdot \mathbf{H} = 0 \\ \nabla \times \mathbf{H} - \varepsilon \mu \frac{\partial \mathbf{E}}{\partial t} = \mathbf{0} \end{array} \right.$$

这样便有:

$$v = \frac{c}{n} = \frac{1}{\sqrt{\varepsilon \mu}} , \quad k = \frac{\omega}{v} = \omega \sqrt{\varepsilon \mu}$$

但是, 如果介质漏电, 且符合  $\mathbf{J} = \sigma \mathbf{E}$ , 那么以上方程就改写为:

$$\left\{ \begin{array}{l} \nabla \cdot \mathbf{E} = 0 \\ \nabla \times \mathbf{E} + \frac{\partial \mathbf{H}}{\partial t} = \mathbf{0} \\ \nabla \cdot \mathbf{H} = 0 \\ \nabla \times \mathbf{H} - \varepsilon \mu \frac{\partial \mathbf{E}}{\partial t} = \mu_0 \sigma \mathbf{E} \end{array} \right.$$

<sup>1</sup>事实上, 只需要亚波长的结构便可造成衰减.

对最后一个式子所引发的不同在光学情况下是很好处理的. 因为单色光总是具有固定的频率  $\omega$ , 那么其实以上式子无非是把  $\frac{\partial}{\partial t}$  变成  $-i\omega$ . 在磁导率被认为几乎等于真空磁导率的情形下, 这相当于说:

$$\left\{ \begin{array}{l} \nabla \cdot \mathbf{E} = 0 \\ \nabla \times \mathbf{E} - i\omega \mathbf{H} = \mathbf{0} \\ \nabla \cdot \mathbf{H} = 0 \\ \nabla \times \mathbf{H} + i\omega \varepsilon \mu \mathbf{E} = \mathbf{0} \end{array} \right. \xrightarrow{\varepsilon \rightarrow \varepsilon + i\frac{\sigma}{\omega}} \left\{ \begin{array}{l} \nabla \cdot \mathbf{E} = 0 \\ \nabla \times \mathbf{E} - i\omega \mathbf{H} = \mathbf{0} \\ \nabla \cdot \mathbf{H} = 0 \\ \nabla \times \mathbf{H} + i\omega \varepsilon \mu \mathbf{E} - i\sigma \mu \mathbf{E} = \mathbf{0} \end{array} \right.$$

所以对于漏电介质<sup>2</sup>, 通常会有复电容率(complex permittivity) 的说法<sup>3</sup>, 它就是以上把电容率和电导率合并以后的新的复常数, 用它第四个方程就与真空中的方程没有任何区别了, 除了系数是一个复数:

$$\varepsilon' = \varepsilon + i\frac{\sigma}{\omega} : \quad \nabla \times \mathbf{H} + i\omega \varepsilon' \mu \mathbf{E} = \mathbf{0}$$

这样一个方程的解是可以完全照搬之前的解的, 因为数学上可以证明复数解具有可解析延拓(analytic continuation) 的特性. 从而容易发现, 复波矢就变为:

$$\mathbf{k}^2 = \omega^2 \varepsilon' \mu$$

这样就得到:

$$\alpha^2 - \beta^2 = \omega^2 \varepsilon \mu, \quad 2\alpha \cdot \beta = 2\alpha \beta \cos \theta = \omega \sigma \mu > 0$$

此时传播波矢  $\alpha$  与衰减波矢  $\beta$  就不一定要垂直了, 它们必须夹锐角, 也就是说, 如果电磁波在漏电介质或者导体中传播, 沿传播方向必须要衰减. 我们最后引入光学中使用最多的复折射率(complex refraction index), 按照原来的看法它意味着  $k^2 = n^2 k_0^2 = n^2 \frac{\omega^2}{c^2} = \frac{\omega^2}{v^2}$ . 现在要更小心些, 因为  $\mathbf{k} = \alpha + i\beta$  已经包含两个不贡献的波矢部分. 故我们先对以下表达式开方:

$$\alpha^2 - \beta^2 + 2i\alpha\beta \cos \theta = \omega^2 \varepsilon' \mu$$

$$\Rightarrow \sqrt{\alpha^2 - \beta^2 + 2i\alpha\beta \cos \theta} = n \frac{\omega}{c} = (n_1 + i n_2) \frac{\omega}{c}$$

通过以上两式对比, 我们能得到两个方面. 第一是复折射率实部  $n_1$  与虚部  $n_2$  分别是这样依赖于相对介电常数  $\varepsilon_r = \varepsilon/\varepsilon_0$  和电导率  $\sigma$  的:

$$n_1 = \sqrt{\varepsilon_r} \cdot \sqrt{\frac{1}{2} \left( 1 + \frac{\sigma^2}{\omega^2 \varepsilon^2} + \sqrt{1 + \frac{\sigma^2}{\omega^2 \varepsilon^2}} \right)}$$

$$n_2 = \frac{\sigma}{\sqrt{\varepsilon_r} \omega \varepsilon_0} \cdot \sqrt{\frac{1}{2} \left( 1 + \frac{\sigma^2}{\omega^2 \varepsilon^2} + \sqrt{1 + \frac{\sigma^2}{\omega^2 \varepsilon^2}} \right)^{-1}}$$

第二组关系式如果已知了介质的两个折射率, 那么在介质中传播的波的两个波矢  $\alpha$  和  $\beta$  需要满足的关系为:

$$\alpha^2 - \beta^2 = (n_1^2 - n_2^2) \frac{\omega^2}{c^2}$$

$$\alpha \beta \cos \theta = n_1 n_2 \frac{\omega^2}{c^2}$$

只有当  $\theta = 0$  才恰有:

$$\alpha = n_1 \frac{\omega}{c}, \quad \beta = n_2 \frac{\omega}{c}$$

<sup>2</sup>即使不漏电, 也会由于有损耗而等效于有复电容率.

<sup>3</sup>对于普通的导体, 也一样可以讨论复电容率, 或者更常见地, 复电导率(complex conductivity). 它不仅包含由于原子实部分极化导致的电容性, 还要包含由于电子运动惯性导致的电感性.

### 27.1.2 经典电子论的解释

下面我们来介绍历史上发挥了重要作用的经典电子论。它虽然不够精确但物理图像十分重要。

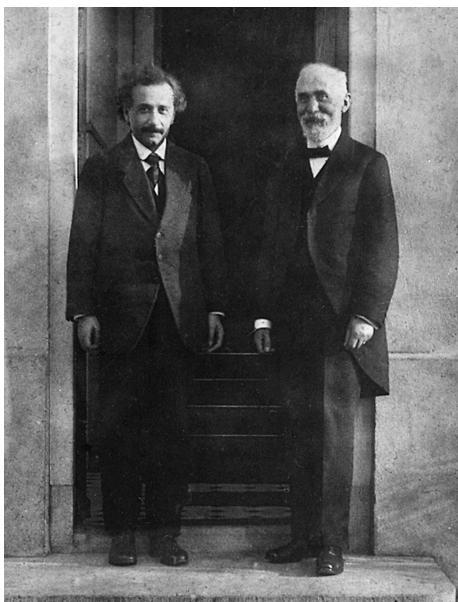


图 27.2: Einstein 与 Lorentz

历史上虽然电子作为粒子的发现是 1897 年汤姆孙 (J. J. Thomson) 的工作。但是早在半个世纪前人们就开始逐渐相信物质由两种带电粒子构成并在导电时可以移动的模型了。人们用这样的模型去想象输运现象, 电磁感应, 热电耦合等现象并取得了丰硕的成果, 经典电磁学早在 1860 年代便在麦克斯韦 (J. C. Maxwell) 的工作中达到顶点进入尾声。剩下来的几十年时间历史让位给了洛伦兹 (H. A. Lorentz), 洛伦兹对电磁学的贡献是深远的。他承前人之大成, 讨论电磁场与电磁介质的相互作用, 讨论电磁学原理与古老的伽利略式相对性原理的结合; 又开近代物理之先河, 电子论自然地导向了量子理论, 而电磁学又自然地促成了狭义相对论的诞生。事实上, 早年间狭义相对论在学界被介绍时便是以洛伦兹和爱因斯坦这一对忘年交命名的洛伦兹-爱因斯坦理论。

电子, 这样一个基本粒子可以说是妇孺皆知的概念, 它带一个单位的负电荷, 其基本参数为:

$$e = 1.602176634 \times 10^{-19} \text{C} \quad m_e = 9.10938356 \times 10^{-34} \text{kg}$$

然而出于各种原因, 在经典电子论中的电子却又是陌生的概念。根据洛伦兹的经典电子论精神, 重的原子实被认为静止, 轻的束缚电子被认为在外场中可以运动, 且主要受电场的影响<sup>4</sup>。电子在外电场中相对原子实的位移造成了电偶极矩。出于好的近似, 将电子视作被绑在连接中心的弹簧的另一端的质点。质量为  $m$  不再被认为与基本粒子的质量一致, 事实上这个质量的本质实际上是虚无的, 因为当考虑到原子与电磁场相互作用的严格理论时图像将完全是量子的而超出了此处可以讨论的范围, 这样的一个为了解释现象去构造的模型就叫唯象模型 (phenomenological model)。但是其电荷量还是  $-e$ , 弹簧劲度系数为  $k$ 。这样子我们就可以写出电子在外电场中的动力学方程:

$$m\ddot{\mathbf{r}} + k\mathbf{r} = -e\mathbf{E}$$

但是我们还忽略了一点, 加速运动的电荷会产生辐射, 也会因为辐射而带走能量和动量, 这就被视作介质对电磁波吸收, 散射的根本原因。也同样为了唯象地描述它, 我们认为电子是受到了一个阻尼力  $f = -\gamma v$ 。这样以上方程就被修改为:

$$m\ddot{\mathbf{r}} + \gamma\dot{\mathbf{r}} + k\mathbf{r} = -e\mathbf{E}$$

命  $\omega_0 = \sqrt{k/m}$  代表共振频率 (resonance frequency) 或待会就会说明的吸收频率 (absorption frequency)。再命入射电磁波为频率  $\omega$  的光, 按光学符号约定  $\mathbf{E} = \mathbf{E}_0 e^{-i\omega t}$ , 那么很容易解出受迫振动解  $\mathbf{r} = \mathbf{r}_0 e^{-i\omega t}$ , 振幅为:

$$\mathbf{r}_0 = \frac{-e\mathbf{E}_0/m}{\omega_0^2 - \omega^2 - i\gamma\omega/m}$$

我们现在就能算出动态的分子极化率来, 它被定义为分子中  $Z$  个价电子产生的偶极矩  $\mathbf{p} = -Z\epsilon_0\mathbf{r}_0$  与外电场  $\mathbf{E}_0$  的比值, 可以发现它也是个复数, 依赖于外光场的频率:

$$\alpha(\omega) = \frac{Ze^2/m}{\omega_0^2 - \omega^2 - i\gamma\omega/m}$$

我们知道, 对于具体介质的极化总是与各个单元的极化有关。我们在此讨论稀薄的无固有偶极矩的气体

<sup>4</sup>根据洛伦兹力公式  $\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B})$ , 而电磁场中  $E = cB$ , 故  $v \ll c$  时磁场力可以忽略。

在外电磁波中的极化<sup>5</sup>, 此时不难想到介质的介电常数应该直接依赖于分子极化率, 事实上, 极化强度为:

$$\mathbf{P} = \chi \varepsilon_0 \mathbf{E} = (\varepsilon_r - 1) \varepsilon_0 \mathbf{E} = n \mathbf{p}$$

从而得到复电容率的值:

$$\varepsilon' = \varepsilon_0 + \frac{n Z e^2 / m}{\omega_0^2 - \omega^2 - i \gamma \omega / m}$$

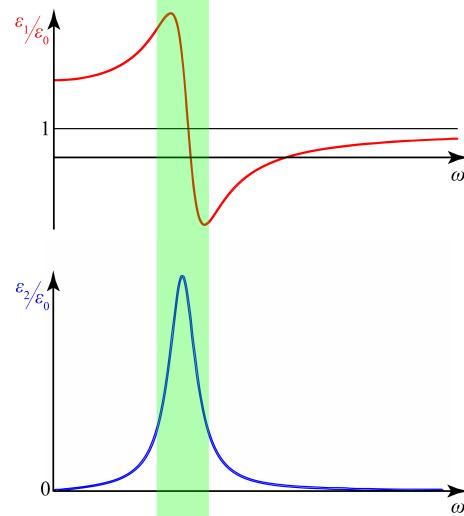
通常, 把  $n Z e^2 / \varepsilon_0 m$  称作等离子体频率(plasma frequency)的平方  $\omega_p^2$ . 那么上式也作:

$$\varepsilon' = \varepsilon_0 + \frac{\omega_p^2}{\omega_0^2 - \omega^2 - i \gamma \omega / m} \varepsilon_0$$

出现虚部其实就意味着损耗, 在振动的过程中电场在持续不断地对振子输入能量而导致了电磁波的衰减. 只不过这里的损耗并不是导电行为导致的. 写出复电容率的实部和虚部  $\varepsilon' = \varepsilon_1 + i \varepsilon_2$ :

$$\varepsilon_1 = \varepsilon_0 \cdot \left[ 1 + \frac{\omega_p^2(\omega_0^2 - \omega^2)}{(\omega_0^2 - \omega^2)^2 + (\gamma \omega / m)^2} \right]$$

$$\varepsilon_2 = \varepsilon_0 \cdot \frac{\gamma \omega_p^2 \omega / m}{(\omega_0^2 - \omega^2)^2 + (\gamma \omega / m)^2}$$



两个函数看上去复杂, 其实都属于近似的洛伦兹型函数(lorentzian function). 图像如右图所示. 在绿色带状区域这个函数有着较大的转折. 这发生在  $\omega \approx \omega_0$  处. 若做近似, 认为反应损耗的  $\gamma$  较小. 则令  $\omega^2 - \omega_0^2 = x$ ,  $\gamma \omega_0 / m = a$ , 那么上式在带状区域附近近似为标准的洛伦兹型:

$$\varepsilon_1 / \varepsilon_0 = 1 - \omega_p^2 \cdot \frac{x}{x^2 + a^2}$$

$$\varepsilon_2 / \varepsilon_0 = \omega_p^2 \cdot \frac{a}{x^2 + a^2}$$

电容率的实部在带状区域由区域外的缓慢增加转为剧烈减少, 而恰好在这样一个区域, 电容率的虚部突然变得很大. 我们习惯上用虚部这种洛伦兹函数增加到最大值的一半的两个点之间的间距作为特征的半高峰宽(full width at half maximum, FWHM). 对于洛伦兹型, 其值恰为  $x = \pm a$  之间的间距  $2a$ , 但是注意到要转化为  $\omega$  对应的间距, 它恰好为:

$$\Gamma = \frac{\gamma}{m}$$

我们可以从这个模型中清楚地看到两个物理现象: 色散(dispersion) 与吸收(absorption). 前者就是说折射率依赖于波长, 这和复电容率的实部依赖于频率只是说法不同本质一样. 而复电容率的虚部, 根据之前的讨论, 也就代表吸收. 我们发现在入射光的频率近似为谐振子的固有频率  $\omega_0$  时, 对应经典力学中发生共振的区域附近, 色散和吸收都变得很显著. 我们知道, 经典电子论给出的解释仅仅是一个唯象模型, 所以相关的参数都要根据实验结果来确定.  $\omega_0$  可以根据发生强烈吸收与色散的波长来确定, 而  $\gamma/m$  则根据吸收峰的半高峰宽  $\Gamma$  来确定, 最后等离子体频率  $\omega_p$  则比较有趣, 我们恰好可以根据零频率处的折射率  $n_0$  来确定它:

$$n_0^2 = \varepsilon_1(0) / \varepsilon_0 = 1 + \omega_p^2 / \Gamma^2$$

那么实际情况是否这么简单呢? 至少通过对右图的常温下水的色散的测量中我们发现, 除了在可见光波段水几乎是透明的而具有大约 1.33 的折射率, 在近红外 (0.8-2.5 μm) 与中红外 (2.5-15 μm) 波段的短波区定性上符合以上模型推导出来的结果. 但是中红外到远红外 (15-1000 μm) 则明显偏离以上结果.

<sup>5</sup>如果是液体或固体, 一方面由于不同的偶极子间的相互影响不可忽略导致公式需要修正, 也因为极化方式还有取向极化等所以需要修正, 但全都只影响定量结果, 定性的图像仍然是适用的.

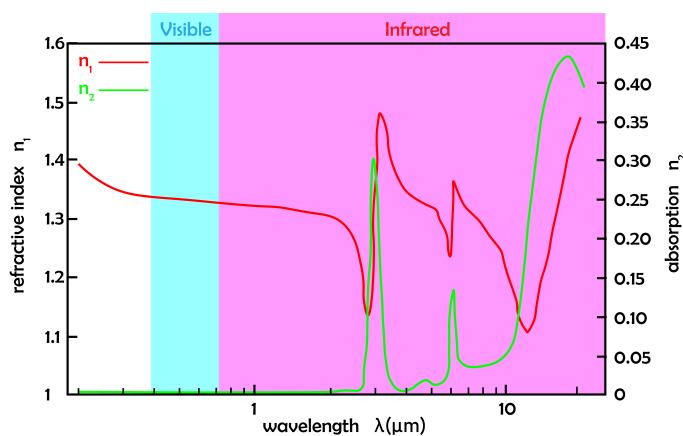


图 27.4: 水的可见-红外色散曲线

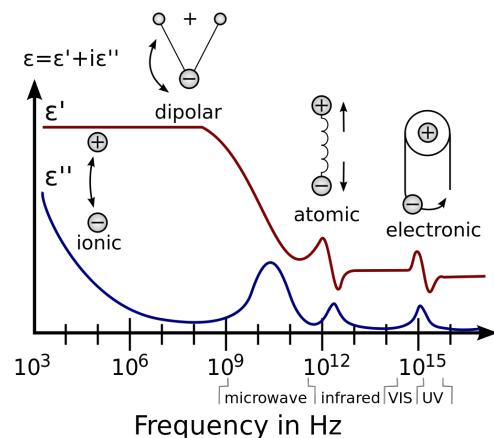


图 27.5: 光与原子作用类型

事实上, 在不同的波段, 适用于不同的光与物质相互作用的基本模型. 微波波段光波长足够长使得晶体的正负离子做大范围的整体相对运动<sup>6</sup>, 就是晶格振动. 红外则开始使得分子振动, 包括转动和振动等形式. 再往下的可见波段才是电子共振, 也包括原子间的能带共振和原子内的能级共振. 最后在紫外波段, 电子甚至能直接被电离, 这就造成了光与物质相互作用问题的复杂性.

但无论哪种相互作用的机制, 我们上述推导得到的描述有着特定共振频率的色散与吸收的最终公式是十分普适的. 它只有三个待定的参数: 共振频率  $\omega_0$ , 半高峰宽  $\Gamma$  和反应共振强度的等离子体频率  $\omega_p$ . 我们接下来要做的, 是把不同振子, 不同吸收效应带来的结果进行求和:

$$\begin{aligned}\varepsilon_1/\varepsilon_0 &= 1 + \omega_p^2 \cdot \sum_i \frac{c_i(\omega_i^2 - \omega^2)}{(\omega_i^2 - \omega^2)^2 + \Gamma_i^2 \omega^2} \\ \varepsilon_2/\varepsilon_0 &= \omega_p^2 \cdot \sum_i \frac{c_i \Gamma_i \omega}{(\omega_i^2 - \omega^2)^2 + \Gamma_i^2 \omega^2}\end{aligned}$$

其中  $\omega_p$  仍然统计各个吸收峰的总强度, 而  $c_i$  系数的和必须为 1 则统计各个吸收峰的相对强度. 我们最后计算远离吸收带处的折射率值. 忽略各个  $\Gamma_i$  后得到:

$$n^2 = 1 + \omega_p^2 \cdot \sum_i \frac{c_i}{\omega_i^2 - \omega^2}$$

于是便得到著名的色散关系:

$$n^2 = 1 + \sum_i a_i \cdot \frac{\lambda^2}{\lambda^2 - \lambda_i^2}, \quad a_i = \frac{c_i \omega_p^2 \lambda_i^2}{4\pi^2}$$

上式依然可以做近似, 把共振的各个波长  $\lambda_i$  从小到大按  $i$  来排列. 不妨设  $\lambda_i \ll \lambda_{i+1}$ , 而  $\lambda$  恰好夹在两者之间, 那么:

$$n^2 =$$

## 27.2 色散, 散射与吸收

## 27.3 群速与展宽

## 27.4 光量子

<sup>6</sup>十分类似于热运动的那种形式, 区别在于这是外场诱导的有规律的振动, 而且正负电荷位移一定反相, 对应热振动的特定光学支.

# 第二十八章 量子论

我们用选自费曼(R. Feynman)先生的物理学讲义的开篇金句来作为本章与下一章近代物理内容的开头:

Each piece, or part, of the whole of nature is always merely an approximation to the complete truth. Therefore, things must be learned only to be unlearned again or, more likely, to be corrected. The test of all knowledge is **experiment**. Experiment is the sole judge of scientific “truth”.

的确,量子理论以其不直观而被近代早期物理学家们所疑惑,这其中不乏一些赫赫有名的大师.直到现在也有很多基础的问题是没有被深刻地理解的:电子的本性与内部结构,基本粒子的类别与参数,对量子非定域性与测量的理解...所以在关于可能会造成问题的领域的学习时,采取先明白实验上的事实,从中理解理论建立的必要性.

## 28.1 黑体辐射

对于**热辐射**(thermal radiation)的讨论是何时进入物理研究的视野的呢?可以肯定的是人类认识到热辐射现象非常的早:光芒万丈的太阳,烧红的木炭与金属都是典型的热辐射的情形.但人们掌握足够的方法去测量它则也是要到19世纪后半期了.热辐射势必涉及到电磁场与电荷的相互作用.而且深入到原子尺度,实际上就是电磁波的发射与吸收.对于电磁波的发射,我们在电磁学中粗略讲过,只要有加速运动的电荷就会导致电磁辐射.之后小节我们将认识到微观电荷不能用“加速”来描述,其状态其实是量子态,处于激发态才会自发向基态去跃迁放出电磁辐射.而对于电磁波的吸收,则在光学中我们简要介绍过洛伦兹电子论中的处理方法.

一个物体如果能够在任何温度下把照射到它上面的任何频率的光都全部吸收掉,那么这个物体就叫做**黑体**(black body).虽然在现实生活中这样的物体并不存在,但石墨和碳黑往往被视作比较理想的黑体,尽管这样,黑体“看上去”也不总是“黑的”.这里有两个主要的原因:

## 28.2 光粒子性

## 28.3 玻尔原子

## 28.4 电子波动性

## 28.5 物质波与波函数



# 第二十九章 物理学尺度

**29.1 宇观**

**29.2 宏观**

**29.3 介观**

**29.4 微观原子**

**29.5 微观亚原子**