

固体物理

吴晋渊

2025 年 4 月 22 日

前置知识

真正重要的前置知识只有量子力学的基本框架（有态矢量、有希尔伯特空间、有算符），哈密顿量和拉格朗日量是什么，以及在此基础上的一些自然推论，比如说量子场论和统计场论的框架，包括格林函数、费曼图、重整化群等。关于平衡态格林函数和费曼图，^[28]的前几章是很好的介绍（但出于某些原因，文小刚使用零温场论的形式表述所有问题，即使是平衡态有限温问题）。

一定的“具体物理”还是需要的，包括单粒子量子力学，主要是关于 $[x_i, p_i] = i\delta_{ij}$ 和自旋自由度的那些东西，以及经典和量子电动力学，当然还有它的低能极限——库伦相互作用。这些可以在任何教科书中找到。

一些好用的技巧，诸如Hubbard-Stratonovich变换，Bogoliubov变换等，可以在各种专门的凝聚态场论教科书中找到，如^[2]。

由于以上形式理论非常容易在各种教科书中找到，本文不再介绍它们，而是将注意力集中在凝聚态问题上。必要的前置知识——尤其是那些涉及记号约定的——将通过“背景知识”块引入。

记号约定

术语与表达式记号

“自由”一词可能具有两个意思：一个表示**自由理论**，即系统的场论哈密顿量中只有场算符的二次型，一个表示**自由电子**，即能量就是 $\mathbf{k}^2/2m$ 的电子。前者包括后者但是不完全就是后者，因为能带电子、紧束缚模型等也属于前者。

本文提到的费米子主要是电子，使用一个三维坐标 \mathbf{r} （或者三位动量 \mathbf{p} ），以及只有向上和向下两种选择的自旋就可以描述一个电子。单电子自旋算符为

$$\mathbf{S} = \sum_{\alpha, \beta} |\alpha\rangle \boldsymbol{\sigma}_{\alpha\beta} \langle\beta|, \quad (0.1)$$

其中 α 和 β 取遍 \uparrow 和 \downarrow ， $\boldsymbol{\sigma}$ 为泡利矩阵。在不涉及自旋-轨道耦合的场合，在书写哈密顿量时我们直接略去自旋的下标，这是合理的，因为只需要把不考虑自旋的哈密顿量中的各个产生湮灭算符根据自旋守恒的性质机械地加上自旋下标再求和就能够得到完整的哈密顿量。在需要实际计算粒子数时就不能这么做了；需要计算总能量时当然也不能这么做。

由于本文不涉及相对论性过程，设 \mathbf{a} 为一个矢量，则使用 a 表示其模长。

在符号够用时，用 $i\omega_n$ 表示松原频率，在符号不够用时，设一个场已有动量标记 \mathbf{k} ，用 ik^0 表示松原频率。此时用 k 表示 (\mathbf{k}, ik^0) 。

对离散格点系统，使用 $\langle i, j \rangle$ 表示最接近的一对格点。（只求和一次，即认为 $\langle i, j \rangle$ 和 $\langle j, i \rangle$ 相同）

我们用 $\{\alpha|\mathbf{a}\}$ 表示三维欧几里得群 E 的成员，它定义为

$$\{\alpha|\mathbf{a}\}\mathbf{r} = \alpha\mathbf{r} + \mathbf{a}, \quad (0.2)$$

其乘法为

$$\{\alpha_i|\mathbf{a}_i\}\{\alpha_j|\mathbf{a}_j\} = \{\alpha_i\alpha_j|\alpha_i\mathbf{a}_j + \mathbf{a}_i\}, \quad (0.3)$$

从而

$$\{\alpha_i|\mathbf{a}_i\}^{-1} = \{\alpha_i^{-1}|\alpha_i^{-1}\mathbf{a}_i\}. \quad (0.4)$$

单位制

本文取普朗克单位制，认为 $\hbar = c = 1$ ，且 $4\pi\epsilon_0 = 1$ ， $k_B = 1$ 。将本文的计算结果恢复到国际单位制需要遵循以下规则：

- 将本文中的 T 替换为 $k_B T$ ；

-

需要注意“比率”（如热容、态密度等）的定义式中的 T, d^3k 等不需要替换。

h.c.表示厄米共轭，c.c.表示复共轭。

若无特殊说明， $f(z)$ 定义为近独立费米子的分布函数，即

$$f(z) = \frac{1}{e^{\beta z} + 1}.$$

主要的字母符号

费米子的产生湮灭算符为 c^\dagger 和 c ，而如果是关于位置的产生湮灭算符，则为 ψ^\dagger 和 ψ 。

第一部分

晶体系统的组成

这一部分是关于晶体系统的组成的。我们将讨论晶体如何形成（第一章），理想晶体的几何结构和对称性，它的点群、空间群中有什么操作，如何据此分析可能的晶格形式，以及如何根据纯粹的几何性质分类晶体，以及准粒子是晶体的几何结构的对称群的表示，这导致很多非常物理的结果，如准粒子应该使用晶格动量标记这一事实（第二章）。基本上，凝聚态系统的基本自由度都可以使用准粒子的观点讨论，将自旋自由度称为准粒子可能有些牵强，但是它们无疑是产生准粒子的场。很多人用准粒子一词表示电子、空穴等和电子结构有关的集体模式而用元激发表示声子等经典极限为波动的集体模式，虽然这个区分并不是很重要——例如，Luttinger液体中的玻色子算是什么呢？还有一种区分方式是将元激发限定为和系统的激发态相关的准粒子。有些准粒子在基态中也会出现，不将它们称为元激发。

我们将讨论能带电子、声子的基质和性质。固体中固有的成分包括电子和晶格。电子之间存在库伦相互作用，同时受到晶格提供的周期性势场作用。在静止晶格的周期性势场作用下，原本能量简并的电子状态的能量会发生偏移，从而，一系列能量完全相同的电子态经过周期势场的扰动，其能量将成为一系列相差非常小的能级。当简并电子态的数目足够多——正如晶体中通常是的那样——就会形成在能谱中连续的能带，对应的电子状态即为能带电子。周期势的全部影响就是裸电子将会有能量修正而成为能带电子，无相互作用的能带电子组成费米气体，很容易处理。库伦排斥和声子介导的吸引相互作用合并为电子-电子相互作用，非常难以处理——电子-电子相互作用可以导致各种奇特的现象，并且其量级通常与费米能的量级一致。

固体中，晶格可以起到两种作用：一种是提供一个周期性势场，一个是晶格的畸变；前者属于电子的单体哈密顿量，可以通过求解自由模型被考虑在内；后者则导致声子。这样，解出电子的近独立理论（如紧束缚模型）、声子的自由理论之后再加入电子、声子间的相互作用，我们就可以用电子和声子组成的相互作用气体完整描述晶体的行为，而不需要再显式考虑晶格的存在。例如，在计算热容等变量时，完全可以认为热容是声子和电子提供的，而不是（裸的）电子和晶格提供的。

离子实的运动让我们得到一个定义在正格子上的场，这个场的激发就是声子。晶体中一定会出现这样的激发，且这是一个Goldstone模式，因为晶体的形成实际上破缺了连续平移对称性，那么必然有一种零质量激发产生。离子实的哈密顿量大致如下：（使用大写字母以和电子区分）

$$H = \sum_n \frac{P_n^2}{2M_n} + \frac{1}{2} \sum_{\langle m,n \rangle} \omega_{mn}^2 (X_m - X_n)^2 + \dots,$$

其中 n 指的是晶格坐标，也就是三元组 (n_1, n_2, n_3) ，

$$\mathbf{X} = \mathbf{R} - \mathbf{R}_0$$

为某个离子实的位移。正则量子化之后，哈密顿量中的前两项给出一个自由场，而离子实之间的非线性相互作用则给出声子-声子散射过程。由于坐标-动量关系是对易关系而不是反对易关系，声子是玻色子。

离子实和电子的相互作用通常取这样的形式：

$$H_{ei} = \sum_{n,i} V_{ei}(\mathbf{r}_i - \mathbf{R}_n) = \sum_{n,i} (\mathbf{R}_n - \mathbf{R}_n^0) \cdot \nabla V_{ei}(\mathbf{R}_n^0 - \mathbf{r}_i) + \dots,$$

其中 \mathbf{R}_n^0 指的是离子实 n 未发生移动时的位置，即 n 号正格子的位置。对位移场 \mathbf{X} 做正则量子

化，并将关于电子的部分做二次量子化，得到

$$H_{ei} = \sum_{n,i} V_{ei}(\mathbf{r}_i - \mathbf{R}_n) = \sum_n \int d^3\mathbf{r} \mathbf{X} \cdot \psi^\dagger(\mathbf{r}) \nabla V_{ei}(\mathbf{R}_n^0 - \mathbf{r}) \psi(\mathbf{r}) + \dots,$$

这就是电子-声子相互作用。可以看出以上哈密顿量对 \mathbf{X} 不具有 $U(1)$ 对称性，因此声子一般来说不是守恒的。

声子和电子的量子化过程很不一样。对电子，我们首先写出一个多体一次量子化哈密顿量，然后做二次量子化；而对声子，我们实际上把使用格点坐标标记的离子实位移和动量当成了场算符（离散的场），然后直接对这两个场算符做正则量子化。换言之，没有声子的一次量子化：定义声子时我们的理论就是二次量子化的。

原则上通过分析电子和声子组成的凝聚态场论可以得到凝聚态系统的一切性质。但是，相互作用和特殊的势场会导致各种眼花缭乱的现象。例如，考虑一个被较强的势场或是库伦排斥固定在空间某一点的电子，它提供一个轨道编号的自由度和一个自旋自由度。相邻的这种电子的自旋自由度之间可能通过交换相互作用等产生等效相互作用，这样的电子周期排列，就形成了一个能够产生一些新的元激发的场。

第一章 晶体

1.1 原子物质的哈密顿量

普通的固体、液体、气体由一系列原子组成。通过实验和计算可以发现，原子的最外层电子在各种过程中容易发生重新排列，称为价电子；内层电子和原子核（合称为离子实）则通常保持为一个整体，也即，其内部状态发生变化的物理过程的描述需要使用QCD，其涉及的能标远高于价电子发生变化涉及的能标。

本文基本上只分析涉及价电子低能运动的物理过程，即只讨论非相对论极限下的电荷-电磁场耦合系统，而忽略强相互作用、弱相互作用和引力。此时，带电粒子由薛定谔场完全描述，系统具有 $U(1)$ 规范对称性且无粒子数生灭，有确定的粒子数，于是可以原则上实物粒子部分可以直接用单粒子量子力学描述。进一步，我们假定系统中没有变化特别快的电磁场，这意味着实际上我们可以积掉电磁场并且得到一个延迟不明显的相互作用，那就是说，我们可以把所有电磁相互作用都用静电学和静磁学处理。由于无论是电子还是离子实都是非相对论性的，电子-电子相互作用、电子-离子实相互作用、离子实-离子实相互作用几乎完全是库伦相互作用。

1.1.1 动能和库伦能

设有 N_e 个价电子， N_i 个离子实（ i 表示离子），不考虑外界扰动的一次量子化哈密顿量为

$$H = H_e + H_i + H_{ei}, \quad (1.1)$$

其中 H_e 表示仅涉及价电子的哈密顿量， H_i 表示仅涉及离子实的哈密顿量，最后一项则是两者的相互作用，所有相互作用是库伦相互作用。诸价电子组成的系统就好像由电子组成的气体，称为相互作用电子气。单体哈密顿量为电子的动能项加上单体势能项。在物质不受外界作用时当然不应该有单体势能项，于是

$$H_{e1} = \frac{\mathbf{p}^2}{2m},$$

在坐标表象下它就是

$$H_{e1} = -\frac{\nabla^2}{2m}.$$

二体哈密顿量为电子两两作用而产生的库伦势能是

$$H_{e2} = \frac{e^2}{|\mathbf{r}_1 - \mathbf{r}_2|},$$

从而价电子本身的能量以及它们之间发生库伦相互作用的能量就是

$$H_e = \sum_{i=1}^{N_e} \frac{p_i^2}{2m_e} + \frac{1}{2} \sum_{i \neq j} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|}. \quad (1.2)$$

使用类似的方法，离子实的组成的系统（如果是晶体那就是晶格）的哈密顿量为

$$H_i = \sum_{\alpha=1}^{N_i} \frac{p_{\alpha}^2}{2m_i} + \frac{1}{2} \sum_{\alpha \neq \beta} V(\mathbf{R}_{\alpha} - \mathbf{R}_{\beta}). \quad (1.3)$$

由于离子实中的内层电子结构复杂，离子实之间的相互作用能写不出特别简单的表达式（我们相当于把内层电子的自由度也积掉了）。请注意这个相互作用能是平移不变的，这是当然的，因为QED是平移不变的；但是实际的固体在短距离上并不是平移不变的，因为在低能下有对称性自发破缺。离子实和价电子的相互作用则是

$$H_{ei} = \sum_{\alpha, i} V_{ei}(\mathbf{r}_i - \mathbf{R}_{\alpha}). \quad (1.4)$$

分别使用*i*表示价电子，用 α 表示离子实；由于价电子和离子实不全同，不需要加上1/2系数。同样我们还是假定了相互作用本身的平移不变性。本文仅仅讨论固体（实际上主要是晶体）的物理，因此我们将假定离子实的位移始终局限在非常小的范围内。

背景知识 1.1: 量子涨落

“量子涨落”一词的含义其实有一些不清楚之处。

由于量子理论中的哈密顿量可以有彼此不对易的项，计算该哈密顿量的配分函数必然会得到一个路径积分，每条路径上，不同虚时间点的物理量可以任意取值，而路径积分权重中将会出现一个Berry相位项（如 $\pi\dot{\phi}$ 项），这个现象就可以称为量子涨落。 $\hbar \rightarrow 0$ 或者说取鞍点近似时，这个Berry相位项就是勒让德变换中的 $\pi\dot{\phi}$ 项。

这个说法的经典极限此时是有一定的微妙之处的，就平衡态统计中的配分函数而言，取 $\hbar \rightarrow 0$ 似乎会让配分函数变成简单的 $\sum e^{-\beta H}$ ，每条路径都没有虚时间演化；但是另一方面，经典动力学中 $\hbar \rightarrow 0$ 时虽然哈密顿量中没有不对易的项，但是物理量的时间演化还是有的。这个微妙之处是必然会出现的，因为量子力学中我们依靠算符的非平庸对易关系 $[H, A]$ 来获得时间演化，求经典极限时，直接关于哈密顿量的配分函数在 $\hbar \rightarrow 0$ 极限下退化为 $\sum e^{-\beta H}$ ，但是海森堡运动方程由于有 $1/\hbar$ 的前置因子，会给出经典的哈密顿运动方程。这也体现出了“经典统计物理”的内在张力：经典统计物理相比量子统计物理更加难以良定义。某种意义上不存在纯粹的经典统计物理：例如，计算服从牛顿力学的经典粒子系统的配分函数时如果同时以 \mathbf{r} 和 \mathbf{p} 为变量，则必须引入相空间离散化，实际上这是用到了坐标和动量的不确定性关系的。

在求解薛定谔方程时，可以将发生得较慢的过程涉及到的物理量当成不变的，或者将量子效应不明显的那部分系统用经典动力学处理，这就称为“忽略量子涨落”，其意义是显然的。

经典物理的微扰求解实际上也是可以用费曼图表示的，只不过这种费曼图中的粒子线的方向必须严格体现“事件发生的顺序”^[16]。因此，量子场论中超出经典场论的费曼图的那部分费曼图也可以称为量子涨落。

还有一种“量子涨落”的定义是物理量在能量本征态下的方差。这种定义的量子涨落和前述“系统构型的虚时间演化”是紧密相关的：如果物理量*A*的方差很大，这意味着它和哈密顿量里面的某些项强烈地不对易，从而它有时间演化。然而，这个对量子涨落的定义会导致一定的混淆。我们也经常将偏离基态的系统构型称为“涨落”，因此会有“低能下的涨落”之说，即系统的低能自由度。然而，如果一个物理量*A*在低能能量本征态下的方差很大，那么它显然不适合用于标记系统的低能自由度。因此，如果*A*“具有很强的量子涨落”，那么*A*“不是系统的低能涨落”！例如，

在 σ^z 铁磁序中， σ^x “有很强的量子涨落”，但是它不是系统的低能涨落。规范场的禁闭相中规范场有很强的量子涨落（否则规范荷不可能被紧紧束缚在一起：例如，设一个规范荷被外场移动了一些，很快强烈的规范场被激发出来将另一个规范荷也移动相同距离，然后规范场强度下降），但是系统的低能涨落中不包括规范场涨落。

在大部分过程中，由于原子核的质量比电子的质量大至少三个数量级，涉及价电子的过程通常比涉及离子实的过程发生得快很多，从而在价电子的时间尺度上，诸离子实的位置可以看成是给定的。从而，在分析价电子时我们可以将 H_i 项直接略去，并忽略离子实位置的量子涨落，而将 $H_{ei}(\mathbf{r}_i - \mathbf{R}_\alpha)$ 项对 \mathbf{R}_α 求和得到 $V_{ion}(\mathbf{r}_i)$ 。这个近似称为玻恩-奥本海默近似。这样一来相互作用电子气的一次量子化哈密顿量在坐标表象下就是

$$H = \sum_{i=1}^{N_e} \left(-\frac{\nabla^2}{2m_e} + V_{ion}(\mathbf{r}_i) \right) + \frac{1}{2} \sum_{i \neq j} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|}, \quad (1.5)$$

从而二次量子化哈密顿量为

$$H = \sum_{\sigma} \int d^3\mathbf{r} \psi_{\sigma}^{\dagger}(\mathbf{r}) \left(-\frac{\nabla^2}{2m} + V_{ion}(\mathbf{r}) \right) \psi_{\sigma}(\mathbf{r}) + \frac{1}{2} \sum_{\alpha, \beta} \int d^3\mathbf{r}_1 \int d^3\mathbf{r}_2 \psi_{\alpha}^{\dagger}(\mathbf{r}_1) \psi_{\beta}^{\dagger}(\mathbf{r}_2) \frac{e^2}{|\mathbf{r}_1 - \mathbf{r}_2|} \psi_{\beta}(\mathbf{r}_2) \psi_{\alpha}(\mathbf{r}_1). \quad (1.6)$$

其中 $\psi^{\dagger}(\mathbf{r})$ 是薛定谔场的场算符，它也是在位置为 \mathbf{r} 的位置产生一个电子的产生算符。这个哈密顿量当然也可以通过QED的低能近似得到，但并没有必要这么做。请注意电子是费米子。(1.6)实际上不是对角的，因为它的单粒子项涉及一个梯度算符。另一方面，离子实的相互作用可以认为遵从由电子总能量确定的等效势，¹可以列写晶格的薛定谔方程求解其位置。然而由于离子实通常比较大，我们很多时候会彻底忽略离子实位置的量子涨落而使用分子动力学方法处理它。

最后我们指出，由于凝聚态介质始终可以和外界交换电子——外界的电子可以进入系统，系统中的电子可以溢出——系统中电子气本身的哈密顿量(1.6)不足以充分描述系统。本文将只研究近平衡系统，因此这种与外界的相互作用可以使用化学势描述，即我们需要在(1.6)中加入一项 $-\mu \psi^{\dagger}(\mathbf{r}) \psi(\mathbf{r})$ ，这样得到的哈密顿量才是完整的。换言之，电子气完整的哈密顿量形如

$$H = \sum_{\mathbf{k}, \sigma} \left(\underbrace{\frac{k^2}{2m}}_{\epsilon_{\mathbf{k}}} - \mu \right) c_{\mathbf{k}\sigma}^{\dagger} c_{\mathbf{k}\sigma} + \int d^3\mathbf{r} V_{ion}(\mathbf{r}) \psi^{\dagger}(\mathbf{r}) \psi(\mathbf{r}) + \frac{1}{2} \sum_{\alpha, \beta} \int d^3\mathbf{r}_1 \int d^3\mathbf{r}_2 \psi_{\alpha}^{\dagger}(\mathbf{r}_1) \psi_{\beta}^{\dagger}(\mathbf{r}_2) \frac{e^2}{|\mathbf{r}_1 - \mathbf{r}_2|} \psi_{\beta}(\mathbf{r}_2) \psi_{\alpha}(\mathbf{r}_1). \quad (1.7)$$

在有些模型中有时也将外势场并入 $\epsilon_{\mathbf{k}}$ 项。为了简便起见，通常用 ξ 表示扣除了化学势的单电子能量，即

$$\xi_{\mathbf{k}} = \epsilon_{\mathbf{k}} - \mu. \quad (1.8)$$

1.1.2 哈密顿量的一般形式

以上我们都是将库伦相互作用加入薛定谔场中，可以说是给出了第一性原理计算需要

¹注意，先计算电子-电子库伦散射导致的能量然后用它计算晶格中离子实的相互作用最后计算离子实振动对电子的影响，这个过程中没有双重计数，虽然我们表面上似乎“积掉了电子而计算等效的原子间相互作用”。

的哈密顿量（虽然实际上从高能物理的角度这远非第一性原理，但对凝聚态理论来说通常已经够用了）。但实际上还有以下机制没有考虑：

- 作用在单体上的外场的束缚，离子实的束缚已经被计入考虑了，但是还有其它外场，比如说或许会有一个磁场，然后哈密顿量中将会有一项 $-\boldsymbol{\mu} \cdot \mathbf{B}$ ；无论如何这会是一个单粒子算符。
- 电子-声子相互作用会引入一个二电子和一个声子发生相互作用的顶角，积掉声子自由度之后会留下一个等效的电子-电子相互作用，这会让(1.7)中电子-电子相互作用的系数发生变化，不再是严格的库伦排斥。

于是我们写出一般形式的相互作用电子气的二次量子化哈密顿量：

$$H = \sum_{\mathbf{k}, \sigma} (T_{\mathbf{k}} - \mu) c_{\mathbf{k}\sigma}^\dagger c_{\mathbf{k}\sigma} + \sum_{\mathbf{k}_1, \mathbf{k}_2, \sigma} V_{\mathbf{k}_1 \mathbf{k}_2 \sigma} c_{\mathbf{k}_1 \sigma}^\dagger c_{\mathbf{k}_2 \sigma} + \sum_{\mathbf{k}_1, \mathbf{k}_2, \mathbf{q}, \alpha, \beta} c_{\mathbf{k}_1 + \mathbf{q}, \alpha}^\dagger c_{\mathbf{k}_2 - \mathbf{q}, \beta}^\dagger V_{\mathbf{q}} c_{\mathbf{k}_2 \beta} c_{\mathbf{k}_1 \alpha}. \quad (1.9)$$

相应的，热力学作用量为

$$S = \sum_n \left(\sum_{\mathbf{k}, \sigma} (-i\omega_n + T_{\mathbf{k}} - \mu) \bar{c}_{\mathbf{k}\sigma} c_{\mathbf{k}\sigma} + \sum_{\mathbf{k}_1, \mathbf{k}_2, \sigma} V_{\mathbf{k}_1 \mathbf{k}_2 \sigma} \bar{c}_{\mathbf{k}_1 \sigma} c_{\mathbf{k}_2 \sigma} + \frac{1}{2V} \sum_{\mathbf{k}_1, \mathbf{k}_2, \mathbf{q}, \sigma} \bar{c}_{\mathbf{k}_1 + \mathbf{q}, \sigma} \bar{c}_{\mathbf{k}_2 - \mathbf{q}, \sigma} V_{\mathbf{q}} c_{\mathbf{k}_2 \sigma} c_{\mathbf{k}_1 \sigma} \right). \quad (1.10)$$

上式中的动能项和相互作用项的形式由对称性确定，具体系数可以暂时不设置具体值（因为直接从(1.7)出发做重整化群计算显然是非常困难的）。我们经常写出一些简化了的模型以复现某个现象，如忽略一部分动能（即忽略一部分电子跃迁方式），或者简化相互作用形式。

背景知识 1.2: 量子多体系统中的理论与数值计算方法

- 费曼图微扰计算（需要注意由于库伦相互作用是瞬时而超距的，通常将库伦相互作用顶角写成用虚线连接的两个顶角，每个顶角有一个电子入射和一个电子出射，虚线可以携带动量），由于相互作用是二体的，在相互作用较弱时计算一到二圈图就可以得到很好的效果，不过实际上相互作用并不总是那么弱，此时需要一些其它近似手段，如RPA近似等各种重求和方法。
- 做平均场计算并与实验或数值计算作比较（可以对原来的模型做平均场也可以对Hubbard-Stratonovich变换之后的辅助场做平均场，两者等价）。考虑平均场之上的涨落，获得一个关于平均值附近涨落的理论，具体方法通常是这样的：由于相互作用项是四阶的，可以使用Hubbard-Stratonovich变换引入一个辅助场，通过适当选取辅助场（通常要和某个有趣的序参量具有同样的对称性）并积掉电子自由度，则接近临界点时，辅助场满足的场论就给出了长程自由度。
- 密度泛函理论方法：凝聚态系统彼此不同的地方其实就在于晶格势场，我们只需要找到系统基态的一些量，如电子数密度等，能够和晶格势场建立一一对应关系即可，然后通过一些办法把系统总能量写成系统基态的这些量的泛函，通过优化该泛函，就得到了不少关于基态的性质，并且在适当的近似下能够计算出关于系统的全部信息。
- 量子蒙特卡洛方法、张量网络方法等数值方法。

量子多体系统通常都足够复杂，且可能有很强的相互作用，以至于分析方法多种多样，但是没有哪一种能够占有支配地位。

1.1.3 外加电磁场

现在讨论外加电磁场导致的哈密顿量变化，或者说“辐射和物质相互作用”导致的哈密顿量变化。一般的，设系统被放置在电磁场 (φ, \mathbf{A}) 中，则一次量子化哈密顿量（使用一次量子化哈密顿量是为了和经典的“一群电子定向移动”的物理图像对应上）为

$$H = \frac{1}{2m} \sum_i (\mathbf{p}_i - q_i \mathbf{A}(\mathbf{r}_i))^2 + \sum_i q_i \varphi(\mathbf{r}_i) + H_{\text{int}}, \quad (1.11)$$

其中 \mathbf{p} 是正则动量， H_{int} 表示粒子间相互作用。本文仅考虑外加电磁场产生的线性响应，于是考虑辐射场不很强以至于 \mathbf{A}^2 可以忽略的情况，也即，仅保留单光子过程，或者说做偶极辐射近似。(1.11)的导出见光学笔记，其中第1.2.1节解释了将辐射场和库伦相互作用分开的方式，第1.4.1节解释了偶极辐射近似。

对电子系统， $q = -e$ ，那么就有

$$\begin{aligned} H &= \frac{1}{2m} \sum_i (\mathbf{p}_i + e \mathbf{A}(\mathbf{r}_i))^2 - e \sum_i \varphi(\mathbf{r}_i) + H_{\text{int}} \\ &= -\frac{1}{2m} \sum_i (\nabla + ie \mathbf{A}(\mathbf{r}_i))^2 - e \sum_i \varphi(\mathbf{r}_i) + H_{\text{int}}. \end{aligned} \quad (1.12)$$

设 Ω 是某个空间区域，电流密度为 \mathbf{J} ，则

$$\int_{\Omega} d^3\mathbf{r} \mathbf{J} = - \sum_i e \mathbf{v}_i, \quad (1.13)$$

其中 \mathbf{v}_i 是电子移动的速度，满足

$$\mathbf{v}_i = \frac{\partial H}{\partial \mathbf{r}_i} = \frac{\mathbf{p}_i + e \mathbf{A}(\mathbf{r}_i)}{m}. \quad (1.14)$$

考虑到 Ω 的任意性，我们就有以下近似表达式：

$$\mathbf{J} = \underbrace{-\frac{e}{m} \sum_i (\delta(\mathbf{r} - \mathbf{r}_i) \mathbf{p} + \mathbf{p} \delta(\mathbf{r} - \mathbf{r}_i))}_j - \underbrace{\frac{e^2}{m} n_e \mathbf{A}}_{j_D}. \quad (1.15)$$

j 项特意被写成了厄米的形式： j_D 已经做了一遍粗粒化了，将诸 \mathbf{A}_i 平均了一遍。通常这是合理的，因为电磁波的波长通常不会特别小，从而不会有很大的空间起伏。（而如果有很大的空间起伏，我们会使用cQED而不是经典电动力学讨论问题了）

现在写出略去高阶项的哈密顿量的形式。选取库伦规范，并认为 $\varphi = 0$ ，此时会发现，实际上我们有

$$H = \frac{1}{2m} \sum_i \mathbf{p}_i^2 + H_{\text{int}} + \frac{e}{m} \sum_i \mathbf{p}_i \cdot \mathbf{A}(\mathbf{r}_i) - e \sum_i \varphi(\mathbf{r}_i),$$

$\mathbf{A}(\mathbf{r}_i)$ 和 \mathbf{p}_i 本来是不对易的，但是库伦规范下它们对易。代入 \mathbf{J} 的表达式并再次略去高阶项，就有

$$H = \frac{1}{2m} \sum_i \mathbf{p}_i^2 + H_{\text{int}} - \int d^3\mathbf{r} \mathbf{J} \cdot \mathbf{A} - e \sum_i \varphi(\mathbf{r}_i).$$

至于含有电势的那一项，注意到电荷密度为

$$\rho = -e \sum_i \delta(\mathbf{r}_i - \mathbf{r}),$$

于是最后得到

$$H = \frac{1}{2m} \sum_i \mathbf{p}_i^2 + H_{\text{int}} + \int d^3\mathbf{r} \varphi \rho - \int d^3\mathbf{r} \mathbf{J} \cdot \mathbf{A}. \quad (1.16)$$

虽然物质本身的哈密顿量不是洛伦兹协变的（因为取了非相对论近似），但是物质和辐射的相互作用项却是洛伦兹协变的——对电磁场的描述一般都是如此。以上哈密顿量实际上仅仅讨论了轨道部分，电子还有自旋磁矩

$$H_{\text{spin}} = \sum_i \boldsymbol{\mu}_i \cdot \mathbf{B}(\mathbf{r}_i),$$

我们可以如法炮制地将它写成

$$H_{\text{spin}} = \int d^3\mathbf{r} \boldsymbol{\mu} \cdot \mathbf{B}.$$

因此完整的哈密顿量实际上是

$$H = -\frac{1}{2m} \sum_i \mathbf{p}_i^2 + H_{\text{int}} + \int d^3\mathbf{r} \varphi \rho - \int d^3\mathbf{r} \mathbf{J} \cdot \mathbf{A} + \int d^3\mathbf{r} \boldsymbol{\mu} \cdot \mathbf{B}. \quad (1.17)$$

所有和单粒子携带的电荷数量有关的量全部被藏在电荷密度和电流密度中了，上式在电荷正反变换下不变。

(1.17)当然也可以非常容易地写成二次量子化的形式。薛定谔场满足 $U(1)$ 对称性，因此通过诺特定理可以得到守恒荷（当然就是电荷）

$$\rho = -e \sum_{\sigma} \psi_{\sigma}^{\dagger} \psi_{\sigma} = -en_e, \quad (1.18)$$

守恒流（也即电流密度）

$$\mathbf{j} = -\frac{ie}{2m} \sum_{\sigma} (\psi_{\sigma}^{\dagger}(\mathbf{r}) \nabla \psi_{\sigma}(\mathbf{r}) - (\nabla \psi_{\sigma}^{\dagger}(\mathbf{r})) \psi_{\sigma}(\mathbf{r})) - \frac{e^2}{m} \mathbf{A}(\mathbf{r}) \sum_{\sigma} \psi_{\sigma}^{\dagger}(\mathbf{r}) \psi_{\sigma}(\mathbf{r}), \quad (1.19)$$

而另一方面自旋磁矩为

$$\boldsymbol{\mu} = \psi_{\alpha}^{\dagger} \boldsymbol{\sigma}_{\alpha\beta} \psi_{\beta}. \quad (1.20)$$

这样就把三个对外加电磁场的线性响应全部写成二次量子化的形式了；我们可以用推迟格林函数计算出有关的响应大小。

由线性响应理论，我们有

$$\begin{aligned} \langle J_i \rangle_A(t) &= \langle J_i \rangle_0 + i \int dt' \int d^3\mathbf{r}' \theta(t-t') \langle [J_i(\mathbf{r}, t), J_j(\mathbf{r}', t')] \rangle A_j(\mathbf{r}', t') \\ &= -\frac{e^2}{m} \langle n_e \rangle A_i + i \int dt' \int d^3\mathbf{r}' \theta(t-t') \langle [j_i(\mathbf{r}, t), j_j(\mathbf{r}', t')] \rangle A_j(\mathbf{r}', t'). \end{aligned}$$

这里的 i, j 为维度脚标，并不表示粒子编号，且使用爱因斯坦求和约定；下标 A 表示有外场 \mathbf{A} 时的期望值； j_i 的无外场期望是零，这是对称性的结果。

实际上，电阻率定义为²

$$J_i(\mathbf{r}, t) = \int d^3\mathbf{r}' \int_{-\infty}^t \sigma_{ij}(\mathbf{r} - \mathbf{r}', t - t') E_j(\mathbf{r}', t'), \quad (1.21)$$

于是为了避免繁琐的时间上的微积分我们切换到频域上，有

现在我们回到二次量子化的框架下，考虑怎么计算有关的推迟格林函数。

² $\mathbf{A}(\mathbf{r}, t)$ 完全可以不是时间、空间平移不变的，但是既然我们将 \mathbf{A} 当成扰动，只需要无扰动的系统的动力学时间、空间平移不变即可。

在虚时间路径积分中不能简单地将(1.11)（从而，(1.17)）做勒让德变换。在这两个哈密顿量中有一个 $U(1)$ 规范场 (φ, \mathbf{A}) ，由于规范对称性的存在，实际上只有 \mathbf{A} 是独立的自由度，因此不能保证在Wick转动中含有 φ 的项是不变的。最简单的从(1.11)推导出虚时间路径积分的方法是求最小耦合。我们知道 $\varphi\rho$ 项会出现本质上是因为加入 $U(1)$ 规范场之后需要将导数替换为协变导数，而可以验证以下替换

$$\partial_\tau \longrightarrow \partial_\tau - ie\varphi, \quad -i\nabla \longrightarrow -i\nabla + e\mathbf{A} \quad (1.22)$$

给出了虚时间场论中的协变导数，于是在自由理论中做这个替换就得到了与电磁场发生相互作用的电子场的虚时间配分函数。

这样会带来一个疑难，就是Wick转动后电势前面多出来了一个负号；但是凝聚态理论中电子可以被放在一个势场中，显然Wick转动后势场前面不需要多出来一个负号。这个疑难的解答是，含有 $ie\varphi\rho$ 项的理论描述了一个电子场和一个电磁场的耦合，而“电子置于势场中”的模型中电磁场已经被积掉了。如果将电磁场积掉，按照高斯积分的原理， $e\varphi\rho$ 项前面会多一个负号，并且要乘以系数 i 的平方，于是我们发现Wick转动后前面不需要多出来一个负号的势场出现了。

在以上的推导中，我们均取 $e > 0$ 为正的元电荷；在分析电子系统时，还有一种记号是取 $e < 0$ ，令 $|e|$ 为元电荷。这样可以让电子系统的配分函数看起来更像是直接从 $U(1)$ 规范不变性得到的（即 \mathbf{p} 替换成 $\mathbf{p} - e\mathbf{A}$ ），从而看起来更加接近高能物理中的记号。此处我们采用第一种记号，认为 $e > 0$ 。

1.2 晶格形成

所谓晶体指的是一种在三个独立的空间方向上具有离散的平移不变性且并没有连续平移不变性的物体。(1.1)显然具有连续的平移不变性，因此晶体的形成必然经历了对称性自发破缺，且在较高的能量下原本的晶体一定会相变成某种更加均匀的东西。本章将展示组成晶体的离子实是具体如何组成晶体的。原则上可以有哪些晶体见第二章。

将固体分解为组成它的组件所需的能量称为内聚能。造成内聚的相互作用方式包括：

- 离子键，离子之间的库伦引力。
- 共价键，相对局域的电子云的重叠导致的等效原子间吸引力。
- 金属键，离域电子组成大范围的电子气，导致等效的原子间吸引力。此时静电屏蔽非常强，从而正电荷可以看成均匀的“凝胶”。很直观地，晶格中原子排列的具体形式没有特别的要求。
- 范德华力，一种分子之间较弱的相互作用力，键能很低，大体上可以分为取向力、诱导力、色散力。
- 氢键，几乎裸露的氢原子和其它原子的吸引力，比范德华力强但比真正的化学键弱。

这些吸引相互作用不能让原子无限制地靠近。这包括两个原因。首先，当然是因为原子核带正电，彼此之间存在排斥，正如氢分子的形成时，电子成键能够降低能量，而原子核靠近会增高能量。然而对稍微大一些的原子，内层电子的屏蔽就很明显，此时阻止原子靠近的主要因素是电子距离足够近时的泡利排斥能。举一个极端的例子：设两个氢原子核距离非常接近，此时1s轨道上似乎应该有四个电子。如果电子不遵循费米统计，多电子态就是

单电子波函数的直积，那么的确电子-电子库伦排斥会增强，但是电子介导的原子核之间的吸引相互作用也会增强。然而，由于费米统计，实际上两个电子要出现在高于1s轨道的轨道上，因此能量大大增大，并且没有什么能够弥补这种能量增大。即使没有电子-电子库伦排斥，费米统计也会导致原子核接近时体系能量增大，这就是泡利排斥能。

可以使用分子轨道理论近似地分析共价键：设有两个（可以不一样的）原子A和B。我们在两个原子周围各自取一个单原子电子波函数（即所谓原子轨道），然后将来自另一个原子的库伦吸引力当成微扰。如果作为出发点的两个原子轨道能量差别不大，那么考虑了来自另一个原子的相互作用之后，将会出现显著的能级劈裂。较低的那个能级就是成键轨道，而较高的那个能级就是反键轨道，它们是彼此正交的、原来两个原子轨道的线性组合。最后，引入电子-电子相互作用，但是仍然假定多电子波函数近似为两个单电子波函数的直积（这就是分子轨道一说的来历：我们假定系统中仍然有定义良好的单电子轨道，尽管此时这个轨道遍布整个分子）。此时电子-电子相互作用导致的一阶能量修正包括一项密度-密度排斥力和一项交换相互作用。前者会让成键轨道上的电子相互排斥，然而无论怎么放置这两个电子，都会有互相排斥；后一项在成键轨道上有两个自旋相反的电子时却会让两个电子相互吸引。如果总的电子-电子相互作用导致的修正没有大过成键轨道能量降低的量，那么一个共价键就形成了：因为电子放在成键轨道上能够降低总能量。直观地看，此时成键轨道上的两个电子的电子云互相重叠，由于没有特别紧密地重叠，库伦排斥并不是特别大，而两个原子中间的这团电子云同时受到两边的原子的库伦吸引，因此就在两个原子之间建立了明显的等效吸引力。反之，如果一开始的两个原子轨道能量相差很大，那么成键轨道和反键轨道和原来的两个原子轨道没有什么差别，电子放在成键轨道上不能降低多少能量，此时也不会有等效的强烈的原子-原子相互作用，也就没有共价键。

金属键弥散到整个金属晶体中，原子-原子间的等效吸引力没有特别明显的方向性之类，原子排列得越密集，库伦能越低。这就是金属中密排结构特别常见的原因。但是另外一方面，原子密排又会导致电子动量增大，以及泡利不相容原理导致的排斥。

关于如何为原子和分子间的各种相互作用建模，见流体和软物质物理笔记中的第2.2节。

1.3 固体系统的表征

表征一个固体系统通常可以使用的方法，或者说固体系统常见的可测量量（这里不是指可观察量算符，而是指真的做量子测量能够测出来的物理量，比如说期望，等等），包括导电性、磁性、对电磁波的响应、力学和热学性质。

1.3.1 热学和固体力学

热学性质——从而固体力学性质——经常可以使用自由粒子的模型来估计。一种无相互作用的粒子的哈密顿量是完全对角化的：

$$H = \sum_{\mathbf{k}, \sigma} \omega_{\mathbf{k}\sigma} \left(n_{\mathbf{k}\sigma} + \frac{1}{2} \right),$$

于是对配分函数的贡献为

$$Z = \sum_{E_i} e^{-\beta E_i} = e^{-\beta E_{\text{eq}}} \prod_{\mathbf{k}, \sigma} e^{-\beta \frac{\omega_{\mathbf{k}\sigma}}{2}} \sum_{n_{\mathbf{k}, \sigma}} e^{-\beta \omega_{\mathbf{k}\sigma} n_{\mathbf{k}\sigma}}, \quad (1.23)$$

因此这种粒子对内能的贡献为

$$U = E_{\text{eq}} + \sum_{\mathbf{k}, \sigma} \left(\frac{1}{2} + \frac{1}{e^{\beta \omega_{\mathbf{k}\sigma}} - 1} \right) \omega_{\mathbf{k}\sigma}. \quad (1.24)$$

这样，等容热容为

$$C_V = \left(\frac{\partial U}{\partial V} \right)_V = \sum_{\mathbf{k}, \sigma} \left(\frac{\omega_{\mathbf{k}\sigma}}{T} \right)^2 \frac{e^{\omega_{\mathbf{k}\sigma}/T}}{(e^{\omega_{\mathbf{k}\sigma}/T} - 1)^2}. \quad (1.25)$$

每种粒子都会对热容有一个贡献，相互作用会修正这个贡献。

1.3.2 电磁响应和输运

如果一种粒子携带某种荷，那么还可以观察到输运。大体上输运过程可以分成两种，一种是粒子平均自由程远小于系统尺寸的过程，此时系统内部足以出现荷的梯度，输运流量和这种梯度有关，可以称为扩散；另一种是粒子平均自由程大于系统尺寸的过程，此时系统内部的粒子基本上不受到任何阻碍，输运流量由系统边界上的性质确定，即所谓弹道输运。

输运经常是依靠外加电磁场而产生的。由于 \mathbf{A} 耦合在 \mathbf{j} 上，而能够决定极化电场的也无非是电子数密度，讨论电子气的电磁响应基本上就是要计算密度-密度关联函数。如果不考虑电子之间的库仑相互作用，那么这是平凡的：诸如“谐振子受到电场策动，然后出现极化”的图像就能工作得很好。在绝缘体中这大体上是正确的，在金属中则不见得是这样。因此，比较精确地计算电子气的电磁响应，实际上就是需要将电子-电子库仑相互作用纳入计算，而这就很自然地要求我们使用一种完整的凝聚态场论。

第二章 理想晶体的晶格结构

本章讨论理想晶体，即可以认为是无限大，以至于其表面的情况几乎不会影响内部电子运动，且各个原子永远位于平衡位置的晶体。

2.1 晶体的几何形状

2.1.1 晶格、晶胞和格点坐标

低能下离子实自发地排成了比较规则的序列，从而虽然晶体服从的物理规律实际上确实是连续平移不变的，近似定律(1.5)却由于离子实排列成了空间重复的序列而只有离散平移不变性而没有连续平移不变性。这种离子实周期性排列形成的结构可以写成两部分，一部分是若干个离子的平衡位置，或者称为基元，一部分是基元的周期性排列方式或者说点阵 (lattice)，称为晶格。基元——或者说全同的结构单元——占据的空间称为原胞 (primitive cell)——是晶体中体积最小的重复性成分。实际的晶格会因为振动热偏离上述周期结构，但是本节暂时不讨论这一点，而将其留到关于声子的讨论中。

在晶格中只要知道了某个格点的位置，就可以计算出其它任何格点的位置。或者，如果知道了某个原胞中某个原子的位置，就可以知道其它任何一个原胞中同种原子的位置。任意两个格点或是同种原子之间的位置矢量形如

$$\mathbf{R}_i = i_1 \mathbf{a}_1 + i_2 \mathbf{a}_2 + i_3 \mathbf{a}_3, \quad \mathbf{i} = (i_1, i_2, i_3) \in \mathbb{N}^3. \quad (2.1)$$

这些位置矢量构造了一架三维网格，这个网格称为布拉伐格子，这些矢量称为布拉伐格矢， $\{\mathbf{a}_i\}$ 称为晶格常数， n 称为格点坐标。（也有很多人将 \mathbf{a}_i 称为格矢，在我们的术语中 \mathbf{a}_i 是“基元格矢”）布拉伐格子中的所有格点周围环境相同，无重叠无遗漏地覆盖整个空间。在每个格子上放置一个基元，即可构造出整个晶体。设基元中有 n 种原子，我们任取其中一种原子而忽略其它原子，则被选中的这种原子本身也组成一个晶格，且这个晶格和整体的晶格是完全一样的。 $n = 1$ 的情况称为简单晶格，而 $n > 1$ 的情况称为复式晶格，因为它实际上是 n 个简单晶格套在一起而得到的。相当神奇的是，即使晶体中只有一种原子，可能仍然无法分割出一个只含有一个原子的原胞，换言之，此时晶体中的物理上完全一样的原子在几何上仍然可以进一步分成若干类，从而晶格是复式晶格。要看出为什么会有这种情况出现，只需要想象取一个普通的复式晶格晶体，然后将其中所有的原子都替换成同种原子即可。一些重要的晶体如石墨烯就具有这种性质，虽然物理上只有一种原子，却仍然是复式晶格。

布拉伐格子的原胞有许多划分方法。可以以 $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ 张成的平行六面体为一个原胞，称为初基原胞。另一种原胞是维格纳-赛兹原胞，它是空间中与某个特定格点的距离小于与任何其它格点的距离的点的轨迹，或者等价地说，它是某个特定格点与相邻格点的连线的

垂直平分面包围出的立体。任意一个空间矢量都可以写成某种原胞中的一个矢量加上一个布拉伐格矢，这可以使用非常直观的方式证明。

原胞有时很难直观地展示晶体的特征。晶体学原胞或者说单胞指的是最大限度反映晶格对称性的最小单元。它是重复性的单元，因此它应该包含整数倍的原胞。原胞中原子位置可以使用原胞基矢，当然也可以使用单胞基矢。例如，面心立方格子的原胞乍一看就是一个形状奇怪的平行六面体，而其单胞则是非常直观的“面心立方格子”。确定一个单胞里面有几个原胞，可以通过计算一个单胞中有几个格点（实际上就是有几个基元）来完成。

由于以下用到晶体离散对称性的地方几乎从来不会用到“原胞是最小的”这一事实，很多时候把“原胞”一词替换成“单胞”是完全可以的。于是我们就模糊地说晶胞（unit cell），即晶体中不需要最小的重复周期。晶体中各个晶胞的坐标同样组成一架周期性格子，和原胞组成的格子的对称性相同。

2.1.2 晶体中的坐标系和方向

在获得了一个平行六面体晶胞之后，可以用确定这个平行六面体的三个矢量作为基矢量，建立一个坐标系。我们通常用 $\mathbf{a}, \mathbf{b}, \mathbf{c}$ 或 $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ 标记这三个矢量，用 α, β, γ 标记它们的夹角，从而确定这个平行六面体的几何形状。

以 \mathbf{a}_i 为基矢量可以得到矢量分量 l^i ，在它们全部都是整数时，它们是一个正格矢在 $\{\mathbf{a}_i\}$ 下的分量。这个正格矢的指向——即所谓晶向——可以用彼此互质的 $[l^1 l^2 l^3]$ 表示，相应的 l^1, l^2, l^3 称为晶向指数。可以通过对称性操作（见后文）彼此转化的晶向使用 $\langle l^1 l^2 l^3 \rangle$ 表示。

晶体中三个不同格点连接而成的平面称为晶面。将晶面在三个 \mathbf{a}_i 方向上的截距（都是整数）的倒数乘以适当的因子，让它们成为互质的整数，用圆括号括起来，就是晶面的米勒指数。米勒指数 $(h_1 h_2 h_3)$ 对应一族彼此平行且平行于

$$h_1 x^1 + h_2 x^2 + h_3 x^3 = 1 \quad (2.2)$$

的晶面，其中 x^i 是 \mathbf{r} 在基矢量 $\{\mathbf{a}_i\}$ 上的分量。

2.1.3 倒格子

各个原胞（我们通常使用原胞而不是单胞或者别的类型的晶胞构造正格子；无论如何正格子的原胞大小总是和用于构造正格子的晶胞大小一样的；以下我们将“晶胞”一词局限在“实际晶体的一个周期（未必是最小周期）”这个意义上；在讨论格点——无论是正格子还是倒格子——时我们只用“原胞”）所在的 \mathbf{R}_n 构成的空间网格（以下称为正格子）上的所有可观察物理量均具有和布拉伐格矢一样的对称性，也即，它们在三个方向上以 $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ 为周期。回顾傅里叶级数的公式，我们有

$$f(x) = \frac{1}{T} \sum_{m=-\infty}^{\infty} e^{i \frac{2\pi m x}{T}} \left(\int dt f(t) e^{-i \frac{2\pi m t}{T}} \right),$$

其三维形式就是

$$f(\mathbf{r}) = \frac{1}{V} \sum_{m_1, m_2, m_3=-\infty}^{\infty} e^{i \mathbf{G}_m \cdot \mathbf{r}} \int_V d^3 \mathbf{r}' f(\mathbf{r}') e^{-i \mathbf{G}_m \cdot \mathbf{r}}, \quad \mathbf{G}_m \cdot \mathbf{a}_i = 2\pi N_i, \quad N_i \in \mathbb{N},$$

其中 V 指的是正格子原胞的大小。 \mathbf{G}_m 满足的条件等价于，对任意的布拉伐格矢都有

$$\mathbf{G}_m \cdot \mathbf{R}_i = 2\pi N, \quad N \in \mathbb{N}, \quad (2.3)$$

这又等价于,

$$\mathbf{G}_m = m_1 \mathbf{b}_1 + m_2 \mathbf{b}_2 + m_3 \mathbf{b}_3, \quad \mathbf{a}_i \cdot \mathbf{b}_j = 2\pi \delta_{ij}. \quad (2.4)$$

因此诸 \mathbf{G}_m 也构成一个布拉伐格子, 我们称它为倒格子, 与正格子相区分, 同样, 称 \mathbf{r} 所在的三维空间为实空间, \mathbf{G} 所在的空间为倒空间. 倒格子的基矢量和正格子的基矢量互为共轭基矢量. 实际上我们应该把 \mathbf{b} 记为 \mathbf{b}^i . 两种格子的基矢量可以通过下式

$$\frac{1}{2\pi} \mathbf{b}_1 = \frac{\mathbf{a}_2 \times \mathbf{a}_3}{\mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)} \quad (2.5)$$

及其轮换相互换算. 在写出倒格子的显式表达式之后, 晶体中具有正格子的周期性的物理量的傅里叶变换就是

$$F(\mathbf{r}) = \sum_{\mathbf{g}} \tilde{F}(\mathbf{g}) e^{i\mathbf{g} \cdot \mathbf{r}}, \quad (2.6)$$

其中 \mathbf{g} 是布拉伐格矢, 且

$$\tilde{F}(\mathbf{g}) = \frac{1}{V} \int_V d^3\mathbf{r} F(\mathbf{r}) e^{-i\mathbf{g} \cdot \mathbf{r}}. \quad (2.7)$$

倒格子的布拉伐格子可以和正格子同一类型, 但是也可以不一样. 但是, 倒格子和正格子的最高点群对称性 (也即, 不考虑基元, 仅考虑格子本身, 或者只讨论球对称的基元) 一定是一样的. 这是因为设 α 是一个点群对称性操作, 则

$$\alpha^{-1} \mathbf{R}_i = \mathbf{R}_i,$$

设 \mathbf{G}_m 是倒格矢, 则对任意 \mathbf{R}_i 都有

$$\mathbf{G}_m \cdot \mathbf{R}_i = 2\pi N, \quad N \in \mathbb{N},$$

由于 $\alpha^{-1} \mathbf{R}_i$ 也是正格矢, 有

$$\mathbf{G}_m \cdot (\alpha^{-1} \mathbf{R}_i) = 2\pi N = (\alpha \mathbf{G}_m) \cdot \mathbf{R}_i, \quad N \in \mathbb{N},$$

因此 $\alpha \mathbf{G}_m$ 也是倒格矢. 因此凡是正格子有的点群操作, 倒格子也有. 反过来也能够证明凡是倒格子有的点群操作, 正格子也有. 因此两者的最高点群对称性是一样的.

倒格子当然也有原胞的概念. 倒格子的维格纳-赛兹原胞称为第一布里渊区, 相应的, 某格点和它所有次近邻格点的垂直平分面包围成的区域称为第二布里渊区, 等等. 只要给定一种倒格子, 其第一布里渊区就是完全确定的, 而和怎样选择倒格子的初基格矢没有关系.

倒格子的原胞实际上是定义在正格子上的函数的傅里叶变换的动量取值范围. 使用 $\mathbf{i} = (i_1, i_2, i_3)$ 表示格点坐标, 则

$$\frac{1}{N} \sum_{\mathbf{R}_i} e^{i(\mathbf{k}-\mathbf{k}') \cdot \mathbf{R}_i} = \sum_{\mathbf{g}} \delta_{\mathbf{k}-\mathbf{k}'+\mathbf{g}}, \quad (2.8)$$

其中 \mathbf{R}_i 指的是 \mathbf{i} 对应的位矢, N 是晶格中总离子数, \mathbf{g} 扫过整个倒格子. 可以看到方程右边是周期性的, 如果限制 \mathbf{k} 在一个倒空间原胞中, 那么就有非常简单的形式:

$$\frac{1}{N} \sum_{\mathbf{R}_i} e^{i(\mathbf{k}-\mathbf{k}') \cdot \mathbf{R}_i} = \delta_{\mathbf{k}\mathbf{k}'}, \quad (2.9)$$

从而得到与之对偶的

$$\frac{1}{N} \sum_{\mathbf{k}} e^{i(\mathbf{R}_i - \mathbf{R}_j) \cdot \mathbf{k}} = \delta_{ij},$$

其中 \mathbf{k} 扫过第一布里渊区中的所有动量点。这又意味着倒格子的一个原胞内的动量取值数目可以认为是 N 个，当然这是正确的，因为实空间中的 N 点离散信号做离散傅里叶变换之后会得到倒空间中的周期性离散信号，其周期正好是 N ，一个原胞正好是一个周期。

并非所有晶体中的物理量具有正格子的周期性（有的周期性强于正格子，有的也许弱于正格子），它们的傅里叶变换中的动量不局限在倒格子上。

倒格矢也可以用于表示晶面。米勒指数为 $(h^1 \ h^2 \ h^3)$ 的晶面的法向量为

$$\nabla(h_1 x^1 + h_2 x^2 + h_3 x^3) = h_1 \frac{1}{2\pi} \mathbf{b}^1 + h_2 \frac{1}{2\pi} \mathbf{b}^2 + h_3 \frac{1}{2\pi} \mathbf{b}^3,$$

正好和分量为 h_1, h_2, h_3 的倒格矢平行。实际上这给出了一种用米勒指数表示倒格矢的方式： $(h_1 \ h_2 \ h_3)$ 表示分量为 h_1, h_2, h_3 的倒格矢。能够验证米勒指数为 $(h^1 \ h^2 \ h^3)$ 的晶面方程为

$$(h_1 \mathbf{b}_1 + h_2 \mathbf{b}_2 + h_3 \mathbf{b}_3) \cdot \mathbf{r} = 2\pi n, \quad n \in \mathbb{Z}, \quad (2.10)$$

这又意味着，设 d 为两个晶面的间距，则有

$$|\mathbf{G}_{h_1 h_2 h_3}|d = 2\pi n, \quad n \in \mathbb{Z}, \quad (2.11)$$

因此 $(h^1 \ h^2 \ h^3)$ 方向上长度最短的倒格矢的长度是

$$G_0 = \frac{2\pi}{d}, \quad (2.12)$$

d 是 $(h^1 \ h^2 \ h^3)$ 晶面族中相邻晶面的间距。 $(h^1 \ h^2 \ h^3)$ 方向上的其它倒格矢的长度都是上式的整数倍。

2.1.4 有限大小的晶体

晶格对电子的吸引比较明显，因此电子自发溢出晶格的概率并不大，从而可以将晶格表面看成一个势阱。晶格表面的形状以及势阱的高度无疑会影响电子气的行为，但由于晶体非常大，这种影响对稍微远离表面的电子都是非常微弱的。（接近表面的电子可能参与表面态，此时关于晶体表面的信息就非常重要了）因此我们认为晶体是长宽高各为 L 的大正方体， L 相对电子、原子的尺度都是非常大的；同时我们简单地施加一个周期性边界条件来表示势阱的存在，即认为

$$\psi(\mathbf{r}) = \psi(\mathbf{r} + L\mathbf{e}_i), \quad i = 1, 2, 3. \quad (2.13)$$

同时我们暂时忽略在晶体外找到电子的概率，因为它相对于在晶体内部找到电子的概率是非常小的。

由于晶体是有限大小的， \mathbf{k} 的取值是离散化的，因为波函数必须满足(2.13)，为了尽可能让 u 容纳较多信息，我们用 \mathbf{k} 来满足这个要求，即

$$e^{i\mathbf{k} \cdot \mathbf{r}} = e^{i\mathbf{k} \cdot (\mathbf{r} + L\mathbf{e}_i)}, \quad i = 1, 2, 3,$$

这样 \mathbf{k} 的取值范围就是一个晶格常数为 $2\pi/L$ 的三维点阵。¹这个三维点阵正是局限在晶体内部的任何函数做空间傅里叶变换得到的波矢的取值范围，且有如下归一化条件：

$$\frac{1}{V} \int d^3\mathbf{r} e^{i(\mathbf{k} - \mathbf{k}') \cdot \mathbf{r}} = \delta_{\mathbf{k}\mathbf{k}'}. \quad (2.14)$$



图 2.1: 三种动量空间，倒格子（黑色）和由于晶体是有限大小的而产生的晶格常数为 $2\pi/L$ 的离散动量点阵（灰色），两者均可以无限延伸；第一布里渊区用淡黄色标记，它的边界垂直平分倒格矢包括诸 \mathbf{b}_i 以及用紫色标出的倒格矢

2.1.5 三种动量空间，它们中的动量本征态和对应的傅里叶变换

现在我们有三种动量空间，这三种动量空间、它们适用于什么样的函数以及它们对应的傅里叶变换列举如下。在这里唯一确定的是坐标空间的场算符 $\psi(\mathbf{r})$ ，它满足

$$[\psi_\alpha(\mathbf{r}), \psi_\beta^\dagger(\mathbf{r}')] = \delta_{\alpha\beta} \delta(\mathbf{r} - \mathbf{r}').$$

实际上由于我们局限在一个大小为 V 的盒子中，这里的 $\delta(\mathbf{r} - \mathbf{r}')$ 函数并非全空间中的 δ 函数，而是实空间区域 V 中的 δ 函数，它能够让任何限制在 V 中的函数 $f(\mathbf{r})$ 满足

$$\int_V d^3\mathbf{r}' f(\mathbf{r}') \delta(\mathbf{r} - \mathbf{r}') = f(\mathbf{r}).$$

但显然 $V \rightarrow \infty$ 时这种“软化”的 $\delta(\mathbf{r} - \mathbf{r}')$ 就趋于无限大空间中的 $\delta(\mathbf{r} - \mathbf{r}')$ ，从而以下我们可以混用两种 $\delta(\mathbf{r} - \mathbf{r}')$ 。动量表象的归一化系数是不确定的。我们需要定义电子波函数的内积，由于 $\langle \psi | \psi \rangle = 1$ ，每种内积定义都会在波函数前面引入一个不同的归一化因子。归一化是很重要的，因为诸如微扰论给出的能级修正等在归一化不对时是不正确的；此外一些物理量真的会正比于 V 或者说 N ，从而归一化常数中的 V 或是 N 是很重要的。

1. 晶格常数为 $2\pi/L$ 的三维点阵是被限制在边长为 L 的实空间方盒中的函数的傅里叶变换的动量取值范围，由于 L 通常很大，这个点阵要比倒格子密得多，在倒格子的尺度上工作时没有必要考虑晶格常数为 $2\pi/L$ 的三维点阵是离散的这一事实。该三维点阵动量空间虽然是离散的，却是无限大的，因为它实际上给出的是一个箱子中所有可能的平面波的波矢，而这当然可以取到无限大。与这个动量空间相关的傅里叶变换的归一化表达式为

$$\frac{1}{V} \int_V d^3\mathbf{r} e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}} = \delta_{\mathbf{k}\mathbf{k}'}, \quad \frac{1}{V} \sum_{\mathbf{k}} e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')} = \delta(\mathbf{r} - \mathbf{r}'), \quad (2.15)$$

¹ 这个点阵不是倒格子；倒格子的晶格常数和实际的物理结构——也就是晶格的结构——有关，而此处的点阵的晶格常数完全是我们强加的，且总是趋于零，使得格点动量看起来几乎是连续的，因此可以被划分成连续的布里渊区，等等。

第二个表达式中的 δ 函数就是动量空间无限大的一个结果。于是有以下几种方案：一种是将动量本征态取为

$$\psi_{\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{V}} e^{i\mathbf{k}\cdot\mathbf{r}}, \quad (2.16)$$

则容易验证我们应该定义内积为

$$\langle \alpha | \beta \rangle = \int d^3\mathbf{r} \langle \alpha | \mathbf{r} \rangle \langle \mathbf{r} | \beta \rangle = \sum_{\mathbf{k}} \langle \alpha | \mathbf{k} \rangle \langle \mathbf{k} | \beta \rangle, \quad (2.17)$$

这样可以满足归一化条件。此时表象变换为

$$\langle \mathbf{k} | \alpha \rangle = \int d^3\mathbf{r} \frac{1}{\sqrt{V}} e^{-i\mathbf{k}\cdot\mathbf{r}} \langle \mathbf{r} | \alpha \rangle, \quad \langle \mathbf{r} | \alpha \rangle = \sum_{\mathbf{k}} \frac{1}{\sqrt{V}} e^{i\mathbf{k}\cdot\mathbf{r}} \langle \mathbf{k} | \alpha \rangle$$

于是通常如此定义傅里叶变换：

$$f(\mathbf{k}) = \frac{1}{\sqrt{V}} \int d^3\mathbf{r} e^{-i\mathbf{k}\cdot\mathbf{r}} f(\mathbf{r}), \quad f(\mathbf{r}) = \frac{1}{\sqrt{V}} \sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}} f(\mathbf{k}), \quad (2.18)$$

这样定义的好处在于，我们知道 $\psi^\dagger(\mathbf{r})$ 和 $c_{\mathbf{k}}^\dagger$ 彼此为傅里叶变换，而如果采用上式的傅里叶变换定义，那么从

$$\{\psi_\alpha(\mathbf{r}), \psi_\beta^\dagger(\mathbf{r}')\} = \delta(\mathbf{r} - \mathbf{r}') \delta_{\alpha\beta}$$

可以得到

$$\{c_{\mathbf{k}\alpha}, c_{\mathbf{k}'\beta}^\dagger\} = \delta_{\mathbf{k}\mathbf{k}'} \delta_{\alpha\beta},$$

于是动量表象下的产生湮灭算符关系也是非常简单的。

另一种方案是取动量本征态为

$$\psi_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}, \quad (2.19)$$

此时内积应该定义为

$$\langle \alpha | \beta \rangle = \frac{1}{V} \int d^3\mathbf{r} \langle \alpha | \mathbf{r} \rangle \langle \mathbf{r} | \beta \rangle = \frac{1}{V} \sum_{\mathbf{k}} \langle \alpha | \mathbf{k} \rangle \langle \mathbf{k} | \beta \rangle, \quad (2.20)$$

这样的好处在于，在 V 很大时，考虑到点阵的晶格常数为 $2\pi/L$ ，可以做替换

$$\frac{1}{V} \sum_{\mathbf{k}} \longrightarrow \int \frac{d^3\mathbf{k}}{(2\pi)^3}, \quad (2.21)$$

从而 \mathbf{k} 基底下实际上没有 V 的依赖，正如动量本征态的定义所展示的那样。此时表象变换为

$$\langle \mathbf{r} | \beta \rangle = \frac{1}{V} \sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}} \langle \mathbf{k} | \beta \rangle = \int \frac{d^3\mathbf{k}}{(2\pi)^3} e^{i\mathbf{k}\cdot\mathbf{r}} \langle \mathbf{k} | \beta \rangle,$$

从而傅里叶变换应该取为

$$f(\mathbf{r}) = \frac{1}{V} \sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}} f(\mathbf{k}) = \int \frac{d^3\mathbf{k}}{(2\pi)^3} e^{i\mathbf{k}\cdot\mathbf{r}} f(\mathbf{k}). \quad (2.22)$$

此时

$$[c_{\mathbf{k}\alpha}, c_{\mathbf{k}'\beta}^\dagger] = (2\pi)^3 \delta_{\alpha\beta} \delta_{\mathbf{k}\mathbf{k}'}. \quad (2.23)$$

在这种记号下对 $2n$ 个坐标空间下的场算符的乘积积分，所得结果用动量空间中的产生湮灭算符表示则一定正比于 V ，因为到最后那个对 \mathbf{r} 的积分给出

$$\int d^3\mathbf{r} e^{i\mathbf{k}\cdot\mathbf{r}} \propto V.$$

相反，在(2.18)的记号下，对 $2n$ 个坐标空间下的场算符的乘积积分，所得结果用动量空间中的产生湮灭算符表示则正比于 V^{1-n} 。(2.18)相当于将体积依赖放到了与动量空间有关的物理量（比如说 $\psi_{\mathbf{k}}$ ）中，而(2.22)相当于将体积依赖放到了与坐标空间有关的物理量（主要是对 $d^3\mathbf{r}$ 积分时总是需要乘以 V ）中。我们通常采用第一种记号，也就是(2.18)，因为在固体物理中实空间的物理定律——如库仑定律——是保证和自由情况下完全一样的。例如我们时常需要从库仑相互作用获得某种等效相互作用通道，那么使用记号(2.18)的话就不需要担心归一化出错。

有时候，我们不使用平面波基底展开波函数，但是这并不影响 \mathbf{r} ，即内积定义(2.17)和(2.20)中涉及坐标的部分在任何时候都是可以用的。我们通常还是使用(2.17)。

顺带一提：这种傅里叶变换实际上就是虚时间场论中时域-频域的傅里叶变换，即

$$\psi(\tau) = \frac{1}{\sqrt{\beta}} \sum_{\omega_n} \psi(i\omega_n) e^{-i\omega_n \tau}, \quad \psi(i\omega_n) = \frac{1}{\sqrt{\beta}} \int_0^\beta d\tau e^{i\omega_n \tau} \psi(\tau). \quad (2.24)$$

2. 倒格子是具有和晶格一样的周期性的连续函数的傅里叶变换的动量取值范围，即通常所谓的 \mathbf{G} 格子。倒格子中，最近邻格点的间距和 V 是无关的。与之相关的归一化表达式为

$$\frac{1}{V_{\text{u.c.}}} \int_{V_{\text{u.c.}}} d^3\mathbf{r} e^{i\mathbf{r}\cdot(\mathbf{k}-\mathbf{k}')} = \delta_{\mathbf{k}\mathbf{k}'}, \quad \frac{1}{V_{\text{u.c.}}} \sum_{\mathbf{k}} e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{r}')} = \delta(\mathbf{r}-\mathbf{r}'), \quad (2.25)$$

其中 \mathbf{k} 取遍整个倒格子而 \mathbf{r} 取遍一个正格子晶胞内部的所有点（因为仅讨论周期函数）。这里我们使用 $V_{\text{u.c.}}$ 表示一个正格子晶胞的大小，来和整块晶体的大小区分开。相应的，傅里叶变换为

$$f(\mathbf{k}) = \frac{1}{\sqrt{V_{\text{u.c.}}}} \int d^3\mathbf{r} e^{-i\mathbf{k}\cdot\mathbf{r}} f(\mathbf{r}), \quad f(\mathbf{r}) = \frac{1}{\sqrt{V_{\text{u.c.}}}} \sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}} f(\mathbf{k}). \quad (2.26)$$

倒格子是无限大的，因此以上展示的傅里叶变换的动量空间同样是无限大的。在很多文献中，会使用 \mathbf{G} 来表示倒格子上的矢量。倒格子对应的实空间自由度是一个晶胞内部的位置矢量。

3. 倒空间的原胞，如第一布里渊区，是定义在正格子上的函数的傅里叶变换的动量取值范围。第一布里渊区内的动量对应的实空间自由度是正格子的格矢，即晶胞以外的位置矢量。归一化表达式为

$$\frac{1}{N} \sum_{\mathbf{R}_i} e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{R}_i} = \delta_{\mathbf{k}\mathbf{k}'}, \quad \frac{1}{N} \sum_{\mathbf{k}} e^{i(\mathbf{R}_i-\mathbf{R}_k)\cdot\mathbf{k}} = \delta_{ij}. \quad (2.27)$$

第二个表达式表明对定义在正格子上的函数，它在动量空间中在一个倒空间原胞中的动量取值有 N 个。在 \mathbf{k} 未必位于第一布里渊区时，有

$$\frac{1}{N} \sum_{\mathbf{R}_i} e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{R}_i} = \sum_{\mathbf{G}_m} \delta_{\mathbf{k}', \mathbf{k}+\mathbf{G}_m}. \quad (2.28)$$

傅里叶变换为

$$f(\mathbf{k}) = \frac{1}{\sqrt{N}} \sum_{\mathbf{i}} e^{-i\mathbf{k} \cdot \mathbf{a}_i} f(\mathbf{i}), \quad f(\mathbf{i}) = \frac{1}{\sqrt{N}} \sum_{\mathbf{k}} e^{i\mathbf{k} \cdot \mathbf{R}_i} f(\mathbf{k}), \quad (2.29)$$

即将动量本征态取为

$$\psi_{\mathbf{k}}(\mathbf{i}) = \frac{1}{\sqrt{N}} e^{i\mathbf{k} \cdot \mathbf{R}_i}, \quad (2.30)$$

从而内积应该定义为

$$\langle \alpha | \beta \rangle = \sum_{\mathbf{i}} \langle \alpha | \mathbf{i} \rangle \langle \mathbf{i} | \beta \rangle = \sum_{\mathbf{k}} \langle \alpha | \mathbf{k} \rangle \langle \mathbf{k} | \beta \rangle \quad (2.31)$$

以满足归一化条件。使用这种记号和(2.18)类似，一个对 $2n$ 个坐标空间中的场算符乘积求和的项如果改用 $c_{\mathbf{k}}$ 表示，将正比于 N^{1-n} 。与前两种情况不同，上式中的动量空间中只有有限个（共 N 个）动量。实际上容易验证，这 N 个动量恰恰是一个倒空间原胞中容纳的全部来自晶格常数为 $2\pi/L$ 的点阵的格点。这样，我们就有

$$N = \frac{2\pi/V_{\text{u.c.}}}{2\pi/V},$$

即

$$V = NV_{\text{u.c.}}, \quad (2.32)$$

正好是预期中的结果。在热力学极限下 $N \rightarrow \infty$ ，此时倒空间原胞中容纳的动量取值趋于连续。这就是前述“在倒格子的尺度上工作时无需考虑晶体有限大（即 L 有限大）导致的动量离散化”的一个直观理解。

还有一个应该注意的细节：以上推导中我们都使用了 $\delta_{\mathbf{k}\mathbf{k}'}$ ，一些时候也会把它写成 $\delta(\mathbf{k} - \mathbf{k}')$ ，从而和坐标空间保持一致，但是要注意，如果我们希望 $\delta(\mathbf{k} - \mathbf{k}')$ 满足

$$\int d^3\mathbf{k} \delta(\mathbf{k} - \mathbf{k}') f(\mathbf{k}) = f(\mathbf{k}'),$$

那么其实应该取

$$\delta(\mathbf{k} - \mathbf{k}') = \frac{V}{(2\pi)^3} \delta_{\mathbf{k}\mathbf{k}'}, \quad \text{as } V \rightarrow \infty. \quad (2.33)$$

2.2 晶体的点群、空间群对称性和分类

2.2.1 晶体结构分类概述

从本节开始我们尝试用对称性分类晶体。晶体的对称性是欧氏空间的等距变换群作用在离散格子上的一个子群，其中的任何一个元素可以标记为 $(R|\mathbf{a})$ ，其中 R 表示一个不包括平移的等距变换（有限大小的对象的对称群只包括这种变换，因为有限大小的对象不可能具有平移对称性）， \mathbf{a} 表示平移的距离。

晶体的对称性是非常有用的，例如，它可以直接用于决定晶体的一些参数的形式。例如，对立方晶体，介电张量是一个标量，而对六角晶体介电张量具有

$$\begin{pmatrix} \epsilon_{\parallel} & & \\ & \epsilon_{\perp} & \\ & & \epsilon_{\perp} \end{pmatrix}$$

的形式。晶体分为晶格和基元，而晶格的对称性是具有这种晶格的晶体能够具有的最高对称性（即基元具有球对称性时晶体具有的对称性）——如果基元在某种不改变晶格的操作下会发生改变，那么整个晶体的对称性实际上低于晶格的对称性。例如，金刚石和闪锌矿具有完全一样的晶格，但是金刚石基元具有沿着化学键中心的反射对称性，而闪锌矿则没有。因此，我们可以首先分析晶格的对称性，然后考虑基元的对称性而得到不考虑平移的晶体的对称性，最后加入平移（它是空间群的不变子群）而得到完整的对称性。

我们总结一下所有的分类：

1. **晶系**，这是最为粗略的分类方法，其依据是晶格的对称轴和对称面的数目，共有7种晶系。晶系给定之后晶格的点群（见下文）可以不同，而可以验证，如果两种晶体分别属于不同的晶系，这两种晶体的点群肯定是不一样的。晶系提供了晶格的点群，是具有这种晶格的晶体可以具有的最大点群（如果基元对称性较低，实际的晶体点群会小于晶格的点群）。
2. **布拉伐晶格**，在每种晶系内，按照一个晶体学单胞（默认8个角上都有格点）中各个格点的位置——简单，面心，体心，底心（共三种）——可以细分出总计14种布拉伐格子。²虽然表面上应该有 $7 \cdot 6 = 42$ 种格子，但其中很大一部分是重复的。在晶系给定之后，布拉伐格子额外提供了关于平移对称性的信息。因此，布拉伐格子概括了**晶格**的对称性：给定一个布拉伐格子类型，对应的晶格的空间对称群就可以写下来了。
3. **晶体点群**或者简称**点群**，虽然后者也常被用于指代一个更大的概念，指保持空间中一个点不动的群，实际上就是描述有限大小的物体的对称性的群。一个晶体群（即晶体的完整对称性，见下文）中除了平移以外的部分一定是 $O(3)$ 的子群，可以写成某个坐标系下的一个矩阵群，矩阵变换下坐标原点始终是不动的，因此点群一定会保持一个点不动，这个点就是如前所述的原点；除此以外的其它点在某个群操作下都会发生变动。晶体点群一定是广义上的点群。然而，并非所有广义上的点群都是晶体点群，因为它们可能会和空间平移不兼容。一个晶体的点群一定是它的晶系的点群的子群。晶体点群和它的晶系的点群之间的差异表达了基元有多不对称。在晶系给定之后，晶体点群和布拉伐格子类型（简单、体心、面心、底心）可以任意组合，彼此无关。
4. **晶体群**或者说**晶体空间群**，最为细致的晶体对称性分类，它是欧氏空间 \mathbb{R}^n 的离散等距变换子群，并且 \mathbb{R}/G 是紧致的（即整个空间中任何一个点都可以等效到某个晶胞中）。数学上可以证明这样的群一定包含 n 种彼此独立的平移变换（本该如此，因为我们需要这些平移变换来产生完整的布拉伐格子），具体这些平移变换是什么由布拉伐格子的类型决定。在晶体点群和布拉伐格子均给定之后仍然不能完全确定晶体空间群，因为一些复式晶格的点群操作和平移操作一起使用时，部分点群操作并不在基元的点群中，然而，先做它们然后附带小的平移操作却能够保持晶体不变。实际上，这意味着“点群”的概念宏观和微观来看会有微妙的不同：典型的例子是金刚石和ZnS晶体，两者都是复式晶格，但是金刚石的几何上不等价的两种格点含有一样的原子，因此，虽然两种晶体的微观点群都是 T_d ，对金刚石， O_h 点群中不是 T_d 的那些操作如果与一段很小的平移结合，仍然可以保持晶体不变，因此宏观地看，金刚石似乎在 O_h 点群下也不变。

²角上的格点以外的格点不能位于简单、面心、体心、底心以外的地方，否则我们总是可以通过重新定义基元来把晶体划归到简单、面心、底心中的一类中去。

2.2.2 点群中的变换

我们来分析点群中可能出现的变换。首先肯定有旋转，并且由于我们讨论的是离散对象上的对称群，旋转一定是离散旋转，即绕着某个轴，一次旋转 $2\pi/n$ ， n 为某个整数。这样的旋转操作构成一个点群的子群，记作 C_n ，相应的转轴称为一条 C_n 轴；一个转角为 $2\pi k/n$ 的离散旋转操作记作 c_n^k 。其次肯定还有空间反演或者说反射，这包括镜面反射和点反演。我们用 σ 标记一个反演平面，用 i 标记一个反演点。在给定一条已知的旋转轴之后，将垂直于它的反演平面记作 σ_h ，将包含它的反演平面记作 σ_v ；如果给定两条对称轴而有一个反演平面正好平分这个角，那么这个反演平面就是 σ_d 。这两个符号也用于指代对应的空间反演操作。简单考虑一下会发现这些操作和它们的乘积足够给出所有可能的点群变换。

由于讨论范围局限于一个点群，总是存在一个点，在所有变换下都不变。因此，一个点群内的旋转操作的旋转轴一定有共同的交点，这个交点就是反演中心，反演平面一定包含这个点。

空间反演和反射操作并非简单地被直积在一起——它们之间会有一些特殊的代数关系。例如，很明显在给定一条旋转轴之后， c_n^k ， σ_h 和 σ_v 之间会有某些特殊的代数关系。我们将 z 轴取在旋转轴上，则

$$c_n^k = \begin{pmatrix} \cos\left(\frac{2\pi k}{n}\right) & -\sin\left(\frac{2\pi k}{n}\right) & 0 \\ \sin\left(\frac{2\pi k}{n}\right) & \cos\left(\frac{2\pi k}{n}\right) & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (2.34)$$

而

$$\sigma_h = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad i = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad (2.35)$$

于是立刻可以看出

$$c_n^k \sigma_h = \sigma_h c_n^k. \quad (2.36)$$

在 $n=2$ 时 $C_2 = \{e, c_2^1\}$ ，记 c_2^1 为 c_2 ，则

$$c_2 = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (2.37)$$

计算可得

$$c_2 \sigma_h = \sigma_h c_2 = i, \quad c_2 i = i c_2 = \sigma_h, \quad \sigma_h i = i \sigma_h = c_2. \quad (2.38)$$

假定有两个二阶轴 C_{2A} 和 C_{2B} 交于 O 点，它们之间的夹角为 φ ，则

$$c_{2A} c_{2B} = c(2\varphi), \quad (2.39)$$

其中 $c(2\varphi)$ 的旋转轴垂直于 C_{2A} 和 C_{2B} 确定的平面。这件事只需要通过

$$c_{2B} = c(\varphi) c_{2A} c(-\varphi) \quad (2.40)$$

即可得到证明。类似地可以证明

$$c_{2B} c_{2A} = c(-2\varphi). \quad (2.41)$$

也可以对反射面证明类似的公式。设有两个反射面 σ_ν 和 σ'_ν 相交于 z 轴，以 z 轴为旋转轴，则有

$$\sigma'_\nu \sigma_\nu = c(2\varphi), \quad \sigma_\nu \sigma'_\nu = c(-2\varphi). \quad (2.42)$$

证明方法是，取和 σ_ν 垂直的 C_2 轴 c_2 ，和 σ'_ν 垂直的 C_2 轴 c'_2 ，则有

$$\sigma_\nu = c_2 i = i c_2, \quad \sigma'_\nu = c'_2 i = i c'_2,$$

于是

$$\sigma'_\nu \sigma_\nu = c'_2 c_2 = c(2\varphi).$$

(2.40)其实说明一件事：一个旋转轴在特定的对称操作下可以被转换成另一个旋转轴。我们先讨论这“特定的对称操作”是另一个旋转的情况。设一个 C_n 轴和一个 C_m 轴相交于 O 点，两者夹角为 φ ，则可以获得这样的新的旋转：

$$c_n^j c_m^i (c_n^j)^{-1},$$

因此通过绕着 C_n 轴的旋转可以产生 n 个 C_m 轴，记作 C_{mj} ，其中的操作为

$$c_{mj}^i = c_n^j c_m^i (c_n^j)^{-1}. \quad (2.43)$$

特别的，如果有一个 C_2 轴与 C_m 轴垂直，有

$$c_n^{-k} = c_2 c_n^k c_2. \quad (2.44)$$

可以验证，将旋转轴和 σ 面做共轭变换也能够得到新的旋转，即

$$\sigma c_m^i \sigma$$

也是一个 C_m 操作，记为 C'_m ，并且可以证明 C_m 的正转被映射为了绕着 C'_m 的反转，即

$$c_m'^{-k} = \sigma c_m^k \sigma. \quad (2.45)$$

特别的，

$$\sigma_\nu c_m^k \sigma_\nu = c_2 i c_m^k i c_2 = c_2 c_m^k c_2 = c_m^{-k}, \quad (2.46)$$

而对 σ_h 有

$$c_n^k = \sigma_h c_n^k \sigma_h. \quad (2.47)$$

看起来这里似乎有一个矛盾，因为等式左边的 k 应该是 $-k$ ，但是在 σ 是 σ_h 时变换前后的旋转轴其实是同一根旋转轴，只不过指向相反，那么如果将它们看成同一根轴，就应该将 $-k$ 换成 k 。旋转操作在点反演的共轭变换下不变，即

$$i c_m^k i = c_m^k. \quad (2.48)$$

以上方程给出了 c_m^k 在各种共轭变换下的变换，从而给出了一个旋转操作的等价类。给定一个旋转轴和它的旋转群，做这些共轭变换，会得到另一个旋转群，从而以上方程实际上指出了在一个点群中的等价旋转轴：对一条旋转轴 C_m ，如果有和它相交的旋转轴 C_n ，那么 C_m 中的等价旋转轴包括 C_{mj} ， j 取遍0到 $n-1$ ；如果有反演面，那么就有等价旋转轴 C'_m ；如果有与之垂直的 C_2 轴，那在 C_m 的旋转群内部就有群元彼此等价。没有更多方式产生等价旋转轴。需注意两个轴等价不代表以它们为轴的所有旋转操作都在一个等价类中；这只能说明，某个绕一个轴的旋转一定和某个绕另一个轴的旋转等价。

也可以依照类似的方法分析反射变换的共轭等价类。对反演面，做旋转的共轭变换有

$$\sigma' = c_n^k \sigma c_n^{-k}. \quad (2.49)$$

2.2.3 点群的分类

本节给出所有可能的晶体点群，以及如何分类它们。

第一类点群

第一类点群包括那些只有纯粹的转动，即没有反演的点群。它们也可以称为离散旋转群。第一类点群中最简单的是**单轴点群**，即只有一条对称轴的点群，它们就是 C_n 。 C_n 是 n 阶阿贝尔群，从而有 n 个不可约表示，每个不可约表示都是一维的。晶体点群必须和离散空间平移对称性兼容，换言之，做完晶体点群变换之后的晶格必须仍然是原来的晶格，格矢的指向不变。正是这个条件导致 C_n 中的 n 不能任意取值。分子点群就不受到这个限制。设 \mathbf{t} 是一个长度最短的格矢（我们正是在这一步用到了“存在一个晶格”的假设，否则没有一个长度最短但是有限的格矢），我们以它的起点为中心施加一个点群中的转动 c_n^1 ，则在转动 c_n^1 之后它必须变成另一个格矢，这等价于说 $\mathbf{t}' - \mathbf{t}$ 是一个格矢。我们有

$$|\mathbf{t}' - \mathbf{t}| = 2|\mathbf{t}| \sin\left(\frac{\pi}{n}\right),$$

由于 \mathbf{t} 长度最短，有

$$2|\mathbf{t}| \sin\left(\frac{\pi}{n}\right) \geq |\mathbf{t}|,$$

于是 n 只能从下面的值中选取：

$$n = 1, 2, 3, 4, 6.$$

我们会发现每个值都是可以的： $n = 1$ 对应一个平庸的情况，即晶体没有什么特别的点群对称性； C_2 和 C_4 可以和正方形晶格兼容， C_3 可以和三角晶格兼容， C_6 可以和蜂窝晶格兼容。总之，晶体点群中，单轴点群只有 C_1, C_2, C_3, C_4, C_6 ，这就是所谓的**晶体局限定理**。实际上，点群同样会选择能够与它搭配的晶格，晶格的形状不能完全任意搭配，一些点群只能和特定的一些晶格一起出现。

另一种第一类点群不包含或者只包含一个高阶轴，除此以外的旋转轴都是 C_2 轴。如果有两个 C_2 轴，那么它们必须垂直，否则按照(2.41)就会产生高阶轴（证明(2.41)时我们是假定存在垂直两根 C_2 轴的轴，但是这没有关系，因为群乘法的封闭性保证了 c_{2A} 和 c_{2B} 相乘的结果总是在原来的点群中）。在两根 C_2 轴垂直时，围绕它们旋转 π 的操作相乘给出绕着垂直于这两根轴确定的平面旋转 π 的操作。因此我们得到一个点群 D_2 。无法再塞入更多的 C_2 轴了——三维空间中彼此相互正交的方向最多能有三个。因此， D_2 就是只含有 C_2 轴但是和 C_n 不同的唯一一个第一类点群。

现在考虑一个只有一个 C_n 高阶轴的第一类点群，剩下的旋转轴只有 C_2 轴。这些 C_2 轴必须和 C_n 轴垂直，否则根据(2.43)将会有两条 C_n 轴。反过来，通过 c_n^k 共轭变换，从一条 C_2 轴可以产生 n 个 C_2 轴。不能有更多的 C_2 轴了，否则将会出现另一条和 C_n 轴重合的高阶轴。于是我们又得到了一个点群，它包含一条 C_n 轴和 n 条与之垂直，相邻夹角为 π/n 的 C_2 轴（注意这 n 条轴都是直线而不是射线，从而，相邻的 C_2 轴的夹角是 $2\pi/2n$ ）。记这样的群为 D_n ，其阶为 $2n$ ，其中有 C_n 轴给出的 n 个（包含了单位元的）操作和 n 个 C_2 轴给出的旋转 π 的操作。 C_2 轴旋转和 C_n 轴旋转的复合给出的也是 C_2 轴旋转。可以看出 D_n 实际上就是正 n 边形的对称群。由于 C_n 中的 n 有限制，因此只有 D_1, D_2, D_4, D_6 能成为晶体点群。

这些 D_n 群的等价类划分可以如下做出。容易验证在 D_n 中，我们有

$$c_2 c_2' c_2 = c_2', \quad c_n^k c_2 c_n^{-k} = c_2', \quad c_2 c_n^k c_2 = c_n^{-k}, \quad c_n^k c_n^{k'} c_n^{-k} = c_n^{k'},$$

因此 D_n 的一个等价类中要么都是 C_n 旋转, 要么都是 C_2 旋转。上式直接给出 C_n 的分类: c_n^k 和 c_n^{-k} 被分为一类。我们发现 $c_n^1 c_2 c_n^{-1}$ 和 c_2 的夹角为 $2\pi/n$, 即通过 c_n^k 对 c_2 做共轭变换, 一次最少导致 $2\pi/n$ 的方向变化。这意味着, 如果 n 是偶数, 那么通过 c_n^k 对 c_2 做共轭变换只能够覆盖一半的 C_2 轴, 因此全体 C_2 轴被分成两个等价类; 而对奇数 n 通过 c_n^k 对 c_2 做共轭变换可以覆盖一半的 C_2 轴。总之, 对奇数 n , 等价类包括: 单位元, 全体 C_2 轴作为一个等价类, 以及

$$(c_n^1, c_n^{n-1}), \dots, (c_n^{(n-1)/2}, c_n^{(n+1)/2})$$

共计 $(n+3)/2$ 个; 而对偶数 n , 等价类包括: 单位元, 两个包含了 C_2 轴的等价类, 以及

$$(c_n^1, c_n^{n-1}), \dots, (c_n^{n/2-1}, c_n^{n/2+1}), c_n^{n/2},$$

共计 $3 + n/2$ 个。这些等价类的数目就是不可约表示的数目, 而全体不可约表示的维数的平方和就是群的阶, 于是不可约表示的维数确定如下:

- D_2 有4个1维表示;
- D_3 有2个1维表示, 1个2维表示;
- D_4 有4个1维表示, 1个2维表示;
- D_6 有4个1维表示, 2个2维表示。

最后我们考虑具有多个高阶轴的第一类点群。设有两根高阶轴相交于 O 点, 分别为 C_m 轴和 C_n 轴, 并且它们之间的夹角最小。使用之前的套路, 可以获得环绕在 C_n 轴周围的 n 根 C_m 轴, 以及环绕在 C_m 轴周围的 m 根 C_n 轴。接着我们从中再次挑选出一对夹角最小的 C_m 轴和 C_n 轴, 重复这个步骤。晶体点群的有限性和封闭性意味着这种迭代最终会停止。一点几何上的直觉就告诉我们, 这样产生的全体对称轴和一个以 O 为球心的球面的交点连接起来将会产生一个正多面体。因此, 具有多个高阶轴的第一类点群必须包括某个正多面体的全部旋转对称性。由于正多面体的高度对称性, 实际上无法向正多面体的旋转对称群中再塞入更多的旋转对称性而仍然得到一个离散旋转群。因此具有多个高阶轴的第一类点群一定是某个正多面体的旋转对称群。

于是我们首先分析正多面体的对称性。一共只有五种正多面体, 其中正四面体的旋转对称群是 T , 正八面体和立方体的旋转对称群是 O , 正二十面体和正十二面体的旋转对称群是 I , 但是 I 含有一个 C_5 轴, 因此不能作为晶体点群。因此, 具有多个高阶轴的第一类点群必须包括 T 或 O 。 T 群是12阶群, 具有四个 C_3 轴和三个 C_2 轴。可以验证 C_3 轴彼此等价, C_2 轴彼此等价, 等价类为 e , 全体 c_2 , 全体 c_3^1 和全体 c_3^2 , 共4个等价类, 从而有3个1维不可约表示和1个3维不可约表示。 O 群是24阶群, 有三个 C_4 轴, 四个 C_3 轴和六个 C_2 轴, 容易验证所有 C_4 轴都是等价的, 所有 C_3 轴和 C_2 轴也是等价的, 并且每个高阶轴都有与之垂直的 C_2 轴, 从而正转和反转相同角度的旋转操作也是等价的。这样, 等价类包括: 单位元, 全体 c_2 , 全体 c_3^1 和全体 c_3^2 构成的单独一个类, 全体 c_4^1 和全体 c_4^3 构成的一个类, 全体 c_4^2 , 共计5个等价类。

总之, 我们得到了11种第一类点群:

$$C_1, C_2, C_3, C_4, C_6, D_2, D_3, D_4, D_6, T, O.$$

第二类点群

现在将反演操作加入。点群中的旋转部分组成一个正规子群，从而，第二类点群可以通过向第一类点群加入反演操作来得到。实际上容易发现如果点群中包括 i ，那么 $\{e, i\}$ 也是一个正规子群。能够加入的反演操作包括面反演操作和中心反演操作。向非平凡的旋转群加入面反演操作 σ 时，如果反射面和一个对称轴既不垂直也没有包含关系，则显然会产生一条新的对称轴，新的对称轴根据(2.43)又可以产生新的对称轴。新的对称轴或是可以没完没了地产生，这样得到的就不是有限大小的点群了，或是在产生一定数目的新对称轴之后停下，此时 σ 或者必须能够让某个对称轴不变，或者让两个对称轴相互转化（几何上的考虑表明，不可能让多个对称轴做置换）。如果反演操作总是将一个对称轴转化为另一个对称轴（并且与此同时让后者也变成前者），则考虑一族围绕在一个 C_m 轴周围的 m 个对称轴，它们均交于一点 O ，则 σ 反射面也必须经过 O ，则几何上的考虑意味着 σ 反射面或者包含 C_m 轴或者垂直于 C_m 轴。而相对于那些围绕在 C_m 轴周围的那些对称轴而论，容易看出 σ 反射面可以是 σ_h 轴，可以是 σ_v 轴也可以是 σ_d 轴。几何上的考虑表明，奇数 C_m 和 i 一起出现会引发矛盾，从而 i 只能出现在所有旋转轴都是偶数的点群中。然而，此时，必定存在一个 c_2 操作，从而根据(2.38)，点群中一定存在一个镜面反演操作，且 i 可以用这个镜面反演和旋转复合给出。

总之，含有 i 的点群一定含有镜面反射，并且所有的镜面反射——无论对哪些轴而言——都或者是 σ_h 轴，或者是 σ_v 轴，或者是 σ_d 轴。可以存在这样的情况：点群中有某个旋转反射的操作，但是这个旋转反射操作中的镜面操作不在点群中，但是因此，我们接下来可以采用这样的方法获得所有第二类点群：给定一个第一类点群，可以向其中插入镜面反演操作和所谓**旋转反射轴**（即将一个旋转轴的所有旋转操作都乘上一个镜面反演操作而得到的一系列操作，但是该旋转轴的纯旋转操作并不都在群内），并通过群乘法获得所有其它群元。旋转反射覆盖了所有可能的“旋转和反演相乘”的操作，因为一个形如 $c_m^k i$ 的操作一定可以写成一个形如 $c_m^{k'} \sigma$ 的操作。

最简单的，向 C_1 群加入反演操作，将得到

$$C_s = \{e, \sigma\}, \quad C_i = \{e, i\} \quad (2.50)$$

两个群，各有两个一维不可约表示。我们无法加入更多的反演操作，因为两个反演操作的复合是一个旋转操作，从而产生了旋转轴，不再是简单的 C_1 群的扩充了。

然后尝试向 C_n 群加入反演操作。如果反演平面垂直于旋转轴，那么我们将获得

$$C_{nh} = C_n \otimes \{\sigma_h, e\} \simeq C_n \otimes C_s. \quad (2.51)$$

其结构是很简单的，包含 $2n$ 个元素，且为阿贝尔群。注意此处我们讨论的点群都是具体的矩阵群而不是抽象的群结构，群乘法表同构的群我们不认为是同一个。特别的，

$$C_{1h} = C_s, \quad C_{2h} = \{e, c_2, \sigma_h, c_2 \sigma_h = i\} = C_2 \otimes C_i. \quad (2.52)$$

如果反演平面包含旋转轴，那么我们将获得 C_{nv} 。在 C_{nv} 中由于 C_n 轴的作用，我们实际上可以获得 n 个 σ_v 平面，它们均包含 C_n 轴。一个有趣的地方是 C_{nv} 实际上和 D_n 是同构的，因为 σ_v 和 C_2 地位等价。实际上，注意到对多边形， σ_v 和 C_2 是同一类操作，就能够发现这一点。 C_{nv} 的不可约表示和等价类和 D_n 均一致。

此外尚可考虑同时向 C_n 加入（一圈共计 n 个） σ_v 操作和 σ_h 操作，不过此时实际上有 n 个垂直于 C_n 轴的 C_2 轴，因为通过

$$c_2 = c_n^2 \sigma_v \sigma_h$$

可以得到一根 C_2 轴，从而得到一根因此我们将这个情况留到 D_n 群的扩充中讨论。

以上得到的几种第二类点群都不包含旋转反射。我们现在考虑包含旋转反射的群。旋转反射中的镜面反演操作必须是 σ_h 。如果是 σ_v ，那么

$$(c_n^{k'} \sigma_v)(c_n^k \sigma_v) = c_n^k (\sigma_v c_n^k \sigma_v) = c_n^{k'} c_n^k,$$

因此这根轴其实不是旋转反射轴：它的所有旋转操作都可以不乘上 σ 就出现在群中。类似的 σ 也不能是歪斜的。因此旋转反射中的镜面反演操作必须是 σ_h 。这又意味着

$$c_n^k \sigma_h = \sigma_h c_n^k. \quad (2.53)$$

旋转反射轴的 n 不能是奇数，因为对奇数阶的旋转操作，每一个操作的阶都是 n ，我们知道

$$(c_n^k \sigma_h)(c_n^{k'} \sigma_h) = c_n^{k+k'},$$

而使用 $c_n^{k+k'}$ 连续自乘可以得到所有的 C_n 操作，从而这根 S_n 轴并不是旋转反射轴——全部纯旋转操作都可以出现在点群中。因此在晶体点群中的旋转反射轴只能是 $n = 4$ 或 $n = 6$ 的。 n 为偶数并不意味着所有 $c_n^k \sigma_h$ 都应该放进点群中，否则，通过

$$(c_n^{k+1} \sigma_h)(c_n^k \sigma_h) = c_n^1$$

就能够得到所有纯旋转操作，从而这根轴不是旋转反射轴。因此我们需要决定哪些 $c_n^k \sigma_h$ 操作需要放入点群中。肯定必须放入 k 为奇数的 $c_n^k \sigma_h$ ，否则我们得到的其实是退化的 $n/2$ 阶旋转反射操作。另一方面，注意到，如果点群中出现了一个 k 为奇数的 $c_n^k \sigma_h$ 操作，那么它的某个奇数次自乘将会给出 $c_n^1 \sigma_h$ ，而使用 $c_n^1 \sigma_h$ 的 k 次自乘就能够得到 $c_n^k \sigma_h$ 。因此我们得出结论：点群中的旋转反射操作包括有且只包括 k 为奇数的 $c_n^k \sigma_h$ 操作，并且它们都可以用 $c_n^1 \sigma_h$ 的自乘给出。于是我们设

$$s_n^k = (c_n^1 \sigma_h)^k, \quad (2.54)$$

在 k 为奇数时它是 $c_n^k \sigma_h$ ，在 k 为偶数时，注意到

$$(c_n^1 \sigma_h)^2 = c_n^2,$$

我们有

$$s_n^k = c_n^k. \quad (2.55)$$

因此与 n 阶旋转反射轴伴随着还有一系列 $C_{n/2}$ 操作。这应当不出乎意料，因为旋转反射操作不能自己构成一个群——它们的行列式都是 -1 ，因此相乘之后会得到行列式为 1 的操作，从而必定是旋转。由于 σ_h 和 c_n^k 是交换的， c_n^k 彼此是交换的，我们发现 s_n^k 实际上是彼此交换的。

我们首先考虑一个生成元由一根 S_n 旋转反射轴提供的的第二类点群，将它记作 S_n ，它显然就是 s_n^1 连续自乘给出的，或者说是生成元为 s_n^1 的循环群。我们只需要考虑 S_4 和 S_6 。简单的计算给出

$$S_4 = \{e, s_4^1, c_2, s_4^3\}, \quad (2.56)$$

以及

$$S_6 = \{e, s_6^1, c_3^1, i, c_3^2, s_6^5\}. \quad (2.57)$$

这里的 i 来自

$$s_6^3 = c_6^3 \sigma_h = c_2 \sigma_h = i. \quad (2.58)$$

这些群都是阿贝尔群。可以看到旋转反射轴和旋转轴的代数性质是非常相似的。

很自然的问题是，有多根旋转反射轴的点群是什么。由于只有 S_4 和 S_6 两种旋转反射轴，而它们分别伴随 C_2 和 C_3 轴出现，如果一个第二类点群具有多根旋转反射轴，那么它也有多根普通旋转轴。如果一个第二类点群具有两根 S_4 旋转反射轴，那么它具有两根位置相同的 C_2 旋转反射轴，即它含有 D_2 为子群。类似的会发现，所有含有多根旋转轴的第二类点群都含有 D_n 群或是多面体群。因此，一个含有旋转反射轴的第二类点群 G 一定包含一个第一类点群 K ，只需要将其中的一些旋转轴替换成旋转反射轴就可以了；任何一个旋转反射轴的空间位置都和 K 中的某个旋转轴的位置是重合的。我们下面还是会使用“往第一类点群中塞入反演变换”来构造第二类点群，但是对每一种情况，都需要检查是否存在这样的旋转反射轴，它的 σ_h 操作本身不在群里（实际上是没有这样的情况）。

对 D_n 群，由于只有一根 C_n 轴，也可以有一样的构造，让我们得到 D_{nh} 。此时 n 根 C_2 轴可以和 σ_h 组合产生 n 个 σ_v 反射面。使用和推导 D_n 群类似的论证，会发现 n 为偶数时这些 σ_v 反射面被分成两个等价类，而 n 为奇数时它们同属一个等价类。除了 C_n 旋转， n 个 C_2 旋转， n 个 σ_v 反演面和一个 σ_h 反演面以外， D_n 群中还有 C_n 旋转和 σ_h 组合形成的旋转反射操作 S_n ；请注意此时即使 k 为偶数，我们还是可以定义

$$s_n^k = \sigma_h c_n^k, \quad (2.59)$$

虽然在 S_n 群中只有奇数 k 才有意义；此外 n 也可以取2或是3。

我们现在可以给出 D_{nh} 群的等价类和不可约表示数目。如果 n 是奇数，那么 D_{nh} 的等价类包括：单位元，全体 C_2 轴作为一个等价类，如下共计 $(n-1)/2$ 个的等价类：

$$(c_n^1, c_n^{n-1}), \dots, (c_n^{(n-1)/2}, c_n^{(n+1)/2}),$$

以及将以上所有等价类都乘上 σ_h 得到的新的等价类，这其中， C_n 操作乘上 σ_h 就得到了那些 S_n 操作，而那些垂直于 C_n 轴的 C_2 轴乘上 σ_h 后就得到了（相对于 C_n 轴而言的） σ_v 操作，它们相对于那些 C_2 轴一半是 σ_h 轴，一半是 σ_v 轴。总之，合计共有 $n+3$ 个等价类。 n 为奇数时可以验证 σ_h 和其它群元均对易，从而

$$D_{nh} \simeq D_n \otimes C_h. \quad (2.60)$$

偶数 n 的 D_{nh} 群的等价类可以使用类似的套路写出，它包括 D_n 的 $3+n/2$ 个等价类，和这些等价类乘上 σ_h 得到的新等价类，共有 $n+6$ 个。这里面 σ_h 和 C_2 轴相乘得到的轴相对于 C_2 轴来说都是 σ_h 轴。 n 为偶数时有群元 $c_n^{n/2}$ ，它实际上是一个 c_2 转动，于是 $i = c_n^{n/2} \sigma_h$ 是在群中的，它和其它所有群元对易，即我们有

$$D_{nh} \simeq D_n \otimes C_i. \quad (2.61)$$

这两种情况下， D_{nh} 的结构都是一个第一类点群直积上 $\{1, -1\}$ ，并且它们中的旋转反射轴都是那条 C_n 轴，没有更多旋转反射轴。

在 D_{nh} 中，垂直于 C_2 轴的那些 σ_h 面（相对于 C_n 轴是 σ_v 面）由 C_2 操作和相对于 C_n 轴的 σ_h 操作相乘得到，言下之意是，如果在一个包含 D_n 的第二类点群 G 中加入垂直于 C_2 的镜面反射面，那么立刻有一个垂直于 C_n 的反射面出现，从而接下来就会发现， G 一定包含 D_{nh} 。我们现在尝试构造一个不包含 D_{nh} ，但是还是包含 D_n 的第二类点群，那么插入 σ 反射面的选择就只有插入相对于 C_2 轴而言的 σ_d 反射面了（插入相对于 C_2 轴而言的 σ_v 在 n 为偶数时会产生 D_{nh} 群，在 n 为奇数时会产生新的旋转轴）。称这样的第二类点群为 D_{nd} 。我们会发现此

时 C_n 轴实际上变成了 S_{2n} 旋转反射轴：我们有

$$s_{2n}^1 = \sigma_d c_2. \quad (2.62)$$

σ_d 轴的存在让所有 C_2 轴都在同一个等价类内；而反过来， $\sigma_d c_2 \sigma_d$ 会给出所有其它的 c_2 操作，即所有 C_2 轴都是在同一个等价类里面的。 D_{nd} 群中不包含 σ_h 群，从而 S_{2n} 轴的存在本身就意味着 $2n = 4, 6$ ，即 $n = 2, 3$ 。 D_{2d} 中的等价类包括单位元， (s_4^1, s_4^3) ，转轴方向和 S_4 轴一致的那个 c_2 旋转，一个包含与 S_4 轴垂直的两个 c_2 旋转的等价类，还有两个 σ_d 组成的等价类，共计5个等价类，共有8个群元，从而有4个一维不可约表示和一个二维不可约表示。 D_{3d} 与之类似，所有旋转反射操作同属一类，所有转轴和旋转反射轴一致的纯旋转同属一类，但是还有一个 i 操作，于是共有6个等价类。计算可得有12个群元，不可约表示中有4个一维不可约表示和2个二维不可约表示。

我们现在看到，向 D_n 中塞入 σ_h 操作必然导致 C_n 轴变成旋转反射轴。剩余的扩充 D_n 的方式有：向 D_{nh} 或 D_{nd} 塞入更多 σ 操作，实际上，这就是要构造一个同时具有 σ_h 和 σ_d 操作，而包含 D_n 为子群的群；又或者，可以尝试让 D_n 中的 C_2 轴能够提供旋转反射操作。让 C_2 轴变成 S_n 这样的旋转反射轴肯定是不可能的，因为如前所述，只有 S_4 和 S_6 ；而如果我们加入某个垂直于 C_2 轴的 σ_h 操作（它相对于 C_n 轴是 σ_v ），那么它和 C_2 旋转复合给出的将是 i ，如果 n 是奇数，矛盾就产生了，而如果 n 是偶数，那么实际上我们得到的将是一个 D_{nh} 群。要求 σ_d 和 σ_h 同时存在同样会产生一个 D_{nh} 群，因为此时容易验证

$$c_2 \sigma_d = i, \quad (2.63)$$

因此 n 不能是奇数，否则会引发矛盾；但是在 n 是偶数时，所谓的 σ_d 实际上就是 D_{nh} 已有的那些 σ 面。总之，除了 D_{nh} 和 D_{nd} 以外，不存在更多扩充 D_n 群的方法。

最后考虑多面体群。我们首先尝试向 T 群加入空间反演操作。为了让空间反演相对所有反射轴都或者是 σ_h ，或者是 σ_d ，或者是 σ_v ，我们只有两种可能：一种是向 O 群加入空间反演操作只有一种可能：反演面垂直于 C_4 轴。

因此现在我们得到了21种第二类点群：

$$C_i, C_s, C_{2h}, C_{3h}, C_{4h}, C_{6h}, C_{2v}, C_{3v}, C_{4v}, C_{6v}, S_4, S_6, D_{2h}, D_{3h}, D_{4h}, D_{6h}, D_{2d}, D_{3d}, T_d, T_h, O_h.$$

至此，我们得到了全部32种点群，这包括10种第一类点群和22种第二类点群。关于这些群的具体的群表、不可约表示、特征标表等可以查阅晶体学用表得到，我们在此处给出了从头计算它们需要的全部信息。

2.2.4 7种晶系和14种布拉伐格子

旋转操作和平移操作兼容的要求给出了晶体局限定理，反过来，点群也会挑选能够和他兼容的晶格。任何一个晶格都必然在某个点群的作用下保持不变。如前所述，晶体的点群对称性总是小于它的晶格，或者说晶体的点群一定是它的晶格的点群的一个子群，因此，分析出晶格有哪些可能的对称性——或者说，晶格的点群有哪些——就能够为晶体分类提供依据。

设 P 是某个晶格 T （这里 T 就是格矢组成的集合，但是很显然，这个集合立刻就确定了晶体的离散平移对称性；之后我们也会用 T 表示晶格的离散平移群）的不需要和任何平移操作联合使用的点群，满足

$$P = \{\alpha | \alpha \mathbf{r} = \mathbf{r}' \in T\}, \quad (2.64)$$

显然 P 是32种点群中的一种。下面我们将推导出 P 需要满足的一些约束条件。

首先空间反演操作 i 肯定在 P 当中，理由是显然的。其次，如果 $C_n \subset P, n \geq 2$ ，则一定有 $\sigma_h \in P$ 。这件事的证明如下：设 S 是一个垂直于 C_n 轴的晶面，设其上最短的格矢为 \mathbf{a}_1 ，设 $\mathbf{a}_2 = c_n^{-1} \mathbf{a}_1$ 。显然， \mathbf{a}_1 和 \mathbf{a}_2 都是格矢。设 \mathbf{a}_3 是一个和这两个格矢都线性无关的格矢，并做分解

$$\mathbf{a}_3 = \mathbf{u} + \mathbf{v},$$

其中 \mathbf{u} 是平行于 C_n 的而 \mathbf{v} 是垂直于 C_n 的。这样我们就有

$$c_n^{-1} \mathbf{a}_3 - \mathbf{a}_3 = (c_n^{-1} \mathbf{u} - \mathbf{u}) + (c_n^{-1} \mathbf{v} - \mathbf{v}) = c_n^{-1} \mathbf{v} - \mathbf{v}.$$

可以确定 $c_n^{-1} \mathbf{a}_3 - \mathbf{a}_3$ 一定在 T 中，而 $c_n^{-1} \mathbf{v} - \mathbf{v}$ 一定在 S 中，从而 $c_n^{-1} \mathbf{v} - \mathbf{v}$ 一定是位于 S 上的某个格点，从而能够找到整数 n_1 和 n_2 使得

$$c_n^{-1} \mathbf{v} - \mathbf{v} = n_1 \mathbf{a}_1 + n_2 \mathbf{a}_2.$$

我们在两边作用 c_n^{-1} ，得到

$$\mathbf{v} - c_n^{-1} \mathbf{v} = n_1 c_n^{-1} \mathbf{a}_1 + n_2 \mathbf{a}_1.$$

两式相减，得到

$$c_n^{-1} \mathbf{v} + c_n^{-1} \mathbf{v} - 2\mathbf{v} = n_2 \mathbf{a}_2 - n_1 c_n^{-1} \mathbf{a}_1 + (n_1 - n_2) \mathbf{a}_1.$$

注意到

$$c_n^{-1} \mathbf{v} + c_n^{-1} \mathbf{v} = 2 \cos \frac{2\pi}{n} \mathbf{v}, \quad c_n^{-1} \mathbf{a}_1 + c_n^{-1} \mathbf{a}_1 = \mathbf{a}_2 + c_n^{-1} \mathbf{a}_1 = 2 \cos \frac{2\pi}{n} \mathbf{a}_1,$$

从而

$$2 \left(\cos \frac{2\pi}{n} - 1 \right) \mathbf{v} = \left(n_1 - n_2 - 2n_1 \cos \frac{2\pi}{n} \right) \mathbf{a}_1 + (n_1 + n_2) \mathbf{a}_2. \quad (2.65)$$

现在我们尝试构造一个 σ_v 。如果选取 σ_v 与 \mathbf{a}_1 垂直，那么，只需要证明 \mathbf{a}_2 和 \mathbf{a}_3 在 σ_h 作用下也是一个格矢，就足够证明晶格的点群中有 σ_v 。对 $n = 3, 4, 6$ 验证，均可发现以上说法正确。因此确 $\sigma_h \in P$ 。

满足这两个条件的点群共有7个，并且我们马上会看到它们是什么晶格的点群。我们说，这7个点群标记了7个晶系，因为如第2.2.1节所说那样，晶系是按照晶体中的对称轴和对称面数目划分的，实际上就是按照晶格的点群划分的。我们下面列出它们和以它们为点群的晶格（以下提到“三个基元格矢”时指的都是六面体晶体学单胞的基元格矢）：

- C_i 给出三斜晶系，三个基元格矢之间没有什么特别明确的关系：它们的长度不一致，并且彼此不垂直，仅有的对称性就是空间点反演对称性；
- C_{2h} 给出单斜晶系，三个基元格矢长度不同，但是它们的三个夹角中有两个是 90° ，从而晶体学单胞是一个长方体朝着某条棱的方向受剪切力而形成的形状，或者说是一个底面为边长不相等的直棱柱， C_{2h} 中的 c_2 操作的转轴平行于棱， σ 操作的平面就是这个直棱柱的底面；
- D_{2h} 给出正交晶系或是斜方晶系，三个基元格矢的长度不同，但是全部彼此正交，晶体学单胞是一个三条棱各不相同的长方体， D_{2h} 群的作用方式是显然的；
- D_{4h} 给出四方晶系，其晶体学单胞是一个正四棱柱， D_{4h} 的作用方式也是显然的；

- D_{3d} 给出三角晶系，这个晶系又可以分成两种类型，其中一种的晶体学单胞是一个棱长全部相同的，三个基元格矢之间夹角相同的平行六面体， D_{3d} 的 C_3 轴就是它的对角线， σ_d 面就是它的对角面；另一种的晶体学单胞也是一个平行六面体，其中一个轴垂直于另外两个，它就是 D_{3d} 的 C_3 轴，那另外两个轴的夹角是 120° ；
- D_{6h} 给出六角晶系，为了更好地展现其对称性，它的晶体学单胞是一个正六棱柱的 $1/3$ ，横截面为边长相等，其中一个角为 60° 的平行四边形， D_{6h} 的作用方式也是显然的，六角晶系的格子和三角晶系的第二种格子完全一样，但是前者有 C_6 轴而后者只有 C_3 轴；
- O_h 给出的立方晶系，顾名思义，其晶体学单胞是一个立方体，并且存在空间镜面反演操作，因此其点群自然就是 O_h 。

这个列表给出了 P 的所有可能取值。

这样的根据点群给出的晶系定义会将三个基元格矢长度相同，彼此之间夹角相同的平行六面体组成的晶格和正六棱柱组成的晶格都和 D_{3d} 匹配，而这两种晶格的几何形状是不同的，且三角晶系和六角晶系共用一种格子。由于“晶系”一词有时候也用来形容晶格的分类，这就引起了混乱：如果根据格子的形状划分晶系，那么应该将“三角晶系”一词用于描述那种三个基元格矢长度相同，彼此之间夹角相同的平行六面体组成的晶格，而“六角晶系”则用于描述那种正六棱柱组成的晶格，此时六角晶系可能具有两种点群，并且“不同晶系的晶体的点群不同”不再成立。无论使用哪种定义都有7种晶系。为了避免混乱，本文将三个基元格矢长度相同，彼此之间夹角相同的平行六面体组成的晶格称为三角格子，而将正六棱柱组成的晶格称为六方格子，而将七种晶格种类称为晶格系统(lattice system)而不是“晶系”(crystal system)。

需要注意的是，晶系给定后晶格形状并不能够唯一确定，因为一个晶体学单胞中可以有多个原胞，从而，除了角上的格点以外，晶体学单胞的面、体内也可以有格点。这些格点在晶格的点群的作用下必须被变换到其它格点上。能够验证，或者没有这些面、体上的格点（称为简单格子），或者只有体心、面心、底心三种情况。一些体心、面心、底心的情况在晶格的点群的作用下无法被变换到其它格点上，一些情况则可以归并入另一些情况，具体来说：

- 三斜晶系只有简单晶格，其它情况均为简单晶格。
- 底心单斜格子的“底”要是长方形面，否则就是简单单斜格子。体心和面心单斜格子满足 C_{2h} 对称性，但是实际上就是底心单斜格子。
- 底心四方格子的“底”如果是正方形，显然就是简单四方格子，如果是长方形，则和 D_{4h} 群矛盾，因为所有长方形面都是等价的。面心四方格子就是体心四方格子。
- 底心三角格子不满足 D_{3d} 对称性，因为 D_{3d} 作用在三角格子上会让各个面都等价。面心和体心三角格子
- 底心、体心六方格子不满足 D_{6h} 对称性：想象我们绕着原胞角上的一个格点做 60° 旋转，这会让角上的格点被旋转到另一个格点上，但是位于底心或是体心的格点会被转动到新的位置上。面心六方格子包括了体心处的格点，因此也不满足 D_{3d} 对称性。
- 底心立方格子不满足 O_h 对称性，因为 O_h 要求各个面都要等价，实际上它是简单四方晶格。

推导结果是，总共有14种可能的晶格，即布拉伐格子一共只有14种。

在进行下面的讨论时，我们有时不将晶胞取为原胞，从而让对称性展现得更加直观。

应当注意，存在这样的可能：由于基元的对称性实在是差，虽然晶格具有某种较高的对称性（比如说六方格子），但是整个晶体只具有低一些的对称性（比如说斜方格子）。我们只能看着点群写下格子，但是不能反过来；或者，换句话说，可以看着格子写下**晶格系统**，但是不能就此确定**晶系**。例如，在简单正方晶格的每个侧面的中心增加一个基元，所得到的晶体的晶格系统仍然是正方的，但是晶系则是四方的。

在小的晶格扭曲之下，晶系和晶格系统均会出现降级。两者的降级方式是基本一致的——注意无论是六方晶系变成三方晶系还是六方晶格变成三方晶格都是可能的。

2.2.5 空间群的结构和分类

将一个空间群 G 中出现的所有的操作的点群部分都提取出来显然构成了一个群 K ，称为 **G 的点群**；空间群的点群未必是空间群的子群，因为其中的一些点群操作也许并不能单独出现在空间群中。然而， K 一定是 P 的子群，从而 K 中的旋转轴方向、反射面位置等都受到很强的限制，因为它们必须和晶格匹配。 K 对应宏观的晶体的点群对称性，因为点群操作加上一个小的平移即可成为空间群的元素，而小的平移宏观上是看不出来的。我们称这个小的平移操作为**分数平移**，因为它不在离散平移群中，不能将它写成基元格矢的整数倍；分数平移可以加上一个任意的格矢，以下不失一般性，我们要求分数平移的长度取最小值。这个空间群中单独出现的点群操作（对应微观的基元的对称性）构成的那个点群是空间群的点群 K 的子群，空间群的点群 K 又是其晶系的点群 P 的子群。

空间群的陪集分解

我们首先分析空间群中会出现什么元素，以及其群结构。空间群 G 可以根据其不变子群 T 做陪集分解，其中的 τ_i 约束为分数平移：

$$G = \sum_{i=1}^n \{\alpha_i | \tau_i\} T, \quad \alpha_i \in K. \quad (2.66)$$

显然其中的某个 $\{\alpha_i | \tau_i\}$ 就是单位元，不妨指定它为 $i = 1$ 的情况：

$$\{\alpha_1 | \tau_1\} = \{e | 0\}. \quad (2.67)$$

对每个 α_i ，我们可以找到一个唯一的 τ_i ，理由如下：设

$$\{\alpha_i | \tau_i\} T = \{\alpha_i | \tau'_i\} T,$$

则

$$\{\alpha_i | \tau_i\} \{e | \mathbf{r}_n\} = \{\alpha_i | \tau'_i\} \{e | \mathbf{r}_{n'}\},$$

即

$$\tau_i + \alpha_i \mathbf{r}_n = \tau'_i + \alpha_i \mathbf{r}'_n.$$

这就意味着

$$\tau_i - \tau'_i = \alpha_i (\mathbf{r}_n - \mathbf{r}'_n) \in T,$$

因为 α_i 肯定在晶格点群 P 中，因为它必定将格矢 $\mathbf{r}_n - \mathbf{r}'_n$ 变换为另一个格矢。然而，如前所述， τ_i 和 τ'_i 都是分数格矢，并且已经取为最短，它们的差不可能是有限大小的格矢，因此

只能有

$$\tau_i - \tau'_i = 0.$$

这就证明了 τ_i 的唯一性。请注意前述推导并没有用到 α_i 的编号 i ——实际上，对不同的 i 和 j ， α_i 和 α_j 也不会重复，因为如果有两个不同的 $\{\alpha_i|\tau_i\}$ 和 $\{\alpha_j|\tau_j\}$ ，而 $\alpha_i = \alpha_j$ ，那么 $\tau_i - \tau_j$ 给出了空间群中的一个平移操作，然而， τ_i 和 τ_j 都是分数平移，它们之差也是，而分数平移是不会出现在空间群中的。这就导致一个矛盾。

不过，这并不是说整个空间群中都不存在 α 相同而 τ 不同的多个 $\{\alpha|\tau\}$ 操作，因为显然可以通过空间平移获得无数多个点群部分为 α 的操作：我们知道

$$\{e|\mathbf{R}\}\{\alpha|\tau\} \in G, \quad \mathbf{R} \in T,$$

这就是说

$$\{\alpha|\tau + \mathbf{R}\} \in G.$$

因此我们有无数多个 $\{\alpha_i|\tau\}$ 形式的操作。

空间群 G 的点群 K 中的操作不会出现在 T 中。由于 α_i 是点群操作，它们 α_i 都来自 K 。而由于 α_i 如前所述不会重复， K 中的每一个操作都给出一个（且只有一个） α_i 。因此我们可以用这样的方式推导出所有的空间群：我们首先指定晶体的 K 点群，然后，给这个 K 点群中的每一个操作 α_i 都指派一个唯一的 τ_i （实际上，只需要给 K 的每个生成元指派一个 τ_i 即可，剩下的都可以算出来），并保证群封闭性、有限性、与某种晶格匹配等条件成立，这样我们就得到了所有的空间群。实际上，注意到

$$G/T = \{\{\alpha_i|\tau_i\}T | \alpha_i \in K\},$$

其中，群乘法

$$(\{\alpha_i|\tau_i\}T)(\{\alpha_j|\tau_j\}T) = \{\alpha_k|\tau_k\}T$$

能够推出

$$\alpha_i\alpha_j = \alpha_k, \quad \tau_k = \alpha_i\tau_j + \tau_i + \mathbf{R}, \quad \mathbf{R} \in T,$$

于是实际上我们可以写出

$$G/T \simeq K, \tag{2.68}$$

从而空间群 G 和它的 K 点群同态，以及群乘法规则

$$(\{\alpha_i|\tau_i\}T)(\{\alpha_j|\tau_j\}T) = \{\alpha_i\alpha_j|\tau_i + \alpha_i\tau_j \mod T\}T, \tag{2.69}$$

或者我们其实可以略去 T 以简化书写。

空间群中的元素

表面上看 τ_i 的指派是很任意的，但是实际上不然，因为有很多限制。首先，如此构造时需要注意去除一些重复，因为我们可以随意从欧几里得群中找一个行列式为正（这点很重要，因为螺旋轴什么的是有手性的）的元素出来，对空间群做共轭变换，变换前后的空间群是等价的，应该看成一个。例如，我们可以用 $\{e|\mathbf{d}\}$ 做共轭变换，这会导致

$$\{e|\mathbf{d}\}\{\alpha|\tau\}\{e|\mathbf{d}\}^{-1} = \{\alpha|\tau + (1 - \alpha)\mathbf{d}\}.$$

实际上这就是把坐标平移了一下：设 $\mathbf{r}' = \mathbf{r} + \mathbf{d}$ ，则

$$\mathbf{r} \longrightarrow \alpha\mathbf{r} + \tau$$

被转化为了

$$\mathbf{r}' \longrightarrow \alpha \mathbf{r}' + \boldsymbol{\tau} + (1 - \alpha) \mathbf{d}. \quad (2.70)$$

我们也可以不局限在平移上，比如说可以用 $\{\alpha|\mathbf{d}\}$ 做共轭变换。因此，对空间群中的纯点群操作，可以通过适当的平移让它具有一个非零的 $\boldsymbol{\tau}$ ，或者当然也可以反过来让一个非零 $\boldsymbol{\tau}$ 变成零；不过，无论怎么做共轭变换， $\{e|0\}$ 一直是不变的。

其次， $\boldsymbol{\tau}$ 的取值也受到限制。对 $\{c_n^1|\boldsymbol{\tau}\}$ ，用 $\boldsymbol{\tau}_{\parallel}$ 表示 $\boldsymbol{\tau}$ 平行于 c_n^1 转轴的分量，用 $\boldsymbol{\tau}_{\perp}$ 表示 $\boldsymbol{\tau}$ 垂直于 c_n^1 的转轴的分量，我们可以验证

$$\{c_n^1|\boldsymbol{\tau}\}^n = \{c_n^n|n\boldsymbol{\tau}_{\parallel} + (c_n^1 + c_n^2 + \cdots + c_n^n)\boldsymbol{\tau}_{\perp}\} \in G,$$

由对称性

$$(c_n^1 + c_n^2 + \cdots + c_n^n)\boldsymbol{\tau}_{\perp} = 0,$$

于是 $n\boldsymbol{\tau}_{\parallel} \in T$ 。这就是说，如果 $\boldsymbol{\tau}$ 不是零，那么 c_n^1 的轴指向必须和某个正格矢一样（这件事本身就对旋转轴的选择施加了很大的限制，因为正如第2.2.4节中展示的那样，晶格就只有那么几种，而在每个晶格上旋转轴的方向都是确定的；反过来，后面 \mathbf{R} 的选择就是根据第2.2.4节中给出的旋转轴方向自动确定的），且设这个方向上最短的正格矢为 \mathbf{R} ，则存在整数 m, n 使得

$$\boldsymbol{\tau}_{\parallel} = \frac{m}{n} \mathbf{R}, \quad m = 0, 1, \dots, n-1. \quad (2.71)$$

请注意 $\boldsymbol{\tau}_{\parallel}$ 是分数平移这件事对 \mathbf{R} 也施加了一个限制：在 $\boldsymbol{\tau}$ 非零时，如果 \mathbf{R} 很大，那么即使 $m = 1$ ，也不能得到合理的 $\boldsymbol{\tau}_{\parallel}$ 。

$\boldsymbol{\tau}_{\perp}$ 的意义还没有分析。如果 α_i 是 c_n^1 操作（只需要考虑 c_n^1 ，因为只需要考虑给生成元指派派的 $\boldsymbol{\tau}$ ），那么(2.70)中的 $1 - \alpha$ 是

$$1 - \alpha = \begin{pmatrix} 1 - \cos \theta & -\sin \theta & 0 \\ \sin \theta & 1 - \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \theta = \frac{\pi}{6}, \frac{\pi}{4}, \frac{\pi}{3}, \frac{\pi}{2}, \pi. \quad (2.72)$$

这个矩阵的秩是2，因此 $(1 - \alpha)\mathbf{d}$ 的取值范围不是三维的，而是二维的，并且容易看出其取值范围是 xy 平面，或者说是垂直于 C_n 转轴的那个平面。因此可以通过适当的平移共轭变换让与 c_n^1 匹配的 $\boldsymbol{\tau}$ 的垂直于转轴的部分取任何的值；而 $\boldsymbol{\tau}_{\parallel}$ 则由(2.71)确定，不能通过平移共轭变换改变。因此，无论怎么平移， $\boldsymbol{\tau}_{\parallel}$ ——或者说 m ——都是固定不变的。没有必要对 G 中的所有操作都做对应的平移共轭变换，否则其它的 $\boldsymbol{\tau}_i$ 会变化，但以上说法意味着我们可以用一种几何意义更加明显的方法确定 $\{c_n^1|\boldsymbol{\tau}_i\}$ ：给定 n ，在垂直于 C_n 转轴的平面上的一个点（这个点确定了 C_n 转轴的空间位置），以及(2.71)中的 m ，就唯一确定了一个 $\{c_n^1|\boldsymbol{\tau}_i\}$ ，反过来，给定 $\{c_n^1|\boldsymbol{\tau}_i\}$ ，用 c_n^1 和 $\boldsymbol{\tau}_{\parallel}$ 可以确定 n, m ，而 $\boldsymbol{\tau}_{\parallel}$ 则给出转轴位置。实际上， $-c_n^1\boldsymbol{\tau}_{\perp}$ 一定是转轴上的一个点，据此和 c_n^1 的位置就确定了转轴。容易看出 c_n^i 的转轴位置和 c_n^1 是一样的。关于转轴位置还值得多说几句。我们知道可以在(2.70)中让 $\boldsymbol{\tau}$ 取任意一个格矢而得到无穷多个 $\{c_n^1|\boldsymbol{\tau}\}$ 操作，而 $-c_n^1\boldsymbol{\tau}_{\perp}$ 是转轴上的一个点，则取几个简单的情况（如 $n = 2$ ）就会发现实际上一个晶胞内可以有多个平移操作增生的旋转轴。在点群中不会有多条不相交而彼此平行的旋转轴，但是在空间群中，由于空间平移对称性，就可以有。

由于 $\boldsymbol{\tau}_{\parallel}$ 在平移下不变，我们称 $\boldsymbol{\tau}_{\parallel}$ 不为零的那些旋转轴为**螺旋轴**，因为它们给出螺旋线的对称性：做完旋转之后还需要沿着旋转轴做一小段平移。第2.2.4节中已经给出了 P 点群中的旋转轴的可能指向和位置——除了三角格子的 C_3 轴是沿着对角线以外，其它都是沿着某个棱，因此 \mathbf{R} 的可能取值是非常有限的。 K 点群中的旋转轴一定是其中之一。

类似的, 对 $\{\sigma|\tau\}$, 设 $\tau_{\parallel} \parallel \sigma$ 而 $\tau_{\perp} \perp \sigma$, 则能够证明, σ 平行于某个正格矢 \mathbf{R} , 并且

$$\tau_{\parallel} = 0, \quad \frac{1}{2}\mathbf{R}. \quad (2.73)$$

$\{i|\tau\}$ 的 τ 没有特殊的限制, 然而由于 i 操作肯定可以用其它两个群元乘出来, 它的 τ 也是可以被别的操作确定的。

如果 α_i 是 σ 操作, (2.70)中的 $1 - \alpha$ 就是 $1 - \sigma$, 从而

$$(1 - \alpha)\mathbf{d} = 2\mathbf{d}_{\perp},$$

从而可以通过适当的平移去掉 τ 中垂直于 σ 的分量, 但是不能改变 τ_{\parallel} 。与旋转轴类似, 我们不去真的做平移, 而是使用 σ 的法向量的指向, σ 面上的一个点 (它和 σ 的法向量的指向一起确定了 σ 面的位置) 和 $\tau_{\parallel} = 0, \mathbf{R}/2$ 确定(2.66)中的一个 $\{\alpha_i|\tau_i\}$, 反过来, 给定 $\{\alpha_i|\tau_i\}$, τ_{\perp} 给出了反射面的位置。由于 τ_{\parallel} 在平移下是不变的, 我们称 $\tau_{\parallel} = \mathbf{R}/2$ 的情况为滑移面。

根据第2.2.4节中的分析, 可以发现, 在所有情况下, σ 面或者平行于晶胞的底面, 或者是对角面, 。逐个检查这些情况, 会发现此时的 \mathbf{R} 的取值是高度受限的。总结起来可以分成这么几种:

- 轴向滑移面, 如果 \mathbf{a}_i 中的某一个正好在滑移面内, 则 $\mathbf{R} = \mathbf{a}_i$ 而 $\tau_{\parallel} = \mathbf{a}_i/2$ 。
- 对角线滑移面,
- 金刚石滑移面

对旋转反射轴指定非零的 τ_i 没有意义, 因为此时(2.70)中的 $1 - \alpha$ 是

$$1 - \alpha = \begin{pmatrix} 1 - \cos \theta & \sin \theta \\ -\sin \theta & 1 - \cos \theta \\ & & 2 \end{pmatrix},$$

这里我们把 z 轴取成 s_n^1 轴。 $1 - \alpha$ 的行列式在 S_4 和 S_6 轴下都不是零, 因此实际上可以通过做平移让 $\tau = 0$, 从而不需要像旋转轴一样, 同时指定转轴位置和 τ_{\parallel} , 对旋转反射轴只需要指定其位置就完全确定了它。

实际上, 我们只需要给 K 点群的每个生成元——旋转, 镜面反射, 旋转反射操作——指派一个 τ , 如果能够保证这样会产生一个合理的空间群, 那就确定了一个空间群。

空间群的分类和推导

如果在适当的坐标系下, 陪集分解(2.66)中没有出现分数平移, 则此时的空间群称为简单空间群或者说点式空间群。因此, 在原胞中至少可以找到一个点 (即坐标原点), 使得简单空间群 G 对它的作用和 G 的点群 K 完全一样。点式空间群就是它的 K 点群和平移群 T 的半直积; 请注意此处 K 和 T 必须要兼容, 即 K 要是 T 的点群的子群。需要注意点式空间群中也是可以有螺旋轴和滑移面的, 例如, 最简单的, 如果某个旋转轴正好沿着某个格矢, 那么把这个旋转操作和格矢的平移操作乘起来总能够得到一个螺旋操作。稍微复杂一些, 设有一个和 C_n 轴不平行的格矢 \mathbf{R} , 则

$$\{c_n^1|0\}\{e|\mathbf{R}\} = \{c_n^1|c_n^1\mathbf{R}\} = \{c_n^1|\mathbf{R}_{\parallel} + c_n^1\mathbf{R}_{\perp}\}$$

也在空间群中。在 \mathbf{R}_{\parallel} 本身不是格矢时, 由于 $c_n^1\mathbf{R}_{\perp}$ 不为零, 实际上我们获得了一根不通过 K 点群的不动点的螺旋轴! 但是无论如何, 对简单空间群, 总是可以不使用它们做陪集分解(2.66)。

点式空间群的推导是比较平凡的，根据点群给出晶胞中所有旋转轴和反射面的位置即可。给定一种格子和一个点群，即可得到对应的点式空间群。

非点式空间群或者说**非简单空间群**顾名思义是那些在陪集分解(2.66)中必须包含一些非零 τ_i 的空间群。然而应当注意到，任何一个点群 K 与和它兼容的晶格放在一起都能够产生简单空间群（只需要机械地做半直积即可），设我们有一个非简单空间群，我们把它的所有 τ_i 都重新设置为零，总能够得到一个合法的简单空间群。因此可以把这个过程反过来，对每个简单空间群，我们尝试向它的一些旋转或是反射操作中按照前面说过的要求——保证旋转轴和某个格矢方向相同， τ_{\parallel} 取特定的值等等——引入非零 τ ，去掉会导致自相矛盾的情况，去掉重复的情况，最终就能够得到所有的非简单空间群。实际上，这个过程还可以简化，因为 τ_{\perp} 实际上同样是高度受限的。从之前的推导可以看出， τ_{\perp} 可以确定各个旋转轴和反射平面的位置，而点群的乘法关系实际上已经将一个点群的空间群的旋转轴和反射平面的位置高度限制了。（例如，旋转轴和反射面的乘积可以产生新的旋转轴和反射面，如果它们的位置不进行适当的匹配，就会没完没了地在一个原胞内产生无数的转轴）通过这种方式无法将所有转轴和反射面的相对位置确定下来，实际上，可以出现这样的情况：将一个非点式空间群的所有 τ_{\parallel} 设为零后无法得到一个合法的点式空间群。然而，由于点群的乘法关系限制了可能的转轴、反射面的相对位置，不需要额外考虑太多情况。总之，我们只需要尝试向点式空间群导入 τ_{\parallel} ，必要时将旋转轴或是反射面移动诸如 $a/2$ 这样的距离，就能够推导出所有的非点式空间群。这正是熊夫利最初推导空间群的方法。

使用这种方法时除了要排除会导致群不封闭等问题的安排以外，还需要小心重复，例如如果向点式空间群导入的那些 τ_{\parallel} 实际上已经存在于点式空间群中了（只需要做一下坐标平移即可），那么就会有重复。

一旦空间群确定了，我们也就确定了原胞中每一个点的对称性，也即，不做整数平移操作，以这个点为中心的群操作构成的对称群。上述讨论——通过其他对称性操作，从一个旋转轴可以到达另一个旋转轴——启发我们，有时若干个点的对称性操作可以通过空间群中的共轭操作联系起来；此时我们就说，它们构成一种Wyckoff位置。例如，对第31号空间群，围绕着形如 $(x, y, z), (-x + 1/2, -y, z + 1/2), (x + 1/2, -y, z + 1/2), (-x, y, z)$ 的四个点的对称性操作彼此共轭等价，围绕着形如 $(0, y, z), (1/2, -y, z + 1/2)$ 的点的对称性操作彼此共轭等价。这说明这种空间群有两种Wyckoff位置。一些人会使用Wyckoff位置作为坐标原点来标记原子位置，因为如果一个位置有原子，那么含有这个位置的一组Wyckoff位置上肯定也有一样的原子。

2.2.6 二维材料

二维材料的晶体结构分析起来要简单很多。可以采用和分析三维晶体的点群和空间群类似的方式做分析，也可以考虑三维晶体的点群和空间群的退化。

二维材料的布拉伐格子可以当成三维材料的布拉伐格子在 z 方向“压扁”的产物，因此只有那些能够画出直棱柱原胞的三维布拉伐格子能够诱导出二维格子。满足该条件的只有单斜、正交、四方、六方四种晶系，因此二维材料也只有这四种晶系，其原胞分别是一个任意的平行四边形、一个长方形、一个正方形和一个有一个角是 60° ，并且四条边长度相等的平行四边形。长方形格子的面心和简单格子不等价，其它几种格子的面心格子和简单格子均等价。因此一共有五种二维布拉伐格子。

类似的，可以发现二维的点群有10种，是五种旋转群以及它们加入反演操作之后得到

的，是

$$C_1, C_2, C_3, C_4, C_6, C_s, C_{2v}, C_{3v}, C_{4v}, C_{6v}.$$

把五种平面布拉伐格子和10种平面点群结合起来，一共能够获得17种二维空间群，或者说平面群，有时候也称为墙纸群，因为墙纸上的重复性图案必定能够分类到这17种平面群中的一种上面。

2.2.7 记号和常用术语

至此我们已经完全推导并分类了理想晶体的所有可能的空间对称性。给定一个晶格，尝试得到一个晶胞，然后分析其格子类型（如金刚石可以分解出一个立方体单元，然后发现它没法画成简单立方格子），最后分析对称群，即确定了晶体类型。本节介绍一些非常常见的晶体结构的“俗名”，以及一种能够让有经验的人一眼看出点群和空间群中包含哪些对称性操作的记号。

常用结构的俗名

fcc和**bcc**分别是面心立方和体心立方的简写，它们互为倒格子。**最密堆积**指的是一系列占据体积的球堆在一起，最为节省空间的排列方式。最密堆积给出一些晶格，可以分成ABAB结构（它是**六方密排结构**，一种复式六方结构）和ABCABC结构（它就是fcc结构）。

金刚石结构顾名思义是金刚石的结构，闪锌矿（一种ZnS晶体）也属于这种结构。它是一种fcc结构。

点群的国际符号

点群的**国际符号**是一种用于简练地表示点群中的对称素，从而标记所有的点群的方法。

空间群的国际符号

我们知道晶体结构确定了它的空间群，而反过来，知道了一个晶体的空间群之后，可以获得晶格的类型，空间群中的对称素也可以提供晶体的基元的一些信息。因此，需要一种能够简要说明空间群的平移部分的种类（即晶格的种类），以及空间群中的生成元的空间群记号。

资料框 2.1: 晶体对称性有关的参考资料

国际晶体学表^[7]（简称“国际表”）是晶体学的最为权威详细的参考资料。

2.3 静态晶格的物理效应

2.3.1 电磁波散射和衍射

设一束电磁波以波矢 \mathbf{k} 照射在晶体上。本节分析晶格中的原子造成的衍射。通常，只有X射线衍射才能够观察到这一类的现象，因为可见光波长远远大于晶格常数。

单个电荷的散射为

$$\mathbf{E} = -\frac{e^2 E_0}{r} \mathbf{e} \times (\mathbf{e} \times \mathbf{e}_0) e^{i\mathbf{k} \cdot \mathbf{R}}, \quad \mathbf{R} = \mathbf{r} - \mathbf{r}'. \quad (2.74)$$

晶体通常不会致密到散射电磁波被多次散射，于是我们假定电磁波只被散射一次，散射波也不会对晶体内部的入射波有非常显著的修正。这样，散射场就是

$$E_{\text{scatter}} = -\frac{e^2 E_0}{r} \mathbf{e} \times (\mathbf{e} \times \mathbf{e}_0) \int d^3 \mathbf{R} n(\mathbf{r}') e^{i\mathbf{k} \cdot \mathbf{R}}. \quad (2.75)$$

在无穷远处散射波肯定是球面波， \mathbf{r} 和 \mathbf{R} 相差不会太大，且波矢 \mathbf{k} 和 \mathbf{r} 方向一致。在固定 r 大小、电场强度等不动，而仅仅改变 \mathbf{k} 指向（也即所谓的弹性散射）时，我们可以用波矢 \mathbf{k} 来标记散射振幅：

$$\begin{aligned} A(\mathbf{k}) &\propto e^{i\mathbf{k} \cdot \mathbf{R}} \int d^3 \mathbf{r}' n(\mathbf{r}') e^{-i\mathbf{R} \cdot \mathbf{k}_0} e^{-i\mathbf{r}' \cdot \mathbf{k}} \\ &\propto \int d^3 \mathbf{r}' n(\mathbf{r}') e^{-i\mathbf{r}' \cdot (\mathbf{k} - \mathbf{k}_0)}. \end{aligned} \quad (2.76)$$

因子 $e^{-i\mathbf{k}_0 \cdot \mathbf{R}}$ 是因为入射光打到不同的位置的相位不同。在晶体中 $n(\mathbf{r}')$ 由周期性的原胞给出，即

$$n(\mathbf{r}') = \sum_{\mathbf{R}_i} n_{\text{u.c.}}(\mathbf{r}' - \mathbf{R}_i), \quad (2.77)$$

代入(2.76)，就得到

$$\begin{aligned} A(\mathbf{k}) &\propto \sum_{\mathbf{R}_i} \int d^3 \mathbf{r}' n_{\text{u.c.}}(\mathbf{r}' - \mathbf{R}_i) e^{-i\mathbf{r}' \cdot (\mathbf{k} - \mathbf{k}_0)} \\ &= \sum_{\mathbf{R}_i} \int d^3 \mathbf{r}' n_{\text{u.c.}}(\mathbf{r}') e^{-i(\mathbf{k} - \mathbf{k}_0) \cdot (\mathbf{r}' + \mathbf{R}_i)} \\ &= \sum_{\mathbf{R}_i} e^{-i(\mathbf{k} - \mathbf{k}_0) \cdot \mathbf{R}_i} \int_{\text{u.c.}} d^3 \mathbf{r}' n_{\text{u.c.}}(\mathbf{r}') e^{-i(\mathbf{k} - \mathbf{k}_0) \cdot \mathbf{r}'} \\ &= N \sum_{\mathbf{G}_m} \delta_{\mathbf{k} - \mathbf{k}_0, \mathbf{G}_m} \int_{\text{u.c.}} d^3 \mathbf{r}' n_{\text{u.c.}}(\mathbf{r}') e^{-i(\mathbf{k} - \mathbf{k}_0) \cdot \mathbf{r}'}. \end{aligned}$$

定义

$$S(\mathbf{p}) = \int_{\text{u.c.}} d^3 \mathbf{r}' n_{\text{u.c.}}(\mathbf{r}') e^{-i\mathbf{p} \cdot \mathbf{r}'} \quad (2.78)$$

为几何结构因子，则散射振幅形如

$$A(\mathbf{k}) \propto N \sum_{\mathbf{G}_m} \delta_{\mathbf{k} - \mathbf{k}_0, \mathbf{G}_m} S(\mathbf{G}_m). \quad (2.79)$$

当晶胞中只有一个原子时这就是原子形状因子，设 $\rho(\mathbf{r})$ 给出了原子中的电子数密度，则

$$S(\mathbf{p}) = F(\mathbf{p}) = \int d^3 \mathbf{r} e^{-i\mathbf{p} \cdot \mathbf{r}} \rho(\mathbf{r}). \quad (2.80)$$

这里要说明一件事：上式的积分是遍及全空间的。 $u_{\text{u.c.}}$ 中不仅包含了本原胞内的电子的数密度，也包含了邻近原胞中的原子附近的电子扩展到本原胞的电子数密度。因此 $n_{\text{u.c.}}$ 中不仅包含本原胞中的原子周围的电子数密度，也要加上邻近原胞的电子数密度在本原胞中的值，我们会发现遍及全空间的积分(2.80)正好是正确的。如果晶格中有多个原子，用 j 标记每个原子，用 \mathbf{r}_j 表示晶胞内各个原子的位置矢量，则

$$\begin{aligned} S(\mathbf{p}) &= \int_{V_{\text{u.c.}}} d^3 \mathbf{r}' \sum_j \rho_j(\mathbf{r}' - \mathbf{r}_j) e^{-i\mathbf{p} \cdot \mathbf{r}'} \\ &= \sum_j e^{-i\mathbf{p} \cdot \mathbf{r}_j} F_j(\mathbf{p}). \end{aligned} \quad (2.81)$$

因此，只有当散射波波矢的变化和某个倒格矢一致时，才有非零的散射振幅，这个条件称为**Laue条件**。但即使散射波波矢为某个倒格矢，由于特殊的对称性可能让结构因子变成零，散射振幅仍然可以是零。这就是说，我们可以将一束非常准直单色的X光打到晶体样品（最好大一些，因为散射振幅正比于 N ）上，然后在不同角度测量散射光光强，这样能够确定晶体中可能的各种倒格矢的方向，从而可以帮助确定晶体晶格的类型。

在光的散射是弹性的时候（即所谓的**布拉格散射**的情况），Laue条件导致

$$|\mathbf{k}| = |\mathbf{k} - \mathbf{G}|,$$

\mathbf{G} 是某个倒格矢。我们看到了熟悉的用于确定布里渊区边界的方程；在这个语境下，垂直平分 \mathbf{G} 的平面称为**布拉格平面**。 \mathbf{k} 和 $\mathbf{k} - \mathbf{G}$ 与布拉格平面的夹角相同，设为 θ ，则

$$|\mathbf{G}| = 2|\mathbf{k}| \sin \theta = 2 \frac{2\pi}{\lambda} \sin \theta.$$

回顾(2.12)，我们发现

$$2 \frac{2\pi}{\lambda} \sin \theta = n \frac{2\pi}{d},$$

即

$$n\lambda = 2d \sin \theta. \quad (2.82)$$

这称为**布拉格衍射条件**。只有满足这个条件才能看到明显的散射。很容易看出，两个相距 nd ，方向平行于 \mathbf{G} 的晶面反射的光发生相长干涉就能够产生满足(2.82)的散射。

2.4 空间群的么正表示和无自旋准粒子

背景知识 2.1: 群表示论和量子力学

如果向量空间 V 上的全体线性算符 $\mathcal{L}(V)$ 的一个子集 Λ 与算符复合构成群，并且存在群同态 $G \rightarrow \Lambda$ ，就称 Λ 是 G 的一个**表示**，称 V 携带了 G 的一个表示，并且记同态映射为 D ，称 $D(g)$ 给出 g 的**表示矩阵**。

以下列出一些常见的群表示论的定理。如果 H 是 G 的子群

设系统哈密顿量 H 在一组构成群的变换下不变，则我们显然有 $[H, g] = 0$ ，其中 g 是任意一个这样的变换。这就意味着，如果 $|n\rangle$ 是一个能量本征态，那么

$$Hg|n\rangle = gH|n\rangle = gE_n|n\rangle,$$

因此 $g|n\rangle$ 是与 $|n\rangle$ 具有同样的能量的能量本征态。于是，设能量为 E_n 的全体能量本征态构成子空间 \mathcal{H}_n ，则 \mathcal{H}_n 携带 G 的一个表示。显然，它能够被拆分成一系列不可约表示的直和：

$$\mathcal{H}_n = \oplus_i a_i \Lambda^i,$$

于是可以用参数 E_n ， \mathcal{H}_n 中的不可约表示的种类 i ，同种不可约表示的编号（1到 a_i 之间的某个正整数 p ），第 p 个 Λ^i 不可约表示的不同基矢量这一系列参数来分类能量本征态。在运气好的时候，通过对称性分析我们能够为每个能量本征态找到唯一的参数组合，因此在没有求解定态薛定谔方程的情况下，我们就能够对希尔伯特空间和系统能谱的结构有一些了解。特别的，能谱实际的简并情况只可能比不可约表示维数所暗示的简并情况高而不可能比它低。

“准粒子”或者说“元激发”，顾名思义，可以用坐标或动量，以及一些离散标签来标记，并且有“单粒子波函数” $\langle \Omega | \psi(\mathbf{r}) | \Psi \rangle$ 。既然准粒子给出了系统的能量本征态，准粒子的多粒子态应该携带空间群的一组表示。特别的，单个准粒子的状态——也就是“单粒子波函数”——携带了空间群的一个么正的场表示，我们下面将这个场表示记作 $D(\{\alpha | \mathbf{t}\})$ 。设 $\{|\mu\rangle\}$ 携带了空间群的 d 维不可约表示 D ，这里 μ 可能包括波矢 \mathbf{k} ，能带编号和自旋， d 维相当于说 d 个波矢是彼此只相差一个点群操作的（从而能量相同的）。简单地通过对称性分析，就可以获得系统中的准粒子的能谱和单粒子波函数的一些性质，而不必关心这些准粒子是怎样的底层机制演生出来的。仅仅依靠对称性我们就能够确定一个多体系统的很多行为，特别是在有其他证据能够说明这些准粒子是什么时候。

本节讨论无自旋的准粒子，它们的波函数只有坐标部分没有自旋部分，空间群作用在波函数的空间部分上。虽然我们时常需要处理有自旋的准粒子——如电子——本节的内容仍然是有用的，首先它是一个好的玩具模型，其次很大一类带自旋的空间群表示也只有空间部分在空间群下会有变化，因此我们常常可以首先分析空间部分然后简单地把自旋加回去。

2.4.1 平移群的表示

设系统的空间群为 G ，由于哈密顿量和 G 中的操作均对易，系统的希尔伯特空间携带 G 的一个表示，且能量本征态同时也是 G 中对称性操作的本征态，而单准粒子波函数也是 G 中对称性操作的本征态。

设 D 是 G 的一个不可约表示。首先考虑 G 的离散平移群部分 T ，由于平移群是阿贝尔群， $D(\{e | \mathbf{r}\})$ 一定是某个复数乘上单位矩阵 I 。显然，我们有

$$T = \{\{e | \mathbf{a}_1\}^{n_1} \{e | \mathbf{a}_2\}^{n_2} \{e | \mathbf{a}_3\}^{n_3} | n_1, n_2, n_3 \in \mathbb{Z}\}, \quad (2.83)$$

由于实际的晶体都是有限大小的，为了便于处理，我们假定

$$\{e | \mathbf{a}_i\}^{N_i} = \{e | 0\}, \quad i = 1, 2, 3, \quad (2.84)$$

很显然这是给晶体设置了周期性边界条件。这样 T 的结构大体上和无限大的平移群一样，但是同时又是有限的。由 Schur 引理，

$$D(\{e | \mathbf{a}_i\})^{N_i} = D(\{e | \mathbf{a}_i\}^{N_i}) = D(\{e | 0\}) = \lambda I,$$

不失一般性地取 $\lambda = 1$ ，就得到

$$D(\{e | \mathbf{a}_i\}) = \exp\left(-\frac{i2\pi}{N_i} p_i\right), \quad p_i = 0, 1, \dots, N_i - 1. \quad (2.85)$$

这意味着 T 的不可约表示的标签一定包括 (p_1, p_2, p_3) ，而 (p_1, p_2, p_3) 确定下来之后， T 的不可约表示就完全确定了：我们有

$$D(\{e | n_1 \mathbf{a}_1 + n_2 \mathbf{a}_2 + n_3 \mathbf{a}_3\}) = \exp\left(-i \sum_{i=1}^3 \frac{2\pi}{N_i} p_i n_i\right). \quad (2.86)$$

我们马上发现，设

$$\mathbf{k} = \frac{p_1}{N_1} \mathbf{b}_1 + \frac{p_2}{N_2} \mathbf{b}_2 + \frac{p_3}{N_3} \mathbf{b}_3, \quad (2.87)$$

则可以使用 \mathbf{k} 标记 D ：设 \mathbf{R}_i 是一个正格矢，则

$$D^{\mathbf{k}}(\{e | \mathbf{R}_i\}) = e^{-i\mathbf{k} \cdot \mathbf{R}_i}. \quad (2.88)$$

由于指数函数的周期性，设 \mathbf{G}_m 是一个任意的倒格矢，那么就有

$$e^{-i\mathbf{k}\cdot\mathbf{r}} = e^{-i(\mathbf{k}+\mathbf{G}_m)\cdot\mathbf{r}},$$

因此，没有必要让 \mathbf{k} 取遍整个空间。我们可以要求 \mathbf{k} 在任何一个倒空间的原胞内部，因为对任何一个 \mathbf{k} ，都存在一个局限在某个原胞内部的 \mathbf{k}' ，使得

$$e^{-i\mathbf{k}\cdot\mathbf{r}} = e^{-i\mathbf{k}'\cdot\mathbf{r}},$$

而另一方面，原胞内部的矢量彼此的差值不可能达到一个倒格子格矢的大小，因此原胞内部的 \mathbf{k} 不会导致重复的 D 。

以上分析都没有指出表示空间是什么。现在我们设 $D^{\mathbf{k}}$ 作用在波函数（无论是哪种粒子的波函数）组成的希尔伯特空间上。显然此时只可能有所谓场表示，因为有平移，即由

$$D(g)\phi(\mathbf{r}) = \phi(g^{-1}\mathbf{r})$$

给出的表示，在这里就是

$$D(\{e|\mathbf{r}'\})\phi(\mathbf{r}) = \phi(\mathbf{r} - \mathbf{r}'). \quad (2.89)$$

设有一组 $\psi(\mathbf{r})$ 是按照(2.89)定义的 D 的本征函数，且它们张成的空间携带了一个不可约表示，则我们可以用 \mathbf{k} 标记它，有

$$\psi(\mathbf{r} - \mathbf{R}_i) = D^{\mathbf{k}}(\{e|\mathbf{R}_i\})\psi(\mathbf{r}) = e^{-i\mathbf{k}\cdot\mathbf{R}_i}\psi(\mathbf{r}), \quad (2.90)$$

或者说

$$\psi(\mathbf{r} + \mathbf{R}_i) = e^{i\mathbf{k}\cdot\mathbf{R}_i}\psi(\mathbf{r}). \quad (2.91)$$

考虑到不同的 \mathbf{k} 都能够让(2.91)成立，且指定了 \mathbf{k} 之后 ψ 还是不止一种，我们发现离散空间平移对称群在波函数上的场表示是可约的，可以使用 \mathbf{k} 和一个额外的标记 n 区分。同一个 \mathbf{k} ，不同 n 的 $\psi_{n\mathbf{k}}$ 组成的波函数组携带 T 的一个不可约表示（但是完整的空间群的表示则会含有多个 \mathbf{k} ，因为空间群操作可以在不同 \mathbf{k} 之间跳转）， n 则区分这个表示空间的不同基矢量。

总之，(2.91)构成晶体的态空间的单准粒子空间的一组基，且可以用好量子数 \mathbf{k} 和另一个好量子数 n 标记它。 \mathbf{k} 称为晶格动量或者准动量。它的性质类似动量，但并不是自由空间中的动量。从以上各式（比如(2.91)）中可以看出，这里的 \mathbf{k} 和给定一个单粒子哈密顿量，通过傅里叶变换对角化时引入的 \mathbf{k} 应该是一样的（至多因为记号不同差一个 -1 ），而没有相差某个尺度变换。 \mathbf{k} 的取值范围如前所述被约束在一个倒空间原胞中，且由于周期性边界条件，其取值是离散的。虽然定义 \mathbf{k} 时用到了分量，但由于晶体中倒格子基矢量本来就是三个特殊方向， \mathbf{k} 是坐标系无关的真正的矢量。满足(2.91)的波函数在空间平移下多出来了一个因子，这是合理的，因为波函数的对称性可以略微低于哈密顿量，只要由它计算出的物理量的对称性和哈密顿量一致就可以。(2.91)又说明，我们可以设

$$\psi_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}u_{\mathbf{k}}(\mathbf{r}), \quad u_{\mathbf{k}}(\mathbf{r} + \mathbf{a}_i) = u_{\mathbf{k}}(\mathbf{r}), \quad i = 1, 2, \dots \quad (2.92)$$

因此，晶体中的单粒子波函数是一个受到一个周期为 $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ 的振幅调制的平面波。这个结论称为Bloch定理，(2.92)称为Bloch波函数。

2.4.2 对称性操作在Bloch态上的作用

在进一步讨论如何构造空间群的表示之前我们先讨论各种对称性操作作用在Bloch波函数上面会有什么样的结果。

时间反演不变性带来的简并

如果没有外界磁场，晶体就具有时间反演对称性。时间反演算符是反幺正的，在时间反演变换下哈密顿量不变， \mathbf{r} 不变，但是动量要变号，自旋也反号。对没有自旋的准粒子，无需考虑自旋，只需要分析空间部分。

我们先来考虑无自旋准粒子的Bloch波函数和能量本征值在时间反演变换下如何变动。设时间反演算符为 T （不要和平移群混淆）。首先，哈密顿量在时间反演变换下保持不变，于是本征方程

$$H\psi_{n\mathbf{k}} = \epsilon_{n\mathbf{k}}\psi_{n\mathbf{k}}$$

经过时间反演之后会得到

$$H\psi_{n\mathbf{k}}^* = \epsilon_{n\mathbf{k}}\psi_{n\mathbf{k}}^*,$$

其中

$$T\psi_{n\mathbf{k}}(\mathbf{r}) = \psi_{n\mathbf{k}}^*(\mathbf{r})$$

是 $\psi_{n\mathbf{k}}$ 的时间反演，它也是以 $\epsilon_{n\mathbf{k}}$ 为能量本征值；我们下面分析 $\psi_{n\mathbf{k}}^*(\mathbf{r})$ 的标签。由于

$$\psi_{n\mathbf{k}}(\mathbf{r})^* = e^{-i\mathbf{k}\cdot\mathbf{r}} u_{n\mathbf{k}}(\mathbf{r})^*,$$

$\psi_{n\mathbf{k}}^*$ 是一个Bloch波矢为 $-\mathbf{k}$ 的Bloch波函数，并且是能量为 $\epsilon_{n\mathbf{k}}$ 的能量本征态，从而必然可以在某条能带上找到它。我们通常认为系统的能谱具有很好的解析性质，因此一条能带 $\epsilon = \epsilon_{n\mathbf{k}}$ 上的各个态经过时间反演变换后应该整齐地被映射到另一条能带上，否则 ϵ 与 \mathbf{k} 之间的关系会出现间断。现在我们发现第 n 号能带上的每个能级经过时间反演变换后，都被映射到一个能量和 n 号能带上的能级一样的能级上，如果 n 号能带没有和另一条能带完全重合，这就说明一件事：时间反演变换实际上将 n 号能带映射成了它自身，只不过重新调整了 \mathbf{k} ，如前所述，是将 \mathbf{k} 变换成 $-\mathbf{k}$ 。如果 n 号能带和另一条能带完全重合，此时能带编号的定义存在任意性，因此我们总是可以重新定义能带编号（与此同时将 \mathbf{k} 相同的态做线性组合），让能带编号在时间反演变换下保持不变，于是时间反演变换还是将 n 号能带映射成了它自身。

综上，时间反演变换将一条能带转化成它自身，即不改变能带编号，而将Bloch波矢取反号，并且有

$$\epsilon_{n\mathbf{k}} = \epsilon_{n,-\mathbf{k}}. \quad (2.93)$$

这个结论的成立不依赖晶体是否具有空间反演对称性，因为它是时间反演对称性的结果。

对自旋1/2的准粒子——比如说电子——在自旋翻转不变性不成立，只有时间反演不变性成立时，没有(2.93)，但是注意到，设 $|\mu\rangle$ 是一个能量本征态，如果 $T|\mu\rangle$ 和 $|\mu\rangle$ 线性相关，则

$$T|\mu\rangle = c|\mu\rangle,$$

于是

$$T^2|\mu\rangle = T(c|\mu\rangle) = c^*T|\mu\rangle = |c|^2|\mu\rangle, \quad |c|^2 > 0,$$

但是对自旋1/2的准粒子， $T^2 = -1$ ，这就产生了矛盾。因此一个准粒子态经过时间反演算符的作用以后是不可能直接回到自身的。由于时间反演不变性成立， $|\mu\rangle$ 和 $T|\mu\rangle$ 是具有同样的能量本征值的，因此只要时间反演不变性成立，自旋1/2的准粒子的能级至少有二重简并。这称为**Kramers定理**。在不存在自旋-轨道耦合，从而系统的能量本征态有确定的自旋标签时，我们有

$$\epsilon_{n\mathbf{k}\uparrow} = \epsilon_{n,-\mathbf{k},\downarrow}. \quad (2.94)$$

对无自旋的粒子，由于有可能 $T|\mu\rangle$ 和 $|\mu\rangle$ 是一回事，不能确定时间反演是不是一定导致能级简并。

空间群操作带来的简并

如果空间群操作 $\{\alpha|\mathbf{t}\}$ 是晶体的对称操作，则它和哈密顿量对易，从而

$$HD(\{\alpha|\mathbf{t}\})\psi_{n\mathbf{k}} = D(\{\alpha|\mathbf{t}\})H\psi_{n\mathbf{k}}.$$

注意到

$$D(\{\alpha|\mathbf{t}\})\psi_{n\mathbf{k}}(\mathbf{r}) = \psi_{n\mathbf{k}}(\alpha^{-1}(\mathbf{r} - \mathbf{t})),$$

由于准粒子波函数携带空间群的表示（注意平移操作保持 \mathbf{k} 不变，因为 \mathbf{k} 标记了平移群的不同不可约表示，但是一般空间群操作不保持 \mathbf{k} 不变）， $D(\{\alpha|\mathbf{t}\})\psi_{n\mathbf{k}}(\mathbf{r})$ 也应该是一个Bloch波函数。由于

$$\begin{aligned} D(\{\alpha|\mathbf{t}\})\psi_{n\mathbf{k}}(\mathbf{r} + \mathbf{R}_i) &= \psi_{n\mathbf{k}}(\alpha^{-1}(\mathbf{r} - \mathbf{t} + \mathbf{R}_i)) \\ &= \psi_{n\mathbf{k}}(\alpha^{-1}(\mathbf{r} - \mathbf{t}))e^{i\mathbf{k}\cdot\alpha^{-1}\mathbf{R}_i} \\ &= \psi_{n\mathbf{k}}(\alpha^{-1}(\mathbf{r} - \mathbf{t}))e^{i(\alpha\mathbf{k})\cdot\mathbf{R}_i}, \end{aligned}$$

因此 $D(\{\alpha|\mathbf{t}\})\psi_{n\mathbf{k}}(\mathbf{r})$ 是Bloch态，Bloch波矢为 $\alpha\mathbf{k}$ ，可以在某条能带上找到它。与分析时间反演变换时类似，空间群操作也只会将一条能带完整地映射到另一条能带上，根据与分析时间反演变换时类似的论证，我们确信空间群操作不会改变能带编号，或者只是将几条完全重叠的能带彼此映射。一个空间群操作 α ，时间反演操作，哈密顿量彼此对易，因此在有多条能带的能量完全相同时，总是可以找到适当的能带编号安排，使得能带编号在 α 和时间反演操作下都保持不变。然而不能保证在一个空间群操作 α 下不变的能带编号安排在另一个空间群操作下仍然保持不变。总之，设 n 是空间群操作 α 下不变的某种能带编号安排方式，则

$$\epsilon_{n\mathbf{k}}D(\{\alpha|\mathbf{t}\})\psi_{n\mathbf{k}} = D(\{\alpha|\mathbf{t}\})H\psi_{n\mathbf{k}} = HD(\{\alpha|\mathbf{t}\})\psi_{n\mathbf{k}} = H\psi_{n,\alpha\mathbf{k}} = \epsilon_{n,\alpha\mathbf{k}}\psi_{n,\alpha\mathbf{k}},$$

于是

$$\epsilon_{n\mathbf{k}} = \epsilon_{n,\alpha\mathbf{k}}. \quad (2.95)$$

换言之，能带的对称性和空间群的 K 点群完全一致，而和分数平移没有关系。

由于空间群操作不会改变能带编号，一条能带中的模式可以写成多个空间群不可约表示的表示空间的直和。一条能带中显然有多个不可约表示，因为设有两个间隔非常小的 \mathbf{k} ，则没有任何空间群操作能够让它们互相转化。因此，如果将 \mathbf{k} 限制在一个不可约体积（见后文）中，那么 \mathbf{k} 和 $\alpha\mathbf{k}$ 不会一起出现，那么每一个 \mathbf{k} 都对应着一份空间群的不可约表示，相互不重叠。

波矢群

现在我们看到了空间群是如何作用在 \mathbf{k} 上的。只有非平移的空间群操作可以让 \mathbf{k} 变动，因此，空间群的一个不可约表示需要多个不同 \mathbf{k} 的平移群的不可约表示拼合在一起而得到。一个空间群的不可约表示的表示空间中包含一系列能带编号相同，波矢之间可以通过某个 α 相互切换的Bloch波函数，它们的能量是一样的。

如果

$$\alpha\mathbf{k} = \mathbf{k} + \mathbf{G}_m, \quad (2.96)$$

我们就说 α 把 \mathbf{k} 变换到了一个等价的波矢上。空间群中所有将波矢 \mathbf{k} 转化到与之等价的波矢上的操作组成的群

$$G_{\mathbf{k}} = \{ \{ \alpha | \mathbf{r} \} \in G | \alpha \mathbf{k} = \mathbf{k} + \mathbf{G}_m \} \quad (2.97)$$

称为 \mathbf{k} 的小群或者波矢群，它是 G 的子群，并且 T 显然是它的不变子群。其点群部分

$$G_{\mathbf{k}}^{(0)} = \{ \alpha | \{ \alpha | \mathbf{r} \} \in G, \alpha \mathbf{k} = \mathbf{k} + \mathbf{G}_m \} \quad (2.98)$$

则是 K 的子群。既然 $G_{\mathbf{k}}^{(0)}$ 是 K 的子群，可以做陪集分解

$$K = G_{\mathbf{k}}^{(0)} \otimes \{ \alpha | 0 \} G_{\mathbf{k}}^{(0)} \otimes \cdots, \quad (2.99)$$

其中陪集代表元数目为

$$m_{\mathbf{k}} = |K| / |G_{\mathbf{k}}^{(0)}|. \quad (2.100)$$

容易看出将 K 中所有的操作作用在 \mathbf{k} 上，得到的彼此不等价（即不仅仅只是差一个倒格矢）的 \mathbf{k} 共有 $m_{\mathbf{k}}$ 个。我们将它们记作 \mathbf{k}^* 。 \mathbf{k}^* 越大则 $G_{\mathbf{k}}^{(0)}$ 越小，即 \mathbf{k} 点的位置越“不对称”。容易看出不同的 \mathbf{k} 对应的 \mathbf{k}^* 要么完全重合，要么完全不同，于是全体 \mathbf{k}^* 实际上构成了第一布里渊区中的全体 \mathbf{k} ——或者等价地说， T 群的不可约表示，共计有 N 个——的一个划分，有

$$\sum_{\mathbf{k}^*} m_{\mathbf{k}} = N. \quad (2.101)$$

能带简并和相容性

能带上的各个模式当然也携带了波矢群的不可约表示。波矢群可以用来分析固定 \mathbf{k} 后存在的简并，即同一个 \mathbf{k} ，不同能带编号的模式怎么就发生简并了。

高对称点具有更大的波矢群。从一个对称性较高的点往对称性较低的点走，相应的 \mathbf{k} 对应的波矢群瞬间变小，然而能带是连续的，这实际上会对能带可能能够取的形式提出了特殊的要求。设 \mathbf{k}_1 点具有比与它非常接近的 \mathbf{k}_2 点更大的波矢群，则 $G_{\mathbf{k}_2}^{(0)}$ 是 $G_{\mathbf{k}_1}^{(0)}$ 的子群。于是， $G_{\mathbf{k}_1}^{(0)}$ 的每个不可约表示都可以拆分成 $G_{\mathbf{k}_2}^{(0)}$ 的一个或几个不可约表示，也只能如此拆分。现在设 \mathbf{k}_1 点有 m 个模式按照 $G_{\mathbf{k}_1}^{(0)}$ 的不可约表示 Γ 变换，而

$$\Gamma = \sum_{i=1}^p \Gamma_i, \quad (2.102)$$

其中 Γ_i 是 $G_{\mathbf{k}_2}^{(0)}$ 的一些不可约表示，则由能带的连续性，在 \mathbf{k}_2 点，能量在前述模式附近处，有 m_1 个模式按照 Γ_1 变换， m_2 个模式按照 Γ_2 变换，等等，且所有 m_i 之和就是 m 。按照 $\Gamma_1, \Gamma_2, \dots, \Gamma_p$ 变换的这些模式的能量可以各不相同。这就导出能带相容性：设我们有 m 条能带，它们在 \mathbf{k}_1 点构成 $G_{\mathbf{k}_1}^{(0)}$ 的一个 m 维不可约表示，因而在 \mathbf{k}_1 点简并，则到了波矢群降低的 \mathbf{k}_2 点，根据 $G_{\mathbf{k}_2}^{(0)}$ 做 Γ 的直和分解，这原本重叠在一起的 m 条能带最多分裂成 p 支，每支上有 m_i 重简并。

也可以反过来考虑这个问题。设在波矢群为 $G_{\mathbf{k}_2}^{(0)}$ 的 \mathbf{k} 点，有 m_1 条能带上的模式构成 $G_{\mathbf{k}_2}^{(0)}$ 的表示 Γ_1 ， m_2 条能带上的模式构成 $G_{\mathbf{k}_2}^{(0)}$ 的表示 Γ_2 ，等等。在这些 \mathbf{k} 点组成的空间上，能带编号可以分成两部分，一部分区分不同的 m_i ，一部分区分每个 Γ_i 内部不同的基矢量。 $G_{\mathbf{k}_2}^{(0)}$ 能够改变后者，但是不能改变前者。但在某个波矢群更大的点 \mathbf{k}_1 ，波矢群中的操作同样能够改变前者，于是这所有的 $\Gamma_1, \Gamma_2, \dots, \Gamma_p$ 混合在了一起，形成了一个大的不可约表示，这 m 个能带在 \mathbf{k}_1 就简并了。

第一布里渊区、不可约体积和高对称点

我们还会发现将倒空间的原胞设置为第一布里渊区实际上是最好的选择。晶体中的准粒子哈密顿量总可以看成某个对称性高一些的、更加简单的哈密顿量 h_0 加上对称性就是 G 的微扰。加入微扰后，倒空间原胞缩小，根据 h_0 计算出的某一个模式的能谱将会因此出现自我交叉，而微扰导致的能级修正会让由此产生的简并消失一部分。对称性为 G 的微扰是周期性的，从而，如果两个模式 \mathbf{k} 和 \mathbf{k}' 之间存在因为微扰导致的跃迁，那么必然

$$\mathbf{k} = \mathbf{k}' + \mathbf{G}_m,$$

而由于它们之间要存在简并，需要有 $\omega_{\mathbf{k}} = \omega_{\mathbf{k}'}$ 。由于能带具有和第一布里渊区一样的对称性以及空间反演不变性，我们会发现（通过穷举各种晶格，或者一些几何上的论证TODO），如果取第一布里渊区作为倒空间原胞，则定义在原布里渊区与新布里渊区的差集中的那部分能谱通过用新的倒格矢做平移变换放进缩小了的第一布里渊区之后，缩小了的第一布里渊区边界上仍然处处有能量简并。这是很不平凡的，因为用新的倒格矢做平移后，原本能量简并、 \mathbf{k} 很接近的两个态会被移开；这要求任意给定一个波矢 \mathbf{k} ，设 α 是空间群的点群中的一个操作，则 $\alpha\mathbf{k}$ 也在同一个倒空间原胞中，直观地看这是要求倒空间原胞尽可能的“像一个球”，这正好就是第一布里渊区的直观概念。此外，不失一般性地，认为 h_0 是空间旋转不变的，则据此计算出的能谱只和 $|\mathbf{k}|$ 有关。现在考虑完整的哈密顿量，则条件

$$\mathbf{k} = \mathbf{k}' + \mathbf{G}_m, \quad \epsilon_{\mathbf{k}} = \epsilon_{\mathbf{k}'}$$

意味着

$$|\mathbf{k}| = |\mathbf{k} - \mathbf{G}_m|,$$

或者说

$$\mathbf{G} \cdot (\mathbf{G} - 2\mathbf{k}) = 0,$$

即微扰导致的简并能级解除或是发生在缩小后的第一布里渊区内部，或是发生在缩小后的第一布里渊区边界上，或是发生在缩小后的第二布里渊区边界上，等等。倒格矢平移会将第 n 布里渊区边界移动到第一布里渊区内部，从而在第一布里渊区内部的一些面上产生简并解除，但注意到这样一个面的两端的模式只差了另一个 \mathbf{G}_n ，简并解除后第一布里渊区内部仍然有连续的能带。因此我们得出结论：第一布里渊区内部的能谱总是连续的（我们确定傅里叶变换能够让哈密顿量对角化，而因为傅里叶变换本身是一个连续的变化， \mathbf{k} 的小幅变化只会让实空间中的波函数有小的变化，而 \mathbf{k} 的小幅变化是从一个能量本征态跳到另一个能量本征态，只要哈密顿量性质良好，两个相邻的 \mathbf{k} 对应非常接近的能量本征态，从而对应非常接近的能量），能谱的不连续总是可以认为是出现在第一布里渊区边界上。（当然，依照前述论证，如果将第一布里渊区内部的能带整体平移到第二、第三、……布里渊区中，第一布里渊区的边界上的能带仍然是连续的，但是这样就重复计数了准粒子模式）因此以下无特殊说明，我们总是将倒空间原胞——无论讨论的是何种准粒子——取为第一布里渊区，并且将布里渊区缩小称为**布里渊区折叠**。

直观地看，布里渊区边界上发生的简并解除可以做如下理解：设想向晶体中引入某种源，来产生一种准粒子流，则动量相差了某个 \mathbf{G}_m 的两个模式满足布拉格散射的Laue条件，从而相应的动量为 \mathbf{k} 和 $\mathbf{k} + \mathbf{G}_m$ 的准粒子流不可能在晶体中存在，因为它们往任何方向走都只能受到强烈的反射。因此，动量为 \mathbf{k} 和 \mathbf{G}_m 而能量相同的准粒子模式是不可能存在的，如果晶格势能引入之前有这样的简并，那么晶格势能必然要破坏这个简并，形成禁带。这个

直观的图像又导致如下的关于能谱形状的自然性条件：群速度在所有布里渊区边界上垂直于这些边界的分量必须是零。这是因为在这些边界上的准粒子模式是禁带的极限情况，在实空间中不应该能够传播，因此代表“跨原胞传播”的群速度的垂直分量应为零（但水平分量可以——而且通常——不为零）。

设晶体点群为 P ，则第一布里渊区实际上可以被分成 $|P|$ 份，从其中一份就可以重构出其它 $|P| - 1$ 份。我们将这 $|P|$ 份第一布里渊区中的任意一份称为不可约体积。我们希望这所有的共计 P 份不可约体积都是全等的，并且是多面体。此时会发现，其中任意一份不可约体积的顶点、棱、面都具有较高的对称性：显然，两个不可约体积中的点可以通过某个 K 点群操作彼此切换，但是不可约体积的顶点和棱为两个不可约体积共享，那么一些点群操作实际上是将它们转化成它们自己——这些顶点、棱、面具有非平庸的 $G_k^{(0)}$ ，可以称为高对称点，高对称点中的顶点彼此连接得到的线条就是不可约体积的棱，称为高对称线，同样还有高对称面。某条高对称线上各点的 $G_k^{(0)}$ 都是一样的，高对称面也是这样。不难看出高对称线的位置由第一布里渊区的对称轴、镜像平面等确定。显然第一布里渊区的边界也是高对称点、线、面确定的，而当外界扰动引入时，这些高对称点、线、面肯定会有变动，出现布里渊区折叠。

最后提一句：倒格子中的格点的分布当然和正格子有关，如果我们选择了一个并非原胞的晶胞，那么根据这个晶胞画出来的倒格子要比正确的倒格子小，倒格子原胞是正确的倒格子原胞的整数分之一倍。因此根据一个并非原胞的晶胞绘制晶体中准粒子的能谱（这种做法是很常见的，如见第7.3.1节）时在第一布里渊区边界上会有大量简并，因为有布里渊区折叠，但并没有什么真的破坏了周期性，因此没有能隙打开。我们后面在讨论晶体中的各种准粒子时还会含糊其辞地使用“晶胞”的概念，此时的倒空间默认就是相应的晶胞对应的倒空间。

晶体的空间群确定后，不可约体积、高对称点什么的全部确定了，而分解(2.102)则可以通过点群的特征标轻易得到。

习惯上，绘制三维晶体的能带时基本上只是把不可约体积的棱过一遍。不同的第一布里渊区的棱有标准的记号，如 Δ, Σ 等，由于每种类型的棱对应一个非平凡的波矢群，实际上可以用这些棱的字母来标记不同点群的表示，如 C_{4v} 群是 Δ 棱的对称群，于是 C_{4v} 的五个表示可以写成 Δ_1 到 Δ_5 。

2.4.3 构造空间群的不可约表示

简单空间群的不可约表示的构造

我们已经分析了平移群和点群的表示了，最后要做的就是把它们放在一起，构造出空间群的表示。我们先从较简单的简单空间群或者说点式空间群开始，具体地讨论如何构造空间群的表示，以及如何直接从空间群获得关于能带的一些信息。例如，既然准粒子实际上是空间群的不可约表示中的基矢量，可以直接通过对称性决定它们的能谱的很大一部分性质，主要是各种能级简并。如果对系统的微扰没有破坏系统的对称性，那么一个不可约表示中的那些Bloch波函数之间的能量简并是不会解除的。简单空间群 G 总是可以写成半直积

$$G = K \otimes T. \quad (2.103)$$

这其中， T 决定了三个倒空间基元矢量的方向，从而决定了第一布里渊区的形状，而 K 则决定了第一布里渊区的不可约体积，高对称点、线、面等性质。

我们使用第2.4.2节中的记号。我们的目标是通过 $G_{\mathbf{k}}^{(0)}$ 的不可约表示 $D_{G_{\mathbf{k}}^{(0)}}^{(p)}$ （已知的，因为 $G_{\mathbf{k}}^{(0)}$ 是32种点群之一，而点群的不可约表示较容易求出）诱导出整个空间群的不可约表示。由于我们有(2.103)，首先构造 K 的表示，然后将它和 T 的表示直积起来，总是能够获得 G 的一个表示。由于 \mathbf{k}^* 是 K 作用在 \mathbf{k} 上得到的不重复的波矢的集合，我们发现 K 的子群 $G_{\mathbf{k}}^{(0)}$ 实际上是作用在 \mathbf{k}^* 上的，根据陪集分解(2.99)，可以获得 K 关于其子群 $G_{\mathbf{k}}^{(0)}$ 的诱导表示，该诱导表示作用在 $D_{G_{\mathbf{k}}^{(0)}}^{(p)} \otimes \mathbf{k}^*$ 上，作用在前一个集合上的方式为基础表示，作用在后一个集合上的方式由 $D_{G_{\mathbf{k}}^{(0)}}^{(0)}$ 给出，于是我们发现该诱导表示可以写成

$$D_K^{(kp)}(\{\alpha|\mathbf{r}_n\})_{\gamma r, \beta s} = \begin{cases} e^{-i(\alpha_\gamma \mathbf{k}) \cdot \mathbf{r}_n} D_{G_{\mathbf{k}}^{(0)}}^{(p)}(\alpha_\gamma^{-1} \alpha \alpha_\beta)_{rs}, & \text{if } \alpha_\gamma^{-1} \alpha \alpha_\beta \in G_{\mathbf{k}}^{(0)}, \\ 0, & \text{otherwise,} \end{cases}$$

其中指标 γ 和 β 取值为1到 $m_{\mathbf{k}}$ ，即取遍所有波矢 \mathbf{k}^* 的成员，而 r 和 s 取遍1到 d_p ，即取遍所有 $D_{G_{\mathbf{k}}^{(0)}}^{(p)}$ 的成员。上标 kp 代表着这一类表示可以使用 \mathbf{k} 和 p 两个参数加以标记。以上表示可以说是实现了 $\{\alpha|\mathbf{r}_n\}$ 的点群部分，在完整的空间群的表示种，对从 βs 到 γr 的矩阵元，其平移部分是作用在 α_γ 上的，于是直积上平移群 T 的（一维）表示 $e^{-i\alpha \mathbf{k} \cdot \mathbf{r}_n}$ ，我们获得了

$$D_G^{(kp)}(\{\alpha|\mathbf{r}_n\})_{\gamma r, \beta s} = \begin{cases} e^{-i(\alpha_\gamma \mathbf{k}) \cdot \mathbf{r}_n} D_{G_{\mathbf{k}}^{(0)}}^{(p)}(\alpha_\gamma^{-1} \alpha \alpha_\beta)_{rs}, & \text{if } \alpha_\gamma^{-1} \alpha \alpha_\beta \in G_{\mathbf{k}}^{(0)}, \\ 0, & \text{otherwise.} \end{cases} \quad (2.104)$$

这是 $d_p m_{\mathbf{k}}$ 维表示。

表面上看，(2.104)只是一些可能的表示。然而我们下面将要论证说，任何一个(2.104)都是不可约的，并且全体(2.104)给出了 G 的全部不可约表示。

背景知识 2.2: 有限维不可约表示的一些性质

用 p, q 等表示不可约表示的编号，用 g 等表示群元，群的阶记为 n 。用 $\chi^{(p)}(g)$ 表示 p 号不可约表示中群元 g 的特征标。

关于不可约表示的数目和维数，我们有：

- 不可约表示的数目就是共轭类的数目。
- 不可约表示维数的平方和就是群的阶（推论：阿贝尔群的不可约表示全部是一维的）。
- 我们有正交关系

$$\frac{1}{n} \sum_{g \in G} \chi^{(p)}(g)^* \chi^{(q)}(g) = \delta_{pq}, \quad (2.105)$$

或者，设共轭类 $[g]$ 一共有 k 个， k_i 表示第 i 个等价类中的元素数目，则

$$\frac{1}{n} \sum_{i=1}^k k_i \chi^{(p)}([g_i])^* \chi^{(q)}([g_i]) = \delta_{pq}. \quad (2.106)$$

推论：特征标的模长平方和大于等于 n ，如果是不可约表示就是 n ，否则比 n 大。

- 在 g 和 g' 在同一个等价类中时

$$\frac{1}{|G|} \sum_{\chi^{(p)}} \chi^{(p)}(g)^* \chi^{(p)}(g') = \frac{1}{|[g]|}, \quad (2.107)$$

否则为零。这也可以用于判断是否找到了全部不可约表示，因为如果 g 为单位元，且 p 扫过了所有不可约表示，那么就有

$$\sum_{\chi^{(p)}} |\chi^{(p)}(e)|^2 = \sum_{\chi^{(p)}} d_p^2 = |G|, \quad (2.108)$$

如果 p 只是扫过了部分不可约表示那么上式左边就要小于 $|G|$ 。

我们将使用(2.105)来论证(2.104)的不可约性。我们有

$$\chi_G^{(kp)}(\{\alpha|\mathbf{r}_n\}) = \sum_{\beta} e^{-i(\alpha_{\beta}\mathbf{k})\cdot\mathbf{r}_n} \chi_{G_{\mathbf{k}}^{(0)}}^{(p)}(\alpha_{\beta}^{-1}\alpha\alpha_{\beta}).$$

计算特征标的平方和，我们有

$$\begin{aligned} \sum_{\mathbf{r}_n, \alpha} |\chi_G^{(kp)}(\{\alpha|\mathbf{r}_n\})|^2 &= \sum_{\mathbf{r}_n, \alpha} \sum_{\beta, \gamma} e^{i(\alpha_{\gamma}\mathbf{k} - \alpha_{\beta}\mathbf{k})\cdot\mathbf{r}_n} \chi_{G_{\mathbf{k}}^{(0)}}^{(p)}(\alpha_{\beta}^{-1}\alpha\alpha_{\beta}) \chi_{G_{\mathbf{k}}^{(0)}}^{(p)}(\alpha_{\gamma}^{-1}\alpha\alpha_{\gamma})^* \\ &= \sum_{\alpha} \sum_{\beta, \gamma} N \delta_{\gamma\beta} \chi_{G_{\mathbf{k}}^{(0)}}^{(p)}(\alpha_{\beta}^{-1}\alpha\alpha_{\beta}) \chi_{G_{\mathbf{k}}^{(0)}}^{(p)}(\alpha_{\gamma}^{-1}\alpha\alpha_{\gamma})^* \\ &= \sum_{\alpha} \sum_{\beta} N \chi_{G_{\mathbf{k}}^{(0)}}^{(p)}(\alpha_{\beta}^{-1}\alpha\alpha_{\beta}) \chi_{G_{\mathbf{k}}^{(0)}}^{(p)}(\alpha_{\beta}^{-1}\alpha\alpha_{\beta})^* \\ &= N m_{\mathbf{k}} \sum_{\alpha} \chi_{G_{\mathbf{k}}^{(0)}}^{(p)}(\alpha) \chi_{G_{\mathbf{k}}^{(0)}}^{(p)}(\alpha)^* \\ &= N m_{\mathbf{k}} |G_{\mathbf{k}}^{(0)}| = N |G_{\mathbf{k}}| = |G|. \end{aligned}$$

第四个等号用到了同一个共轭类中的元素特征标相同这一事实。从上式可以看出，一个形如(2.104)的表示的确是一个不可约表示。

现在我们再使用(2.108)分析是否(2.104)给出了全部不可约表示。所有表示的维数的平方和（注意属于同一 \mathbf{k}^* 的波矢对应的(2.104)相同）为

$$\sum_{p, \mathbf{k}^*} (d_p m_{\mathbf{k}})^2 = \sum_{\mathbf{k}^*} m_{\mathbf{k}}^2 |G_{\mathbf{k}}^{(0)}| = \sum_{\mathbf{k}^*} m_{\mathbf{k}} |K| = N |K| = |G|.$$

第一个等号是因为 p 标记的表示均为不可约表示，从而

$$\sum_p d_p^2 = |G_{\mathbf{k}}^{(0)}|.$$

因此，我们发现(2.104)确实是给出了所有的不可约表示。

总之，在简单空间群中，小群 $G_{\mathbf{k}}$ 是构造整个空间群的不可约表示的关键。直观地说，(2.104)中的 $e^{-i\mathbf{k}\cdot\mathbf{r}}$ 因子给出了Bloch态的外部的空间变化，而小群部分则给出了若干简并的能带的“内部结构”：小群操作让对应小群的表示 $D_{G_{\mathbf{k}}^{(0)}}^{(p)}$ 的表示空间的不同基矢量的Bloch波函数彼此转化，从而导致一些能带能量完全相同。相同的构造也用于粒子物理中，在那里小群保持粒子动量（真正的动量，而不是差一个格矢就认为彼此等价的晶格动量）不变，从而给出粒子的自旋结构。

非简单空间群的不可约表示

由于空间群操作作用在 \mathbf{k} 上时只有其点群部分才有效果，如果一个非点式空间群和一个点式空间群具有一样的布拉伐格子（或者说 T ）和 K 点群，则两者的第一布里渊区形状、高对称点、线、面位置等倒空间性质是完全一样的。

2.5 空间群的么正表示和自旋1/2粒子

2.5.1 双群及其表示

双群的构造

自旋1/2粒子实际上是 $SU(2)$ 群而不是 $SO(3)$ 群的表示。我们前面分析的点群和空间群都是欧几里得群的子群，其中的旋转操作是 $SO(3)$ 而不是 $SU(2)$ 。 $SU(2)$ 是 $SO(3)$ 的双覆盖，后者相比前者多出来的无非是一个“旋转 360° ”的操作；在 $SO(3)$ 中这个操作就是单位元 e 但是 $SU(2)$ 中它不是单位元；我们将它记作 \bar{e} 。对每个 $SO(3)$ 中的群元 α ， $\bar{\alpha} = \bar{e}\alpha = \alpha\bar{e}$ 都是另一个 $SU(2)$ 群元。因此，晶体中的自旋1/2粒子的对称群实际上是将空间群中的每个 $SO(3)$ 旋转操作 α 扩充为 $SU(2)$ 中对应的两个操作—— α 和 $\bar{\alpha}$ ——而得到的。我们将这样的群称为**双群**，而将前几节中讨论的点群、空间群称为**单群**。单群 G 对应的双群用 G^D 表示。由于旋转 360° 此时不是单位元，我们会有诸如 $c_2c_2 = \bar{e}$ 而非 e 等群乘法关系，但是另一方面， $c_2^4 = e$ 。因此，自旋1/2粒子的对称群的群结构不能通过诸如“将所有 e 换成 \bar{e} ”之类的方法机械地从单群的群结构得到。这也意味着单群 G 不是对应的双群 G^D 的子群，虽然 G^D 含有 G 的全部群元，但是后者与前者贡献的群元彼此之间的乘法关系可能不同。类似的，双群的等价类数目未必是单群的两倍。

实际上，由于凝聚态系统中的自旋1/2粒子基本上都是经过某种修正的电子， \bar{e} 是所谓的费米子parity对称性的生成元；费米子parity指的是系统中的总费米子数目的奇偶性，即 $(-1)^{n_f}$ 。这个量必须守恒，因为二次量子化的基本原则告诉我们，哈密顿量中的每一项都应该有偶数个费米子算符。正常的费米子系统都具有费米子parity对称性，或者说，常规的费米子系统不可能真的一点对称性也没有。当我们向费米子系统施加对称性时，得到的对称群将是施加的对称群和费米子parity放在一起得到的群扩张，我们要从可能的群扩张中找到实际上的物理的对称群。如同存在非点式空间群一样，这个群扩张未必是直积：向费米子系统引入 $SO(3)$ 对称性，得到的完整的对称群是 $SU(2)$ ，引入 $SU(2)$ ，由于 \bar{e} 只有作用在自旋1/2的粒子的波函数上才是 -1 ，而此处自旋1/2的粒子就是某种修正后的电子，我们发现 \bar{e} 起到的就是费米子parity的作用，所以，最终的对称群就是 $SU(2)$ ，无需额外引入独立的费米子parity操作。

由于凝聚态系统中没有自旋统计定理，原则上，自旋1/2粒子也可以是玻色子，此时 \bar{e} 的物理意义就需要做新的阐释。

由于从 G^D 到 G 存在群同态， G 的不可约表示也是 G^D 的不可约表示，而 G^D 的不可约表示甚至未必是 G 的表示（例如， G^D 的一个忠实表示的 $D(\bar{e})$ 肯定不是单位元，但是 \bar{e} 被同态映射到 G 中后却是单位元，因此 D 绝对不可能是 G 的表示）。因此， G^D 的特征标表会比 G 多出来几行。

哈密顿量的形式和旋量波函数携带的表示

自旋1/2的准粒子的波函数是旋量，同时带有自旋和轨道标签。现在我们讨论自旋1/2且具有一定的晶格对称性的准粒子系统的波函数和双群表示之间的关系。

如果系统具有自旋翻转不变性，那么，就哈密顿量而言，哈密顿量中只能够出现 \mathbf{S}^2 ，这是常数，因此实际上哈密顿量只有空间项，即只关于 \mathbf{x} 和 \mathbf{p} 。这种情况下不存在自旋-轨道耦合。因此，可以先求解出准粒子波函数的空间部分，然后手动贴上自旋 \uparrow 和 \downarrow 的标签，就得到了全部能量本征态，并且实际上自旋无论取值为什么能量都不会变化。这样的系统的

自旋1/2准粒子能量本征态形如

$$\psi_{n\mathbf{k}}(\mathbf{r}\sigma) = \varphi_{n\mathbf{k}}(\mathbf{r})\chi_{n\mathbf{k}}(\sigma), \quad (2.109)$$

它们携带的 G^D 的表示显然只可能是 $\varphi_{n\mathbf{k}}(\mathbf{r})$ 携带的 G^D 的表示和 $\chi_{n\mathbf{k}}(\sigma)$ 携带的 G^D 的表示的直积。前者显然是 G 的表示，而后者则是一个二维表示，就是将 $SO(3)$ 群的二维旋量表示限制在 G^D 中有的操作上得到的 $D^{1/2}$ 表示。需要注意，一个 G 的不可约表示和 G^D 的 $D^{1/2}$ 表示直积一般是可约的，但是这是合理的，因为此时自旋自由度和轨道自由度解耦，两者的标签是彼此独立的而没有混合：理论上我们可以分别对自旋部分和轨道部分施加 G^D 中的操作。

自旋翻转对称性的破缺可能是通过向哈密顿量加入 $\mathbf{L} \cdot \mathbf{S}$ 这样的自旋-轨道耦合项，此时时间反演不变性仍然保持，也可能是通过加入磁场，此时时间反演不变性破缺。如果系统不具有自旋翻转不变性但是仍然没有自旋-轨道耦合，此时自旋带来的能级简并解除。不过外加磁场的情况实际上比较复杂。如果只有外加磁场对自旋的影响是值得关注的，那么除了一个简单的能级劈裂以外什么也不会发生。这时候不失一般性地认为 z 方向是磁场方向，则自旋向上和向下的准粒子其实可以分别看成两种无自旋粒子处理。（虽然时间反演变换之后自旋反号，但马上可以附加一个自旋翻转变换把自旋变回来）如果外加磁场对轨道的影响也是值得关注的，那么我们将得到量子霍尔态。对这一问题的分析太过复杂，我们将它留到专门的章节——而且，本节是在讨论“系统内部”的元激发，显然是没有外加磁场的。

然后我们考虑存在自旋-轨道耦合的情况。在存在自旋-轨道耦合时，我们不再能够说“空间群操作只作用在空间部分”，因为根本就没有截然分开的空间部分波函数和自旋部分波函数。此时就可以将一个轨道部分的 G 的表示和一个自旋部分的 $D^{1/2}$ 表示的直积分解到 G^D 的各个不可约表示上，使用这些不可约表示（其中自旋自由度的标签和轨道自由度的标签混在一起）作为系统的能量本征态的标签。需注意存在自旋-轨道耦合系统中的自旋1/2准粒子携带的有可能是空间群的表示，而未必是空间双群特有的表示，因为轨道部分的 G 的表示和一个自旋部分的 $D^{1/2}$ 表示的直积做直和分解之后得到的一些表示可能就是空间群的表示。可以看到，虽然自旋-轨道耦合看起来是千变万化的，自旋角动量和轨道角动量都是角动量这一事实却意味着存在自旋-轨道耦合的能带说白了也是某种“有限旋转群”的表示，从而，并没有特别突破无自旋粒子作为空间群表示的理论。

对含有自旋-轨道耦合的系统，无自旋准粒子的能谱的各种性质仍然是成立的，只不过要将空间群换成空间双群。

2.6 磁群

自旋1/2粒子可以存在磁性序。这可以算一种对称性自发破缺，因为在空间群下等价的原子由于其附近的电子的自旋指向不同，实际上应该认为是不同类型的原子，因此，磁性序形成后，原胞扩张，第一布里渊区缩小。

磁群是每一点上都有一个（通常是二值的）离散变量的晶体的对称群。

2.7 时间反演和附加简并

背景知识 2.3: 复表示

设 $D(G)$ 是群 G 的表示，则 $D^*(G)$ 显然也是——实际上，联系这两者的就是(2.113)。 $D(G)$ 和 $D^*(G)$ 可以不等价，如果等价，那么对每个 $g \in G$ ，我们有

$$D(g)^* = CD(g)C^{-1}.$$

实际上，可以证明，我们有 $C^\top = \pm C$ 。在 $C^\top = C$ 时，我们说 $D(G)$ 是实表示，此时可以将 $D(G)$ 和 $D^*(G)$ 都做共轭变换到一个实数表示上；在 $C^\top = -C$ 时做不到这一点。

背景知识 2.4: 哈密顿量的“built-in”对称性

哈密顿量 H 的时间反演版本是 H^* 。应注意这个对称性在正则量子化框架下是无法使用么正算符表示出来的。应当注意哈密顿量是否是非厄米的和哈密顿量是否具有时间反演不变性是完全无关的。

应注意时间反演算符对不同的粒子是不同的。对无自旋粒子，有

$$T\psi = \psi^*, \quad (2.110)$$

对狄拉克费米子，在质量基下有

$$T\psi = i \begin{pmatrix} \sigma^y & \\ & \sigma^y \end{pmatrix} \psi^*, \quad (2.111)$$

于是对凝聚态物理中的电子，有

$$T\psi = i\sigma^y\psi^*. \quad (2.112)$$

时间反演对称性和空间群（或空间双群——下面的推理没有用到任何空间群的具体信息）对称性配合能够产生额外的结构。时间反演算符不会直接改变 \mathbf{r} 而空间群只会改变 \mathbf{r} ，于是两者对易，即

$$TD(\{\alpha|\mathbf{t}\}) = D(\{\alpha|\mathbf{t}\})T,$$

则

$$D(\{\alpha|\mathbf{t}\})T|\mu\rangle = TD(\{\alpha|\mathbf{t}\})|\mu\rangle = T \sum_{\nu} |\nu\rangle D_{\nu\mu}(\{\alpha|\mathbf{t}\}) = \sum_{\nu} T|\nu\rangle D_{\nu\mu}^*(\{\alpha|\mathbf{t}\}), \quad (2.113)$$

因此 $\{T|\mu\rangle\}$ 同样携带了空间群的一个不可约表示，只不过是 $D^*(\{\alpha|\mathbf{t}\})$ 而不是 $D(\{\alpha|\mathbf{t}\})$ 。因此我们有以下分类：

- 如果空间群的表示 D 和 D^* 不等价，那么 $\{|\mu\rangle\}$ 和 $\{T|\mu\rangle\}$ 是不同的，但是它们都是能量本征态，并且能量本征值全部相同。因此，如果 $\{|\mu\rangle\}$ 是 d 维的，那么实际上有 $2d$ 个简并能级。这称为**附加简并**。
 - 对没有自旋的准粒子，如果 $|\mu\rangle$ 是 $\psi_{n\mathbf{k}}$ ，那么 $T|\mu\rangle$ 是 $\psi_{n,-\mathbf{k}}$ 。这种情况下 $\psi_{n\mathbf{k}}$ 和 $\psi_{n,-\mathbf{k}}$ 不在空间群的同一个表示当中，但是它们的能量是一样的，因为时间反演对称性将 $\psi_{n\mathbf{k}}$ 变换成 $\psi_{n,-\mathbf{k}}$ 。

- 对自旋1/2、无自旋轨道耦合但是没有自旋翻转不变性的准粒子（如磁场中的电子），如果 $|\mu\rangle$ 是 $\psi_{n\mathbf{k}\uparrow}$ ，那么 $T|\mu\rangle$ （差一个 ± 1 的因子）是 $\psi_{n,-\mathbf{k},\downarrow}$ 。
- 如果准粒子有自旋1/2且具有自旋翻转不变性，如果 $|\mu\rangle$ 是 $\psi_{n\mathbf{k}}$ ，那么 $T|\mu\rangle$ 是 $\psi_{n,-\mathbf{k}}$ ，同时还可以任意地加入 \uparrow 和 \downarrow 的标签，总共有 $4d$ 个简并能级。这 $4d$ 的简并数量中，因子 d 来自群 G ，一个因子 2 来自时间反演不变性，一个因子 2 来自自旋翻转不变性。
- 如果空间群的表示 D 和 D^* 等价，且是实表示，则不失一般性地可以对 D 和 D^* 分别做共轭变换，使得 $D = D^*$ 。因此 $T|\mu\rangle$ 和 $|\mu\rangle$ 实际上携带了空间群的同样的表示。尚有待确定的是，它们携带的是不是同一个表示——可以有这样的情况，即虽然 $T|\mu\rangle$ 和 $|\mu\rangle$ 携带的表示是一样的，它们有另一个不同的标签，使得它们处在不同的表示空间中（见后文）。请注意 D 是实表示这件事意味着，时间反演算符和每一个空间群操作都对易，于是时间反演算符在空间群中如同一个标量因子。没有附加简并的情况等价于 T 将 $|\mu\rangle$ 映射回到 $\{|\mu\rangle\}$ 中，即 $T|\mu\rangle$ 和 $\{|\nu\rangle\}$ 线性相关，由于此时 T, D, D^* 都作用在同一个空间上，且 T 是标量算符， $T|\mu\rangle$ 如果要和 $\{\nu\}$ 线性相关，只可能是 $|\mu\rangle$ 和 $T|\mu\rangle$ 都在 μ 列，彼此只相差一个复数因子，即

$$T|\mu\rangle = \lambda|\mu\rangle,$$

从而

$$T^2|\mu\rangle = \lambda^*T|\mu\rangle = |\lambda|^2|\mu\rangle.$$

- 对自旋1/2的准粒子，正如我们证明Kramers定理时看到的那样，

$$T|\mu\rangle = \lambda|\mu\rangle$$

是不可能的，因为 $|\lambda|^2 > 0$ ，但是对自旋1/2的粒子，依照定义这个因子应该是 -1 ，因此虽然 $\{|\mu\rangle\}$ 和 $\{T|\mu\rangle\}$ 携带了空间群的一样的表示，它们必定不在同一个空间中。因此， $\{|\mu\rangle\}$ 和 $\{T|\mu\rangle\}$ 是两个不同的表示空间，存在附加简并，有 $2d$ 个简并。也可以换一个角度，我们将波函数的空间部分当成没有自旋的准粒子求解后将自旋加回去，上下自旋的能量相同，用 T 可以在它们之间切换，因此有 $2d$ 个简并。

- 类似的构造—— D 和 D^* 的表示空间不同，两个空间用 T 来回切换——也可以在没有自旋的准粒子中实现，但是这不是必须的，因为我们允许

$$T|\mu\rangle = |\mu\rangle^* = |\mu\rangle.$$

因此我们说，无自旋粒子在 $D = D^*$ 时不会有附加简并。我们可以考虑一个有三个格点的玩具模型，三个格点上的局域态记作 $|i\rangle, i = 1, 2, 3$ 。设

$$D = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad d = 2.$$

我们可以想象两种情况，一种情况下，能量本征态是

$$\frac{1}{\sqrt{2}}(|1\rangle + |2\rangle), |3\rangle,$$

D 交换这两个态，显然，此时只有两重简并， T 对称性不引入新的简并。另一种情况下， D 作用在

$$\frac{1}{\sqrt{2}}(|1\rangle + i|2\rangle), |3\rangle$$

张成的空间中而 D^* 作用在

$$\frac{1}{\sqrt{2}}(|1\rangle - i|2\rangle), |3\rangle$$

张成的空间中, 此时前一个态存在 T 引入的额外两重简并, 而后一个态没有, 于是总共有三重简并。当然, 也可以想象一个有 $2d = 4$ 重简并的能谱。我们可以看到, 对无自旋粒子, 与时间反演对称性有关的简并是可以通过哈密顿量的具体形式精细调控的, 而不是一定有 $2d$ 重, 也即, 没有时间反演对称性导致的必然的简并, 因此我们说, 此时没有附加简并。

- 如果空间群的表示 D 和 D^* 彼此等价但是不是实表示, 那么有

$$D^*(g) = CD(g)C^{-1}, \quad C^\top = -C,$$

在这里矩阵 C 就是时间反演算符 T 在 $\{|\mu\rangle\}$ 下的矩阵元。如果 T 将 $|\mu\rangle$ 映射到 $\{|\mu\rangle\}$ 中, 即 $\{|\mu\rangle\}$ 同时负载 D 和 D^* , 那么就有

$$T|\mu\rangle = \sum_{\nu} |\nu\rangle C_{\nu\mu},$$

注意此时因为无论怎么做共轭变换, $D \neq D^*$, 我们是不能够保证 $T|\mu\rangle$ 和 $|\mu\rangle$ 在 C 的同一列的。从而

$$T^2|\mu\rangle = \sum_{\nu} C_{\nu\mu}^* T|\nu\rangle = \sum_{\nu, \alpha} C_{\nu\mu}^* C_{\alpha\nu} |\alpha\rangle = \sum_{\nu} |\nu\rangle (CC^*)_{\nu\mu}.$$

我们容易看出

$$D(g) = C^* D(g)^* C^{*-1} = C^* CD(g)C^{-1}C^{*-1},$$

从而

$$D(g)C^*C = C^*CD(g),$$

因此通过Schur引理,

$$C^*C = \lambda I, \quad C^\dagger C = -\lambda I,$$

并且由 $C^\dagger C$ 的形式, 应该有 $-\lambda > 0$ 即 $\lambda < 0$ 。因此我们有

$$T^2|\mu\rangle = \lambda|\mu\rangle, \quad \lambda < 0.$$

- 对无自旋准粒子, 这是不可能的, 因此 $T|\mu\rangle$ 不能将 $|\mu\rangle$ 放回 $\{|\mu\rangle\}$, 即存在附加简并。
- 对自旋1/2的准粒子, 与之前 $D = D^*$ 的情况类似, 与时间反演不变性对应的能级简并是可以通过调整哈密顿量的形式来精细调节的, 因此没有简并重数严格为 $2d$ 这个要求, 于是我们说没有附加简并。需要注意的是这和Kramers定理其实没有矛盾, 因为设 l 是 D 的维数, 由于

$$(\lambda)^l = \det(C^*C) = |\det C|^2 > 0$$

而 λ 是负的, 只能有 l 是偶数, 因此 $\{|\mu\rangle\}$ 是偶数维的, 因此能量简并也是(非零的)偶数个。总之, 自旋1/2的准粒子在 D 和 D^* 不等价时每个能级至少有两重简并, $D = D^*$ 时, 由于附加简并, 每个能级至少两重简并, 在 D 等价于 D^* 然而不是实表示时, 由于 λ 需要是偶数, 每个能级还是至少两重简并。于是我们从另一个角度证明了Kramers定理。

2.8 对理想晶体的偏离

最后我们讨论哪些物理现象是理想晶体之外的。

- 组成晶体的各个原子的位置可以有涨落。这导致声子出现，见第三章。声子谱的对称性分析和电子谱类似，不过声子没有自旋。
- 晶体的表面也会有一些元激发，即表面态。表面态破缺了三维离散平移对称性的一个方向，但是保留了另一些方向。由于晶体表面的晶格相比于晶体内部（称为体态（bulk））可能会有变形，表面会有特殊的组分等，表面物理是非常复杂的。
- 可能平衡时的体态本身就不是理想晶体，即存在缺陷、杂质等。这会导致局域模式的出现。
- 理想晶体的所有性质都是局部的，但是晶体可能有全局的几何性质，通常是各种拓扑性质。

第三章 晶格振动和声子

在本章中我们需要考虑晶格的变形，因此， \mathbf{R}_i 不再是简单的 \mathbf{i} 以初基格矢为基矢量张成的矢量，而会在此基础上有小的偏离。使用 $\mathbf{R}^{(0)}$ 表示其它章节中的 \mathbf{i} 以初基格矢为基矢量张成的 \mathbf{R}_i 。

3.1 举例：谐振子链导出的自由声子

3.1.1 一维简谐振子链

首先考虑一个一维的、均质的晶格，并假定非线性效应很微弱，即不存在声子-声子散射，从而

$$H = \sum_n \frac{P_n^2}{2M} + \frac{M}{2} \sum_n \omega^2 (X_{n+1} - X_n)^2. \quad (3.1)$$

记晶格常数为 a ，做傅里叶变换

$$P_n = \frac{1}{\sqrt{N}} \sum_k e^{ikna} P_k, \quad X_n = \frac{1}{\sqrt{N}} \sum_k e^{ikna} X_k,$$

由于 \mathbf{X} 和 \mathbf{P} 都是实场，将 k 换成 $-k$ 只会导致共轭转置。可以计算出对易关系

$$[X_k, P_{k'}] = i\delta_{k+k',0}, \quad [X_k, X_{k'}] = [P_k, P_{k'}] = 0, \quad (3.2)$$

并将哈密顿量写成

$$H = \frac{1}{2M} \sum_k P_{-k} P_k + \sum_k \frac{M\omega_k}{2} X_{-k} X_k,$$

其中

$$\omega_k^2 = 2(1 - \cos ka)\omega_0^2, \quad \omega_k > 0. \quad (3.3)$$

做变换

$$b_k = \sqrt{\frac{M\omega_k}{2}} X_k + i \frac{1}{\sqrt{2M\omega_k}} P_k, \quad b_k^\dagger = \sqrt{\frac{M\omega_k}{2}} X_k - i \frac{1}{\sqrt{2M\omega_k}} P_k, \quad (3.4)$$

则有

$$H = \sum_k \omega_k \left(b_k^\dagger b_k + \frac{1}{2} \right). \quad (3.5)$$

这样就得到了声子。

另一种更加一般的方法是傅里叶变换。注意此时 ω 表面上看有两个解，一正一负，但是因为 \mathbf{X} 是实场，其正频部分和负频部分不是独立的，即所谓“增加一个频率为 $-\omega$ 的声子”就是“消灭一个频率为 ω 的声子”。因此，只需要取 $\omega > 0$ 的情况，就覆盖了所有的声子模式。用场论的话说，声子由于是实场的量子化产生的，其反粒子就是声子。

背景知识 3.1: 反粒子

设已有粒子产生算符 a^\dagger ，取其复共轭 $b^\dagger = a$ ，则适当为 a 粒子和 b 粒子分别选择合适的基态（如 b 的产生算符会消灭 a 的真空态），会发现 b 粒子的能量正好是 a 粒子的能量的相反数；考虑动量守恒，我们会发现 $a_{\mathbf{k}}^\dagger$ 在适当选择基态之后似乎和 $b_{-\mathbf{k}}$ 具有一样的行为。我们称 b 粒子为 a 粒子的反粒子。在规范场论中一种粒子的荷和它的反粒子也相反。但是反粒子和粒子的质量和自旋是一样的。

一个复场的量子化会产生两种粒子，彼此互为反粒子，但是一个实场产生的粒子的反粒子就是它自身，即在一个理论中交换 a^\dagger 和 a ，并适当地在各个波函数前乘上一些相位因子，所得结果并无差别。注意从(3.4)中可以看出无法在 $a_{\mathbf{k}}^\dagger$ 和某个 $a_{\mathbf{k}'}$ 之间划等号（例如 $\mathbf{k}' = -\mathbf{k}$ 肯定不行），但是交换它们却并不能改变我们的理论。

从自由场的傅里叶展开中可以看到，交换 a 和 a^\dagger ，将 \mathbf{x} 变更正负号，然后同时将 t 也变更正负号，整个展开式一点也没有变。这是CPT对称性的一个平庸的例子。

3.1.2 三维单成分正方晶格

现在考虑一个三维的正方晶格，各个离子实还是一模一样的。用下标标记晶格坐标，用上标标记矢量的分量，得到(3.1)的三维版本（已经略去了一个无关紧要的基态能量 V_0 ）：

$$H = \frac{1}{2M} \sum_i P_i^2 + \frac{1}{2} \sum_{\langle i,j \rangle} \frac{\partial^2 V}{\partial R^a \partial R^b} (X_i^b - X_j^b). \quad (3.6)$$

将 \mathbf{X}_i 和 \mathbf{P}_i 做如下展开：

$$\mathbf{X}_i = \frac{1}{\sqrt{N}} \sum_{\mathbf{k}, \lambda} \lambda \mathbf{X}_{i\lambda} e^{i\mathbf{k} \cdot \mathbf{R}_i^0}, \quad \mathbf{P}_i = \frac{1}{\sqrt{N}} \sum_{\mathbf{k}, \lambda} \lambda \mathbf{P}_{i\lambda} e^{i\mathbf{k} \cdot \mathbf{R}_i^0}, \quad (3.7)$$

其中 λ 是对称矩阵 $\frac{\partial^2 V}{\partial R^a \partial R^b}$ 的特征方向，我们用它表示不同的偏振方向（由于 \mathbf{X} 是矢量，声子会有一个额外的偏振自由度）。在系统对称性比较好的时候，无论 \mathbf{k} 方向如何，都可以找到三个合适的 λ ，一个平行于 \mathbf{k} ，两个垂直于 \mathbf{k} ，即可以区分出明确的纵波模式和横波模式，从而可以对不同的 \mathbf{k} 选取不同的 λ ，记作 $\lambda_{\mathbf{k}}$ 。这样，哈密顿量就是

$$H = \frac{1}{2M} \sum_{\mathbf{k}, \lambda} P_{-\mathbf{k}\lambda} P_{\mathbf{k}\lambda} + \frac{M}{2} \sum_{\mathbf{k}, \lambda} \omega_{\mathbf{k}\lambda}^2 X_{-\mathbf{k}\lambda} X_{\mathbf{k}\lambda}. \quad (3.8)$$

容易看出，由于 H 的厄米性， $\omega_{\mathbf{k}\lambda} = \omega_{-\mathbf{k}\lambda}$ 。做变换

$$b_{\mathbf{k}\lambda} = \sqrt{\frac{M\omega_{\mathbf{k}\lambda}}{2}} X_{\mathbf{k}\lambda} + i \frac{1}{\sqrt{2M\omega_{\mathbf{k}\lambda}}} P_{\mathbf{k}\lambda}, \quad b_{\mathbf{k}\lambda}^\dagger = \sqrt{\frac{M\omega_{\mathbf{k}\lambda}}{2}} X_{-\mathbf{k}\lambda} - i \frac{1}{\sqrt{2M\omega_{\mathbf{k}\lambda}}} P_{-\mathbf{k}\lambda}, \quad (3.9)$$

哈密顿量就成为

$$H = \sum_{\mathbf{k}, \lambda} \omega_{\mathbf{k}\lambda} \left(b_{\mathbf{k}\lambda}^\dagger b_{\mathbf{k}\lambda} + \frac{1}{2} \right), \quad (3.10)$$

而由于

$$\sqrt{2M\omega_{\mathbf{k}\lambda}} X_{\mathbf{k}\lambda} = b_{\mathbf{k}\lambda} + b_{-\mathbf{k}, \lambda}^\dagger,$$

形变量为

$$\mathbf{X}_i = \frac{1}{\sqrt{2MN\omega_{\mathbf{k}\lambda}}} \sum_{\mathbf{k}, \lambda} (\lambda b_{\mathbf{k}\lambda} e^{i\mathbf{k} \cdot \mathbf{R}_i^0} + \text{h.c.}). \quad (3.11)$$

3.2 理想晶体中声子的一般理论

3.2.1 声学声子和光学声子

如果晶格是复式晶格，那么 \mathbf{X} 需要两个量子数来标记，其一是格点坐标，其二是表明它是哪种原子的位移的一个标签。这样，如果晶格有 n_i 种原子，那么它就有 $3n_iN$ 个自由度，因子3来自三个振动方向； N 个原胞在哈密顿量对角化后转化为连续的Bloch波矢，因此做完正则量子化之后，这 $3n_iN$ 个自由度会产生 $3n_i$ 条连续的声子能带或者说声子谱。

如果晶格中只有一种原子，那么显然只有3条谱。另一方面Goldstone定理预言了由于连续平移对称性的破缺，一定会出现三个模式，那么这三个模式就是前述的那三个模式，并且这些模式在动量很小时色散关系近似为线性的。

声子作为一种准粒子，当然也服从第2.4.2节中的说法，其“动量”也是局限在第一布里渊区中的。一个复式的晶格可以看成单式晶格加入扰动之后的结果。我们将扰动视为微扰，从而获得复式晶格的声子谱的一些定性性质。首先，扰动会让晶格的对称性下降，从而实空间中晶格常数增大，从而第一布里渊区缩小，即产生所谓布里渊区折叠¹。于是，类似于图4.1的现象会出现在声子谱上，出现“声子能带”，一些声子模式变得有能隙，当然还有一些声子模式是无能隙的。但是，Goldstone模式——在动量很小时色散关系为线性，没有能隙——还是只有那三个，我们称这些模式为声学声子，将剩下的 $3n-3$ 个模式称为光学声子。显然，在长波极限下——即这个动量的声子足够多时，足以在宏观上产生效应的情况下——声子退化为固体力学振动，即声波，因为晶体中的宏观弹性波当然来自晶格振动，通过弹性力学的论证会发现它们的色散关系是线性的，并且宏观弹性波的频率通常都很低，因此三种弹性波或者说声波确实就是三个声学声子支。由于不同的声子模式彼此解耦（或者至少，相互作用足够弱），三个方向上的声波应该能够对应到单一的声子模式上。容易看出，它们只可能对应到声学声子上，这就是“声学声子”这个名称的由来。另一方面，由于光学声子不对应弹性波，光学声子即使很多，也不会导致固体出现宏观的形变（但这不代表没有和光学声子对应的宏观场——比如说后面所说的电子密度就可以，光学声子的经典极限称为光学波²），但是此时一个晶胞中的原子之间出现相对位移，电子密度改变，从而介质的光学性质会改变；如果晶体是离子晶体，那么甚至会有明显的声子-光子散射。这就是“光学声子”这个名称的由来。

长波声学声子的理论和第3.1.2节是一致的，由于晶格整体移动，在这些模式中可以将晶格中的全体原子看成一个单一的原子，从而，做替换（其中 ρ 是晶体密度）

$$MN \rightarrow \rho V \quad (3.12)$$

就可以。

应当说明，复式晶格可能仍然保留了足够的对称性，使得一些模式具有简并，从而能谱上可能没有 $3n$ 条那么多的分支。

¹我们还会不止一次看到布里渊区折叠现象出现，在这里，布里渊区折叠是来自我们设定的扰动，而在一些低温电子配对情况下布里渊区折叠是自发出现的。

²光学波一词的英文直译optical wave通常指的是电磁波。本文中的“光学波”则是指（介质中的）和光学有关的波，即光学声子。

3.2.2 声子格林函数和路径积分

首先做从哈密顿量到拉格朗日量的勒让德变换:

$$\dot{X}_{q\lambda} = \frac{P_{q\lambda}}{M},$$

于是

$$Z = \int \mathcal{D}Q_{q\lambda} \mathcal{D}Q_{q\lambda}^* \exp \left(- \int d\tau \sum_{q,\lambda} \frac{M}{2} \left| \frac{dQ_{q\lambda}}{d\tau} \right|^2 + \frac{M\omega_{q\lambda}^2}{2} |Q_{q\lambda}|^2 \right). \quad (3.13)$$

在虚频率空间中做计算, 我们有

$$Z = \int \prod_{\omega_n} dQ_{q\lambda}^*(\omega_n) dQ_{q\lambda}(\omega_n) \exp \left(- \sum_{q,\lambda,\omega_n} \frac{M}{2} (\omega_n^2 + \omega_{q\lambda}^2) |Q_{q\lambda}(\omega_n)|^2 \right). \quad (3.14)$$

我们故意省去了积分上限 β 以及它导致的离散的松原频率, 因为这里是在用虚时间技术算零温格林函数。

3.2.3 声子谱的测定

可以通过一些声子和其它粒子的耦合来测定声子谱。通常我们需要的是一个有一条声子线和两条探测粒子线的相互作用顶角, 此时按照能量和动量守恒我们有 (其中 \mathbf{p} 表示探测粒子的动量而 \mathbf{q} 表示声子动量——实际上是晶格动量)

$$\mathbf{p}' - \mathbf{p} = \pm \mathbf{q} + \mathbf{G}, \quad \omega_{\mathbf{p}'}^{\text{detect}} - \omega_{\mathbf{p}}^{\text{detect}} = \pm \omega_{\mathbf{q}}^{\text{phon}}, \quad (3.15)$$

其中 \mathbf{G} 是一个任意的倒格矢。通过设置不同的 \mathbf{p} 并观察不同 \mathbf{p}' 方向上的探测粒子散射概率和出射探测粒子的能量, 即可扫描得到声子谱。显然, 探测粒子的能量和动量应该和声子在一个量级上。

Raman散射是指一种入射光和出射光频率不同的现象。输出光频率低于入射光, 其成因是泵浦光光子的能量一部分被用于激发出介质中的某个模式, 另一部分输出。Raman散射只适合测长波长光学声子。

3.2.4 态密度

声子谱中一条带上的态密度为

$$g(\omega) = \frac{1}{V} \frac{dn}{d\omega} = \frac{1}{(2\pi)^3} \int \frac{dS}{|\nabla_{\mathbf{q}} \omega|}. \quad (3.16)$$

这第二个等号是因为 dn 正比于 ω 和 $\omega = d\omega$ 两个等能面之间的状态数, 所以

$$\begin{aligned} dn &= \frac{V}{(2\pi)^3} \int_{\omega=\omega(\mathbf{q})} dS dq_{\perp} \\ &= \frac{V}{(2\pi)^3} \int_{\omega=\omega(\mathbf{q})} dS \frac{d\omega}{|\nabla_{\mathbf{q}} \omega|}. \end{aligned}$$

总的态密度只需要将各支谱上的态密度加起来即可。

对长波极限下的声学声子——其实就是声波——每一支谱上的态密度为

$$g(\omega) d\omega = \frac{1}{(2\pi)^3} 4\pi q^2 \frac{1}{c} d\omega,$$

即

$$g(\omega) = \frac{q^2}{2\pi^2 c} = \frac{\omega^2}{2\pi^2 c^3}. \quad (3.17)$$

如果实验测得的声子态密度在一些频率出现了很高的峰，即作为一个分布函数看起来有些奇异，这意味着这个频率处群速度为零，这种情况处的 \mathbf{k} 称为van Hove奇点。第一布里渊区的高对称点处的群速度通常不会高，因为如果一个方向上具有某个增长趋势，作用一个点群操作后的方向上就有一模一样的增长趋势，于是这个高对称点就至少是一个鞍点。因此通过观察态密度曲线的尖锐峰或谷的位置就可以估计出高对称点处的声子频率。非晶态物质没有第一布里渊区，也不会有这样的尖锐峰。

3.2.5 热容

暂时忽略声子-声子散射和声子-电子散射，从而声子成为自由粒子，则声子对配分函数的贡献为

$$Z = \sum_{\{n_{\mathbf{k}\sigma}\}} e^{-\beta \sum_{\mathbf{k},\sigma} \omega_{\mathbf{k}\sigma} n_{\mathbf{k}\sigma}}, \quad (3.18)$$

其中 σ 标记不同的声子模式（实际上这并不是自旋，虽然在较简单的晶体中这基本上就是偏振，和光子偏振很像）。给定声子能谱就能够计算出这部分配分函数，从而计算出声子贡献的热容。自由声子近似下我们就有

$$C_V = \sum_{\mathbf{k},\sigma} \left(\frac{\omega_{\mathbf{k}\sigma}}{T} \right)^2 \frac{e^{\omega_{\mathbf{k}\sigma}/T}}{(e^{\omega_{\mathbf{k}\sigma}/T} - 1)^2}. \quad (3.19)$$

高温极限：Dulong-Petit定律

当 $T \rightarrow \infty$ 时对 $\omega_{\mathbf{k}\sigma}/T$ 做小量展开，发现热容正比于晶格的自由度数

$$C_V = \sum_{\mathbf{k},\sigma} 1 = 3N, \quad (3.20)$$

其中 N 为晶格中的原子数目。

实际上，晶格振动有物理效应这件事首先是因为比热容而发现的。在温度不太低也不太高的时候，晶体的比热似乎是 $3Nk_B$ 。晶格有 N 个原子， $3N$ 个移动方向，每个方向上有 X 和 P 两个变量，如果它们没有耦合，即哈密顿量对这些变量是二次型，则有能量均分定理比热是 $k_B/2 \times 3 \times 2$ 。因此，晶体的比热似乎主要就是由晶格振动贡献的，而事实的确如此。

随着温度下降，自由度冻结开始出现，能量均分定理失效，晶体热容相对于经典值明显降低。例如，对有能隙的声子模式， $T \rightarrow \infty$ 时 $\omega_{\mathbf{k}\sigma}/T \rightarrow \infty$ ，从而根据(3.19)我们有

$$C_V = \sum_{\mathbf{k},\sigma} \left(\frac{\omega_{\mathbf{k}\sigma}}{T} \right)^2 e^{-\omega_{\mathbf{k}\sigma}/T}. \quad (3.21)$$

这时必须对声子做完整的量子处理。

爱因斯坦模型

爱因斯坦模型是爱因斯坦提出的一个极度简化的模型，假定声子能谱为平带，频率均为 ω_0 ，这等价于说各个原子之间其实没有任何耦合，无论 \mathbf{k} 怎么变，频率都是 ω_0 。这样根

据(3.19)就有

$$C_V = 3N \left(\frac{\omega_0}{T} \right)^2 \frac{e^{\omega_0/T}}{(e^{\omega_0/T} - 1)^2}. \quad (3.22)$$

这个看起来非常简陋的模型，在经过和实验数据对比之后，只要选择了适当的 ω_0 就能够很好地展现温度降低时热容的降低。爱因斯坦模型在较低温下符合得很好。当温度进一步降低时它衰减得太快了：我们会发现在 $T \rightarrow 0$ 时大体上有

$$C_V \sim \frac{1}{T^2} e^{-\omega_0/T},$$

C_V 关于 T 的曲线一开始几乎是平的，然后才开始缓慢上升；然而，在极低温下的热容与温度之间的关系是可以写成多项式形式的，通常是正比于 T^3 。因此爱因斯坦模型在极低温下不成立，此时需要使用更精确的处理。注意虽然表面上(3.21)在 $T \rightarrow 0$ 时的行为和爱因斯坦模型一样，由于(3.21)中的 $\omega_{\mathbf{k}}$ 可以变化，将所有项加起来之后会出现类似于干涉的现象，使得 C_V 对 T 的依赖不是简单的 $e^{-\omega_0/T}/T^2$ 。³

低温极限：德拜模型

在温度低到爱因斯坦模型失效时，主要的热容贡献来自声学声子：这里的关键点在于(3.21)此时是失效的，因为声学声子无能隙，从而总能够找到一个 \mathbf{k} 使得 $\omega_{\mathbf{k}\sigma}/T$ 并不很大。**德拜模型**假定晶体中只有三个声学声子分支，它们严格对应于晶体中的三个宏观弹性波模式，即一个波速为 c_l 的纵波模式，以及两个波速为 c_t 的横波模式。忽略 \mathbf{k} 增大后声子能谱会偏离直线这件事。当然，这样就有无数多个 (\mathbf{k}, σ) 了，因此我们做一个截断：我们直接忽略频率高于某个**德拜频率** ω_D 的全部声子模式，并且认为德拜频率以下的声子模式只有前述的三个色散关系和宏观弹性波一模一样的模式。这些假设在温度真的非常低的时候的确是适用的，德拜频率基本上就描述了“热涨落能够激发频率多高的声子”。

根据(3.17)，我们有

$$g(\omega) = \frac{\omega^2}{2\pi^2 c_l^3} + 2 \times \frac{\omega^2}{2\pi^2 c_t^3}, \quad (3.23)$$

于是

$$\begin{aligned} C_V &= \int_0^{\omega_D} d\omega V \left(\frac{\omega^2}{2\pi^2 c_l^3} + 2 \times \frac{\omega^2}{2\pi^2 c_t^3} \right) \left(\frac{\omega}{T} \right)^2 \frac{e^{\omega/T}}{(e^{\omega/T} - 1)^2} \\ &= \frac{3V}{2\pi^2 \bar{c}^3} T^3 \int_0^{\omega_D/T} dx \frac{x^4 e^x}{(e^x - 1)^2}, \end{aligned} \quad (3.24)$$

其中

$$\frac{3}{\bar{c}^3} = \frac{1}{c_l^3} + \frac{2}{c_t^3}. \quad (3.25)$$

在 $T \rightarrow 0$ 时可以看到 $C_V \propto T^3$ ，和实验一致，即所谓**德拜 T^3 定律**。

德拜频率以下的声子模式数目大体上和 N 同个量级。如果我们真的相信德拜模型在任何时候都成立，那么根据(3.17)就有

$$3N = V \int_0^{\omega_D} d\omega \left(\frac{\omega^2}{2\pi^2 c_l^3} + 2 \times \frac{\omega^2}{2\pi^2 c_t^3} \right),$$

³例如

$$\sum_n e^{-n\omega_0/T} = \frac{1 - e^{-N\omega_0/T}}{1 - e^{-\omega_0/T}},$$

分母正比于 $1/T$ ，因此求和后整个式子正比于 T 。求和化积分可以更加容易地算出这一点。

从而

$$\omega_D = \bar{c} \left(\frac{6\pi^2 N}{V} \right)^{1/3}, \quad (3.26)$$

上式不依赖于温度，代入(3.24)之后就得到

$$C_V = 9N \left(\frac{T}{\omega_D} \right)^3 \int_0^{\omega_D/T} dx \frac{x^4 e^x}{(e^x - 1)^2}, \quad (3.27)$$

因此，我们可以在不同温度下测量热容，然后通过上式计算出 ω_D ，结果发现不同温度对应的 ω_D 是不同的，因此德拜模型并非在任何温度下都成立。在 $T \rightarrow 0$ 时上面的积分可以被严格计算出来，是

$$C_V = \frac{12\pi^4}{5} N \left(\frac{T}{\omega_D} \right)^3, \quad (3.28)$$

虽然上式建立在错误的假设（真的所有声子模式都遵从弹性波的色散关系）上，但是其形式是正确的（正比于粒子数，德拜 T^3 定律），因此可以作为 ω_D 的实验定义。

3.3 声子-光子耦合

3.3.1 长光学波的黄方程

对一个每个原胞中只含有两个离子，一正一负的离子晶体，声子-光子相互作用可以用一个非常直观的方法描述。我们考虑长波极限，即 $\mathbf{k} \ll 1$ 的情况，此时我们可以同时使用“晶格离子的位移场”的概念和“极化矢量”的概念，后者是一个宏观概念，所以我们要求前者也是宏观概念。光学声子自由度大体上是 $\mathbf{u}_+ - \mathbf{u}_-$ ，我们做一个归一化，记

$$\mathbf{W} = \sqrt{\frac{M}{V_{u.c.}}} (\mathbf{u}_+ - \mathbf{u}_-), \quad (3.29)$$

其中 M 是正负离子的约化质量，则光学声子自由度的动能就是

$$\sum \frac{1}{2} M (\dot{\mathbf{u}}_+ - \dot{\mathbf{u}}_-)^2 = \frac{1}{2} \int d^3\mathbf{r} \dot{\mathbf{W}}^2.$$

很显然光学声子模式会导致一个电偶极矩，具体是多大并不重要，但是总之是正比于 \mathbf{W} ；电场当然也会拉开电偶极矩。于是总之电极化矢量 \mathbf{P} 正比于 \mathbf{W} 和 \mathbf{E} ，设

$$\mathbf{P} = b_{21} \mathbf{W} + b_{22} \mathbf{E}, \quad (3.30)$$

于是电磁场和物质的耦合能量密度为

$$\int_0^{\mathbf{E}} \mathbf{P} \cdot d\mathbf{E} = -\frac{1}{2} b_{22} \mathbf{E} \cdot \mathbf{E} - b_{21} \mathbf{W} \cdot \mathbf{E}.$$

此外通过分析弹性系数矩阵的形式，会发现 $(\mathbf{u}_+ - \mathbf{u}_-)$ 是其本征态，于是光学声子模式有确定的弹性势能，形如

$$\sum \frac{1}{2} k (\mathbf{u}_+ - \mathbf{u}_-)^2 = -\frac{1}{2} \int d^3\mathbf{r} b_{11} \mathbf{W}^2.$$

于是最后，系统拉氏量为

$$\begin{aligned} L &= \sum \frac{1}{2} M (\dot{\mathbf{u}}_+ - \dot{\mathbf{u}}_-)^2 - \int_0^{\mathbf{E}} \mathbf{P} \cdot d\mathbf{E} - \sum \frac{1}{2} k (\mathbf{u}_+ - \mathbf{u}_-)^2 \\ &= \frac{1}{2} \int d^3\mathbf{r} (\dot{\mathbf{W}}^2 + b_{22} \mathbf{E}^2 + 2b_{21} \mathbf{E} \cdot \mathbf{W} + b_{11} \mathbf{W}^2). \end{aligned} \quad (3.31)$$

按照这个拉氏量，就得到 \mathbf{W} 的运动方程。于是我们就得到 \mathbf{P} 的本构方程和 \mathbf{W} 的运动方程联立得到的黄方程：

$$\mathbf{P} = b_{21}\mathbf{W} + b_{22}\mathbf{E}, \quad \ddot{\mathbf{W}} = b_{11}\mathbf{W} + b_{12}\mathbf{E}, \quad b_{21} = b_{12}. \quad (3.32)$$

直接通过“ \mathbf{W} 和 \mathbf{E} 之间的关系是线性的”也可以得到以上方程，但是 $b_{12} = b_{21}$ 不能得到——这个关系来自“系统可以被一个只含有 \mathbf{W} 和 \mathbf{P} 的拉格朗日动力学描述”这一假设。

这个唯象模型(3.32)中的参数和实际可测的参数有明确关系。在电场恒定时我们有

$$\mathbf{P} = \left(b_{22} - \frac{b_{12}^2}{b_{11}} \right) \mathbf{E},$$

又有

$$\mathbf{P} = (\epsilon(0) - 1)\mathbf{E},$$

于是

$$\epsilon(0) - 1 = b_{22} - \frac{b_{12}^2}{b_{11}}. \quad (3.33)$$

如果电场频率远高于晶格振动频率， \mathbf{W} 基本上没有响应，即 $\mathbf{W} = 0$ ，于是就有

$$\mathbf{P} = b_{22}\mathbf{E},$$

于是

$$\epsilon(\infty) - 1 = b_{22}. \quad (3.34)$$

我们现在得到了一个平滑化之后的光学波以及支配它的方程。这些东西相对于光学声子，就好像固体力学相对于声学声子，宏观电动力学相对于光子。

我们考虑一个简单的情况，即 \mathbf{W} 做长波长振荡，没有不均匀的空间分布。实际的 $\sqrt{-b_{11}}$ 一般都是很小的，因为声子频率再高不会比电磁波频率还要高，从而我们可以近似认为由 \mathbf{W} 激发的电磁场是库伦场，这样我们可以将由 \mathbf{W} 激发的电磁场当成库伦场。⁴于是我们联立求解

$$\begin{aligned} \mathbf{P} &= b_{21}\mathbf{W} + b_{22}\mathbf{E}, \quad \ddot{\mathbf{W}} = b_{11}\mathbf{W} + b_{12}\mathbf{E}, \\ \nabla \cdot (\mathbf{E} + \mathbf{P}) &= 0, \quad \nabla \times \mathbf{E} = 0. \end{aligned}$$

对横波模式 \mathbf{W}_T ，其散度为零，而我们注意到

$$\nabla \times \ddot{\mathbf{W}}_T = b_{11} \nabla \times \mathbf{W}_T,$$

因此就有⁵

$$\ddot{\mathbf{W}}_T = b_{11}\mathbf{W}_T, \quad (3.35)$$

即长波长极限下横波模式的频率为

$$\omega_{TO}^2 = -b_{11}, \quad (3.36)$$

有时也记为 ω_0 。另一方面，对纵波模式，旋度为零，且有

$$\begin{aligned} \nabla \cdot \ddot{\mathbf{W}}_L &= b_{11} \nabla \cdot \mathbf{W}_L + b_{12} \nabla \cdot \mathbf{E} \\ &= b_{11} \nabla \cdot \mathbf{W}_L - b_{12} \nabla \cdot \mathbf{P} \\ &= b_{11} \nabla \cdot \mathbf{W}_L - b_{12}^2 \nabla \cdot \mathbf{W} - b_{12}b_{22} \nabla \cdot \mathbf{E}, \end{aligned}$$

⁴ b_{11} 什么的当然也是库伦场导致的，所以这里似乎有重复计数的疑难。不过其实这里并没有什么问题： \mathbf{W} 是经过粗粒化的，计算 b_{11} 其实是要做这么一件事：将介质分割成很多小的区块，将每个区块内部的原子之间的库伦相互作用归入 b_{11} ，而将区块之间的相互作用用电场表达。

⁵这里用到了(3.35)两边取旋度和散度都成立这一条件。无穷大空间中的调和场一定是零，从而(3.35)也成立。后面推纵波模式同理。



图 3.1: 横长光学波与电磁波的耦合

而

$$0 = \nabla \cdot (\mathbf{E} + b_{21}\mathbf{W} + b_{22}\mathbf{E}),$$

于是

$$\nabla \cdot \ddot{\mathbf{W}}_L = \left(b_{11} - \frac{b_{12}^2}{1 + b_{22}} \right) \nabla \cdot \mathbf{W}_L,$$

就得到

$$\ddot{\mathbf{W}}_L = \left(b_{11} - \frac{b_{12}^2}{1 + b_{22}} \right) \mathbf{W}_L, \quad (3.37)$$

于是纵波模式的频率为

$$\omega_{LO}^2 = -b_{11} + \frac{b_{12}^2}{1 + b_{22}}. \quad (3.38)$$

这样我们可以得到一个实验中确实测到了的关系：对长光学波，有

$$\frac{\omega_{LO}}{\omega_{TO}} = \sqrt{\frac{\epsilon(0)}{\epsilon(\infty)}}. \quad (3.39)$$

这称为**LST**关系。

导致长光学波的纵波和横波模式产生频率差异的原因就在于 b_{12} ，即在于 \mathbf{W} 产生 \mathbf{E} 的能力上。纵波模式由于能够激发出库伦电场，会有额外的回复力，从而频率会相对比较高。对非离子晶体， $b_{12} = 0$ ，长光学波的纵波和横波频率一致。

下面我们可以将黄方程和麦克斯韦方程直接联合求解，得到电磁波进入晶体后和长光学波耦合而产生的激发。

总之会求解得到三支谱，包括看起来没有发生什么变化的纵波模式，以及两个横波模式。一些书上会说：“纵波模式没有和电磁波耦合，因为观察黄方程的形式，会发现只有横波模式能够和电磁波耦合，而电磁波是横波”。这个说法是正确的，但是要说明，虽然纵波模式显然不会有通常意义上的、满足横波条件 $\nabla \cdot \mathbf{E} = 0$ 的电磁波伴随它，它仍然和电磁场耦合，这个电磁场就是先前讨论的那个纵波模式激发出来的、产生额外回复力的库伦电场，是宏观的库伦电场，不是已经被归为“回复力”而不放进宏观的 \mathbf{E} 中的那部分库伦电场。

实际上，直接将黄方程和麦克斯韦方程联立求解，由纵波条件能够严格证明 $\nabla \times \mathbf{E} = 0$ ；这其实说明黄方程本身的形式就蕴含了“介质激发出来的电场近似看成静电场”这一条件。介质中的极化如果正比于 \mathbf{E} ，极化就相当于只是修正了一下 ϵ_0 ，从而介质中的能够长距离传播的电磁场模式——即所谓“电磁波”——和真空中没有太大区别，横波条件总是满足的。但在电磁波和黄方程耦合的情况下，多出来的 \mathbf{W} 场可以提供不满足横波条件的 \mathbf{P} ，于是仍然能有纵波和电磁场的耦合。此时 \mathbf{W} 满足纵波条件会让 $\nabla \times \mathbf{E} = 0$ ，因此与 \mathbf{W} 耦合的电磁场基本上是静电场（准确地说是似稳场，并且没有宏观电流），正好给出先前得到的 $\omega = \omega_{LO}$ 的模式。这个模式中的电磁场虽然是宏观的，但是不能“逃逸”出晶体，因为库伦场总是局限在电荷附近的，因此它不能通过很远处产生的、在空间中走过一段距离后传入晶体的（横场的）电磁波激发，这就是我们不称它为电磁波的原因。

求解得到的另外两支谱，一支在 $\mathbf{k} \rightarrow 0$ 时，一支在 $\mathbf{k} \rightarrow \infty$ 时会展现出类似于电磁波的色散关系，但应当注意此时 \mathbf{W} 始终是存在的，类似于远离共振频率的受迫振子；在介质中横光学波一定会和电磁波耦合，电磁波一定会激发出横光学波。表面上，某一支谱在 $\mathbf{q} \rightarrow \infty$ 时存在仅有横光学波而没有电磁波的模式，但是这其实是因为电磁波频率和横光学波固有频率非常接近，从而共振导致的，此时实际上横光学波的耗散已经不能忽略了。能够激发出这个模式的电磁波将会被强烈吸收，以及在界面处被强烈地反射。然而，即使是在 $\mathbf{q} \rightarrow \infty$ 时，这个模式中实际上也同时有光学波（格波）和电磁波。这些模式称为极化激元（polariton），它们是电磁波和（与极化直接相关的横场的）光学声子的混合，属于玻色型激发。

实际上，我们完全可以从黄方程(3.32)中消去 \mathbf{W} 场，而得到一个有频率依赖、方向依赖的介电常数张量。将这个介电张量代入麦克斯韦方程并求解，同样可以得到上面提到的三支谱。所以从最一般的介质中的无源麦克斯韦方程中是可以求解出纵波来的，但那是因为源被藏在了介电常数里面。

3.4 声子-声子散射

由于实际上，晶格中的离子实之间的相互作用并不是线性的，哈密顿量中会有诸如 \mathbf{X} 和 \mathbf{P} 的高阶项，从而会有 b, b^\dagger 的高于二次的项，也就是说，出现了声子-声子相互作用。

不同类型的晶格中，声子-声子相互作用可能有很大的不同。

3.4.1 相互作用和热膨胀

本节分析晶格的等压热膨胀。一般来说，如果晶体表面上有压力，那么晶体内部将会出现形变且形变有梯度以产生应力来和外表面压力抵消。换言之，基态中 \mathbf{X} 将有一个非零经典值，即处在相干态中，声子相当于处在玻色-爱因斯坦凝聚中。然而大气压通常不会让材料有特别大的形变。因此，下面我们在宏观的热力学方程中令 $p = 0$ （并且在求导时忽略下标 p ），而在微观模型中假定基态中晶格没有任何形变，这两个说法等价。

设没有声子时晶体体积为 V_0 。假定晶格内能主要还是由二次型的哈密顿量贡献的（通常都是如此——无论如何，Dulong-Petit定律是成立的），则根据近独立玻色子系统的热力学有

$$F = U + T \sum_i \left(\frac{\omega_i}{2T} + \ln(1 - e^{-\omega_i/T}) \right),$$

而

$$p = - \left(\frac{\partial F}{\partial V} \right)_T = - \left(\frac{\partial U}{\partial V} \right)_T - \sum_i \left(\frac{1}{2} + \frac{\omega_i}{e^{\omega_i/T} - 1} \right) \frac{d\omega_i}{dV}.$$

我们考虑一类常见的情况，即格林爱森常数

$$\gamma = - \frac{d \ln \omega_i}{d \ln V} \quad (3.40)$$

大体上不变的情况（无声子-声子散射的情况显然满足这个条件，且 $\gamma = 0$ ）。此时我们有

$$0 = p = - \left(\frac{\partial U}{\partial V} \right)_T + \gamma \frac{U}{V}. \quad (3.41)$$

这就给出了晶体的状态方程。设没有任何声子激发时体积为 V_0 ，显然

$$\left(\frac{\partial U}{\partial V} \right)_T \Big|_{V_0} = 0,$$

在这个点附近展开(3.41)，有

$$\left(\frac{\partial^2 U}{\partial V^2} \right)_T \Big|_{V_0} \Delta V = \gamma \frac{U}{V},$$

考虑到体膨胀系数的定义

$$\alpha = \frac{\partial}{\partial T} \left(\frac{\Delta V}{V_0} \right)_{V_0}$$

以及体变模量的定义

$$K_0 = V_0 \left(\frac{\partial^2 U}{\partial V^2} \right)_T \Big|_{V_0},$$

得到

$$\alpha = \frac{\gamma}{K_0} \frac{C_V}{V}. \quad (3.42)$$

此即格林爱森关系。

3.4.2 热输运

声子是热输运的载体之一，但不是唯一载体，如在金属中热量主要是电子输运导致的。对声子热输运的一个粗糙的估计如下。假定声子数目不变，考虑声子平均自由程远小于晶体尺寸的情况，从而输运属于扩散，即声子流完全来自晶体内部的声子密度梯度，从而完全来自晶体内部的温度梯度。设 c 为单电子携带的能量，则

$$j_\theta = - \langle vnc \Delta T \rangle = - \left\langle vnc \frac{\partial T}{\partial x} v \tau \right\rangle = - \frac{1}{3} C_V v \lambda \frac{\partial T}{\partial x},$$

即

$$\mathbf{j}_\theta = -\kappa \nabla T, \quad \kappa = \frac{1}{3} C_V v \lambda. \quad (3.43)$$

由于被激发的主要是声学声子，此处的 v 可以近似看成一个常数，就是晶体中的声速。

平均自由程 λ 的大小由很多因素确定。首先，声子-声子散射会限制声子的平均自由程。声子-声子散射有两种，其一是正规过程，即初末动量相同，另一种是Umklapp过程，其中初末动量相差一个倒格矢。正规过程影响了动量的分布，但热流的方向大体上是不变的，而Umklapp过程会导致两个向前的声子合并后反而向后，即会导致热流方向反转，从而是主要的阻碍传热的因素。此外，声子与杂质、缺陷、界面的散射也会影响平均自由程。

声子平均自由程和温度有非常密切的关系。在高温下,或者说 $T > \omega_D$ 时,可以把晶格当成经典统计系统,此时

$$n_{\sigma q} = \frac{1}{e^{\omega_{\sigma q}/T} - 1} \approx \frac{T}{\omega_{\sigma q}}, \quad (3.44)$$

由于 $1/\tau$ 正比于散射几率而散射几率正比于声子数,有

$$\tau \propto \frac{1}{T} \quad \text{when } T > \omega_D. \quad (3.45)$$

在低温下,即 $T \ll \omega_D$ 时,高动量声子的数量指数衰减,于是Umklapp过程的发生几率也随之衰减,有

$$\tau \propto e^{\omega_D/\alpha T}. \quad (3.46)$$

此时的声子平均自由程主要由样品大小和杂质密度确定,即声子在晶体中弹道输运,只有撞上边界或杂质才改变方向,因此 λ 和时间无关。

3.5 电子-声子耦合

电子-声子耦合是因为电子在晶格产生的库伦势中运动。显然我们有

$$V \sim \int d^3r \rho_e(\mathbf{r}) V_{\text{ion}}(\mathbf{r}),$$

我们取最低阶的近似,认为 $V_{\text{ion}}(\mathbf{r})$ 正比于 \mathbf{r} 点的离子实数密度,由于离子实并非一个理想的正电荷,而是受核外电子屏蔽的一个正电荷,离子实产生的电势随着距离上升指数衰减,因此这样假设是合理的。于是

$$V \sim \int d^3r \rho_e(\mathbf{r}) \rho_{\text{ion}}(\mathbf{r}).$$

如果 $\{\mathbf{X}_n\}$ 完全没有偏离理想晶格,即晶格没有发生任何形变,那么上式实际上给出了电子质量的一个修正,可以归入电子质量项。当晶格发生形变,但是形变很小时,根据输运方程我们有

$$\Delta \rho_{\text{ion}} \sim \nabla \cdot \mathbf{X},$$

这里我们将原本定义在格子上的 \mathbf{X} 做了一个延拓,假定它在空间中是连续变化的,这是将离散的、一个一个尖峰的离子实密度连续化的常用方法。于是,扣除了背景项(背景项导致了电子质量的修正)之后,相互作用为

$$V \sim \int d^3r \rho_e(\mathbf{r}) \nabla \cdot \mathbf{X}.$$

我们可以直接把上式量子化并切换到动量空间。我们有

$$\rho_{eq} = \sum_{\mathbf{k}, \sigma} c_{\mathbf{k}+\mathbf{q}, \sigma}^\dagger c_{\mathbf{k} \sigma}, \quad (3.47)$$

并注意到 \mathbf{X} 实际上是 b, b^\dagger 的线性组合,例如对三维单成分正方晶格,我们根据(3.9),得到

$$H_{\text{int}} = \gamma \sum_{\mathbf{q}, \lambda} \frac{i\mathbf{q} \cdot \hat{\lambda}_{\mathbf{q}}}{\sqrt{2M\omega_{\mathbf{q}\lambda}}} (b_{\mathbf{q}\lambda} + b_{-\mathbf{q}\lambda}^\dagger) \sum_{\mathbf{k}, \sigma} c_{\mathbf{k}+\mathbf{q}, \sigma}^\dagger c_{\mathbf{k} \sigma}. \quad (3.48)$$

这里, γ 是由 $\rho_{\text{ion}}(\mathbf{r})$ 和 $V_{\text{ion}}(\mathbf{r})$ 之间的关系确定的 $\rho_e(\mathbf{r})$ 和 $\nabla \cdot \mathbf{X}$ 的耦合常数,因子 $i\mathbf{q} \cdot \hat{\lambda}_{\mathbf{q}}$ 来自对 \mathbf{X} 的散度。于是我们就得到了二电子、一声子的相互作用,在费曼图中体现为一个二电子线一进一出、一声子线的顶角。

电子-声子散射远不止这一种类型。值得注意的电子-声子相互作用包括电子和声学声子的短程散射，极化晶体中电子和光学声子的散射等。电子-声子散射会导致多种物理效应，包括有效电子-电子相互作用（电子交换声子产生有效电子-电子吸引）和电子弛豫（电子与热态的声子碰撞导致能量损失）等。电子-声子散射是电阻的一个重要来源。在温度高于德拜温度时可以认为经典的能量均分定理成立，于是晶格振幅平方正比于温度，而电子-声子散射顶角通常正比于声子场的一次方，于是

$$\frac{1}{\tau} \propto |M|^2 \propto \mathbf{X}^2 \propto T,$$

因此电阻正比于 T 。在低温下会被激发起来的声子模式数目正比于 T^3 ，而。电子在离子晶体中移动时，将让周围的正负离子产生相对位移，形成介质局域极化，即激发出光学声子，由此产生一个电子和光学声子混合的模式，称为**极化子（polaron）**。注意极化子是费米型激发，不是玻色型的极化激发。电子-声子散射也会反过来影响声子自身的行为，如声子谱在一些点会出现强烈的扭折，这是来自这些点上强烈的电子-声子相互作用，称为**Kohn反常**。电子-声子散射导致的物理过程如此之多，我们将不在本节一一列举，而是将具体计算留到后面。

3.6 局域振动

3.7 关于非晶体的振动模式

非晶体没有晶格，也就没有晶格动量；但是非晶体当然也有振动本征模式，而且也有 $3N$ 个。我们仍然可以测定非晶体中的振动态密度，因为无论如何非晶体的振动都有固定的频率。观察实验测定结果（Raman光谱，红外吸收谱等仍然可以大体反映）会发现非晶体的振动态密度通常比较光滑，这在预期之中，因为非晶体没有布里渊区，没有高对称点，自然没有态密度非常集中的地方。

第四章 电子能带

本章分析能带电子，所谓能带电子通常是在两个假设下计算出来的：首先是静态近似，即暂时不考虑晶格振动，只是将晶格当成提供了周期性势场的一个背景；其次是近独立近似，由于很难一上手就处理带有复杂相互作用的电子气，因此我们首先处理近独立电子气，也就是电子之间近似没有相互作用的电子气。近独立假设下单电子哈密顿量为

$$H = \frac{\mathbf{p}^2}{2m_e} + V(\mathbf{r}), \quad (4.1)$$

整团电子气的哈密顿量是关于各个电子的哈密顿量之和。

静态近似需要使用电子-声子相互作用修正。至于近独立近似，虽然在很多系统中相互作用的影响不能忽略，但能带理论仍然提供了一个很好的“自由理论”，可以在此之上讨论相互作用带来的影响。例如，(4.1)中的 $V(\mathbf{r})$ 可以不是物理意义明确（比如由原子核施加的库伦能）的势能，而有可能是相互作用电子气在一定情况下产生的等效势能。实际上，对任何一个相互作用系统，都可以找到势 $V(\mathbf{r})$ 使得其能量近似可以写成(4.1)的形式（见(4.4)），因为总是可以将相互作用用电子自能体现，从而自洽求解一个单电子问题，这个单电子问题中的一些条件是由整个晶体中的所有电子确定的。如果相互作用很强，即使往系统中放入一个电子，势 $V(\mathbf{r})$ 的形式也会发生很大改变。电子之间的库伦相互作用的作用是不确定的，如果它不会产生全新的准粒子，那么它主要提供电子自能修正，如朗道费米液体理论，其中可以保留单电子图像，而将相互作用的影响当成体系总能量的一个修正。总之，在很大一类情况下，经过重整化的电子仍然可以近似看成能带电子，其能谱为一条条分离的能带，并且有时可以近似认为不存在相互作用，有时可以认为仅存在较弱的密度-密度相互作用。但也有可能会得到强关联系统，此时将系统看成近独立电子气并无意义。例如在一维情况下能带结构是不稳定的，此时可以做玻色化而得到一类非常不同的准粒子。

背景知识 4.1: 电子自能和多体系统中的单电子波函数

对电子-电子两点松原格林函数，我们有如下Dyson方程

$$G = G^0 + G^0 \Sigma G, \quad (4.2)$$

其中 $-\Sigma$ 是1PI自能图之和，或者说是正规自能图之和。通过求和我们认为有必要考虑的那些正规自能图，可以计算出了 Σ ，然后即可计算出两点格林函数

$$G(\omega) = \frac{1}{\omega - \Sigma(\omega) - H_0},$$

其中 H_0 表示单粒子哈密顿量，同时包括电子动能和周期势场。这里我们为了和单电子非相对论性量子力学的记号保持一致，使用实时理论，用 ω 而非 $i\omega_n$ ；这样的定义和 $-i\mathcal{T} \langle \psi \psi^\dagger \rangle$ 是一致的。

一个一般的费米子两点格林函数具有如下谱结构：

$$G(\mathbf{r}, \mathbf{r}', E) = \sum_j \varphi_j(\mathbf{r}) \varphi_j^*(\mathbf{r}') \left(\frac{\theta(\epsilon_j - \mu)}{\omega - \epsilon_j + i0^+} + \frac{\theta(\mu - \epsilon_j)}{\omega - \epsilon_j - i0^+} \right), \quad (4.3)$$

对自由系统， $\varphi_j(\mathbf{r})$ 就是哈密顿量对角化之后给出的单电子波函数，而 ϵ_j 是单电子能量，而即使对相互作用系统，把 $\varphi_j(\mathbf{r})$ 当成单电子波函数计算单体算符的期望值是不会出错的。在极点附近，有

$$(\epsilon_j - \Sigma(E) - H_0) |\varphi_j\rangle = 0,$$

于是通过求解

$$\left(-\frac{\nabla^2}{2m} + V_{\text{ion}}(\mathbf{r}) \right) \varphi(\mathbf{r}) + \int d^3\mathbf{r}' \Sigma(\mathbf{r}, \mathbf{r}', \epsilon) \varphi(\mathbf{r}') = \epsilon \varphi(\mathbf{r}) \quad (4.4)$$

能够得到各个 φ_j 和对应的 ϵ_j 。这个改头换面的Dyson方程看起来和普通的薛定谔方程非常类似，不同之处在于 $\Sigma(\mathbf{r}', \mathbf{r}, E)$ 可以是非常复杂的，并且由于它显含 E ，且和诸如系统中的电子总数等量有关，需要自洽求解(4.4)和(4.3)。对单电子近似适用的系统，以上步骤给出了重整化后的单电子能谱。

这种写出电子自能然后自洽求解的方法不仅仅适用于单电子波函数和能谱的求解。我们也可以在选定一个自能之后通过格林函数的时间演化方程求解格林函数，或者将格林函数约化到一些近似的动理学方程上。

4.1 周期势场中的自由电子气模型

4.1.1 自由电子气

现在我们讨论最为简单的近独立电子气，也就是 $V(\mathbf{r})$ 在物体内部为常数（可以看成零）的情况。

能谱和本征态

我们在坐标表象下处理问题。计算单个电子的波函数：

$$-\frac{\nabla^2}{2m_e} \psi(\mathbf{r}) = \epsilon \psi(\mathbf{r}),$$

这种方程的解当然是平面波解的线性组合。一个这样的平面波解形如

$$\psi(\mathbf{r}) \propto e^{i\mathbf{k} \cdot \mathbf{r}}.$$

只能保证这个式子在物体内部成立，因为物体边界处 $V(\mathbf{r})$ 不可能是常数。然后我们归一化这些平面波。电子可以自发地溢出物体，但是这样的概率并不大，所以我们可以简单地认为电子只会出现在物体内部（也即，物体被放置在一个无限深势阱当中）。设物体体积为 V ，以下我们均使用记号(2.17)，就有

$$\int d^3\mathbf{r} |\psi(\mathbf{r})|^2 = 1,$$

于是

$$\psi(\mathbf{r}) = \frac{1}{\sqrt{V}} e^{i\mathbf{k} \cdot \mathbf{r}}, \quad \epsilon = \frac{k^2}{2m_e}.$$

很容易看出这些波函数实际上是动量算符的本征态， \mathbf{k} 实际上就是动量。另一方面，这些波函数定义在坐标空间中，坐标空间中的一切都和自旋算符对易，因此这些波函数也是自旋本征态。于是动量和自旋的一组共同正交本征函数为

$$\psi_{\mathbf{k},\sigma}(\mathbf{r}) = \frac{1}{\sqrt{V}} e^{i\mathbf{k} \cdot \mathbf{r}}, \quad \epsilon_{\mathbf{k},\sigma} = \frac{k^2}{2m_e}. \quad (4.5)$$

\mathbf{k} 能够取什么值取决于边界条件。由于物体通常比较大，具体取什么样的边界条件对物体内部的过程毫无影响。

零温状态

自由电子气的费米面是球状的（即费米球）。对三维系统，动量空间中大小为 $d^3\mathbf{k}$ 的区域内的状态数为

$$dN = 2 \frac{V}{(2\pi)^3} d^3\mathbf{k}, \quad (4.6)$$

因子2是因为电子有两个自旋。积掉无用的动量分布角自由度就得到

$$dN = \frac{V k^2 dk}{\pi^2}. \quad (4.7)$$

在本问题中，使用“相空间仍然由动量和坐标确定，只不过被划分为相格”的假设也可以推导出正确的 dN ：我们有

$$dN = 2 \frac{d^3\mathbf{r} d^3\mathbf{k}}{(2\pi)^3},$$

而积掉无用的坐标自由度和动量角自由度之后得到

$$dN = \frac{V k^2 dk}{\pi^2}.$$

我们根据能谱计算态密度。由于

$$\epsilon = \frac{k^2}{2m},$$

可以推导出

$$dN = \frac{V(2m)^{3/2} \sqrt{\epsilon} d\epsilon}{2\pi^2},$$

即单位能量间隔中的态有

$$D(\epsilon) = \frac{V(2m)^{3/2} \sqrt{\epsilon}}{2\pi^2} \quad (4.8)$$

这么多。这样就可以计算出总粒子数和费米能（即零温化学势）之间的关系：

$$N = \int_{\epsilon=0}^{\epsilon_F} dN = \frac{V(2m)^{3/2}}{3\pi^2} \epsilon_F^{3/2}, \quad (4.9)$$

以及总能量

$$E = \int_{\epsilon=0}^{\epsilon_F} \epsilon dN = \frac{V(2m)^{3/2}}{5\pi^2} \epsilon_F^{5/2} = \frac{3}{5} N \epsilon_F. \quad (4.10)$$

对二维或者一维的自由电子气也可以使用类似的方法得到系数不同的结论。

热力学性质

我们现在计算有限温下自由电子气的一些热力学性质。设 $Q(\epsilon)$ 为 ϵ 下方的所有状态数，即

$$Q(\epsilon) = \int_0^\epsilon d\epsilon' D(\epsilon'), \quad (4.11)$$

于是总电子数为

$$N = \int_0^\infty d\epsilon D(\epsilon) f(\epsilon) = \int_0^\infty d\epsilon Q(\epsilon) \left(-\frac{\partial f}{\partial \epsilon} \right).$$

我们有

$$-\frac{\partial f}{\partial \epsilon} = \frac{\beta}{(e^{\beta(\epsilon-\mu)} + 1)(e^{-\beta(\epsilon-\mu)} + 1)},$$

它在温度不高时是 $\epsilon = \mu$ 附近的一个尖峰，并且相对于 $\epsilon - \mu$ 是偶函数，从而可以当成一个经过某些修正的 δ 函数。这样，我们可以将 N 中的积分下限拓展到 $-\infty$ 而对结果没有太大影响。这样我们有

$$\begin{aligned} N &= \int_{-\infty}^\infty d\epsilon Q(\epsilon) \left(-\frac{\partial f}{\partial \epsilon} \right) \\ &= \int_{-\infty}^\infty d\epsilon \left(Q(\mu) + Q'(\mu)(\epsilon - \mu) + \frac{1}{2}Q''(\mu)(\epsilon - \mu)^2 + \dots \right) \left(-\frac{\partial f}{\partial \epsilon} \right). \end{aligned}$$

这其中，第二项一定是零，因为它关于 $\epsilon - \mu$ 是奇函数。第一项就是

$$-\int_{-\infty}^\infty d\epsilon Q(\mu) \frac{\partial f}{\partial \epsilon} = Q(\mu)(f(-\infty) - f(\infty)) = Q(\mu),$$

而第三项是

$$\begin{aligned} &\frac{1}{2}Q''(\mu) \int_{-\infty}^\infty d\epsilon (\epsilon - \mu)^2 \frac{\beta}{(e^{\beta(\epsilon-\mu)} + 1)(e^{-\beta(\epsilon-\mu)} + 1)} \\ &= \frac{1}{2\beta^2}Q''(\mu) \int_{-\infty}^\infty d\xi \frac{\xi^2}{(e^{-\xi} + 1)(e^\xi + 1)} \\ &= \frac{\pi^2}{6}Q''(\mu)T^2. \end{aligned}$$

因此在 T^2 精度下我们有

$$N = Q(\mu) + \frac{\pi^2}{6}Q''(\mu)T^2. \quad (4.12)$$

我们现在尝试写出 μ 的一个显式表达式。首先 $Q(\mu)$ 和 N 其实是很接近的，可以做展开

$$\begin{aligned} Q(\mu) - N &= Q(\epsilon_F) + Q'(\epsilon_F)(\mu - \epsilon_F) - N \\ &= D(\epsilon_F)(\mu - \epsilon_F), \end{aligned}$$

因此我们有

$$\mu - \epsilon_F \sim \frac{Q''(\mu)}{D(\epsilon_F)}T^2.$$

这又意味着展开式

$$\frac{\pi^2}{6}Q''(\mu)T^2 = \frac{\pi^2}{6}(Q''(\epsilon_F) + Q'''(\epsilon_F)(\epsilon_F - \mu) + \dots)T^2$$

中的第一项就达到了 T^2 的精度，从而如果我们只需要计算到 T^2 精度，只需要求解

$$D(\epsilon_F)(\mu - \epsilon_F) + \frac{\pi^2}{6}Q''(\epsilon_F)T^2 = 0$$

即可，最终计算得到

$$\mu = \epsilon_F \left(1 - \frac{\pi^2}{12} \left(\frac{T}{\epsilon_F} \right)^2 + \cdots \right). \quad (4.13)$$

类似地计算热容，电子贡献的内能为

$$U = \int_0^\infty d\epsilon \epsilon f(\epsilon) D(\epsilon),$$

类似地引入

$$R(\epsilon) = \int_0^\infty d\epsilon \epsilon D(\epsilon),$$

做分部积分，泰勒展开，

$$C_V = \frac{n\pi^2}{2} \frac{T}{T_F}, \quad (4.14)$$

与经典Drude模型非常不一样。直观地看，这是因为只有费米面附近的电子才会被激发。

4.1.2 周期势

在Bloch定理中我们没有看到周期势场具体是怎样让能带产生的，以及能带编号 n 到底标记了哪些自由度。下面我们将自由电子放在周期势场中做微扰计算，来更清楚地看到能带如何形成。本节开始给出自由电子放在周期势场中而形成能带这一事实的具体计算。

背景知识 4.2: 定态微扰论

设 $\{|n\rangle\}$ 是 H_0 的一组能量本征态。非简并微扰论中外加哈密顿量 V 导致的一阶波函数修正为：

$$|n^{(1)}\rangle = \sum_{m \neq n} \frac{|m^{(0)}\rangle \langle m^{(0)}|V|n^{(0)}\rangle}{E_n^{(0)} - E_m^{(0)}},$$

能量修正为

$$E_n^{(1)} = \langle n^{(0)}|V|n^{(0)}\rangle,$$

二阶能量修正为

$$E_n^{(2)} = \sum_{m \neq n} \frac{|\langle n|V|m\rangle|^2}{E_n^{(0)} - E_m^{(0)}}.$$

上述修正发散或接近发散（例如：能量修正比原本的能级差还要大）时，需要使用（近）简并微扰，它倾向于让本来能量相近的两个态分开。

设周期势场为 $V(\mathbf{r})$ ，它导致的一阶能量修正为

$$\epsilon_{\mathbf{k}}^{(1)} = \langle \mathbf{k}|V|\mathbf{k}\rangle = \bar{V}, \quad (4.15)$$

就是周期势的空间平均。因此如下我们将 \bar{V} 归入 $\epsilon_{\mathbf{k}}^{(0)}$ 中，或者也可以归入化学势中，而认为周期势场的空间平均是零。二阶非简并微扰论给出

$$\epsilon_{\mathbf{k}}^{(2)} = \sum_{\mathbf{k} \neq \mathbf{k}'} \frac{|\langle \mathbf{k}'|V|\mathbf{k}\rangle|^2}{\epsilon_{\mathbf{k}}^{(0)} - \epsilon_{\mathbf{k}'}^{(0)}}, \quad (4.16)$$

由于周期性，有（我们其实是在使用(2.18)，但是由于是在展开算符，归一化常数为 $(1/\sqrt{V})^2 = 1/V$ ）

$$\begin{aligned}\langle \mathbf{k}' | V | \mathbf{k} \rangle &= \frac{1}{V} \int d^3\mathbf{r} V(\mathbf{r}) e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}} \\ &= \frac{1}{V_{\text{u.c.}}} \int_{\text{u.c.}} d^3\mathbf{r} V(\mathbf{r}) e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}} \frac{1}{N} \sum_m e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{R}_m} \\ &= \frac{1}{V_{\text{u.c.}}} \sum_n \delta_{\mathbf{k}-\mathbf{k}'+\mathbf{G}_n} \int_{\text{u.c.}} d^3\mathbf{r} V(\mathbf{r}) e^{-i\mathbf{G}_n\cdot\mathbf{r}} =: \sum_n \delta_{\mathbf{k}-\mathbf{k}'+\mathbf{G}_n} V_n,\end{aligned}\quad (4.17)$$

这里我们定义了

$$V_n = \frac{1}{V_{\text{u.c.}}} \int_{\text{u.c.}} d^3\mathbf{r} V(\mathbf{r}) e^{-i\mathbf{G}_n\cdot\mathbf{r}}. \quad (4.18)$$

于是计算得到

$$\epsilon_{\mathbf{k}}^{(2)} = \sum_n \delta_{\mathbf{k}-\mathbf{k}'+\mathbf{G}_n} \frac{|V_n|^2}{\epsilon_{\mathbf{k}}^{(0)} - \epsilon_{\mathbf{k}'}^{(0)}}. \quad (4.19)$$

因此，只要非简并微扰论适用，周期势场做的就是修正一下单电子能级。修正了的波函数为

$$\psi_{\mathbf{k}}^{(1)} = \sum_{\mathbf{G}_n} e^{i(\mathbf{k}+\mathbf{G}_n)\cdot\mathbf{r}} \frac{V_n}{\epsilon_{\mathbf{k}}^{(0)} - \epsilon_{\mathbf{k}+\mathbf{G}_n}^{(0)}}. \quad (4.20)$$

然而，在

$$|\mathbf{k}|^2 = |\mathbf{k} + \mathbf{G}_n|^2 \quad (4.21)$$

时，非简并微扰论发散。此时应当使用简并微扰。由于满足这一条件的 \mathbf{k} 和 $\mathbf{k} + \mathbf{G}_n$ 在三维情况下很多，我们此处不做一般的计算。(4.21)的几何意义是 \mathbf{k} 在起点位于原点的 $-\mathbf{G}_n$ 的垂直平分面上。回顾布里渊区的概念，我们发现此时 \mathbf{k} 一定在某两个布里渊区的交界面上。我们现在将每个电子模式的 \mathbf{k} 适当地加上或者减去一个倒格矢，将动量在第一布里渊区以外的所有模式都移动到第一布里渊区以内，即于将单电子能谱平移任意一个倒格矢，然后将第一布里渊区以外的部分全部抹除掉。此时，一些简并出现在第一布里渊区边界上，一些简并出现在第一布里渊区内部。现在加上周期势，在没有简并的地方，能级只是发生小的修正，在第一布里渊区边界处或是在第一布里渊区内部、对应于高阶布里渊区边界的地方，简并微扰导致交叉点“退开”，自我交叉的单一能带分离成一系列不相交的波浪线，这就得到了一系列不同的能带。此时我们说，能带之间打开了能隙或者说回避交叉，让原本交叉的能带之间出现了间隙。

一维情况下，在加入周期势之后，能带在第一布里渊区边界变平了。高维空间中设 \mathbf{n} 是第一布里渊区的法向量，则 $\mathbf{n} \cdot \nabla_{\mathbf{k}} \epsilon_{\mathbf{n}\mathbf{k}} = 0$ 。为什么？TODO

因此，布里渊区的概念在能带论中是一定会出现的：只有在布里渊区边界上，会出现必须使用简并微扰论、在自由电子情况下简并的能级打开能隙（从而连续的能谱断开）等现象。在远离布里渊区边界的地方，周期势场只会修正能级能量，不会改变能带的连续性。我们通过一个具体例子看到了为什么通常将第一布里渊区作为倒空间原胞。

以上使用的 \mathbf{k} 都对应着某个平面波的真正的波矢（我们称它为广延波矢，和约束在某个倒空间原胞中的晶格波矢相区分）。在周期势场下第一布里渊区边界上简并解除，原本连续的自由电子能谱分离成若干条能带，从而我们可以使用能带编号 n 和通过加减特定倒格矢回到第一布里渊区中的 \mathbf{k} 标记各个电子模式。晶格动量就是把自由电子的广延动量加减特定倒格矢挪到第一布里渊区中，因此也可以称为简约动量。波函数的微扰论（非简并微扰有(4.20)，简并微扰发生在彼此差一个倒格矢的模式之间）显示一个Bloch波函数的简约动

量和构成它的各个自由电子波函数分量的广延动量之间只差一个倒格矢，因此，在相差一个倒格矢的意义上，可以认为电子的简约动量和广延动量是一回事。同时，无论是非简并微扰的微扰后波函数(4.20)还是简并微扰的微扰后波函数都满足

$$\psi(\mathbf{r} + \mathbf{R}_i) = e^{i\mathbf{k} \cdot \mathbf{R}_i} \psi(\mathbf{r}),$$

因此简约波矢 \mathbf{k} 实际上就是Bloch定理中的那个 \mathbf{k} ，即晶格动量。

4.1.3 赝势

前述的周期势场 $V(\mathbf{r})$ 的来源没有明确指出。如果我们将它取为原子核提供的库伦势场，能量最低的那些能带对应的电子状态通常都是非常局域化的原子内层电子，它们的行为和没有形成晶体时的原子的内层电子应该不会有太大区别。这些电子和原子核就形成了所谓离子实。电子是费米子，因此外层电子不会占据这些轨道，这就产生了一个等效的排斥力。实际上，这个等效的排斥反而能够简化计算，因为这意味着外层电子感受不到原子核的势场最为奇异的那部分，从而外层电子波函数在接近原子核的区域的高频振荡可以忽略，从而使用几个平面波就能够足够精确地描写它，微扰论总是适用，等等。这个等效的排斥力是泡利不相容原理的一个例子——即使不考虑电子-电子库伦排斥，内层电子的存在本身就能够导致一个等效排斥力。

本节给出一种比较系统的考虑内层电子轨道对外层电子的影响，但是又不需要真的计算内层电子轨道的作用的方法。离子实中的核心电子轨道和价电子轨道同时都是单电子哈密顿量的本征态：

$$H |\psi_c\rangle = E_c |\psi_c\rangle, \quad H |\psi_v\rangle = E_v |\psi_v\rangle, \quad (4.22)$$

这里 c 标记了各个内层电子轨道。定义

$$|\psi_v^{\text{PS}}\rangle = |\psi_v\rangle + \sum_c |\psi_c\rangle \langle \psi_c | \psi_v^{\text{PS}} \rangle, \quad (4.23)$$

则计算得到

$$(H - E_v) |\psi_v^{\text{PS}}\rangle = \sum_c (E_c - E_v) |\psi_c\rangle \langle \psi_c | \psi_v^{\text{PS}} \rangle,$$

从而

$$\left(H - \sum_c (E_c - E_v) |\psi_c\rangle \langle \psi_c| \right) |\psi_v^{\text{PS}}\rangle = E_v |\psi_v^{\text{PS}}\rangle. \quad (4.24)$$

这意味着可以定义一个等效的“势”

$$V^{\text{PS}} = V - \sum_c (E_c - E_v) |\psi_c\rangle \langle \psi_c|, \quad (4.25)$$

从这个势出发做计算能够得到正确的能谱，而价电子波函数则和实际情况不同。在 \mathbf{r} 比较大的地方， $\psi_c(\mathbf{r})$ 很小，从而 V^{PS} 和 V 没有差别，在相应的位置上 $\psi_v^{\text{PS}}(\mathbf{r})$ 和 $\psi_v(\mathbf{r})$ 也没有区别。

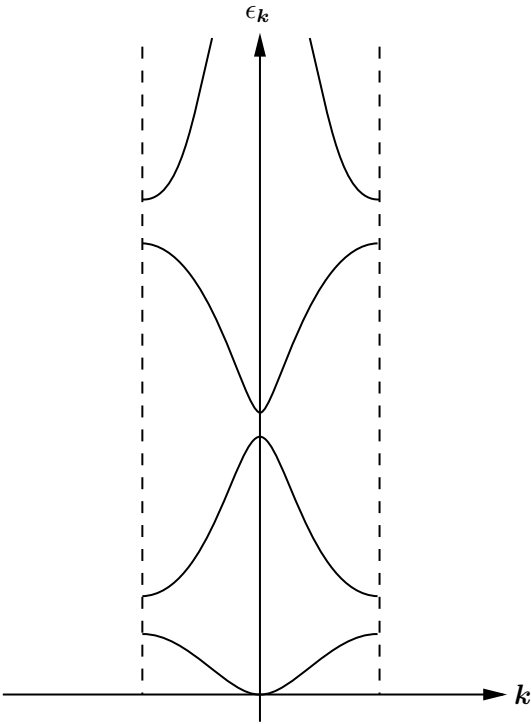
到这里我们都没有做任何近似。内层电子的行为的复杂性体现在 V^{PS} 显含 E_v 这件事上。然而， E_v 完全由 $|\psi_v\rangle$ 的标签确定，单单依靠对称性分析我们就能够知道这些标签是什么：主量子数、角量子数、磁量子数和自旋量子数。因此，(4.24)左边完全可以写成一个不显含 E_v 的形式，因此看起来似乎不像本征值问题的(4.24)实际上真的就是一个本征值问题

$$\left(\frac{\mathbf{p}^2}{2m} + V^{\text{PS}} \right) |\psi_v^{\text{PS}}\rangle = E_v |\psi_v^{\text{PS}}\rangle. \quad (4.26)$$



(a) 自由空间中的单电子能谱

(b) 由于晶格的周期性，出现第一布里渊区折叠，能谱在第一布里渊区外的部分移动回到第一布里渊区中，在简并微扰中彼此会有影响的态体现为第一布里渊区边界上的简并



(c) 相互作用打开能隙，形成分离的能带

图 4.1: 能带结构

然而，和通常的定态薛定谔方程不同，以上方程中 V^{PS} 是依赖于 $|\psi_v^{\text{PS}}\rangle$ 的诸如晶格动量（在计算能带时）或是 l （在求解单原子问题时）的标签的。实际上这也意味着 V^{PS} 在坐标表象下的矩阵元一般来说是 $V(\mathbf{r}, \mathbf{r}')$ ，而不是局域的 $V(\mathbf{r})$ 。

更一般的，可以先计算得到电子能级 E^{AE} 和波函数 $\psi^{\text{AE}}(\mathbf{r})$ ，然后将单电子轨道波函数 $\psi^{\text{AE}}(\mathbf{r})$ 对球谐函数基底展开，做截断，即只取前几个球谐波，得到 $\psi^{\text{PS}}(\mathbf{r})$ ，然后拟合出一个 V^{PS} 使得

$$-\frac{\nabla^2}{2m}\psi^{\text{PS}}(\mathbf{r}) + V^{\text{PS}}(\mathbf{r})\psi^{\text{PS}}(\mathbf{r}) = E^{\text{AE}}\psi^{\text{PS}}(\mathbf{r}), \quad (4.27)$$

就得到了一个赝势。在远离原子核的地方 l 较低的球谐函数占据主导地位，因此在那里 ψ^{PS} 和 ψ^{AE} 是一样的，从而 V^{PS} 和 V^{AE} 是一样的。在接近原子核的地方由于 ψ^{PS} 含有更少的球谐波，有更少的振荡，相应的， V^{PS} 相较于实际的 V^{AE} 要平滑得多。

由于远离原子核时势场快速下降，在晶体中，非常接近一个原子核处的波函数的形式对另一个原子的势场导致的能级修正而言是不重要的。因此可以预期，只要赝势和实际的势场不相符合的区域相比于原子间距充分小，就不会出现内层电子行为由于其它原子的存在发生很大改变的情况，用从单原子全电子计算中获得的赝势做晶体的能带计算就能够得到正确的结果；这实际上就是“只有价电子才是重要的”这一假设。

4.1.4 $\mathbf{k} \cdot \mathbf{p}$ 模型和有效质量

设 \mathbf{k}_0 是某一条能带 $\epsilon_{n\mathbf{k}}$ 的极值点。我们总是可以在它附近做泰勒展开，以有效质量张量（显然它是对称的实张量）

$$\left(\frac{1}{m^*}\right)_{\alpha\beta} = \frac{\partial^2 \epsilon_{n\mathbf{k}}}{\partial k_\alpha \partial k_\beta} \quad (4.28)$$

的主轴为坐标轴，有

$$\epsilon_{n\mathbf{k}} = \epsilon_{n\mathbf{k}_0} + \frac{(k_x - k_{0x})^2}{2m_x^*} + \frac{(k_y - k_{0y})^2}{2m_y^*} + \frac{(k_z - k_{0z})^2}{2m_z^*}, \quad (4.29)$$

我们将在第5.2.1节中看到有效质量的应用。将Bloch波函数的形式(2.92)代入单电子薛定谔方程

$$\left(-\frac{\nabla^2}{2m} + V_{\text{ion}}(\mathbf{r})\right) \psi_{n\mathbf{k}} = \epsilon_{n\mathbf{k}} \psi_{n\mathbf{k}},$$

就得到

$$\left(-\frac{\nabla^2}{2m} + V_{\text{ion}}(\mathbf{r})\right) u_{n\mathbf{k}} - \frac{i\mathbf{k} \cdot \nabla}{m} u_{n\mathbf{k}} = \left(\epsilon_{n\mathbf{k}} - \frac{\mathbf{k}^2}{2m}\right) u_{n\mathbf{k}}, \quad (4.30)$$

其边界条件为

$$u(\mathbf{r}) = u(\mathbf{r} + \mathbf{R}), \quad (4.31)$$

其中 \mathbf{R} 是任意一个正格矢。注意到同一个 \mathbf{k} 不同的 n 的电子模式是彼此正交的，从而

$$\int d^3\mathbf{r} \psi_{n0}^*(\mathbf{r}) \psi_{m0} = \delta_{mn},$$

即

$$\int d^3\mathbf{r} u_{n0}^*(\mathbf{r}) u_{m0} = \delta_{mn}.$$

因此 $\{u_{n0}\}$ 构成了边界条件(4.31)下的一组正交归一化基底。虽然 $u_{n\mathbf{k}}$ 不是真正的电子波函数，(4.30)意味着，实际上可以通过微扰论来从 u_{n0} 出发计算 \mathbf{k} 不是很大的电子模式，只要 $\mathbf{k} = 0$ 是能带的一个极值点。例如由于 $\mathbf{k} = 0$ 是能量极值点，能量无一阶微扰，我们有

$$\epsilon_{n\mathbf{k}} - \frac{\mathbf{k}^2}{2m} = \sum_{n'} \frac{\langle n0 | \mathbf{k} \cdot \mathbf{p} / m | n'0 \rangle \langle n'0 | \mathbf{k} \cdot \mathbf{p} / m | n0 \rangle}{\epsilon_{n0} - \epsilon_{n'0}},$$

从而

$$\epsilon_{n\mathbf{k}} = \frac{\mathbf{k}^2}{2m} + \frac{1}{m^2} \sum_{n'} \frac{\langle n0|p_i|n'0\rangle \langle n'0|p_j|n0\rangle}{\epsilon_{n0} - \epsilon_{n'0}} k^i k^j, \quad (4.32)$$

将 i, j 取到有效质量张量的主轴上, 就有

$$\frac{1}{m_i^*} = \frac{1}{m} + \frac{2}{m^2} \sum_{n', i} \frac{1}{m^2} \sum_{n'} \frac{\langle n0|p_i|n'0\rangle \langle n'0|p_i|n0\rangle}{\epsilon_{n0} - \epsilon_{n'0}}. \quad (4.33)$$

$u_{n\mathbf{k}}$ 也可以类似地计算。

更一般的, 对任意的一个能带极值点 \mathbf{k}_0 , 我们定义

$$\chi_{n\mathbf{k}}(\mathbf{r}) = e^{i(\mathbf{k}-\mathbf{k}_0)\cdot\mathbf{r}} \psi_{n\mathbf{k}_0} = e^{i\mathbf{k}\cdot\mathbf{r}} u_{n\mathbf{k}_0}(\mathbf{r}), \quad (4.34)$$

我们来验证正交性和完备性均成立。首先

$$\begin{aligned} \int d^3\mathbf{r} \chi_{n\mathbf{k}}^*(\mathbf{r}) \chi_{m\mathbf{k}'}(\mathbf{r}) &= \sum_{\mathbf{i}} \int_{\text{u.c.}} d^3\mathbf{r} e^{i(\mathbf{k}'-\mathbf{k})\cdot(\mathbf{r}+\mathbf{R}_i)} u_{n\mathbf{k}_0}^*(\mathbf{r}) u_{m\mathbf{k}_0}(\mathbf{r}) \\ &= N \delta_{\mathbf{k}\mathbf{k}'} \int_{\text{u.c.}} d^3\mathbf{r} e^{i(\mathbf{k}'-\mathbf{k})\cdot\mathbf{r}} u_{n\mathbf{k}_0}^*(\mathbf{r}) u_{m\mathbf{k}_0}(\mathbf{r}) \\ &= N \delta_{\mathbf{k}\mathbf{k}'} \int_{\text{u.c.}} d^3\mathbf{r} u_{n\mathbf{k}_0}^*(\mathbf{r}) u_{m\mathbf{k}_0}(\mathbf{r}) \\ &= \delta_{\mathbf{k}\mathbf{k}'} \delta_{mn}, \end{aligned}$$

即正交性成立。另一方面通过数自由度也可以发现完备性应该成立。因此, $\chi_{n\mathbf{k}}$ 构成晶体中电子的一组正交完备基底。现在设

$$\psi_{n\mathbf{k}}(\mathbf{r}) = e^{i(\mathbf{k}-\mathbf{k}_0)\cdot\mathbf{r}} \varphi_{n\mathbf{k}}(\mathbf{r}), \quad (4.35)$$

则薛定谔方程会转化为

$$\left(-\frac{\nabla^2}{2m} + V_{\text{ion}}(\mathbf{r}) \right) \varphi_{n\mathbf{k}} - \frac{i(\mathbf{k} - \mathbf{k}_0) \cdot \nabla}{m} \varphi_{n\mathbf{k}} = \left(\epsilon_{n\mathbf{k}} - \frac{(\mathbf{k} - \mathbf{k}_0)^2}{2m} \right) \varphi_{n\mathbf{k}}, \quad (4.36)$$

边界条件为

$$\varphi_{n\mathbf{k}}(\mathbf{r} + \mathbf{R}) = e^{i\mathbf{k}_0\cdot\mathbf{r}} \varphi_{n\mathbf{k}}(\mathbf{r}). \quad (4.37)$$

因此, 可以将 $\varphi_{n\mathbf{k}}$ 用 $\{\chi_{n\mathbf{k}_0}\}_n$ 做基底展开, 然后后面微扰论的步骤是类似的。

以上做的微扰都是非简并微扰, 有简并微扰时对 u_{n0} 或是 $\varphi_{n\mathbf{k}}$ 做简并微扰即可。

以上方法就称为 **$\mathbf{k} \cdot \mathbf{p}$ 模型**, 这个名字的来源是显然的。它可以在已经求出了某一晶格动量点的全套Bloch波函数之后用于计算相邻点的Bloch波函数, 也可以用于获得有效质量张量, 这有助于在不计算特别多的动量点的情况下就对体系的能带结构获得一个比较精确的估计。能带越宽, 有效质量越轻, 能带越窄, 有效质量越重。**重费米子系统**指的是有效质量非常大的系统。这样的系统常常是强关联系统。

4.2 Bloch波函数和Wannier波函数

本节给出周期性势场中的近独立电子气的一些一般性质。我们将从对称性出发得到Bloch电子, 并且展示一种“部分定域”的表象: Wannier波函数。

4.2.1 Bloch电子

Bloch定理和Bloch波函数

第2.4.1节中的Bloch定理当然适用于近独立电子气模型，因此这种模型中的电子的波函数自然服从(2.92)。我们在处于这种状态的电子称为Bloch电子。能带电子总是可以认为是自由电子加入了足够多的微扰之后得到的，而我们在第4.1.2节最后提到过，(2.92)中的那个晶格动量 \mathbf{k} 和组成Bloch电子的诸多平面波分量的波矢（即自由电子的真正的动量，或者说延展动量）最多差一个倒格矢。因此，晶格动量为 \mathbf{k} 的Bloch电子可以认为是携带了真实的动量，这动量是 \mathbf{k} ， $\mathbf{k} + \mathbf{G}_i$ ， $\mathbf{k} + \mathbf{G}_j$ 等一系列延展动量的混合，而自由电子做了足够多的微扰之后得到的波函数其实就是这种系统的Bloch波函数。

我们现在来看 n 代表什么。Bloch电子显然满足一个周期性边界条件，因此它是某个等效单电子哈密顿量的周期性解。将(2.92)代入

$$-\frac{\nabla^2}{2m_e}\psi(\mathbf{r}) + V(\mathbf{r})\psi(\mathbf{r}) = \epsilon\psi(\mathbf{r}),$$

得到

$$\left(-\frac{(\nabla + i\mathbf{k})^2}{2m_e} + V(\mathbf{r})\right)u_{\mathbf{k}}(\mathbf{r}) = \epsilon u_{\mathbf{k}}(\mathbf{r}). \quad (4.38)$$

在 \mathbf{k} 已知的情况下求解该本征值问题，可以得到一组 u 以及对应的 E 。由于先前要求 $u(\mathbf{r})$ 是周期函数，该本征值问题必定给出离散谱，也即，我们会获得一组 $(\psi_{n\mathbf{k}}, \epsilon_{n\mathbf{k}})$ ，使得

$$\left(-\frac{(\nabla + i\mathbf{k})^2}{2m_e} + V(\mathbf{r})\right)u_{n\mathbf{k}}(\mathbf{r}) = \epsilon_{n\mathbf{k}}u_{n\mathbf{k}}(\mathbf{r}), \quad (4.39)$$

n 是另一个（离散的）好量子数，我们现在看到它是单个晶胞中的离散谱的能级标记。

通过数自由度我们还可以看出 n 的另一个意义： \mathbf{k} 对应着晶格，从而 n 对应着一个晶胞中不同的原子以及每个原子周围的电子轨道。当然，原则上原子周围有无数个电子轨道，从而 n 可以任意的大，但是实际上我们通常会在某一个能量尺度上截断，物理上，这是因为特别高的空轨道对低能过程没有什么影响。一旦 \mathbf{k} 和 n 给定，Bloch波函数及其时间演化就完全求解出来了，从而 \mathbf{k} 和 n 是坐标空间中的Bloch波函数的全部好量子数。我们称取不同 n 值的电子处于不同的能带上， n 为能带标记。¹能带中的电子在空间上非常不定域，它们实际上是一系列格点上的定域态叠加而成的结果。也可以这么理解这一点：写出一个格点上的哈密顿量之后为了计算能带能量肯定要做对角化，对角化得到的本征向量肯定是把一系列格点上的定域态都线性组合起来了。总之，能带中的Bloch电子常常称为巡游电子。

由于晶体近似在一个无限深势阱中，波函数在晶体边界处快速衰减为零。这就意味着 \mathbf{k} 实际上是离散的。然而，由于晶体的尺度通常远大于原子的尺度， \mathbf{k} 近似可认为是连续的。于是可以写出函数

$$\epsilon = \epsilon_n(\mathbf{k}),$$

由于 \mathbf{k} 加上任何一个倒格矢之后给出同样的Bloch波函数， E 相对于 \mathbf{k} 应该具有周期性，那么它必定是有界的。这就是“能带”这个名称的来源：穷举一个倒空间原胞内部的所有 \mathbf{k} ，得到的所有能量组成一条有限宽的条带。

两条不同的能带之间的间隙提供了一个自然的能量截断，因此在我们已经知道了系统的能带之后，如果需要一个低能有效理论，可以只考虑能量较低的能带，而将粒子跃迁到能量较高的能带再跃迁回来作为微扰，进行能量修正，即可以很自然地将高能能带积掉。

¹注意到，能带的导出实际上并未用到太多晶体的性质（空间周期性等），因此非晶体很多时候也有能带。

最后我们指定波函数的归一化方式。我们这里做的内积操作是第2.1.5节中的第一种情况，即对全空间积分。可以使用约定(2.17)。此时由

$$\int d^3\mathbf{r} \psi_{n\mathbf{k}}^*(\mathbf{r}) \psi_{m\mathbf{k}'}(\mathbf{r}) = \delta_{mn} \delta_{\mathbf{k}\mathbf{k}'} \quad (4.40)$$

可以得出

$$\begin{aligned} \delta_{mn} \delta_{\mathbf{k}\mathbf{k}'} &= \sum_i \int_{\text{u.c.}} d^3\mathbf{r} e^{-i(\mathbf{R}_i+\mathbf{r})\cdot\mathbf{k}} u_{n\mathbf{k}}^*(\mathbf{r}) e^{i(\mathbf{R}_i+\mathbf{r})\cdot\mathbf{k}'} u_{m\mathbf{k}'}(\mathbf{r}) \\ &= N \delta_{\mathbf{k}\mathbf{k}'} \int_{\text{u.c.}} d^3\mathbf{r} e^{-i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}} u_{n\mathbf{k}}^*(\mathbf{r}) u_{m\mathbf{k}'}(\mathbf{r}) \\ &= N \delta_{\mathbf{k}\mathbf{k}'} \int_{\text{u.c.}} d^3\mathbf{r} u_{n\mathbf{k}}^*(\mathbf{r}) u_{m\mathbf{k}'}(\mathbf{r}), \end{aligned}$$

于是

$$\int_{\text{u.c.}} d^3\mathbf{r} u_{n\mathbf{k}}^*(\mathbf{r}) u_{m\mathbf{k}}(\mathbf{r}) = \frac{1}{N},$$

或者考虑 $u_{n\mathbf{k}}$ 的周期性，有

$$\int d^3\mathbf{r} u_{n\mathbf{k}}^*(\mathbf{r}) u_{m\mathbf{k}}(\mathbf{r}) = \delta_{mn}. \quad (4.41)$$

如果使用(2.20)，则

$$\frac{1}{V} \int d^3\mathbf{r} \psi_{n\mathbf{k}}^*(\mathbf{r}) \psi_{m\mathbf{k}'}(\mathbf{r}) = \delta_{mn} \delta_{\mathbf{k}-\mathbf{k}'}, \quad (4.42)$$

从而让简单的平面波 $\exp(i\mathbf{k} \cdot \mathbf{r})$ 不需要乘上归一化因子就能够成为归一化本征态。设 $V_{\text{u.c.}}$ 是单个晶胞的大小，则

$$V = NV_{\text{u.c.}},$$

从而可以得到

$$\frac{1}{V_{\text{u.c.}}} \int_{\text{u.c.}} d^3\mathbf{r} u_{n\mathbf{k}}^*(\mathbf{r}) u_{m\mathbf{k}}(\mathbf{r}) = \delta_{mn}. \quad (4.43)$$

只需求解出一组满足以上条件的 $\{u_{n\mathbf{k}}\}$ ，就得到了一组正交归一化波函数 $\{\psi_{n\mathbf{k}}\}$ 。归一化系数使用 $V_{\text{u.c.}}$ 是非常合理的，因为如(2.25)所示这正是具有正格子的周期性的函数通常使用的归一化系数，这里 $u_{n\mathbf{k}}$ 的地位和 $e^{i\mathbf{k} \cdot \mathbf{r}}$ 相同。

(4.42)意味着Bloch波函数是正交归一化波函数且对应的积分测度为

$$\frac{1}{\sqrt{V}} \int d^3\mathbf{r},$$

记 $c_{n\mathbf{k}}^\dagger$ 为位于能带 n 、晶格动量为 \mathbf{k} 的Bloch电子的产生算符，那么²

$$c_{n\mathbf{k}}^\dagger = \frac{1}{\sqrt{V}} \int d^3\mathbf{r} \psi_{n\mathbf{k}}(\mathbf{r}) \psi^\dagger(\mathbf{r}), \quad \psi^\dagger(\mathbf{r}) = \frac{1}{\sqrt{V}} \sum_{\mathbf{k}} \psi_{n\mathbf{k}}(\mathbf{r})^* c_{n\mathbf{k}}. \quad (4.44)$$

这和(2.18)非常类似，不同之处在于此时的积分变换核中多了一个周期性的 $u_{n\mathbf{k}}$ ，并且除了 \mathbf{k} 以外还有离散的标签 n 。哈密顿量在这一组基下是对角化的，于是写出二次量子化哈密顿量（已经加入化学势项）

$$H = \sum_{n,\mathbf{k}} (\epsilon_{n\mathbf{k}} - \mu) c_{n\mathbf{k}}^\dagger c_{n\mathbf{k}}. \quad (4.45)$$

²其中的 $1/\sqrt{V}$ 的因子是因为二次量子化场算符通常使用全空间的积分为内积的定义，在此定义下，归一化的波函数是 $\psi_{n\mathbf{k}}/\sqrt{V}$ 而不是 $\psi_{n\mathbf{k}}$ 。换句话说，在归一化波函数 $\psi_{n\mathbf{k}}$ 时我们使用了约定(2.22)，而归一化场算符时我们使用了约定(2.18)。

\mathbf{k} 的取值局限在第一布里渊区内部，作为对比，不考虑周期势的边长为 L 的正方体势阱中的电子的 \mathbf{k} 可以取遍所有位于那个边长为 $2\pi/L$ 的格点。但实际上两者的自由度是一样的，因为我们还有 n 标记各个能带，也即，我们相当于把所有能带中的动量都移动到了第一布里渊区内部。

以上讨论并没有说明我们应该将 \mathbf{k} 约束在哪个倒格子原胞中。随着晶胞数 N 增大， \mathbf{k} 在原胞中的取值会变得非常密集，从而 \mathbf{k} 可以认为是连续变化的。但是，这并不能保证 $\epsilon_{n\mathbf{k}}$ 也是连续变化的。我们希望 n 固定后 $\epsilon_{n\mathbf{k}}$ 随着 \mathbf{k} 连续变化，从而可以将能量的不连续变化完全归入 n 跳变中，让能带真的成为连续的“带”，其数目就是 n 可能的取值的数目。前面的对称性分析告诉我们应该将 \mathbf{k} 约束在第一布里渊区内，而在做近自由电子在周期势场下的微扰时我们非常直接地看到了原因。

基态

背景知识 4.3: 近独立费米气体

使用巨正则系综（当然，我们认为系统能够达到统计平衡，就意味着电子之间不可能真的完全没有相互作用，否则能量无法传递），对很大的近独立费米子系统，处在能量本征态 $|n\rangle$ 上的粒子数的平均值为（以下使用 ϵ 表示单个电子的能量而使用 E 表示系统总能量）

$$\langle n_n \rangle = f(\epsilon_n) = \frac{1}{e^{\beta(\epsilon_n - \mu)} + 1}. \quad (4.46)$$

我们让能量尽可能低，那就是要让 $T \rightarrow 0$ ，也就是让 $\beta \rightarrow \infty$ ，此时就有

$$\langle n_n \rangle = \begin{cases} 1, & \epsilon_i \leq \mu, \\ 0, & \epsilon_i > \mu. \end{cases} \quad (4.47)$$

这意味着， $T = 0$ 时电子占据的所有状态就是

$$\epsilon_i = \mu \quad (4.48)$$

以内的所有能量本征态。

化学势是多少需要根据

$$\mu_i = \frac{\partial U}{\partial n_i} \quad (4.49)$$

计算。当然，化学势和粒子数、温度等因素都有关系。在 $T = 0$ 且电子数 N 给定时，常用的做法是显式地写出所有能量本征态，从小到大排列 N 个电子，从而计算出费米能量，然后我们就知道了 $T = 0$ 时的化学势。能量正好为化学势的所有能量本征态共同组成的能级称为**费米能级**，其能量称为**费米能量**。能量在费米能级以上的态无占据，能量在费米能级以下的态有占据。

不同粒子数对应的费米能量是不同的；并且，在分析有限温问题时，化学势不再是费米能量。温度不很高、粒子数很大时，不同粒子数对应的费米能量相差不大，并且化学势和费米能量（也就是 $T = 0$ 时的化学势）相差不大，因此有时会使用费米能量近似作为化学势。

由于能量本征态可以使用晶格动量标记，Bloch电子组成的近独立电子气的基态中所有电子有确定的动量空间，即动量空间中有确定的粒子占据。由于能量和晶格动量之间的关



图 4.2: 化学势和电子填充, 描黑的态有电子填充

系是连续的, 动量空间中被电子占据的态形成连续的几何体, 即存在费米海或者说费米球, 其表面称为费米面, 其上的态即为费米能级态, 与这些态对应的动量称为费米动量。基态的表达式就是一个乘积态, 为

$$|\Psi\rangle = \prod_{|\mathbf{k}| < k_F} c_{\mathbf{k}}^\dagger |0\rangle. \quad (4.50)$$

关于本节的论述要着重指出一点: 虽然我们采用了统计物理的论证来表明必然存在着一个费米面, 从而有对应的费米能量, 但统计物理的论证仅仅为我们提供了系统基态的性质, 而无论系统是不是需要使用平衡态系综描述, 它一定有一个基态。因此, 费米面、费米能级等概念在任何情况下——无论是平衡态还是非平衡态、纯态还是混合态——全部是适用的。这些概念并不依赖统计物理的框架。

大部分材料的费米能都远高于室温, 因此在正常的温度下, 费米分布函数 f 可以就认为是阶跃函数。然而, 对诸如电子比热之类的强烈依赖费米面性质的物理量 (因为只有费米面附近的电子会被激发), 费米分布函数一点对阶跃函数的偏离就会产生很大的影响。

化学势的大小由系统中的电子数决定, 或者也可以反过来说, 化学势的高度决定了系统基态中哪些位置被电子填充。这个对应关系形象地展示如图4.2, 化学势越高, 被描黑的态——也就是基态有电子填充的态——就越多。再次强调, 这个“电子能量形式固定, 改变化学势往系统中填充电子”的物理图像不适用于强关联系统, 因为在那里等效的电子能量形式(4.1)随着电子填入会发生剧烈变化; 相应的, 从这个物理图像衍生出来的理论, 如能带理论也不再适用。

没有什么保证费米面一定是第一布里渊区中的一个封闭曲面。它可以和第一布里渊区边界相交, 从而和另一侧的边界上的自己相连。甚至可以有不连通的费米面, 这出现在三维空间中, 由于适当的能带形状, 费米面附近的能带均没有完全被填充, 此时可以出现多个彼此不连通的费米面。那些同时为多个费米面包裹的动量空间区域中有动量相同, 能带编号不同的多个被占据的电子模式。所谓“费米面包裹的动量空间区域”的说法本身也需要评注, 因为完全可以是费米面以外的那些动量模式被填充; 这仍然可以视为“包裹”, 因为第一布里渊区两侧实际上是连通的。

空穴

虽然多电子态是将场算符作用在真空态上得到的, 但由于本文讨论的电子都由薛定谔场描述, 在任何一个物理过程中电子数都是守恒的。换言之, 设有一个 N 电子的系统,

这个系统实际可能具有的波函数只是态空间的一小部分，即保持电子数为 N 的部分。这个 N 电子空间 \mathcal{H}_N 当然可以使用一次量子化理论来描述，但能否使用二次量子化的语言描述？费米子的特性让这一点成为可能。请注意由于泡利不相容原理，

$$c_{\mathbf{k}}^\dagger |\Psi\rangle = 0, \quad \epsilon_{\mathbf{k}} < \epsilon_F,$$

而由产生湮灭算符的性质显然有

$$c_{\mathbf{k}} |\Psi\rangle = 0, \quad k > k_F,$$

因此如果定义³

$$b_{\mathbf{k}}^\dagger = \begin{cases} c_{-\mathbf{k}}, & k < k_F, \\ c_{\mathbf{k}}^\dagger, & k > k_F, \end{cases} \quad (4.51)$$

那么基态 $|\Psi\rangle$ 实际上是 b^\dagger 产生的准粒子的真空态。 b^\dagger 产生的是什么？当 $k > k_F$ 时它产生的就是费米面以上的电子，而 $k < k_F$ 时它产生的是费米海之内的空穴。空穴的动量就是它占据的态如果有电子的话，这个电子的动量的相反数。这样，我们可以将空穴看成电子的反粒子，虽然这种“反粒子”并不像高能物理中那样，来自洛伦兹群的表示。使用 b^\dagger 写出的哈密顿量在省略一个无关紧要的常数之后为

$$H = \sum_{k > k_F} \xi_{\mathbf{k}} b_{\mathbf{k}}^\dagger b_{\mathbf{k}} - \sum_{k < k_F} \xi_{\mathbf{k}} b_{\mathbf{k}}^\dagger b_{\mathbf{k}}, \quad (4.52)$$

因此一个空穴的能量为 $-\xi_{\mathbf{k}}$ 。以上的哈密顿量不是正定的，但这不会导致负能量疑难，因为费米海虽然很大，但大小有限，因此不会出现能量无限下降的问题。

使用 b^\dagger 的结果是，保持电子数不变的相互作用需要被看成粒子数可变的，例如一个费米海中的电子被激发到费米海之上就意味着产生了一个电子-空穴对。

值得注意的是，使用 c^\dagger 和 b^\dagger 时的“单粒子能谱”是不同的： c^\dagger 表象下低于费米面的那部分模式的能谱在 b^\dagger 表象下被加了一个负号，翻到了费米面上方。在 c^\dagger 表象下，低于费米面的那部分模式是被占据的能带，基态里面已经有一些粒子占据了，将低于费米面的某个模式中的电子移除会让能量上升；而在 b^\dagger 表象下，没有什么已经占据了的低于费米面的模式，基态是“空”的，但是在费米面上方有两种能带，一种是电子能带，一种是空穴能带，在这两种能带中填充粒子都会让能量上升。这两种描述是等价的，因为“在空穴能带中填充一个空穴”和“在已被占据的能带中填充一个空穴”显然是一回事。

这里的构造不仅适用于能带电子，实际上适用于固体中任何看起来比较像电子的准粒子，对这些准粒子我们都可以定义“空穴”和“空穴能带”。

表面上看这里有一个矛盾之处：我们从费米海中移除一个电子，按理说系统能量应该降低，因为系统能量是所有电子的能量之和；然而按照(4.52)，从费米海中移除一个电子，即在费米面以下的电子能带中产生了一个空穴，即在费米面以上的空穴能带中填充了一个空穴，系统能量应该升高。这个佯谬是由于两种说法中的能量是不同的：前一种说法中的能量是指 $\epsilon_{\mathbf{k}}$ ，而后一种说法中的能量是指 $\xi_{\mathbf{k}}$ 。从费米海中移除一个电子确实让晶体内部的电子的能量降低了，但是由于系统和环境的接触和化学势的存在，环境中的电子会获得向系统中跃迁的势能，因此晶体内部电子的能量加上与晶体接触的那部分电子的能量反而升高了。我们讨论系统能量时通常是把两部分能量放在一起讨论的，因为这样系统才能够看成是一个孤立的系统。

³费米面上的态相对来说非常少，因此忽略。 b 的记号只是临时性的，因为它和声子的记号冲突。

我们会看到，如果以 $\xi_{\mathbf{k}}$ 为能量衡量基准，那么空穴实际上是电子的反粒子（见背景知识3.1）。历史上，正电子的存在曾经就用这样的方式解释过：在高能物理中，假定真空态中实际上已经有一个电子组成的费米海，那么从中拿一个电子扔掉，就得到一个电荷为正、动量相反的“粒子”，这当然就是正电子，即电子的反粒子。后来的高能物理学家弄清楚了场的量子化必然导致粒子和反粒子，因此无需构造一个不自然的充满电子的费米海，就能够接受正电子的存在。凝聚态物理中的情况正好相反：费米海是真的存在的，我们考虑它附近的态，发现基本自由度是电子激发和空穴激发，如果以费米海为基态，以它为能量零点，那么空穴就是电子的反粒子（而反之，以“真正的”真空态为基态，那么正电子才是电子的反粒子）。然而应当注意，此处定义的“粒子-反粒子变换”一来和高能物理中不同，二来其实并无太多用处，因为费米子的性质决定了，如前所述，已填充能带上的低能激发就是空穴，则我们实际上并无必要讨论已填充能带上的电子；反之，也没有必要讨论未填充能带上的空穴。反粒子出现时（费米面以下），粒子就不出现，粒子出现时（费米面以上），反粒子就不出现，那么实际上并无引入反粒子这一概念的必要。

4.2.2 Wannier波函数

定义

前面看到，我们使用了两个标签来标记一个电子模式：一个是“晶格动量” \mathbf{k} ，一个是能带编号。前者的可能取值的数量有多少？晶格动量加上一个倒格矢之后与之前等价，因此晶格动量的取值数目为

$$\frac{2\pi/V}{2\pi/V_{\text{u.c.}}} = N,$$

N 为晶胞数目。当然这就是第一布里渊区中的动量数目。一个晶格内可以有多种原子，但是这个信息已经被能带编号考虑在内了，因为电子“看到”原子种类的方式就是原子施加的势场，而周期势导致能带出现。

任何一种非Bloch波函数的电子波函数基底也应该有同样的数目，例如，它们可以用一个“坐标”和一个类似于能带编号的离散标签标记。在正交归一化条件成立时这种电子状态称为**Wannier波函数**。我们使用 $\mathbf{i}, \mathbf{j}, \dots$ 表示电子位置， m, n, \dots 用于区分定域在 \mathbf{i} 号晶胞附近的电子的电子云的“形状”。这是一个比较模糊的说法，例如定域在某个格点附近的电子可以非常定域，也可以比较定域，有两种模式。在没有自旋-轨道耦合时，自旋取向也可以计入 m, n, \dots 中，此时自旋起到的唯一作用就是让系统中的电子模式的数目是轨道模式的数目的两倍。而有自旋-轨道耦合时自旋取向肯定会会计入 m, n, \dots 中，且此时波函数的形式不再是 $\psi(\mathbf{r})$ 而是 $\psi(\mathbf{r}\sigma)$ 了。下面为了书写方便我们还是将波函数记作 $\psi(\mathbf{r})$ ，或者说把波函数表示成由 $\psi(\mathbf{r}, \uparrow)$ 和 $\psi(\mathbf{r}, \downarrow)$ 排成的列向量。 \mathbf{R}_i 表示 \mathbf{i} 处的格点的位矢，则定域在 \mathbf{i} 号晶胞附近的Wannier电子的波函数可以用 $w_{n\mathbf{i}}(\mathbf{r})$ 表示，可以写成如下形式：

$$w_{n\mathbf{i}}(\mathbf{r}) = w_n(\mathbf{r} - \mathbf{R}_i), \quad \int d^3\mathbf{r} w_n(\mathbf{r})^* w_m(\mathbf{r}) \propto \delta_{mn}. \quad (4.53)$$

此外，由于Wannier波函数的定域性，有

$$\int d^3\mathbf{r} w_m(\mathbf{r} - \mathbf{R}_i)^* w_m(\mathbf{r} - \mathbf{R}_j) \propto \delta_{ij}, \quad (4.54)$$

因为Wannier波函数足够局域，因此不同的格点附近的Wannier波函数可认为没有交叠。(4.53)和(4.54)实际上构成了Wannier波函数这一概念的**定义**，符合它们的 $\{w_{n\mathbf{i}}(\mathbf{r})\}$ 即可称

为Wannier波函数。通常会要求

$$\frac{1}{V_{\text{u.c.}}} \int d^3\mathbf{r} |w_{n\mathbf{i}}(\mathbf{r})|^2 = 1, \quad (4.55)$$

这样归一化常数是一个局域的晶胞体积，符合我们的要求。这样的约定满足(2.31)，其中Wannier波函数的地位和约束在晶胞中的 $e^{i\mathbf{k}\cdot\mathbf{r}}$ 一样。在归一化Bloch波函数时我们使用了积分测度

$$\frac{1}{V} \int d^3\mathbf{r},$$

从而导致归一化条件中出现了关于整块晶体的参数 V 。对Bloch波函数这是合理的，因为它的好量子数是动量，因此是非常不定域的波函数，但Wannier波函数使用格点坐标标记，因此是非常定域的波包，因此我们希望Wannier波函数的归一化因子应该是一个常数而不包含任何关于系统大小的信息。Wannier波函数既然是一组正交归一化基底，就可以定义对应的二次量子化算符，即

$$c_{n\mathbf{i}}^\dagger = \frac{1}{\sqrt{V_{\text{u.c.}}}} \int d^3\mathbf{r} w_{n\mathbf{i}}^*(\mathbf{r}) \psi^\dagger(\mathbf{r}). \quad (4.56)$$

实际上“位置”和“形状”这两个词都不是非常良定义的，比如说 \mathbf{i}, \mathbf{j} 等可以仅仅标记晶胞位置，也可以标记原子位置，后者相比前者，需要区分一个晶胞中的不同类型的原子。“形状”一词的含义也不是非常确定，它可以和能带编号一致也可以就是单原子周围轨道的编号，或者介于两者之间；此外如果 \mathbf{i} 仅仅标记了晶胞位置，那么电子在一个晶胞中的哪一个原子附近这件事也需要被当成“形状”信息。

在 \mathbf{i} 仅仅区分不同晶胞时， \mathbf{i} 的取值个数也是 N ，因此我们可以猜测， \mathbf{i} 与Bloch电子的晶格动量标记了同样的自由度，而 n 与能带编号标记了同样的自由度。我们后面在分析Wannier表象下的哈密顿量的对角化时会看到这是正确的。

Wannier表象下的哈密顿量

之前我们用(4.45)给出了Bloch基下的能带理论哈密顿量。现在来看同样的理论在Wannier基下会呈现什么形式。设能带理论下单电子哈密顿量为 \hat{h} ，则可以做表象变换

$$H = \sum_{m\mathbf{i}, n\mathbf{j}} c_{m\mathbf{i}}^\dagger \langle m\mathbf{i} | \hat{h} | n\mathbf{j} \rangle c_{n\mathbf{j}},$$

其中我们令

$$-t_{m\mathbf{i}, n\mathbf{j}} = \langle m\mathbf{i} | \hat{h} | n\mathbf{j} \rangle = \frac{1}{V_{\text{u.c.}}} \int d^3\mathbf{r} w_m^*(\mathbf{r} - \mathbf{R}_i + \mathbf{R}_j) \hat{h} w_n(\mathbf{r}), \quad (4.57)$$

则

$$H = - \sum_{m\mathbf{i}, n\mathbf{j}} t_{m\mathbf{i}, n\mathbf{j}} c_{m\mathbf{i}}^\dagger c_{n\mathbf{j}}.$$

记号 $\{m\mathbf{i}, n\mathbf{j}\}$ 表示对一组顺序不限的 $\{m\mathbf{i}, n\mathbf{j}\}$ 组合只求和一次，就得到

$$H = - \sum_{\{m\mathbf{i}, n\mathbf{j}\}} (t_{m\mathbf{i}, n\mathbf{j}} c_{m\mathbf{i}}^\dagger c_{n\mathbf{j}} + \text{h.c.}). \quad (4.58)$$

我们得到了一个非对角化的哈密顿量，这当然是正确的，因为系统显然应该在动量基——即Bloch基——下对角化。但(4.58)提供了非常清晰的物理图像：电子会从一个格点跃迁到另一个格点； $t_{m\mathbf{i}, n\mathbf{j}}$ 为所谓跃迁系数。

现在我们有两种表示能带的方法，一种是Bloch基底下，用 \mathbf{k} 做标签的表示法，一种是Wannier基底下用(4.58)表示的。(4.58)中的跃迁系数 $t_{mi,nj}$ 具体是什么实际上是没有一种唯一的诠释方式的，因为(4.58)中的 i, j 等只需要取值范围正好是晶格中全部格点，并且满足空间平移不变性就可以。实际上，在有电磁场或是杂质时，空间平移不变性也可以不满足。具体两种表述之间的关系我们将在下一节分析。

4.3 Wannier波函数的获得和跃迁系数的计算

本节讨论如何选定Wannier波函数，以及选定了一种Wannier波函数之后要如何计算(4.58)中的跃迁系数。

4.3.1 Bloch波函数诱导出的Wannier波函数

对Bloch波函数做傅里叶变换

现在我们考虑一种特殊的Wannier波函数定义，是直接从系统的正确的Bloch波函数出发做傅里叶变换得到的：

$$w_{ni}(\mathbf{r}) \propto \sum_{\mathbf{k}} \psi_{n\mathbf{k}}(\mathbf{r}) e^{-i\mathbf{k} \cdot \mathbf{R}_i},$$

这里， n 直接对应于能带编号（如果是复式晶格，那么一个晶胞中的不同原子周围的电子会被重新组合而成为不同能带，体现在能带标签中），而 \mathbf{R}_i 对应晶格格点。上式前面的归一化系数尚未选定，但容易计算出

$$\frac{1}{V} \int d^3\mathbf{r} \frac{1}{N} \left(\sum_{\mathbf{k}} \psi_{n\mathbf{k}}^*(\mathbf{r}) e^{i\mathbf{k} \cdot \mathbf{R}_i} \right) \left(\sum_{\mathbf{k}} \psi_{m\mathbf{k}}(\mathbf{r}) e^{-i\mathbf{k} \cdot \mathbf{R}_j} \right) = \delta_{mn} \delta_{ij},$$

取

$$w_{ni}(\mathbf{r}) = \frac{1}{\sqrt{N}} \sum_{\mathbf{k}} \psi_{n\mathbf{k}}(\mathbf{r}) e^{-i\mathbf{k} \cdot \mathbf{R}_i}, \quad \psi_{n\mathbf{k}} = \frac{1}{\sqrt{N}} \sum_{\mathbf{R}_i} e^{i\mathbf{k} \cdot \mathbf{R}_i} w_{ni}(\mathbf{r}), \quad (4.59)$$

这样归一化条件就是

$$\frac{1}{V_{\text{u.c.}}} \int d^3\mathbf{r} |w_{ni}(\mathbf{r})|^2 = 1, \quad (4.60)$$

符合我们的要求。

现在来分析Wannier波函数在实空间中的具体形状。考虑到 $u_{n\mathbf{k}}$ 的周期性，我们有

$$u_{n\mathbf{k}}(\mathbf{r}) = u_{n\mathbf{k}}(\mathbf{r} - \mathbf{R}_i),$$

于是

$$w_{ni}(\mathbf{r}) = \frac{1}{\sqrt{N}} \sum_{\mathbf{k}} u_{n\mathbf{k}}(\mathbf{r} - \mathbf{R}_i) e^{i\mathbf{k} \cdot (\mathbf{r} - \mathbf{R}_i)}, \quad (4.61)$$

这表明Wannier波函数实际上是 $\mathbf{r} - \mathbf{R}_i$ 的函数，可以写成

$$w_{ni}(\mathbf{r}) = w_n(\mathbf{r} - \mathbf{R}_i), \quad (4.62)$$

遵从条件(4.53)。在 \mathbf{r} 远离 \mathbf{R}_i 时，指数因子快速振荡，导致整个求和基本上为零，因此Wannier函数只在 \mathbf{R}_i 附加有较明显的值，因此它确实定域在格点 \mathbf{R}_i 附近。

无能带间跃迁的哈密顿量

Bloch波函数下的哈密顿量为(4.45)，不同能带之间没有跃迁。(4.59)没有混合不同能带，因此(4.59)中的不同 n 之间也没有跃迁。于是在 $t_{mi,nj}$ 中我们取 $m = n$ ，此时有

$$H = - \sum_{n, \{i,j\}} t_{n,ij} c_{ni}^\dagger c_{nj} + \text{h.c.}, \quad (4.63)$$

显然我们对格点坐标 i, j 做了傅里叶变换之后得到的就是以 n 为能带编号的能带，即 n 是能带编号，不同能带之间没有跃迁。

既然不同能带之间没有跃迁，我们暂时略去能带标签 n 。(4.63)是最为一般的，原则上可以描述一切晶体中的电子能带。例如，原则上，自由电子气可以通过引入非常远的跃迁矩阵元来实现。如果晶格的吸引作用是如此之强，以至于电子是高度定域的，这样只有最近邻的格点之间才会跃迁，即

$$\sum_{i,j,\sigma} c_{i\sigma}^\dagger t_{ij} c_{j\sigma} = \sum_{\langle i,j \rangle, \sigma} t_{ij} c_{i\sigma}^\dagger c_{j\sigma},$$

从而

$$H = - \sum_{\langle i,j \rangle} t_{ij} c_i^\dagger c_j + \text{h.c.} \quad (4.64)$$

这称为紧束缚模型。在没有耦合磁场或者杂质等时，平移不变性意味着

$$H = - \sum_{\langle i,j \rangle} t c_i^\dagger c_j + \text{h.c.} \quad (4.65)$$

我们当然还可以引入次近邻跃迁等来修正紧束缚模型。在空间平移对称性成立，但是紧束缚近似未必成立时，我们用 t_l 表示从格点 \mathbf{R}_i 到 $\mathbf{R}_i + \mathbf{R}_l$ 的跃迁系数，此时

$$H = - \sum_{i,l} t_l c_{i+l}^\dagger c_i + \text{h.c.}, \quad (4.66)$$

厄米性要求

$$t_l = t_{-l}^*. \quad (4.67)$$

请注意 $i + l$ 是正格矢 $\mathbf{R}_i + \mathbf{R}_l$ 的编号。

从坐标表象到晶格动量表象的变换是(4.44)，而(4.63)相当于是(4.44)的一个中间结果：(4.44)中从 \mathbf{r} 表象到 \mathbf{k} 表象的积分可以写成

$$\frac{1}{\sqrt{V}} \int d^3\mathbf{r} \psi_{n\mathbf{k}}(\mathbf{r})^* = \frac{1}{\sqrt{N}} \sum_{\mathbf{i}} e^{-i\mathbf{k} \cdot \mathbf{R}_i} \frac{1}{\sqrt{V_{\text{u.c.}}}} \int_{\text{u.c.}} d^3\mathbf{r} u_{n\mathbf{k}}(\mathbf{r})^*,$$

从坐标表象切换到Wannier波函数相当于首先完成了在晶胞内的积分，然后从Wannier波函数的表象（即(4.63)）计算能带就是完成了剩下的对 \mathbf{i} 的求和，即(2.29)。从坐标表象切换到 \mathbf{i} 表象相当于对哈密顿量做了一点对角化但没有完成，从 \mathbf{i} 表象切换到 \mathbf{k} 表象（这是一个傅里叶变换，我们将这个过程称为计算Bloch和）则完成了对角化。这意味着给定一个形如(4.58)的哈密顿量，对 c_{mi}^\dagger 做傅里叶变换(2.29)，(4.58)对角化，这个过程产生的傅里叶变换中的 \mathbf{k} 实际上就是直接从周期势计算Bloch波函数而产生的那个简约波矢 \mathbf{k} 。⁴对角化结果

⁴这个结论并不是平凡的，因为存在这样的可能：对同一个能量本征态，用傅里叶变换得到的 \mathbf{k} 表象的那个 \mathbf{k} 和直接从自由电子受到周期势场扰动出发计算得到的 \mathbf{k} 不一样。例如，它们是否可能差一个正负号？但仔细检查我们此处的计算过程，这种可能性是不存在的。

在解析计算时基本上我们写下来的每个式子都非常“连续”，从而不会有“ \mathbf{k} 的值被缩放、顺序被打乱”这种情况，但是在做数值计算时这些是需要考虑到。一个新手在编写对角化近独立电子哈密顿量的程序时就有可能用错误的 \mathbf{k} 标记各个本征态，从而产生莫名其妙的错误。

是

$$H = \sum_{\mathbf{k}} \epsilon_{\mathbf{k}} c_{\mathbf{k}}^{\dagger} c_{\mathbf{k}}, \quad \epsilon_{\mathbf{k}} = - \sum_{\mathbf{R}_l} t_l e^{-i\mathbf{k} \cdot \mathbf{R}_l}. \quad (4.68)$$

显然，这意味着，如果我们已经有了一个单电子能谱，可以通过傅里叶变换

$$t_l = -\frac{1}{N} \sum_{\mathbf{k}} e^{i\mathbf{k} \cdot \mathbf{R}_l} \epsilon_{\mathbf{k}} = -V_{\text{u.c.}} \int \frac{d^3 \mathbf{k}}{(2\pi)^3} e^{i\mathbf{k} \cdot \mathbf{R}_l} \epsilon_{\mathbf{k}} \quad (4.69)$$

得到对应的跃迁系数；因子 $V_{\text{u.c.}}$ 是必要的，因为

$$V_{\text{u.c.}} \sum_{\mathbf{R}_l} \cdots \propto V.$$

因此，给定一个能带之后，我们其实总是可以写出一个(4.58)形式的哈密顿量来复刻这个能带，只要引入足够多的 t_l ，就能够得到任何能谱。反过来，将电子的产生湮灭算符 $c_{m\mathbf{i}}^{\dagger}$ 做傅里叶变换(2.29)，(4.58)即可写成对角化的，立刻可以看出能带。

根据第2.1.5节中的论证，(4.58)做完傅里叶变换后不会有 $1/N$ 或是 $1/V$ 之类的因子出现——直接从自由电子气出发加入周期势场计算得到的对角化哈密顿量中也没有这样的因子。因此两种 \mathbf{k} 表象的归一化也是一致的。总之，(2.29)给出的 \mathbf{k} 表象和从自由电子出发得到的限制在第一布里渊区中的简约波矢标记的Bloch电子是完全一样的表象。

请注意上面的推导完全没有用到一个晶胞里面有几种原子之类的条件： i, j 等都是**晶格格点**，而不是具体的原子的位置。对简单晶格，这是非常合理的，而对复式晶格，我们的推导给出了一种把它近似地看成简单晶格的方法。

4.3.2 一般的Wannier波函数

在(4.63)中我们让 $n = m$ ，现在则讨论一般的不同的 m 和 n 之间存在跃迁的模型。在这样的模型中我们指派了一个其实并非我们的体系的能量本征态的Wannier波函数，然后以此为表象写出一个修正后的紧束缚模型，这样就会有带间跃迁了。这种含有带间跃迁的紧束缚模型在对角化之后能带标签将会重新设定。

我们要做的事情大体上是图4.3展示的那样。首先设我们有一组大体上定域在各个原子附近的波函数（对简单晶格如图4.3(a)所示，对复式晶格如图4.3(c)所示），记作 $v_n(\mathbf{r} - \mathbf{R}_i)$ ，这里 i 标记它定域在的晶胞的位置， n 提供进一步的信息，如它是一个晶胞中的哪一种原子附近的哪一个轨道，或者是哪两个原子之间的成键轨道或是反键轨道，等等。同一个 i ，不同 n 的 $v_n(\mathbf{r} - \mathbf{R}_i)$ 之间的内积通常是零，但是不同 i 的 $v_n(\mathbf{r} - \mathbf{R}_i)$ 内积通常不是零（图4.3(a)直观地展示了这一点），需要对不同 i 的 $v_n(\mathbf{r} - \mathbf{R}_i)$ 做正交化，让它们变得更加尖锐，这就得到了Wannier波函数（如图4.3(b)和图4.3(d)中的圆点所示），据此可以将哈密顿量写成(4.58)的形式（见图4.3(b)和图4.3(d)中的跃迁箭头）。正交化不会混合不同的 n ，因此大体上可以认为标签 i 和 n 在正交化前后没有变。

总之我们要做的事情就是获得一组相比于用Bloch波函数算出来的Wannier波函数稍微劣质一些的Wannier波函数。这种稍微劣质一些的Wannier波函数可以通过将不同原子附近的轨道波函数正交化得到，也可以通过线性组合单原子的几个不同的轨道波函数然后再将不同原子附近的这些轨道波函数正交化得到，需要线性组合是因为原子中的电子受其它原子作用而发生能级重组，即轨道杂化。甚至还可以计算单个晶胞内的成键轨道和反键轨道，然后把成键轨道和反键轨道正交化之后当作Wannier波函数。这和坐标表象、 i 表象和晶格动量表象之间的关系很像：从坐标表象到单原子轨道相当于稍稍对哈密顿量做了对角化，

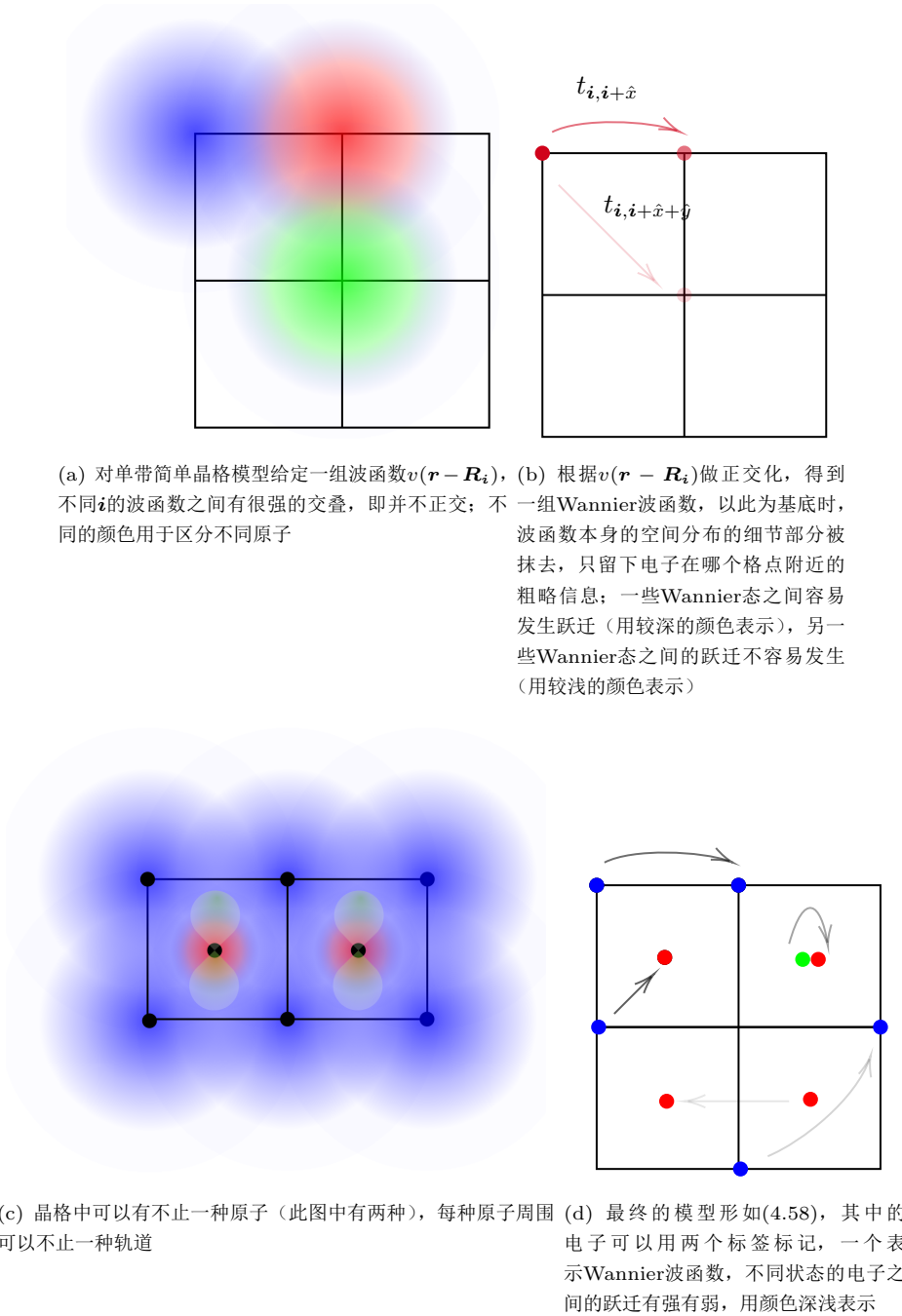


图 4.3: 从给定的局域波函数获得Wannier波函数并在此表象下计算哈密顿量

从单原子轨道到有杂化的轨道相当于又做了一些对角化，到最后得到正确的能带编号，就是完成了能带标签部分的对角化。

正交化：从任意的局域波函数获得Wannier波函数

在给定一组 $v_n(\mathbf{r} - \mathbf{R}_i)$ 之后我们尝试做正交化。由于不同 n 的 $v_n(\mathbf{r} - \mathbf{R}_i)$ 已经是正交的，不需要变动这个标签；我们下面暂时略去 n 。首先定义

$$\phi_{\mathbf{k}}(\mathbf{r}) = \frac{A_{\mathbf{k}}}{\sqrt{N}} \sum_i e^{i\mathbf{k} \cdot \mathbf{R}_i} v(\mathbf{r} - \mathbf{R}_i), \quad (4.70)$$

根据归一化条件

$$1 = \int d^3\mathbf{r} |\phi_{\mathbf{k}}(\mathbf{r})|^2,$$

可以计算得到

$$\frac{1}{|A_{\mathbf{k}}|^2} = \sum_i e^{i\mathbf{k} \cdot \mathbf{R}_i} \int d^3\mathbf{r} v(\mathbf{r})^* v(\mathbf{r} - \mathbf{R}_i). \quad (4.71)$$

我们会发现 $\{\phi_{\mathbf{k}}(\mathbf{r})\}$ 是正交归一化的，因此对它们做反向的傅里叶变换得到的

$$w_i(\mathbf{r}) = \frac{1}{\sqrt{N}} \sum_{\mathbf{k}} e^{-i\mathbf{k} \cdot \mathbf{R}_i} \phi_{\mathbf{k}}(\mathbf{r}) \quad (4.72)$$

也是正交归一化的；不过此时我们有

$$\delta_{ij} = \int d^3\mathbf{r} w_i(\mathbf{r})^* w_j(\mathbf{r}), \quad (4.73)$$

和(4.55)有所不同，在本节的习惯下面应有

$$-t_{mi,nj} = \langle m\mathbf{i} | \hat{h} | n\mathbf{j} \rangle = \int d^3\mathbf{r} w_m^*(\mathbf{r} - \mathbf{R}_i + \mathbf{R}_j) \hat{h} w_n(\mathbf{r}). \quad (4.74)$$

由于 $A_{\mathbf{k}}$ 存在动量依赖， $w_i(\mathbf{r})$ 和 $v(\mathbf{r} - \mathbf{R}_i)$ 未必是一样的——数学上可以证明 $w_i(\mathbf{r})$ 比 $v(\mathbf{r} - \mathbf{R}_i)$ 更加局域。在 $v(\mathbf{r})$ 已经足够局域的时候，

$$\int d^3\mathbf{r} v(\mathbf{r} - \mathbf{R}_i)^* v(\mathbf{r} - \mathbf{R}_j) = \delta_{ij}, \quad (4.75)$$

此时 $w_i(\mathbf{r})$ 就是 $v(\mathbf{r} - \mathbf{R}_i)$ 。

一般的Wannier表象下的哈密顿量及其对角化

现在我们考虑基底 $\{w_{ni}(\mathbf{r})\}$ 下的哈密顿量及其对角化。形式地我们总是可以写出

$$-\frac{\nabla^2}{2m} w_{ni} + U(\mathbf{r} - \mathbf{R}_i) w_{ni} = \epsilon_n w_{ni}, \quad (4.76)$$

不过要注意此时 U 未必在坐标表象下是对角化的； U 与晶体的周期势场 $V(\mathbf{r})$ 需要区分。由于此时有不同的 n 之间的跃迁，我们使用记号 $t_{mn,l}$ 表示 $t_{m(i+l),ni}$ 。跃迁系数计算为

$$\begin{aligned} -t_{mn,l} &= \int d^3\mathbf{r} w_{m(i+l)}(\mathbf{r})^* \left(-\frac{\nabla^2}{2m} + V(\mathbf{r}) \right) w_{ni}(\mathbf{r}) \\ &= \int d^3\mathbf{r} w_{m(i+l)}(\mathbf{r})^* (\epsilon_n - U(\mathbf{r} - \mathbf{R}_i) + V(\mathbf{r})) w_{ni}(\mathbf{r}), \end{aligned}$$

第一项只有在 $l = 0$ 时才有非零值，是 ϵ_n ；第二项和第三项可以使用周期性写成

$$\begin{aligned} & \int d^3\mathbf{r} w_{m(i+l)}(\mathbf{r})^* (-U(\mathbf{r} - \mathbf{R}_i) + V(\mathbf{r})) w_{ni}(\mathbf{r}) \\ &= \int d^3\mathbf{r} w_m(\mathbf{r} - \mathbf{R}_l)^* (-U(\mathbf{r}) + V(\mathbf{r})) w_n(\mathbf{r}), \end{aligned}$$

因此我们有

$$-t_{nn,0} = \epsilon_n + \underbrace{\int d^3\mathbf{r} w_n(\mathbf{r})^* (V(\mathbf{r}) - U(\mathbf{r})) w_n(\mathbf{r})}_{-J_{n,0}}, \quad (4.77)$$

而

$$-t_{mn,l} = \int d^3\mathbf{r} w_m(\mathbf{r} - \mathbf{R}_l)^* (V(\mathbf{r}) - U(\mathbf{r})) w_n(\mathbf{r}), \quad m \neq n \text{ or } l \neq 0. \quad (4.78)$$

我们下面要求 i 仅仅标记晶胞位置，晶胞内部有不同原子，每个原子周围有不同形状的Wannier波函数等事实全部使用 n 标记；这样 i 的自由度数日和 \mathbf{k} 一致，这里的对角化过程相比直接从Bloch波函数获得的Wannier波函数表象下的对角化过程，唯一多出来的就是重新安排能带编号。因此，如果我们采用这样的方案对角化哈密顿量：对 c_{mi}^\dagger 做傅里叶变换(2.29)，然后对 \mathbf{k} 相同， n 不同的不变子空间做对角化，那么得到的 \mathbf{k} 表象同样就是Bloch波函数表象。将 i 表象切换到 \mathbf{k} 表象，给出

$$H = \sum_{\mathbf{k}} \sum_{m,n} \sum_l t_{mn,l} e^{-i\mathbf{k} \cdot \mathbf{R}_l} c_{m\mathbf{k}}^\dagger c_{n\mathbf{k}}, \quad (4.79)$$

因此只需要对角化以上哈密顿量——这只不过是线性组合一下不同 m, n 而已——就得到了电子能带。

跃迁系数大小和能带形态

我们现在来分析跃迁系数的大小受什么因素影响。有两种参数可以影响跃迁系数，一种是晶格常数，一种是 $V(\mathbf{r})$ 。由于Wannier波函数非常局域， $U(\mathbf{r} - \mathbf{R}_i)$ 总是非常“深”的，即其各个矩阵元基本上是负的且绝对值很大； $V(\mathbf{r})$ 的“尖锐”与否从而影响了 $V(\mathbf{r}) - U(\mathbf{r})$ 的大小。

跃迁系数会显著影响能带形态。能带形态对电子行为显然有很大的影响。例如，能带更窄的系统态密度更大，从而库伦散射更明显，从而也更加容易在相互作用加入之后变成强关联系统（但实际上有没有强关联系统，或是强关联效应是否有趣，则是说不好的，例如强烈的库伦散射也可能让电子就定域在本地，低能下看就像一个自旋自由度，或者干脆就是离子实的一部分）。

对只有最近邻或是次近邻的跃迁的系统，我们有一个非常直观的物理图像：似乎可以用“某个电子正在某个格点附近”来标记一个电子，即电子似乎被紧紧地束缚在晶格上，无法长距离移动，这就是“紧束缚”一词的来历。或者，也有可能电子并不是特别局域，但是原子间距很大，以至于在系统中电子通常具有的动能下，非常远距离的跃迁似乎无法一次完成。这就是说，紧束缚模型中，只要 t 不是很大，电子的动能不应该能够到达特别大，从而也不应该有太大的能带带宽（的确应该考虑带宽而不是能量的绝对大小，因为总是可以通过化学势来任意增减能量，而上面的物理图像和化学势毫无关系）。还可以从另一个角度看这个现象：我们知道自由电子的动能项在动量表象下是对角化的，这意味着它在坐标表象下是非常不对角化的，体现在 i 表象中就是有非常远的跃迁。另一方面，晶格势能项在坐标表象下是对角化的，体现在 i 表象中就是 $\epsilon_i c_i^\dagger c_i$ 型的项占据压倒性优势。因此，单电子

越是被束缚，其哈密顿量就越像晶格势能项，只涉及很近邻的跃迁，而越是自由，其哈密顿量就越像自由电子的哈密顿量，涉及很远的跃迁。也可以从傅里叶变换的角度看这个问题：此时(4.69)中只有一个 l 分量，而自由电子则有很多相长干涉的 l 分量，从而紧束缚模型通常能带带宽不宽。总之，以下几个说法基本上是等价的：

1. 系统可以用紧束缚模型或者考虑了次近邻、再次近邻的紧束缚模型加上一个相互作用项描述；
2. 系统中的电子跃迁能力不大；
3. 系统中电子动能有较低的上限。

满足以上等价条件的电子系统具有较窄的能带。另一方面，我们有(4.69)，由于傅里叶变换是线性的，如果动量空间中的两个函数非常接近，那么实空间中它们的对应物也应该足够接近；特别的，如果有一条带是平带，那么根据傅里叶变换的唯一性，我们必然发现 t_l 在 $l \neq 0$ 时为零。于是，似乎如果能带较窄的话，电子的跃迁能力——通过 t_l 表征——也应该较差。然而应注意以下微妙的因素：出现在(4.69)中的Wannier函数未必足够局域；如果这些Wannier函数非常不局域，那么即使 t_l 在 $l \neq 0$ 时为零，我们也不能说电子的跃迁能力真的很差：我们只能说电子的跃迁方式很特别而已。例如，设有一维晶体，其每个原胞包含两个地位不同的原子，费米面附近的能带由如下紧束缚模型描述：

$$H = -t \sum_i c_{i+1,A}^\dagger c_{iB} + \text{h.c.}, \quad (4.80)$$

其中的 $c_{i,A/B}$ 涉及的Wannier函数已经经过了适当的局域化。这个哈密顿量在傅里叶变换后是

$$H = -t' \sum_{\mathbf{k}} \begin{pmatrix} c_{\mathbf{k}A}^\dagger & c_{\mathbf{k}B}^\dagger \end{pmatrix} \begin{pmatrix} 0 & e^{-ika} \\ e^{ika} & 0 \end{pmatrix} \begin{pmatrix} c_{\mathbf{k}A} \\ c_{\mathbf{k}B} \end{pmatrix}. \quad (4.81)$$

对角化这个矩阵，我们清楚地看到有两条平带，分别位于 $\epsilon = \pm t'$ ； $\epsilon = t'$ 对应Bloch波函数

$$\psi_{\mathbf{k},\text{low}} = \frac{1}{\sqrt{2}}(-\psi_{\mathbf{k}A} + e^{ika}\psi_{\mathbf{k}B}), \quad (4.82)$$

而 $\epsilon = -t'$ 对应

$$\psi_{\mathbf{k},\text{high}} = \frac{1}{\sqrt{2}}(\psi_{\mathbf{k}A} + e^{ika}\psi_{\mathbf{k}B}), \quad (4.83)$$

对这两个Bloch波函数做离散傅里叶变换，我们得到

$$w_{i,\text{high}} = \frac{1}{\sqrt{2}}(-w_{iA} + w_{iB}), \quad w_{i,\text{low}} = \frac{1}{\sqrt{2}}(w_{iA} + w_{iB}). \quad (4.84)$$

可是这并不能够说明这个体系中的电子“跃迁能力差”：事实上，这个体系中的电子可以跨原胞跃迁，于是我们看到，由能带中的Bloch态直接得到的Wannier波函数的确并不局域。如果我们只保留一条能带——比如说能量较低的能带——那么我们获得的紧束缚模型和单原子极限下电子不从一个原子跃迁到另一个原子，从而形成单带的那种紧束缚模型的形式是完全一致的，可是这两个体系有着巨大的差异，这种差异体现在它们使用的Wannier波函数上。反过来说，如果我们（通过第一性原理计算发现）某条能带转化而来的Wannier波函数并不能做到局域，这就意味着，如果我们想用一个紧束缚模型描述它，并且这个紧束缚模型基于足够局域的Wannier基底，则我们必须引入这条能带以外的能带（在上文的例子中，可以引入 $\epsilon = t'$ 那条能带）。

以上提到的现象——表面上看起来相同的紧束缚模型中Wannier函数可以具有不同的局域性——经常出现在**拓扑能带论**中：拓扑非平庸的能带在被连续变动到一条平带后，必然和普通的单原子极限的能带有着本质不同，从而，如果我们要求Wannier函数足够局域，拓扑非平庸的能带一般来说需要通过一个多带模型来捕捉。不过没有明显的拓扑性质的能带同样可以有这种性质，此时我们说，这条能带具有**fragile topology**。

窄带一般意味着强关联效应，因为此时库仑相互作用的强度压倒了能带电子的动能（实际上是裸的电子的动能加晶格势能）。不过这倒未必总是导致有趣的效应：原子的内层电子基本上不跃迁，因此能带也非常窄。然而原子的内层电子是离子实的一部分，所以在一般的凝聚态体系中并不导致什么强关联效应。对较低能标的理论而言，窄带只有在费米面附近才有关价值。

4.3.3 原子轨道线性组合和紧束缚模型

设 $w_s(\mathbf{r} - \mathbf{R}_i)$ 为我们选定的Wannier波函数，其中标签 s 标记轨道而我们用 s 而不是 n 做标签，因为不能确定在 $\{w_s(\mathbf{r} - \mathbf{R}_i)\}$ 下不同 s 之间是否有很强的跃迁，从而不能确定 s 标签是不是足够精确地标记了不同的能带。应注意虽然在构造紧束缚模型时，我们经常说某个 n 对应 s 电子或者 p 电子，但严格而言， s 和 n 总是不同的：单原子的轨道理论上说构成了单电子波函数的一组完备基，现在如果我们在不同的空间点上放置多组完备基，显然是不正确地扩大了希尔伯特空间。在纳入考虑的单原子轨道并不多的情况下，将原子轨道做线性组合并构造Wannier波函数是一个合理的近似，但这个近似无法通过增加纳入考虑的单原子轨道的数目来系统性地提升精度。因此，在需要第一性原理地计算能带结构时，一般不以紧束缚模型为出发点。

简单晶格的原子轨道线性组合（LCAO）

在简单晶格中，如果不同原子的轨道基本上没有重叠，那实际上单原子轨道波函数就是Wannier波函数。设不同原子的轨道没有什么重叠，则有

$$\int d^3\mathbf{r} \varphi_n^*(\mathbf{r} - \mathbf{R}_i) \varphi_n(\mathbf{r} - \mathbf{R}_j) = \delta_{ij}, \quad (4.85)$$

$i = j$ 时上式显然是1， $i \neq j$ 时由于两个波函数没有重叠，就是零，因此上式的确是成立的。 $\varphi_n(\mathbf{r} - \mathbf{R}_i)$ 满足

$$\left(-\frac{\nabla^2}{2m} + V_{\text{atom}}(\mathbf{r} - \mathbf{R}_i) \right) \varphi_n(\mathbf{r} - \mathbf{R}_i) = \epsilon_n \varphi_n(\mathbf{r} - \mathbf{R}_i), \quad (4.86)$$

而能带电子的波函数满足

$$\left(-\frac{\nabla^2}{2m} + V(\mathbf{r}) \right) \psi_{n\mathbf{k}} = \epsilon_{n\mathbf{k}} \psi_{n\mathbf{k}}, \quad (4.87)$$

我们现在做展开

$$\psi_{n\mathbf{k}}(\mathbf{r}) = \sum_i a_i \varphi_n(\mathbf{r} - \mathbf{R}_i), \quad (4.88)$$

就得到

$$\sum_i a_i (\epsilon_n + V(\mathbf{r}) - V_{\text{atom}}(\mathbf{r} - \mathbf{R}_i)) \varphi_n(\mathbf{r} - \mathbf{R}_i) = \epsilon_{n\mathbf{k}} \sum_i a_i \varphi_n(\mathbf{r} - \mathbf{R}_i),$$

两边作用(4.85)，就得到

$$(\epsilon_{n\mathbf{k}} - \epsilon_n)a_i = \sum_j a_j \int d^3\mathbf{r} \varphi_n(\mathbf{r} - \mathbf{R}_i)^*(V(\mathbf{r}) - V_{\text{atom}}(\mathbf{r} - \mathbf{R}_i))\varphi_n(\mathbf{r} - \mathbf{R}_j). \quad (4.89)$$

总之，在原子轨道波函数之间真的没有什么交叠的情况下，原子轨道波函数就是Wannier波函数。然而，在大多数情况下，原子轨道波函数还是比较不局域的，我们需要先从原子轨道波函数出发计算得到Bloch波函数，然后做傅里叶变换得到比原子轨道波函数更加局域的Wannier波函数。

复式晶格的LCAO

可以先假装轨道波函数真的是Wannier波函数，然后做杂化

4.3.4 坐标表象作为一种Wannier表象

如果用以自行设定的用不同电子轨道线性组合得到的Wannier波函数为表象，这个表象下的波函数只能够展示电子在不同晶胞上的概率振幅分布，保留了一些“一个电子是哪些轨道的混合”的信息，能带编号比较清晰。以自行设定的用不同电子轨道线性组合得到的Bloch波函数为表象。波函数不显示任何电子的空间分布信息，但是保留了一些“一个电子是哪些轨道的混合”的信息。

4.3.5 表象变换总结

总结一下，形如(4.58)的近独立电子模型最大的好处在于提供了介于哈密顿量完全对角化的Bloch波函数表象（或者说晶格动量表象）和完全没有做过任何处理的坐标表象之间的许多表象。我们上面提到过的表象从坐标表象到晶格动量表象形成如下序列：

1. (1.5)，完全就是“自由电子气动能加势能”，没有做任何处理。最容易写下哈密顿量，最难以理解其物理。这种表象下的波函数清楚地展示了电子在空间各点的概率振幅分布，但是没有提供关于 \mathbf{k} 和 n 的任何信息。
2. 以单原子轨道为表象。相对容易写下来哈密顿量。这种表象下的波函数无法展示每个轨道具体长什么样的信息（因为轨道编号只是一个抽象的量子数），但电子出现在不同原子附近的概率振幅的信息保留了。第一次出现了和能带编号有关的标签（轨道编号）。
3. 以单原子内部的杂化轨道为表象。电子出现在不同原子附近的概率振幅的信息保留了。杂化轨道的编号是能带编号的更好的近似，即能带编号开始清晰起来。
4. 以原子间的成键和反键轨道为表象。电子出现在不同原子附近的概率振幅的信息被模糊化了，但是能带编号变得更清晰了。
5. (4.63)，离完全对角化的 $\epsilon = \epsilon_{\mathbf{k}}$ 只有一步之遥。这个表象下的波函数只能够展示电子在不同晶胞上的概率振幅分布，能带编号非常清晰，而原来的单原子轨道完全模糊了。
6. 能带 $\epsilon = \epsilon_{n\mathbf{k}}$ 。波函数不展示任何电子的空间分布信息，也没有任何关于单原子轨道的信息，只有确定的晶格动量和能带编号。

从上到下，坐标变得模糊而动量变得清晰：先是单原子波函数的细节模糊了，然后是电子在晶格内的哪个原子附近（或者说，复式晶格中在哪套子格子附近）变得模糊了，最后电子在哪个晶胞附近也模糊了，而 \mathbf{k} 变得清晰了。单电子轨道标签先是出现（作为单原子波函数的细节模糊的补偿），然后变得模糊了：先是杂化，然后是成键、反键，单原子轨道的信息逐步丢失，而越来越像正确的能带标签。凝聚态理论研究中通常不会采取这个序列两端的表象，因为它们过于复杂而难以做近似。这就是紧束缚模型的重要之处，虽然当代已经没人用它做任何第一性原理计算了。

到目前为止，所谓“保留一些信息，丢弃另一些”只不过是普通的表象变换：总的信息量都是守恒的。然而一些时候，一种表象下的哈密顿量比另一种更加对角化，从而在这种表象下舍弃或者更加精确的，“积掉”一些自由度，能够获得良好的近似，因此表象变换确实和基于哈密顿量的重整化有着密切的关系。我们将在讨论张量网络方法时用到这个想法。

4.4 电磁场耦合

将电子和一个满足库伦规范的磁矢势 \mathbf{A} 耦合，那么需要将动量替换为正则动量，这个修正在波函数上引入如下的相位变化：

$$\theta = \int d\mathbf{l} \cdot \mathbf{A}. \quad (4.90)$$

在格点模型中，电子仅仅出现在格点上。我们知道紧束缚模型的哈密顿量（即跃迁项）实际上就是动能，因此加入磁场意味着紧束缚模型的 t_{ij} 出现变化，考虑相位变化，则磁场会导致以下修正：

$$t_{i,j} = \exp\left(-ie \int_i^j d\mathbf{r} \cdot \mathbf{A}\right), \quad (4.91)$$

相应的，设一个格点上的闭合路径为 C ，通过它的磁通量为 Φ ，则

$$e^{i\Phi} = \prod_C t_{ij}. \quad (4.92)$$

4.5 相互作用通道

4.5.1 Bloch表象下的相互作用通道

本节讨论从自由电子模型出发如何加入相互作用。在自由电子模型中

$$c_{\mathbf{k}}^\dagger = \frac{1}{\sqrt{V}} \int d^3\mathbf{r} e^{i\mathbf{k} \cdot \mathbf{r}} \psi^\dagger(\mathbf{r}). \quad (4.93)$$

对库伦相互作用，有

$$\begin{aligned} H_{\mathbf{k}_4 \mathbf{k}_3 \mathbf{k}_2 \mathbf{k}_1} &= \frac{1}{2} \int d^3\mathbf{r} \int d^3\mathbf{r}' \langle \mathbf{k}_4 | \mathbf{r} \rangle \langle \mathbf{k}_3 | \mathbf{r}' \rangle V(\mathbf{r} - \mathbf{r}') \langle \mathbf{r}' | \mathbf{k}_2 \rangle \langle \mathbf{r} | \mathbf{k}_1 \rangle \\ &= \frac{1}{2V^2} \int d^3\mathbf{r} \int d^3\mathbf{r}' e^{i(\mathbf{k}_2 - \mathbf{k}_3) \cdot \mathbf{r}'} e^{i(\mathbf{k}_1 - \mathbf{k}_4) \cdot \mathbf{r}} V(\mathbf{r} - \mathbf{r}') \\ &= \frac{1}{2V^2} \int d^3\mathbf{r} \int d^3\mathbf{r}' e^{-i(\mathbf{k}_3 + \mathbf{k}_4 - \mathbf{k}_1 - \mathbf{k}_2) \cdot \mathbf{r}'} e^{-i(\mathbf{k}_2 - \mathbf{k}_3) \cdot \mathbf{r}} V(\mathbf{r}) \\ &= \frac{1}{2V} \delta_{\mathbf{k}_3 + \mathbf{k}_4, \mathbf{k}_1 + \mathbf{k}_2} \int d^3\mathbf{r} e^{-i(\mathbf{k}_2 - \mathbf{k}_3) \cdot \mathbf{r}} V(\mathbf{r}), \end{aligned}$$

我们定义

$$\mathbf{q} = \mathbf{k}_2 - \mathbf{k}_3,$$

就得到

$$H_{\mathbf{k}_1+\mathbf{q}, \mathbf{k}_2-\mathbf{q}, \mathbf{k}_2, \mathbf{k}_1} = \frac{1}{2V} \int d^3\mathbf{r} e^{-i\mathbf{q}\cdot\mathbf{r}} V(\mathbf{r}), \quad (4.94)$$

或者说

$$H_{\text{int}} = \frac{1}{2V} \sum_{\mathbf{k}_1, \mathbf{k}_2, \mathbf{q}} V_{\mathbf{q}} c_{\mathbf{k}_1+\mathbf{q}}^\dagger c_{\mathbf{k}_2-\mathbf{q}}^\dagger c_{\mathbf{k}_2} c_{\mathbf{k}_1}, \quad V_{\mathbf{q}} = \int d^3\mathbf{r} e^{-i\mathbf{q}\cdot\mathbf{r}} V(\mathbf{r}). \quad (4.95)$$

如果考虑自旋，就有

$$H_{\text{int}} = \frac{1}{2V} \sum_{\alpha, \beta} \sum_{\mathbf{k}_1, \mathbf{k}_2, \mathbf{q}} V_{\mathbf{q}} c_{\mathbf{k}_1+\mathbf{q}, \alpha}^\dagger c_{\mathbf{k}_2-\mathbf{q}, \beta}^\dagger c_{\mathbf{k}_2, \beta} c_{\mathbf{k}_1, \alpha}. \quad (4.96)$$

$$V(\mathbf{q}) = \frac{4\pi e^2}{|\mathbf{q}|^2}. \quad (4.97)$$

4.5.2 单能带系统的相互作用哈密顿量

对单带模型，在Wannier表象下，库仑相互作用可以表示成如下矩阵元形式：

$$\begin{aligned} V_{ii'jj'} &= \frac{1}{2} \int d^3\mathbf{r} \int d^3\mathbf{r}' \langle i|\mathbf{r} \rangle \langle i'|\mathbf{r}' \rangle V(\mathbf{r} - \mathbf{r}') \langle \mathbf{r}'|j' \rangle \langle \mathbf{r}|j \rangle \\ &= \frac{1}{2} \int d^3\mathbf{r} \int d^3\mathbf{r}' w_i^*(\mathbf{r}) w_{i'}^*(\mathbf{r}') V(\mathbf{r} - \mathbf{r}') w_{j'}(\mathbf{r}') w_j(\mathbf{r}), \end{aligned} \quad (4.98)$$

于是总的哈密顿量为

$$H = \sum_{i, i', \sigma} c_{i\sigma}^\dagger t_{ii'} c_{i'\sigma} + \sum_{i, i', j, j', \sigma, \sigma'} V_{ii'jj'} c_{i\sigma}^\dagger c_{i'\sigma'}^\dagger c_{j'\sigma'} c_{j\sigma}. \quad (4.99)$$

对上式做傅里叶变换，使用(2.29)会有 $1/2N$ 的因子，然而第2.1.5节也论证了，从 \mathbf{r} 表象切换到动量表象后，二体相互作用项会有 $1/2V$ 的因子。第4.3.3节中已经论证了，无论是自由电子气模型加入周期势扰动，然后做傅里叶变换得到的 \mathbf{k} 表象和对上式做傅里叶变换得到的 \mathbf{k} 表象就是一个东西，所以看起来似乎有矛盾。实际上两者没有矛盾，因为第2.1.5节中的二体相互作用的耦合常数直接就是 $V_{\mathbf{q}}$ ，但是本节中的二体相互作用的耦合常数是 $V_{ii'jj'}$ 傅里叶变换而来，根据(4.56)，两者的确应该差一个 $V_{\text{u.c.}}$ 因子。

与自由电子气模型或凝胶模型不同，紧束缚模型对相互作用是非常敏感的，不能保证加入（即使比较弱的）相互作用后，系统能够用费米液体理论（见第6.3节）描述，甚至不能保证能带仍然是具有明确物理意义的。这是因为紧束缚模型的能带通常更加窄，从而态密度更大；我们也可以如此理解紧束缚模型对相互作用的敏感性：由于电子动能不能取特别大的值，相比之下相互作用能量应该占据主导，从而系统容易具有强关联效应。（见第4.3.2节）一些数值计算方法，如DMRG，需要将所有自由度都定义在晶格上，从而，它们可以毫无困难地用于模拟很大一类系统，无论它们有无强关联，但是却无法用于有效地模拟普通的费米液体系统。

需要注意，并非所有强关联系统都出现在紧束缚模型中。费米液体系统加入了适当的相互作用之后也可以出现强关联效应，此时使用基于格点的计算方法就不能获得很好的效果了。

4.5.3 紧束缚近似成立时的单带相互作用通道

原则上 $U_{ii'jj'}$ 对很多不同的 i, i', j, j' 组合都有非零值。物理上这是很好理解的, 因为库伦相互作用可以将任意原子附近的两个电子散射到任意的其它地方, 由于高度定域的电子动量不确定, 这里无所谓动量守恒的约束。然而, 在紧束缚模型成立时, 足够大的矩阵元的 i, i', j, j' 中, j 和 j' 不是最近邻, 就是完全一样(即所谓on-site repulsion)。由于Wannier波函数高度定域, 从(4.98)中可以看出, 仅有的可能是 $i = i' = j = j'$, 或是 $i = j \neq i' = j'$, 或是 $i = j' \neq i' = j$ 。于是就有

$$\begin{aligned} & \sum_{i, i', j, j', \sigma, \sigma'} V_{ii'jj'} c_{i\sigma}^\dagger c_{i'\sigma'}^\dagger c_{j'\sigma'} c_{j\sigma} \\ &= \sum_{i, \sigma, \sigma'} V_{iiii} c_{i\sigma}^\dagger c_{i\sigma'}^\dagger c_{i\sigma'} c_{i\sigma} + \sum_{i, j, \sigma, \sigma'} V_{ijji} c_{i\sigma}^\dagger c_{j\sigma'}^\dagger c_{j\sigma'} c_{i\sigma} + \sum_{i, j, \sigma, \sigma'} V_{ijij} c_{i\sigma}^\dagger c_{j\sigma'}^\dagger c_{i\sigma'} c_{j\sigma}, \end{aligned}$$

第一项为

$$\begin{aligned} \sum_{i, \sigma, \sigma'} V_{iiii} c_{i\sigma}^\dagger c_{i\sigma'}^\dagger c_{i\sigma'} c_{i\sigma} &= 2 \sum_i V_{iiii} n_{i\uparrow} n_{i\downarrow} + \sum_i U_{iiii} n_i \\ &= \sum_i U_i n_{i\uparrow} n_{i\downarrow} + \frac{1}{2} \sum_i U_i n_i, \end{aligned}$$

这里我们重新定义

$$U_i = 2V_{iiii}. \quad (4.100)$$

第二项实际上是密度-密度相互作用, 即为

$$\sum_{i, j, \sigma, \sigma'} V_{ijji} c_{i\sigma}^\dagger c_{j\sigma'}^\dagger c_{j\sigma'} c_{i\sigma} = \sum_{\langle i, j \rangle} V_{ij} n_i n_j,$$

其中我们重新定义了 $V_{ij} = V_{ijji}$ 。泡利矩阵的点乘有下面的有用结果:

$$\sigma_{\alpha\beta} \cdot \sigma_{\alpha'\beta'} = 2\delta_{\alpha\beta'} \delta_{\beta\alpha'} - \delta_{\alpha\beta} \delta_{\alpha'\beta'}, \quad (4.101)$$

于是可以验证

$$\sum_{i, j, \sigma, \sigma'} V_{ijij} c_{i\sigma}^\dagger c_{j\sigma'}^\dagger c_{i\sigma'} c_{j\sigma} = -2 \sum_{\langle i, j \rangle} V_{ijij} \left(\mathbf{S}_i \cdot \mathbf{S}_j + \frac{1}{4} n_i n_j \right).$$

我们把上式右边的第二项吸收进 V_{ij} 中, 并且重新定义 $2V_{ijij} = J_{ij}$, 于是总的相互作用哈密顿量就是

$$\sum_i U_i n_{i\uparrow} n_{i\downarrow} + \sum_{\langle i, j \rangle} V_{ij} n_i n_j - \sum_{\langle i, j \rangle} J_{ij} \mathbf{S}_i \cdot \mathbf{S}_j + \sum_i U_i n_i.$$

由空间平移对称性, on-site repulsion肯定是均一的, 这样我们可以将 $\sum_i U_i n_i$ 丢进化学势中; 而如果它不是均一的, 就意味着系统实际上不具有完美的离散平移对称结构, 即系统中存在无序, 直观地看, 就是散布的杂质对电子产生散射, 这基本上是一个单体算符, 于是我们就得到了紧束缚模型最一般的哈密顿量:

$$H = - \sum_{\langle i, j \rangle, \sigma} t_{ij} c_{i\sigma}^\dagger c_{j\sigma} + \sum_i U_i n_{i\uparrow} n_{i\downarrow} + \sum_{\langle i, j \rangle} V_{ij} n_i n_j - \sum_{\langle i, j \rangle} J_{ij} \mathbf{S}_i \cdot \mathbf{S}_j - \mu \sum_i n_i + \sum_{i, j, \sigma, \sigma'} \epsilon_{ij\alpha\beta} c_{i\alpha}^\dagger c_{j\beta}. \quad (4.102)$$

这里的每一项都是有意义的, 从左到右分别是:

1. 动能项，衡量电子在格点之间跳跃的可能性；实际上它是真正的动能加上晶格势能，但是我们通常还是将它称为动能，因为它在模型中的地位 and 动能更加相似：和坐标不对易，驱动电子运动，等等；
2. on-site repulsion，这一项是宏观经典模型中唯一重要的密度-密度库伦排斥，仅考虑这个相互作用得到的模型是所谓**Hubbard模型**；
3. 密度-密度相互作用项，可能导致系统中出现持续的、不能平息的电子密度涨落，即所谓电荷密度波；
4. 自旋-自旋相互作用项，这一项来自所谓的交换相互作用（一个例子是做Hartree-Fock近似时的交换能，见(6.9)），可能导致出现自旋密度波，这一项意味着在有多个轨道的情况下，每个轨道上放置一个电子，且所有电子的自旋都平行时，系统能量最低，这是洪特规则的一个例子，不过通常说的“洪特相互作用”很多时候指的同个原子的各个壳层电子之间的相互作用；
5. 化学势，调节系统中电子个数；
6. 无序，可能来自杂质或是晶格的缺陷。

可以看到，和基于电子场 $\psi^\dagger(\mathbf{r})$ ，使用连续的坐标标记的模型不同，紧束缚模型是一个**格点模型**。格点系统在物理中的重要性是由于以下事实：首先晶格是非常常见的系统，从而晶格上的物理的低能有效理论往往就包括格点理论；在我们这里，紧束缚模型中的格点就是原胞或是原子的编号。其次，但凡物理问题涉及无穷而出现不良定义的情况，我们总是可以把问题离散化，人为地将问题放在一个晶格上。在这种用途中需要保证离散化之后的问题能够保证原问题的重要特征，如维数、局域性等。例如，我们总是可以往格点模型中加入一些次近邻的相互作用，适当安排这些相互作用可以改变系统的普适类的维数。

4.5.4 紧束缚近似成立时的多能带相互作用通道

以上的讨论集中在单能带模型中。在有多能带时，我们需要使用八个标签标记相互作用通道：有四个标签标记格点，四个标签标记能带编号。对格点的标签，可以使用第4.5.3节开头的论证，仅有的可能是 $i = i' = j = j'$ ，或是 $i = j \neq i' = j'$ ，或是 $i = j' \neq i' = j$ ，所以我们只需要讨论不同能带之间的相互作用情况。

对on-site repulsion项，即 $i = i' = j = j'$ 的情况，容易看出其多能带版本是

$$\sum_{i,\sigma,\sigma'} \sum_{m,m',n,n'} V_{iiii,mm'nn'} c_{im\sigma}^\dagger c_{in'\sigma'}$$

实际的原子内部的洪特规则应该就是在能够拿到

4.5.5 能带理论何时失效？

基本自由度不能写成经过相互作用修饰的能带电子的系统通常称为**强关联系统**（电子间的相互作用导致一个电子受到扰动别的电子也会跟着发生变化就称为**电子关联**），因为电子和电子之间的关系是如此剧烈，以至于单电子图像根本就是不可靠的。本节给出一些能够暗示能带理论失效的判据，并且列举强关联系统通常会有什么性质。

如果能带计算出现平带，即非常窄的能带，则容易出现强关联系统。可以从好几个角度看这个现象：态密度反比于群速度，从而能带越窄，态密度越大，自然相互作用也越

大；能带窄也意味着电子动能低，因此相互作用能不需要太大就能够产生显著的效果。如前所述，紧束缚模型通常具有这种特征，向紧束缚模型加入相互作用是容易产生强关联系统的。

即使是能带理论失效的情况下，能带的概念也是有意义的，因为它毕竟提供了一个合理的“自由理论”，我们可以向它加入各种相互作用来得到合适的强关联理论。

第二部分

普通导体和绝缘体

这一部分讨论那些重要的基本自由度是电子或某种重整化后的电子（即，近独立电子近似成立）的系统，这样的系统中的电子的行为仍然比较像单电子。我们将在第五章讨论能带论基础上关于金属和绝缘体的理论。各种来自对称性自发破缺的相变，在第八章中讨论。这些话题构成了所谓的传统凝聚态物理的很大一部分。所谓传统凝聚态物理包括三个部分：

- 能带理论，即系统中的电子基本上是某种受到相互作用一定修饰的能带电子。
- 费米液体理论，给出前述相互作用修饰的主要影响。费米液体理论适用于，不过不仅仅适用于能带电子系统，因此我们将它和能带理论分开说；但是应该注意，如果费米液体理论不适用，系统的基本自由度不是电子型自由度，那么此时关于电子的能带论不能直接拿来分析系统行为，实际上失去了意义，当然，如果系统的基本自由度仍然是点状粒子，那么这些粒子还是有能带结构。
- 对称性自发破缺，这是一种相对平凡的费米液体理论失效的情况，即在低温下由于特殊的相互作用通道，电子发生凝聚，产生比费米液体态更加稳定的，但是对称性低于费米液体态的状态，即从而出现对称性自发破缺。对称性自发破缺相内部的基本自由度是电子凝聚体的涨落，而电子型激发通常是有能隙的（因为需要拆开一部分凝聚体才能够产生单独的电子）。

我们还将第七章中讨论固体电子结构的密度泛函理论（DFT）计算。虽然原则上DFT可以计算一切电子系统，实际上还是主要还是用于处理能带理论成立的系统。

第五章 金属与能带绝缘体中的电子态

在几个冗长的、本质上是在回顾单粒子量子力学和群表示论的章节之后，我们开始分析一些实际的系统。本章讨论那些电子重整化后形成可以辨认的电子型准粒子的系统，基本上就是传统意义上的“金属”、“绝缘体（为了和因为相互作用导致的或是杂质导致的绝缘体区分，通常称为能带绝缘体）”、“半导体”等。本章的内容将足够解释常见的这些材料的行为——实际上，“有费米面的导体”几乎可以成为金属的一个定义，而能带绝缘体因为只含有满带，其费米海的形状和第一布里渊区完全一样，而且费米能级附近根本没有电子模式，无法将费米能级附近的态在动量空间中排列成一个几何体，总之，对能带绝缘体谈不上有费米面。

仔细想一下，金属中能够有可以辨认的电子型准粒子这件事是非常奇怪的：金属中的电子的库仑相互作用的量级通常和电子动能一样，似乎没有什么理由认为在如此强的相互作用下单电子模型能够适用。然而实验的确告诉我们金属的行为在很多方面非常像费米气体，即实际上描述金属的低能有效理论基本上就是一个费米气体加上一定的相互作用修饰。

为什么对金属单电子图像有效仍然是并不完全清楚的问题。相互作用特别强烈以至于单电子图像不适用的系统是确实存在的（我们将在后文讨论它们），那么单电子图像适用就不是库仑相互作用电子系统必然会有性质，具体怎样的晶体能够让单电子图像适用，怎样的晶体不行，仍然是有待研究的。

无论如何，实验事实告诉我们，对寻常的金属和能带绝缘体，我们可以以某种电子气——自由电子气，或者紧束缚模型等——为自由理论，以库仑相互作用为微扰，或者，可以使用这种凝聚态场论简化后得到的动理学，通常就是玻尔兹曼方程，这样可以避开完整的Keldysh场论而讨论一些非平衡问题。

本章将首先介绍能带电子的波包动力学和全量子力学处理，并首先使用玻尔兹曼方程、然后使用格林函数方法，计算电子间库伦排斥产生的影响，最后表明对相当一类系统，电子间库伦排斥只会让能带电子形成的费米气体变成费米液体。费米液体在粒子数涨落不大的情况下，单粒子能谱和费米气体一样，热容、磁化率等和费米气体具有类似的形式，这反过来说明了单纯使用能带理论就能够足够精确地描述寻常的金属和能带绝缘体的原因，因此我们说，费米液体理论是能带论的基础，这里的能带论指的是“系统的基本自由度可以近似看成能带电子”这一假设。

5.1 一些常见的材料的电子结构

5.2 外加电磁场下能带电子的准经典处理

本节做如下近似：电子可以看成波包，要求它在动量空间比较局域，在实空间相对于

晶格常数比较扩展，但是在宏观尺度上仍然足够局域。这通常要求原胞内的离子势没有剧烈变化，空间势的变化比较缓慢，外场没有破坏能带，外场变化足够慢，碰撞是瞬时发生的，散射较弱。它们的作用在后文的推导中可以看到。

5.2.1 Bloch电子波包，波包动力学和半经典近似

考虑一个可以被看成晶格动量为 \mathbf{k}_0 的Bloch电子波包。显然这个波包由同一个能带中的不同晶格动量成分组成（不同能带的电子具有非常不同的色散关系，它们各自运行的速度非常不同，用它们是无法组成一个比较稳定的波包的；以下略去能带标记 n ）。注意波包的存在性本身就要求空间势的变化不能太剧烈，否则Bloch波函数在一个晶胞中将基本上定域在原子周围，无法形成比较“柔和”的波包。波包中各个成分的晶格动量 \mathbf{k}' 均很接近 \mathbf{k}_0 ，假定大体上

$$-\frac{\Delta}{2} \leq k_i = k_{0i} - k'_i \leq \frac{\Delta}{2}, \quad (5.1)$$

从而可以做展开

$$\epsilon_{\mathbf{k}'} = \epsilon_{\mathbf{k}_0} + \mathbf{k} \cdot (\partial \epsilon_{\mathbf{k}} / \partial \mathbf{k})|_{\mathbf{k}_0}. \quad (5.2)$$

在没有外场时Bloch波函数是能量本征态，其时间演化是平凡的，它们叠加而成的波包的形式形如

$$\psi(\mathbf{r}, t) = \sum_{\mathbf{k}'} a_{\mathbf{k}} \psi_{\mathbf{k}'}(\mathbf{r}, t) = \sum_{\mathbf{k}'} a_{\mathbf{k}} e^{i(\mathbf{k}' \cdot \mathbf{r} - \epsilon_{\mathbf{k}'} t)} u_{\mathbf{k}'}(\mathbf{r}, t),$$

其中 $a_{\mathbf{k}}$ 满足归一化条件。代入 $\epsilon_{\mathbf{k}'}$ 的展开式，能够得到以下近似表达式

$$\begin{aligned} \psi(\mathbf{r}, t) &= e^{i(\mathbf{k}_0 \cdot \mathbf{r} - \epsilon_{\mathbf{k}_0} t)} \sum_{\mathbf{k}} a_{\mathbf{k}} u_{\mathbf{k}'}(\mathbf{r}) e^{i\mathbf{k} \cdot (\mathbf{r} - (\partial \epsilon_{\mathbf{k}} / \partial \mathbf{k})|_{\mathbf{k}_0} t)} \\ &= e^{i(\mathbf{k}_0 \cdot \mathbf{r} - \epsilon_{\mathbf{k}_0} t)} u_{\mathbf{k}_0}(\mathbf{r}) \sum_{\mathbf{k}} a_{\mathbf{k}} e^{i\mathbf{k} \cdot (\mathbf{r} - (\partial \epsilon_{\mathbf{k}} / \partial \mathbf{k})|_{\mathbf{k}_0} t)}, \end{aligned}$$

因此可以看出，波包前进的速度是 $\partial \epsilon_{\mathbf{k}} / \partial \mathbf{k}$ 在 $\mathbf{k} = \mathbf{k}_0$ 时的值。这个对 \mathbf{k} 的求和可以通过以下方式近似求出：我们认为

$$\psi(\mathbf{r}, t) = e^{i(\mathbf{k}_0 \cdot \mathbf{r} - \epsilon_{\mathbf{k}_0} t)} u_{\mathbf{k}_0}(\mathbf{r}) \quad (5.3)$$

Δ 是很小的，即 $\Delta \ll 1/a$ ，即

$$\frac{1}{\Delta} \gg a, \quad (5.4)$$

或者说，为了让波包描述有效，波包必须远大于原胞。这是合理的，因为如果使用波包的描述，我们是看不到晶格的，因此必须保证我们讨论的问题的尺度确实远大于晶格常数。

因此，结论是，一个“晶格动量大体上是 \mathbf{k}_0 ”的电子可以使用一个坐标 \mathbf{r} 以及它的波函数的形状描述，其波函数是一个波包，在没有外场——从而系统的本征态就是Bloch电子——时，可以认为这个波包的运动速度由 $\partial \epsilon_{\mathbf{k}} / \partial \mathbf{k}$ 给出。

现在考虑施加了外场的情况。我们之前假设过外场的空间变化不大，那么外场可以近似通过以下势场引入：

$$V = \mathbf{r} \cdot \mathbf{F}.$$

与自由电子的情况不同，Bloch波矢 \mathbf{k} 不能写成一个解析形式非常简单的算符的本征值，因此我们采取一个略微迂回的方法，考虑某个初基格矢 \mathbf{a} 方向上的平移算符

$$T|\mathbf{r}\rangle = |\mathbf{r} + \mathbf{a}\rangle,$$

Bloch波函数是它的本征态，有

$$T\psi_{n\mathbf{k}} = e^{i\mathbf{k}\cdot\mathbf{a}}\psi_{n\mathbf{k}},$$

于是通过计算 $\langle T \rangle$ 的演化方程就能够计算出外场存在时波包的 \mathbf{k} 的演化方程。我们有

$$\frac{d\langle T \rangle}{dt} = \langle i[H, T] \rangle = i\langle [V, T] \rangle \Theta$$

通过对 $|\mathbf{r}\rangle$ 的作用可以验证

$$[x_i, T] = aT,$$

其中 x_i 指的是 \mathbf{a} 方向上的 \mathbf{r} 分量；其它 \mathbf{r} 分量和 T 对易。于是

$$[V, T] = F_i a T.$$

由于我们假定了波包在实空间非常舒展、在倒空间比较狭窄，可以认为

$$\langle T \rangle = e^{i\mathbf{k}\cdot\mathbf{a}},$$

于是 $\langle T \rangle$ 的时间演化方程最终转化为

$$\frac{d}{dt}e^{i\mathbf{k}\cdot\mathbf{a}} = iF_i a e^{i\mathbf{k}\cdot\mathbf{a}},$$

就有

$$i\mathbf{a} \cdot \dot{\mathbf{k}} = iF_i a = i\mathbf{a} \cdot \mathbf{F}.$$

将 \mathbf{a} 取遍三个初基格矢就得到

$$\dot{\mathbf{k}} = \mathbf{F}.$$

于是，我们就得到了

$$\mathbf{v} = \frac{\partial \epsilon_{\mathbf{k}}}{\partial \mathbf{k}}, \quad \frac{d\mathbf{k}}{dt} = \mathbf{F}, \quad (5.5)$$

联立求解这组方程就获得了外场下的能带电子运动情况。实际上，我们有

$$\frac{d\mathbf{v}}{dt} = \frac{d}{dt} \frac{\partial \epsilon_{\mathbf{k}}}{\partial \mathbf{k}} = \frac{d\mathbf{k}}{dt} \cdot \frac{\partial^2 \epsilon_{\mathbf{k}}}{\partial \mathbf{k}^2},$$

即

$$\frac{d\mathbf{v}}{dt} = \frac{1}{m^*} \mathbf{F}. \quad (5.6)$$

这里的 m^* 正是(4.28)中的有效质量。上式是各向异性版本的牛顿第二定律，不过要注意 m^* 不是把 $[\partial^2 \epsilon_{\mathbf{k}} / \partial k_i \partial k_j]_{ij}$ 中的每个元素取倒数，并且可能有对 \mathbf{k} 的依赖；此外这里的 \mathbf{k} 一直就是Bloch波矢而不是广延波矢。

应注意以上的运动方程对空穴也是成立的（我们可以想象，动量空间中，空穴周围的电子都在按照(5.5)移动，则等效的，空穴也在如此移动），不需要改变任何东西的正负号。特别的，无需改变电场力 $\mathbf{F} = -e\mathbf{E}$ 中单电子电荷 $-e$ 的正负号。所谓空穴看起来像是带正电的粒子乃是因为空穴通常出现在带顶，从而电子的有效质量 m^* 是负的，于是

$$\frac{d\mathbf{v}}{dt} = \frac{-e\mathbf{E}}{-|m^*|} = \frac{e\mathbf{E}}{|m^*|}, \quad (5.7)$$

因此如果我们希望空穴的质量应该是正的，那么的确空穴的电荷应该取为正的才能得到正确的运动方程。也可以换一个角度看问题：电磁场和电子的相互作用哈密顿量在费米面以下的部分正比于 $c^\dagger c$ ，由于变换(4.51)，这部分哈密顿量也可以说是正比于 $-b^\dagger b$ ，于是的确

对空穴来说电荷变成正的了，而由于在变换(4.51)下已填充的电子能带被翻到了费米面以上，成为了未填充的空穴能带，空穴的有效质量也此时可以看成正的，于是我们又一次得到了上面的空穴运动方程。另一方面，计算空穴导致的电流时，空穴的电荷的确应该看成正的。于是空穴可以看成质量为导带带顶的电子有效质量的相反数、带正电的粒子。

最后，动量的正负号也应当注意；如果一个空穴在能带图上位于 \mathbf{k} ，那么它对体系总动量的贡献应当是 $-\mathbf{k}$ ；因此在书写动量守恒关系的时候要特别注意一个动量是关于电子的还是关于空穴的。一些作者会人为定义一个空穴的动量是它在能带图上的位置的相反数，于是这样动量守恒关系里面就不需要考虑正负号；此外可以验证，此时 m^* 和 q 都应该取成正的，并且在书写能量守恒定律时空穴能量也应该算成正的，即

$$\epsilon_{\mathbf{k}^{\text{hole}}}^{\text{hole}} = -\epsilon_{\mathbf{k}} = -\epsilon_{-\mathbf{k}^{\text{hole}}}, \quad (5.8)$$

此处 $\epsilon_{\mathbf{k}}$ 是空穴在能带图上的色散关系，即价带电子的能量，扣除费米能后是负的，而 $\epsilon_{\mathbf{k}^{\text{hole}}}^{\text{hole}}$ 是在体系中产生一个空穴的能量，是正的；在将空穴电荷设置成正的、在能带图上的 \mathbf{k} 和 ϵ 前面加上负号之后，可以验证空穴的运动方程是正确的。

在本节给出的准经典理论中不存在从一个能带到另一个能带的跃迁。运动学上看，这是因为含有多个能带分量的波包会很快分裂开来，而只含有一个能带分量的波包则是相对稳定的。考虑电子的能谱，这是因为不同能带之间的能隙可以看成是一个势垒，从而，一个波包到了能带边界会很快被弹回来。在一些情况下能带之间的隧穿是很重要的，此时本节给出的理论必须加以修正。

本节使用同时确定 \mathbf{p} 和 \mathbf{r} 标记电子状态，这当然是有条件的，具体来说是电子的德布罗意波长要远小于外场发生变化的特征长度，由于会被激发的电子集中在费米面附近，有

$$1/k_F \ll L, \quad (5.9)$$

其中 L 为外场发生变化的特征尺度；常常将外场在空间上做傅里叶变换，则上式等价于

$$k_F \gg k, \quad (5.10)$$

其中 k 是外场的动量尺度，大体上是 $1/L$ 。

5.2.2 外加电场下的电子振荡

现在设一个恒定电场被加到晶体上，则会发生电子速度和晶格动量的振荡。例如，考虑一个带底位于 $\mathbf{k} = 0$ 处的能带，设 $t = 0$ 时电子在 $\mathbf{k} = 0$ 处，由于是带底，有 $\partial\epsilon_{\mathbf{k}}/\partial\mathbf{k} = 0$ ，即电子无速度，而 $m^* > 0$ ，于是电子速度增加。而到了布里渊区边界，由于晶格周期势场，能带变平，此时为带顶，由于 $\partial\epsilon_{\mathbf{k}}/\partial\mathbf{k}$ 为零，电子速度为零，而由于是带顶， $m^* < 0$ ，于是电子被反向加速；注意此时 \mathbf{k} 的变化方向还是沿着 \mathbf{E} 的， \mathbf{k} 走出了第一布里渊区，或者等价地说，从第一布里渊区的另一侧进入了第一布里渊区。这样，电子会从带底出发，首先被加速，到了一个 $m^* = \infty$ 的地方时速度最大，然后减速，到了带顶速度正好降到零，然后 \mathbf{k} 冲出第一布里渊区，而从第一布里渊区的另一侧回到第一布里渊区内，而 \mathbf{v} 的方向则反了过来，最后回到带底。在倒空间，电子从带底运动到带顶，然后从第一布里渊区的另一侧（也是带顶）回到带底；在实空间，出现能带倾斜，这就是所谓的**Bloch振荡**。

在实际实验中这个过程很难观察到，因为电子在运动中会不断受到介质中的其它准粒子和杂质的散射，这些本节没有考虑，而当电场强到电子运动得足够快，从而来不及被散射，让这种振荡得以发生时，往往或是材料被击穿，或是能带间隧穿会变得很明显，

前一种情况下根本不能将介质看成晶体，后一种情况下波包动力学不可靠。实际上，由于 \mathbf{r} 和 \mathbf{p} 的不对易性，在引入电场后严格来说就不能够让 \mathbf{r} 和 \mathbf{p} 同时出现。我们此处所谓“能带在实空间中倾斜”的说法是建立在系统中没有任何东西有尖锐的空间变化，从而短距离看，似乎加入电场前后电子都处在Bloch态这一基础上的。在这个假设被打破时——比如说，电势在空间中有急剧的变化，或者说有极强的电场时——波包动力学本来就不适用。

Bloch振荡的周期的量级和能带结构关系不大，而主要取决于晶格常数。电子在倒空间的运动速度为

$$\frac{dk}{dt} = eE,$$

而第一布里渊区宽度大体上为 $2\pi/a$ ，因此

$$\omega \sim eEa. \quad (5.11)$$

Bloch振荡实际上指出了一件事：超高纯度的金属外加恒定电场实际上是无法远距离导电的，因为这样的金属中的载流子——电子和空穴——由于Bloch振荡，会在一个小范围内打转。通常的服从欧姆定律的导电性必须要求一定的对载流子的散射。实际上，在宏观电动力学中就是如此：欧姆定律的存在本身就意味着有能量持续不断地从电磁场能量和介质中电子能量里耗散出去。

5.2.3 外加磁场下的电子振荡

现在我们转而考虑有外加磁场时电子的响应，同样使用准经典的波包理论，并且忽略碰撞。此时需要求解

$$\mathbf{v} = \frac{\partial \epsilon_{\mathbf{k}}}{\partial \mathbf{k}}, \quad \frac{d\mathbf{k}}{dt} = -e\mathbf{v} \times \mathbf{B}. \quad (5.12)$$

第二个方程意味着沿着磁场方向的 \mathbf{k} 分量不发生改变，且由于洛伦兹力不做功，电子在 \mathbf{k} 空间的运动轨迹是垂直磁场的平面与等能面的交线。因此我们不做任何计算就确定了电子在倒空间中的行为。应当注意，从 \mathbf{k} 的轨迹可以推出 \mathbf{v} 的轨迹，反之亦然，从而，如果我们测量有外加磁场的样品的电流响应——后者能够反映 \mathbf{v} ——是可以重建出此时的 \mathbf{k} 的轨迹，从而得到费米面形状的。特别的，如果费米面和第一布里渊区边界连接，则可以观察到方向随着时间推移从不倒转的电流。

例如，对自由电子可以解析求解以上方程，得到的是我们熟知的回转频率为

$$\omega_0 = \frac{eB}{m} \quad (5.13)$$

的螺旋运动。

5.2.4 经典霍尔效应

首先考虑经典电动力学预言的霍尔效应。在有自由载流子的固体材料中电导率是一个张量。设我们在 y 方向施加一个电场 E_y ，并施加一个在 $-z$ 方向的磁场 B 。设载流子为电子，则稳态时电子应当朝左运动，且

$$evB = eE_y,$$

而电流方向朝右，大小为

$$j_x = -n_e ev = \frac{en_e}{B} E_y.$$

我们将 σ_{xy} 这种横向的电导称为霍尔电导，记为

$$\sigma_H = \frac{e}{B} n_e. \quad (5.14)$$

定义磁通量子（它是超导中磁通量子化的单位乘以2）

$$\Phi_0 = \frac{h}{e}, \quad (5.15)$$

一个电子绕着磁通量子整数倍的一个磁通量转一圈，不会产生可观测的效果（因为Berry相的变化是 2π 的整数倍）。磁通量子是体系磁通量的一个自然的尺度。相应地，定义归一化磁通密度为

$$n_\Phi = \frac{B}{\Phi_0}, \quad (5.16)$$

这给出了一个自然的电子数尺度，于是定义

$$\nu = \frac{n_e}{n_\Phi} = \frac{n_e h}{B e}. \quad (5.17)$$

使用以上归一化之后的物理量，就可以写出霍尔电导为

$$\sigma_H = \nu \frac{e^2}{h}. \quad (5.18)$$

当然，到这里， ν 其实还是可以连续取值的。

5.3 能带电子的经典动理学

5.3.1 电子运动和碰撞的玻尔兹曼方程描述

背景知识 5.1: 玻尔兹曼方程

关于玻尔兹曼方程的推导和其意义的诠释见流体和软物质物理笔记中的第2.1.1节中的内容。我们要注意的虽然玻尔兹曼方程本身的导出是非常经典的，它实际上是某个格林函数运动方程退化而来的，从而可以在一些情况下描述量子系统。

现在将电子-电子相互作用纳入考虑。

由于填满的能带中同时具有Bloch波矢为 \mathbf{k} 和 $-\mathbf{k}$ 的电子，外加电场时它们运动方向相反，无导电性。

Drude模型

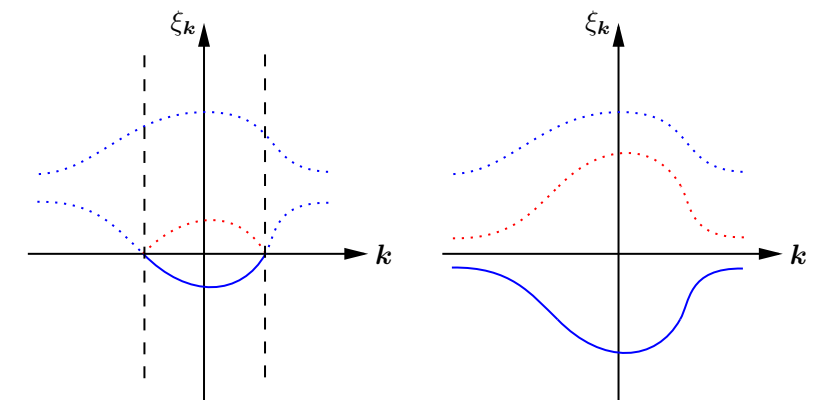
5.3.2 热力学性质

金属中的准粒子主要就是电子和声子。按照第3.2.5节，在极低温下声子能提供形如 T^3 的热容，而作为对比，极低温下电子热容正比于 T （见(4.14)）。因此在极低温下金属的比热主要由电子贡献，因为此时 T 相比 T^3 要明显。

5.3.3 导电性和能带形式的分类

能带论能够解释为什么绝缘体不导电。自由电子气模型无法解释为什么绝缘体不导电——绝缘体中同样有大量的电子，似乎本应该导电。

现在可以根据导电性将较为寻常的、可以使用能带论解释的系统分类如下：



(a) 金属的能带结构，费米能级在某一条能带内，有费米面，载流子在费米面上无能隙
(b) 能带绝缘体的能带结构，费米能级不在任何一条能带内，无费米面，载流子有能隙

图 5.1: 金属和能带绝缘体的能带结构，蓝色实线表示已经被占据的电子态，蓝色虚线表示没有被占据的电子态，红色虚线和蓝色实线沿着 x 轴轴对称，表示满的电子能带就是空的空穴能带

- 电子的能带被完全填满的系统是绝缘体，称为能带绝缘体。
- 金属就是有一些有电子的能带只是部分被填满的系统。
- 半金属 (semimetal)是一些有电子的能带只有少量电子的系统，其导电性不好，但是仍然呈现一些金属的性质。
- 半导体是掺杂了的绝缘体，如果杂质能够形成能量大体上是在满带顶附近的局域空穴态，以及能量大体上是在空带底附近的局域电子态，那么热涨落就会让一些电子填充进空带中，而让满带中出现空穴，这就同时形成了两种载流子。不能依靠热涨落让电子从满带跳往空带，因为这要求费米温度量级的温度；我们需要掺杂才能够形成半导体。

在金属中，根据能带形状又可以分成这么几类：

使用 ξ_k 作为系统能量的定义，金属中的载流子——电子、空穴——在费米面附近是没有能隙的，因为 ξ_k 可以低到零，而能带绝缘体中的载流子是有能隙的。本节讨论的都是近独立能带电子系统，因此载流子有能隙则系统有能隙，载流子无能隙则系统无能隙。金属是无能隙的系统，因此其输运性质比有能隙的能带绝缘体要好。

5.4 外加电磁场下能带电子的量子理论

5.4.1 强磁场下电子的量子理论

忽略晶格周期势场的半经典模型中磁场下的电子处于束缚态，因此可以预期，强磁场下将会出现明显偏离(5.12)的行为，半经典模型不再适用，因而值得单独拿出来处理。

自由电子在磁场中的朗道能级

我们采用对称规范，设磁场方向为 z 方向，且设

$$A_x = -\frac{1}{2}By, \quad A_y = \frac{1}{2}Bx, \quad (5.19)$$

这样通过

$$\mathbf{B} = \nabla \times \mathbf{A}$$

就得到一个 z 方向上大小为 B 的磁场。自由电子被放进磁场后，哈密顿量为

$$H = \frac{\hbar^2}{2m} \left(-i\nabla + \frac{e}{\hbar} \mathbf{A} \right)^2. \quad (5.20)$$

我们把 \hbar 放回了定义中，因为本节将需要频繁地用到这个能量尺度。我们忽略了自旋与磁场的耦合，但是我们后面会发现这是无关紧要的。(5.20)中的导数被协变导数替代了，以满足 $U(1)$ 对称性。由于 z 方向没有 \mathbf{A} 分量，实际上可以将(5.20)当成二维电子气的哈密顿量，电子在 xy 平面上处于束缚态，而在 z 方向和自由电子没有区别。因此接下来我们讨论加入了磁场的二维自由电子气。

通过量纲分析可以发现(5.20)中有一个特征长度

$$l_0 = \sqrt{\frac{\hbar}{eB}}, \quad (5.21)$$

同时可以看出磁通量子确实是磁通量的一个尺度：

$$\Phi_0 = \frac{h}{e} \sim 2B\pi l_0^2.$$

这样，两个方向上的协变导数为

$$D_x = \partial_x + \frac{i}{2l_0^2}y, \quad D_y = \partial_y - \frac{i}{2l_0^2}x. \quad (5.22)$$

哈密顿量就转化为

$$H = -\frac{\hbar^2}{2m}(D_x^2 + D_y^2).$$

容易验证 D_x 和 D_y 之间具有对易关系现在定义

$$D^\pm = D_x \pm iD_y, \quad (5.23)$$

并定义复变量 z 为

$$z = x + iy, \quad (5.24)$$

则可以得到

$$[D^-, D^+] = -\frac{2}{l_0^2}. \quad (5.25)$$

这意味着它们差一个常数就是一对升降算符。定义

$$a = i\frac{l_0}{\sqrt{2}}D^-, \quad (5.26)$$

就有

$$H = \frac{\hbar^2}{ml_0^2} \left(a^\dagger a + \frac{1}{2} \right) = \hbar\omega_0 \left(a^\dagger a + \frac{1}{2} \right). \quad (5.27)$$

因此系统具有分立的能级，称为朗道能级，其中

$$\omega_0 = \frac{eB}{m} \quad (5.28)$$

正是经典图景中电子在磁场下做匀速圆周运动的圆频率。在磁场很弱的时候，这个能谱接近连续，此时可以使用(5.20)。而在磁场很强时就有明显的分立的能级，且能级间距远大于 $\hbar B$ ，从而自旋-磁场耦合导致的能级分裂可以忽略。把 z 方向的运动加上，就是

$$H = \frac{\hbar^2}{ml_0^2} \left(a^\dagger a + \frac{1}{2} \right) + \frac{p_z^2}{2m} = \hbar\omega_0 \left(a^\dagger a + \frac{1}{2} \right) + \frac{\hbar^2 k_z^2}{2m}. \quad (5.29)$$

实际上，朗道能级对应的波函数可以写成一个全纯函数乘以一个高斯因子。设 Ψ 是基态波函数，我们做拟设

$$\Psi(x, y) = f(z, z^*) e^{-zz^*/4l_0^2}, \quad (5.30)$$

由于

$$D^- \Psi = 0,$$

展开计算可以发现

$$\frac{\partial f}{\partial z^*} = 0,$$

这表明 f 一定是全纯函数，即得所求证。

二维电子气在强磁场下的行为

我们现在考虑 z 方向高度受限的一个系统，即考虑一个真正的自由二维电子气在强磁场下的行为。自由二维电子气的 \mathbf{k} 分布在平面上，其“费米面”实际上是费米圆。现在讨论二维电子气的朗道能级的简并度。有一点是可以确定的：朗道能级的简并度非常高，因为原本准连续的 k_x, k_y 被“压缩”到了分立的朗道能级上。二维平面上自由电子计入自旋的态密度为

$$\frac{dn}{d\epsilon} = 2 \times \int dl \frac{1}{|\partial \epsilon_{\mathbf{k}} / \partial \mathbf{k}|} = \frac{mS}{\pi \hbar^2}, \quad (5.31)$$

单个朗道能级的简并度是

$$D = \frac{mS}{\pi \hbar^2} \times \hbar\omega_0 = \frac{Se}{\pi \hbar} B. \quad (5.32)$$

直观地看，原本连续的二维自由电子能带被分割成了若干长度为 $\hbar\omega_0$ 的小段，每个小段被“压缩”到一个朗道能级上；每个小段上的能量平均值和对应的朗道能级的能量刚好一样。

一个能够赋予明确的半经典意义的数量级估计如下。(5.30)给出了一个特征长度为 l_0 的波包。当然，将一个波包做平移之后还是可以得到一个波包。两个波包如果不重叠，就是彼此正交的系统的本征态。设体系总面积为 S ，我们可以认为体系被分割成了一系列大小为 πl_0^2 的圆，于是简并度为

$$\frac{S}{\pi l_0^2} \sim S \frac{eB}{h} = \frac{\Phi}{\Phi_0}.$$

其中定义

$$\Phi_0 = \frac{h}{e} \quad (5.33)$$

为磁通量子。

实际上，根据半经典理论，有

$$\frac{\hbar^2 k^2}{2m} = \hbar\omega_0 n,$$

因此第 n 能级在动量空间中画出的圆的面积是

$$S_k^n = \pi k_n^2 = \frac{2\pi e}{\hbar} B n, \quad (5.34)$$

因此相邻两个能级的倒空间中的圆的面积差（是一个环）是

$$\Delta S_k = \frac{2\pi e}{\hbar} B. \quad (5.35)$$

既然运动方程为

$$\hbar d\mathbf{k} = -e d\mathbf{r} \times \mathbf{B},$$

有

$$\Delta S_r = \frac{2\pi\hbar}{eB}. \quad (5.36)$$

注意到由于 $\Psi(\mathbf{r})$ 可以被恰当地归一化，它在电子气占据的范围内不应该有任何奇点，于是可以做泰勒展开

$$f(z) = c_0 + c_1 z + c_2 z^2 + \cdots,$$

实际上可以证明， $\{z^m e^{-|z|^2/4l_0^2}\}$ 构成一组完备正交基。现在寻找波函数幅值最大的位置，即计算

$$\frac{\partial |\Psi_n(\mathbf{r})|^2}{\partial r} = 0$$

的解，其解为

$$r_n = \sqrt{2n} l_0.$$

换言之我们得到了一组年轮一样的基矢量。设基态简并度为 n 则

$$S = \sqrt{2n} l_0,$$

从而

$$n = S \cdot \frac{eB}{\hbar c} = \frac{\Phi}{\Phi_0}.$$

在磁场很大时，我们有若干高度简并，同时彼此之间有明显间距的离散能级。 B 增大会让能级间距线性增大，同时让单能级简并度线性增大，于是总的模式数目保持不变。

二维电子气的费米能级实际上会导致一个很有趣的现象：缓慢增大或是减小磁场然后待系统达到平衡，系统能量会发生变化。如果在某个磁场值 B_0 下正好前 λ 个朗道能级被填满，通过数模式数目会发现这些电子模式在无磁场的情况下占据的最高能级的能量正好就是 $(\lambda + 1)\hbar\omega_0$ ，计算系统总能量，发现和磁场为 B_0 时的系统总能量正好一样。现在略微降低磁场，朗道能级间距缩短，从而一些电子在平衡时需要占据 $\lambda + 1$ 能级，因此加入磁场后系统能量升高了。容易看出当第 $\lambda + 1$ 能级刚好占据了一半的时候系统能量比加入磁场前增大得最多。进一步增大磁场，会发现一部分在自由情况下能量高于 $\lambda + 1$ 能级的能量的电子也需要填充在 $\lambda + 1$ 能级上，不过这部分电子的数目少于自由情况下能量低于 $\lambda + 1$ 的电子的数目，因此此时加入磁场后系统能量仍然上升，但是上升的量和第 $\lambda + 1$ 能级刚好占据了一半的情况比起来不那么大了。

总之，如此周而复始，只要加入磁场，系统的平衡态能量就不会低于自由时的能量，且每当有某个磁场使得对应的某个朗道能级满足（ ϵ_F 是加入磁场前的费米能级）

$$\left(n + \frac{1}{2}\right) \hbar\omega_0 = \epsilon_F \quad (5.37)$$

时系统能量相对于加入磁场前的增量达到峰值，而

$$n\hbar\omega_0 = \epsilon_F \quad (5.38)$$

或者说 nD 是总电子数 N_e 时系统能量和加入磁场前相比没有什么变化。代入 ω_0 会发现

$$n \frac{Se}{\pi\hbar} B = N_e,$$

因此实际上以上系统能量的变化相对于 $1/B$ 是周期性的，周期为

$$\Delta\left(\frac{1}{B}\right) = \frac{Se}{\pi\hbar N_e} = \frac{2\pi e}{\hbar S_F}, \quad (5.39)$$

其中

$$S_F = 2\pi^2 \frac{N_e}{S} \quad (5.40)$$

是二维电子气费米圆的面积。

很多东西都依赖于系统总能量，如磁矩定义为 $-\partial E/\partial B$ ，照以上推导，它也应该关于 $1/B$ 以周期(5.39)变化。这称为de Hass-van Alphen效应。比热等物理量也有这个关系。

三维电子气在强磁场下的行为

一个 z 方向可以充分延展的系统，即一个三维电子气，在强磁场下的行为还要复杂一些。在与磁场垂直的 $k_z = \text{const}$ 的平面内轨道按照(5.29)是分立的，而 z 方向上的 k_z 是连续的。 xy 平面上没有良定义的Bloch波矢，但是还是可以根据二维情况中的说法计算出准经典的 k_x 和 k_y 。于是我们可以将三维电子气中的各个模式形象地在倒空间中展示为一系列同心圆柱面，一个模式的 \mathbf{k} 总是固定在某个同心圆柱面上，其 k_z 不变，而 k_x 和 k_y 随着时间变化连续旋转，没有确定的值。好量子数有两个，一个是区分不同同心圆柱面的 n ，一个是 z 方向的 k_z 。

一维电子气的轨道态密度是

$$g(\epsilon_{k_z}) = \frac{L\sqrt{2m}}{2\pi\hbar} \epsilon_{k_z}^{-1/2}, \quad (5.41)$$

由于强磁场下可以忽略自旋导致的能级分裂而有两种自旋这件事已经充分体现在了 D 当中，所有子带的态密度（同时包括自旋和轨道）的总和就是

$$g(\epsilon) = D \frac{L\sqrt{2m}}{2\pi\hbar} \sum_n \left(\epsilon - \left(n + \frac{1}{2} \right) \hbar\omega_0 \right)^{-1/2}. \quad (5.42)$$

在三维电子气中我们也可以观察到de Hass-van Alphen效应，虽然不是严格成立的。当费米球和某个圆柱面相切时系统能量增量最大。

以上讨论的都是完全没有周期势场的电子气在强磁场下的行为。在晶格周期势场相对于磁场很强时，电子是普通的能带电子，有明确定义的 k_x 和 k_y ，而

5.4.2 能带电子的线性响应和格林函数

背景知识 5.2: 线性响应理论中的格林函数

凝聚态理论中计算最多的是线性响应理论中的格林函数。所谓线性响应理论是说, 设时间平移不变的系统受到外加哈密顿量

$$H_{\text{ext}} = h(t)A \quad (5.43)$$

扰动, 只考虑系统的密度矩阵正比于 $h(t)$ 的变化, 则算符 B 的期望值由

$$y(t) = \int dt' G_{AB}^{\text{ret}}(t, t') h(t'), \quad y(t) = \langle B \rangle(t) - \langle B \rangle_0 \quad (5.44)$$

给出, 其中

$$G_{AB}^{\text{ret}}(t - t') = -i\Theta(t - t') \langle [B(t), A(t')] \rangle_0 \quad (5.45)$$

是有限温度推迟格林函数, 其中 $\langle \cdot \rangle$ 表示考虑了 H_{ext} 的期望而 $\langle \cdot \rangle_0$ 表示没有考虑 H_{ext} 的期望。不过因为我们在做实际计算时通常都使用没有考虑 H_{ext} 的哈密顿量, 以下会略去下标0。

推迟格林函数不易直接计算, 因为它不是编时格林函数, 因此不能直接使用Wick定理。通常会使用松原格林函数辅助计算。松原格林函数为

$$G_{AB}(\tau) = -\mathcal{T} \langle A(\tau)B(0) \rangle, \quad G_{AB}(i\omega_n) = \int_0^\beta d\tau e^{i\omega_n \tau} G_{AB}(\tau), \quad (5.46)$$

从频域切换回虚时间的方法是

$$G_{AB}(\tau) = \frac{1}{\beta} \sum_{\omega_n} G_{AB}(i\omega_n) e^{-i\omega_n \tau}. \quad (5.47)$$

对自由理论, 有

$$G^0(i\omega_n) = - \int_0^\beta d\tau e^{i\omega_n \tau} \mathcal{T} \langle c(\tau) c^\dagger(0) \rangle = \frac{1}{i\omega_n - \epsilon}. \quad (5.48)$$

做费曼图计算时应当注意传播子是 $-G_{AB}^0(i\omega_n)$ 。从松原格林函数延拓到实时间推迟格林函数的方法是将 $i\omega_n$ 替换成 $\omega + i0^+$ 。

形式上, 我们可以对 τ 做Wick转动 $\tau = it$, 这样的理论和高能物理中常用的零温场论形式上完全一样, 后者的格林函数定义为

$$iG_{AB}(t) = \mathcal{T} \langle A(t)B(0) \rangle, \quad G_{AB}(\omega) = \int dt e^{i\omega t} G_{AB}(t), \quad (5.49)$$

对自由理论, 有

$$G^0(\omega) = \frac{1}{\omega - \epsilon}. \quad (5.50)$$

近独立电子气的格林函数和它的谱函数是可以比较容易地算出的。如果基态为真空, 松原格林函数为

$$G(\mathbf{k}, i\omega_n) = - \int_0^\beta d\tau e^{i\omega_n \tau} \mathcal{T} \langle c_{\mathbf{k}}(\tau) c_{\mathbf{k}}^\dagger(0) \rangle = \frac{1}{i\omega_n - \epsilon_{\mathbf{k}}}, \quad (5.51)$$

推迟格林函数为

$$G^{\text{ret}}(\mathbf{k}, \omega) = \frac{1}{\omega - \epsilon_{\mathbf{k}} + i0^+}, \quad (5.52)$$

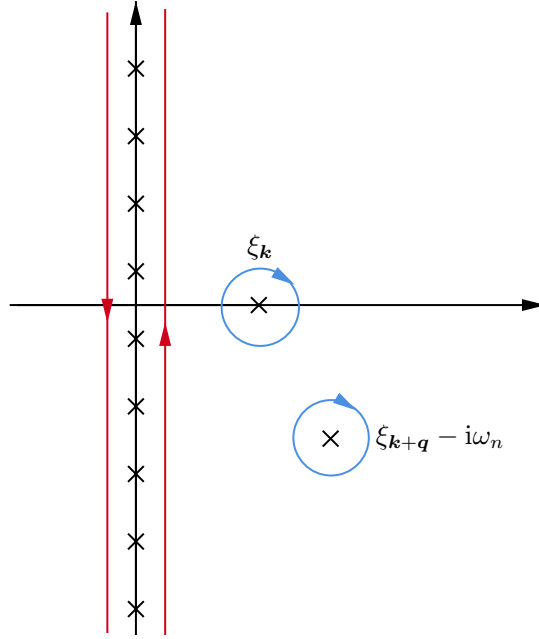


图 5.2: (5.58)中被积函数在复平面上的解析结构

谱函数为

$$A(\mathbf{k}, \omega) = \delta(\omega - \epsilon_{\mathbf{k}}). \quad (5.53)$$

这些都和单电子计算出来的结果完全一样。如果基态中填充了一个费米球，化学势为 μ ，那么就有

$$G(\mathbf{k}, i\omega_n) = \frac{1}{i\omega_n - \xi_{\mathbf{k}}}. \quad (5.54)$$

可以这么考虑这个表达式的含义：如果以费米面上方的粒子和费米面下方的空穴为表象，则粒子能谱为 $\xi_{\mathbf{k}}$ ，空穴能谱为 $-\xi_{\mathbf{k}}$ 。对粒子，格林函数就是(5.54)，对空穴，空穴的格林函数是 $1/(i\omega_n - (-\xi_{\mathbf{k}}))$ ，现在切换回费米面上下的模式都是粒子的表象，则 ω_n 作为四维动量的一部分变号，由于粒子、空穴算符都是费米子算符，整个格林函数也变号，则得到 $-1/(-i\omega_n - (-\xi_{\mathbf{k}}))$ ，正好就是(5.54)。

格林函数实际上就是一个概率系统（无论是经典概率还是量子的密度矩阵）中重要的量的各阶矩。系统性质足够良好时实际上完全可以通过格林函数来完全描述一个系统，而不需要知道哈密顿量或者拉格朗日量。这就好比观察一个赌场：要知道赌场中最后哪些人能够赚钱，你只需要让一个人带着一大笔钱去赌博，看最后大家手头的钱怎么变化就可以了；赌场规则的文本会决定大家手头的钱怎么变化，不过要直观地看到赌场的“行为”，盯着规则看可能并没有什么帮助。

密度-密度格林函数

电子的密度-密度格林函数在线性响应理论中的作用是指出外加电势下电子气的密度会如何变化。外加电势导致如下的哈密顿量：

$$H_{\text{ext}} = \int d^3\mathbf{r} n(\mathbf{r}, t) V_{\text{ext}}(\mathbf{r}, t),$$

而由于

$$\int d^3\mathbf{r} n_0(\mathbf{r}, t) V_{\text{ext}}(\mathbf{r}, t)$$

虽然随着时间发生变化, 但是和其它任何算符都对易, 它不能够产生任何动力学效果, 所以也可以认为外加电势引入的哈密顿量是

$$H_{\text{ext}} = \int d^3\mathbf{r} \delta n(\mathbf{r}, t) V_{\text{ext}}(\mathbf{r}, t),$$

则它对电子数密度的影响可以使用(各向同性不含时体系的)推迟格林函数

$$G_{nn}^{\text{ret}}(\mathbf{r} - \mathbf{r}', t - t') = -i\Theta(t - t') \langle [\delta n(\mathbf{r}, t), \delta n(\mathbf{r}', t')] \rangle$$

描述, 而响应就是¹

$$\langle \delta n(\mathbf{r}, t) \rangle = \int dt' G_{nn}^{\text{ret}}(\mathbf{r} - \mathbf{r}', t - t') V_{\text{ext}}(\mathbf{r}', t'), \quad (5.55)$$

在频域中就是

$$\langle \delta n(\mathbf{k}, \omega) \rangle = G_{nn}^{\text{ret}}(\mathbf{k}, \omega) V_{\text{ext}}(\mathbf{k}, \omega). \quad (5.56)$$

于是我们计算密度-密度松原格林函数即可。这里我们用 $G_{nn}^{0, \text{ret}}(\mathbf{k}, \omega)$ 作为 $G_{\delta n \delta n}^{0, \text{ret}}(\mathbf{k}, \omega)$ 的简写依照定义, 近独立电子的密度-密度松原格林函数为

$$\begin{aligned} G_{nn}^0(\mathbf{r} - \mathbf{r}', \tau - \tau') &= -T_\tau \langle \delta n(\mathbf{r}, \tau) \delta n(\mathbf{r}', \tau') \rangle \\ &= -T_\tau \langle n(\mathbf{r}, \tau) n(\mathbf{r}', \tau') \rangle + \langle n(\mathbf{r}, \tau) \rangle \langle n(\mathbf{r}', \tau') \rangle \\ &= \sum_{\sigma, \sigma'} T_\tau \langle \psi_{\sigma'}(\mathbf{r}', \tau') \psi_\sigma^\dagger(\mathbf{r}, \tau) \rangle T_\tau \langle \psi_\sigma(\mathbf{r}, \tau) \psi_{\sigma'}^\dagger(\mathbf{r}', \tau') \rangle, \end{aligned}$$

其中最后一个等号用到了Wick定理。由于自旋守恒性, 只有 $\sigma = \sigma'$ 的项才非零, 我们得到

$$G_{nn}^0(\mathbf{r} - \mathbf{r}', \tau - \tau') = \sum_{\sigma} G_{\sigma}(\mathbf{r}' - \mathbf{r}, \tau' - \tau) G_{\sigma}(\mathbf{r} - \mathbf{r}', \tau - \tau'). \quad (5.57)$$

上式具有卷积的形式, 切换到频谱, 就是

$$G_{nn}^0(\mathbf{q}, i\omega_n) = \sum_{\sigma} \frac{1}{V} \sum_{\mathbf{k}} \frac{1}{\beta} \sum_{\nu_n} G_{\sigma}(\mathbf{q}, i\nu_n) G_{\sigma}(\mathbf{k} + \mathbf{q}, i(\omega_n + \nu_n)),$$

其中

$$\nu_n = \frac{\pi(2n+1)}{\beta}, \quad \omega_n = \frac{2\pi n}{\beta},$$

前者是费米子的松原频率, 后者则是玻色子的, 因为电子配对形成玻色子。近独立电子的谱函数为

$$A(\mathbf{k}, \omega) = \delta(\omega - \xi_{\mathbf{k}}),$$

格林函数为

$$G_{\sigma}(\mathbf{r}, i\omega_n) = \frac{1}{i\omega_n - \xi_{\mathbf{k}}}.$$

于是, 有

$$G_{nn}^0(\mathbf{q}, i\omega_n) = \frac{2}{V} \sum_{\mathbf{k}} \frac{1}{\beta} \sum_{\nu_n} \frac{1}{i\nu_n - \xi_{\mathbf{k}}} \frac{1}{i\omega_n + i\nu_n - \xi_{\mathbf{k}+\mathbf{q}}}.$$

¹我们没有将 V_{ext} 写进 H 中而是把它看成了外界扰动, 这样电子气本身的哈密顿量仍然是空间平移不变的。

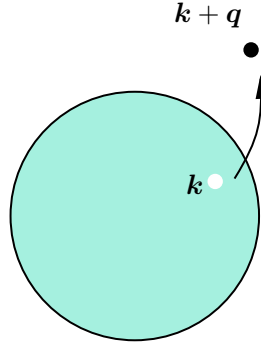


图 5.3: 电子-空穴对

要计算上式，常用的一个技巧是将求和转化为围道积分。注意到 $\{i\nu_n\}$ 正是 $f(z)$ 的全部极点，而且这些极点都是一阶极点。容易看出这些极点的留数都是 $-1/\beta$ ，于是对一个在虚轴上解析的函数 $F(z)$ ，有

$$\frac{1}{\beta} \sum_{\nu_n} F(i\nu_n) = - \oint \frac{dz}{2\pi i} f(z) F(z),$$

于是设围绕虚轴的环路为 C ，它实际上就是从 $-i\infty+0^+$ 到 $i\infty+0^+$ ，再从 $i\infty-0^+$ 到 $-i\infty-0^+$ ，有

$$\begin{aligned} G_{nn}^0(\mathbf{q}, i\omega_n) &= -\frac{2}{V} \sum_{\mathbf{k}} \oint_C \frac{dz}{2\pi i} f(z) \frac{1}{z - \xi_{\mathbf{k}}} \frac{1}{z + i\omega_n - \xi_{\mathbf{k}+\mathbf{q}}} \\ &= \frac{2}{V} \sum_{\mathbf{k}} \left(\frac{f(\xi_{\mathbf{k}})}{\xi_{\mathbf{k}} + i\omega_n - \xi_{\mathbf{k}+\mathbf{q}}} + \frac{f(\xi_{\mathbf{k}+\mathbf{q}} - i\omega_n)}{\xi_{\mathbf{k}+\mathbf{q}} - i\omega_n - \xi_{\mathbf{k}}} \right). \end{aligned} \quad (5.58)$$

负号消失了是因为绕极点 $z = \xi_{\mathbf{k}}$ 和 $z = \xi_{\mathbf{k}+\mathbf{q}} - i\omega_n$ 是顺时针而不是留数定理通常是的逆时针（见图5.2）。由 $\{\omega_n\}$ 的定义， $f(\xi_{\mathbf{k}+\mathbf{q}} - i\omega_n)$ 就是 $f(\xi_{\mathbf{k}+\mathbf{q}})$ ，于是得到松原格林函数

$$G_{nn}^0(\mathbf{q}, i\omega_n) = \frac{2}{V} \sum_{\mathbf{k}} \frac{f(\xi_{\mathbf{k}}) - f(\xi_{\mathbf{k}+\mathbf{q}})}{i\omega_n + \xi_{\mathbf{k}} - \xi_{\mathbf{k}+\mathbf{q}}}. \quad (5.59)$$

做解析延拓就得到推迟格林函数

$$G_{nn}^{\text{ret}, 0}(\mathbf{q}, \omega) = \frac{2}{V} \sum_{\mathbf{k}} \frac{f(\xi_{\mathbf{k}}) - f(\xi_{\mathbf{k}+\mathbf{q}})}{\omega + \xi_{\mathbf{k}} - \xi_{\mathbf{k}+\mathbf{q}} + i0^+}. \quad (5.60)$$

这就给出了能谱为 $\xi_{\mathbf{k}}$ 的（有自旋的）近独立电子气对外加势场的响应，傅里叶变换回到实空间即可。

我们马上可以注意到在

$$\omega = \xi_{\mathbf{k}+\mathbf{q}} - \xi_{\mathbf{k}} \quad (5.61)$$

处密度-密度推迟格林函数发散，表明存在频率为上式的集体模式。考虑密度-密度格林函数的意义，这种集体模式应该是： \mathbf{k} 点的电子数密度下降而 $\mathbf{k} + \mathbf{q}$ 的电子数密度上升，或者直观地说，一个动量为 \mathbf{k} 的电子被激发到 $\mathbf{k} + \mathbf{q}$ 处。因此这个集体模式实际上是电子-空穴对，它产生后系统能量的变化是 $\xi_{\mathbf{k}+\mathbf{q}} - \xi_{\mathbf{k}}$ ，正好是(5.61)。

一个电子-空穴对需要两个动量来标记：一个是 \mathbf{q} ，为整体的动量，一个是 \mathbf{k} ，指出空穴的动量。(5.60)左边只有一个动量 \mathbf{q} ，因此电子-空穴对的能谱在(5.60)的解析结构上留下的不是一个极点，而是一个割线。同样，如果绘制能谱， ω - \mathbf{q} 关系不是一条曲线，而是一块区域。

对自由电子，有

$$\xi_{\mathbf{k}} = \frac{\mathbf{k}^2}{2m} - \mu,$$

我们可以做一些更加具体的计算。此时电子-空穴对的能量为

$$\omega = \frac{2\mathbf{k} \cdot \mathbf{q} + \mathbf{q}^2}{2m}, \quad (5.62)$$

在 \mathbf{q} 给定的情况下， ω 最大值和最小值分别为

$$\omega_{q,\max} = \frac{2k_F q + q^2}{2m}, \quad \omega_{q,\min} = \max\left(\frac{-2k_F q + q^2}{2m}, 0\right). \quad (5.63)$$

在 $q < 2k_F$ 时 ω_{\min} 是0，相当于是说存在这样的 \mathbf{k} 使得 $\xi_{\mathbf{k}+\mathbf{q}} = \xi_{\mathbf{k}}$ ，等于是说存在一个与费米球同心，并且在费米球内的球，使得矢量 \mathbf{q} 是这个球的弦，这又等价于，费米球平移 \mathbf{q} 后和原来的费米球重合。反之，如果 $q > 2k_F$ ，费米球平移后和原来的费米球不重合。

对费米海并非球形的情况，从(5.60)的形式中我们能够读取出一个可能的密度波长程序。我们考虑一种特殊情况：如果存在一片连续的费米面区域，它们平移某个矢量 \mathbf{Q} 后得到了另一片连续的费米面区域，我们就说存在费米面嵌套 (nesting)， \mathbf{Q} 为嵌套矢量。零频、在费米面嵌套矢量附近处的密度-密度松原格林函数为

$$G_{nn}^{\text{ret}, 0}(\mathbf{Q}, 0) = \frac{2}{v} \sum_{\mathbf{k}} \frac{0}{0} = ???$$

简单的尝试无法计算出推迟格林函数，这暗示 $\mathbf{q} = \mathbf{Q}, \omega = 0$ 处很可能存在一定奇异性。

这意味着，存在费米面嵌套的系统

第六章 金属中的库仑相互作用

6.1 Hartree-Fock近似

我们先前处理过电子间相互作用，但是使用的是玻尔兹曼方程，有浓重的半经典色彩，从本节开始我们在场论框架下比较系统地分析相互作用对能带的修正。

电子系统哈密顿量中的三项——动能、晶格势能、电子间库伦排斥——都会对电子结构产生影响。首先考虑动能和库伦排斥能之间的关系。我们定义平均每个电子占据的球状体积的半径（所谓Wigner-Seitz半径）

$$\frac{4\pi}{3}r_s^3 = \frac{1}{\rho}, \quad (6.1)$$

它和电子数密度一一对应。不考虑晶格存在的话，这是电子气中唯一一个距离尺度。平均每个电子的库伦排斥能量级为

$$\text{P.E.} \sim \frac{e^2}{r_s},$$

每个电子的动能的量级，根据德布罗意关系，是

$$\text{K.E.} \sim \frac{p^2}{2m} \sim \frac{(1/r_s)^2}{m},$$

从而

$$\frac{\text{P.E.}}{\text{K.E.}} = me^2 r_s. \quad (6.2)$$

因此 r_s 很大，或者说电子数密度很小时，电子间库伦排斥占据主导地位，而当 r_s 很小，或者说电子数密度很大时，电子动能占据主导地位。

这一说法看起来有些匪夷所思，因为如果将电子想象成钢球，那么高电子密度时电子根本无法移动，而低电子密度时电子可以自由移动，似乎低电子密度时才是动能占据主导地位。然而库伦排斥恰恰不是钢球势，而是与之非常不同的一种长程势。电子密度低时，两个相隔一定距离的电子仍然有明显的散射，从而不能保证电子真的可以自由移动；电子密度高时，一个电子同时受到四面八方的电子的排斥，结果是运动起来反而不会受到多少阻碍。因此，从硬球气体获得的物理图像不能用于估计电子动能还是库伦排斥能占据主导地位。

虽然这么说，硬球气体的图像并非毫无用处，因为它可以用于形象地展示哪一类费曼图的贡献最明显。电子数密度较高的系统中主要的费曼图是所谓的项链图（见第6.2.2节），直观地看，就是由于电子之间间距很小，一个电子发射的光子很快会被相邻的电子吸收，再发射，再吸收……相反，电子数密度很低的系统中主要的费曼图是梯形图，直观地看，就是因为粒子数少，一阶微扰是两个粒子发生一次散射，二阶微扰是两个粒子发生一次散射后再发生一次散射……

一个库伦排斥占据主导地位的库伦排斥电子气中显然不会有良定义的“单电子”。这种系统通常称为**强关联系统**，或者也可以说这样的系统有强关联效应。

然后我们再来分析晶格的影响。晶格常数和晶格势场的大小都是重要的因素。如果晶格势能足够将电子定域在原子核周围，那么电子动能相比库伦排斥势很小，体系很可能是强关联的。我们将在第十章中讨论一些这种系统。另一方面，如果电子数密度足够高，那么晶格影响是可以忽略的，因为被原子核束缚的电子会马上屏蔽晶格势能，于是其它电子可以认为只有动能没有晶格势能。

6.1.1 费曼规则

我们先给出凝聚态介质中的电子的场论。我们在虚时间场论和倒空间（即 (ω, \mathbf{k}) ，其中 \mathbf{k} 是简约波矢）下工作，则费曼规则为：

- 传播子为

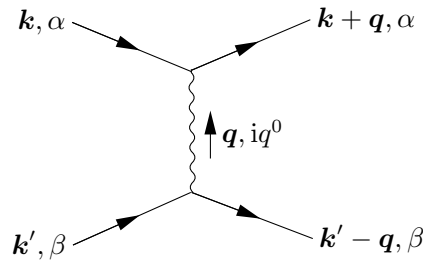
$$\alpha \xrightarrow{\mathbf{k}} \beta = -\frac{\delta_{\alpha\beta}}{i\omega_n - \xi_{\mathbf{k}}}, \quad (6.3)$$

其中对近自由电子系统，有

$$\xi_{\mathbf{k}} = \frac{\mathbf{k}^2}{2m} - \mu \quad (6.4)$$

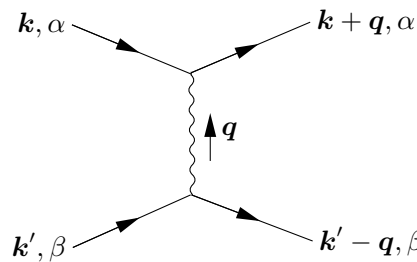
对一般的能带电子就是能带。

- 相互作用使用



$$= -\frac{1}{V} \frac{4\pi e^2}{|\mathbf{q}|^2}, \quad (6.5)$$

表示，其中 \mathbf{q} 可以随意取值（见第4.5.1节）。需要对 \mathbf{q} 做求和，并注意由于库仑相互作用被假定为瞬时发生的，光子频率 iq^0 可以任意选取，其具体的值通过能量守恒条件得到（注意到，在时域上，库仑相互作用正比于 $\delta(t - t')$ ，从而在频域上，它没有任何 ω 依赖）。光子由于已经被积掉了，不贡献传播子。可以将(6.5)整体看成一个给出库仑相互作用势能的顶角，不过更常见的做法是将光子线称为**相互作用线**，将相互作用线和电子线的连接点称为顶角。两个顶角被一条相互作用线连接则贡献一个库仑相互作用势能和一个 -1 因子。通常系统均足够大，此时有



$$= -\frac{1}{\beta} \sum_{\omega_n} \int \frac{d^3 \mathbf{q}}{(2\pi)^3} \frac{4\pi e^2}{|\mathbf{q}|^2}. \quad (6.6)$$

- 闭合电子回线需要乘上 -1 。

不要忘记按照松原格林函数的定义，将需要计算的所有图求和后需要加上负号才能够得到松原格林函数。

我们仍然使用波浪线表示库仑相互作用，表示这是光子介导的。在之后的模型中，相互作用势会被多次修正，得到一个形式复杂的，(1.10)形式的理论，那时将使用虚线表示电子间排斥。

6.1.2 正规自能计算和Hartree-Fock近似

背景知识 6.1: 单粒子格林函数的自能修正

做自能计算时我们有

$$\frac{1}{\text{格林函数}} = \frac{1}{\text{自由格林函数}} - \text{1PI图},$$

即

$$\frac{1}{G} = \frac{1}{G^0} + \text{1PI图},$$

或者说

$$\frac{1}{G} = \frac{1}{G^0} - \Sigma, \quad -\Sigma = \text{1PI图}, \quad (6.7)$$

即单电子（或者别的什么粒子）带相互作用的格林函数的极点展示的能量相比自由格林函数要减去1PI图或者说加上 Σ ，这就是所谓自能修正。

对零温场论同样有自能修正，为

$$\frac{1}{iG} = \frac{1}{iG^0} + i\Sigma, \quad -i\Sigma = \text{1PI图}. \quad (6.8)$$

1PI图的计算常常采用骨架图的做法，即画出1PI图中最低阶的那些图，然后将其中的粒子线和相互作用顶角替换成相互作用修饰后的，这样可以覆盖全部1PI图。

背景知识 6.2: 平均场近似

一种常用的近似是，在1PI图中只将我们想要计算的那部分单粒子格林函数对应的粒子线替换成相互作用粒子线，于是得到一个形如“我们想要计算的单粒子格林函数=一系列包含我们想要计算的单粒子格林函数的费曼图的和”这样的方法，即可解出我们想要计算的单粒子格林函数。形式上，这等价于将相互作用哈密顿量做类似于这样的分解：

$$c^\dagger c^\dagger c c \approx \langle c^\dagger c \rangle c^\dagger c,$$

其中 $\langle c^\dagger c \rangle$ 是我们要计算的格林函数（有时称为**平均场序参量**，见背景知识8.2），然后用这个“单电子哈密顿量”形式地求解 $\langle c^\dagger c \rangle$ ，就得到了一个自治方程，解之就得到想要的格林函数。形式上看，这相当于将相互作用看成了系统中其它粒子形成的大小为 $\langle c^\dagger c \rangle$ 的一个外场，所以也可以叫做平均场近似。

以上介绍的是需要求解自洽方程的平均场近似，当然作为最粗糙的估计，这个自洽方程的右边可以代入自由情况下我们要计算的格林函数，这就是非自洽的平均场方法，通常效果不好。

平均场近似可以用于分析各种各样的问题。它可以用于处理某种粒子的相互作用修饰（正如我们马上要做的那样），可以用于分析对称性自发破缺（见背景知识8.2），还可以用于分析部分子构造产生的准粒子的动力学。所有这些应用中“平均场”一词都指的是同一个意思：从费曼图的角度，重求和一类图形而丢弃其它图形，从哈密顿量的角度，将一个相互作用哈密顿量中的几个算符的乘积替换为某种期望值（“平均场”）。平均场近似总是低估系统的量子涨落，但是如果做得适当，仍然是一种重要的工具。

计算单电子格林函数实际上就是要计算自能。本节考虑仅仅保留一阶自能图的近似，这个近似通常称为**Hartree-Fock近似**。Hartree-Fock近似又可以分成两种，一种是**自洽Hartree-Fock近似**，即将一阶自能图中的所有电子线修正为相互作用电子线，不对相互作用线和顶角做修正，这是库仑相互作用以动量表象下格林函数为平均场序参量的平均场理论。另一种是**非自洽Hartree-Fock近似**，一阶自能图中的所有电子线都是无相互作用电子线。非自洽Hartree-Fock近似基本上是不太准确的，只能做定性分析，而自洽Hartree-Fock近似在化学上已经能够做一些定量预言了（虽然还是很不准确，因为被忽略的电子关联能的量级和键能等相同，因此需要更高阶的修正）。

无论是哪一种，Hartree-Fock近似实际上都是将一个相互作用理论转化为一个单粒子理论，因为电子-电子格林函数中存在无有限大小的虚部的极点，因此系统的低能自由度包括和自由电子类似的费米型激发。虽然我们这里是从Hartree-Fock推导出这个结论的，但是这个结论的适用范围显然很广，因为只要系统中电子-电子格林函数的自能能够微扰计算，系统的低能自由度都是一些费米型激发。第6.3节中仔细分析了这类系统的性质。

6.1.3 零温下的自洽Hartree-Fock方程组

Hartree-Fock近似为数值计算系统中的电子型激发。实际上这种激发就是第6.3节中的费米型准粒子，是经过相互作用修饰的能带电子。根据Bloch定理，我们将这些经过相互作用修饰的能带电子用 $n\mathbf{k}$ 标记。

我们遵从自洽Hartree-Fock近似的要求，考虑了两个自能图，并且将自能图中的电子线取为相互作用修饰之后的。根据零温场论的费曼规则可以写出

$$\begin{aligned}
 -i\Sigma(\mathbf{r}'\sigma', \mathbf{r}\sigma) &= \text{Diagram 1} + \text{Diagram 2} \\
 &= \delta(\mathbf{r} - \mathbf{r}')\delta_{\sigma\sigma'} \sum_{n,\mathbf{k}} \sum_{\sigma''} \int d^3\mathbf{r}'' \theta(\mu - \epsilon_{n\mathbf{k}}) |\varphi_{n\mathbf{k}}(\mathbf{r}''\sigma'')|^2 \frac{-ie^2}{|\mathbf{r} - \mathbf{r}''|} \\
 &\quad - \delta_{\sigma\sigma'} \frac{-ie^2}{|\mathbf{r} - \mathbf{r}'|} \sum_{n,\mathbf{k}} \theta(\mu - \epsilon_{n\mathbf{k}}) \varphi_{n\mathbf{k}}(\mathbf{r}'\sigma') \varphi_{n\mathbf{k}}^*(\mathbf{r}\sigma),
 \end{aligned}$$

这里的负号来自格林函数中湮灭算符在前产生算符在后，和粒子数密度算符正好差一个负号。所谓单粒子波函数或者能量云云都是在背景知识4.1的意义下说的。代入(4.4)，有

$$\begin{aligned}\epsilon_{n\mathbf{k}}\varphi_{n\mathbf{k}}(\mathbf{r}\sigma) = & -\frac{\nabla^2}{2m}\varphi_{n\mathbf{k}}(\mathbf{r}\sigma) + V_{\text{ion}}(\mathbf{r})\varphi_{n\mathbf{k}}(\mathbf{r}\sigma) \\ & + \sum_{n',\mathbf{k}'} \sum_{\sigma''} \int d^3\mathbf{r}'' \Theta(\mu - \epsilon_{n'\mathbf{k}'}) |\varphi_{n'\mathbf{k}'}(\mathbf{r}''\sigma'')|^2 \frac{e^2}{|\mathbf{r} - \mathbf{r}''|} \varphi_{n\mathbf{k}}(\mathbf{r}\sigma) \\ & - \int d^3\mathbf{r}' \sum_{n',\mathbf{k}'} \Theta(\mu - \epsilon_{n'\mathbf{k}'}) \varphi_{n'\mathbf{k}'}(\mathbf{r}\sigma) \varphi_{n'\mathbf{k}'}^*(\mathbf{r}'\sigma) \frac{e^2}{|\mathbf{r} - \mathbf{r}'|} \varphi_{n\mathbf{k}}(\mathbf{r}'\sigma).\end{aligned}\quad (6.9)$$

求解出上式（上式，或者其离散化版本，称为Rootaan方程）就求解出了自洽Hartree-Fock近似下的费米型激发的能谱。(6.9)的推导是非常一般的，因此可以将 $V_{\text{ion}}(\mathbf{r})$ 用赝势代替，或是引入自旋-轨道耦合等等。方程右边的第三、四项来自电子-电子库伦相互作用，第三项就是将电子数密度当成一个外场写出的库伦势，称为经典库伦能或者Hartree项，而第四项没有特别明确的意义，通常称为交换能或者Fock项，我们会发现只有自旋相同的电子之间才有交换能，而且交换能是负的，即电子自旋平行会让系统能量下降。可以看到Hartree项和Fock项显然分别来自两种自能图，前者有时也称为蝌蚪图。

在确定没有自旋-轨道耦合时，只要求解轨道部分的波函数 $\varphi_{n\mathbf{k}}(\mathbf{r})$ ，且状态 $|n\mathbf{k}\uparrow\rangle$ 和 $|n\mathbf{k}\downarrow\rangle$ 要么都有电子，要么都没有电子，从而(6.9)化简为

$$\begin{aligned}\epsilon_{n\mathbf{k}}\varphi_{n\mathbf{k}}(\mathbf{r}) = & -\frac{\nabla^2}{2m}\varphi_{n\mathbf{k}}(\mathbf{r}) + V_{\text{ion}}(\mathbf{r})\varphi_{n\mathbf{k}}(\mathbf{r}) \\ & + 2 \sum_{n',\mathbf{k}'} \int d^3\mathbf{r}'' \Theta(\mu - \epsilon_{n'\mathbf{k}'}) |\varphi_{n'\mathbf{k}'}(\mathbf{r}'')|^2 \frac{e^2}{|\mathbf{r} - \mathbf{r}''|} \varphi_{n\mathbf{k}}(\mathbf{r}) \\ & - \int d^3\mathbf{r}' \sum_{n',\mathbf{k}'} \Theta(\mu - \epsilon_{n'\mathbf{k}'}) \varphi_{n'\mathbf{k}'}(\mathbf{r}') \varphi_{n'\mathbf{k}'}^*(\mathbf{r}') \frac{e^2}{|\mathbf{r} - \mathbf{r}'|} \varphi_{n\mathbf{k}}(\mathbf{r}').\end{aligned}\quad (6.10)$$

这里应当注意，相互作用电子气的电子-电子格林函数中的化学势和关闭相互作用时计算出来的化学势通常是不一样的（前者实际上是费米液体中的化学势，见第6.3节），但是在做自洽Hartree-Fock计算时其实根本不需要显式使用化学势，因为只需要根据本轮计算得到的 $\epsilon_{n\mathbf{k}}$ 自下而上填充 N_e 个电子即可。

实际上也可以从另一个角度获得(6.9)。我们可以写下一个自由电子气的波函数，计算库伦相互作用在这个波函数下的期望值，并且将这个期望值当成系统能量（可见Hartree-Fock近似确实是一种平均场近似方法），同样能够得到(6.9)。这实际上是预期之中的，因为Hartree-Fock近似只是修正了单电子传播子，从而它可以认为是一个单电子的模型。

系统实际的能量减去通过以上方法计算出来的Hartree-Fock近似下的能量通常称为关联能，因为这部分能量的计算无法通过将基态拟设为无相互作用波函数计算得到。“电子关联”的概念其实带有一些模糊性。如果我们将电子关联视为“一个电子的引入会影响其它电子”，那么Hartree-Fock近似当然也考虑了一些电子关联效应：例如，自能修正就和所有电子均有关，因为它依赖于 k_F （见(6.14)）。然而，虽然上述Hartree-Fock考虑到了一定“电子关联”，但由于Hartree-Fock近似下，基态波函数仍然是近独立电子气的基态波函数的形式（虽然其中的一些参数和全体电子均有关），它无法捕捉诸如Wick定理失效这样超越近独立电子气图景的现象。Hartree-Fock近似中，所有物理量仍然可以使用单电子格林函数计算，虽然单电子格林函数依赖于全体电子，然而实际的库伦排斥电子气中则不是这样。从这个意义上讲，实际的库伦排斥电子气中的电子关联不仅仅包括Hartree-Fock近似能够覆盖的“单电子状态取决于其它电子”，也包括“并非所有物理量都能够使用单电子格林函数

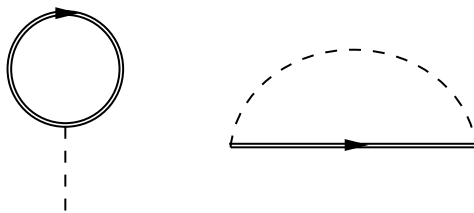


图 6.1: 一般的电子-电子散射也有Hartree项和Fock项

完成此积分就得到

$$\Sigma_{\mathbf{k}\alpha} = -\frac{e^2 k_F}{\pi} \left(1 + \frac{1-y^2}{2y} \ln \left| \frac{1+y}{1-y} \right| \right), \quad y = \frac{k}{k_F}. \quad (6.14)$$

我们马上会发现一个不妙的事实，就是 $k \rightarrow k_F$ 时电子等效群速度

$$v_k^* = v_k + \frac{\partial \Sigma_{\mathbf{k}\alpha}}{\partial k} \quad (6.15)$$

对数发散。实验中当然没有观察到这种现象——别忘了，对金属，单纯的能带理论就足够解释大部分现象了，而能带理论中当然没有什么发散的费米速度。在费米面处的群速度发散还意味着相互作用修正后的态密度在费米面处严格为零（回顾态密度的定义，它大体上是群速度的倒数），而这当然不是事实。因此我们发现，接近费米面处，非自洽Hartree-Fock近似失效。对数发散的出现显然来自对 $1/q^2$ 的积分，因此，这一发散实际上来自我们使用没有经过屏蔽修正的裸的库伦排斥来做Hartree-Fock近似这一事实。我们将在后面讨论RPA近似，来得到经过屏蔽的短程库伦相互作用，它将消除这一发散。

Hartree-Fock近似中没有出现有限大小的自能虚部，因为无论是Hartree图还是Fock图，切成两半都不能够形成完整的代表一个动量的电子激发出别的自由度（在库仑相互作用电子气系统中，就是激发出电子-空穴对）而衰变到更低能量的过程的费曼图。

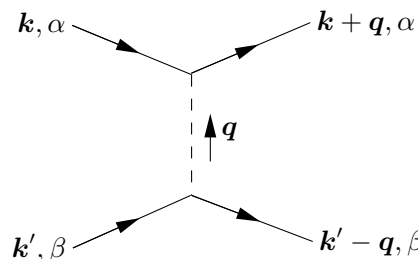
6.1.5 基态能量计算

从基态能量的费曼图可以看到Fock项相比Hartree项更加“量子”。经典电动力学中不存在“电子场”自由度，电子数密度是一个独立的变量。费曼图中电子数密度体现为闭合电子回线，Hartree项包含两个闭合电子回线，因此是经典电动力学就能够捕捉到的。相反，Fock项的闭合电子回线含有内部结构（一条库伦相互作用线或者说虚光子线），是经典电动力学捕捉不到的，只有有电子场自由度，并且不要求“粒子线的指向和时间顺序一致”的理论才能够捕捉到，在经典电动力学中是没有的。换一个角度，我们也可以说，只有Hartree项的理论的主要的涨落都是粒子数，在宏观下对应流体动力学所研究的那种涨落，从而是非常经典的。

6.1.6 Hartree-Fock近似的用处以及它何时失效

(6.9)是化学中最早的可用的不加入任何经验参数，从头计算电子态的方法。求解它得到的单电子波函数称为分子轨道。Hartree-Fock近似得到的能量和实际能量偏差并不大，但由于化学反应中的能量变化也不大，通常单独依靠Hartree-Fock近似无法非常精确地预言化学反应行为。然而，由于Hartree-Fock近似本身有清楚的场论意义，很容易以其为基础进一步开发更加精确的算法，因此它堪称量子化学的基石。

Hartree-Fock近似的用处其实不只是处理库仑相互作用的。显然对(1.10)中一般的电子-电子相互作用，由于存在动量转移，从高能物理的角度说，由于有光子交换，电子-电子相互作用顶角仍然需要包含相互作用线（此时绘制为虚线，因为没有特别良定义的“光子”概念），为



$$= -\frac{1}{\beta} \sum_{\omega_n} \int \frac{d^3 q}{(2\pi)^3} V_q, \quad (6.16)$$

因此(1.10)关于单电子格林函数的平均场近似同样包含Hartree项和Fock项（见图6.1），同样可以称为Hartree-Fock近似。

Hartree-Fock近似成立的条件是相互作用线修正和顶角修正可以略去，在这些说法不适用的情况下，自然就不正确了。与大部分平均场近似方法类似，自洽Hartree-Fock近似何时失效并不是特别清楚的，只有当计算结果出现了明显不对头的地方时我们才能够确定某些地方出问题了。例如，非自洽Hartree-Fock近似没有考虑任何屏蔽效应。自洽Hartree-Fock近似实际上通过Hartree项考虑了一部分屏蔽效应，但是Fock项当然也应该做同样级别的近似，从而应当将Fock项内部的库仑相互作用线替换成经过了Hartree项级别的屏蔽的有效相互作用线——见第6.2.2节结尾的讨论。在均匀电子气的例子中，非自洽Hartree-Fock近似会导致费米面附近的电子群速度发散，这是可以预料到的，因为费米面附近电子很多，小的外加电荷会快速被屏蔽，屏蔽效应不能忽略。高密度电子气中无疑有很强的屏蔽效应，因此Hartree-Fock近似确定不能对高密度电子气使用。然而另一方面，在关于关联能的讨论中我们可以看出，Hartree-Fock近似，即只计算单电子传播子修正，实际上相当于在单电子图像允许的范围内最大限度地考虑相互作用修正，因此Hartree-Fock近似在强关联系统中是必定失效的，因此单纯的Hartree-Fock近似又不能用于电子数密度太低的系统。总而言之，单纯的Hartree-Fock近似适用于这样的系统：电子数密度足够高，使得。

6.2 凝胶模型，高密度电子气的静电屏蔽和RPA近似

本节讨论高密度电子气的行为。高密度电子气中最明显的物理效应就是屏蔽，从而不能再使用Hartree-Fock近似。高密度电子气由于强烈的屏蔽效应，凝胶模型总是适用的。所谓凝胶模型是一种将金属中的原子当成均匀的正电荷凝胶的金属模型。这个假设对那些在其中的电子很像自由电子的金属系统是确实成立的，因为这样的系统中晶格对电子的作用比较小，从而能量微扰可以只考虑最低阶的(4.15)，其物理直观就是均匀的正电荷凝胶，它让所有电子的能量整体上升或是下降了一定量。由于高密度电子气强烈的屏蔽效应，晶格势场作用不明显，因此应该可以使用凝胶模型。我们可以把这部分能量变化整体归入化学势，因此，到最后，所谓凝胶模型实际上就是在讨论自由电子加上可以微扰处理的库仑相互作用之后会发生什么。

对高密度电子气，费曼图微扰展开等都是适用的，因为如同第6.1节开头说的那样，此时库仑相互作用确实不占据主导地位。然而由于库仑相互作用是长程的，计算有限个费曼图仍然会产生发散。实验上，金属和近独立电子气的行为确实很相似（对这件事的解释和讨论见第6.3节），因此至少非微扰效应不会让系统基态相比于近独立电子气面目全非，做

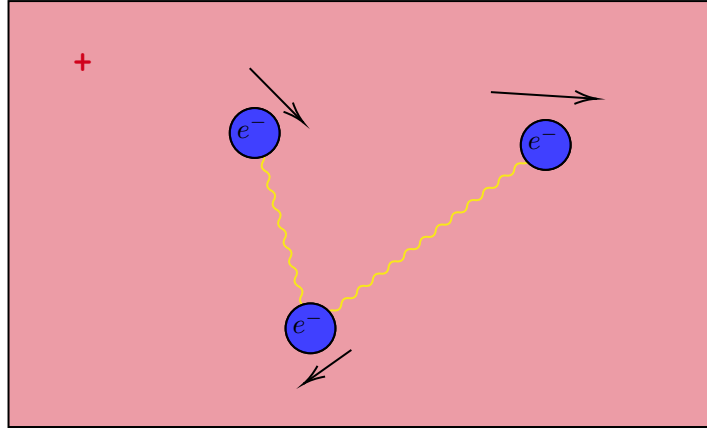


图 6.2: 凝胶模型: 原子实带来均一的、实际上不会对约束在晶体中的电子产生任何影响的正电荷背景, 主要的物理是电子以及它们的电磁相互作用

适当的费曼图重求和修正电子能带应该能够有足够好的精度。

按照

$$N_e \frac{4\pi}{3} r_s^3 = V$$

估算, 实际金属的 r_s 通常落在 $1.8 < r_s < 6$ 的范围内。因此, 库仑相互作用能的量级和电子动能大部分时候相差不大, 从而RPA近似仍然是不够的。

6.2.1 托马斯-费米静态屏蔽势

本节讨论的系统满足所谓托马斯-费米近似: 在其中, 电子充分稠密而均匀, 从而系统状态可以使用电子数密度 $n(\mathbf{r})$ 标记², 且相互作用能可以完全认为是来自密度-密度相互作用即Hartree项。这样可以完全照搬经典静电学, 即认为一个电子的全部能量是其动能加上电势能, 而不考虑任何交换能或者关联能, 因为直观地看, 交换能或者说Fock项真的对应一个两个电子交换位置的费曼图, 但是既然电子如此稠密, 这样的过程大概是不太能够发生的, 从而, 如果考虑这样的过程, 那么它的影响局限于电子自能修正。

想象我们在一个满足托马斯-费米近似的近独立电子气当中放入一个电荷。显然, 异号电荷会移动到前者附近而产生一个屏蔽效应。在电子气规模很大时 (我们所研究的固体总是在热力学极限之下, 因此这是成立的), 屏蔽效应很强, 这让外加电荷对电子气的扰动实际上局限在外加电荷的一个小邻域内。于是求解外加电荷的影响就是求解以下自洽问题:

1. 外加电荷导致外加电势能

$$\phi_{\text{ext}} = \frac{Q}{r}, \quad (6.17)$$

2. 外加电势与屏蔽电荷形成的电势 $\delta\phi$ 叠加, 导致总的电势变化为

$$\phi_{\text{eff}} = \phi_{\text{ext}} + \delta\phi, \quad (6.18)$$

3. 总电势变化 ϕ_{eff} 导致了电荷密度变化 δn , 电荷密度变化给出屏蔽电势 $\delta\phi$ 。

自洽求解以上问题就可以确定所有物理量。

²很明显由于 \mathbf{r} 和 \mathbf{p} 不对易, 如果我们想要有有意义的 \mathbf{k} 表象, 系统中所有依赖于 \mathbf{r} 的量在空间上都必须变化得足够平缓。这和前面的波包动力学中外加电场时的条件是一个道理。

以上自洽求解步骤还需要一个方程： V_{eff} 是怎么影响电荷密度的。 V_{eff} 改变粒子排布的方式是这样的：单电子的能量 $\epsilon_{\mathbf{k}}$ 会由于 V_{eff} 的引入发生一个小的变化，且这个小的变化在不同位置通常是不一样的。由于托马斯-费米近似成立，我们认为每个 \mathbf{r} 位置附近的诸电子都可以认为是组成了一个正则系综³，且 \mathbf{k} 和 \mathbf{r} 可以同时确定，于是

$$n(\mathbf{r}) = \sum_{\mathbf{k}, \sigma} \frac{1}{1 + e^{\beta(\epsilon_{\mathbf{k}}(\mathbf{r}) - \mu)}} = 2 \sum_{\mathbf{k}} \frac{1}{1 + e^{\beta(\epsilon_{\mathbf{k}}(\mathbf{r}) - \mu)}}, \quad \epsilon_{\mathbf{k}}(\mathbf{r}) = \epsilon_{\mathbf{k}}^0 + V_{\text{eff}}(\mathbf{r}),$$

由于体系足够大，屏蔽作用让外加电荷的作用几乎是局域的，则 $\delta n(\mathbf{r})$ 可以写成 $V_{\text{eff}}(\mathbf{r})$ 的函数而不涉及长程相互作用。而由于外加电荷很小，可以采取线性近似，于是（第二个等号是因为费米-狄拉克分布的形式）

$$\delta n = \frac{dn}{dV_{\text{eff}}} V_{\text{eff}} = \sum_{\mathbf{k}} \frac{\partial n}{\partial \epsilon_{\mathbf{k}}} V_{\text{eff}} = -\frac{\partial n}{\partial \mu} V_{\text{eff}},$$

关于能量的态密度（即单位 $d\epsilon$ 、单位体积中有多少可能的态）为

$$N(\mu' - \mu) = \frac{\partial n|_{\mu \rightarrow \mu'}}{\partial \mu'}, \quad (6.19)$$

其中能量从费米面量起，且我们将 n 的表达式中的 μ 换成了变量 μ' ，而使用 μ 表示实际系统的化学势。使用该记号，则

$$\delta n(\mathbf{r}) = -N(0)V_{\text{eff}}(\mathbf{r}). \quad (6.20)$$

这就是从 V_{eff} 计算 δn 的方法。通过 δn 计算 $\delta \phi$ 的方程为泊松方程

$$-\nabla^2 \phi_{\text{eff}} = 4\pi (Q\delta(\mathbf{r}) - e\delta n(\mathbf{r})),$$

其中我们不失一般性地将外加电荷放在原点上。使用傅里叶变换并且使用 δn 和 V_{eff} 之间的线性关系(6.20)，可以得到

$$k^2 V_{\text{eff}}(\mathbf{k}) = -4\pi eQ - 4\pi e^2 N(0) V_{\text{eff}}(\mathbf{k}),$$

解出 $V_{\text{eff}}(\mathbf{k})$ 为

$$V_{\text{eff}}(\mathbf{k}) = -\frac{4\pi eQ}{k^2 + \kappa_{\text{TF}}^2},$$

做傅里叶逆变换得到

$$V_{\text{eff}}(\mathbf{r}) = -\frac{eQ}{r} e^{-\kappa_{\text{TF}} r}, \quad (6.21)$$

其中

$$\kappa_{\text{TF}}^2 = 4\pi e^2 N(0). \quad (6.22)$$

这个结果就是托马斯-费米屏蔽。

6.2.2 RPA近似下的动态屏蔽响应

托马斯-费米屏蔽的推广

托马斯-费米屏蔽是静态的。本节分析更加一般的动态屏蔽响应。回顾第5.4.2节开头的说法，计算密度-密度松原格林函数即得到最为一般的电子气对外加电势的响应。可以直接

³这么做总是可以的，因为我们讨论近平衡态理论，确实允许电荷密度在空间上出现分布，虽然完全的平衡态电子气应该是均匀的。

用费曼图计算出格林函数 G_{nn}^{ret} ，但是本节不这么处理问题，因为如果真的这么做很快就会遇到数学上的困难：费曼图求和需要经过一些特殊的重求和步骤才能够给出有意义的结果。因此本节只是做一个近似，将最为奇异的那部分费曼图做一个求和，然后发现它确实能够给出有意义的结果。

为简化问题我们做RPA (Random Phase Approximation) 近似⁴，这个近似要求以下两点：

- δn 和 V_{eff} 之间的响应函数近似是自由电子的 δn - δn 格林函数，即我们认为 V_{eff} 中来自电子屏蔽的部分在其它电子上的作用和一个外加势场完全没有区别；
- 只考虑经典静电能，认为 V_{eff} 中来自电子屏蔽的部分完全就是使用 $\langle \delta n(\mathbf{r}, t) \rangle$ 激发出来的；不考虑诸如Fock交换能这样的来源。

熟悉费曼图的话，容易看出实际上这两个近似就是在做环形图近似：一个闭合电子圈就是 $\langle \delta n(\mathbf{r}, t) \rangle$ ，而自洽地计算泊松方程给出的 V_{eff} 和自由电子的密度-密度格林函数等价于计算库仑相互作用线的“真空极化”；两个近似放在一起就给出环形图近似；我们之后将显式地验证这一点。

RPA近似看起来似乎没有什么明确的意义，但它实际上把最为奇异剧烈效应考虑在内，并且给出了定性上正确的屏蔽库伦势（从一个长程的相互作用变成了一个短程的相互作用）。如果RPA近似不准，只需要多算几个微扰就可以。这样，自洽方程可以通过下面的方法给出。

1. 外加电荷引入外加电势能，
2. 外加电势能加上屏蔽电荷产生的电势能等于总的电势能变化 V_{ext} ，
3. 总的电势能变化可以反推出屏蔽电荷。

粒子数的平均值近似为一个经典的数密度，从而 δV 可以使用泊松方程写成

$$-\nabla^2 \delta V(\mathbf{r}, t) = -4\pi e^2 \langle \delta n(\mathbf{r}, t) \rangle,$$

也即

$$k^2 \delta V(\mathbf{k}, \omega) = 4\pi e^2 \langle \delta n(\mathbf{k}, \omega) \rangle. \quad (6.23)$$

于是第二步完成了。

第三步通过RPA近似的基本假设得到：

$$\langle \delta n(\mathbf{k}, \omega) \rangle = G_{nn}^{0, \text{ret}}(\mathbf{k}, \omega) V_{\text{eff}}(\mathbf{k}, \omega). \quad (6.24)$$

这样就得到自洽方程

$$V_{\text{ext}}(\mathbf{k}, \omega) + \frac{4\pi e^2}{k^2} \langle \delta n(\mathbf{k}, \omega) \rangle = (G_{nn}^{0, \text{ret}}(\mathbf{k}, \omega))^{-1} \langle \delta n(\mathbf{k}, \omega) \rangle,$$

从而

$$G_{nn}^{\text{ret}}(\mathbf{k}, \omega) = \frac{G_{nn}^{0, \text{ret}}(\mathbf{k}, \omega)}{1 - \frac{4\pi e^2}{k^2} G_{nn}^{0, \text{ret}}(\mathbf{k}, \omega)} = \frac{G_{nn}^{0, \text{ret}}(\mathbf{k}, \omega)}{\epsilon(\mathbf{k}, \omega)}, \quad (6.25)$$

其中 $\epsilon(\mathbf{k}, \omega)$ 是介电常数，为

$$\epsilon(\mathbf{k}, \omega) = \frac{V_{\text{ext}}}{V_{\text{eff}}} = 1 - \frac{4\pi e^2}{k^2} G_{nn}^{\text{ret}, 0}(\mathbf{k}, \omega). \quad (6.26)$$

⁴这个名称最早来自核物理，在那里它是通过对某个随机相位的求和得到的。

这样，计算出近独立电子的密度-密度格林函数之后问题就完全解决了。而近独立电子的密度-密度格林函数是已经计算出来的(5.60)，因此问题完全解决了。

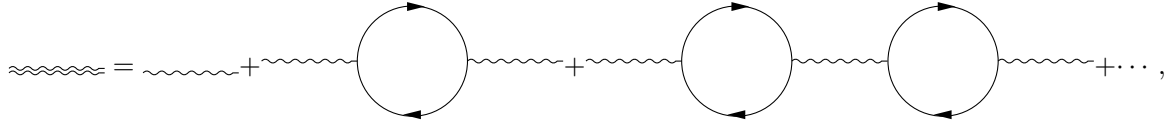
(5.60)可以推导出托马斯-费米屏蔽。在 V_{eff} 在时间和空间上都变化非常缓慢时，只需要考虑 \mathbf{q} 接近零的那部分傅里叶分量。让 ω 和 \mathbf{q} 趋于零，我们有

$$G_{nn}^{\text{ret}, 0}(\mathbf{q}, i\omega_n) = 2 \sum_{\mathbf{k}} \frac{df(\xi_{\mathbf{k}})}{d\xi_{\mathbf{k}}} = -2 \sum_{\mathbf{k}} \frac{df(\xi_{\mathbf{k}})}{d\mu} = -\frac{\partial n}{\partial \mu} = -N(0), \quad (6.27)$$

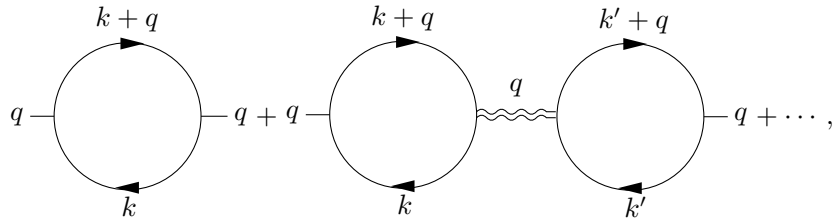
这正是我们在托马斯-费米屏蔽中做的近似：这表明RPA近似在低频下的确和托马斯-费米屏蔽一致。这个事实的原因下面马上会讲到。

RPA近似的场论表述

实际上从(6.25)就可以看出RPA近似的实质。我们考虑环形图近似，即在电子自能、真空极化（即相互作用线的修正）和顶角修正之中，只考虑相互作用线的修正，则有



而环形图近似下的密度-密度格林函数在扣除了无涨落的 $\langle n_{\mathbf{k}\sigma} \rangle \langle n_{\mathbf{k}'\sigma'} \rangle$ 并乘以 -1 之后，是



注意到根据(5.57)， $G_{nn}^{0, \text{ret}}(\mathbf{k}, \omega)$ 就是电子-空穴组成的环中两个传播子的乘积（没有加闭合费米子回线导致的 -1 因子），环形图近似下的密度-密度格林函数在扣除了无涨落的 $\langle n_{\mathbf{k}\sigma} \rangle \langle n_{\mathbf{k}'\sigma'} \rangle$ 之后由

$$-G_{nn}(\mathbf{q}, i\omega_n) = -G_{nn}^0(\mathbf{q}, i\omega_n) + G_{nn}^0(\mathbf{q}, i\omega_n) \left(-\frac{4\pi e^2}{|\mathbf{q}|^2} \right) G_{nn}^0(\mathbf{q}, i\omega_n) + \dots$$

确定。另一方面，将(6.25)做级数展开，我们有

$$G_{nn}^{\text{ret}}(\mathbf{q}, \omega) = G_{nn}^{0, \text{ret}}(\mathbf{q}, \omega) + G_{nn}^{0, \text{ret}}(\mathbf{q}, \omega) \frac{4\pi e^2}{|\mathbf{q}|^2} G_{nn}^{0, \text{ret}}(\mathbf{q}, \omega) + \dots,$$

和前一式只相差一个Wick转动。因此我们就发现(6.25)实际上就是环形图近似或者说项链图近似。

(6.25)实际上给出了 G_{nn}^{ret} 中最为奇异的部分，因为环形图会产生次数非常高的 $1/k^n$ 型的项，从而发散得最厉害；但是将所有这些发散的图求和起来却能够得到一个收敛的结果。环形图近似足够精确的条件是粒子数密度很高，从而主要的相互作用形式是电子通过库仑相互作用，激发出一些电子-空穴对（即环形图中的环），即主要的相互作用是环形图。也可以从另一个角度理解这件事：粒子数密度较高意味着屏蔽作用很强，所以 $1/k^n$ 发散得不那么厉害的那些项没有什么贡献。RPA近似下没有出现这样的图：一个电子回线内部有一条光子线，即在单电子回线的图中没有出现Fock项，这和托马斯-费米屏蔽的假设是一致的，因此确实可以预期，当 $\omega \rightarrow 0$ 时RPA近似退化到托马斯-费米屏蔽上：托马斯-费米屏蔽

和RPA近似的唯一差别就是前者只考虑 $\delta n(\mathbf{r}, t) = \delta n(\mathbf{r})$ 的情况，从费曼图的角度出发，两者计入的图的结构是一样的。

如果我们只计算两个电子交换一个光子，即发生一次库伦散射的图，那么会得到发散；RPA近似提供了一个足够好的方法能够消除这种发散。由于RPA近似不再只讨论单电子传播子的修正，它可以用于讨论电子关联。只要高密度电子气的假设成立，RPA近似就是足够精确的。

最后我们注意到，以上讨论中的单电子格林函数都是没有经过任何修正的，然而这并不是必须的：完全可以用经过修正的单电子格林函数做RPA近似，不存在任何重复计数的问题。上述计算中我们都是在考虑外加电荷激发出的响应，并且我们都是在对库仑相互作用线做环形图修正，那么当然，在计算单电子自能时应该做类似的修正。使用经过相互作用修正的——也许就是经过Hartree-Fock近似修正的，虽然此时库伦排斥最好替换为屏蔽后的库伦排斥——单电子格林函数会让 $\epsilon(\mathbf{k}, \omega)$ 的零点位置发生偏移。这种同时做Hartree-Fock近似和RPA近似的思路是很重要的，因为在计算系统基态能量时，Fock项的修正毕竟要远大于RPA近似带来的关联能修正；实际上，Hartree项加上RPA修正后的Fock项给出的就是所谓的GW近似，它可以相当精确地计算弱关联体系的能带结构。

RPA计算基态能量

我们还可以进一步，根据环形图近似把高密度电子气的基态能量计算出来。将有限个费曼图加起来计算基态能量是注定要失败的，因为正如第6.1.6节中提到的那样，由于屏蔽效应没有被系统考虑，很可能出现各种能量发散。

RPA近似能够计算基态关联能，它需要叠加在Hartree项和Fock项上才给出精度足够的总能量。

6.2.3 等离子体和等离激元

运动方程

等离激元的能谱

考虑 ω 非常高的极限，此时做展开

$$\frac{1}{\omega + \xi_{\mathbf{k}} - \xi_{\mathbf{k}+\mathbf{q}} + i0^+} = \frac{1}{\omega} - \frac{\xi_{\mathbf{k}} - \xi_{\mathbf{k}+\mathbf{q}}}{\omega^2} + \dots,$$

展开到前两阶，代入(5.60)，得到

$$\begin{aligned} G_{nn}^{\text{ret}, 0}(\mathbf{q}, \omega) &= -\frac{2}{\omega^2} \sum_{\mathbf{k}} (f(\xi_{\mathbf{k}}) - f(\xi_{\mathbf{k}+\mathbf{q}}))(\xi_{\mathbf{k}} - \xi_{\mathbf{k}+\mathbf{q}}) \\ &= \frac{2q^2}{m\omega^2} \sum_{\mathbf{k}} f(\xi_{\mathbf{k}}), \end{aligned}$$

于是就得到

$$G_{nn}^{\text{ret}, 0}(\mathbf{q}, i\omega_n) = \frac{q^2 n_e}{m\omega^2}, \quad (6.28)$$

相应的，

$$\epsilon(\mathbf{k}, \omega \rightarrow \infty) = 1 - \frac{4\pi e^2 n_e}{m\omega^2},$$

定义

$$\omega_p^2 = \frac{4\pi e^2 n_e}{m}, \quad (6.29)$$

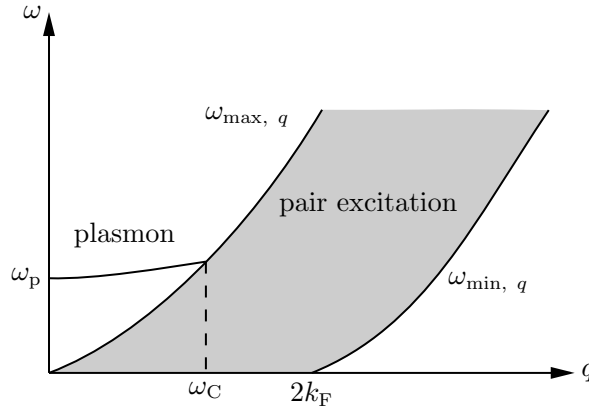


图 6.3: RPA近似下金属中的密度波模式

就得到

$$\epsilon(\mathbf{k}, \omega \rightarrow \infty) = 1 - \frac{\omega_p^2}{\omega^2}. \quad (6.30)$$

考虑到(6.25)，当 $\omega = \omega_p$ 时格林函数出现奇点，即此处有集体激发。 $G_{nn}^{\text{ret}}(\mathbf{k}, \omega)$ 是一个密度-密度格林函数，对任意的动量，在 $\omega = \omega_p$ 处都有一个密度波的集体激发模式。我们此处的计算是在频率非常大，以至于介质几乎是等离子体时得到的，因此这种集体激发实际上正是等离激元。物理地看，等离激元是密度波这件事实际上说明了等离子体中的电子运动方式，即所有电子和空穴共同集体振荡。这个集体激发的频谱是平的，这当然是因为我们假定光速无穷大，完全使用库仑相互作用。如果考虑到光速有限，使用麦克斯韦方程可以推出

$$\omega_{\mathbf{k}}^2 = \omega_p^2 + c^2 k^2 / \epsilon(\infty), \quad (6.31)$$

即光子的一个模式和固体发生耦合，打开了一个能隙。在假定光速无限时，丢弃不可见、恒定的无穷大项（相当于将元激发的产生湮灭算符的时间演化加上一个因子 e^{ickt} ）就得到

$$\omega_{\mathbf{k}} = \omega_p,$$

即我们算出的结果。

等离激元是密度波集体模式，它的出现是库仑相互作用引入的结果（这点可以从其频率有 e 的依赖看出）。观察(6.25)的形式，会发现近独立电子气中有的电子-空穴对在引入库仑相互作用之后仍然存在，且能谱没有发生变化。因此和密度-密度关联有关的模式如图6.3所示。灰色区域是电子-空穴对的能谱，从 ω_p 出发的那一支能谱是等离激元的能谱。等离激元的能谱画到和 ω_{max} 相交处就不再画了，因为等离激元和电子-空穴对都是密度波模式，等离激元如果具有和电子-空穴对相同的 ω 和 \mathbf{q} ，它就立刻可以分散成许多 \mathbf{k} 不同， \mathbf{q} 相同的电子-空穴对，因此此时等离激元不稳定。

我们最后指出等离激元计算中的一个重要细节。自由电子密度格林函数(5.60)含有一个 $f(\xi_{\mathbf{k}}) - f(\xi_{\mathbf{k}+\mathbf{q}})$ ，这个因子实际上可以重写成

$$f(\xi_{\mathbf{k}})(1 - f(\xi_{\mathbf{k}+\mathbf{q}})) - (1 - f(\xi_{\mathbf{k}}))f(\xi_{\mathbf{k}+\mathbf{q}}),$$

也就是说，自由电子密度格林函数实际上包含两个部分：一个是所谓的“激发”（excitation）部分，在它的极点上 $\omega > 0$ ，还有一个是所谓的“去激发”（deexcitation）部分，在它的极点上 $\omega < 0$ 。这两个部分是预期之中要有的，因为电子密度的振荡是实玻色型激发，而实玻色型激发的格林函数一般都会有一个产生部分和一个湮灭部分。在前述的泰勒展开 $G_{nn}^{\text{ret}, 0}(\mathbf{q}, \omega)$ 的

过程中, 因子 $f(\xi_{\mathbf{k}}) - f(\xi_{\mathbf{k}+\mathbf{q}})$ 是非常重要的: 如果我们在 $G_{nn}^{\text{ret}, 0}(\mathbf{q}, \omega)$ 中只保留激发或者去激发支, 算出来的玻色型激发的能量是大错特错的。因此, 非常有趣的是, 在体系中出现等离激元时, 电子密度格林函数中的产生部分和湮灭部分被库仑相互作用大大地混合起来了。这种事情在别的体系里面也有出现, 如超导准粒子的 Bogoliubov 变换。相反, 在计算激子的时候, 我们发现在相互作用修正后的电子密度格林函数中, 仍然有定义良好的激发和去激发支。假定相互作用修正后的电子密度格林函数有定义良好的激发和去激发支, 这个假定称为 Tamm-Dancoff 近似。这个近似下我们可以分别考虑激发支和去激发支受到的修正, 而不考虑两者被耦合在一起。Tamm-Dancoff 近似对激子适用, 对等离激元不适用, 这正好体现了后者是一个更加“多体”的现象。

6.2.4 Kohn 异常

一些三维金属中, 对

6.3 金属和朗道费米液体

朗道费米液体是一个这样的低能有效理论, 其基本假设为:

1. 费米液体的状态可以和费米气体一样, 使用某种费米型自旋 $1/2$ 准粒子的占据数标记, 并且系统能量可以写成占据数的函数。这个假设对哈密顿量的形式做出了很强的限制, 因为此时的哈密顿量已经在粒子数表象下被严格对角化了。(但是这个哈密顿量仍然可以不是自由哈密顿量, 因为可以有密度-密度相互作用) 不过我们将会看到, 这样的限制是合理的。
2. 当相互作用趋于零 (“被关闭”) 时, 费米液体态回退到实际的费米气体态, 我们假设此时费米液体态的占据数和实际的费米气体态的占据数相同。这个假设实际上是说, 费米液体中的占据数所描述的准粒子实际上是经过某种重整化的电子, 和电子可以一一对应。

可以从费米液体中准粒子的行为中看到很明显的普通电子的影子: 准粒子场和电子场的对称性相同, 准粒子也是费米子, 准粒子能谱和电子能谱形式类似, 自旋均为 $1/2$, 带电量相同, 等等。唯象的讨论中可以直接将费米液体中的准粒子当成电子。

到这里我们其实并不知道费米液体中的准粒子到底是什么。可以通过这样的方式想象费米液体可以怎样产生: 对无相互作用的能带电子⁵, 我们可以绝热地将相互作用加入, 如果没有出现诸如费米子配对之类让系统能谱结构发生很大变化的情况, 那么第二个假设就已经满足了。

相互作用会导致电子出现自能修正, 由于库仑相互作用的顶角是保持粒子数守恒的, 不存在一个电子衰变成多个其它粒子的过程, 因此自能修正是实数, 即单电子态的确是本征态。⁶在很多粒子数守恒的理论——如 ϕ^4 理论或是 QED——中, 可以有稳定的单粒子、二

⁵应当注意能带电子这个名称本身有些模糊不清。只计算周期性势场的影响可以得到能带, 只计算经过平滑处理的赝势也可以得到能带, 将一部分相互作用作为自能修正而留下最为显著的相互作用通道作为碰撞也可以得到能带 (虽然还需要额外计算碰撞的影响, 如费米液体中那样), 对很多系统, 其实相互作用全都考虑进去了还是可以得到能带——如同用 DFT 计算出来的那样。例如, 在费米液体中, 如果粒子数涨落不大, 完全可以用粒子数平均值代替相互作用项中的准粒子占据数。即使相互作用大一些, 也可以理解成“能带被拉歪了”。

⁶虽然如此, 如果我们将高于费米面的电子称为电子而将低于费米面的电子称为空穴, 那么的确可以出现一个电子和费米球中的空穴发生相互作用, 而衰变成电子和空穴的过程。见第 6.3.1 节。

粒子、三粒子……本征态，虽然 n 粒子态的能量不是 n 个粒子的能量简单相加，但是无论如何，一个本征态的能量可以写成各个动量上的粒子的占据数的函数，因此量子态可以用占据数标记，能量也可以用粒子数标记，即第一个假设也是成立的。

对那些确定是从相互作用电子系统产生出来的费米液体，准粒子的动量就是电子的动量。这是因为，在关闭相互作用时费米液体“无缝地”退化到实际的费米气体上，因此在无相互作用点处费米液体中准粒子的动量就是实际的能带电子的动量。现在缓慢地加上相互作用，则费米液体准粒子的动量可以发生连续变化，但有限体系中动量实际上是分立的，从而动量只能不变。因此，准粒子的动量实际上就是电子的动量。然而动量守恒和粒子数没有关系，设 $|\Psi\rangle$ 是费米液体的基态而 $c_{\mathbf{k}}^\dagger$ 是Bloch电子的产生算符，则 $c_{\mathbf{k}}^\dagger|\Psi\rangle$ 不是 $n_{\mathbf{k}}$ 精确地加一的态，而是多出来一个电子位于 \mathbf{k} 上、多出来两个电子位于 \mathbf{k} 上和一个空穴位于 $-\mathbf{k}$ 上、多出来三个电子位于 \mathbf{k} 上和两个空穴位于 $-\mathbf{k}$ 上这一系列态的叠加。费米液体中的准粒子是能带电子重整化得到的，这个过程可以线性组合 $c_{\mathbf{k}}$ 和 $c_{\mathbf{k}}^\dagger$ 。这也意味着无法保证费米液体的准粒子的费米面和产生费米液体的含有库伦散射的多电子系统在不考虑库伦散射时的费米面高度是一样的。

在什么情况下费米液体理论失效？如果相互作用是吸引的，那么低温下可能出现费米子配对，此时总是会出现相变。还有一些情况下相互作用会导致电子在空间上排列成一定的模式（这样的模式可以看成“电子晶体”，即“费米液体结晶了”），此时也会出现相变，因为出现了新的序。我们将在第八章中讨论这些情况。如果相互作用较强，能带的图像甚至可能就不适用了，此时系统的元激发可能都不是经过修正的电子，也无所谓费米液体。令人震惊的是，虽然大部分实际体系中库仑相互作用的确很强，并且显然不只有密度-密度相互作用通道，费米液体图像仍然是适用的。

6.3.1 基态附近的激发

与高能物理中不同，实际的凝聚态体系的基态中都有大量准粒子。由于准粒子是费米子，系统基态中有一个费米球。由于粒子数守恒，系统基态附近的激发态就是一些准粒子从费米球内部被打到费米球外部之后形成的，并且距离费米面均不远。因此其实我们可以忽略费米海而认为系统中实际上有两种元激发：“电子”和“空穴”，而所谓的费米面就是区分电子和空穴的一个边界。

在费米液体理论中通常也只分析费米面附近的物理，部分原因在于费米海的结构可以非常复杂，因此只考虑费米面附近的物理是比较容易的，也是比较有实际意义的（因为是低温近似），部分原因在于只有这里才确定有稳定的准粒子——通常准粒子的寿命在接近费米面时比较长，因此看起来像是“真正的”粒子（否则会有非常明显的能级展宽）。

这件事的原因如下。首先应当注意，虽然准粒子的自能修正总是实数，从而，在没有其它任何准粒子时，单个准粒子可以永远稳定存在而不会衰变，但与高能物理中的情况不同，费米球的存在意味着如果费米面上方出现了一个准粒子，它会激发费米海内部的准粒子，从而产生一个空穴，正如图6.4所示。因此，在基态（不是真空态，而是带有费米球的态），费米面上方的准粒子的确会发生衰变。将费米面上方的准粒子和费米球内部的空穴看成两种元激发可以更加清楚地看到这一点：此时费米面上方的准粒子的个数根本不是守恒的（从而，积掉费米海会产生一个非幺正的理论），从而其自能修正不可能是实数。这是含有相互作用的多粒子态的一般特征：在这种态中没有非常“尖锐”的单粒子极点。

设准粒子寿命为 τ ，则 τ 反比于散射速率，而散射速率正比于库伦散射的强度。完整地做这个计算是很困难的，因为涉及到静电屏蔽等复杂的效应。散射的过程可以概括为：一

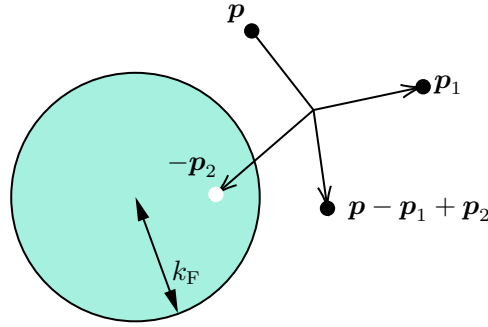


图 6.4: 一个准粒子衰变成两个准粒子和一个准空穴, 或者说准粒子费米海中的一个准粒子被激发到费米面以上

个动量为 p 的费米面外部的准粒子的能量降低, 变成了动量为 p_1 的准粒子, 同时激发了一个费米面内的动量为 p_2 的准粒子。结果是, 动量为 p 的费米面以外的准粒子衰变成了两个准粒子, 动量分别为 p_1 和 $p - p_1 + p_2$, 还有一个动量为 $-p_2$ 的空穴。如果这种散射实际上很少发生, 那么就能够有稳定的准粒子。⁷

设我们通过若干轮重整化群计算, 已经得到了一种有费米型自旋1/2的准粒子, 且符合费米液体的第二个假设的有效理论。假设准粒子之间的相互作用强度不非常强地依赖于动量, 不妨记作 M (纯粹的库仑相互作用不满足这个条件, 它强烈依赖入射电子的动量差, 因此即使在费米面附近, 相互作用强度也不能当作常数, 因此能带电子加入库仑相互作用后并不严格就是稳定的准粒子, 而仍然需要经过几轮重整化, 让 M 变得“柔和”一些)。我们至此还没有确定这个系统是费米液体, 因为其中准粒子之间有散射, 不同动量上的准粒子数目有涨落。在这样的系统中, 动量分别为 p_1 和 $p - p_1 + p_2$ 的两个准粒子和动量为 p_2 的空穴的联合态密度在当前温度下的期望值为 n , 由费米黄金法则有

$$\frac{1}{\tau} \propto \text{transition rate} \sim |M|^2 n.$$

由于系统中的准粒子非常多, 不同能级上准粒子数的涨落可以略去, 即认为不同能级上不多不少正好就有费米-狄拉克分布给出的粒子个数,⁸那么就有

$$n = \int d^3 p_1 \int d^3 p_2 (1 - f(\epsilon_{p_1})) f(\epsilon_{p_2}) (1 - f(\epsilon_{p - p_1 + p_2})) \delta(\epsilon_p - \epsilon_{p_1} + \epsilon_{p_2} - \epsilon_{p - p_1 + p_2}).$$

因子 $(1 - f(\epsilon_{p_1}))$ 表示动量为 p_1 的粒子应该占据一个空态 (或者说在接近零温时应该在费米面以外), 因子 $f(\epsilon_{p_2})$ 表示空穴一定来自一个已有的粒子, 最后的 δ 函数强制要求能量守恒。我们不严格计算这个积分, 而是做一些数量级估计。由于 p_2 在费米面以下而 $p - p_1 + p_2$ 在费米面以上, 容易写出以下不等式

$$0 < \xi_{p_1} < \xi_p, \quad 0 < -\xi_{p_2} < \xi_p - \xi_{p_1} < \xi_p,$$

对 n 有贡献的 p_1 和 p_2 均满足这些不等式, 这些不等式给出了两个宽度为 ξ_p 的球壳, 因此

$$n \leq (4\pi k_F^2 \xi_p)^2,$$

⁷应注意此处所谓的碰撞或者散射可以发生在零温下: 加入库仑相互作用后, $|n_{k_1} = 1, n_{k_2} = 0\rangle$ 和 $|n_{k_1} = 0, n_{k_2} = 1\rangle$ 都可以出现在基态中; 换言之, 在基态, k_1 处的电子也可以被散射到 k_2 处。使用一种陈词滥调的说法, 这些散射部分来自量子涨落而非热涨落。但应注意一般假定费米液体的基态同样具有平移不变性, 从而各个形如 $|n_{k_1} = 1, n_{k_2} = 0\rangle$ 和 $|n_{k_1} = 0, n_{k_2} = 1\rangle$ 的分量的系数应当遵从适当的条件, 使得 $\langle c_{k_1} c_{k_2}^\dagger \rangle$ 只有在 $k_1 = k_2$ 时才非零; 否则, 基态电子数密度在实空间中将会存在条纹。

⁸这是热力学系统的一般性质: 系统规模大时涨落可略去。由于本文涉及的系统都是多体系统, 总是可以做这样的近似。

于是

$$\frac{1}{\tau} \lesssim \xi_p^2. \quad (6.32)$$

因此，如果准粒子非常接近费米液体的费米面，那么它是非常稳定的，因为此时 ξ_p 很小。从物理图像上看，此时的准粒子虽然会和费米海中的准粒子发生相互作用，但其能量不足以产生粒子-空穴对，因此也不会衰变。

总之，准粒子之间的碰撞不可忽略意味着费米面以上的准粒子会衰变。然而，在费米面附近，这种衰变是非常缓慢的，以至于我们可以认为任何一种动量确定的准粒子分布都可以长期稳定存在，即哈密顿量在以动量标记的粒子数表象下是对角化的，从而费米液体理论的第一个假设成立。

从上面的论述可以看出，对相当大的一类相互作用电子气系统，**只有在费米面附近第一个假设才是严格成立的**，在稍微远离费米面的地方，准粒子之间仍然有碰撞。显然，准粒子完全没有碰撞的情况只适用于这种电子气系统的红外极限。几乎所有凝聚态系统都处在接近热平衡的状态，温度提供了一个好的能标，因此费米液体理论的适用条件就转化成了这样一个问题：热涨落是不是足够让准粒子之间的碰撞体现出来？换句话说，动量确定的准粒子寿命是否长到**热涨落**的时间尺度？或者等价地说，由于碰撞导致的准粒子能级展宽相比于费米-狄拉克分布函数关于能量的变化尺度是不是可以忽略的？

我们首先估算电子平均自由时间。按照费米-狄拉克分布， $\langle n \rangle$ 在 ϵ_F 附近一个大约长为 T 的区域内明显偏离阶跃函数；只有这个“弥散区域”内的电子会发生散射（费米球内部的准粒子不怎么会被激发，费米球外面根本没有准粒子）。发生一次散射需要两个来自弥散区域的电子，因此

$$\frac{1}{\tau} \sim T^2,$$

即

$$\tau \sim \frac{1}{T^2}, \quad (6.33)$$

准粒子平均自由时间本身和温度有关。另一方面，碰撞导致能量本身出现弥散，为⁹

$$\Delta E \sim \frac{1}{\Delta t} \sim \frac{1}{\tau},$$

其中 τ 为准粒子平均自由时间。如前所述，有稳定准粒子等价于

$$\Delta E \ll \frac{1}{T},$$

也即

$$\frac{1}{\tau} \ll T.$$

最后我们发现我们有

$$T \ll 1. \quad (6.34)$$

因此费米液体图像只在低温下适用。

6.3.2 哈密顿量和前向散射相互作用

费米液体的能级结构和费米气体完全一致并不意味着库仑相互作用完全不产生任何影响，因为每个能级具体的能量大小可以经历修正。这意味着准粒子的哈密顿量并不是简单的

$$H = \sum_{\mathbf{k}, \sigma} \epsilon_{\mathbf{k}} n_{\mathbf{k}\sigma},$$

⁹注意不确定性关系在混合态中也是有的。

而有一些高阶项，它们表示密度-密度相互作用。

考虑一个能量本征态，其中准粒子在费米面之上的数量为 δn （ δ 表示相对基态的偏离，正表示有粒子，负表示有空穴）。考虑到费米液体理论的第一条假设，把能量本征值相对于零温平衡态（由于费米海的结构可以非常复杂，零温能量反而是算不出来的，我们也不需要计算它）的变化写成以下级数展开（ \mathbf{k} 在费米面附近）：

$$\delta E = \underbrace{\sum_{\mathbf{k}, \sigma} \epsilon_{\mathbf{k}}^0 \delta n_{\mathbf{k}\sigma}}_{\delta E_1} + \underbrace{\frac{1}{2V} \sum_{\mathbf{k}, \mathbf{k}', \sigma, \sigma'} f_{\sigma\sigma'\mathbf{k}\mathbf{k}'} \delta n_{\mathbf{k}\sigma} \delta n_{\mathbf{k}'\sigma'}}_{\delta E_2}, \quad (6.35)$$

其中 δn 表示准粒子数目相对基态的变化。把能量写成粒子数的函数假定了自旋守恒。对动量求和化积分，就得到

$$\frac{\delta E}{V} = \underbrace{\sum_{\sigma} \int \frac{d^3\mathbf{k}}{(2\pi)^3} \epsilon_{\mathbf{k}}^0 \delta n_{\mathbf{k}\sigma}}_{\delta E_1/V} + \underbrace{\frac{1}{2} \sum_{\sigma, \sigma'} \int \frac{d^3\mathbf{k}}{(2\pi)^3} \int \frac{d^3\mathbf{k}'}{(2\pi)^3} f_{\sigma\sigma'\mathbf{k}\mathbf{k}'} \delta n_{\mathbf{k}\sigma} \delta n_{\mathbf{k}'\sigma'}}_{\delta E_2/V} \quad (6.36)$$

我们保留到二阶项，一阶项给出准粒子的近独立理论，二阶项给出准粒子的相互作用。这种相互作用并不会让粒子数发生涨落或是让单个准粒子的动量发生涨落，而只会对能级做修正，是所谓的“前向散射”。会让准粒子动量发生变化的那部分相互作用或者说“碰撞”，正如上一节所说，需要略微远离一些费米面才看得到，没有体现在(6.35)中。

重整化群可能给出更高阶项，而也许二阶项实际上可以忽略，而绕了一大圈之后发现 $|M|$ 很小，从而实际上在很大的区域内费米液体图像均适用。为了表明我们保留到二阶项是正确的，我们给出一个数量级估计，说明一阶项和二阶项的量级是同阶的，而低于更高阶项。我们知道由于相互作用的存在，总能量 E 肯定不是单粒子哈密顿量（比如说 $k^2/2m$ 这种形式）的期望值简单加起来的的结果，但是显然能量具有可加性，设想我们改变了准粒子数分布，这样应该有

$$\delta E = \sum_{\mathbf{k}, \sigma} \epsilon_{\mathbf{k}\sigma} \delta n_{\mathbf{k}\sigma},$$

其中 $\epsilon_{\mathbf{k}}$ 是在有限温度下的近平衡态激发一个准粒子的能量，它的一部分是单准粒子能量，一部分是其它准粒子给它的相互作用能之和。因此， $\epsilon_{\mathbf{k}}$ 会依赖准粒子数分布。由于我们只研究二体相互作用，我们有

$$\epsilon \sim \sum_{\mathbf{k}'} \text{something} \cdot n_{\mathbf{k}'},$$

于是设

$$\delta \epsilon_{\mathbf{k}\sigma} = \frac{1}{V} \sum_{\mathbf{k}', \sigma'} f_{\sigma\sigma'\mathbf{k}\mathbf{k}'} \delta n_{\mathbf{k}'\sigma'},$$

记 $\epsilon_{\mathbf{k}}^0$ 为 $n_{\mathbf{k}}$ 一概为零的 $\epsilon_{\mathbf{k}}$ ，代入 δE 中就得到(6.35)；第二项的1/2因子是因为一对粒子会被计数两次，所以要除以2；由于我们假定准粒子分布相对于零温只有微小的偏离， $\epsilon_{\mathbf{k}}$ 被取为零温的值。虽然 $\epsilon^0 \delta n$ 看起来比 $f \delta n \delta n$ 大，但需要注意到我们在巨正则系综中工作，则真的有意义的应该是 $E - \mu N$ （且由于是近平衡，应有 $\delta E = \mu \delta N$ ），而

$$\sum_{\mathbf{k}} (\epsilon_{\mathbf{k}}^0 - \mu) \delta n_{\mathbf{k}} \sim \delta n^2,$$

于是 $\epsilon^0 \delta n$ 项和 $f \delta n \delta n$ 项的贡献是同阶的，都需要考虑。更高阶相互作用则涉及 δn 的更高阶项，于是不考虑。因此，(6.35)的确是成立的。

在已经知道了 E 的表达式之后（比如说微扰计算出了体系能量），可以用变分计算出各个物理量：

$$\epsilon_{\mathbf{k}} = \frac{\delta E}{\delta n_{\mathbf{k}\sigma}}, \quad f_{\sigma\sigma'\mathbf{k}\mathbf{k}'} = V \frac{\delta^2 E}{\delta n_{\mathbf{k}\sigma} \delta n_{\mathbf{k}'\sigma'}}, \quad \mu = \frac{\partial E}{\partial N}. \quad (6.37)$$

(6.35)中的一阶项可以看成是一个等效的单粒子能量。由于只讨论费米面附近的理论，我们让能量从费米面量起，即使用 ξ 代替 ϵ ， $k = k_F$ 时 $\xi_{\mathbf{k}}^0$ 就是零，在假定费米面具有旋转对称性的情况下可以做展开

$$\xi_{\mathbf{k}}^0 = \frac{k_F}{m^*} (k - k_F).$$

我们仿照自由粒子的能量

$$\xi_{\mathbf{k}} = \frac{k^2}{2m} - \frac{k_F^2}{2m} \approx \frac{k_F}{m} (k - k_F)$$

得到了一个等效质量 m^* 。能够像上面这样做的前提是准粒子能谱要足够光滑，如果像声子那样，就没法定义任何等效质量。¹⁰如果温度很高，以至于不能保证有趣的行为仅仅发生在费米面附近，那有效质量的概念也没有什么用；实际上此时费米液体的理论就没有什么用。请注意(6.35)完全是形式上的：诸如晶格势能造成的单粒子能量修正已经被纳入了 δE_1 中，而只要费米面保持旋转对称性，就可以引入等效质量的概念，没有旋转对称性也可以引入有效质量张量。并且，在只有费米面附近才有明显的激发的情况下，可以不失一般性地设

$$\epsilon_{\mathbf{k}}^0 = \frac{k^2}{2m^*},$$

因为真正有意义的是 $\epsilon_{\mathbf{k}} - \mu$ ，只需要同时调节 $\epsilon_{\mathbf{k}}$ 和 μ 就可以让准粒子的 $\epsilon_{\mathbf{k}}$ 取自由粒子的形式。再次强调，调节 $\epsilon_{\mathbf{k}}$ 和 μ 之类的操作只适用于费米面附近；因此对一个费米液体我们通常避免讨论费米球深处有什么——这些东西对费米面附近的行为并不重要。

对二阶项，假定系统具有自旋旋转不变性，则 f 的取值完全由 $f_{\uparrow\uparrow\mathbf{k}\mathbf{k}'}$ 和 $f_{\uparrow\downarrow\mathbf{k}\mathbf{k}'}$ 决定。实际上，由于 \mathbf{k} 局限在费米面附近，我们有

$$f_{\alpha\beta\mathbf{k}\mathbf{k}'} = f_{\alpha\beta}(\vartheta),$$

ϑ 是 \mathbf{k} 和 \mathbf{k}' 的夹角。这样，设

$$\begin{aligned} f_{\uparrow\uparrow}(\vartheta) &= f^S(\vartheta) + f^A(\vartheta), \\ f_{\uparrow\downarrow}(\vartheta) &= f^S(\vartheta) - f^A(\vartheta), \end{aligned} \quad (6.38)$$

或者等价地可以设

$$\hat{f}(\vartheta) = f^S(\vartheta) + \boldsymbol{\sigma}\boldsymbol{\sigma}' f^A(\vartheta), \quad (6.39)$$

从而将自旋守恒这一事实一并表示出来（ σ^z 就是 z 方向的泡利矩阵），并将 $f^S(\vartheta)$ 和 $f^A(\vartheta)$ 展开成无量纲常数：

$$\frac{k_F m^*}{\pi^2} f^{S,A}(\vartheta) = F^{S,A}(\vartheta) = \sum_{l=0}^{\infty} F_l^{S,A} P_l(\cos \vartheta). \quad (6.40)$$

于是，给定参数 m^* ， k_F 以及 $\{F_l^{S,A}\}$ ，费米液体服从的物理规律就给定了。在这里，我们实际上又把准粒子当成了可以彼此散射、有相互作用的粒子，“准粒子动能” $k^2/2m^*$ 和“准粒子势能” $f_{\alpha\beta}(\vartheta)$ 是“单个准粒子能量” $\epsilon_{\mathbf{k}}$ 的两部分；单单一个 $k^2/2m^*$ 肯定和 $\epsilon_{\mathbf{k}}$ 是不一样的。

¹⁰应注意此处的等效质量和“激发有能隙，是有质量的”中的“质量”是不同的：前者并不代表有一个能隙，而只是 $\epsilon_{\mathbf{k}}$ 的 k^2 项的系数而已。

6.3.3 费米液体的参数和裸电子的参数的对应关系

如果一个费米液体系统可以确定是一个实际的费米气体加入相互作用的结果，并且如前所述，能够保证准粒子个数和实际电子的个数一样，自旋相同，等等，并且保证自旋旋转不变性、空间平移不变性、空间各向同性，那么费米液体中准粒子的等效质量和实际电子的质量有一个简单的，使用 $\{F_l^{S,A}\}$ 写出的关系。由于

$$E - E_0 = \sum_{\mathbf{k}} \delta n_{\mathbf{k}} \epsilon_{\mathbf{k}},$$

由动量为 \mathbf{k} 的一个准粒子的运动速度为

$$\mathbf{v}_i = \frac{\partial E}{\partial \mathbf{p}_i} = \frac{\partial \epsilon_{\mathbf{k}}}{\partial \mathbf{k}},$$

上式的量子版本就是

$$\mathbf{v} = \frac{\partial \epsilon_{\mathbf{k}}}{\partial \mathbf{k}}.$$

我们于是可以将 \mathbf{v} 当成费米液体中准粒子的流速。设 ρ 是某种守恒荷的密度，则任意一个流算符的期望可以写成

$$\langle \rho \mathbf{v} \rangle = V \int \frac{d^3 \mathbf{k}}{(2\pi)^3} \left\langle \rho \frac{\partial \epsilon}{\partial \mathbf{k}} \right\rangle,$$

准粒子和实际的费米子数量相同，准粒子的粒子数流密度算符就是实际费米子的粒子数流密度算符，且由于动量守恒，一个准粒子的总动量就是与它对应的实际费米子的总动量，而实际费米子的总动量就是总质量流（因为 $\mathbf{p} = m\mathbf{v}$ ）；由于动量和自旋守恒，我们将费米子占据数算符用 \mathbf{k} 和 σ 标记，于是我们有

$$\sum_{\mathbf{k}, \sigma} \int \frac{d^3 \mathbf{k}}{(2\pi)^3} \mathbf{k} n_{\mathbf{k}\sigma} = \text{Tr} \int \frac{d^3 \mathbf{k}}{(2\pi)^3} m \frac{\partial^2 \epsilon}{\partial \mathbf{k} \partial n}.$$

同样 ϵ 也适用一样的推导，于是就有

$$\int \frac{d^3 \mathbf{k}}{(2\pi)^3} m \frac{\partial \epsilon_{\mathbf{k}\sigma}}{\partial \mathbf{k}} n_{\mathbf{k}\sigma} = \int \frac{d^3 \mathbf{k}}{(2\pi)^3} \mathbf{k} n_{\mathbf{k}\sigma},$$

对上式求变分，就有

$$\begin{aligned} \int \frac{d^3 \mathbf{k}}{(2\pi)^3} \mathbf{k} \delta n_{\mathbf{k}\sigma} &= \delta \int \frac{d^3 \mathbf{k}}{(2\pi)^3} m \frac{\partial \epsilon_{\mathbf{k}\sigma}}{\partial \mathbf{k}} n_{\mathbf{k}\sigma} \\ &= m \int \frac{d^3 \mathbf{k}}{(2\pi)^3} \frac{\partial \epsilon_{\mathbf{k}\sigma}}{\partial \mathbf{k}} \delta n_{\mathbf{k}\sigma} + m \int \frac{d^3 \mathbf{k}}{(2\pi)^3} \int \frac{d^3 \mathbf{k}'}{(2\pi)^3} n_{\mathbf{k}\sigma} \delta n_{\mathbf{k}'\sigma} \frac{\partial f_{\sigma}(\vartheta)}{\partial \mathbf{p}} \\ &= m \int \frac{d^3 \mathbf{k}}{(2\pi)^3} \frac{\partial \epsilon_{\mathbf{k}\sigma}}{\partial \mathbf{k}} \delta n_{\mathbf{k}\sigma} - m \int \frac{d^3 \mathbf{k}}{(2\pi)^3} \int \frac{d^3 \mathbf{k}'}{(2\pi)^3} \delta n_{\mathbf{k}\sigma} \frac{\partial n_{\mathbf{k}'\sigma}}{\partial \mathbf{k}'} f_{\sigma}(\vartheta). \end{aligned}$$

第三个等号交换了 \mathbf{k} 和 \mathbf{k}' ，但这是合理的，因为 f 只和这两者的夹角有关。考虑到 δn 的任意性，就有

$$\frac{\mathbf{k}}{m} = \frac{\partial \epsilon_{\mathbf{k}\sigma}}{\partial \mathbf{k}} - \int \frac{d^3 \mathbf{k}'}{(2\pi)^3} \frac{\partial n_{\mathbf{k}'\sigma}}{\partial \mathbf{k}'} f_{\sigma}(\vartheta).$$

在上式两边点乘 \mathbf{k} ，代入 n 是阶跃函数这一事实，并且注意到动量几乎总是在费米面上，从而 $\mathbf{k} = k_F \mathbf{n}$ ，就得到

$$\frac{1}{m} = \frac{1}{m^*} + \frac{k_F}{(2\pi)^3} \int d\Omega \cos \vartheta f_{\sigma}(\vartheta). \quad (6.41)$$

上式的形式似乎有些奇怪——它似乎是在说“ f 和 m 决定了 m^* ”。但这种理解实际上是正确的： f 对应于重整化后的电子-电子散射顶角¹¹，而 m^* 和自能修正有关，而顶角修正和自能修正之间的关系不是任意的：事实上，通过顶角修正能够计算出自能，于是确实是通过 f 和 m 确定 m^* 。这里还有一个疑难，就是在使用电子动量计算总动量时我们直接用了 $p = mv$ ，而以费米液体的观点计算总动量时我们却没有这么做。这是因为，只是根据(6.35)计算出的动量并不是真正的总动量，庞大的费米海的动量被忽略了：当一个费米液体中的准粒子被激发起来时，实际的系统中的费米海会受到扰动，从而会有额外的动量。 f 通常和费米能级有关，因此可以提供一些关于“费米海有多重”的信息，这就是(6.41)中会出现 f 的原因：(6.41)来自动量守恒关系，动量守恒方程的一边含有 m ，另一边含有 m^* 和通过 f 反解出的费米海的动量，经过化简就得到(6.41)。当然，如果实际系统中根本没有电子间排斥，那么 f 肯定一直是零。

6.3.4 费米液体的平衡态宏观性质

使用以上参数： m^* ， k_F 以及 $\{F_l^{S,A}\}$ ，可以计算费米液体的各种宏观性质。

首先考虑零温附近的比热。费米气体的比热在低温极限下正比于温度，费米液体实际上也一样。能量由(6.35)给出，随着 T 增大，一些粒子从费米海溢出，从而能量增大，产生一个热容。实际上，在零温极限附近，(6.35)中的 E_2 部分没有贡献。这是因为

$$E_2 = \sum_{\sigma, \sigma'} \underbrace{\frac{1}{2V} \sum_{\mathbf{k}} f_{\sigma\sigma'}(\theta) \delta n_{\mathbf{k}\sigma} \delta n_{\mathbf{k}'\sigma'}}_{\text{constant}},$$

被大括号括起来的部分和 \mathbf{k} 无关，而显然

$$\sum_{\mathbf{k}} \delta n_{\mathbf{k}\sigma} = 0,$$

因此 E_2 对总能量没有贡献。这样费米液体的热容和费米气体的热容就是完全一致的，为

$$C_V = \frac{1}{3} m^* k_F T. \quad (6.42)$$

这个公式在实验上非常重要，如果确定一个体系是费米液体（如发现低温下热容正比于温度），那么就可以据此测出粒子的有效质量。

也可以计算费米液体的磁化率。考虑弱场近似，则磁化率

$$\chi = \frac{\partial M}{\partial H}$$

近似为

$$\chi = \frac{M}{H},$$

其中 M 表示磁化强度， H 表示磁场强度（不是哈密顿量），而磁化强度为

$$M = \frac{\partial E}{\partial H},$$

于是得到

$$\frac{1}{\chi} = \frac{\partial^2 E}{\partial M^2}.$$

¹¹ 此处“顶角”一词指的是二电子进、二电子出的费曼图部件；有些文献遵从高能物理的习惯，将做在库仑相互作用线上的修正称为真空极化，而将做在电子线和库仑相互作用线连接部的修正称为顶角修正。此处和 f 有关的顶角修正是后一种定义中的真空极化和顶角修正的乘积。

这样只需要使用 M 表示出 E 就可以了。记自旋向上（以磁场方向为 z 轴）和向下的粒子数为 N_{\uparrow} 和 N_{\downarrow} ，则

$$M = \mu_B(N_{\uparrow} - N_{\downarrow}),$$

其中 μ_B 为玻尔磁子。磁场导致自旋向上和向下的粒子数发生变化的原因是，自旋和磁场一致的粒子的费米面会扩大，自旋和磁场相反的粒子的费米面会缩小，从而让 N_{\uparrow} 变大， N_{\downarrow} 缩小。由于粒子数不变，有

$$\delta N_{\uparrow} = -\delta N_{\downarrow},$$

而没有磁场时向上和向下的粒子数一样，于是

$$M = 2\mu_B \delta N_{\uparrow}.$$

δN_{\uparrow} 和费米动量的变化之间的关系是

$$\delta N_{\uparrow} = \int_{k_F < k < k_F + \delta k_F} \frac{V}{(2\pi)^3} d^3\mathbf{k} = \frac{V k_F^2 \delta k_F}{2\pi^2}.$$

现在可以将 M 用 δk_F 表示出来了。接下来将能量写成 δk_F 的函数。对动能部分 E_1 ，我们有

$$\delta E_1 = \sum_{\sigma, \mathbf{k}} \frac{k_F}{m^*} (k - k_F) \delta n_{\mathbf{k}\sigma},$$

$n_{\mathbf{k}\uparrow}$ 仅有的变化是在 $k_F < k < k_F + \delta k_F$ 的区域内从0变成1， $n_{\mathbf{k}\downarrow}$ 仅有的变化是在 $k_F - \delta k_F < k < k_F$ 的区域内从1变成0。这样就有

$$\begin{aligned} \delta E_1 &= \int_{k_F < k < k_F + \delta k_F} \frac{V}{(2\pi)^3} d^3\mathbf{k} \frac{k_F}{m^*} (k - k_F) + \int_{k_F - \delta k_F < k < k_F} \frac{V}{(2\pi)^3} d^3\mathbf{k} \frac{k_F}{m^*} (k - k_F)(-1) \\ &= \frac{V k_F^3}{2\pi^2 m^*} (\delta k_F)^2. \end{aligned}$$

最后，得到 δE_1 和 M 的关系：

$$\delta E_1 = \frac{\pi^2}{2m^* \mu_B^2 V k_F} M^2.$$

同理，可以计算得到（计算的关键点在于意识到对全空间计算积分，则只有零阶勒让德多项式能够给出非零结果）

$$\delta E_2 = \frac{\pi^2}{2m^* \mu_B^2 V k_F} F_0^A M^2.$$

这样就得到了 δE 关于 M 的表达式，从而最终得到

$$\chi = \frac{1}{1 + F_0^A} \frac{m^* \mu_B^2 V k_F}{\pi^2}. \quad (6.43)$$

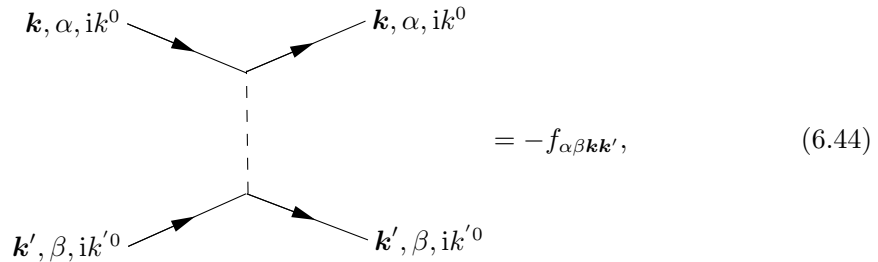
6.3.5 等效单粒子图像，动理学和费米液体的集体模式

单粒子自能修正和动理学方程

现在我们开始分析费米液体的动态响应。应当注意(6.35)在稍微远离费米面的地方就失效了，因为此时真的能够改变准粒子动量的碰撞是需要考虑的。在碰撞有物理效应时，不能使用(6.35)作为哈密顿量而计算格林函数。实际上，这也体现了“纯净的”费米液体哈密顿量(6.35)虽然表面上看非常合乎直觉（例如它可以通过实际的库仑相互作用电子气的Hartree-Fock近似得到辩护，见第6.3.8节），但是有非常反直觉的地方：我们知道碰撞会导致液体的粘滞（见流体和软物质物理笔记中第2.1.4节和第3.2.3节），而库仑相互作用电

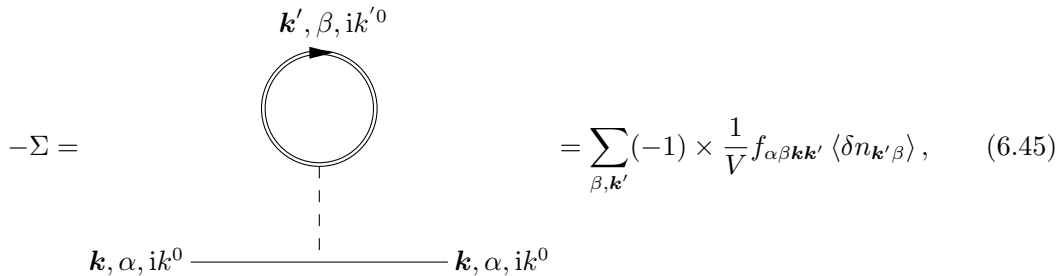
子气中库仑相互作用能的量级和电子动能是同阶的，那么似乎库仑相互作用电子气演生出的液体态（是否相互作用电子气的低能有效理论真的和通常意义上的液体一样是需要辩护的，不过后面我们会看到费米液体有动理学方程，从而它确实是液体）应该是极为粘稠的。然而，从前面的推导过程可以看出，实际上费米液体中准粒子之间的碰撞并没有那么强烈，正是在这个假设下我们导出了(6.35)。这也产生了一个问题：是否存在这样的电子态，它大体上仍然是一种费米液体，但是有明显的粘滞，从而其行为和普通的分子液体非常类似，遵从带有粘滞项的纳维-斯托克斯方程（例如，在狭窄的管道中的层流中速度场分布为抛物线）？文献^[11]中讨论了二维电子气中电子的流体态。

一种替代的方式是使用玻尔兹曼方程。需要注意玻尔兹曼方程需要单粒子能量。(6.35)中的两项的量级是一样的，从而单粒子能量不能取为 ϵ_k^0 ；但(6.35)中的第二项也不能纳入碰撞项。玻尔兹曼方程（以及很多其它动理学方程）是格林函数运动方程的近似，玻尔兹曼方程中的单粒子分布函数实际上是单粒子格林函数（“单粒子密度矩阵”）的一种简化，因此(6.35)中的第二项应该通过计算电子自能加以考虑。(6.35)中的第二项在动量空间中给出



$$= -f_{\alpha\beta k k'}, \quad (6.44)$$

由于是前向散射，入射出射准粒子动量不变，虚线上没有三维动量交换，松原频率交换由于是瞬时相互作用也为零。它给出的全部正规自能图可以概括为



$$-\Sigma = \sum_{\beta, \mathbf{k}'} (-1) \times \frac{1}{V} f_{\alpha\beta k k'} \langle \delta n_{\mathbf{k}'\beta} \rangle, \quad (6.45)$$

其中闭合电子回线意味着 $i\omega_n = 0$ ，-1因子来自相互作用线，闭合费米子回线由于给出了 $\langle c^\dagger c \rangle$ ，不贡献-1因子。Fock项仅仅在 $\mathbf{k} = \mathbf{k}'$ 时有非零值；可是实际上(6.44)也并不是完整的费曼规则，因为它对应的哈密顿量形如 $f c^\dagger c^\dagger c c$ ，而不是费米液体能量泛函中的 $f c^\dagger c c^\dagger c$ ；由于费米子算符的交换反对易性，实际上在费曼规则中还会多出一个形如 $f_{\mathbf{k}\mathbf{k}} \delta n_{\mathbf{k}}$ 的“自能修正”，它和Fock项正好抵消了：理应如此，因为我们可以把(6.44)中上部的两条电子线看成一条玻色线，下部的两条电子线看成一条玻色线，那么单独(6.44)一条规则以及“Fock项不计入考虑”的约束就很合理了。总之，加入了自能修正的单粒子能量就是

$$\epsilon_{\mathbf{k}\alpha} = \epsilon_{\mathbf{k}}^0 + \frac{1}{V} \sum_{\beta, \mathbf{k}'} f_{\alpha\beta k k'} \langle \delta n_{\mathbf{k}'\beta} \rangle. \quad (6.46)$$

期望值 $\langle n_{\mathbf{k}'\beta} \rangle$ 是考虑了(6.35)中第二项之后的，本身不容易直接算出，它将被用于做自洽计算。实际上也可以更加直观地得到(6.46)： $\epsilon_{\mathbf{k}\alpha}$ 大体上是向费米液体基态打入一个新准粒子后的能量，那让 E 对 $n_{\mathbf{k}\alpha}$ 求导就是了。

以上费米液体的Hartree-Fock近似当然也可以通过对(6.35)做关于粒子数的平均场近似更加直观地得到。Fock项不重要这个事实意味着费米液体实际上是一个比较经典的模型。实际上,我们完全可以将哈密顿量中的那些粒子数算符都替换成(计算了期望值之后的)粒子分布函数。只有Hartree项也意味着费米液体主要的涨落应该是粒子数密度的涨落(因为显然此时我们可以将费曼图中的全部闭合粒子回线替换成电子数密度场),即它的宏观行为可以使用某种流体动力学来描述。

现在列写玻尔兹曼方程。设单粒子分布函数为 $n_{\mathbf{k}\alpha}(\mathbf{r}, t)$ ¹², 玻尔兹曼方程为

$$\frac{\partial n_{\mathbf{k}\alpha}}{\partial t} + \frac{\partial n_{\mathbf{k}\alpha}}{\partial \mathbf{r}} \cdot \frac{\partial \epsilon_{\mathbf{k}\alpha}}{\partial \mathbf{k}} + \frac{\partial n_{\mathbf{k}\alpha}}{\partial \mathbf{k}} \cdot \left(\mathbf{F} - \frac{\partial \epsilon_{\mathbf{k}\alpha}}{\partial \mathbf{r}} \right) = I[n_{\mathbf{k}\alpha}]. \quad (6.47)$$

这个方程形式上和普通的玻尔兹曼方程一样,但是 ϵ 和 n 之间存在关系(6.46)。注意此时 $\epsilon_{\mathbf{k}\alpha}$ 也有 \mathbf{r} 的依赖,且(6.46)中的二次量子化粒子数算符 $\delta n_{\mathbf{k}'\beta}$ 被单粒子分布函数偏离平衡态的量 $\delta n_{\mathbf{k}'\beta}$ 取代,即此时我们就只需要写出

$$\epsilon_{\mathbf{k}\alpha} = \epsilon_{\mathbf{k}}^0 + \frac{1}{V} \sum_{\beta, \mathbf{k}'} f_{\alpha\beta\mathbf{k}\mathbf{k}'} \delta n_{\mathbf{k}'\beta}. \quad (6.48)$$

有时我们将(6.47)联立上(6.48)称为**朗道方程**。朗道动理学方程的稳态或者说费米液体的热力学基态当然形式上和费米气体是完全一样的,是

$$n_{\mathbf{k}\alpha} = \frac{1}{e^{\beta\epsilon_{\mathbf{k}\alpha}} + 1}, \quad (6.49)$$

但同样需要联立(6.46)才能求解。应注意(6.49)中的 $\epsilon_{\mathbf{k}\alpha}$ 依赖于 $n_{\mathbf{k}\alpha}$,而且就依赖于(6.49)给出的这个 $n_{\mathbf{k}\alpha}$ 。这是符合直觉的,但表面上看似乎存在理论上的难题,即能谱似乎应该是一个完全纯态的概念,不应该有温度依赖,但应当注意到(6.45)在有限温度场论中同样适用,而它当中的费米子圈正是(6.49)。事实上,完全可以从(6.49)和(6.48)的联立出发来建立费米液体理论,这正是朗道十卷的凝聚态理论卷卷首的处理方法,虽然对不熟悉有限温度场论中的自能修正的读者来说这样的处理方法似乎无法理解。

有关此处的玻尔兹曼方程有必要强调一点:实际上,由于我们考虑了 $n_{\mathbf{k}}$ 会有空间依赖,却与此同时仍然承认(6.48)的正确性,实际上我们已经没有再使用(6.35),而是已经往其中加入了一定的非前向散射了。要看出为什么这么说,注意到

$$n_{\mathbf{k}}(\mathbf{r}) \simeq \int d^3\mathbf{x} e^{-i\mathbf{x}\cdot\mathbf{k}} \int d^3\mathbf{k}_1 \int d^3\mathbf{k}_2 e^{i\mathbf{x}\cdot\frac{\mathbf{k}_1+\mathbf{k}_2}{2}} e^{i(\mathbf{k}_1-\mathbf{k}_2)\cdot\mathbf{r}} G(\mathbf{k}_1, \mathbf{k}_2), \quad (6.50)$$

其中格林函数中的 t_1 和 t_2 需要做正确的无穷小位移,使得最后能够给出粒子数期望值。现在如果要让(6.48)成立,并且 f 作用到 $n_{\mathbf{k}}(\mathbf{r})$ 上面以后输出 $n_{\mathbf{k}'}(\mathbf{r})$,则准粒子相互作用顶角应当形如

$$\int d^3\mathbf{r} e^{i\mathbf{r}\cdot(\mathbf{k}_3-\mathbf{k}_4)} \delta\left(\frac{\mathbf{k}_3+\mathbf{k}_4}{2} - \mathbf{k}'\right) f_{\mathbf{k}\mathbf{k}'} \delta\left(\frac{\mathbf{k}_1+\mathbf{k}_2}{2} - \mathbf{k}\right) e^{i\mathbf{r}\cdot(\mathbf{k}_1-\mathbf{k}_2)},$$

而这显然不是一个真正的前向散射相互作用。实际上,只有基于(6.48),并恰当地考虑了准粒子碰撞(玻尔兹曼方程中的碰撞积分来自格林函数中的自能修正的虚部,一般而言不是零)的玻尔兹曼方程——或等价地说,电子格林函数运动方程——才是完整的费米液体理论;能量泛函(6.35)仅仅是它的一个低能有效理论而已。(不过,由于动量弥散并不太大, $n_{\mathbf{k}}(\mathbf{r})$ 随着 \mathbf{r} 变化得足够平缓,直接将(6.35)中的 f 函数代入(6.48)中是合理的;因此能量

¹²其实 \mathbf{r} 和 \mathbf{k} 的不对易比较明显时,也可以定义单粒子分布函数的,只不过此时它是Wigner函数,是单粒子密度矩阵做了一个傅里叶变换而得到的;在 \mathbf{r} 和 \mathbf{k} 的不对易性不明显时,经典概率分布函数、Wigner函数、密度矩阵基本上就是一回事,我们也不再区分它们。

泛函表述和玻尔兹曼方程表述携带了同样多的信息，虽然玻尔兹曼方程表述并不能直接通过将(6.35)代入运动方程得到，而是要将其中的 f 函数适当地延拓到非前向散射以后才能得到。)的确，我们下面将证明费米液体中存在一种称为零声的玻色型激发，并且将要导出零声的色散，却从来没有将它引入到热容的计算中：因为我们此处（根据(6.35)）处理的并非真正完整的费米液体的热容，而是其低温极限，在这个低温极限中零声没有贡献。零声由于是线性色散，其热容遵从德拜 T^3 定律，而准电子贡献的热容则正比于 T ，因此低温下忽略零声的贡献的确是合理的。

下面我们要根据电子的玻尔兹曼方程的振动模式寻找电子-电子或电子-空穴束缚而成的玻色型激发。这是一个普遍适用的方法，因为玻尔兹曼方程的振动模式就是 $n_{\mathbf{k}}(\mathbf{r})$ 对外场的响应函数的极点，而通过费曼图方法可以证明， $n_{\mathbf{k}}(\mathbf{r})$ ——实际上就是单电子格林函数——对外场的响应正是由基态上的二电子格林函数给出。因此，玻尔兹曼方程的振动模式对应于二电子格林函数的极点，从而对应二电子构成的玻色型激发的能谱。再次强调，此处的玻尔兹曼方程并没有在描述基态行为，而是在描述激发态行为；我们可以想象， $n_{\mathbf{k}}(\mathbf{r})$ 之所以在振荡，是因为过去某个时间段我们用电场去扰动了费米液体。

费米液体中的集体振荡模式

由于动理学方程实际上是准粒子两点格林函数的运动方程的一个近似，它给出的 $n_{\mathbf{k}\alpha}$ 如果随着时间有某种振荡模式，做完傅里叶变换就会有极点，即存在一个准粒子集体运动形成的元激发。通过分析动理学方程的稳定周期解，就能够找到很大一类费米液体中的元激发。

我们采用弛豫时间近似，即

$$\frac{\partial n_{\mathbf{k}\alpha}}{\partial t} + \frac{\partial n_{\mathbf{k}\alpha}}{\partial \mathbf{r}} \cdot \frac{\partial \epsilon_{\mathbf{k}\alpha}}{\partial \mathbf{k}} + \frac{\partial n_{\mathbf{k}\alpha}}{\partial \mathbf{k}} \cdot \left(\mathbf{F} - \frac{\partial \epsilon_{\mathbf{k}\alpha}}{\partial \mathbf{r}} \right) = -\frac{n_{\mathbf{k}\alpha} - n_{\mathbf{k}\alpha}^0}{\tau}, \quad (6.51)$$

由于弛豫是粒子碰撞导致的， $\tau \propto T^{-2}$ 。注意(6.47)的碰撞项中是**没有**温度依赖的，并且大体上正比于 $n_{\mathbf{k}\alpha}$ 的平方；然而， $n_{\mathbf{k}\alpha}$ 本身是有温度依赖的，在系统偏离平衡态不多时，可以将碰撞项写成 $n_{\mathbf{k}\alpha}/\tau$ 这样的弛豫项，此时的 τ 就是有温度依赖的。形象地看，弛豫项 $n_{\mathbf{k}\alpha}/\tau$ 会把准粒子分布“拖”向平衡态。

系统状态偏离平衡态首先会导致的就是(6.48)中的 n 不再是由(6.49)和(6.48)联合确定的那个费米分布状态，然后玻尔兹曼方程(6.51)又会根据(6.48)算出的 ϵ 改变 n 。我们采取如下求解策略：设 $n_{\mathbf{k}\alpha}^0$ 为将实际的（暂时还不确定的） $n_{\mathbf{k}\alpha}$ 代入(6.48)，将得到的真实 $\epsilon_{\mathbf{k}\alpha}$ 代入(6.49)右边计算出来的分布函数，然后将 $\hat{n}_{\mathbf{k}}$ 在 $\hat{n}_{\mathbf{k}}^0$ 附近展开到平衡态附近的一阶（由于只考虑近平衡态），得到将(6.51)线性化的关于

$$\delta n_{\mathbf{k}\alpha} = n_{\mathbf{k}\alpha} - n_{\mathbf{k}\alpha}^0 \quad (6.52)$$

的演化方程。注意到 $n_{\mathbf{k}\alpha}^0$ 不依赖于时间和坐标，而 $\epsilon_{\mathbf{k}\alpha}$ 不依赖于坐标，仅保留正比于 $\delta n_{\mathbf{k}\alpha}$ 的项，得到

$$\frac{\partial \delta n_{\mathbf{k}\alpha}}{\partial t} + \frac{\partial \epsilon_{\mathbf{k}\alpha}^0}{\partial \mathbf{k}} \cdot \frac{\partial \delta n_{\mathbf{k}\alpha}}{\partial \mathbf{r}} - \frac{\partial n_{\mathbf{k}\alpha}^0}{\partial \mathbf{k}} \cdot \frac{\partial \delta \epsilon_{\mathbf{k}\alpha}}{\partial \mathbf{r}} = -\frac{\delta n_{\mathbf{k}\alpha}}{\tau}. \quad (6.53)$$

(6.51)描述的系统能够传播的波大致上可以分成两种，一种是普通的声波，一种是所谓的零声。设波动的圆频率为 ω ，如果 $\omega\tau \ll 1$ ，那么准静态近似成立，即有波动传递的费米液体的每一点几乎都是热力学平衡的。这样的波动就是普通的声波（虽然寻常的传递声波的液体中的粒子没有费米统计而费米液体中的粒子有，但两者都可以在适当的情况下使用玻尔兹曼方程描述，从而都有声波）。应注意在 ω 很小时并无明显的损耗，因为此时一点

的振动有充足的时间传播到下一点，此时的声波可以认为几乎是没有任何损耗的。 $\omega\tau$ 稍微大一些时，准静态近似不再成立，在热力学上这意味着会有明显的熵产生，换句话说，在费米液体中激发出的任何振动模式都会快速因为碰撞而受到阻尼，没有能够稳定传播的模式。在温度非常低时，基本上没有声波能够传递。这是正确的，因为声波的传播需要准粒子碰撞，而温度非常低时没有什么碰撞事件发生。¹³

但这并不是说 $T \rightarrow 0$ 时费米液体中没有任何集体振动模式。当 $\omega\tau \gg 1$ 时，一幅非常奇怪的图景出现了：此时弛豫项几乎可以忽略，因此，外界扰动而产生的 $n_{\mathbf{k}\alpha}$ 涨落传播到的地方不会发生任何弛豫现象，无法达到热平衡，且波动不会受到任何阻尼。此时的波动称为**零声**，因为它看起来很像声波，但首先频率通常高于普通的声波，其次通常需要 τ 比较大才能观察到，即需要在非常接近绝对零度的系统中才能观察到。

我们还要指出，费米液体有玻尔兹曼方程的动理学方程这件事意味着费米液体准粒子的位移的宏观理论是某种类似于纳维-斯托克斯方程的东西（见流体和软物质物理笔记中第2.1.4节和第3.2.3节），因此费米液体真的可以看成一种液体。出于这样的原因，凝聚态物理中通常将有明确的准粒子激发，并且相互作用不甚重要的体系称为气体（如“声子气体”），将有明确的准粒子激发，而相互作用比较重要，但是其中仍然有良定义的准粒子的体系称为液体（如“费米液体”——需要注意费米液体中相互作用主要是通过自能而不是通过玻尔兹曼方程的碰撞项引入的），尽管这里的“粒子”和通常意义上的原子、分子等粒子可以很不一样。除了费米液体以外我们还会看到其它“液体”，如Luttinger液体，自旋液体等等。一个量子液体可以经历对称性自发破缺；一些情况下这会导致原本被识别为准粒子的自由度被局域化；例如，分子、原子液体经历这样的过程以后形成固体；此时在原来的准粒子的表象下，这个体系就不再是“液体”了，但是新的低能自由度——如声子——仍然能够形成另一种液体或者气体。

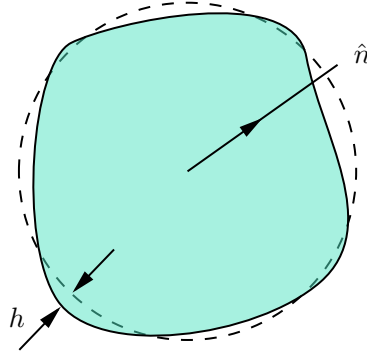
一个对“液体”或者准确地说“量子液体”的更加宽泛的定义是任何没有空间中的长程序的系统。一个有长程序的系统在加入更多格点之后，我们能够将新加入的格点带来的自由度增多定位到一个确定的格点上（如晶体中加入原子就加入了这个原子的位移自由度），但是一个没有长程序的系统加入更多格点后，新的自由度“溶解”在了系统中，我们无法将新的自由度定位到确定的空间点上，正如一滴水滴入海洋中后我们无法将新增的自由度定位到任何一个地方，而只能泛泛地说“系统体积增大了，流场范围增大了”。这样的液体的基本自由度到底是什么则是没有限制的：它可以就是点粒子，此时这些点粒子可以自由移动，我们就得到了前文中定义的那种液体；它也可以是更加复杂的自由度，比如说弦网——我们将在第十九章中讨论这种系统。

6.3.6 零声

费米液体在绝对零度的低能激发

我们来计算 $T = 0$ 时的零声模式，实际上就是要处理 $T = 0$ 时的费米液体的低能激发。此时(6.53)右边为零。我们考虑一个 $f_{\alpha\beta\mathbf{k}\mathbf{k}'}$ 中无自旋依赖的系统，因为与自旋相关的那种涨落通常称为自旋密度波。为方便起见我们将 \mathbf{k} 给定而携带一个自旋指标的分布函数记作 $\hat{n}_{\mathbf{k}}$ ，同理 f 记作 $\hat{f}_{\mathbf{k}\mathbf{k}'}$ 。

¹³这里“零温”一词可能引起混淆；我们不是说普通声波是一个纯粹有限温的现象——它当然可以在纯态（“零温”）系统中发生。但是，对处在热态中而温度为零（另一种“零温”）的系统——即状态和费米球相差不大的系统——声波是很难产生的。



(a) 变形的费米面，费米液体在 $T = 0$ 时的低能激发普遍具有这个形式

图 6.5: 费米面的变形和零声

我们有

$$\frac{\partial \hat{n}_{\mathbf{k}}^0}{\partial \mathbf{k}} = -\delta(\epsilon_F - \epsilon_{\mathbf{k}}) \mathbf{v}, \quad \frac{\partial \epsilon_{\mathbf{k}}^0}{\partial \mathbf{k}} \approx \mathbf{v}, \quad \delta \epsilon_{\mathbf{k}} = \frac{1}{V} \text{Tr}' \sum_{\mathbf{k}'} \hat{f}_{\mathbf{k}\mathbf{k}'} \delta \hat{n}_{\mathbf{k}'} = \text{Tr}' \int \frac{d^3 \mathbf{k}'}{(2\pi)^3} \hat{f}_{\mathbf{k}\mathbf{k}'} \delta \hat{n}_{\mathbf{k}'},$$

这里我们使用惯常的群速度定义

$$\mathbf{v} = \frac{\partial \epsilon_{\mathbf{k}}}{\partial \mathbf{k}} \approx \frac{\partial \epsilon_{\mathbf{k}}^0}{\partial \mathbf{k}} = \frac{\mathbf{k}}{m^*} \approx v_F \hat{\mathbf{k}}, \quad (6.54)$$

因此 \mathbf{v} 和 \mathbf{k} 的指向是一致的，记它们的指向为 $\hat{\mathbf{n}}$ 。这样，(6.53) 化为

$$\frac{\partial \delta \hat{n}_{\mathbf{k}}}{\partial t} + \mathbf{v} \cdot \frac{\partial \delta \hat{n}_{\mathbf{k}}}{\partial \mathbf{r}} + \delta(\epsilon_F - \epsilon_{\mathbf{k}}) \mathbf{v} \cdot \text{Tr}' \int \frac{d^3 \mathbf{k}'}{(2\pi)^3} \hat{f}_{\mathbf{k}\mathbf{k}'} \frac{\partial \delta \hat{n}_{\mathbf{k}'}}{\partial \mathbf{r}} = 0. \quad (6.55)$$

注意到上式具有时间和空间平移不变性，即可以做傅里叶变换，且第三项中的 $\delta(\epsilon_F - \epsilon_{\mathbf{k}})$ 无法通过对 \mathbf{r} 或是 t 求导得到，于是不失一般性地，可以 $\delta \hat{n}_{\mathbf{k}}$ 具有以下形式

$$\delta \hat{n}_{\mathbf{k}} = e^{i(\mathbf{q} \cdot \mathbf{r} - \omega t)} \delta(\epsilon_F - \epsilon_{\mathbf{k}}) \hat{\nu}_{\mathbf{k}},$$

ω 和 \mathbf{q} 分别是零声的频率和波矢， $\hat{\nu}$ 不依赖于 \mathbf{r} 。此外，由于我们只关心费米面附近的行为，可以认为 \mathbf{k} 提供的信息主要是关于其指向而不是其大小的（大小就是 k_F ；由于(6.55)中没有对 \mathbf{k} 的导数，无需考虑费米面附近 $\delta \hat{n}_{\mathbf{k}}$ 的变化趋势），从而

$$\delta \hat{n}_{\mathbf{k}}(\mathbf{r}) = e^{i(\mathbf{q} \cdot \mathbf{r} - \omega t)} \delta(\epsilon_F - \epsilon_{\mathbf{k}}) \hat{\nu}(\hat{\mathbf{n}}). \quad (6.56)$$

注意(6.55)中 \mathbf{r} 具有空间各向同性而 \mathbf{k} 未必具有（ f 可以破坏动量空间的各向同性）。

(6.55) 的形式看起来很奇怪，不过实际上这说明费米液体中如果有波矢为 \mathbf{q} ，频率为 ω 的零声存在，那么仍然有一个费米面，只不过费米面的形状发生了扭曲。我们反过来，通过指出费米面形状扭曲以后的分布函数真的是(6.55)的解来论证这件事。 $T = 0$ 时的分布函数满足

$$\delta \hat{n} \propto \Theta(\epsilon_F - \epsilon_{\mathbf{k}}) \approx \Theta\left(\frac{k_F}{m^*}(k(\hat{\mathbf{n}}) - k_F)\right),$$

设费米面在方向 $\hat{\mathbf{n}}$ 上发生了按照动量来说，大小为 $h(\hat{\mathbf{n}})$ 的位移，则此时的分布函数为

$$\hat{n} \propto \theta\left(\frac{k_F}{m^*}(k(\hat{\mathbf{n}}) - k_F - h(\hat{\mathbf{n}}))\right) \approx \Theta(\epsilon_F - \epsilon_{\mathbf{k}}) + \frac{k_F}{m^*} \delta(\epsilon_F - \epsilon_{\mathbf{k}}) h(\hat{\mathbf{n}}),$$

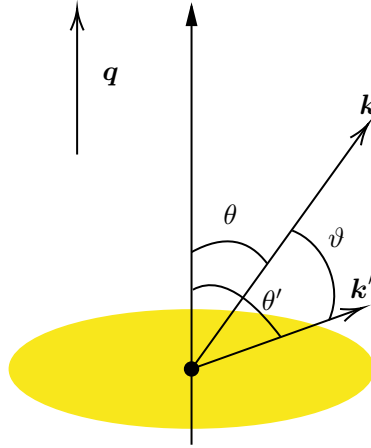


图 6.6: 计算过程中建立的坐标系, θ 和 θ' 是 \mathbf{k} 和 \mathbf{k}' 的极角, ϑ 是这两者之间的夹角

即

$$\delta\hat{n} \propto \delta(\epsilon_F - \epsilon_{\mathbf{k}})h(\hat{\mathbf{n}}).$$

这和(6.56)的形式完全一样。 $\delta\hat{n}$ 随着时间变化会振动,即 ω 给出了这种振动的频率为 ω 的分量而 $\hat{\nu}(\hat{\mathbf{n}})$ 其实给出了频率 ω 的分量的振幅。费米液体中有零声模式时,费米面就以不同频率的 $\hat{\nu}(\hat{\mathbf{n}})$ 叠加而成的位移做膨胀-收缩的“呼吸”运动。作为对比,费米液体中有普通声波时,粒子速度周期性地增加和减少,从而对应的费米面运动模式是费米面保持体积不变,位置来回移动。

将(6.56)代入(6.55),得到

$$(\omega - \mathbf{v} \cdot \mathbf{q})\hat{\nu}(\hat{\mathbf{n}}) = \mathbf{v} \cdot \mathbf{q} \text{Tr}' \int \frac{d^3\mathbf{k}'}{(2\pi)^3} \hat{f}_{\mathbf{k}\mathbf{k}'} \delta(\epsilon_{\mathbf{k}'} - \epsilon_F) \hat{\nu}(\hat{\mathbf{n}}').$$

上式右边只和 \mathbf{k}' 在费米面附近处时 $f_{\mathbf{k}\mathbf{k}'}$ 的值有关,我们不妨设 \mathbf{k} 和 \mathbf{k}' 均在费米面附近的 $f_{\mathbf{k}\mathbf{k}'}$ 为 $f(\hat{\mathbf{n}}, \hat{\mathbf{n}}')$,则有

$$\text{Tr}' \int \frac{d^3\mathbf{k}'}{(2\pi)^3} \hat{f}_{\mathbf{k}\mathbf{k}'} \delta(\epsilon_{\mathbf{k}'} - \epsilon_F) \hat{\nu}(\hat{\mathbf{n}}') = \frac{1}{(2\pi)^3} \text{Tr}' \int k_F^2 d\Omega' \frac{1}{|\frac{\partial \epsilon_{\mathbf{k}}}{\partial \mathbf{k}}|} \hat{f}(\hat{\mathbf{n}}, \hat{\mathbf{n}}') \hat{\nu}(\hat{\mathbf{n}}'),$$

因此

$$(\omega - v_F \hat{\mathbf{n}} \cdot \mathbf{q})\hat{\nu}(\hat{\mathbf{n}}) = \hat{\mathbf{n}} \cdot \mathbf{q} \frac{k_F^2}{(2\pi)^3} \text{Tr}' \int d\Omega' \hat{f}(\hat{\mathbf{n}}, \hat{\mathbf{n}}') \hat{\nu}(\hat{\mathbf{n}}').$$

现在将上式两边取Tr运算,有

$$(\omega - v_F \hat{\mathbf{n}} \cdot \mathbf{q})\nu(\hat{\mathbf{n}}) = \hat{\mathbf{n}} \cdot \mathbf{q} \frac{k_F^2}{(2\pi)^3} \text{Tr} \text{Tr}' \int d\Omega' \hat{f}(\hat{\mathbf{n}}, \hat{\mathbf{n}}') \hat{\nu}(\hat{\mathbf{n}}'),$$

这里 $\nu(\hat{\mathbf{n}}) = \text{Tr} \hat{\nu}(\hat{\mathbf{n}})$, 而从(6.39)会发现

$$\text{Tr} \text{Tr}' \hat{f}(\hat{\mathbf{n}}, \hat{\mathbf{n}}') \hat{\nu}(\hat{\mathbf{n}}') = \frac{2\pi^2}{k_F m^*} F^S(\vartheta) \nu(\hat{\mathbf{n}}'),$$

于是

$$(\omega - v_F \hat{\mathbf{n}} \cdot \mathbf{q})\nu(\hat{\mathbf{n}}) = \mathbf{q} \cdot \hat{\mathbf{n}} v_F \int \frac{d\Omega'}{4\pi} F^S(\vartheta) \nu(\hat{\mathbf{n}}'). \quad (6.57)$$

现在以 \mathbf{q} 的方向为 z 轴,建立球坐标系以确定 $\hat{\mathbf{n}}$ 的方向,记

$$u_0 = \omega/|\mathbf{q}|, \quad s = u_0/v_F, \quad (6.58)$$

就有

$$(s - \cos \theta) \nu(\theta, \varphi) = \cos \theta \int \frac{d\Omega'}{4\pi} F^S(\vartheta) \nu(\theta', \varphi'). \quad (6.59)$$

在 \mathbf{q} 给定后, 求解此积分方程(这实际上是一个本征值问题)即可得到 $\nu(\hat{\mathbf{n}})$ 和 ω 。

在 $s \leq 1$ 时 $\nu(\theta, \varphi)$ 是有极点的, 于是(6.59)右边的积分中也是有极点的, 因此积分路径在复平面上必定需要用某种方式绕过该极点, 则(6.59)的左右两边都会出现虚部, 因此 ω 有虚部, 因此零声存在衰减。

零声模式和能谱

我们来分析一些常见情况下的零声模式和能谱。首先考虑 $F^S(\vartheta)$ 各向同性的情况, 即(6.40)中只有 $l = 0$ 的分量的情况。此时(6.59)右边的积分没有 θ 和 φ 的依赖, 因此左边也没有, 因此

$$\nu(\theta, \varphi) = \text{const} \times \frac{\cos \theta}{s - \cos \theta}, \quad (6.60)$$

这是一个各向异性的振动, 虽然此时费米液体的整个哈密顿量都是各向同性的。当然这没有引起任何矛盾, 因为说到底这是激发态, 考虑到 \mathbf{q} 的方向可以任意选取, 各向同性是能够保证的。在(6.60)知道之后, (6.59)就等价于

$$1 = \int \frac{2\pi \sin \theta' d\theta'}{4\pi} F_0^S \frac{\cos \theta'}{s - \cos \theta'},$$

也即

$$\frac{s}{2} \ln \frac{s+1}{s-1} - 1 = \frac{1}{F_0^S}. \quad (6.61)$$

因此 F_0^S 给定后 s 是常数, 因此零声是线性色散的。

6.3.7 费米液体的玻色化和流体动力学化

请注意 $n_{\mathbf{k}\sigma}$ 本身是粒子数的期望的某种经典化, 而粒子数期望的严格运动方程是所谓Schwinger-Dyson方程。因此, 朗道方程(6.47)——实际上就是粒子数的经典运动方程——和粒子数期望的严格运动方程(的Wigner函数形式)之间实际上只相差一些由于交换算符而产生的 δ 函数项, 通常这意味着粒子数算符满足的海森堡绘景运动方程的形式和零温朗道方程是一样的。由于朗道方程代入 $\epsilon_{\mathbf{k}\alpha}$ 的定义之后是自洽的, 这实际上是说, 我们可以通过朗道方程(6.47)就得到只关于朗道费米液体准粒子的某个二次型的涨落(显然, 它们是玻色子, 并且可以是也可以不是密度涨落)的等效哈密顿量, 而且是货真价实的、是算符的、做了勒让德变换之后能够出现在路径积分中的哈密顿量。对那些只有这些涨落才重要的系统, 实际上我们确实只需要一个关于玻色子的等效哈密顿量。因此通过观察(6.47), 我们实际上可以将费米液体玻色化。

在 $\tau \rightarrow \infty$ 极限下得到的费米液体的玻色模式中不包括普通声波, 或者说不包括第一声, 因为弛豫项的存在实际上意味着准粒子存在碰撞, 这样才能够有声波, 而 $\tau \rightarrow \infty$ 时没有弛豫项, 没有准粒子碰撞, 那么自然也没有声波。这其实可以看成Hubbard-Stratonovich变换的一个特例(关于玻色化和Hubbard-Stratonovich变换之间的关系见^[29]): 关于零声或第一声的理论可以通过这样的方式得到: 引入一个玻色模式, 将费米液体中能够产生零声或第一声的相互作用通道等效地用这个玻色模式传递, 然后积掉准粒子, 得到的玻色场论就是描述零声或第一声的理论了。但是, 并没有零声以外的能产生第一声的相互作用通道, 那么自然也没有独立于零声的能谱的第一声。从另一个角度, 第一声如果存在, 那么必然在

关于费米液体的玻色模式的哈密顿量中占据一部分能量，可是既然哈密顿量中没有给出准粒子碰撞的那部分能量，那么自然也就不会有第一声的能量，所以也就没有独立于零声的第一声。因此我们得出结论：第一声实际上就是被有限温度效应扭曲的零声，只是这个有限温度效应是如此的明显，以至于它让第一声的物理图像和零声的物理图像完全不一样了。虽然如此，第一声和零声的能谱永远是彼此连接的。

最后，我们指出，在有限温度时何为“一个零声声子”或者一个“第一声声子”理论上是存疑的，因为体系的响应谱的峰值和单粒子态的色散曲线的对应关系在零温下严格成立，可是有限温下似乎不容易定义一个“单粒子态”。不过，如果我们不考虑频率较高的第一声或是频率很低的零声等耗散极为明显的情况，将有限温第一声和零声当成可以直接通过对角化哈密顿量来得到的元激发来看待似乎并不会产生定性的错误：只要我们不经过零声-第一声过渡发生的那块能谱，热涨落的主要的贡献是修正系统的能谱（并且加入一个小的虚部），即使我们知道能够出现简直是定性上就和零声不一样的第一声，热涨落是必要的。这件事在很多别的领域——如量子光学——中也能够看到。

二维无自旋费米液体的流体动力学

考虑自旋可以忽略的二维无自旋费米液体，且假定温度充分接近零温，无电子-电子散射，从而 $\tau \rightarrow \infty$ ，此时可以使用一个单一的、没有下标的 $\delta n_{\mathbf{k}\alpha}$ 描述系统状态，动力学方程(6.55)化为

$$\frac{\partial \delta n_{\mathbf{k}}}{\partial t} + \mathbf{v} \cdot \frac{\partial \delta n_{\mathbf{k}}}{\partial \mathbf{r}} + \delta(\epsilon_F - \epsilon_{\mathbf{k}}) \mathbf{v} \cdot \int \frac{d^2 \mathbf{k}'}{(2\pi)^2} f_{\mathbf{k}\mathbf{k}'} \frac{\partial \delta n_{\mathbf{k}'}}{\partial \mathbf{r}} = 0. \quad (6.62)$$

设 $\hat{\mathbf{k}}$ 方向（可以用在动量空间中的某个极坐标系下的辐角 θ 标记）上费米面的运动距离为 $h(\theta)$ ，则

$$h(\theta) \cdot k_F d\theta = \int k dk d\theta \delta n_{k\hat{\mathbf{k}}},$$

从而有

$$h(\theta) = \frac{(2\pi)^2}{k_F} \tilde{\rho}(\theta), \quad \tilde{\rho}(\theta) = \int \frac{k dk}{(2\pi)^2} \delta n_{k=k\hat{\mathbf{k}}}. \quad (6.63)$$

我们用 $\tilde{\rho}$ 表示我们讨论的这些模式是密度涨落的“分量”；我们有

$$\rho = \frac{1}{V} \sum_{\mathbf{k}} \delta n_{\mathbf{k}} = \int \frac{d^2 \mathbf{k}}{(2\pi)^2} \delta n_{\mathbf{k}} = \int d\theta \tilde{\rho}(\theta). \quad (6.64)$$

低温下费米面的扰动非常小，从而我们可以认为所有事情都在未扰动的费米面上发生，从而单粒子能量为

$$\epsilon_{\mathbf{k}} = \epsilon_{\mathbf{k}}^0 + \int d\theta' f(\theta, \theta') \tilde{\rho}(\theta'), \quad (6.65)$$

其中

$$f(\theta, \theta') = f_{k_F \hat{\mathbf{k}}, k_F \hat{\mathbf{k}}'}. \quad (6.66)$$

(6.62)现在是

$$\frac{\partial \delta n_{\mathbf{k}}}{\partial t} + v_F \hat{\mathbf{k}} \cdot \frac{\partial \delta n_{\mathbf{k}}}{\partial \mathbf{r}} + \delta(\epsilon_F - \epsilon_{\mathbf{k}}) v_F \hat{\mathbf{k}} \cdot \frac{\partial}{\partial \mathbf{r}} \int d\theta' f(\theta, \theta') \tilde{\rho}(\theta'),$$

两边乘上 $\int k dk / (2\pi)^2$ ，并注意到

$$\int \frac{k dk}{(2\pi)^2} \delta(\epsilon_F - \epsilon_{\mathbf{k}}) v_F \hat{\mathbf{k}} = m^* v_F \hat{\mathbf{k}} = k_F \hat{\mathbf{k}},$$

我们有

$$\frac{\partial \tilde{\rho}}{\partial t} + v_F \hat{\mathbf{k}} \cdot \frac{\partial \tilde{\rho}}{\partial \mathbf{r}} + \frac{k_F}{(2\pi)^2} \int d\theta' f(\theta, \theta') \hat{\mathbf{k}} \cdot \frac{\partial \tilde{\rho}}{\partial \mathbf{r}} = 0. \quad (6.67)$$

这就是 $\tilde{\rho}$ 的运动方程。

(6.67)是从玻尔兹曼方程导出的，不过应当注意它实际上也是一个算符的零温运动方程，因为 $n(\mathbf{r}, \mathbf{k})$ 实际上是“单电子密度矩阵” $\langle \psi^\dagger(\mathbf{r}_1) \psi(\mathbf{r}_2) \rangle$ 的Wigner函数，而 $\tilde{\rho}$ 是将 $\delta n(\mathbf{r}, \mathbf{k})$ 的 \mathbf{k} 变量积掉一部分得到的，因此 $\tilde{\rho}$ 是一个良定义的算符的期望值，(6.67)是该算符的期望值的运动方程，它也必定有一个算符运动方程的版本。

费米液体的玻色化

(6.67)既然是费米液体的某个低能模式的运动方程，那么必定可以写出关于它的有效哈密顿量。尚且有待确定的是，(6.67)中的模式 $\tilde{\rho}(\theta)$ 是否给出了低能的费米液体的所有模式。这个问题并不是空泛而没有物理意义的：例如，如果 $\tilde{\rho}(\theta)$ 没有给出所有的模式，那么关于 $\tilde{\rho}(\theta)$ 的有效哈密顿量算出来的比热就会偏小。我们这里从另一条路径出发导出费米液体的玻色性低能有效理论。费米液体的玻色化在文献^[9, 15, 28]中讨论过，我们这里介绍他们的方案。应注意这种玻色化方案本身的意义有限，因为纯的费米液体理论是一个非常平凡、严格可解的理论，而纯的费米液体理论导出的流体动力学理论是没有任何阻尼的“干水”（见流体和软物质物理笔记中的第4.3节），同样是一个没什么有趣的性质的模型。考虑了电子-电子散射导致的弛豫的费米液体导出的就是有一定粘滞的纳维-斯托克斯方程^[26]，这两者都是相当复杂难以系统解决的。因此，费米液体的这种玻色化方案本身并没有太多理论上的价值。第十一章中的玻色化则有很多理论上的意义，因为它明确地表明了一维电子气一般来说不是费米液体，能够指出很大一类相互作用的结果（修正Luttinger液体中的参数 K ），并且提出了明确的判断一个系统是不是费米液体的判据（见(11.22)）。另一方面，费米液体的玻色化或者说流体力学化得到的理论参数——液体平均密度，粘度，压强和密度的关系——大多和费米液体的具体性质连续相关，从而无法通过测量这些参数来得知一个系统是不是费米液体。

实际上，由于流体动力学可以通过一些宏观守恒律加上对“隐藏的自由度”的假设（如应力各向同性等等）直接推导出来，原则上任何一种量子液体都能够玻色化；而且玻色化也不会丢失该量子液体中是否有费米子这一类的信息，因为总是可以通过适当的构造，从玻色型的“流体动力学自由度”中恢复出费米子自由度（如在Luttinger液体中的(11.21)；实际上即使系统中没有低能的费米型激发，原则上也可以从纯玻色型激发通过弦算符构造得到费米型激发，如背景知识13.1）。这个意义上玻色化本身是一个非常普适但是也非常无用的技巧；Luttinger液体这样能够一下子让我们看清楚相互作用量子液体的行为的玻色化是可遇不可求的。

流体动力学行为

以上的玻色化做法实际上预言了金属系统中的一些奇异的电子输运行为。电子在外场驱动下最常见的行为包括弹道输运和欧姆输运。

在电子-电子散射导致的电子动量分布弛豫可以忽略时，金属中的仅有的时间尺度是杂质和电子-声子散射导致的弛豫时间，如果它远大于 l/v_F ，其中 l 是体系线度，那么等效地可以认为电子从来不会受到杂质或是声子的散射，这就是弹道输运，即电子可以认为是直挺挺地如同子弹一般从系统的一个界面跑到另一个界面，然后弹射回来。这个情景下，对电子做流体力学化就会得到所谓的干水。

另一方面，如果金属中杂质、电子-声子散射导致的弛豫时间远远小于 l/v_F ，那么在较大的时间尺度下，可以认为电子流密度可以直接写成外场的某个函数，而无需考虑电子的惯性。这就是欧姆输运。欧姆输运也可以被认为是一种流体模型，但是这里，耗散项直接就是一个 $-k\mathbf{u}$ 项，而没有 $\eta\nabla^2\mathbf{u}$ 这种类型的项。简单的计算表明在外加电场是弱场时我们有

$$-k\mathbf{v} + \frac{e}{m}\mathbf{E} = 0, \quad (6.68)$$

其中漂移项 $\mathbf{u} \cdot \nabla\mathbf{u}$ 由于是电场的二阶项而略去；上式意味着流体中的输运行为遵从欧姆定律——这就是“欧姆输运”这个名称的来历。

然而，在温度足够低，以至于没有什么热涨落产生的声子，材料足够纯，以至于没有什么电子-杂质散射，而弹道输运又没有出现时，电子-电子散射将支配电子流体的行为，这就是**水动力学输运**。水动力学输运中，电子流速遵循的方程是纳维-斯托克斯方程，这意味着这种输运中存在**动量扩散**：在一块区域、一个方向上加载外场，在该区域两侧会出现漩涡。

欧姆输运中主要的损耗应该来自电子-声子散射和电子-杂质散射，而水动力学输运中主要的损耗来自电子-电子散射导致的动量扩散。

应注意水动力学输运虽然是新奇的行为，但是它未必意味着对朗道费米液体理论的违反，因为“纯的”，不考虑电子-电子散射的费米液体理论(6.35)仅仅在费米面附近成立，稍微远离费米面的地方电子-电子散射可以有可见效应。

6.3.8 费米液体理论的合理性

费米液体的合理性实际上仍然不是特别明确，主要是因为我们还无法确定什么时候费米液体理论就不适用了，自然也无法确定何时费米液体确实适用。确定一个系统是否适用费米液体理论首先需要确定系统的低能激发是否真的与电子的行为类似，因为原则上完全可以出现诸如电子配对等现象，或者自旋电荷分离，等等。这件事一般来说是非常困难的，因为直接从相互作用电子气出发做微扰计算得到的当然是修正后的电子，即严格地判断系统的低能激发到底是什么首先需要确定系统基态，而这就多少需要一些非微扰的方法。

在确定系统的低能激发和电子行为类似之后，费米球的存在性，以及低能激发主要是费米面附近的准粒子和空穴就可以确定。此时需要讨论的主要是相互作用的类型，即是否前向散射占据主要地位。这件事倒是不难：我们下面展示多种方法确定这一点。

重整化群

在重整化群计算中，一个二粒子散射项的强弱主要由这个散射项中允许的的粒子动量的取值多少确定，可能的动量占据的空间体积越大，对应的相互作用通道就越强。

懒得写了，总之就是BCS相互作用和密度-密度相互作用是特别强的两种。

格林函数计算

格林函数计算也可以给费米液体理论提供一定的辩护。我们计算单电子格林函数，实际上就是要计算自能修正。我们考虑一个任意的(1.10)中的相互作用势能，考虑一圈图，两

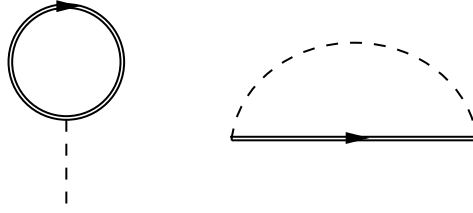


图 6.7: 自洽Hartree-Fock近似下的两个自能图

个自能图是

$$-\Sigma_{\mathbf{k}\sigma\omega}^{\text{Hartree}} = \text{Diagram} = \frac{1}{V} \sum_{\mathbf{k}', \alpha} (-V_0) n_{\mathbf{k}'\alpha}^0, \quad (6.69)$$

和

$$-\Sigma_{\mathbf{k}\sigma\omega}^{\text{Fock}} = \text{Diagram} = \frac{1}{V} \sum_{\mathbf{q}, \alpha} (-V_{\mathbf{q}}) (-n_{\mathbf{k}-\mathbf{q}, \alpha}^0 \delta_{\sigma\alpha}), \quad (6.70)$$

它们均没有 ω 依赖，即重整化因子 $Z = 1$ 。因此对一个做单圈自能修正就足够的费米子系统，等效单粒子能量为

$$\xi_{\mathbf{k}\alpha} = \xi_{\mathbf{k}\alpha}^0 + \sum_{\mathbf{k}', \beta} (V_0 - V_{\mathbf{k}-\mathbf{k}'} \delta_{\alpha\beta}) n_{\mathbf{k}'\beta}^0, \quad (6.71)$$

且单电子格林函数的形式和自由电子格林函数完全一致，区别在于要将 $\xi_{\mathbf{k}\alpha}^0$ 替换成 $\xi_{\mathbf{k}\alpha}$ 。这意味着，首先，重整化后电子仍然是有费米面的；其次，如果近似认为 n^0 就是 n ，(6.35)可以看成重整化后电子的等效哈密顿量。这就说明相互作用电子气的低能理论就是费米液体，并且费米液体中的准粒子就是相互作用修饰后的电子。

上面的推导中没有体现出一件事：如果 n 变了，那么 ϵ 会跟着发生变化，这在我们推导朗道方程(6.47)并且分析其行为时体现得尤其重要。这是因为(6.69)和(6.70)都是最低阶的图，其中的电子线自己没有经过相互作用修饰。完整的图——做了正确的电子自能、真空极化、相互作用顶角修饰——中的电子线给出 $n_{\mathbf{k}\alpha}$ ，因此的确如果 n 变了，那么 ϵ 会跟着发生变化；由于真空极化、相互作用顶角修饰也依赖于 n ，此时的系统总能量实际上相比(6.35)还要多出来一些项。容易看出，直接导致(6.35)的是自洽Hartree-Fock近似，其中自由电子线被修正为相互作用电子线（见图6.7），从而 n^0 被替换成 n 。因此费米液体理论实际上可以看成一种平均场理论。当然，从(6.46)的形式其实就可以看到这一点了。它和(6.9)中的自洽Hartree-Fock近似的不同之处在于 $V_{\mathbf{q}}$ 可以不是库伦排斥能，例如，可以做RPA近似得到等效屏蔽势，然后再做自洽Hartree-Fock近似。因此费米液体理论在给定了合适的 ϵ^0 和 f 之后能够非常精确地描述电子气的行为，而不像直接对库仑相互作用做Hartree-Fock近似一样很多时候计算不准确。

6.3.9 费米液体理论的意义，以及它何时失效

费米液体理论解释了为什么很大一类系统的基本自由度看起来就是经过修饰的单电子，虽然相互作用很强。近平衡态理论中的粒子数涨落通常不会太大，因此，准粒子等效能量 $\epsilon_{\mathbf{k}}$ 在探测单电子行为的实验——ARPES，导电性测试等——中基本上可以看成常数。库仑相互作用很强这件事的确会有明显的物理效果，然而这种物理效果只不过是修饰了单电子能量而已： f 不可忽略，但是因为 $n_{\mathbf{k}\sigma}$ 不怎么偏离 $n_{\mathbf{k}\sigma}^0$ ， $\epsilon_{\mathbf{k}}$ 是可以看成常量的。因此能带论对寻常的金属和绝缘体工作良好，由此我们说，费米液体理论是能带论成立的基础。

费米液体的思想是传统凝聚态物理的基础：在费米液体提出之后的很长一段时间，分析凝聚态系统的标准方式是将系统的基本激发当成某种经过重整化的电子（如费米液体理论中的准粒子），写出形式各异的无相互作用的能带，然后，在真的需要的情况下，加入占主导地位的相互作用通道，通常这是为了观察有没有某种相变。

费米液体理论的失效有多种情况。最容易想到的是，由于某些特殊类型的相互作用，先前所谓的准粒子形成一个费米海的状态其实并非真正的基态；当然，这里发生的就是一个对称性自发破缺，在费米液体理论的语境下，这种情况可以称为**费米面的不稳定性**，因为直观地看，对称性自发破缺形成的序可以认为是某种准粒子大量凝聚的状态，费米液体中有准粒子对应着费米海的“波澜”（零声就是一个例子），而某种准粒子大量凝聚意味着费米海被强烈地扭曲了，偏离了第6.3.1节所描述的样子；通常这种大量凝聚的准粒子是电子形成的，因此费米海不仅会扭曲，还会缩小，极端情况下费米面会直接消失。一些与费米面的不稳定性有关的情况见第八章。这些仍然在传统凝聚态物理的范围内。

还有一些情况下，相互作用会让系统的基本自由度**不是**电子型的准粒子。例如，一些较新的系统，如Luttinger液体，完全无法归入费米液体的框架：一点相互作用就足以破坏能带，从而让系统的低能自由度中没有能带电子激发，取而代之的是电子密度波。

第七章 密度泛函理论计算方法

本章讨论一种常用的求解固体电子结构的方法：密度泛函理论（DFT）。直接求解介质中的所有电子在库仑相互作用下的状态是极为困难的，但是我们将看到，原则上总是可以找到一个不依赖于具体系统的泛函，使得任何一个哈密顿量为(1.5)的系统的任何性质都可以通过最小化这个泛函得到。这样的计算原则上不需要引入任何经验参数，因此属于第一性原理计算。（还有很多其它属于第一性原理计算的方法）虽然这么说，在实际的第一性原理计算中还是需要一些经验性的东西，如泛函的选取；并且，很多时候难以通过适当选取泛函来捕捉到一些特别近距离而强烈的相互作用（如Hubbard相互作用），从而需要使用所谓的DFT+ U 方法等。DFT计算如今也经常和Hartree-Fock近似、PRA近似等方法联合使用。

具体来说我们将主要介绍KSDFT，即基于Kohn-Sham方程的DFT方法，因为它一方面是基于能量泛函的，一方面能够直接给出一些关于单电子波函数的信息。

7.1 KSDFT的理论基础与基本方程

7.1.1 为什么需要密度泛函理论

DFT的出现让凝聚态体系的无经验参数模拟成为可能。直接求解多体薛定谔方程基本上是不现实的，因为自由度实在太多。直接求解 10^{23} 量级的电子的波函数本身完全是胡扯；利用周期性条件，求解一个晶胞中的 N 个电子的波函数，则需要求解一个有 $3N$ 个坐标的薛定谔方程，还是非常困难。

在第6.2.1节中我们使用电子数密度标记系统状态并获得了正确的静态结果。一种想法是，有没有可能一般的相互作用电子气的行为也可以完全用它的基态电子数密度来确定。如果果真如此，我们可以首先将能量表示成电子数密度的泛函，然后最小化这个泛函，就得到了关于该相互作用电子气的一切信息。电子数密度满足的方程中只有3个坐标，这样做，如果可行的话，将大大降低计算量。

7.1.2 Hohenberg-Kohn定理

我们首先说明前述泛函的存在性。本节将通过两个定理，说明的确可以找到一个关于基态电子密度的泛函，使得我们只要最小化这个泛函即可原则上获得关于系统的所有性质。

我们有Hohenberg-Kohn第一定理：基态非简并、动能项和相互作用势能项固定、外加势能项是只和位置有关的单粒子算符的哈密顿量(1.6)的基态电子密度可以唯一确定哈密顿量，即基态电子密度和这类哈密顿量有一一对应的关系。（或者，说得更加明确一些，基

态电子密度和(1.5)中的 $V(\mathbf{r})$ 是一一对应的) 其中, 电子密度为

$$\rho(\mathbf{r}) = \langle \Psi | \psi^\dagger(\mathbf{r}) \psi(\mathbf{r}) | \Psi \rangle. \quad (7.1)$$

这个定理的证明如下。如果两个哈密顿量是相同的那么它们当然会给出一样的基态电子密度。而如果两个哈密顿量不相同却给出了一样的基态电子密度, 设这两个哈密顿量是

$$H_1 = T + V_{\text{int}} + V_1$$

和

$$H_2 = T + V_{\text{int}} + V_2,$$

且记每个粒子的外加势能分别是 $V_1(\mathbf{r})$ 和 $V_2(\mathbf{r})$, 它们的基态分别是 $|\Psi_1\rangle$ 和 $|\Psi_2\rangle$, 基态能量为 E_1 和 E_2 , 则

$$\langle \Psi_1 | T + V_1 | \Psi_1 \rangle = E_1, \quad \langle \Psi_2 | T + V_2 | \Psi_2 \rangle = E_2,$$

且由基态唯一性有

$$\langle \Psi_2 | T + V_1 | \Psi_2 \rangle > E_1, \quad \langle \Psi_1 | T + V_2 | \Psi_1 \rangle > E_2,$$

两式相减就得到

$$\langle \Psi_1 | V_2 - V_1 | \Psi_1 \rangle > E_2 - E_1, \quad \langle \Psi_2 | V_1 - V_2 | \Psi_2 \rangle > E_1 - E_2,$$

而

$$\langle \Psi | V | \Psi \rangle = \int d^3\mathbf{r} \langle \Psi | \psi^\dagger(\mathbf{r}) V(\mathbf{r}) \psi(\mathbf{r}) | \Psi \rangle = \int d^3\mathbf{r} \rho(\mathbf{r}) V(\mathbf{r}),$$

于是上式等价于

$$\int d^3\mathbf{r} \rho_1(\mathbf{r}) (V_2(\mathbf{r}) - V_1(\mathbf{r})) > E_2 - E_1, \quad \int d^3\mathbf{r} \rho_2(\mathbf{r}) (V_1(\mathbf{r}) - V_2(\mathbf{r})) > E_1 - E_2.$$

既然 $\rho_1(\mathbf{r}) = \rho_2(\mathbf{r})$, 以上两个不等式意味着 $0 > 0$, 而这当然是不正确的, 因此如果基态电子密度一样, 那么哈密顿量就不可能不同。这就证明了Hohenberg-Kohn第一定理。

Hohenberg-Kohn第一定理意味着, 只需要基态电子密度就足够确定一个(1.6)型体系, 因为(1.5)中的动能项都是一样的。因此这样一个体系的所有性质都可以写成基态电子密度的泛函。这个看起来不可思议的结论来自电子-原子核体系(1.5)只是所有可能的体系中的很小一部分这一事实。实际上, 我们看到, 事情的关键在于电子气系统和外界的耦合是通过哈密顿量 $\int d^3\mathbf{r} \rho(\mathbf{r}) V(\mathbf{r})$ 完成的, 而外加势场如此和电子气系统耦合又可以追溯到量子电动力学是一个规范理论这一事实 (见相对论性量子场论笔记关于规范理论的部分的说明)。既然我们原则上总是可以使用“外场变化后电荷密度如何响应”来描述系统 (既然我们也没有更多可观测量), 而外场和电荷密度又是线性耦合的, 那么Hohenberg-Kohn第一定理成立也就没有那么奇怪了: 能量泛函无非是一个类似于“量子有效作用量”的东西。

实际上, 还可以将Hohenberg-Kohn第一定理推广到可能具有简并基态的哈密顿量上。定义Levy-Lieb泛函:

$$\begin{aligned} E_{\text{LL}}[V_{\text{ion}}(\mathbf{r}), \rho(\mathbf{r})] &= \underbrace{\min_{\rho[\Psi]=\rho(\mathbf{r})} \langle \Psi | T + V_{\text{int}} | \Psi \rangle}_{E_{\text{LL}}} + \int d^3\mathbf{r} \rho(\mathbf{r}) V_{\text{ion}}(\mathbf{r}) \\ &= \min_{\rho[\Psi]=\rho(\mathbf{r})} \langle \Psi | T + V_{\text{int}} + V_{\text{ion}} | \Psi \rangle, \end{aligned} \quad (7.2)$$

它的取值肯定不会低于系统的基态能量，而如果 $\rho(\mathbf{r})$ 正好是外加势 $V_{\text{ion}}(\mathbf{r})$ 对应的电子密度，那么 $|\Psi\rangle$ 的取值范围当中肯定包括所有基态，于是 E_{LL} 给出了基态能量，而且这是该泛函的极小值。既然如此，记 $\rho_0(\mathbf{r})$ 为 $V_{\text{ion}}(\mathbf{r})$ 对应的电子密度，固定外加势不动对Levy-Lieb泛函做优化，其中 $\rho(\mathbf{r})$ 满足约束

$$\int d^3\mathbf{r} \rho(\mathbf{r}) = 1,$$

那么由拉格朗日乘法，一定有¹

$$\left. \frac{\delta E_{\text{LL}}}{\delta \rho(\mathbf{r})} \right|_{\rho(\mathbf{r})=\rho_0(\mathbf{r})} = \lambda,$$

这也就是说

$$\left. \frac{\delta F_{\text{LL}}}{\delta \rho(\mathbf{r})} \right|_{\rho(\mathbf{r})=\rho_0(\mathbf{r})} + V_{\text{ion}} = \lambda,$$

而可以通过别的优化方程计算出 λ ，这样实际上我们已经写出了 V_{ion} 关于 $\rho_0(\mathbf{r})$ 的表达式（可以差一个常数），因此可以从 $\rho_0(\mathbf{r})$ 把 V_{ion} 恢复出来，也即所有 V_{ion} 和所有可能的基态电子密度是一一对应的。因此Hohenberg-Kohn第一定理对具有简并基态的(1.6)型系统也使用。总之，求解出了基态电子密度，我们就获得了一个系统的所有信息，关于这个系统的所有物理量都可以通过基态电子密度的某个泛函计算出来。

一旦有了以上结论，就得到**Hohenberg-Kohn第二定理**，也即，基态电子密度是固定了 V_{ion} 之后的Levy-Lieb泛函（此时称为**能量泛函**）的极小值，可以使用变分原理求出。这个定理的证明只不过是把上面的论述倒过来使用而已：上面的论述表明给定 $\rho_0(\mathbf{r})$ ，通过泛函变分可以计算出 V_{ion} ，给定了 V_{ion} 通过泛函变分也可以计算出 $\rho_0(\mathbf{r})$ ，既然基态电子密度让能量泛函最小。实际上，记外加势能为 $V(\mathbf{r})$ 的能量泛函为 $E[\dots]$ ，用于求解基态电子密度的变分问题就是

$$\delta \left(E[\rho(\mathbf{r})] - \mu \left(\int d^3\mathbf{r} \rho(\mathbf{r}) - N \right) \right) = 0. \quad (7.3)$$

这个 μ 的形式看起来很眼熟，似乎就是化学势。实际上如果近独立电子近似成立，它真的就是零温化学势，即费米能。这一点我们在第7.4.1节中可以看到。

可以看到上面的推导实际上根本没有用到多少关于 T 和 V_{int} 的信息。对 T 并非 $\mathbf{p}^2/2m$ ， V_{int} 也并非库仑相互作用的系统，实际上同样有密度泛函理论，只不过由于外加势场未必具有 $V_{\text{ion}}(\mathbf{r})$ 这样的形式，可能并不使用电子数密度来标记一个系统，而可能使用诸如粒子流密度等等。例如，我们就能够将 V_{int} 仍然是库仑相互作用，但 T 对应狄拉克方程的问题也用密度泛函理论描述——这个理论对应的是重原子的非常内层的电子。核物理中的一些问题也可以使用它们的密度泛函理论（“协变密度泛函理论”）描述。

最后我们应当注意到，系统基态（多体）波函数、电子数密度 $\rho(\mathbf{r})$ 和 $V_{\text{ion}}(\mathbf{r})$ 三个量包含完全相同的信息，原则上知道了一个就可以知道其它的。因此， $E[\rho(\mathbf{r})]$ 的显式表达式可以显含 $|\Psi\rangle$ 或是 $V_{\text{ion}}(\mathbf{r})$ ，这没有超出DFT的框架。

7.1.3 无自旋系统的Kohn-Sham方法

现在的问题是写出 $E[\rho(\mathbf{r})]$ 并求解最优化问题(7.3)。本节将介绍所谓Kohn-Sham方法，它能够将能量泛函

$$E[\rho(\mathbf{r})] = \langle \Psi | H | \Psi \rangle = \langle \Psi | T + V_{\text{ion}} + V_{\text{int}} | \Psi \rangle \quad (7.4)$$

的最优化问题转化为一个单粒子薛定谔方程求解问题。

¹注意 λ 是一个常数而不是一个场，而变分通常都是一个场。

首先, 设 ψ_i^\dagger 算符能够产生一系列单电子态, 记 ψ_i^\dagger 产生的单粒子波函数为 $\phi(\mathbf{r})$ 。这样, 就有

$$\rho(\mathbf{r}) = \langle \Psi | \psi^\dagger(\mathbf{r}) \psi(\mathbf{r}) | \Psi \rangle = \sum_{i,j} \langle \Psi | \psi_i^\dagger \psi_j | \Psi \rangle \phi_i^*(\mathbf{r}) \phi_j(\mathbf{r}).$$

不失一般性地我们可以要求 ψ_i 的 i 标签体现了系统的守恒量, 则

$$\rho(\mathbf{r}) = \sum_i f_i |\phi_i(\mathbf{r})|^2, \quad (7.5)$$

其中

$$\langle \Psi | \psi_i^\dagger \psi_j | \Psi \rangle = \delta_{ij} f_i \quad (7.6)$$

是所谓单粒子密度矩阵, 我们取了能够让它对角化的一组基底。换言之, 我们要求 ψ_i^\dagger 的 i 标签能够让单粒子密度矩阵对角化, 这给出了 ψ_i^\dagger 的定义。单粒子密度矩阵总是能够定义的, 又总是能够对角化的, 因此一定可以找到这样的 ψ_i^\dagger 算符。我们应当注意, 就此处的条件, 是不能确定 $\{\phi_i(\mathbf{r})\}$ 就是自能计算(见背景知识4.1)意义下的单电子波函数的。 $\{\phi_i(\mathbf{r})\}$ 的物理意义在此处是存疑的, 实际上至今仍然是存疑的——至少现在, 它只不过是用来拟合电子数密度的一个工具而已。KSDFT的计算过程中存在“单粒子波函数 $\phi_i(\mathbf{r})$ ”并不意味着密度泛函理论是一个单粒子理论, 这些单粒子波函数或者说Kohn-Sham波函数只是计算中的辅助量而已。

下面我们分析能量泛函。动能部分为

$$\begin{aligned} \langle \Psi | T | \Psi \rangle &= - \langle \Psi | \int d^3\mathbf{r} \psi^\dagger(\mathbf{r}) \frac{\nabla^2}{2m} \psi(\mathbf{r}) | \Psi \rangle \\ &= - \sum_{i,j} \langle \Psi | \psi_i^\dagger \psi_j | \Psi \rangle \int d^3\mathbf{r} \phi_i^*(\mathbf{r}) \frac{\nabla^2}{2m} \phi_j(\mathbf{r}) \\ &= - \sum_i f_i \int d^3\mathbf{r} \phi_i^*(\mathbf{r}) \frac{\nabla^2}{2m} \phi_i(\mathbf{r}), \end{aligned}$$

我们记之为 T_s , 即定义

$$T_s = - \frac{1}{2m} \sum_i f_i \int d^3\mathbf{r} \phi_i^*(\mathbf{r}) \nabla^2 \phi_i(\mathbf{r}), \quad (7.7)$$

这里下标s的意思我们之后会说明。单粒子外加势能项可以类似的求出

$$V = \langle \Psi | V_{\text{ion}} | \Psi \rangle = \sum_i f_i \int d^3\mathbf{r} V_{\text{ion}}(\mathbf{r}) \phi_i^*(\mathbf{r}) \phi_i(\mathbf{r}) = \int d^3\mathbf{r} \rho(\mathbf{r}) V_{\text{ion}}(\mathbf{r}). \quad (7.8)$$

而二粒子的库伦势却会带来很大麻烦。当然, 如果采用托马斯-费米近似, 则唯一剩下的一项就是Hartree项

$$\langle \Psi | V_{\text{int}} | \Psi \rangle \approx V_H = \frac{1}{2} \int d^3\mathbf{r} \int d^3\mathbf{r}' \frac{e^2 \rho(\mathbf{r}) \rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|}, \quad (7.9)$$

但是正如我们在Hartree-Fock近似中看到的那样, 会有一个交换能, 而且Hartree-Fock近似与真实能量还存在偏差——称为**关联能**, 这种误差肯定会有因为Hartree-Fock近似是一个平均场理论。 $\langle \Psi | V_{\text{int}} | \Psi \rangle$ 只依赖于 $|\Psi\rangle$ 而 $|\Psi\rangle$ 由 V_{ion} 确定, 而 V_{ion} 和 $\rho(\mathbf{r})$ 一一对应, 我们确定, 它是 $\rho(\mathbf{r})$ 的泛函; 另一方面上式也是 $\rho(\mathbf{r})$ 的泛函。于是, 交换能和关联能之和也应该是 $\rho(\mathbf{r})$ 的泛函, 虽然我们现在写不出它的解析表达式。因此实际工作中必须猜测一个**交换-关联能** E_{XC} ,

从而写出能量泛函

$$\begin{aligned} E[\rho(\mathbf{r})] &= T_s[\rho(\mathbf{r})] + V[\rho(\mathbf{r})] + V_H[\rho(\mathbf{r})] + E_{XC}[\rho(\mathbf{r})] \\ &= -\frac{1}{2m} \sum_i f_i \int d^3\mathbf{r} \phi_i^*(\mathbf{r}) \nabla^2 \phi_i(\mathbf{r}) + \int d^3\mathbf{r} \rho(\mathbf{r}) V_{\text{ion}}(\mathbf{r}) \\ &\quad + \frac{1}{2} \int d^3\mathbf{r} \int d^3\mathbf{r}' \frac{e^2 \rho(\mathbf{r}) \rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + E_{XC}[\rho(\mathbf{r})]. \end{aligned} \quad (7.10)$$

将(7.10)对 $\phi_i^*(\mathbf{r})$ 做优化, 即求解

$$\delta \left(E[\rho(\mathbf{r})] - \sum_i \lambda_i \left(\int d^3\mathbf{r} \phi_i^*(\mathbf{r}) \phi_i(\mathbf{r}) - 1 \right) \right) = 0, \quad (7.11)$$

得到

$$\left(-\frac{1}{2m} \nabla^2 + \int d^3\mathbf{r}' \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + V_{\text{ion}}(\mathbf{r}) + V_{XC}(\mathbf{r}) \right) \phi_i(\mathbf{r}) = \epsilon_i \phi_i(\mathbf{r}), \quad (7.12)$$

其中

$$\lambda_i = \epsilon_i f_i \quad (7.13)$$

而

$$V_{XC}(\mathbf{r}) \phi_i(\mathbf{r}) = \frac{\delta E_{XC}[\rho(\mathbf{r})]}{\delta \phi_i^*(\mathbf{r})} = \frac{\delta E_{XC}[\rho(\mathbf{r})]}{\delta \rho(\mathbf{r})} \phi_i(\mathbf{r}). \quad (7.14)$$

(7.12)称为**Kohn-Sham方程**。这是一个本征值问题, 和薛定谔方程形式完全一样, 因此也有成熟的求解方法。它与(7.5)联立并要求

$$\int d^3\mathbf{r} |\phi_i(\mathbf{r})|^2 = 1, \quad (7.15)$$

就给出了所有的 $\psi_i(\mathbf{r})$ 。Kohn-Sham方程的形式基本上就是“一个电荷背景中的单电子”, 只不过电子-电子库伦相互作用不仅包括平方反比律(实际上就是Hartree项), 还包括一个(形式难以一般地写出的) V_{XC} , 即我们需要求解有效势能

$$V_{\text{eff}}(\mathbf{r}) = V_{\text{ion}}(\mathbf{r}) + V_{XC}(\mathbf{r}) + \int d^3\mathbf{r}' \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \quad (7.16)$$

如果近独立电子近似成立, ϵ_i 可能就是电子能级。我们将在第7.4.1节中看到这一点。需注意Kohn-Sham方程的解仅有的确凿无疑的物理意义就是电荷密度(根据(7.5)算出)和基态能量(将前 N 个 ϵ_i 加在一起); ϵ_i 和 ϕ_i 有什么意义是不完全确定的。定性地说它们在单电子图像适用时给出了单电子波函数和能量的估计, 因为Kohn-Sham方程的形式和(4.4)类似。我们也注意到, 如果我们正在分析的系统是寻常的金属或是绝缘体, 那么待计算的系统的基态应该比较接近一个无相互作用电子气的基态, 即和Kohn-Sham多体波函数拟设相似。既然使用Kohn-Sham方法表示出的电子数密度和实际的电子数密度接近, 如果系统基态和电子数密度之间的关系比较光滑, 计算得到的Kohn-Sham波函数和实际的单电子波函数也应该是比较接近的。但是这个估计有多准确是不知道的。我们将在第7.4.1节中讨论详细情况。

现在我们还差一件事没有做: 我们还需要知道 f_i , 就能够立即求出 $\rho(\mathbf{r})$, 从而完全解出了系统的一切性质。原则上 f_i 是 $\rho(\mathbf{r})$ 的函数, 但是我们写不出它的显式形式。常见的做法包括:

- Kohn-Sham一开始的工作中直接假定

$$|\Psi\rangle = \prod_{i=1}^N \psi_i^\dagger |0\rangle, \quad (7.17)$$

这意味着

$$\rho(\mathbf{r}) = \sum_{i=1}^N |\phi_i(\mathbf{r})|^2, \quad (7.18)$$

即 f_i 要么是0要么是1。换句话说，他们假定，在实际的基态 $|\Psi\rangle$ 附近总是找得到一个无相互作用的系统的基态 $|\Psi'\rangle$ ，它是 N 个 ψ_i^\dagger 乘在一起产生的Slater波函数，两者具有差不多的电子数密度，从而

$$\langle \Psi | T + V_{\text{ion}} + V_{\text{int}} | \Psi \rangle \approx \langle \Psi' | T + V_{\text{ion}} + V_{\text{int}} | \Psi' \rangle,$$

而我们实际上是在以 $|\Psi'\rangle$ 为变分参数来优化能量泛函，而不是以 $|\Psi\rangle$ 为变分参数，而以 $|\Psi\rangle$ 为变分参数来优化能量泛函只需要求解出所有 $\phi_i(\mathbf{r})$ 即可。

至于哪些 f_i 要设成1哪些要设成0，容易看出让 f_i 设成1的 i 有 N 个。(7.12)当然不止 N 个解，不过注意到在(7.12)两边乘上 $\phi_i^*(\mathbf{r})$ 并积分再对 i 求和，方程坐标就是交换关联泛函，而右边则是所有 ϵ_i 相加，因此我们取 ϵ_i 最小的 N 个 $\phi_i(\mathbf{r})$ 作为(7.5)中的 $\phi_i(\mathbf{r})$ 即可。

在原始的Kohn-Sham理论下， T_s 和实际的动能期望值之间确定会有系统的偏差，因为 $\phi_i(\mathbf{r})$ 说到底不是对角化单电子密度矩阵得到的。这就是下标s的来历。但是， T_s 和真实的 $\langle \Psi | T | \Psi \rangle$ 有差别这件事不会有什么影响，因为反正它们的差最后会在 E_{XC} 中得到补偿。(因此我们应当记住， E_{XC} 中是应当包含一部分实际电子的动能的！)

$|\Psi\rangle$ 附近是不是真的存在一个电子数密度足够接近的Slater波函数是可疑的。满足这个条件的系统称为**v-representable system**。我们前面论证过DFT适用于一切凝聚态系统，但是对基本自由度不是近独立的准电子的系统——即费米液体理论失效的情况——显然系统的基态波函数长得不会像Slater波函数。有时，对费米液体理论失效的凝聚态系统，可以使用更加合理的基态波函数拟设计算电子数密度并做DFT计算。此时最优化能量泛函不再等价于求解形式和电子薛定谔方程完全一致的Kohn-Sham方程，而等价于求解相应的波函数拟设中的准粒子的波动方程。

- 我们可以放宽对 f_i 的要求，允许它根据 ϵ_i 有一定的“模糊”或者说“分数占据”，可能是高斯型模糊、费米分布函数型模糊或者别的模糊方式。这样有助于改善收敛性。然而，这样计算出来的 $\phi_i(\mathbf{r})$ 算出来的 T_s 同样和动能期望值有系统性偏差。

需注意我们通常会要求 f_i 总是服从一定的形式，这个形式带有一些可调参数（如如果 f_i 服从费米分布，那么温度就是可调参数），这会带来额外的约束，从而被最优化的其实不是能量泛函减去 μN ，而是某种意义下的自由能。

在原始的Kohn-Sham理论和允许分数占据的Kohn-Sham理论中，Kohn-Sham DFT的过程可以概括为算法1。

- 更加精确的做法中，我们可以将单电子密度矩阵也作为变分参数求解，不过这样需要对密度矩阵施加一些约束，计算量增大不少。这种方法是所谓的**RDMFT (reduced density matrix fu**这样的好处则是，首先，此时动能项是准确的，不需要像经典的Kohn-Sham方法一样，指望交换关联泛函补偿动能，其次， f_i 的分数占据是有非常明确的物理意义的，而不是单纯的改善收敛性的一个意义不明的技巧。

Kohn-Sham方程并非做DFT计算的唯一方法。例如，我们可以有Orbital-free DFT，这是一种完全不使用任何波函数的DFT方法，需要写下交换关联泛函 E_{XC} 以及一个用电子密度表示动能的泛函，即**动能泛函**，然后做最优化。动能泛函通常比交换关联泛函还要难以

算法 1: Kohn-Sham方程的自洽求解

输入: 初始电子密度 $\rho_0(\mathbf{r})$, 容差 ϵ , 交换关联泛函选取 $E_{\text{XC}}[\rho(\mathbf{r})]$, 电子数 N

输出: Kohn-Sham波函数 $\phi_i(\mathbf{r})$ 和对应的本征值 ϵ_i

```

1  $i = 1$ 
2 将 $\rho_0(\mathbf{r})$ 代入(7.12)求解得到 $\phi_n^{(1)}$ 和 $\epsilon_n^{(1)}$ 
3 将 $\phi_n^{(1)}$ 代入(7.5)计算得到 $\rho_1(\mathbf{r})$ 
4 while  $\rho_i(\mathbf{r})$ 和 $\rho_{i-1}(\mathbf{r})$ 的差别大于容差 $\epsilon$  :
5   将 $\rho_i$ 代入(7.12)求解得到 $\{\phi_n^{(i+1)}\}$ 和 $\{\epsilon_n^{(i+1)}\}$ 
6   根据 $\epsilon_i$ 计算 $f_i$ 
7   用(7.5)计算 $\rho_{i+1}(\mathbf{r})$ 
8    $i = i + 1$ 
9 return 波函数 $\phi_n^{(i)}$ 和本征值 $\epsilon_n^{(i)}$ 

```

写出。的确，可以仿照计算均匀电子气交换关联泛函的方法计算均匀电子气的动能泛函，但是直接把这些动能泛函拿去算固体系统效果是不太好的。另一方面，数值结果表明大部分体系中，Hartree项加均匀电子气Fock项就足够覆盖大部分能量，因此相对而言，获得交换关联泛函不是那么困难的一件事。因此OFDFT长期以来不受重视，但近年由于找到了一些可靠的动能泛函，OFDFT的高效率让它再次受到量子化学界的关注。

以上算法也可以体现出DFT方法的一个重大缺陷，就是在不和实验对比时很难评估计算精度，也很难系统地提升计算精度。求解Kohn-Sham方程是一个完全确定性的算法，我们无法使用量子蒙特卡洛中模拟中观察计算结果方差的方法估计本次计算是不是“对头”；另一方面， $E_{\text{XC}}[\rho(\mathbf{r})]$ 中各项的物理意义不是非常明晰，因此也很难评估如何改进计算精度。

7.1.4 自旋和相对论效应

以上讨论完全没有涉及自旋。很多凝聚态系统中电子自旋并不重要，从而，自旋指标除了让电子多了一倍出来以外什么也没有做，将这样的系统看成无自旋的和把它看成有自旋的是完全一样的；但是也有自旋很重要的系统，如诸如电子-电子交换相互作用导致的磁性，以及自旋-轨道耦合（相对论效应）都将自旋牵扯进来。对这两种情况，哈密顿量里面都要出现自旋指标，磁性是来自库仑相互作用中的交换相互作用，如Hartree-Fock近似中的(6.9)，而相对论效应会导致 $\mathbf{L} \cdot \mathbf{S}$ 项。

含有自旋的系统的外加势场可以不是仅仅和坐标有关的 $V_{\text{ion}}(\mathbf{r})$ ，而可以和自旋有关，于是第7.1.2节中的推理不再适用了，不过我们会发现，只需要把第7.1.2节中的Hohenberg-Kohn定理中的“ $\rho(\mathbf{r})$ ”替换成四个密度：电子数密度， x, y, z 方向的自旋密度，就能够建立含自旋系统的密度泛函理论。或者等价的，我们也可以考虑带有自旋指标的密度

$$\rho_{\alpha\beta}(\mathbf{r}) = \langle \Psi | \psi_{\alpha}^{\dagger}(\mathbf{r}) \psi_{\beta}(\mathbf{r}) | \Psi \rangle, \quad (7.19)$$

它也有四个独立自由度。此时的Kohn-Sham方程中的 V_{XC} 也是带有两个自旋指标的。自旋-轨道耦合其实可以归入 T 中，其它相对论修正也可以归入 T 中。实际上，库仑相互作用的狄拉克方程也有密度泛函理论，因此这种把相对论修正归入 T 的做法是合理的。综上，含自

旋DFT的Kohn-Sham方程可以写成

$$-\frac{\nabla^2}{2m}\phi_i(\mathbf{r}\sigma) + (V_{\text{ion}}(\mathbf{r}) + V_{\text{H}}(\mathbf{r}) + V_{\text{XC}})\phi_i(\mathbf{r}\sigma) - \boldsymbol{\mu}_{\sigma\sigma'} \cdot (\mathbf{B}(\mathbf{r}) + \mathbf{B}_{\text{XC}})\phi_i(\mathbf{r}\sigma') = \epsilon_i\phi_i(\mathbf{r}\sigma). \quad (7.20)$$

在系统中大部分电子的自旋可以视为指向相同或相反的方向（或者说自旋没有量子涨落）时，可以将四个自由度的 $\rho_{\alpha\beta}(\mathbf{r})$ 用简单的 $\rho_{\uparrow}(\mathbf{r})$ 和 $\rho_{\downarrow}(\mathbf{r})$ 代替。

7.1.5 有限温度DFT

我们之前讨论过， $|\Psi\rangle$ 可以不必像原始的Kohn-Sham方案中一样是一个Slater波函数，可以有分数占据。我们知道热平衡态系统就是有分数占据的，因此很自然的想法是，能不能在有限温度情况下建立DFT理论。一方面这可以用于提高收敛性，一方面显然也有巨大的实用价值：对普通的凝聚态物理中的系统，当温度升高到电子占据数明显偏离阶跃函数时，由于外压普遍不强，一般来说系统已经是经典流体了，但诸如天体物理中可能需要研究所谓warm dense matter，此时有限温度DFT是大有好处的。

在有限温度、与外界没有做功的系统中自由能是取极小值的。（ $T \rightarrow 0$ 时自由能和内能是一样的，而内能就是基态能量，这就退化到了零温DFT上）我们只需要证明自由能够写成^[19]

有限温度DFT中被最优化的是自由能而不是内能，如果将有限温度DFT当成一种加快收敛的手段计算系统的基态性质，那么不能忘记从自由能中减除 TS ，但是这样获得的也是有限温度的内能；比较好的做法是计算多个有限温度的内能，通过插值得到基态能量。

在 TS 项没有占据压倒性优势时，它做的无非是将能量泛函的最小值位置从零温能量泛函的最小值位置切换到零温能量泛函的某个极小值位置附近。因此，零温下的DFT计算中的局部极小值很多时候都是有一定物理意义的，如一种材料的能量泛函可能同时有顺磁序、铁磁序、反铁磁序三个能量极小值，彼此能量相差不大，如果随着温度改变，自由能极小值位置在这三个零温能量泛函的极小值附近切换，就意味着存在有限温度相变。

7.1.6 TDDFT

Runge-Gross (RG) 定理的地位和Hohenberg-Kohn定理是一样的。

7.2 泛函选择

本节讨论如何构造和选择合适的交换关联泛函 $E_{\text{XC}}[\rho(\mathbf{r})]$ ；在DFT的语境下，交换关联泛函经常就简称“泛函”。库仑定律的形式告诉我们，交换关联泛函的形式肯定是不局域的，因为有 $|\mathbf{r} - \mathbf{r}'|$ 这样的因子，并且同时要对 \mathbf{r} 和 \mathbf{r}' 两个变量求积分。但实际上在大部分时候，由于屏蔽作用等这种空间上的非局域性衰减得很快，于是密度关联泛函总是可以写成以下的局域表达式：²

$$E_{\text{XC}}[\rho(\mathbf{r})] = \int d^3\mathbf{r} f(\rho(\mathbf{r}), \nabla\rho(\mathbf{r}), \dots),$$

²如果非局域性衰减得很快，那么可以做多级展开

$$f(\mathbf{r} - \mathbf{r}') = f(\mathbf{r}) + (\mathbf{r} - \mathbf{r}') \cdot \nabla f + \dots,$$

这样就可以先积掉 \mathbf{r}' 这个变量，得到

$$\int d^3\mathbf{r} d^3\mathbf{r}' f(\mathbf{r} - \mathbf{r}') = \int d^3\mathbf{r} g(f(\mathbf{r}), \nabla f(\mathbf{r}), \dots).$$

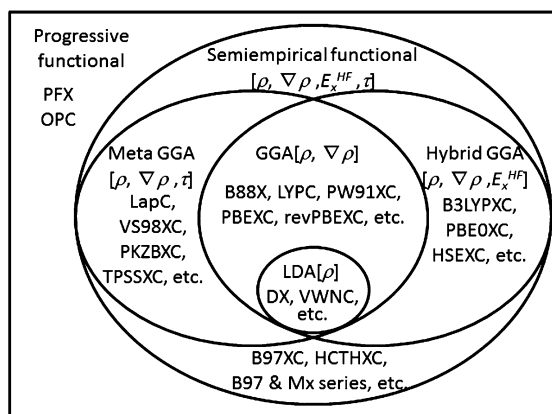


图 7.1: 交换-关联泛函的分类

而由于交换能与 $\rho(\mathbf{r})$ 的平方同阶，通常会设

$$E_{XC}[\rho(\mathbf{r})] = \int d^3\mathbf{r} \rho(\mathbf{r}) \epsilon_{XC}(\rho(\mathbf{r}), \nabla \rho(\mathbf{r}), \dots). \quad (7.21)$$

这里的 $\rho(\mathbf{r})$ 在不考虑自旋时就是单纯的电子数密度，在考虑自旋时就是 $[\rho_{\alpha\beta}(\mathbf{r})]_{\alpha\beta}$ 。

常用的交换-关联泛函的形式包括以下几种：

- **局域密度近似 (LDA)**： $\epsilon_{XC}[\rho(\mathbf{r})] = \epsilon_{XC}(\rho(\mathbf{r}))$ ；
- **广义梯度近似 (GGA)**： $\epsilon_{XC}[\rho(\mathbf{r})] = \epsilon_{XC}(\rho(\mathbf{r}), \nabla \rho(\mathbf{r}))$ ；
- **Meta-GGA**：在GGA近似中加入一个动能密度修正（基本上是 $\nabla^2 \rho(\mathbf{r})$ ）；
- **混合泛函**：Hartree-Fock近似和别的一些东西的线性组合；
- **经验泛函**：有一大堆参数，根据实际数据微调参数。

不同泛函的分类如图7.1所示。

目前并没有能够准确计算一切体系的 E_{XC} ，没有也是很正常的，因为很明显，这个泛函需要的复杂程度和“看着库仑相互作用想出一切可能的凝聚态现象”是一样的，后者眼下不可能做到，前者当然不可能做到。但是对特定的体系，一些相互作用通道可以忽略，一些误差可以互相抵消，找到一个复杂程度适中的泛函描述它们是可以做到的。适当地构造和选择泛函非常需要技术含量和经验。

最后，虽然原则上一切库仑相互作用导致的物理效应都可以通过 $E_{XC}[\rho(\mathbf{r})]$ 得到，这件事实际上做起来并不那么容易。如果一个泛函被证明能够模拟一些相互作用通道导致的物理而无法模拟另一些，我们就需要手动向一个没有可调参数的 $E_{XC}[\rho(\mathbf{r})]$ 引入一些修正项。这些修正项的大小往往无法固定下来，因为不同情况下 $E_{XC}[\rho(\mathbf{r})]$ 对对应的相互作用通道的忽略可能也是不一样的。请注意交换关联泛函是通过计算电子数密度具有某种特点的系统获得的，而不是通过分析哪些费曼图要保留哪些可以忽略获得的，一个泛函不能良好地对应到一种通常意义上的保留一些相互作用而忽略另一些的有效模型上，交换关联泛函到底把什么纳入了考虑是缺乏直观的物理意义的，这也是DFT难以系统地提升精度的原因：在使用更加接近微扰场论方法的那些算法（如各种post-HF方法）时我们是清楚地知道自己忽略了什么的，在DFT中我们并不知道自己忽略了什么。手动引入Hubbard排斥（所谓LDA+U，涉及非常短程的相互作用，不容易被比较光滑的泛函模拟）、范德瓦尔斯力（长程，不容易被比较局域的泛函模拟）就是例子。

7.2.1 LDA近似

本节考虑自旋不重要的系统，即无需考虑自旋-轨道耦合，无磁序等等的系统。对这类系统，可以完全忽略Kohn-Sham波函数的自旋指标，写下

$$\rho(\mathbf{r}) = 2 \sum_i f_i |\phi_i(\mathbf{r})|^2. \quad (7.22)$$

当电子密度变化非常平缓时， $\nabla\rho, \nabla^2\rho$ 等全部可以看成是零，此时交换关联泛函的形式是

$$E_{XC}[\rho(\mathbf{r})] = \int d^3\mathbf{r} \rho(\mathbf{r}) \epsilon_{XC}(\rho(\mathbf{r})). \quad (7.23)$$

这就是所谓的**局域密度近似 (LDA)**：每单位的交换关联能只和该地点的电子密度有关。注意 ϵ_{XC} 和 V_{XC} 不是一回事，两者的关系是

$$V_{XC}(\mathbf{r}) = 2\epsilon_{XC}(\rho(\mathbf{r})) + 2\rho(\mathbf{r}) \frac{d\epsilon_{XC}}{d\rho}. \quad (7.24)$$

因子2来自 $\rho(\mathbf{r})$ 相比 $\sum_i |\phi_i(\mathbf{r})|^2$ 要乘以2。一种常见的习惯是干脆将这个2吸收进 $\epsilon_{XC}(\rho(\mathbf{r}))$ 中，从而，使用这种记号的文献的 $\epsilon_{XC}(\rho(\mathbf{r}))$ 相比本文要多一倍。

我们主要将尝试寻找一个确定适用于均匀电子气的交换关联泛函。均匀电子气指的是没有晶格、也没有外场时的库伦相互作用电子气，其电荷密度是处处均匀的。对充分大的盒子中的均匀电子气， E 正比于系统体积，从而 E/V 可以认为只依赖于电子数密度，和 V 没有直接的依赖关系。对这样的体系，显然我们有

$$\frac{N}{V} \epsilon_{XC}(N/V) \times V = E_{XC}. \quad (7.25)$$

由于无论是均匀电子气还是电子密度较平缓但是的确有晶格周期势的电子气理论上均共享一种LDA近似适用的 E_{XC} ，而 $\epsilon_{XC}(\rho(\mathbf{r}))$ 只能看到 \mathbf{r} 点的电子数密度，而不能够分辨电子数密度是否存在空间起伏，我们确定， ϵ_{XC} 可以按照如下方式得到：

$$\epsilon_{XC}(\rho(\mathbf{r})) = \left(\frac{E_{XC, \text{homogeneous}}}{N} \right) \Big|_{N/V \rightarrow \rho(\mathbf{r})}. \quad (7.26)$$

换句话说，我们只需要计算具有不同电子数密度的均匀电子气的基态能量，按照上式即可拟合出 $\epsilon(\rho(\mathbf{r}))$ 的一个表达式。因此LDA近似实际上是非常唯一的，不同的LDA泛函的区别仅仅在于拟合函数的形式不同。

在计算不同电子数密度的均匀电子气的基态能量时常常使用 r_s 作为更加方便的记号。通常拟合 E 和 r_s 之间的关系，而不是 E 和 ρ 之间的关系。

Hartree-Fock近似和托马斯-费米-狄拉克泛函

既然Kohn-Sham波函数和实际的单电子波函数定性符合，我们用 $\phi_i(\mathbf{r})$ 做Hartree-Fock近似。自由电子气的基态就是费米面以内的每个轨道都占据了一上一下两个电子的状态，我们设共有 N 个电子，那么这些电子占据了 $N/2$ 个不同的轨道。仅考虑库伦相互作用引入的一阶微扰，则有

$$E = \sum_{\sigma, \sigma'} \langle \Psi | \frac{1}{2} \int d^3\mathbf{r} d^3\mathbf{r}' \psi_{\sigma}^{\dagger}(\mathbf{r}) \psi_{\sigma'}^{\dagger}(\mathbf{r}') \frac{e^2}{|\mathbf{r} - \mathbf{r}'|} \psi_{\sigma'}(\mathbf{r}') \psi_{\sigma}(\mathbf{r}) | \Psi \rangle.$$

由于出现了两个自旋指标 σ, σ' ，不能简单地假装电子没有自旋然后把所有计算结果乘以2，而必须小心。可以使用Wick定理把上式展开，这就是在做Hartree-Fock近似。请注意由于均匀电子气中没有自旋翻转，任何形如

$$\langle \Psi | \psi_{\sigma}^{\dagger} \psi_{\sigma'} | \Psi \rangle, \quad \sigma \neq \sigma'$$

的项都是零，而

$$\langle \Psi | \psi_{\sigma}^{\dagger}(\mathbf{r}) \psi_{\sigma}(\mathbf{r}') | \Psi \rangle = \sum_i \phi_i^*(\mathbf{r}) \phi_i(\mathbf{r}'),$$

于是最后能量修正为

$$E = \frac{1}{2} \int d^3\mathbf{r} d^3\mathbf{r}' \frac{e^2}{|\mathbf{r} - \mathbf{r}'|} \left(\underbrace{4 \sum_i |\phi_i(\mathbf{r})|^2 \sum_i |\phi_i(\mathbf{r}')|^2}_{\text{classical coulomb energy, Hartree term}} - \underbrace{2 \sum_i \phi_i^*(\mathbf{r}) \phi_i(\mathbf{r}') \sum_i \phi_i^*(\mathbf{r}') \phi_i(\mathbf{r})}_{\text{exchange energy}} \right).$$

交换能项前面的因子是2而不是4是因为自旋不同的电子之间的交换能全部抵消了。这样交换能或者说Fock项就是

$$E_X = - \int d^3\mathbf{r} d^3\mathbf{r}' \frac{e^2}{|\mathbf{r} - \mathbf{r}'|} \left| \sum_i \phi_i^*(\mathbf{r}) \phi_i(\mathbf{r}') \right|^2. \quad (7.27)$$

对箱归一化的均匀电子气，有

$$\phi_i(\mathbf{r}) = \frac{1}{\sqrt{V}} e^{-i\mathbf{k}_i \cdot \mathbf{r}},$$

于是就有

$$\sum_i \phi_i^*(\mathbf{r}) \phi_i(\mathbf{r}') = \sum_{\text{occupied } \mathbf{k}} \frac{1}{V} e^{i\mathbf{k} \cdot (\mathbf{r} - \mathbf{r}')} = \frac{1}{(2\pi)^3} \int_{|\mathbf{k}| < k_F} d^3\mathbf{k} e^{i\mathbf{k} \cdot (\mathbf{r} - \mathbf{r}')}, \quad (7.28)$$

而电子密度的一半（因为只考虑了轨道自由度）为

$$n(\mathbf{r})/2 = \frac{k_F^3}{6\pi^2}.$$

于是我们通过 k_F 引入了电子数密度。将(7.28)代入(7.27)就得到交换能的形式。由于我们做了LDA近似，就是将每一个体积元看成一个箱子，这个箱子里面是均匀电子气，而忽略箱子和箱子之间的衔接，只有 $\mathbf{r} - \mathbf{r}'$ 不超出一个箱子时(7.28)才有明显的非零值。每个箱子都有自己的费米动量，记作 $k_F(\mathbf{r})$ （具体是关于 \mathbf{r} 还是 \mathbf{r}' 无关紧要因为两者总是在同一个箱子内；它们本质上都是 $\rho(\mathbf{r})$ 的函数）。令

$$\mathbf{s} = \mathbf{r} - \mathbf{r}',$$

则

$$\begin{aligned} E_X &= -e^2 \int d^3\mathbf{r} d^3\mathbf{s} \frac{1}{s} \left| \frac{1}{(2\pi)^3} \int_{|\mathbf{k}| < k_F} d^3\mathbf{k} e^{i\mathbf{k} \cdot \mathbf{s}} \right|^2 \\ &= -e^2 \int d^3\mathbf{r} d^3\mathbf{s} \frac{1}{s} \left(\frac{3 \sin t - t \cos t}{t^3} \rho(\mathbf{r}) \right)^2 \quad (t = k_F(\mathbf{r})s) \\ &= -9e^2 \pi \int d^3\mathbf{r} \int dt \frac{t}{k_F(\mathbf{r})^2} \rho(\mathbf{r})^2 \left(\frac{\sin t - t \cos t}{t^3} \right)^2 \\ &= -e^2 \frac{3}{4} \left(\frac{3}{\pi} \right)^{1/3} \int d^3\mathbf{r} \rho(\mathbf{r})^{4/3}. \end{aligned}$$

当然, 根据(7.26), 可以不使用这种繁琐的方法, 直接假定 $\rho(\mathbf{r})$ 是常数, 计算完之后再将它换成 $\rho(\mathbf{r})$ 。总之, 我们得到了托马斯-费米-狄拉克泛函:

$$E_X[\rho(\mathbf{r})]/e^2 = -\frac{3}{4} \left(\frac{3}{\pi} \right)^{1/3} \int d^3\mathbf{r} \rho(\mathbf{r})^{4/3} = -0.7386 \int d^3\mathbf{r} \rho(\mathbf{r})^{4/3}. \quad (7.29)$$

对均匀或不均匀但缓变的电子气这个泛函确定适用。等价地说, 我们有

$$\epsilon_X(\rho(\mathbf{r})) = -\frac{3}{4} \left(\frac{3}{\pi} \right)^{1/3} \rho(\mathbf{r})^{1/3}, \quad (7.30)$$

或者使用 r_s 表示为

$$\epsilon_X = -\frac{0.458}{r_s}. \quad (7.31)$$

在很多文献中上式中的因子需要乘以2, 因为(7.24)中的因子2被吸收到了 ϵ_{XC} 中。

对一般的系统, 托马斯-费米-狄拉克泛函当然是不那么精确的。首先Hartree-Fock近似就不是什么特别精确的近似, 其次将Kohn-Sham波函数当成Hartree-Fock近似求解出的波函数本身也有误差, 实际上(7.29)甚至没有完整地考虑Fock项, 因为它是在假定电子气是均匀的前提下计算出来的。一般的自旋不重要的系统的Fock项, 也就是(6.10), 难以解析地写成电子数密度的泛函的, 因为它依赖于多电子波函数的细节。虽然原则上后者也是电子数密度的泛函, 但是知道了这两者的关系等于说是将整个密度泛函理论中各个量之间的关系全部解析地写出来了, 显然是不可能做到的。不过, 托马斯-费米-狄拉克泛函我可以作为更精确的泛函的起点。下面我们尝试寻找一个 E_C 使得

$$E_{XC}[\rho(\mathbf{r})] = E_X[\rho(\mathbf{r})] + E_C[\rho(\mathbf{r})]. \quad (7.32)$$

这个 E_C 就称为关联泛函。和自洽Hartree-Fock中的情况很类似, 关联泛函这个名字本身其实也不恰当, 因为一部分关联效应已经体现在 $E_X[\rho(\mathbf{r})]$ 中了。

关联泛函的高电子密度极限和低电子密度极限

关联泛函一般无法解析计算。在电子数密度特别高或是特别低的时候, 关联泛函还是能够写出非常精确的表达式的。

关联泛函的量子蒙特卡洛计算

背景知识 7.1: 量子蒙特卡洛方法

大体上说要处理一个统计系统就是要计算其配分函数。在没有量子涨落时, 如果能够设计一个随机过程, 使得其稳定解在理论上满足

$$p(W) \propto e^{-\beta H[W]},$$

或者说

$$p(W' \rightarrow W) e^{-\beta H[W']} = p(W \rightarrow W') e^{-\beta H[W]},$$

那么只需要真的运行这个随机过程足够长的时间, 就能够得到系统中各个构型的概率分布, 从而通过

$$\langle O \rangle = \sum_c \frac{W(c)}{Z} O[c] = \frac{1}{N} \sum_i O[\mathcal{C}_i]. \quad (7.33)$$

可以计算各个物理量的期望、方差等一切信息。对量子系统，我们需要将 βH 替换成作用量，将 W 替换成虚时间演化路径或者诸如此类的东西，然后构造随机过程，重复以上步骤，然后只要能够将物理量表示成 W 的函数，就能够计算出系统的一切行为。这种方法就是**蒙特卡洛方法**，当它涉及量子系统时称为**量子蒙特卡洛方法**。

蒙特卡洛方法最大的优点可能是能够比较容易地估算误差：只需要看随机过程是不是有很大涨落即可。

比较糟糕的是，配分函数中的权重并不能够保证是正数，甚至不能保证是实数，因此，实际上我们并不能够得到一个真正的概率分布。这可能导致蒙特卡洛算法不收敛，或者虽然收敛但由于分母非常小（正负抵消）而精度很差。这就是所谓的**符号问题**—— $W(n)$ 的正负号不定导致模拟困难。设哈密顿量是正定的，如果在一组基底上所有非对角元或者等于零或者小于零，那么能够保证没有符号问题。直观地看这很合理，因为根据Forbiden's (?) 定理，此时基态波函数在不同基态上的投影的权重都是一样的，因此系统在我们选定的基态下非常“经典”。当然，原则上，在哈密顿量的本征态下非对角元都是零，从而原则上只要恰当选取基底就不会有符号问题，但是如果这组基底高度非局域（或者说纠缠很大），那么基本上没有可能使用蒙特卡罗方法模拟这个系统。

蒙特卡洛方法的另一个问题是时候有待计算的物理量和选取的 W 之间没有特别便利的解析关系。比如说，蒙特卡洛计算纠缠熵就多半很难做到。

量子蒙特卡洛方法有多种方案，适用于不同系统；通常需要针对不同系统设计不同方案。

对电子数密度并不特别高或是特别低的系统，通常使用量子蒙卡方法计算均匀电子气的基态能量，将它表示成 r_s 的函数做曲线拟合，就能够得到 ϵ_C 。

常见的形式包括Perdew-Zunger, VWN等等。

以上获得LDA泛函的步骤很清楚地展示了密度泛函理论的弱点：因为物理量的解析表达式通常都是依赖于多电子波函数的细节的，将一切都使用电子数密度表示，实际上是不太容易做解析计算的，那也就很难分析 $E_{XC}[\rho(\mathbf{r})]$ 中各项的物理意义。

LDA的局限性

普通的LDA泛函虽然将Hartree-Fock近似中的Fock项考虑了，但是由于要写出和 $\rho(\mathbf{r})$ 的显式依赖，LDA泛函给出的仅仅是均匀电子气的交换能，那么实际上会高估电子的离域特性而低估电子的定域特性。如果我们要分析的系统中有电子比较定域，仅仅使用LDA泛函可能是不太可靠的。

LDA近似倾向于低估原子间距，理由也是显然的：如果电子比较离域，化学键可以比较强，从而原子被更加紧密地绑定在一起。

另一个LDA泛函的局限性在于它存在**自相互作用误差**：Hartree项假定库伦相互作用可以写成(7.9)的形式，那么如果系统基态有电子的波函数同时在 \mathbf{r} 和 \mathbf{r}' 处都有可观的振幅，(7.9)就会导致一个奇怪的“电子自己和自己相互作用”的能量。³这当然是错误的——它必须在交换关联泛函中得到修正。完整的、依赖于单电子波函数的Hartree-Fock近似是能够

³我们的确会讨论电子自能，但(7.9)给出的自相互作用是非物理的：在体系中只有一个电子时，库伦相互作用项不导致基态能量修正（因为只有有两个电子时才有库伦相互作用：这和 ϕ^4 理论不同），但是(7.9)仍然会给出基态能量修正。

消除自相互作用误差的。但由于LDA泛函中的 E_X 只有在电子密度足够均匀时才和Hartree-Fock近似中的Fock项一致，实际上LDA泛函不能够完全消除自相互作用误差。

7.2.2 LSDA近似

现在转而考虑自旋系统的LDA近似。含自旋版本的托马斯-费米-狄拉克泛函需要将两种自旋分开计算，因为它们在(6.9)中就是分开的（不然也就不会导致等效的自旋-自旋相互作用了）。回顾 E_X 的推导过程，我们先计算了轨道波函数的电子数密度，将它乘以2得到总的电子数密度，然后将(7.1.2)中的Fock项（是用轨道波函数计算出Fock项以后乘以2得到的）表示成总的电子数密度的泛函。对向上自旋重复以上步骤，则单独的向上自旋的电子贡献的交换能应该是 $E_X[2\rho_\uparrow(\mathbf{r})]/2$ ，因此含自旋的交换能是

$$E_{X,\text{spin}}[\rho_\uparrow, \rho_\downarrow] = \frac{1}{2}(E_X[2\rho_\uparrow(\mathbf{r})] + E_X[2\rho_\downarrow(\mathbf{r})]). \quad (7.34)$$

7.2.3 GGA近似

从LDA近似的推导中可以清楚地看到，一旦离开了均匀电子气模型，LDA近似的可靠性就不好说了。对那些电子密度确定会有比较大的空间起伏的系统——比如说在我们需要精确计算分子结构时——LDA近似的精度不是特别好，此时有必要引入GGA近似。

GGA近似的局限性

与LDA近似相似，GGA近似同样无法非常好地处理电子气有非常明显的空间起伏的系统。

7.2.4 杂化泛函：将Hartree-Fock近似结合进DFT

LDA近似和GGA近似都是所谓的“显式密度泛函”，即它们和电子数密度之间真的能够写下一个解析式。但是我们也注意到，Kohn-Sham波函数 $\phi_{n\mathbf{k}}(\mathbf{r}\sigma)$ 本身其实是电子数密度和 V_{ion} 的泛函，因此，如果 E_{XC} 依赖于Kohn-Sham波函数，其实是无伤大雅的：因为这这不过是说， E_{XC} 同时依赖于 $\rho(\mathbf{r})$ 和 $V_{\text{ion}}(\mathbf{r})$ ，而我们知道 $\rho(\mathbf{r})$ 和 $V_{\text{ion}}(\mathbf{r})$ 包含了同样的信息。因此其实我们可以不拘泥于写下显式的 $E_C[\rho(\mathbf{r})]$ ，完全可以让它依赖于Kohn-Sham波函数，这种做法仍然在密度泛函理论的框架中。

允许交换关联泛函依赖于Kohn-Sham波函数有很大好处。既然眼下写出万能的泛函不太可能，我们往往需要针对不同的体系，从一族泛函中找出来一个。手动地根据不同的系统选择不同的泛函是很常见的（如对电子数密度有很大空间起伏的系统最好不要用LDA），而允许交换关联泛函依赖于Kohn-Sham波函数，可能可以让泛函自动地适合我们要计算的系统。

例如，一个立刻可以注意到的结论就是，自洽Hartree-Fock近似实际上可以看成DFT的一种特例。我们称它对应的泛函为**Hartree-Fock泛函**，记作 E_{HF} （这里不写出 $\rho(\mathbf{r})$ ，因为它和 $\rho(\mathbf{r})$ 的依赖关系是隐式的）。

对有非常局域的电子的系统，LDA甚至GGA泛函可能是不够准确的，如前所述，它们会高估电子的离域特性，或者说，相比于实际情况，LDA或是GGA中的电子会被**过离域化**。自洽Hartree-Fock近似可能能够给出更好的结果，虽然它完全忽略了电子关联，但是至少其Fock项没有做“电子数密度均匀”的假设，因此没有过离域化。此外，Hartree-Fock泛

函由于是诸 $\phi_i(\mathbf{r})$ 组成的Slater行列式态，不存在自相互作用误差。Hartree项和Fock项均存在自相互作用误差，但是它们加起来之后，自相互作用误差完全就没有了。例如，在只有一个电子，且该电子自旋极化时，(6.9)中的Hartree项和Fock项都非零，但是它们彼此抵消了。既然LDA或GGA泛函存在过离域化问题、存在自相互作用误差，但是能够捕捉到电子关联，而Hartree-Fock泛函没有过离域化问题、没有自相互作用误差，但是没有电子关联，一种很自然能够想到的办法是，不如将Hartree-Fock泛函和某些普通的泛函按照一定比例混合起来，可能能够兼顾关联能和电子的定域特性。这就是**杂化泛函**。

虽然杂化泛函实际上仍然可以看成是KS-DFT框架下的，但如何将普通的LDA或是GGA泛函和HF泛函组合起来并没有特别明确的标准。相反，“传统的DFT”中的泛函——LDA、GGA等方法——至少是相对步骤清晰地计算满足特定条件的电子气的能量和电子数密度的关系而得到的。一些人会用DFT来指代后者，而将杂化泛函当成一种“标准的DFT以外”的东西。

7.2.5 DFT+U

DFT+U方法是另一种解决电子过离域化的方法。

7.2.6 范德瓦尔斯力

7.2.7 交换关联泛函的理论研究

关于DFT为什么能够工作，以及它什么时候不能工作的理论研究已有不少。

交换关联洞

在^[21]和^[13]的有关章节对交换关联洞做了简单的介绍。

Hubbard模型的密度泛函理论

Hohenberg-Kohn定理的证明显示出，我们不仅仅可以对库伦排斥电子气定义密度泛函理论，实际上也可以对其它定理定义密度泛函理论。这样做有一个好处，就是，如果我们研究一个和库伦排斥电子气关系紧密，但是相对简单的理论的DFT，那么可以观察DFT是怎么工作的，并顺便测试它什么时候会失灵，从而对真正第一性原理计算中的DFT何时适用何时需要小心有一定的理解。

在^[8]中，作者们为Hubbard模型构建了一个密度泛函，在不同情境下使用该密度泛函理论预测Hubbard模型的有关性质。

7.3 Kohn-Sham方程的离散化和求解

算法1没有给出如下具体内容：如何引入凝聚态系统中原子核的位置、携带的电荷（从而确定材料中的电子总量）等信息，即如何适当地将 V_{ion} 和 μ 引入以列写Kohn-Sham方程；(7.12)本身是微分方程，它需要做适当的离散化才能被计算机求解，由于它是本征值问题，我们可以将 $\phi_i(\mathbf{r})$ 函数在适当的基底（量子化学中通常称为**基组**）下展开，这样只需要求解矩阵的本征值问题。

对一个一般的凝聚态系统，(7.12)中涉及的单电子波函数的个数要比系统中电子的数目（ 10^{23} 量级）还要多，否则 $\rho(\mathbf{r})$ 的总量不可能正确。介质中的电子数目是非常多的，从而直

接求解(7.12)并不现实。然而晶体有周期性结构，这就大大简化了计算，因为只需要计算一个晶胞内的情况即可。实际上，由于大部分情况下相互作用衰减得很快，我们完全可以假定液体等也有周期性边界条件，只不过此时的晶胞要取得非常大。于是以下我们将DFT计算中被认为是周期性的单元称为**超胞 (supercell)**，对固体通常要求超胞取最简单的形式，即原胞，对单分子、液体或是有杂质的系统等则需要取一个较大的超胞。对杂质，超胞应该包括杂质周围足够大的体态。对单分子，当超胞足够大时第一布里渊区足够小，从而计算出的所谓“能带”实际上和单独的一个个能级毫无区别。我们需要指定超胞的几何形状，并且在超胞内给出各个原子核的位置。

凝聚态研究中，由于系统的离散平移对称性，(7.12)此时也具有离散平移对称性，它的解——也就是 $\phi_i(\mathbf{r})$ ——可以看成某个以我们设定的超胞为原胞的晶体中的单电子波函数，从而Bloch定理(2.92)成立，且 $\phi_i(\mathbf{r})$ 中的 i 应该替换成Bloch波矢 \mathbf{k} ，能带编号 n 和自旋 σ 。（同样，到这里，所谓“能带编号”也未必对应着实际的系统中的能带电子的能带编号；如何诠释它还是不确定的）

在下面介绍的各种方法中可以看到，虽然DFT没有任何经验参数，并且将每一条可能的能带都纳入考虑，截断能、 \mathbf{k} 点采样、赝势的选择、泛函的选择会显著影响计算精度，对它们的选择是需要一些经验的，虽然我们的确不需要根据实验数据拟合一些经验参数输入计算程序——或者，有时候，DFT模拟中其实也会出现经验参数。

7.3.1 基于超胞和平面波的Kohn-Sham方程朴素求解方法

平面波基组下的Kohn-Sham方程

求解 $\phi_{n\mathbf{k}}(\mathbf{r})$ 就是要求解 $u_{n\mathbf{k}}(\mathbf{r})$ 。由于(2.92)中的 $u(\mathbf{r})$ 具有周期性，它可以被展开成用倒格矢标记的一系列傅里叶分量，因此我们有（为书写方便暂时略去自旋指标，即 ϕ 和 c 都是二分对象；需注意由于可能有自旋-轨道耦合，自旋不能被看作是附加在能带指标上的标签）

$$\phi_{n\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{V_{\text{u.c.}}}} \sum_{\mathbf{G}} c_{n,\mathbf{k}+\mathbf{G}} e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}}, \quad (7.35)$$

或者，考虑到 \mathbf{k} 局限在第一布里渊区中而 \mathbf{G} 是倒格矢，从 $\mathbf{k} + \mathbf{G}$ 可以唯一确定 \mathbf{k} 和 \mathbf{G} ，有

$$\phi_{n\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{V_{\text{u.c.}}}} \sum_{\mathbf{G}} c_{n,\mathbf{k}+\mathbf{G}} e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}}. \quad (7.36)$$

这种选取 ϕ_i 的方法称为**平面波DFT (PWDFT)**。用平面波做基底的好处在于，首先，对自由电子气模型适用的体系它有直接的物理意义，其次，将波函数展开为平面波有成熟的快速傅里叶变换算法，这样效率很高。这样做也有坏处，首先是只能处理晶格系统（否则需要不可数无限多个平面波），单分子必须放在一个比较大的晶格中才能够正确计算；其次，对高度局域的电子态以及原子核附近波函数的剧烈振荡，需要大量平面波，效率低下，这点可以部分通过赝势解决。

除了平面波以外还有很多常用的基底，比如高斯波包。以高斯波包为基底是可以不需要任何赝势的，它能够更加精确高效地处理低轨道电子，从而相较于PWDFT能够更好地处理过渡金属等。

(7.36)中的 \mathbf{G} 分量在(7.12)的动能项部分为 $(\mathbf{k} + \mathbf{G})^2/2m$ ，由于实际能够存储的傅里叶分量的个数是有限的，通常会引入一个**截断能**，对每个 \mathbf{k} ，我们只保留对动能的贡献小于截断能的那些 \mathbf{G} 。这意味着我们忽略了空间尺度特别小的现象，并且忽略了一部分能量，但是如果对应的 $c_{n,\mathbf{k}+\mathbf{G}}$ 不大就没有什么问题。

实际能够处理的晶体都是有限大小的，或者说实际的 \mathbf{k} 的取值是第一布里渊区中离散的一些点。这就是说，我们需要给出一种“第一布里渊区中的动量采样”。显然，采样越密越好；可以使用 $\mathbf{k} \cdot \mathbf{p}$ 来简化计算。

最后我们处理至今尚未提及的 V_{ion} 。原则上可以将 V_{ion} 取成不同位置的原子的原子核的库伦势场的叠加，这相当于将材料中的所有电子都做了计算，即做了**全电子计算**，当然，这样在原子核附近会有非常剧烈的波函数变化，并且高度局域的内层电子都会出现，从而截断能必须取得非常高，基本上是不现实的。一般来说用不到全电子计算，因为化学反应的能级不足以激发内层电子，而如第4.1.3节所述，外层电子感受到的势场可以用赝势取代：可以只计算外层电子，完全不计算内层电子，从而自始至终 V_{ion} 都是用赝势组合而成的，此时截断能可以设置得比较小，因为 $\phi_{n\mathbf{k}}(\mathbf{r})$ 在靠近原子核的位置也不会发生剧烈振荡。但这样就有一个问题：密度泛函理论需要**全空间**的电子数密度，但是使用了赝势之后，原子核附近的电子数密度是不准确的。如果需要使用赝势，Kohn-Sham方程中能够直接组合得到电子数密度的 $\phi_i(\mathbf{r})$ 实际上**不是**计算中实际求解出的**赝Kohn-Sham波函数**，赝Kohn-Sham波函数服从的方程相比原本的Kohn-Sham方程是需要修改的。本节采用最为朴素的办法，直接将 V_{ion} 当成库伦势处理，因此本节的朴素的算法其实是不现实的：纯平面波方法基本上都是赝势方法。

综上，朴素的PWDFT计算需要的输入信息有：

- 晶体结构，这包括超胞以及各个原子在超胞内部的位置以及各个原子的类型；
- 各个原子的原子序数，用以确定核电荷；
- 第一布里渊区采样，即(7.36)中 \mathbf{k} 的密集程度；
- 和材料的结构、原子、 \mathbf{k} 点采样都无关的东西，如截断能、泛函选取等。

不考虑自旋的朴素PWDFT算法

我们现在可以给出完整的平面波计算的方程和步骤了。首先考虑一个没有必要考虑自旋极化的系统，此时只需要求解轨道部分即可。以下我们给出一个最简单的PWDFT算法，不考虑 f ，则电子数密度为

$$\rho(\mathbf{r}) = 2 \sum_{n, \mathbf{k} \text{ occupied}} \phi_{n\mathbf{k}}(\mathbf{r}), \quad (7.37)$$

而 $\phi_{n\mathbf{k}}$ 展开为

$$\phi_{n\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{V_{\text{u.c.}}}} \sum_{(\mathbf{k}+\mathbf{G})/2m \leq E_{\text{cut}}} c_{n, \mathbf{k}+\mathbf{G}} e^{i(\mathbf{k}+\mathbf{G}) \cdot \mathbf{r}}, \quad (7.38)$$

在此平面波基组下，Kohn-Sham方程是

$$\begin{aligned} & \sum_{(\mathbf{k}+\mathbf{G}')/2m \leq E_{\text{cut}}} \left(\frac{(\mathbf{k}+\mathbf{G}')^2}{2m} \delta_{\mathbf{G}\mathbf{G}'} + V_{\text{ion}}(\mathbf{G}-\mathbf{G}') + V_{\text{H}}(\mathbf{G}-\mathbf{G}') + V_{\text{XC}}(\mathbf{G}-\mathbf{G}') \right) c_{n, \mathbf{k}+\mathbf{G}'} \\ & = \epsilon_{n, \mathbf{k}+\mathbf{G}} c_{n, \mathbf{k}+\mathbf{G}}, \end{aligned} \quad (7.39)$$

其中动量空间的势能由

$$V_{\text{ion}}(\mathbf{r}) = \frac{1}{\sqrt{V_{\text{u.c.}}}} \sum_{\mathbf{G}} V(\mathbf{G}) e^{i\mathbf{G} \cdot \mathbf{r}} \quad (7.40)$$

定义， $V_{\text{H}}(\mathbf{G})$ 和 $V_{\text{XC}}(\mathbf{G})$ 类同。将以上细节填充进算法1中，我们得到算法2，即所谓**静态自治计算**。（说是“静态”是因为有时候我们对晶体结构只有一个大概的猜测，此时需要做结构优化，即

算法 2: 给定晶格和原子赝势, 不考虑自旋的最朴素的PWDFT静态自洽计算

输入: 初始电子密度 $\rho_0(\mathbf{r})$, 交换关联泛函选取 $E_{\text{XC}}[\rho(\mathbf{r})]$, 不同原子的原子序数 $\{Z_i\}$, 原胞结构, \mathbf{k} 点采样 $K = \{\mathbf{k}\}$, 原胞中电子总数 N_e , 截断能 E_{cut} , 最大迭代次数 N_{max} , 所考虑的最大能带数目 n_{max} , 容差 ϵ

输出: Kohn-Sham波函数 $\phi_{n\mathbf{k}}(\mathbf{r})$ 和对应的本征值 $\epsilon_{n\mathbf{k}}$, 迭代次数 i

```

1 for  $\mathbf{k} \in K$  :
2   | 根据  $\frac{(\mathbf{k}+\mathbf{G})^2}{2m} < E_{\text{cut}}$  确定需要考虑的 $\mathbf{G}$ 的取值范围 $G$ 
3    $i = 1$ 
4   使用不同原子的原子序数和原胞结构构造 $V_{\text{ion}}(\mathbf{r})$ , 并按照(7.40)做傅里叶变换
5   使用 $\rho_0(\mathbf{r})$ 计算 $V_{\text{H}}$ 和 $V_{\text{XC}}$ , 并按照(7.40)做傅里叶变换
6   将 $V_{\text{ion}}, V_{\text{H}}, V_{\text{XC}}$ 代入(7.39)求解, 对每个 $\mathbf{k}$ 求前 $n_{\text{max}}$ 个解, 得到 $\{c_{n,\mathbf{k}+\mathbf{G}}^{(1)}\}$ 和 $\{\epsilon_{n,\mathbf{k}+\mathbf{G}}^{(1)}\}$ 
7   将 $\{c_{n,\mathbf{k}+\mathbf{G}}^{(1)}\}$ 代入(7.38)计算出 $\{\phi_{n,\mathbf{k}+\mathbf{G}}^{(1)}\}$ 
8   将 $\{\phi_{n,\mathbf{k}+\mathbf{G}}^{(1)}\}$ 代入(7.37)计算得到 $\rho_1(\mathbf{r})$ 
9   while  $\rho_i(\mathbf{r})$ 和 $\rho_{i-1}(\mathbf{r})$ 的差别大于容差 $\epsilon$ 且 $i < N_{\text{max}}$  :
10  | 使用 $\rho_i(\mathbf{r})$ 计算 $V_{\text{H}}$ 和 $V_{\text{XC}}$ , 并按照(7.40)做傅里叶变换
11  | 将 $V_{\text{ion}}, V_{\text{H}}, V_{\text{XC}}$ 代入(7.39)求解, 对每个 $\mathbf{k}$ 求前 $n_{\text{max}}$ 个解, 得到 $\{c_{n,\mathbf{k}+\mathbf{G}}^{(i+1)}\}$ 和 $\{\epsilon_{n,\mathbf{k}+\mathbf{G}}^{(i+1)}\}$ 
12  | 将 $\{c_{n,\mathbf{k}+\mathbf{G}}^{(i+1)}\}$ 代入(7.38)计算出 $\{\phi_{n,\mathbf{k}+\mathbf{G}}^{(i+1)}\}$ 
13  | 将 $\{\phi_{n,\mathbf{k}+\mathbf{G}}^{(i+1)}\}$ 代入(7.37)计算得到 $\rho_{i+1}(\mathbf{r})$ 
14  |  $i = i + 1$ 
15 return  $i$ , 波函数 $\{\phi_{n,\mathbf{k}+\mathbf{G}}^{(i)}\}$ 和本征值 $\{\epsilon_{n,\mathbf{k}+\mathbf{G}}^{(i)}\}$ 

```

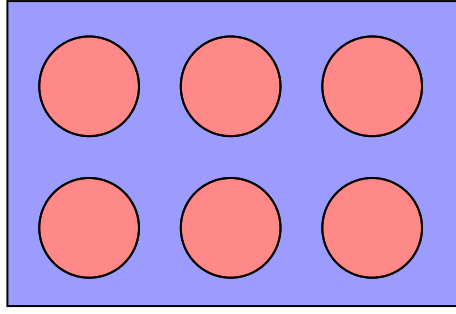


图 7.2: 蛋糕模具基组，蓝色的区域为平面波，红色的区域为角部分为球谐函数、径向部分为贝塞尔函数的一系列函数的线性组合

所谓结构优化自洽计算，这个是要变动晶体结构的，而静态自洽计算不变动晶体结构）只要输入的晶体信息是准确的，交换关联泛函的质量足够好，我们就得到了外层电子的模拟结果。

算法2有不少可以扩充的地方。在实际的DFT软件包中，通常不会将各次迭代的波函数和本征值留到计算结束再输出，而是在每次循环中把计算结果写入一个输出文件中，这样如果某一步计算失败了，已经写入的数据还可以用；也可以不以 $\rho_0(\mathbf{r})$ 为初始值而是以一组波函数为初始值；更新电子数密度时也可以将新的电子数密度和旧的电子数密度做一个混合，而不是用旧的电子数密度代替新的电子数密度。

7.3.2 全电子计算的APW系列方法

APW方法

如前所述，使用平面波方法计算全电子问题有现实中难以克服的困难，即需要非常多的平面波来模拟原子核附近的电子行为。那么一种很自然的想法就是，可以构造蛋糕模具基组：如图7.2所示，每个原子核附近的一个球形区域内（蛋糕模具区域），基函数的值由各向同性的函数——某个径向函数乘以球谐函数，然后对 l, m 线性组合——给出，而除此以外的区域（间隙区域）中则为平面波。这样的基组是所谓缀加平面波（augmented plane-wave, APW）。可以设

$$\phi_{Elm}(\mathbf{r}) = Y_{lm}(\hat{\mathbf{r}})R_{El}(r), \quad (7.41)$$

其中 E 是一个连续参数；需注意由于我们没有像求解真正的单原子问题一样，要求无穷远处波函数衰减得足够快，蛋糕模具区域内的Kohn-Sham波函数是需要不可数无穷多个基函数的。通常可以这样产生 $\phi_{Elm}(\mathbf{r})$ ：指定一个球对称势场 $V(r)$ ，设

$$-\frac{1}{r^2} \frac{d}{dr} \left(r^2 \frac{dR_{El}}{dr} \right) + \left(\frac{l(l+1)}{r^2} + V(r) \right) R_{El}(r) = ER_{El}(r). \quad (7.42)$$

现在给定一个晶胞中各个原子的位置 \mathbf{R}_α 后，我们将Kohn-Sham波函数 $\phi_{n\mathbf{k}}(\mathbf{r})$ 按照如下以 \mathbf{G} 标记的基底展开：

$$\phi_{\mathbf{k}+\mathbf{G}} = \begin{cases} e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}}, & \text{in the interstitial region,} \\ \sum_{l,m} a_{lm} \phi_{Elm}(\mathbf{r} - \mathbf{R}_\alpha), & \text{in muffin-tin balls,} \end{cases} \quad (7.43)$$

其中 a_{lm} 指定为

$$a_{lm} = 4\pi e^{i\mathbf{k}\cdot(\mathbf{r}-\mathbf{R}_\alpha)} i^l Y_{lm}^*(\hat{\mathbf{k}}) j_l(k|\mathbf{r}-\mathbf{R}_\alpha|) / R_{El}(|\mathbf{r}-\mathbf{R}_\alpha|), \quad (7.44)$$

这样可以满足蛋糕模具区域边界上波函数值连续的条件（导数则未必连续）。然后我们发现，只要蛋糕模具球的半径不太大，能够保证球内本原子的库伦吸引势能占据主导地位，球内的 $V_{\text{eff}}(\mathbf{r})$ 的三个部分——晶格势能（此时就是单个原子的库伦势 Ze/r ）、Hartree项、交换关联等效势能——都可以认为是只依赖于 $|\mathbf{r} - \mathbf{R}_\alpha|$ ，具有球对称性。而如果我们选取(7.43)中的 E 为 $\epsilon_{n\mathbf{k}}$ ，就能够保证各个蛋糕模具球中Kohn-Sham方程自动成立，只需要求解间隙区域内的Kohn-Sham方程即可。因此我们就得到了一种自洽求解全电子Kohn-Sham方程的方法：首先从某个试探电子数密度出发计算 V_{XC} ，将结果和试探 E 代入(7.42)中，对不同的 \mathbf{k} 计算得到 $\{\phi_{\mathbf{k}+\mathbf{G}}\}_{\mathbf{G}}$ ，然后据此基底求解Kohn-Sham方程，更新电子数密度和 E 。

APW方法的主要不足在于除了电子数密度以外， E 也需要自洽求解，或者说Kohn-Sham方程在此框架下是非线性的（ $\phi_{\mathbf{k}+\mathbf{G}}$ 非线性地依赖于 E ）。这个方法现在已经极少使用了，取而代之的是更加高效的LAPW方法。

LAPW方法

为了避免自洽求解 E ，我们不妨做一个 $\phi_{Elm}(\mathbf{r})$ 和 E 的关系的拟设。我们选定一个参考能量 E_0 ，在它附近计算 $\phi_{\mathbf{k}+\mathbf{G}}$ 和 $\partial\phi_{\mathbf{k}+\mathbf{G}}/\partial E$ ，以此计算 E 点的 $\phi_{\mathbf{k}+\mathbf{G}}$ 的线性估计值（所谓的线性化缀加平面波（LAPW）），然后用它展开 $\phi_{n\mathbf{k}}$ 并求解Kohn-Sham方程。为了增加一点灵活性，我们不固定 $\phi_{\mathbf{k}+\mathbf{G}}$ 和 $E - E_0$ 之间的关系，并可以对每个 l 都选取一个不同的 E_0 （通常取为主要由轨道角动量为该 l 的轨道构成的能带的中间能量），设

$$\phi_{\mathbf{k}+\mathbf{G}} = \begin{cases} e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}}, & \text{in the interstitial region,} \\ \sum_{l,m} (a_{lm}\phi_{E_0lm}(\mathbf{r} - \mathbf{R}_\alpha) + b_{lm} \frac{\partial\phi_{E_0lm}}{\partial E} \Big|_{E=E_0, \mathbf{r}=\mathbf{R}_\alpha}), & \text{in muffin-tin balls.} \end{cases} \quad (7.45)$$

由于实际的Kohn-Sham波函数通常是比较光滑的，我们不妨用 $\phi_{\mathbf{k}+\mathbf{G}}$ 在蛋糕模具球边界上的导数连续的条件确定 a_{lm} 和 b_{lm} 。然后我们要求 E 总是 $\epsilon_{n\mathbf{k}}$ ，即可复用APW方法的求解过程。

7.3.3 赝势方法

全电子计算常常是不必要的，因此现在我们转而考虑怎么将赝势引入DFT计算。做DFT计算实际上就是求解(7.39)，这形式上仍然是一个单电子方程，因此总的来说，“计算单原子波函数，然后把外层电子高频振荡抹平而拟合出等效势”的做法仍然是适用的。第4.1.3节中的做法不考虑任何电子-电子相互作用而计算内层电子的轨道，即把介质中的原子当成类氢原子，但是这样无疑精度非常糟糕。我们要做的是在Kohn-Sham方程求解中适当地引入赝势。主要的问题在于，由于密度泛函理论依赖全空间的电子数密度，使用赝势求解得到的赝波函数和 $\rho(\mathbf{r})$ 之间没有非常简单的(7.5)这样的关系，因为在原子核附近，用赝Kohn-Sham波函数根据(7.5)计算出来的电子数密度是不正确的，这也意味着赝Kohn-Sham波函数服从的方程中的 V_{XC} 不是简单的 $\delta E_{\text{XC}}/\delta\rho$ 。因此，除了构造柔和的势能以外，我们还需要找到一种从赝Kohn-Sham波函数获得正确的电子数密度的方法。不同种类的赝势的从赝Kohn-Sham波函数获得正确的电子数密度的方法当然也是不一样的，或者说，不同种类的赝势对应的赝Kohn-Sham方程中 V_{XC} 的形式是不一样的。

我们做**核心电子冻结近似**，即认为晶体中其它原子的存在不会改变内层电子的行为（特别的，不会改变原子核附近的电子数密度），从而，使用单原子计算结果拟合出来的赝势在晶体中仍然是适用的。我们实际上也隐含地假定Kohn-Sham方程求解出来的本征值较低的Kohn-Sham波函数对应着内层电子；这件事倒是没有什么争议，因为这些本征值较低

的Kohn-Sham波函数贡献的电子数密度确实集中在原子核附近，它们肯定对应着内层电子的电子数密度，而由核心电子冻结近似，把它们丢弃是没有问题的，因为反正内层电子的电子数密度不怎么变化。

设我们对单原子已经求解出了如下的Kohn-Sham方程：

$$-\frac{\nabla^2}{2m}\phi_{nlm}(\mathbf{r}) - \frac{Ze^2}{r}\phi_{nlm}(\mathbf{r}) + V_{XC}[\rho(\mathbf{r})]\phi_{nlm}(\mathbf{r}) = \epsilon_n\phi_{nlm}(\mathbf{r}), \quad \rho(\mathbf{r}) = 2 \sum_{n,l,m} |\phi_{nlm}(\mathbf{r})|^2. \quad (7.46)$$

这里的 n, l, m 和原子物理中惯用的记号完全一致，由于系统具有旋转不变性， $\rho(\mathbf{r})$ 也具有旋转不变性，因此 V_{XC} 也具有，因此Kohn-Sham本征值只和主量子数 n 有关。我们将电子分成内层电子和外层电子（或者说价电子），前者被认为固定不动的，可以计算出其数密度：

$$\rho_{\text{core}}(\mathbf{r}) = 2 \sum_{\text{core } n, l, m} |\phi_{nlm}(\mathbf{r})|^2. \quad (7.47)$$

完整的数密度就是用价电子Kohn-Sham波函数计算出来的数密度加上 $\rho_{\text{core}}(\mathbf{r})$ 。进一步，将 r_C 内的 $\phi_{nlm}(\mathbf{r})$ “抹平”，得到 $\tilde{\phi}_{nlm}(\mathbf{r})$ ，然后根据 $\phi_{nlm}(\mathbf{r})$ 和 $\tilde{\phi}_{nlm}(\mathbf{r})$ 找到一种能够将赝Kohn-Sham波函数映射到原Kohn-Sham波函数并由此得到全空间的电子数密度的方法，并根据

$$-\frac{\nabla^2}{2m}\tilde{\phi}_{nlm}(\mathbf{r}) + V_{PP}\tilde{\phi}_{nlm}(\mathbf{r}) + V_{XC}[\rho(\mathbf{r})]\tilde{\phi}_{nlm}(\mathbf{r}) = \epsilon_n\tilde{\phi}_{nlm}(\mathbf{r}), \quad \rho(\mathbf{r}) = \rho_{\text{core}}(\mathbf{r}) + \rho_v[\{\tilde{\phi}_i(\mathbf{r})\}] \quad (7.48)$$

计算出赝势 V_{PP} （注意它可能不是局域的）。于是，求解晶体中的Kohn-Sham方程

$$-\frac{\nabla^2}{2m}\phi_{n\mathbf{k}}(\mathbf{r}) + V_{\text{ion}}(\mathbf{r})\phi_{n\mathbf{k}}(\mathbf{r}) + V_{XC}[\rho(\mathbf{r})]\phi_{n\mathbf{k}}(\mathbf{r}) = \epsilon_{n\mathbf{k}}\phi_{n\mathbf{k}}, \quad (7.49)$$

$$\rho(\mathbf{r}) = 2 \sum_{\text{occupied } n, \mathbf{k}} |\phi_{n\mathbf{k}}(\mathbf{r})|^2$$

就转化为求解

$$-\frac{\nabla^2}{2m}\tilde{\phi}_{n\mathbf{k}}(\mathbf{r}) + V_{PP}(\mathbf{r})\tilde{\phi}_{n\mathbf{k}}(\mathbf{r}) + \frac{\delta V_{XC}[\rho(\mathbf{r})]}{\delta \rho(\mathbf{r})} \frac{\delta \rho_v}{\delta \tilde{\phi}_{n\mathbf{k}}^*(\mathbf{r})} = \epsilon_{n\mathbf{k}}\tilde{\phi}_{n\mathbf{k}}, \quad (7.50)$$

$$\rho(\mathbf{r}) = \rho_v[\{\tilde{\phi}_{n\mathbf{k}}\}] + \sum_{\mathbf{i}, \alpha} \rho_{\text{core}}(\mathbf{r} - \mathbf{R}_{\mathbf{i}\alpha}),$$

其中 α 标记一个晶胞中不同的原子。只要核心电子冻结近似成立，求解以上赝Kohn-Sham方程组就充分地考虑了内层电子对价电子的排斥作用、充分地考虑了内层电子对原子核的库伦势场的屏蔽，等等，求解得到的在 r_C 外的Kohn-Sham波函数和全电子计算一致，并且可以使用比较少的平面波。

模守恒赝势

超软赝势

相较于模守恒赝势，超软赝势需要更少的计算资源。然而，超软赝势给出的外层电子波函数不能强行归一化，否则在蛋糕模具区域以外的部分反而不正确了。

7.3.4 PAW方法

赝势方法可以用很少的平面波非常快速地计算各种系统，但是无法做全电子计算，而且合理性存疑；LAPW等方法可以做全电子计算，但是非常慢，而且合理性同样存疑。

投影缀加波法 (projector augmented wave method, PAW) 是一种能够同时兼顾两种需求的方法, 超软赝势和LAPW方法都是它的特例。实际上, PAW方法的推导是如此简单和清晰, 以至于我们甚至可以反过来, 用它作为对赝势方法和LAPW方法的辩护。

我们总是可以在赝Kohn-Sham波函数和真实的Kohn-Sham波函数 (在PAW方法的语境下这称为全电子Kohn-Sham波函数, 因为它和全电子计算算出来的是一样的; 或者说, 全电子波函数不是多体波函数) 之间建立一个线性映射, 不妨写为

$$\phi_{n\mathbf{k}} = \mathcal{T} \tilde{\phi}_{n\mathbf{k}}. \quad (7.51)$$

如果我们要求赝波函数在蛋糕模具球以外和全电子波函数完全一致, \mathcal{T} 算符就可以写作

$$\mathcal{T} = 1 + \sum_{\alpha} \mathcal{T}_{\alpha}, \quad (7.52)$$

其中 \mathcal{T}_{α} 将 α 号原子周围的蛋糕模具球中的赝波函数映射成全电子波函数。

设我们DFT求解孤立原子已经得到了Kohn-Sham波函数

$$\phi_{nlm}(\mathbf{r}) = R_{nl}(r)Y_{lm}(\hat{\mathbf{r}}), \quad (7.53)$$

我们将它在蛋糕模具球以外截断。通过一定方法我们可以让 $R_{nl}(r)$ “柔化”, 减少它的零点, 这样能够得到

$$\tilde{\phi}_{nlm}(\mathbf{r}) = \tilde{R}_{nl}(r)Y_{lm}(\hat{\mathbf{r}}). \quad (7.54)$$

由于 $\tilde{\phi}_{nlm}$ 之间未必正交, 我们需要求出一组投影波函数 $\{d_{nlm}\}$, 满足

$$\langle p_{nlm} | \tilde{\phi}_{n'l'm'} \rangle = \delta_{nn'} \delta_{ll'} \delta_{mm'}, \quad (7.55)$$

则一种很好的选择是

$$\mathcal{T}_{\alpha} = \sum_{n,l,m} (|\phi_{nlm}\rangle - |\tilde{\phi}_{nlm}\rangle) \langle p_{nlm} | |_{\mathbf{r} \rightarrow \mathbf{r} - \mathbf{R}_{\alpha}}. \quad (7.56)$$

这就得到了 \mathcal{T} 。

我们现在指定了一种线性的从赝Kohn-Sham波函数获得全电子Kohn-Sham波函数。现在任何使用全电子Kohn-Sham波函数计算的形如 $\langle \phi | A | \phi' \rangle$ 的式子可以全部改写成关于赝Kohn-Sham波函数的式子了——只需要将 ϕ 换成 $\tilde{\phi}$, 将 A 换成 $\mathcal{T}^{\dagger} A \mathcal{T}$ 即可。特别的, 电子数密度现在是

$$\rho(\mathbf{r}) = \sum_{n,\mathbf{k}} f_{n\mathbf{k}} \int d^3\mathbf{r}' |\psi_{n\mathbf{k}}(\mathbf{r}')|^2 \quad (7.57)$$

可以看到PAW方法是能够给出全电子计算精度的结果的。如果没有任何电子被划为“冻结”的核心电子, 那么PAW就是全电子计算; 如果有一些电子被划为冻结的核心电子, 那么PAW算出的价电子仍然是全电子精度的。因此PAW方法已经被当成全电子方法了, 但是它似乎也可以看成一种赝势方法。一个PAW赝势需要包含蛋糕模具球里面的 ϕ_{nlm} 和 $\tilde{\phi}_{nlm}$, 以及核心电子的Kohn-Sham波函数。

7.3.5 Kohn-Sham方程的求解

由于赝势不是局域的, V_{eff} 也不是。

资料框 7.1: VASP软件包

VASP是维也纳大学维护的一个商业DFT包，基于PWDF，支持的赝势类型包括模守恒赝势、超软赝势和PAW赝势。其输入文件有四个，正好对应四种输入：

- POSCAR，给出一个平行六面体超胞的三个初基格矢，以及各个原子的类型、数目，用笛卡尔坐标系或是晶体坐标系给出的坐标。
- POTCAR，赝势文件。
- KPOINTS， \mathbf{k} 动量采样（注意不是 \mathbf{G} 采样， \mathbf{G} 采样由根据POSCAR计算出来的倒格矢和INCAR中给定的截断能规定），需要指定第一布里渊区中被纳入考虑的 \mathbf{k} 。VASP提供多种自动生成这些 \mathbf{k} 点的方法。
- INCAR，计算中的参数，截断能、泛函选择、最大计算步数、能带数目、容差、蛋糕模具球半径等。

关于VASP使用的详细信息可以参考VASP笔记。

7.4 物理量测量和后处理

本节讨论Kohn-Sham方程求解完成后怎样从计算结果解读出我们需要的各个物理量，在第一性原理计算的语境下也可以称为“非自洽计算”，因为就是拿到了电荷密度，然后算各种东西，无需迭代求解Kohn-Sham方程。这件事是没那么容易的，基本上和基态电子密度有解析形式关系的量就是基态能量，基态能量实际上就是能量泛函作用在基态电子密度 $\rho(\mathbf{r})$ 之后的结果。除此以外的东西——比如第一激发态能量——和Kohn-Sham方程求解结果之间都没有特别平凡的关系。

7.4.1 单电子诠释和能带

Kohn-Sham方程的解的物理意义

在待计算的系统确实适用单电子诠释时，我们将(7.12)和(4.4)对比，很容易看出两者的相似之处。如果Hartree项加上 V_{XC} 构成对自能 Σ 的一个良好模拟，那么(7.12)解出的 ϵ_i 就是经过相互作用修饰的能带电子的能量，或者说是费米液体中的 $\epsilon_{\mathbf{k}}$ 。如果实际情况确实如此，那么(7.3)中的 μ 实际上就是费米能，因为对比(7.3)和(7.11)会发现

$$\mu = \sum_{i=1}^N \epsilon_i. \quad (7.58)$$

当然，如前所述，Kohn-Sham方程的推导的每一步都没有给出足够的条件让我们知道Kohn-Sham波函数到底是什么东西。Kohn-Sham方程解得的本征值和本征函数到底是什么实际上是仍然有疑问的。的确，如果待模拟的系统实际上是一个费米液体，其低能激发适用能带理论，那么其基态的确大体上可以看成某种费米子算符——经过相互作用修正的能带电子——的Fock波函数，和Kohn-Sham拟设形式相同。因此，Kohn-Sham波函数大体上就是“单电子波函数”。但我们不知道这个近似的精度如何。如果Kohn-Sham波函数和单电子波函数只是大体上相似，而在一些细节上有差异呢？在量子化学中，这样的差异

可能就是关键的。如果能带理论会带来一个方向的误差，而将计算得到的Kohn-Sham波函数当成单电子波函数会产生相反方向的误差呢？

有一点是确定的，就是一般来说，Kohn-Sham波函数和本征值**不严格**是单电子能量和波函数，因为已经证实，把它们当成单电子能量和波函数会产生系统的误差。这其中最为著名的就是**半导体能隙问题**，即将DFT计算出来的Kohn-Sham本征值当作电子能量的话会倾向于系统地低估半导体的能隙，从而实验上明确确定是半导体的体系用DFT计算出来的能带结构看起来似乎是金属的。关于这一误差到底在哪里没有特别明确的结论，因为DFT中不确定的、物理意义不清楚的东西非常多，首先我们不知道LDA和GGA这些非常局域的方法以及它们的推广是否足够精确，其次我们也不知道Kohn-Sham方程的解的物理意义是什么。

无论如何，Kohn-Sham方程的解至少提供了对单电子能量和波函数的一种估计，这意味着它可以作为物理意义清楚的场论方法的初始值。如果真的如果需要非常可靠的能谱，总是可以以Kohn-Sham方程的解为初始值做诸如GW这样的进一步计算。

由于绘制能带通常只绘制高对称点上的能带，如果直接将自洽求解Kohn-Sham方程的结果拿来用，要么浪费大量计算资源（因为Kohn-Sham波函数的 \mathbf{k} 必须比较均匀地覆盖第一布里渊区），要么精度不够（因为 \mathbf{k} 如果只是比较稀疏地覆盖第一布里渊区，那么其中落在高对称点上的 \mathbf{k} 个数不会很多）。不过，假定我们已经自洽求解得到了一个电子数密度 $\rho(\mathbf{r})$ （或者是含自旋的KSDF中的自旋密度矩阵 $\rho_{\alpha\beta}(\mathbf{r})$ ），如果我们将Kohn-Sham方程的解看成对单电子波函数的一个良好估计，那么实际上可以根据 $\rho(\mathbf{r})$ 构造出一个Kohn-Sham方程之后，**不做**自洽求解，而是只求解高对称点上的那些 \mathbf{k} 对应的Kohn-Sham波函数和本征值，这样就得到了高对称点上的能带。

如果赝势提供了“原子轨道波函数”，根据Kohn-Sham波函数还可以判断每条能带上的每个态有多少s轨道组成部分，多少p轨道组成部分，判断蛋糕模具球中的内层电子轨道的占据数、磁矩等。

GW计算

GW计算实际上是一种独立的数值计算方法，但是因为它要求一个比较接近实际情况的单电子波函数作为起点，我们仍然可以将它视为DFT的后处理的一部分，即将DFT计算出来的Kohn-Sham波函数作为GW计算的初始值。GW计算也是一种基于自能的计算方法，其中，它

7.4.2 原子受力

在计算单原子受力时我们最好把晶胞扩张得大一些。如果使用原胞做计算，那么将某个原子移动一小段得同时，其它原胞中与它地位一致的原子也发生了移动，因此通过这种方式计算原子移动前后的能量变化并进一步计算力是不准确的。相反，如果晶胞被扩张了，那么移动一个原子之后，与之最为接近的那些原胞中与之地位一致的原子并没有发生移动，会给出相对可靠的结果。

声子谱计算

7.5 结构优化和分子动力学模拟

有时候我们对晶体结构只有一个大概的猜测，此时需要做结构优化，即所谓结构优化自洽计算。

ab initio MD指的是不使用半经验的分子力场，直接使用密度泛函理论计算各个原子受力的分子动力学模拟。

第八章 低温下的电子序

由于不同的相互作用通道，凝聚态体系中可以有各种各样的电子凝聚：自旋密度波¹、电荷密度波、超导配对等。此时的系统不再是一个普通的费米液体了，相互作用改变了系统的元激发谱，我们也可以说，**特殊的相互作用通道让费米液体不稳定**。超导配对顾名思义，可能导致可以不受阻碍地随意移动的“超导电子”出现；所谓密度波实际上就是一种长程序，它在空间中有一种周期性振荡，可以是相位的振荡也可以是大小的振荡，也即，这样的长程序对应的序参量形如

$$\Delta(\mathbf{r}) \sim \Delta_0 e^{i\mathbf{Q} \cdot \mathbf{r}}. \quad (8.1)$$

请注意这个序参量实际上是一种二电子集体运动模式，即相邻的两个电子的自旋总是保持一上一下的这种运动模式。这暗示我们，如果需要超越平均场近似的理论，只需要把这种模式定义成一个新的（玻色）场就可以。本节讨论的这些现象对应的序参量均对应于Hubbard-Stratonovich参量（见背景知识8.3）的不同选择，即我们只讨论序参量能够写成两个电子场乘积的线性组合的那些序。更加物理地说，我们只讨论那些在对称性自发破缺发生之后，仍然有普通电子的系统，只不过这些电子受到由于对称性自发破缺，一些电子形成序而产生的较低对称性的势场的作用，能谱发生改变。相互作用电子体系中的不同的相互作用通道对应不同的Hubbard-Stratonovich参量。写出这些参量的场论后，我们会发现可能有多个参量都有对称性自发破缺，它们相互竞争，最终哪一种参量最终发生凝聚——从而，单单考虑它得到的相变出现——取决于很多条件。

这个图像并不能覆盖所有相变。例如在

8.1 BCS超导

本节介绍一种常见的超导机制，即交换声子导致电子出现有效吸引相互作用而产生的BCS超导。我们将主要讨论所谓的s波超导，其具体含义后面会提到。

8.1.1 交换声子导致的有效电子吸引相互作用

声子介导的电子-电子相互作用

电子-声子相互作用的顶角为一个电子入射，一个电子出射，产生/消灭一个声子。现在尝试积掉声子。我们先将电子场当成给定的，则可以从(3.48)写出声子场加上电子-声子相互作用的虚时间作用量。本节仅考虑晶格是简单正方晶格的情况，于是没有各向异性，

¹为了避免引起混淆我们要区分自旋密度波和自旋波。后者是一个自旋系统中的现象，前者是一个费米子系统中的现象。或者，更加形象地说，后者是固定在格点上的自旋的空间涨落，前者是可以运动的、携带自旋的费米子密度的空间涨落。

我们有（关于 τ 的傅里叶变换见(2.24)）

$$\begin{aligned} S_{\text{ph}} + S_{\text{int}} &= \int d\tau \sum_{\mathbf{q}, \lambda} \left(\bar{\phi}_{\mathbf{q}\lambda} (\partial_\tau + \omega_{\mathbf{q}\lambda}) \phi_{\mathbf{q}\lambda} + \gamma \frac{i q_\lambda}{\sqrt{2NM\omega_{\mathbf{q}\lambda}}} (\phi_{\mathbf{q}\lambda} + \bar{\phi}_{-\mathbf{q}\lambda}) \underbrace{\sum_{\mathbf{k}, \sigma} \bar{\psi}_{(\mathbf{k}+\mathbf{q})\sigma} \psi_{\mathbf{k}\sigma}}_{\rho_{\mathbf{q}}} \right) \\ &= \sum_{\mathbf{q}, \lambda} \left(\bar{\phi}_{\mathbf{q}\lambda} (-i\omega_n + \omega_{\mathbf{q}\lambda}) \phi_{\mathbf{q}\lambda} + \gamma \frac{i \mathbf{q} \cdot \hat{\boldsymbol{\lambda}}}{\sqrt{2NM\omega_{\mathbf{q}\lambda}}} (\phi_{\mathbf{q}\lambda} + \bar{\phi}_{-\mathbf{q}\lambda}) \sum_{\mathbf{k}, \sigma} \bar{\psi}_{(\mathbf{k}+\mathbf{q})\sigma} \psi_{\mathbf{k}\sigma} \right) \end{aligned}$$

其中 ϕ 表示声子， ψ 表示电子， M 是原子质量。使用处理自由场受到线性激励的方法，配平方并积掉二次项，忽略产生的因子，得到

$$\begin{aligned} S_{\text{eff}} &= - \sum_{\omega_n, \mathbf{q}, \lambda} \frac{\gamma^2 |\mathbf{q}|^2 \rho_{\mathbf{q}} \rho_{-\mathbf{q}}}{2\beta NM \omega_{\mathbf{q}\lambda} (\omega_{\mathbf{q}\lambda} - i\omega_n)} = - \sum_{\omega_n, \mathbf{q}, \lambda} \frac{\gamma^2 |\mathbf{q}|^2 \rho_{\mathbf{q}} \rho_{-\mathbf{q}}}{2\beta NM (\omega_{\mathbf{q}\lambda}^2 + \omega_n^2)} \\ &= - \frac{\gamma^2}{2\beta NM} \sum_{\omega_n, \mathbf{q}, \mathbf{k}, \mathbf{k}', \alpha, \beta} \frac{|\mathbf{q}|^2}{\omega_n^2 + \omega_{\mathbf{q}\lambda}^2} \bar{\psi}_{(\mathbf{k}+\mathbf{q})\alpha} \bar{\psi}_{(\mathbf{k}'-\mathbf{q})\beta} \psi_{\mathbf{k}'\beta} \psi_{\mathbf{k}\alpha}. \end{aligned} \quad (8.2)$$

第二个等号是考虑到每个 ω_n 有一个对应的 $-\omega_n$ 而得到的。总之，电子之间可以通过交换声子来产生一个四电子相互作用。相互作用(8.2)是推迟相互作用，因此原则上不能够仅仅使用一个哈密顿量描述。不过，在推迟不明显时我们还是可以近似写出一个哈密顿量。这里有一个微妙的地方：(8.2)是定义在虚时间下的，即算符的时间演化为 $e^{\omega t}$ ，而我们需要一个实时间下的哈密顿量，所以需要首先做频率上的Wick转动 $\omega = i\omega_n$ （但是另一方面作用量形式上仍然是虚时间场论的作用量，即作用量和哈密顿量之间不相差负号），得到

$$S_{\text{eff}} = - \frac{\gamma^2}{2\beta NM} \sum_{\omega} \sum_{\mathbf{q}, \mathbf{k}, \mathbf{k}', \alpha, \beta} \frac{|\mathbf{q}|^2}{-\omega^2 + \omega_{\mathbf{q}\lambda}^2} \bar{\psi}_{(\mathbf{k}+\mathbf{q})\alpha} \bar{\psi}_{(\mathbf{k}'-\mathbf{q})\beta} \psi_{\mathbf{k}'\beta} \psi_{\mathbf{k}\alpha}.$$

上式中的所有 ψ 都是 ω 的函数，可以将它们转换到时域，这将给出一个有推迟的作用量，从而不能简单地用一个哈密顿量描述电子-电子等效相互作用。上式的无推迟近似是它的无时间变化的傅里叶分量。因此实际上我们只需要简单地取

$$\omega = \epsilon_{\mathbf{k}} - \epsilon_{\mathbf{k}+\mathbf{q}}, \quad (8.3)$$

就得到了无推迟的近似²

$$\begin{aligned} H &= - \frac{\gamma^2}{2NM} \sum_{\mathbf{q}, \mathbf{k}, \mathbf{k}', \alpha, \beta} \frac{|\mathbf{q}|^2}{-\omega^2 + \omega_{\mathbf{q}\lambda}^2} c_{(\mathbf{k}+\mathbf{q})\alpha}^\dagger c_{(\mathbf{k}'-\mathbf{q})\beta}^\dagger c_{\mathbf{k}'\beta} c_{\mathbf{k}\alpha} \\ &= \frac{\gamma^2}{2NM} \sum_{\mathbf{q}, \mathbf{k}, \mathbf{k}', \alpha, \beta} \frac{|\mathbf{q}|^2}{(\epsilon_{\mathbf{k}} - \epsilon_{\mathbf{k}+\mathbf{q}})^2 - \omega_{\mathbf{q}\lambda}^2} c_{(\mathbf{k}+\mathbf{q})\alpha}^\dagger c_{(\mathbf{k}'-\mathbf{q})\beta}^\dagger c_{\mathbf{k}'\beta} c_{\mathbf{k}\alpha}. \end{aligned} \quad (8.4)$$

应注意，在不采取无推迟相互作用的近似时，关于 ω 我们仍然有能量守恒关系，但此时我们有的是

$$\omega = k^0 - (k+q)^0,$$

这是关于离壳的虚频率的能量守恒，而不是像(8.3)这样的在壳的能量守恒。

²通过简单地将

$$\sum_{\mathbf{k}, \mathbf{k}', \mathbf{q}} f(q^0, \mathbf{q}) \bar{c}_{\mathbf{k}+\mathbf{q}} \bar{c}_{\mathbf{k}'-\mathbf{q}} c_{\mathbf{k}'} c_{\mathbf{k}}$$

变换到虚时间时域就可以看出为什么作用量中耦合系数显含 q^0 会导致推迟相互作用，以及为什么取(8.3)就给出了这个推迟相互作用的瞬时近似。下同。

更加一般地, 设声子频率为 $\omega_{\mathbf{q}}$, 电子-声子相互作用为

$$H = \frac{1}{\sqrt{N}} \sum_{\mathbf{k}, \mathbf{q}, \lambda} M_{\mathbf{q}\lambda} c_{\mathbf{k}+\mathbf{q}}^\dagger c_{\mathbf{k}} (b_{\mathbf{q}\lambda} + b_{-\mathbf{q}\lambda}^\dagger), \quad (8.5)$$

由厄米性, 有

$$M_{-\mathbf{q}\lambda} = M_{\mathbf{q}\lambda}^*. \quad (8.6)$$

重复以上步骤, 频域下(动量空间加松原频率)的作用量是

$$S = \sum_{\mathbf{q}, \lambda} \bar{b}_{\mathbf{q}\lambda} (-iq^0 + \omega_{\mathbf{q}\lambda}) b_{\mathbf{q}\lambda} + \frac{1}{\sqrt{\beta N}} \sum_{\mathbf{k}, \mathbf{q}, \lambda} M_{\mathbf{q}} \bar{c}_{\mathbf{k}+\mathbf{q}} c_{\mathbf{k}} (b_{\mathbf{q}\lambda} + \bar{b}_{-\mathbf{q}\lambda}), \quad (8.7)$$

利用高斯积分

$$\int d(x, x^\dagger) e^{x^\dagger A x + b^\dagger x + x^\dagger b} \simeq e^{-b^\dagger A b},$$

积掉声子自由度, 得到

$$\begin{aligned} S_{\text{eff}} &= -\frac{1}{\beta N} \sum_{\mathbf{q}, \lambda} M_{\mathbf{q}\lambda} \sum_{\mathbf{k}, \sigma} \bar{c}_{\mathbf{k}+\mathbf{q}, \sigma} c_{\mathbf{k}\sigma} \frac{1}{-iq^0 + \omega_{\mathbf{q}\lambda}} M_{-\mathbf{q}\lambda} \sum_{\mathbf{k}', \sigma'} \bar{c}_{\mathbf{k}'-\mathbf{q}, \sigma'} c_{\mathbf{k}'\sigma'} \\ &= -\frac{1}{\beta N} \sum_{\mathbf{q}, \lambda} M_{\mathbf{q}\lambda} \sum_{\mathbf{k}, \sigma} \bar{c}_{\mathbf{k}+\mathbf{q}, \sigma} c_{\mathbf{k}\sigma} \frac{1}{2} \left(\frac{1}{-iq^0 + \omega_{\mathbf{q}\lambda}} + \frac{1}{iq^0 + \omega_{\mathbf{q}\lambda}} \right) M_{-\mathbf{q}\lambda} \sum_{\mathbf{k}', \sigma'} \bar{c}_{\mathbf{k}'-\mathbf{q}, \sigma'} c_{\mathbf{k}'\sigma'} \\ &= -\frac{1}{\beta N} \sum_{\mathbf{q}, \lambda} \sum_{\mathbf{k}, \mathbf{k}', \alpha, \beta} |M_{\mathbf{q}}|^2 \frac{\omega_{\mathbf{q}\lambda}}{\omega_{\mathbf{q}\lambda}^2 + (q^0)^2} \bar{c}_{\mathbf{k}+\mathbf{q}, \alpha} \bar{c}_{\mathbf{k}'-\mathbf{q}, \beta} c_{\mathbf{k}'\beta} c_{\mathbf{k}\alpha}. \end{aligned} \quad (8.8)$$

Wick转动会让 iq^0 变成 ω 。注意 $c_{\mathbf{k}\alpha}$ 这些算符中的 k^0 未必是 $\epsilon_{\mathbf{k}}$ 。将上式切换回到频域, 得到的是推迟相互作用。然而实际上晶格存在各种弛豫机制, 声子被一个电子产生之后不可能维持特别长的时间再和另一个电子散射。在前面的推导中我们没有将声子弛豫体现出来, 但是声子的寿命有限这件事让推迟相互作用的时间延迟有一个截断。因此我们仅仅考虑上式中那些瞬时发生的相互作用通道, 为此必须让相互作用耦合系数没有任何 k^0, k'^0, q^0 的依赖, 因此我们只考虑最重要的

$$\omega = \epsilon_{\mathbf{k}} - \epsilon_{\mathbf{k}+\mathbf{q}}, \quad (8.9)$$

切换到时域得到

$$H_{4e} = \frac{1}{2N} \sum_{\mathbf{k}, \mathbf{k}', \mathbf{q}'} \sum_{\alpha, \beta} |M_{\mathbf{q}}|^2 \frac{\omega_{\mathbf{q}\lambda}}{(\epsilon_{\mathbf{k}} - \epsilon_{\mathbf{k}+\mathbf{q}})^2 - \omega_{\mathbf{q}}^2} c_{(\mathbf{k}+\mathbf{q})\alpha}^\dagger c_{(\mathbf{k}'-\mathbf{q})\beta}^\dagger c_{\mathbf{k}'\beta} c_{\mathbf{k}\alpha}. \quad (8.10)$$

在只考虑二声子-电子的相互作用顶角时, 任何更复杂的含声子的费曼图都可以化归为二电子和声子的相互作用和一些电子相互作用的组合, 因此(8.10)就给出了完整的没有声子的有效理论。(8.10)本身是瞬时的, 但即使我们不知道它来自一个推迟相互作用, (8.10)含有电子能量差而(8.3)也有这一事实也暗示它有可能实际上代表一个推迟相互作用。另一方面, 库伦势则没有推迟(当然这实际上做了近似, 我们只不过不考虑相对论效应而已, 但是电磁相互作用的推迟远小于声子传递导致的推迟)。

在以上推导过程中应当注意一点: 我们在这里使用的实际上不是标准的声子传播子。(8.7)中的声子自由部分是直接根据声子哈密顿量(类似于(3.5))写出来的, 在此过程中我们没有考虑产生声子的场是实场这件事, 而是将 $b_{\mathbf{q}\lambda}$ 傅里叶变换到实空间中的结果当成了产生声子的场, 而这显然是一个复场。因此, 在作用量(8.7)中, 积掉声子得到的有效电子-电子作用量不需要乘以1/2。更加常见的, 声子传播子会被写成

$$\frac{i2\omega_{\mathbf{q}}}{\omega^2 - \omega_{\mathbf{q}}^2 + i0^+},$$

这个传播子实际上是 $\phi_{q\lambda} + \bar{\phi}_{-q,\lambda}$ 的格林函数，即我们将声子的产生湮灭算符打包成了一个实场，此时纯声子的作用量中将会出现 ω 的二次项以及一个1/2因子，而电子-声子相互作用顶角将只有一个，此时的高斯积分是

$$\int dx e^{\frac{1}{2}xAx+bx} \simeq e^{-\frac{1}{2}bAb},$$

于是(8.8)的第一行将会变成

$$\frac{1}{2} \frac{2\omega_{q\lambda}}{(q^0)^2 + \omega_{q\lambda}^2},$$

最终结果完全一致。注意，无论是实场还是复场， q^0 都可以取负值。

费米球附近的过程

现在考虑(8.10)的低能有效理论，其中我们需要考虑的过程全部发生在费米面附近，入射电子和出射电子能量都接近于 ϵ_F 。费米面是球形的，通过几何关系可以看出 $\mathbf{k} + \mathbf{k}' = 0$ 的过程是最重要的。如果 \mathbf{k} 和 \mathbf{k}' 是两个任意的动量，那么动量守恒加上电子在费米面附近的条件意味着可能的出射态是非常小的，但是如果 $\mathbf{k} + \mathbf{k}' = 0$ ，那么可能的出射态遍布整个费米面。因此可以判断， $\mathbf{k} + \mathbf{k}' = 0$ 的相互作用通道是最重要的相互作用通道。这样(8.10)就简化为

$$H_{4e} = \frac{1}{2N} \sum_{\mathbf{k}, \mathbf{q}'} \sum_{\alpha, \beta} \underbrace{|M_{\mathbf{q}}|^2 \frac{\omega_{\mathbf{q}}}{(\epsilon_{\mathbf{k}} - \epsilon_{\mathbf{k}+\mathbf{q}})^2 - \omega_{\mathbf{q}}^2}}_{V(\mathbf{q}, \epsilon_{\mathbf{k}} - \epsilon_{\mathbf{k}+\mathbf{q}})} c_{(\mathbf{k}+\mathbf{q})\alpha}^\dagger c_{(-\mathbf{k}-\mathbf{q})\beta}^\dagger c_{-\mathbf{k}\beta} c_{\mathbf{k}\alpha}. \quad (8.11)$$

由于电子能量差非常小，显然

$$V(\mathbf{q}, \epsilon_{\mathbf{k}} - \epsilon_{\mathbf{k}+\mathbf{q}}) = |M_{\mathbf{q}}|^2 \frac{\omega_{\mathbf{q}}}{(\epsilon_{\mathbf{k}} - \epsilon_{\mathbf{k}+\mathbf{q}})^2 - \omega_{\mathbf{q}}^2} < 0,$$

因此这个声子中介的相互作用是吸引相互作用。

库伦相互作用是排斥的，电-声子相互作用是吸引的；绝对强度显然是前者强。然而，实际上库伦相互作用很容易被屏蔽，因为把低能电子看成外加电荷，那么高能电子就会来屏蔽它（见第6.2.1节），因此把高能电子积掉之后得到的屏蔽库仑相互作用并不强，因此在重整化下，库仑相互作用实际上只是对电子能带的修正（这和费米液体的想法很相似：渐染地加入一个不大的相互作用仅仅会导致电子自能修正而已）。另一方面，电-声子相互作用是不容易屏蔽的——(8.10)有一个非常物理解释：一个电子因为电子-声子散射，在晶格中运行时产生了一系列格波，电子很快被散射到别处，但是另一个电子受到先前的那列格波作用，这就建立了长程的电子-电子吸引；实际上推迟相互作用就是不容易屏蔽的。因此，(8.11)在重整化之后是主要的电子间相互作用。还有一种看待这个问题的角度是，电-声子相互作用会导致有效的吸引，在重整化下一个吸引相互作用通常会导致粒子配对取代单粒子自由度成为主要的自由度，因此吸引相互作用不可能只是对能带的修正，而必然导致严重的费米面不稳定；库仑相互作用在这里是排斥的，不会造成粒子配对，因此正如Hartree-Fock近似中的那样，仅仅对电子能级产生了一个修正。

声子的频率只出现在0和德拜频率 ω_D 之间，换言之吸引相互作用只出现在

$$\omega = |\epsilon_{\mathbf{k}} - \epsilon_{\mathbf{k}+\mathbf{q}}| < \omega_D \quad (8.12)$$

时。数值计算可以表明对超导现象而言最重要的是建立起电子之间的吸引相互作用，其具体形式并不重要，这是因为会参与超导的实际上只有费米面附近的非常小的一个能量范围

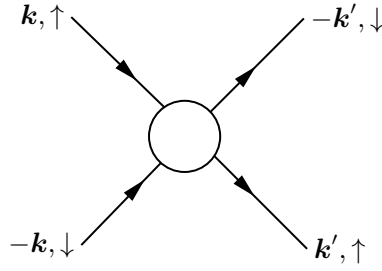


图 8.1: 球形费米面附近最为重要的相互作用通道

内的电子，因此 V 的具体形式根本不重要，实际发挥作用的只有费米面上的 V 值。对本节讨论的球形费米面，在要求(8.12)实际上就等价于这样我们设

$$V(\mathbf{q}, \omega) = \begin{cases} -V_0, & \omega < \omega_D, \\ 0, & \omega > \omega_D \end{cases}. \quad (8.13)$$

其中 ω_D 是一个硬截断。这是一个非常粗糙的截断，其合理性在于，由于我们考虑的过程全部发生在费米面附近，在电子动能与费米能量之差普遍不高时，能够激发出来的声子模式能量显然也很低，因此实际上只有声学声子能够被激发，德拜模型适用，这就是为什么我们认为 $\omega > \omega_D$ 时无有效电子-电子吸引相互作用，因为能够介导这种吸引相互作用的声子根本就不能被激发出来；在 $\omega < \omega_D$ 时，还是由于电子能量很低，可以用。这样系统的行为可以通过一个大幅简化的哈密顿量

$$H = \sum_{\mathbf{k}, \alpha} (\epsilon_{\mathbf{k}} - \mu) c_{\mathbf{k}\alpha}^\dagger c_{\mathbf{k}\alpha} - \frac{V_0}{2} \sum_{\mathbf{k}, \mathbf{q}} \sum_{\alpha, \beta} c_{(\mathbf{k}+\mathbf{q})\alpha}^\dagger c_{(-\mathbf{k}-\mathbf{q})\beta}^\dagger c_{-\mathbf{k}\beta} c_{\mathbf{k}\alpha} \quad (8.14)$$

加以描述。 \mathbf{k} 和 $\mathbf{k} + \mathbf{q}$ 都在费米面附近，并且能量相差必须足够小，小于 ω_D 。为了便于做计算，我们仅仅考虑那些这样的 \mathbf{k} 和 $\mathbf{k} + \mathbf{q}$ ：它们都在费米面附近的这样一个区域中，使得从这个区域中任取两个动量 \mathbf{k} 和 \mathbf{k}' ，它们对应的能量之差的绝对值都小于 ω_D 。由于费米面是球形的，这样一个区域显然由不等式 $\xi_{\min} < \xi_{\mathbf{k}} < \xi_{\max}$ 确定。我们是要计算满足对任意 \mathbf{k} 和 \mathbf{k}' ，均有

$$|\xi_{\mathbf{k}} - \xi_{\mathbf{k}'}| < \omega_D$$

的最大 ξ_{\max} 和最小 ξ_{\min} 。简单的计算会表明我们应当取

$$\xi_{\max} = -\xi_{\min} = \frac{\omega_D}{2},$$

即只考虑满足

$$-\frac{\omega_D}{2} < \xi_{\mathbf{k}} < \frac{\omega_D}{2} \quad (8.15)$$

的电子模式。因此以下我们认为(8.14)仅仅适用于费米面附近宽度为 ω_D 的那些电子模式。

马上可以发现(8.14)中 α 和 β 其实需要是不同的，如果它们相同，由于 \mathbf{k} 和 $-\mathbf{k}$ 都参与求和，根据费米子的性质，

$$c_{-\mathbf{k}\uparrow} c_{\mathbf{k}\uparrow} + c_{\mathbf{k}\uparrow} c_{-\mathbf{k}\uparrow} = 0,$$

因此 $\alpha = \beta$ 的相互作用通道实际上是不存在的。实际上，我们做了“电子能量非常低”的假设之后，这就是必然的，因为慢散射极限下只有自旋不同的电子才能够散射。（很显然，如果两个电子可以离得很近，那么它们的自旋就不能一样）因此(8.14)实际上就是

$$H = \sum_{\mathbf{k}} \xi_{\mathbf{k}} (c_{\mathbf{k}\uparrow}^\dagger c_{\mathbf{k}\uparrow} + c_{\mathbf{k}\downarrow}^\dagger c_{\mathbf{k}\downarrow}) - V_0 \sum_{\mathbf{k}, \mathbf{k}'} c_{-\mathbf{k}'\downarrow}^\dagger c_{\mathbf{k}'\uparrow}^\dagger c_{\mathbf{k}\uparrow} c_{-\mathbf{k}\downarrow}, \quad (8.16)$$

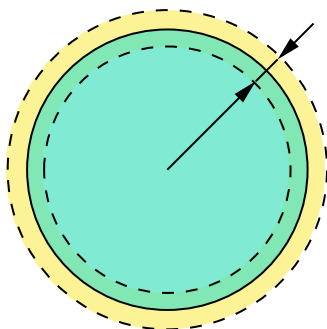


图 8.2: 电子模式在动量空间中的分布；只有两个虚线之间的那部分电子模式被考虑

这里没有 $1/2$ 因子，这是直接计算的结果；换一个角度，如果将 c_{\uparrow} 和 c_{\downarrow} 看成两种类型的场，则的确没有任何必要加入对称性因子。通常我们设

$$V_0 = \frac{g}{V}, \quad (8.17)$$

其中 g 是一个和系统体积无关的因子。

(8.14)中有电子-电子吸引相互作用，这在第五章中从来没有出现过，那里我们处理的都是排斥相互作用，从而不必担心电子凝聚在一起，但现在有了吸引相互作用，则可能出现二电子吸在一起的束缚态。在(8.14)下，发生相互作用前两个电子必须在费米面的两端，相互作用之后两个电子还是位于费米面的两端。这提示我们，费米面两端的两个电子可以发生配对。这就是下一节将要讨论的内容：库伯对。

在(8.16)之上还可以把被我们忽略的那些 $\mathbf{k} + \mathbf{k}' \neq 0$ 的相互作用通道加回来。本节仅讨论所谓的s波超导，所谓s波超导指的是假定有效相互作用形式就是

$$H = \sum_{\mathbf{k}} \xi_{\mathbf{k}} (c_{\mathbf{k}\uparrow}^{\dagger} c_{\mathbf{k}\uparrow} + c_{\mathbf{k}\downarrow}^{\dagger} c_{\mathbf{k}\downarrow}) - V_0 \sum_{\mathbf{k}, \mathbf{k}', \mathbf{q}} c_{-\mathbf{k}'\downarrow}^{\dagger} c_{\mathbf{k}'+\mathbf{q},\uparrow}^{\dagger} c_{\mathbf{k}+\mathbf{q},\uparrow} c_{-\mathbf{k}\downarrow}, \quad (8.18)$$

即忽略等效电子-电子耦合系数对声子动量 \mathbf{q} 的依赖。s波超导形容的是库伯对的对称性，具体它指的是什么将在第8.1.2节中讨论，而(8.18)为什么能够给出s波超导可以从它做Hubbard-Stratonovich变换之后得到的关于库伯对的有效理论的二次型部分(8.69)中看出来。我们这里讨论的都是单带模型，不过如果强行用 $c_{\mathbf{k}}$ 做傅里叶逆变换得到实际上只包含了一条带上的电子的“实空间场算符” $\psi(\mathbf{r})$ ，(8.18)对应着

$$H_{\text{bcs}} = \frac{g}{V} \int d^3\mathbf{r} \psi_{\uparrow}^{\dagger} \psi_{\downarrow}^{\dagger} \psi_{\downarrow} \psi_{\uparrow}. \quad (8.19)$$

这个相互作用的瞬时性来自声子本身只能持续很短时间这一事实，而它的局域性（不像库伦相互作用那样长程）则来自声子速度其实很慢这个事实。很多教科书从上式出发引入BCS理论，而并不详细讨论其来源。

8.1.2 库伯对

$SU(2)$ 的表示

我们从对称性的角度出发，讨论库伯对的一般形式。库伯对写成算符形式就是 $c^{\dagger}c^{\dagger}$ 的形式，或者等价的 cc 形式。如果其期望值不为零，那么 $U(1)$ 对称性就破缺了，即电荷守恒对称性被破缺了（物理图像是，一部分电荷被封存到了库伯对中，不再以独立电子的形式流动）。电荷守恒的破缺实际上是低温超导中最重要的物理，因为它意味着系统中的费米型

准粒子不存在或者相较于费米液体变少了，多出来的电荷封印在库伯对中，而库伯对是可以无电阻移动的，这就导致超导现象。

库伯对序参量的一般形式是 $\langle c_{\mathbf{k}\alpha} c_{\mathbf{k}'\beta} \rangle$ ，或者写成实空间的形式，是 $\langle \psi_\alpha(\mathbf{r}) \psi_\beta(\mathbf{r}') \rangle$ 。我们要根据没有被破缺的对称性来分析 α 和 β 之间的关系，以及 \mathbf{k} 和 \mathbf{k}' 之间的关系。可以发现(8.14)具有自旋旋转不变性（这个对称性没有被破缺掉），因此相应的序参量 $\langle c_{\mathbf{k}\alpha} c_{\mathbf{k}'\beta} \rangle$ 应该是一个二分量的自旋协变的对象。由于(8.18)中忽略了自旋-轨道耦合， $\langle c_{\mathbf{k}\alpha} c_{\mathbf{k}'\beta} \rangle$ 正比于一个关于 \mathbf{k} 和 \mathbf{k}' 的函数乘以一个关于 α 和 β 的函数，即

$$\langle c_{\mathbf{k}\alpha} c_{\mathbf{k}'\beta} \rangle \propto \Delta_{\text{orbital}}(\mathbf{k}, \mathbf{k}') \Delta_{\text{spin}}(\alpha, \beta).$$

二粒子配对对应的 $SU(2)$ 的表示为

$$\frac{1}{2} \otimes \frac{1}{2} = 0 \oplus 1,$$

即可能有自旋单态也可能有自旋三重态。如果该库伯对是自旋单态，由于自旋单态只有一个基矢量

$$|0, 0\rangle = \frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle),$$

将它在 $\frac{1}{2} \otimes \frac{1}{2}$ 基底写下写成矩阵形式就是 $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ ，于是应该有

$$\langle c_{\mathbf{k}\alpha} c_{\mathbf{k}'\beta} \rangle \propto \epsilon_{\alpha\beta} \propto \delta_{\alpha+\beta, 0},$$

即

$$\langle c_{\mathbf{k}\alpha} c_{\mathbf{k}'\beta} \rangle = \Delta(\mathbf{k}, \mathbf{k}') \epsilon_{\alpha\beta}, \quad (8.20)$$

其中 ϵ 为所谓的旋量度规。而如果该库珀对为自旋三重态，由于三重态的基底包括

$$|1, -1\rangle = |\downarrow\downarrow\rangle, \quad |1, 0\rangle = \frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle), \quad |1, 1\rangle = |\uparrow\uparrow\rangle,$$

我们有

$$[\langle c_{\mathbf{k}\alpha} c_{\mathbf{k}'\beta} \rangle]_{\alpha\beta} \propto c_1 \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + c_2 \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} + c_3 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$

或者也可以写成

$$[\langle c_{\mathbf{k}\alpha} c_{\mathbf{k}'\beta} \rangle]_{\alpha\beta} \propto g_1 \sigma^x \sigma^y + g_2 \sigma^y \sigma^y + g_3 \sigma^z \sigma^x,$$

于是能够找到一个 $\mathbf{d}(\mathbf{k}, \mathbf{k}')$ 使得

$$\langle c_{\mathbf{k}\alpha} c_{\mathbf{k}'\beta} \rangle = \langle (\mathbf{d}(\mathbf{k}, \mathbf{k}') \cdot \boldsymbol{\sigma}_{\alpha\beta}) \sigma^y \rangle, \quad (8.21)$$

或者也可以直接写

$$\langle c_{\mathbf{k}\alpha} c_{\mathbf{k}'\beta} \rangle = \mathbf{d}(\mathbf{k}, \mathbf{k}') \cdot \boldsymbol{\sigma}_{\alpha\beta}. \quad (8.22)$$

背景知识 8.1: 二体轨道波函数

设有二体轨道波函数 $\psi(\mathbf{r}_1, \mathbf{r}_2)$ ，其中 \mathbf{r}_1 和 \mathbf{r}_2 地位相同。设

$$\mathbf{R} = \frac{1}{2}(\mathbf{r}_1 + \mathbf{r}_2), \quad \mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2. \quad (8.23)$$

做傅里叶变换，有

$$\begin{aligned}\psi(\mathbf{r}_1, \mathbf{r}_2) &= \frac{1}{V} \sum_{\mathbf{k}_1} \sum_{\mathbf{k}_2} \psi(\mathbf{k}_1, \mathbf{k}_2) e^{i\mathbf{k}_1 \cdot \mathbf{r}_1} e^{i\mathbf{k}_2 \cdot \mathbf{r}_2} \\ &= \frac{1}{V} \sum_{\mathbf{k}_1, \mathbf{k}_2} e^{i(\mathbf{k}_1 + \mathbf{k}_2) \cdot \mathbf{R}} e^{i(\mathbf{k}_1 - \mathbf{k}_2) \cdot \mathbf{r}/2} \psi(\mathbf{k}_1, \mathbf{k}_2).\end{aligned}$$

这样就有

$$\psi(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{V} \sum_{\mathbf{q}} e^{i\mathbf{q} \cdot \mathbf{R}} \sum_{\mathbf{k}'} e^{i(2\mathbf{k}' - \mathbf{q}) \cdot \mathbf{r}/2} \psi(\mathbf{k}', \mathbf{q} - \mathbf{k}'),$$

或者做一个变量代换得到

$$\psi(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{V} \sum_{\mathbf{q}} e^{i\mathbf{q} \cdot \mathbf{R}} \sum_{\mathbf{k}} e^{i\mathbf{k} \cdot \mathbf{r}} \psi(\mathbf{k} + \mathbf{q}/2, \mathbf{q}/2 - \mathbf{k}). \quad (8.24)$$

在连续空间平移不变性严格成立时，只有 $\mathbf{q} = 0$ 分量，此时

$$\psi(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{V} \sum_{\mathbf{k}} e^{i\mathbf{k} \cdot (\mathbf{r}_1 - \mathbf{r}_2)} \psi(\mathbf{k}, -\mathbf{k}). \quad (8.25)$$

可以看到交换 \mathbf{r}_1 和 \mathbf{r}_2 意味着将 \mathbf{k} 取负号，或者说设

$$\psi(\mathbf{r}_1, \mathbf{r}_2) = \psi'(\mathbf{r}_2, \mathbf{r}_1),$$

则

$$\psi(\mathbf{k} + \mathbf{q}/2, \mathbf{q}/2 - \mathbf{k}) = \psi'(-\mathbf{k} + \mathbf{q}/2, \mathbf{q}/2 + \mathbf{k}).$$

直观地看， \mathbf{q} 是“质心动量”，而 \mathbf{k} 是其中一个粒子相对质心的动量。

我们改用总动量 \mathbf{q} 和其中一个电子相对库伯对质心的动量 \mathbf{k} 标记库伯对，记作 $\Delta(\mathbf{k}, \mathbf{q})$ 。由于库伯对可以看成是一个二电子波函数，它必须是反对称的（将库伯对看成是一个整体，它们构成的含有多个库伯对的多体波函数是交换对称的，但是库伯对相对于组成它的电子则是交换反对称的）。如果该库伯对是自旋单态的，那么其自旋部分是反对称的，则其轨道部分就是对称的，也即

$$\Delta(\mathbf{k}, \mathbf{q}) = \Delta(-\mathbf{k}, \mathbf{q}),$$

从而固定 \mathbf{q} 不变，可以将 $\Delta(\mathbf{k}, \mathbf{q})$ 展开成一系列s波、d波等对称球谐函数的线性组合；而如果该库伯对是自旋三重态的，那么其自旋部分就是对称的，于是轨道部分满足

$$\Delta(\mathbf{k}, \mathbf{q}) = -\Delta(-\mathbf{k}, \mathbf{q}),$$

此时固定 \mathbf{q} 不变， $\Delta(\mathbf{k}, \mathbf{q})$ 是一系列反对称球谐函数的线性组合。容易看出单态或者三重态的出现和费米面对称性的关系，比如如果费米面不对称，那么s波配对就不能产生。

库伯对作为载流子

通常的金属中载流子都是电子和空穴。两个电子组成的玻色型准粒子库伯对携带 $-2e$ 的电荷，是一种玻色型载流子。

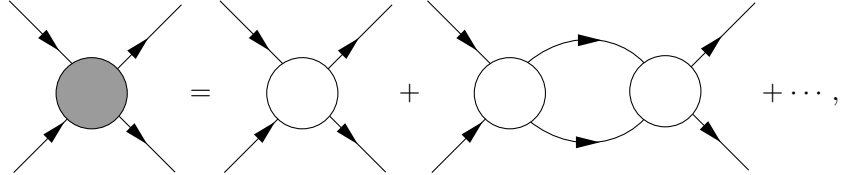
电子在金属中运动时会受到晶格散射，从而有电阻。

用库伯对关联函数探测s波库伯对的稳定性

为了判断库伯对的稳定性，我们计算库伯对关联函数

$$C^{\text{cooper}}(\mathbf{q}, i\omega_n) = \sum_{k, k'} \langle \bar{c}_{-k'\downarrow} \bar{c}_{k'+q, \uparrow} c_{k+q, \uparrow} c_{-k, \downarrow} \rangle. \quad (8.26)$$

这里的 β 和 V 因子是为了让这个关联函数看起来像相互作用哈密顿量一点。与前述寻常金属中的费曼图计算类似，为了避免发散，不能只计算有限个费曼图而必须做一定的无穷求和。做梯形图近似



$$= \text{white circle} + \text{two circles with lines} + \dots, \quad (8.27)$$

能够计算得到顶角函数

$$\Gamma_q = -\frac{g}{1 - \frac{g}{\beta V} \sum_p G_{p+q}^0 G_{-p}^0}, \quad (8.28)$$

从而

$$C^{\text{cooper}}(\mathbf{q}, i\omega_n) = \sum_{k, k'} \langle \bar{c}_{-k'\downarrow} \bar{c}_{k'+q, \uparrow} c_{k+q, \uparrow} c_{-k, \downarrow} \rangle_0 + \frac{g}{1 - \frac{g}{\beta V} \sum_p G_{p+q}^0 G_{-p}^0}. \quad (8.29)$$

上式的第一项可以通过Wick定理，按照第6.2.2节的方法可以计算出来。第二项计算如下。我们有

$$-\frac{1}{\beta V} \sum_p G_{p+q}^0 G_{-p}^0 = -\frac{1}{\beta V} \sum_{p^0, \mathbf{p}} \frac{1}{i(p^0 + q^0) - \xi_{\mathbf{p}+q}} \frac{1}{-ip^0 - \xi_{\mathbf{p}}},$$

这里用到了 $\xi_{\mathbf{p}}$ 和 $\xi_{-\mathbf{p}}$ 一样这一事实。同样使用第6.2.2节中使用过的方法，根据

$$\frac{1}{\beta} \sum_{i\omega_n} F(i\omega_n) = -\frac{1}{2\pi i} \oint dz F(z) f(z),$$

我们有

$$\begin{aligned} -\frac{1}{\beta V} \sum_{p^0, \mathbf{p}} \frac{1}{i(p^0 + q^0) - \xi_{\mathbf{p}+q}} \frac{1}{-ip^0 - \xi_{\mathbf{p}}} &= -\sum_{\mathbf{p}} \frac{1}{V} \frac{1}{2\pi i} \oint dz f(z) \frac{1}{z + iq^0 - \xi_{\mathbf{p}+q}} \frac{1}{z + \xi_{\mathbf{p}}} \\ &= \frac{1}{V} \sum_{\mathbf{p}} \left(\frac{f(-\xi_{\mathbf{p}})}{-\xi_{\mathbf{p}} + iq^0 - \xi_{\mathbf{p}+q}} + \frac{f(\xi_{\mathbf{p}+q} - iq^0)}{\xi_{\mathbf{p}+q} - iq^0 + \xi_{\mathbf{p}}} \right) \\ &= \frac{1}{V} \sum_{\mathbf{p}} \frac{1 - f(\xi_{\mathbf{p}}) - f(\xi_{\mathbf{p}+q})}{iq^0 - \xi_{\mathbf{p}} - \xi_{\mathbf{p}+q}}, \end{aligned}$$

即

$$G^{\text{cooper}}(q) - C_0^{\text{cooper}}(q) = \frac{1}{1 - \frac{g}{V} \sum_{\mathbf{p}} \frac{f(\xi_{\mathbf{p}}) + f(\xi_{\mathbf{p}+q}) - 1}{iq^0 - \xi_{\mathbf{p}} - \xi_{\mathbf{p}+q}}}. \quad (8.30)$$

上式右边的奇异性实际上有温度依赖。考虑 $q = 0$ 的情况，此时

$$\frac{g}{V} \sum_{\mathbf{p}} \frac{f(\xi_{\mathbf{p}}) + f(\xi_{\mathbf{p}+q}) - 1}{iq^0 - \xi_{\mathbf{p}} - \xi_{\mathbf{p}+q}} = g \int d\epsilon D(\epsilon) \frac{1 - 2f(\epsilon)}{2\epsilon}.$$

这里有几个截断。首先根据(8.15), ϵ 的上下限为 $\pm\omega_D/2$, 且 $1 - 2f(\epsilon)$ 在 $\epsilon > 0$ 和 $\epsilon < 0$ 时差一个正负号, 因此

$$\begin{aligned} g \int d\epsilon D(\epsilon) \frac{1 - 2f(\epsilon)}{2\epsilon} &= g \int_{-\omega_D/2}^{\omega_D/2} d\epsilon D(\epsilon) \frac{1 - 2f(\epsilon)}{2\epsilon} \\ &= g \int_0^{\omega_D/2} d\epsilon D(\epsilon) \frac{1 - 2f(\epsilon)}{\epsilon}. \end{aligned}$$

其次, $f(\epsilon)$ 只有在 ϵ 和0不那么接近时才有明显的非零值, 在 $\epsilon \rightarrow 0^+$ 时 $f(\epsilon)$ 快速衰减, 因此我们需要做一个大小大体上正比于 T 的红外截断, 即

$$g \int_0^{\omega_D/2} d\epsilon D(\epsilon) \frac{1 - 2f(\epsilon)}{\epsilon} \sim gD(0) \int_T^{\omega_D/2} d\epsilon \frac{1}{\epsilon} \sim gD(0) \ln \frac{\omega_D}{T},$$

从而(8.30)右边的分母在

$$1 \sim gD(0) \ln \frac{\omega_D}{T},$$

即

$$T \rightarrow T_c \sim \omega_D \exp\left(-\frac{1}{gD(0)}\right) \quad (8.31)$$

时变为零, 关联函数发散。以上数量级估计中的近似之处在于 T 前面可以乘以一个常数, 除此以外没有什么近似, 因此

$$T_c = \text{const} \times \omega_D \exp\left(-\frac{1}{gD(0)}\right) \quad (8.32)$$

前面的常数只要正确, 就是足够精确的。关联函数的发散大体上是

$$\frac{1}{1 - gD(0) \ln \frac{\omega_D}{T}} \sim -\frac{T_c}{gD(0)(T - T_c)}.$$

以上推导暗示两件事: 首先, 在 $T < T_c$ 时仅仅做梯形图近似是不够的, 即, 如果以费米海为基态, 那么需要纳入考虑的量子涨落非常大, 这暗示, 此时我们其实不应该以费米海为基态, 即费米面出现了不稳定性, 一个对称性自发破缺发生了; 且做梯形图近似会出现发散这件事实际上还暗示了对称性自发破缺的序参量, 这点我们在(8.71)前后的讨论中会提到。其次, $q = 0$ 时 $T \rightarrow T_c$ 时关联函数出现极点, 这是相变点附近的典型行为, 即 $T = T_c$ 处可能存在相变。

在远离 $T = T_c, q = 0$ 点处, 关联函数(8.29)显然有有限大虚部以上几个暗示放在一起, 似乎说明, 在低温下系统会发生对称性自发破缺, 库伯对是稳定的, 而且系统基态可能就是由大量库伯对构成的。

8.1.3 s波BCS超导的平均场研究

平均场序参量哈密顿量

背景知识 8.2: 用平均场近似研究对称性自发破缺

在背景知识6.2中我们引入了平均场方法, 很容易看出, 平均场方法可以用于处理对称性自发破缺问题。我们只需要写下一个在被破缺的那部分对称性存在时一定为零的格林函数, 然后用平均场近似计算它就可以, 如果在一些条件下它不为零, 那么就可能发生了对称性自发破缺。这个格林函数出于显然的原因, 被称为“平均场序参量”。需要注意, 在以费米海为基态的费曼规则中看起来不太对头的那些格林函

数（如 $\langle cc \rangle$ ，画成图是两个相对的箭头，根据 $U(1)$ 对称性为零）也可以做平均场序参量，因为无法保证对称性自发破缺后这些格林函数真的还是零。

单纯地使用有限阶微扰论不足以研究对称性自发破缺，因为对称性自发破缺会让基态发生变化，在假的基态附近做微扰是无法给出可靠的结果的。平均场方法则将无数个费曼图重求和了，实际上是一种非微扰方法，因此可以捕捉到对称性自发破缺。然而平均场方法认为存在对称性自发破缺的情况下未必真的有对称性自发破缺，因为平均场方法忽略了许多其它的图，倾向于低估量子涨落。即使平均场方法预言的对称性自发破缺是真实的，它给出的相变点位置通常也不可靠，因此只能作为定性方法使用。另一个平均场方法不可靠的地方是，平均场方法假定序参量 $\langle cc \rangle$ 是没有任何涨落的，系统中仅有的涨落是费米场，而这当然不是事实：序参量本身是可以涨落的，而且这些涨落是系统的低能激发的重要成分。

最后应当指出，其实平均场近似不仅仅适用于费米子系统，它对各种自由度的模型都是适用的。

平均场近似忽略了库伯对的涨落，即我们假定库伯对总是在它们的基态上。此时由于空间平移对称性没有破缺，背景知识8.1中的 $\mathbf{q} = 0$ ，即我们可以认为

$$\langle c_{\mathbf{k}\alpha} c_{\mathbf{k}'\beta} \rangle \propto \delta(\mathbf{k} + \mathbf{k}').$$

换言之，我们只考虑总动量为零的配对 $\langle c_{\mathbf{k}\alpha} c_{-\mathbf{k}\beta} \rangle$ 。这个做法的合理性也可以通过注意到最重要的相互作用通道是(8.16)来得到辩护——实际上这样做更加合理，因为在存在电子凝聚时一般很难判断空间平移对称性是不是真的没被破缺，例如我们在第8.2节中就能够看到一个空间平移对称性破缺的例子。

对(8.14)做平均场近似，有

$$\begin{aligned} & c_{(\mathbf{k}+\mathbf{q})\alpha}^\dagger c_{(-\mathbf{k}-\mathbf{q})\beta}^\dagger c_{-\mathbf{k}\beta} c_{\mathbf{k}\alpha} \\ & \approx \langle c | c_{(\mathbf{k}+\mathbf{q})\alpha}^\dagger c_{(-\mathbf{k}-\mathbf{q})\beta}^\dagger | c \rangle c_{-\mathbf{k}\beta} c_{\mathbf{k}\alpha} + c_{(\mathbf{k}+\mathbf{q})\alpha}^\dagger c_{(-\mathbf{k}-\mathbf{q})\beta}^\dagger \langle c_{-\mathbf{k}\beta} c_{\mathbf{k}\alpha} \rangle - \langle c_{(\mathbf{k}+\mathbf{q})\alpha}^\dagger c_{(-\mathbf{k}-\mathbf{q})\beta}^\dagger c_{-\mathbf{k}\beta} c_{\mathbf{k}\alpha} \rangle, \end{aligned}$$

重新选定求和哑指标，并略去对体系动力学没有影响的常数项，得到

$$H_{\text{MF}} = \sum_{\mathbf{k}, \alpha} (\epsilon_{\mathbf{k}} - \mu) c_{\mathbf{k}\alpha}^\dagger c_{\mathbf{k}\alpha} - \frac{V_0}{2} \sum_{\mathbf{k}, \mathbf{k}', \alpha, \beta} \left(\langle c_{-\mathbf{k}\beta} c_{\mathbf{k}\alpha} | c \rangle c_{\mathbf{k}'\alpha}^\dagger c_{-\mathbf{k}'\beta}^\dagger + \text{h.c.} \right). \quad (8.33)$$

的确，电荷守恒对称性被破缺了，原因是电子形成库伯对之后看起来就像被“冻结”了一样，因此不再被计入 c^\dagger 自由度中，而是被计入序参量 $\langle c_{-\mathbf{k}\beta} c_{\mathbf{k}\alpha} \rangle$ 中。

现在我们考虑单态、s波的库伯对，这样发生配对的就是一个 \mathbf{k}, \uparrow 态的电子和一个 $-\mathbf{k}, \downarrow$ 态的电子，或者做一个自旋旋转。总之，序参量可以选取为

$$\Delta = -\frac{V_0}{2} \sum_{\mathbf{k}} (\langle c_{-\mathbf{k}\uparrow} c_{\mathbf{k}\downarrow} \rangle - \langle c_{-\mathbf{k}\downarrow} c_{\mathbf{k}\uparrow} \rangle) = -V_0 \sum_{\mathbf{k}} \langle c_{-\mathbf{k}\uparrow} c_{\mathbf{k}\downarrow} \rangle, \quad (8.34)$$

则(8.33)中只有尖括号内满足 $\alpha = -\beta$ 的项才能够保留下来，计算得到

$$H_{\text{MF}} = \sum_{\mathbf{k}, \alpha} \xi_{\mathbf{k}} c_{\mathbf{k}\alpha}^\dagger c_{\mathbf{k}\alpha} + \Delta \sum_{\mathbf{k}} c_{-\mathbf{k}\downarrow}^\dagger c_{\mathbf{k}\uparrow}^\dagger + \Delta^* \sum_{\mathbf{k}} c_{\mathbf{k}\uparrow} c_{-\mathbf{k}\downarrow}. \quad (8.35)$$

平均场近似下的系统的费米型激发

下面我们用一个么正变换重新定义一组准粒子（这就称为Bogoliubov变换），使这组准粒子本身是费米子，并且能够让(8.35)对角化（从而这组准粒子的能谱就是(8.39)）。设这

组Bogoliubov准粒子为 $\gamma_{\mathbf{k}\alpha}^\dagger$ ，我们要求费米子的对易关系

$$\{\gamma_{\mathbf{k}_1\alpha}, \gamma_{\mathbf{k}_2\beta}^\dagger\} = \delta_{\mathbf{k}_1\mathbf{k}_2} \delta_{\alpha\beta}, \quad \{\gamma_{\mathbf{k}_1\alpha}, \gamma_{\mathbf{k}_2\beta}\} = 0 \quad (8.36)$$

成立。可以看到，以下正交变换

$$\begin{pmatrix} \gamma_{\mathbf{k}\uparrow} \\ \gamma_{-\mathbf{k}\downarrow}^\dagger \end{pmatrix} = \begin{pmatrix} u_{\mathbf{k}} & -v_{\mathbf{k}} \\ v_{\mathbf{k}} & u_{\mathbf{k}} \end{pmatrix} \begin{pmatrix} c_{\mathbf{k}\uparrow} \\ c_{-\mathbf{k}\downarrow}^\dagger \end{pmatrix}, \quad u_{\mathbf{k}}^2 + v_{\mathbf{k}}^2 = 1 \quad (8.37)$$

能够给出正确的对易关系。于是我们将(8.35)改写成

$$H_{\text{MF}} = \sum_{\mathbf{k}} \begin{pmatrix} c_{\mathbf{k}\uparrow}^\dagger & c_{-\mathbf{k}\downarrow} \end{pmatrix} \begin{pmatrix} \xi_{\mathbf{k}} & -\Delta \\ -\Delta^* & -\xi_{\mathbf{k}} \end{pmatrix} \begin{pmatrix} c_{\mathbf{k}\uparrow} \\ c_{-\mathbf{k}\downarrow}^\dagger \end{pmatrix}. \quad (8.38)$$

将其中的矩阵分解成 $U^\dagger \text{diag}(E_1, E_2)U$ 的形式，其中 U 是正交矩阵，就能够得到Bogoliubov准粒子。³

然而(8.38)中的矩阵做特征分解得到的 U 是否是正交矩阵是存疑的。我们能够确定 U 是么正的，但是是否有 $u_{\mathbf{k}}^2 + v_{\mathbf{k}}^2 = 1$ 成立呢？我们做一个大胆的假设，假定 Δ 为实数。实际情况如何，需要通过系统地分析序参量的场论和对称性自发破缺之后的基态和其上的涨落才能够得到，我们将在第8.1.4节中验证这一点。这里我们姑且假定 Δ 为实数，计算所有能够计算的东西，如果一切都符合得很好，那么这个假设是有道理的。 Δ 为实数时(8.38)中的矩阵是实对称矩阵，一定能够做特征分解，且特征向量是实的，彼此正交。此时计算本征值得到

$$E_{\mathbf{k}} = \pm \sqrt{\xi_{\mathbf{k}}^2 + |\Delta|^2}. \quad (8.39)$$

取 $E_{\mathbf{k}}$ 为正值，特征分解得到

$$\begin{pmatrix} \xi_{\mathbf{k}} & -\Delta \\ -\Delta & -\xi_{\mathbf{k}} \end{pmatrix} = \begin{pmatrix} u_{\mathbf{k}} & v_{\mathbf{k}} \\ -v_{\mathbf{k}} & u_{\mathbf{k}} \end{pmatrix} \begin{pmatrix} E_{\mathbf{k}} & \\ & -E_{\mathbf{k}} \end{pmatrix} \begin{pmatrix} u_{\mathbf{k}} & -v_{\mathbf{k}} \\ v_{\mathbf{k}} & u_{\mathbf{k}} \end{pmatrix},$$

其中

$$u_{\mathbf{k}} = \sqrt{\frac{E_{\mathbf{k}} + \xi_{\mathbf{k}}}{2E_{\mathbf{k}}}}, \quad v_{\mathbf{k}} = \sqrt{\frac{E_{\mathbf{k}} - \xi_{\mathbf{k}}}{2E_{\mathbf{k}}}}. \quad (8.40)$$

这就得到了Bogoliubov变换的显式形式：

$$\begin{cases} \gamma_{\mathbf{k}\uparrow} = u_{\mathbf{k}} c_{\mathbf{k}\uparrow} - v_{\mathbf{k}} c_{-\mathbf{k}\downarrow}^\dagger, \\ \gamma_{\mathbf{k}\downarrow} = u_{\mathbf{k}} c_{\mathbf{k}\downarrow} + v_{\mathbf{k}} c_{-\mathbf{k}\uparrow}^\dagger, \end{cases} \quad (8.41)$$

$\gamma_{\mathbf{k}\uparrow}^\dagger$ 粒子的能量是 $E_{\mathbf{k}}$ ， $\gamma_{\mathbf{k}\downarrow}^\dagger$ 粒子的能量是 $-E_{\mathbf{k}}$ 。逆变换为

$$\begin{cases} c_{\mathbf{k}\uparrow} = u_{\mathbf{k}} \gamma_{\mathbf{k}\uparrow} + v_{\mathbf{k}} \gamma_{-\mathbf{k}\downarrow}^\dagger, \\ c_{\mathbf{k}\downarrow} = u_{\mathbf{k}} \gamma_{\mathbf{k}\downarrow} - v_{\mathbf{k}} \gamma_{-\mathbf{k}\uparrow}^\dagger. \end{cases} \quad (8.42)$$

这个结果具有粒子-空穴对称性，但是这并不具有太多物理意义，因为它实际上是对角化时交换了一对产生湮灭算符，从而把一部分粒子自由度写成了空穴而已。

没有 Δ 时只考虑轨道波函数的话有一条能带，就是 $E = \xi_{\mathbf{k}}$ ，考虑有一上一下两种自旋那就有两条能带。对金属 $\xi_{\mathbf{k}}$ 是无能隙的，即 $\xi_{\mathbf{k}}$ 存在零点。(8.39)中，序参量 Δ 的存在让我们

³从电子到Bogoliubov准粒子的变换未必是正交变换，有时候真的要用一个非正交变换；关键在于对易关系必须正确，不是所有的线性变换形式都能够给出正确的对易关系。

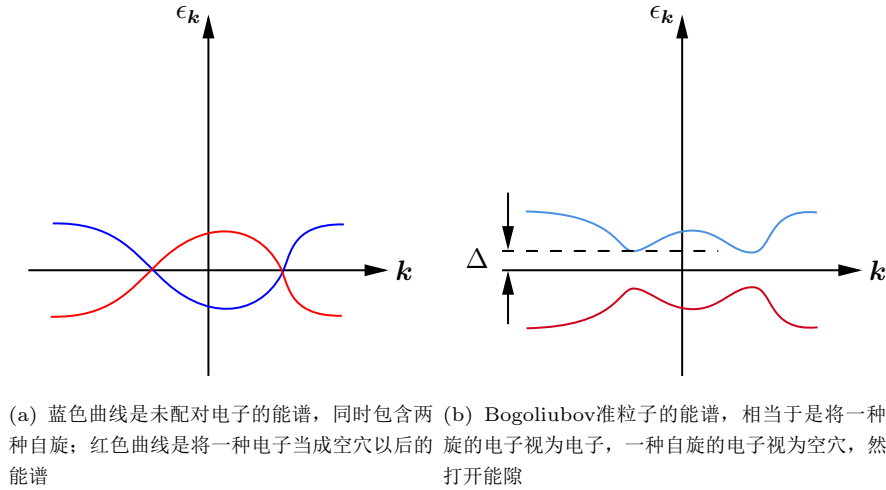


图 8.3: Bogoliubov准粒子的能谱

得到了两条彼此存在能隙的能带。由费米分布， $E = -E_{\mathbf{k}}$ 那一条带，即 $\gamma_{\mathbf{k}\downarrow}$ 那一条带完全被占据（我们已经将化学势算在 $\xi_{\mathbf{k}}$ 中了，无需重复考虑化学势）。因此基态附近的费米型激发就是 $\gamma_{\mathbf{k}\uparrow}$ 准粒子和 $\gamma_{\mathbf{k}\downarrow}$ 准空穴，它们的能谱都是 $E = E_{\mathbf{k}}$ ，并且有一个大小为 Δ 的能隙。（见图8.3）

系统的费米型激发存在能隙这件事是值得注意的。我们这里以 $\xi_{\mathbf{k}}$ 为电子能量，即已经扣除了化学势，如果系统是普通意义上的导体，那么不应该有能隙，因为化学势在某一条能带内部，费米面附近的电子可以被任意小的能量激发。但是(8.14)的费米型激发是有能隙的，这意味着在零温处根本不存在普通的能带电子，所有的电子都被“封印”在库伯对里面了，并且小的扰动只会让库伯对“海”发生小的扭曲，不会让费米型激发出现，因为费米型激发是有能隙的。此时也不存在费米面，因为有费米面则费米面附近的费米型载流子是无能隙的，而这并非这里的情况。略微考虑一下我们就会发现， Δ 正是拆开一个库伯对，从中释放出电子需要的能量，可以说就是库伯对的结合能，系统的费米型激发有能隙即 $\Delta \neq 0$ 意味着库伯对是稳定的，不会被任意小的扰动拆散。

平均场自洽计算

得到平均场理论后我们来做自洽计算。将(8.42)代入(8.34)，并利用近独立电子气的数目

$$\langle \gamma_{\mathbf{k}\alpha}^\dagger \gamma_{\mathbf{k}\alpha} \rangle = f(E_{\mathbf{k}}) = \frac{1}{e^{\beta E_{\mathbf{k}}} + 1},$$

得到

$$\Delta = -V_0 \sum_{\mathbf{k}} u_{\mathbf{k}} v_{\mathbf{k}} (2f(E_{\mathbf{k}}) - 1).$$

可以验证，

$$u_{\mathbf{k}} v_{\mathbf{k}} = \frac{\Delta}{2E_{\mathbf{k}}},$$

于是最终得到

$$V_0 \Delta \sum_{\mathbf{k}} \frac{1}{2E_{\mathbf{k}}} \frac{e^{\beta E_{\mathbf{k}}} - 1}{e^{\beta E_{\mathbf{k}}} + 1} = \Delta. \quad (8.43)$$

如果 Δ 不为零，即出现超导现象，就可以把 Δ 消掉。 $E_{\mathbf{k}}$ 依赖于 Δ ，于是给定一个温度就可以把 Δ 解出。

例如, 在 $T = 0$ 时, 我们有

$$1 = V_0 \sum_{\mathbf{k}} \frac{1}{2E_{\mathbf{k}}} = V_0 \int d\epsilon N(\epsilon) \frac{1}{2\sqrt{\epsilon^2 + \Delta^2}},$$

这里我们在对单个电子的能量做积分。当然由于电子能量过高时第8.1.1节中的机制不再适用, 积分能量肯定有一个截断。(8.13)给出了截断 ω_D 。我们假定 ω_D 相对费米能非常小, 也即发生库伯配对的电子只是费米面附近非常小的一个能量范围内的, 则近似有

$$1 = V_0 N(0) \int_{-\omega_D}^{\omega_D} d\epsilon \frac{1}{\sqrt{\epsilon^2 + \Delta^2}} = N(0) V_0 \sinh^{-1} \left(\frac{\omega_D}{\Delta} \right) \approx N(0) V_0 \ln \frac{2\omega_D}{\Delta},$$

于是

$$\Delta = 2\omega_D \exp \left(-\frac{1}{N(0)V_0} \right). \quad (8.44)$$

显然, Δ 对 V_0 的依赖比对 ω_D 的依赖要强得多。幸好如此, 因为 ω_D 实际上是一个非常粗糙的硬截断。现在我们看到, 硬截断(8.13)是合理的, 因为截断只是给 Δ 提供了一个能量尺度而已, 不会影响更为复杂的行为。我们也可以看出, 只要有相互作用, 不管多强, 都会产生一个非零的 Δ , 因此只要有相互作用就会出现超导转变, 并打开能隙。这意味着(8.10)具有非微扰行为。

(8.43)中总是有一个平庸解 $\Delta = 0$, 消掉因子 Δ 之后得到的方程就不总是有解。消掉因子 Δ 之后得到的方程有没有解就区分了超导相和非超导相。现在我们计算临界温度, 即 Δ 从非零的一侧趋于零时的温度:

$$\begin{aligned} 1 &= \lim_{\Delta \rightarrow 0} V_0 \sum_{\mathbf{k}} \frac{1}{2E_{\mathbf{k}}} \frac{e^{\beta E_{\mathbf{k}}} - 1}{e^{\beta E_{\mathbf{k}}} + 1} \\ &= V_0 \sum_{\mathbf{k}} \frac{1}{2\xi_{\mathbf{k}}} \frac{e^{\beta \xi_{\mathbf{k}}} - 1}{e^{\beta \xi_{\mathbf{k}}} + 1} \\ &= V_0 V \int_{-\omega_D}^{\omega_D} d\epsilon N(\epsilon) \frac{1}{2\epsilon} \frac{e^{\beta \epsilon} - 1}{e^{\beta \epsilon} + 1} \\ &\approx gN(0) \int_0^{\omega_D} d\epsilon \frac{1}{\epsilon} \frac{e^{\beta \epsilon} - 1}{e^{\beta \epsilon} + 1}. \end{aligned}$$

最后一个积分中的因子 $(e^{\beta \epsilon} - 1)/(e^{\beta \epsilon} + 1)$ 的作用在于在 ϵ 接近0时压低 $1/\epsilon$ 的值从而避免发散, 它是一个特征尺度为 T 的红外截断, 于是

$$1 \sim gN(0) \int_T^{\omega_D} d\epsilon \frac{1}{\epsilon},$$

最后得到

$$T_c \sim \omega_D \exp \left(-\frac{1}{gN(0)} \right). \quad (8.45)$$

更为精确的计算会给出

$$T_c = 1.14\omega_D \exp \left(-\frac{1}{gN(0)} \right), \quad (8.46)$$

不过实际上以上公式自身意义不大, 因为首先很难精确计算 g , 其次 g 的微小变化会带来很大误差。真正会在实验上验证的通常是从(8.44)和(8.46)推出的

$$\frac{2\Delta(T=0)}{T_c} = 3.5. \quad (8.47)$$

如果实验测量出来的比值远离3.5, 那就可以确定这个超导现象不来自BCS机制。

8.1.4 s波超导的金斯堡-朗道理论

对称性分析

现在我们采取另一条路，尝试直接从 $U(1)$ 对称性被破缺这件事来获得关于超导的一些解释。设 Φ 为库伯对序参量，我们知道它是一个复标量场。有时称它为超导波函数，虽然它并不是任何粒子的波函数，但它服从的方程和薛定谔方程形式一致。我们假定超导相中仅有的重要的自由度是这个序参量，其他量（比如单个电子）全部不重要。这个假设是整个理论中最需要物理直觉的部分，因为并非对全部体系都有这个结论，例如做一维电子的玻色化时就不能只用一个标量场，在分析二维正方晶格的自旋波时也需要同时考虑序参量和电子。在讨论BCS系统时能用这个假设是因为如前所述，破缺的是 $U(1)$ 对称性，而 $U(1)$ 对称性被破缺是因为电子在低温下配对为库伯对，电子的全部行为都被库伯对反映了，因此只需要考虑库伯对序参量即可。

设外加一个大小为 \mathbf{A} 的磁矢势，按照规范不变性我们写下一个理论

$$F = \int d^3\mathbf{r} \left(|(\nabla - ie\mathbf{A})\Phi|^2 + r|\Phi|^2 + u|\Phi|^4 \right). \quad (8.48)$$

这里我们已经使用协变导数代替了原有的导数；系数为 $2e$ 是因为一个库伯对带有两个负电荷，或者更加数学地说，一个库伯对包含两个湮灭算符，因此做 $U(1)$ 变换时会有两个复数因子而不只是一个。(8.48)在相变点附近保证成立，一方面，相变点附近库伯对序参量不大，可以级数展开取前几项，另一方面，通过量纲分析可以发现

$$[\Phi] \sim [L]^{-1/2},$$

容易验证没有提到的项在重整化不动点附近全部是不相关的，在重整化群作用下会被压低。我们遵从做朗道-金斯堡理论时的惯例，适当调节单位制（或者说把无关的常数融合进 Φ 中）来让梯度平方项前面的系数为1。

计算(8.48)的极小值点，可以得到

$$-\frac{(\nabla - ie\mathbf{A})^2}{4m^*}\Phi + \alpha\Phi + \beta|\Phi|^2\Phi = 0. \quad (8.49)$$

这个方程的形式和薛定谔方程非常接近，当然这更多是对称性带来的结果，即薛定谔场也具有 $U(1)$ 对称性。

$U(1)$ 对称性给出的守恒流为

$$\mathbf{j} \propto \Phi^* \nabla \Phi - \Phi \nabla \Phi^*, \quad (8.50)$$

当然这就是电流。当然，由于我们只是在做对称性分析，并不能明确地得出式子右边的系数。上式说明超导的电流来自序参量的梯度，这和通常导电的机制（电场下费米面发生移动）不同。由于序参量是复的，即使不存在振幅的空间变化，仅仅靠不同点相位不同就足够产生持续电流。这是一个稳态解，所以不存在能量消耗。这也就是超导体唯象地看起来似乎有某种运动全然不会受到阻碍的“反常电子”的原因。

有效哈密顿量(8.16)的Hubbard-Stratonovich变换

背景知识 8.3: 金斯堡-朗道理论和Hubbard-Stratonovich变换

很多集体行为和相变均可以通过经典的金斯堡-朗道理论加以描述：相变的出现是因为某个对称性（由于一些特殊的相互作用）被破缺了，然后我们可以使用某个序参量来描述相变，而序参量的涨落就给出了相变之后系统的低能自由度。写出序参量满足的场论，就能够分析相变点附近的行为，此即金斯堡-朗道理论。

可以直接通过被破缺的对称性获得金斯堡-朗道理论，但是这样无法确定理论中的系数。标准的做法是使用Hubbard-Stratonovich变换将二体相互作用解耦。所谓Hubbard-Stratonovich变换是指引入一个玻色型辅助场 λ （称为Hubbard-Stratonovich参量），让一个电子-电子相互作用项被分解成电子和 λ 的相互作用。例如我们有

$$\exp(-\bar{\mathcal{O}}[\psi]\mathcal{O}[\psi]) = \text{const} \times \int \mathcal{D}\lambda \exp\left(|\lambda|^2 + \lambda\bar{\mathcal{O}}[\psi] + \bar{\lambda}\mathcal{O}[\psi]\right), \quad (8.51)$$

在 \mathcal{O} 是实的时， λ 应当取成实场。适当选取Hubbard-Stratonovich参量使之和序参量对应，然后，如果可能的话，积掉电子自由度，这样就得到了关于 λ 的有效理论。对凝聚态系统中常见的电子-电子相互作用，可以看到Hubbard-Stratonovich能够直接得到的序参量应该是两个电子场算符的乘积及其线性组合。

容易证明 λ 和 $\mathcal{O}[\psi]$ 的期望值成正比，且在空间点彼此距离足够远时， λ 和 $\mathcal{O}[\psi]$ 的多点关联函数也是成正比的，因此可以将 λ 看成 $\mathcal{O}[\psi]$ 的粗粒化版本。

在远离相变点处，短距离涨落对系统行为仍然有比较大的影响，因此隐隐约约还是能够从系统行为中看到低能细节——如晶格常数——因此完全关于 λ 的场论也许并不能很好地描述系统行为。然而，在相变点处，系统关联长度发散，只有极长程的涨落才是重要的，此时描述系统的理论就是完全关于 λ 的金斯堡-朗道理论，且是这个理论的重整化群不动点。

原则上Hubbard-Stratonovich变换是严格的，但如果 λ 涨落强烈，做完Hubbard-Stratonovich变换之后的作用量中可能有高阶 λ^n 项，从而让解析计算变得不可能。在相变点附近，如果 λ 选择适当，可以认为它接近零。因此，我们可以将做完Hubbard-Stratonovich变换之后的作用量中的高阶 λ^n 项舍弃，如果发现存在 λ 非零的鞍点，且涨落不会破坏这个序，那么系统就有对称性自发破缺，算符 $\mathcal{O}[\psi]$ 大体上就是序参量，且做完Hubbard-Stratonovich变换之后得到的关于 λ 的场论舍弃了高阶 λ^n 项之后就是一个好的相变点附近的金斯堡-朗道理论。需注意这里我们要求 λ 接近零，否则得到的理论一来难以解析计算，二来不在相变点附近，但是如果最后发现作用量只有一个 λ 严格为零的鞍点，那么是没有对称性自发破缺的。

Hubbard-Stratonovich变换和平均场近似关系密切，容易看出如果忽略Hubbard-Stratonovich参量的涨落，即计算Hubbard-Stratonovich变换后的作用量的鞍点，那就是在做平均场近似。实际上平均场近似还要强一些，因为Hubbard-Stratonovich变换后的作用量如果加入一些外场作用，其鞍点允许序参量做经典的振动，但是平均场近似中序参量没有任何时间演化；例如，要分析外加电磁场下序参量如何变动，平均场近似是不靠谱的；另外请注意Hubbard-Stratonovich变换前后的鞍点近似给出的经典运动方程未必是一样的，如果不一样，就说明系统很可能有对称性自发破缺，且我们选取的参量能够描述相应的对称性自发破缺。

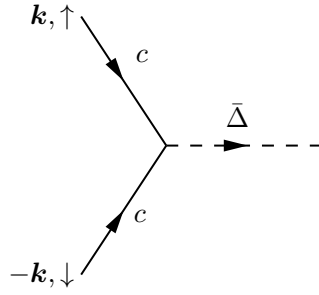


图 8.4: (8.16)做完Hubbard-Stratonovich变换后的相互作用顶角；将所有箭头倒过来， c 换成 \bar{c} ， Δ 换成 $\bar{\Delta}$ 可以得到另一个顶角

Hubbard-Stratonovich变换可以看成平均场理论的扩展。它比平均场理论多出来的东西包括：原则上是准确的（虽然选择正确的Hubbard-Stratonovich参量是非常重要的）；变换后的作用量可以系统地反映平均场近似之上的涨落，而如果使用背景知识8.2引入平均场近似，是很难看出如何改进平均场近似的，使用我们这里的办法则可以很好地考虑平均场之上的涨落；更进一步，如果要分析系统中剩下的费米型激发，可以暂时将Hubbard-Stratonovich参量看成固定的，然后分析费米型激发的行为，最后对Hubbard-Stratonovich参量的不同取值做平均，这为我们仔细分析平均场理论得到的费米子哈密顿量的做法提供了辩护。

从上一段的说法可以看到能够使用Hubbard-Stratonovich变换分析的系统中仍然存在电子型的准粒子（即所谓剩下的费米型激发）。那些在未发生相变时基本自由度就不是电子型激发的系统当然无法使用Hubbard-Stratonovich变换分析。反过来，基本自由度是电子型激发，并且大体上“相变后电子受到序参量产生的一个平均场作用”的图像成立的系统则可以使用Hubbard-Stratonovich变换分析。

模型(8.16)的路径积分形式是

$$S = \int_0^\beta d\tau \left(\sum_{\mathbf{k}, \alpha} \bar{c}_{\mathbf{k}\alpha} (\partial_\tau + \xi_{\mathbf{k}}) c_{\mathbf{k}\alpha} - \frac{g}{V} \sum_{\mathbf{k}, \mathbf{k}'} \bar{c}_{-\mathbf{k}'\downarrow} \bar{c}_{\mathbf{k}'\uparrow} c_{\mathbf{k}\uparrow} c_{-\mathbf{k}\downarrow} \right), \quad Z = \int \mathcal{D}[c, \bar{c}] e^{-S}, \quad (8.52)$$

对 $\sum_{\mathbf{k}} c_{\mathbf{k}\uparrow} c_{-\mathbf{k}\downarrow}$ 做Hubbard-Stratonovich变换，有

$$Z = \int \mathcal{D}[\Delta, \bar{\Delta}] \mathcal{D}[c, \bar{c}] \exp \left(- \int d\tau \left(\sum_{\mathbf{k}, \sigma} \bar{c}_{\mathbf{k}\sigma} (\partial_\tau + \xi_{\mathbf{k}}) c_{\mathbf{k}\sigma} + \bar{\Delta} A + \bar{A} \Delta + \frac{\bar{\Delta} \Delta}{V_0} \right) \right),$$

其中 $\Delta = \Delta(\tau)$ 是一个复场而

$$A = \sum_{\mathbf{k}} c_{\mathbf{k}\uparrow} c_{-\mathbf{k}\downarrow}.$$

引入南部旋量记号

$$\Psi_{\mathbf{k}} = \begin{pmatrix} c_{\mathbf{k}\uparrow} \\ \bar{c}_{-\mathbf{k}\downarrow} \end{pmatrix}, \quad (8.53)$$

则可以将配分函数写成更加紧凑的形式

$$Z = \int \mathcal{D}[\Delta, \bar{\Delta}] \mathcal{D}[\Psi, \bar{\Psi}] \exp \left(- \int d\tau \left(\sum_{\mathbf{k}} \bar{\Psi}_{\mathbf{k}} (\partial_\tau + h_{\mathbf{k}}) \Psi_{\mathbf{k}} + \frac{\bar{\Delta} \Delta}{V_0} \right) \right), \quad (8.54)$$

其中

$$h_{\mathbf{k}} = \begin{pmatrix} \xi_{\mathbf{k}} & \Delta(\tau) \\ \bar{\Delta}(\tau) & -\xi_{\mathbf{k}} \end{pmatrix}. \quad (8.55)$$

积掉电子场 Ψ , 得到

$$Z = \int \mathcal{D}[\Delta, \bar{\Delta}] \exp \left(- \int d\tau \frac{\bar{\Delta}\Delta}{V_0} + \text{tr} \ln(\partial_\tau + h_{\mathbf{k}}) \right). \quad (8.56)$$

至此我们已经得到了在长波极限上和库伯对求和 $\sum_{\mathbf{k}} c_{-\mathbf{k}\downarrow} c_{\mathbf{k}\uparrow}$ 表现完全一致的场 Δ 的严格理论。将(8.56)展开成 Δ 的多项式, 就得到了s波BCS超导的金斯堡-朗道理论, 而且是系数全部都算出来的金斯堡-朗道理论。由于 Δ 没有空间起伏, 这实际上就是纯粹的朗道理论。由于 $\sum_{\mathbf{k}} c_{-\mathbf{k}\downarrow} c_{\mathbf{k}\uparrow}$ 是具有空间平移对称性的, 参量 Δ 自己也就具有空间平移对称性 (而不仅仅是理论具有空间平移对称性), 从而它没有任何空间涨落; 直观地看, 只保留相互作用通道(8.16), 则不存在电子到 Δ 的动量转移, 从而 Δ 在频域上只有 $\mathbf{k} = 0$ 一个分量。(见图8.4)

我们计算(8.56)的鞍点。假设其鞍点处 Δ 不存在时间演化, 则只需要计算

$$\begin{aligned} & \arg \min \left(-\beta \frac{\bar{\Delta}\Delta}{V_0} + \ln \det(\partial_\tau + h_{\mathbf{k}}) \right) \\ &= \arg \min \left(-\beta \frac{\bar{\Delta}\Delta}{V_0} + \ln \det(-i\omega_n + h_{\mathbf{k}}) \right) \\ &= \arg \min \left(-\beta \frac{\bar{\Delta}\Delta}{V_0} + \ln \prod_{\mathbf{k}, \omega_n} (\omega_n^2 + \xi_{\mathbf{k}}^2 + |\Delta|^2) \right), \end{aligned}$$

对 $\bar{\Delta}$ 求导, 即得到

$$-\frac{\beta}{V_0} \Delta + \sum_{\mathbf{k}, \omega_n} \frac{\Delta}{\omega_n^2 + \xi_{\mathbf{k}}^2 + |\Delta|^2} = 0. \quad (8.57)$$

这就是鞍点需要满足的方程。由于 ω_n 也出现在求和中, 尚不容易看出上式和(8.43)之间的关系, 然而经过计算会发现上式的非平庸解正好就是(8.44), 因此我们发现(8.56)的鞍点解的确就是平均场近似给出的。

序参量的空间涨落

实际的BCS超导系统中的序参量是有空间起伏的, 因为库伯对中的两个电子的动量不会完全一样, 例如, 它们可能被电场定向加速, 且相互作用通道也不止(8.16)中的这一种。考虑到这些, 我们就需要一个有动量依赖的 Δ 来完成Hubbard-Stratonovich变换, 从而得到一个序参量真的有空间变化的金斯堡-朗道理论。

考虑(8.18)的路径积分

$$S = \int_0^\beta d\tau \left(\sum_{\mathbf{k}, \alpha} \bar{c}_{\mathbf{k}\alpha} (\partial_\tau + \xi_{\mathbf{k}}) c_{\mathbf{k}\alpha} - \frac{g}{V} \sum_{\mathbf{k}, \mathbf{k}', \mathbf{q}} \bar{c}_{-\mathbf{k}'\downarrow} \bar{c}_{\mathbf{k}'+\mathbf{q}, \uparrow} c_{\mathbf{k}+\mathbf{q}, \uparrow} c_{-\mathbf{k}\downarrow} \right), \quad (8.58)$$

做Hubbard-Stratonovich变换得到

$$\begin{aligned} Z = \int \mathcal{D}[\Delta, \bar{\Delta}] \mathcal{D}[c, \bar{c}] \exp & \left(- \int d\tau \left(\sum_{\mathbf{k}, \alpha} \bar{c}_{\mathbf{k}\alpha} (\partial_\tau + \xi_{\mathbf{k}}) c_{\mathbf{k}\alpha} + \sum_{\mathbf{q}} \frac{1}{g} \bar{\Delta}_{\mathbf{q}} \Delta_{\mathbf{q}} \right. \right. \\ & \left. \left. + \frac{1}{\sqrt{V}} \sum_{\mathbf{k}, \mathbf{q}} \bar{c}_{\mathbf{k}+\mathbf{q}, \uparrow} \bar{c}_{-\mathbf{k}\downarrow} \Delta_{\mathbf{q}} + \frac{1}{\sqrt{V}} \sum_{\mathbf{k}, \mathbf{q}} \bar{\Delta}_{\mathbf{q}} c_{-\mathbf{k}\downarrow} c_{\mathbf{k}+\mathbf{q}, \uparrow} \right) \right), \end{aligned} \quad (8.59)$$

重复之前的手续, 引入南部旋量

$$\Psi = \begin{pmatrix} [c_{\mathbf{k}\uparrow}]_{\mathbf{k}} \\ [\bar{c}_{-\mathbf{k}\downarrow}]_{\mathbf{k}} \end{pmatrix}, \quad (8.60)$$

并设矩阵

$$\Delta = \frac{1}{\sqrt{V}} [\Delta_{\mathbf{q}} \delta_{\mathbf{k}', \mathbf{k}+\mathbf{q}}]_{\mathbf{k}' \mathbf{k}}, \quad (8.61)$$

配分函数写作

$$Z = \int \mathcal{D}[\Psi, \bar{\Psi}] \mathcal{D}[\Delta, \bar{\Delta}] \exp \left(- \int d\tau \left(\sum_{\mathbf{q}} \frac{1}{g} \bar{\Delta}_{\mathbf{q}} \Delta_{\mathbf{q}} + \bar{\Psi} (\partial_{\tau} + h) \Psi \right) \right),$$

其中

$$h = \begin{pmatrix} [\xi_{\mathbf{k}}]_{\mathbf{k}} & \Delta \\ \bar{\Delta} & [-\xi_{\mathbf{k}}]_{\mathbf{k}} \end{pmatrix}. \quad (8.62)$$

积掉电子场, 得到

$$Z = \int \mathcal{D}[\Psi, \bar{\Psi}] \mathcal{D}[\Delta, \bar{\Delta}] \exp \left(- \int d\tau \sum_{\mathbf{q}} \frac{1}{g} \bar{\Delta}_{\mathbf{q}} \Delta_{\mathbf{q}} + \ln \det(\partial_{\tau} + h) \right). \quad (8.63)$$

请注意这里的tr运算不仅作用在所有空间指标上, 也作用在时间上。为了避免定义不清和归一化出错, 我们将虚时间也切换到频域, 这可以通过将(8.61)改写成

$$\Delta = \frac{1}{\sqrt{V\beta}} [\Delta_{\mathbf{q}, i\mathbf{q}^0} \delta_{\mathbf{k}', \mathbf{k}+\mathbf{q}} \delta_{k'^0, k^0+q^0}]_{\mathbf{k}' \mathbf{k}}, \quad (8.64)$$

并相应地将(8.59)中时域上的场替换成频域的。

背景知识 8.4: 矩阵求对数再求迹的泰勒展开

设矩阵 B 较小, 则有展开

$$\text{tr} \ln(A + B) = \text{tr} \ln A + \text{tr} \ln(1 + A^{-1}B) = \text{tr} \ln A + \sum_{n \geq 1} \frac{(-1)^{n-1}}{n} \text{tr}((A^{-1}B)^n). \quad (8.65)$$

不需要 A, B 对易。这个公式在获取金斯堡-朗道理论时非常有用。

我们现在要计算

$$\ln \det(\partial_{\tau} + h) = \text{tr} \ln(\partial_{\tau} + h).$$

在 Δ 很大时, 将这一项对 Δ 做泰勒展开没有什么意义, 但是在临界点附近 Δ 很小, 从而取前几项就足够精确描写系统了。设没有相互作用时的单电子格林函数为 G_k^e 而单空穴格林函数为 G_k^h , 即

$$G_k^e = \frac{1}{-ik^0 + \xi_{\mathbf{k}}}, \quad G_k^h = \frac{1}{-ik^0 - \xi_{\mathbf{k}}}, \quad (8.66)$$

则

$$\partial_{\tau} + h = \begin{pmatrix} \partial_{\tau} + [\xi_{\mathbf{k}}]_{\mathbf{k}} & \Delta \\ \bar{\Delta} & \partial_{\tau} - [\xi_{\mathbf{k}}]_{\mathbf{k}} \end{pmatrix} = \begin{pmatrix} [G_k^e]_{\mathbf{k}}^{-1} & \Delta \\ \bar{\Delta} & [G_k^h]_{\mathbf{k}}^{-1} \end{pmatrix},$$

设

$$G_0^{-1} = \begin{pmatrix} [G_k^e]_{\mathbf{k}}^{-1} & 0 \\ 0 & [G_k^h]_{\mathbf{k}}^{-1} \end{pmatrix}, \quad (8.67)$$

我们有

$$\ln(\partial_\tau + h) = \ln \left(G_0^{-1} + \begin{pmatrix} 0 & \Delta \\ \bar{\Delta} & 0 \end{pmatrix} \right) = \ln G_0^{-1} - \sum_{n \geq 1} \frac{1}{2n} \text{tr} \left(\left(G_0 \begin{pmatrix} 0 & \Delta \\ \bar{\Delta} & 0 \end{pmatrix} \right)^{2n} \right), \quad (8.68)$$

这里用到了 $G_0 \begin{pmatrix} 0 & \Delta \\ \bar{\Delta} & 0 \end{pmatrix}$ 无迹这一事实。 $\ln G_0^{-1}$ 一项是常数，可以略去。 $n = 1$ 项为

$$\begin{aligned} -\frac{1}{2} \text{tr} \left(\left(G_0 \begin{pmatrix} 0 & \Delta \\ \bar{\Delta} & 0 \end{pmatrix} \right)^2 \right) &= -\text{tr}([G_k^e]_k \Delta [G_k^h]_k \bar{\Delta}) \\ &= -\frac{1}{V\beta} \sum_{k, k', q, q'} G_{k'}^e \Delta_q \delta_{k', k+q} G_k^h \bar{\Delta}_{q'} \delta_{k', k+q'} \\ &= -\frac{1}{V\beta} \sum_{q, k} \Delta_q \bar{\Delta}_q G_{k+q}^e G_k^h, \end{aligned}$$

推导时应小心指标，尤其注意到 $\bar{\Delta}$ 矩阵相对于 Δ 做了转置，因此(8.64)中的 k, k' 在 $\bar{\Delta}$ 的定义中需要交换顺序。因此作用量中， Δ^2 阶的部分是

$$S_{\Delta^2} = \sum_q \left(\frac{1}{g} + \frac{1}{V\beta} \sum_k G_{k+q}^e G_k^h \right) \bar{\Delta}_q \Delta_q = \sum_q \left(\frac{1}{g} - \frac{1}{V\beta} \sum_k G_{k+q}^e G_{-k}^e \right) \bar{\Delta}_q \Delta_q, \quad (8.69)$$

其中

$$G_k^h = \frac{1}{i(-k^0) - \xi_{-k}} = -G_{-k}^e. \quad (8.70)$$

S_{Δ^2} 也可以通过一种更加直观的方法得到：只考虑二顶角图（此处的顶角为图8.4，但允许两个电子动量不同， Δ 有动量），那么修正后的 Δ 传播子是

$$= = \Rightarrow = = = - - - \Rightarrow - - - + - \Rightarrow - - - \text{ (loop) } - - - \Rightarrow - - -, \quad (8.71)$$

正好就是(8.69)对应的传播子（负号可以解释成闭合费米子回路导致的）。注意到 $\sum_k G_{k+q} G_{-k}$ 也出现在了(8.29)中，这当然不是偶然的：通过图形很容易发现电子的梯形图近似实际上就是库伯对的环形图近似（或者，注意到电子的梯形图近似(8.28)修正了 g ，而 $\int d\tau \sum_q 1/g \bar{\Delta}_q \Delta_q$ 正好是未经过相互作用修饰的库伯对的作用量），而(8.71)是库伯对的环形图近似的最多保留两个顶角的低阶项。以上两种等价的近似实际上也等价于只计算库伯对的最低阶自能，即认为电子-库伯对转化顶角导致的自能修正为

$$-\Sigma_q = \frac{1}{V\beta} \sum_k (-G_{k+q}^e)(-G_{-k}^e). \quad (8.72)$$

按照(8.30)，这就是

$$\Sigma_q = \frac{1}{V} \sum_p \frac{1 - f(\xi_p) - f(\xi_{p+q})}{iq^0 - \xi_p - \xi_{p+q}}. \quad (8.73)$$

我们现在切换回实空间。前面说过，能量最低的场构型中 Δ 在时间和空间上都是均匀的，即参与库伯配对的两个电子动量基本上只差一个负号，从而，基态附近的 q 通常都很小，类似的 q^0 也很小，这样我们将对 q 展开，由于考虑的是s波超导，

$$\Sigma_q = \Sigma_0 + iq^0 \left. \frac{\partial \Sigma_q}{\partial (iq^0)} \right|_{q=0} + \frac{1}{6} |q|^2 \nabla_q^2 \Sigma_q|_{q=0} + \cdots,$$

但是更进一步的计算会发现

$$\left. \frac{\partial \Sigma_q}{\partial (iq^0)} \right|_{q=0} =$$

因此最后通过适当地重新定义 Δ 以囊括一些因子，就能够得到

$$S_{\Delta^2} = \int d\tau \int d^3\mathbf{r} \left(\frac{c}{2} |\nabla \Delta|^2 + \frac{r}{2} |\Delta|^2 \right). \quad (8.74)$$

(8.68)中的 $n = 2$ 项则是 Δ 的一个四阶项，同样，它也有 $|\mathbf{q}|^2$ 和 iq^0 的依赖，但是既然我们认为 \mathbf{q} 和 Δ 都是小量，实际上只有 $|\Delta|^4$ 的量级是足够客观的，于是可以直接写出

$$S = \int d\tau \int d^3\mathbf{r} \left(\frac{c}{2} |\nabla \Delta|^2 + \frac{r}{2} |\Delta|^2 + \frac{u}{4} |\Delta|^4 \right). \quad (8.75)$$

我们可以发现以上计算过程给出的有效作用量可以有量子涨落，但是自身没有动力学，或者说BCS序参量 Δ 在没有外场时没有时间演化。

在温度稍微高一些时量子涨落可以忽略，此时我们得到所谓的s波超导体的经典金斯堡-朗道理论

$$S = \beta \int d^3\mathbf{r} \left(\frac{c}{2} |\nabla \Delta|^2 + \frac{r}{2} |\Delta|^2 + \frac{u}{4} |\Delta|^4 \right). \quad (8.76)$$

这是朗道

8.1.5 s波BCS超导体的电磁学性能

伦敦方程

背景知识 8.5: 规范场

设 G 是一个李群， S 是它的一个表示，使用参数 α 标记。在底流形 M 上每一点放置一个 S ，并让 α 可以随着时空点的不同而变化。设 M 上的场 $\psi(x)$ 的目标空间（即某一点上场值的取值范围）携带了 S 的一个表示。如此构造之后，就能够定义规范变换

$$\psi'(x) = S(x)\psi(x).$$

现在设我们有了一个关于 $\psi(x)$ 的拉氏量 \mathcal{L}_ψ 。如果我们希望让 \mathcal{L}_ψ 成为规范不变的，即如果另外找任意一个 $S'(x)$ ，根据 $S'(x)$ 做规范变换，拉氏量不变，那么可以验证，只需要将 \mathcal{L}_ψ 中的所有导数替换为以 $-ieA_\mu$ 为联络的协变导数即可，其中 $A_\mu(x)$ 是 x 点的 $S(x)$ 的 $\alpha(x)$ 发生无穷小变化而产生的 $S(x)$ 的变化，从而是 S 表示的生成元的一个线性组合。

因此我们获得了一种批量生产拉氏量的方式：给定一些场，选取特定的规范群 G ，使得前述的场能够携带 G 的某个表示，即可自动生成一个规范场（实际上是一个多分量的场），将前述的各种场耦合在一起。高能物理中最为成功的标准模型就是通过这种方式构造出来的。在凝聚态物理中，以上构造方式也可以大大简化推导，例如要推导BCS理论中库伯对的作用量，我们只需要知道库伯对也携带了规范群 $U(1)$ 的某个表示，就知道了它和电磁场耦合的方式，从而无需从与电磁场耦合的库伦排斥电子气的场论出发从头推导一遍就能够知道库伯对和电磁场耦合的方式。唯一尚有待分析的是一个库伯对的电荷量，但是这也可以通过电荷守恒很自然地得到（库伯对的流和对应的电子的流必须给出一样的电荷流）。

引入规范场的一个直接结果是与规范场耦合的粒子（即场 ψ 产生的粒子）获得Berry相位。

虽然以上构造是在一个底流形 M 上进行的，但是把 M 换成一个格点系统也可以；不过此时要注意，规范联络应该定义在 ψ 场对应的粒子能够跃迁的**边**上而不是格点上，虽然诸 $S(x)$ 还是定义在格点上的。

8.1.6 超导和超流

(8.75)是一个有排斥相互作用的玻色场论。这样的系统可能会呈现一个**超流相**。这是BCS超导的一个直观解释：电子配对形成的库伯对组成的玻色流体在低温下进入超流相，从而可以没有阻碍地移动。

背景知识 8.6: 近独立玻色气体的玻色-爱因斯坦凝聚

近独立玻色气体的基态是所谓的**玻色-爱因斯坦凝聚**。可以从玻色分布函数的形式看出这一点： $\epsilon < \mu$ 时分布函数就是负的，不物理，从而一定有 $\epsilon \geq \mu$ 。而另一方面，简单的积分就能够说明， $\epsilon > \mu$ 这部分能级能够容纳的玻色子数目是有限的，随着温度降低而降低。因此，在零温下，压倒性多数的玻色子停留在 $\epsilon = \mu$ 能级上，进一步的分析表明它们构成一个相干态。

玻色液体的超流相

在超流相中，粒子数流密度为

$$\mathbf{j} = \frac{i}{2m}(\psi \nabla \bar{\psi} - \bar{\psi} \nabla \psi),$$

在 $U(1)$ 对称性破缺相中就有

$$\mathbf{j} = \frac{\rho_0}{m} \nabla \theta. \quad (8.77)$$

这就意味着， $U(1)$ 对称性破缺相中，只需要序参量 θ ——也就是玻色子场的相角——在空间上有一个涨落，就**一定**会有粒子数流密度，而且是没完没了、不会衰减的流密度。因此 $U(1)$ 对称性破缺相就是**超流相**。

超流中的拓扑激发

8.1.7 BCS以外的超导机制

配对机制可以不是声子介导的有效吸引相互作用。

有效吸引相互作用在一些时候可能不能被看成瞬时的。如果一个声子被创生之后可以在空间中激发出别的声子，且声子能够在材料中产生诸如电流等涉及范围较大且短时间内不容易被弛豫掉的模式，那么有效的电子-电子相互作用就根本不应该看成瞬时的。

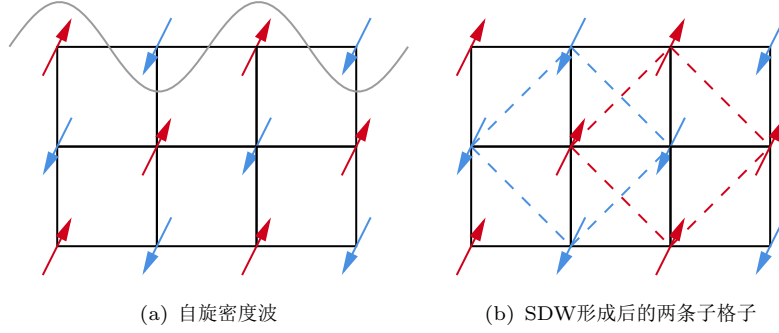


图 8.5: 二维正方晶格上的反铁磁序

8.2 相邻自旋-自旋相互作用导致的二维正方晶格的反铁磁长程序

8.2.1 二维正方晶格上的自旋

现在考虑一个二维正方晶格，它可以有一个反铁磁相，也即，相邻格点的自旋倾向于变得反平行，或者说形成一个自旋密度波。一种能够产生反铁磁序的哈密顿量为

$$H = \sum_{\mathbf{k}, \alpha} \xi_{\mathbf{k}} c_{\mathbf{k}\alpha}^\dagger c_{\mathbf{k}\alpha} + J \sum_{\langle i, j \rangle} \mathbf{S}_i \cdot \mathbf{S}_j, \quad (8.78)$$

其中

$$\mathbf{S}_i = \sum_{\alpha, \beta} c_{i\alpha}^\dagger \boldsymbol{\sigma}_{\alpha\beta} c_{i\beta} \quad (8.79)$$

为格点 i 的自旋矢量。这个模型本身其实并不非常现实，因为自旋相互作用通常来自交换能，但是交换能通常在绝缘系统中比较重要，那么就不应该有一个动能项。设反铁磁序的序参量为 ϕ ，一个不错的选择是

$$\langle \mathbf{S}_i \rangle = (-1)^i \phi, \quad (8.80)$$

这里的 $(-1)^i$ 实际上是一种滥用记号：它实际上是

$$(-1)^i = (-1)^{i_x + i_y}$$

的简写。为了方便起见，我们把 $(i_x + i_y)$ 为奇数的格点的全体记为 A ，将 $(i_x + i_y)$ 为偶数的格点的全体记为 B ，于是 A 中任何一个格点的近邻格点都在 B 中，反之亦然。如果 ϕ 非零，那么显然 $\langle \mathbf{S}_i \rangle$ 在 $i \in A$ 时和 $\langle \mathbf{S}_i \rangle$ 在 $i \in B$ 时差一个负号，即相邻的自旋一定是反向的，正好意味着形成了反铁磁序。

不失一般性地令晶格常数为1，则第一布里渊区为 $[-\pi, \pi]^2$ 。在形成了一个完整的反铁磁序之后序参量 Δ 的周期应该是两个格点，于是应有 $\mathbf{Q} = (\pi, \pi)$ 。

8.2.2 平均场近似

在(8.78)中做平均场近似

$$\mathbf{S}_i \cdot \mathbf{S}_j = \langle \mathbf{S}_i \rangle \cdot \mathbf{S}_j + \mathbf{S}_i \cdot \langle \mathbf{S}_j \rangle - \langle \mathbf{S}_i \rangle \cdot \langle \mathbf{S}_j \rangle, \quad (8.81)$$

忽略平均场近似引入的常数项（我们这里不做自洽计算而只是分析相变点具有的性质），自旋相互作用项为

$$\begin{aligned} H_{\text{int}} &= J \sum_{\langle i, j \rangle} \mathbf{S}_i \cdot \mathbf{S}_j \sim J \sum_{\langle i, j \rangle} (\langle \mathbf{S}_i \rangle \cdot \mathbf{S}_j + \mathbf{S}_i \cdot \langle \mathbf{S}_j \rangle) \\ &= 2J \sum_{\langle i, j \rangle} \langle \mathbf{S}_j \rangle \cdot \mathbf{S}_i \\ &= J \sum_i \sum_{j \in \text{nn of } i} \langle \mathbf{S}_j \rangle \cdot \mathbf{S}_i, \end{aligned}$$

其中nn表示“最近邻”，第二个等号中因子2消失是因为 $\langle i, j \rangle$ 对一对近邻只求和一次，而第三行中一对近邻实际上被求和了两次。代入(8.79)，注意到两个相邻格点的 $(-1)^i$ 差一个负号，有

$$\begin{aligned} H_{\text{int}} &= J \sum_i \sum_{j \in \text{nn of } i} (-1)^j \phi \cdot \mathbf{S}_i \\ &= -J \sum_i (-1)^i \sum_{j \in \text{nn of } i} \phi \cdot \mathbf{S}_i, \end{aligned}$$

由于系统具有自旋旋转不变性，不失一般性地要求 ϕ 指向 z 轴，并设

$$\Delta = 4J\phi, \quad (8.82)$$

则

$$H_{\text{int}} = - \sum_i (-1)^i S_i^z \Delta = - \sum_i (-1)^i \Delta (c_{i\uparrow}^\dagger c_{i\uparrow} - c_{i\downarrow}^\dagger c_{i\downarrow}). \quad (8.83)$$

设 $\mathbf{Q} = (\pi, \pi)$ ，则相互作用哈密顿量可以写成以下傅里叶变换的形式：

$$H_{\text{int}} = - \sum_i e^{i\mathbf{Q} \cdot \mathbf{R}_i} \Delta (c_{i\uparrow}^\dagger c_{i\uparrow} - c_{i\downarrow}^\dagger c_{i\downarrow}) = -\Delta \sum_{\mathbf{k}} (c_{(\mathbf{k}+\mathbf{Q})\uparrow}^\dagger c_{\mathbf{k}\uparrow} - c_{(\mathbf{k}+\mathbf{Q})\downarrow}^\dagger c_{\mathbf{k}\downarrow}). \quad (8.84)$$

于是最后平均场哈密顿量就是

$$H_{\text{MF}} = \sum_{\mathbf{k}, \alpha} (\xi_{\mathbf{k}} c_{\mathbf{k}\alpha}^\dagger c_{\mathbf{k}\alpha} - \alpha \Delta c_{(\mathbf{k}+\mathbf{Q})\alpha}^\dagger c_{\mathbf{k}\alpha}). \quad (8.85)$$

这里指定 \uparrow 对应1， \downarrow 对应-1，虽然电子的自旋为 $\pm 1/2$ 而不是 ± 1 。(8.85)两边乘上2，并注意对 \mathbf{k} 求和等价于对 $\mathbf{k} + \mathbf{Q}$ 求和，且 \mathbf{k} 等价于 $\mathbf{k} + 2\mathbf{Q}$ ，(8.85)给出的能谱等价于以下矩阵

$$\begin{pmatrix} \xi_{\mathbf{k}} & -\alpha\Delta \\ -\alpha\Delta & \xi_{\mathbf{k}+\mathbf{Q}} \end{pmatrix}$$

的本征值，也就是

$$E_{\mathbf{k}} = \frac{\xi_{\mathbf{k}} + \xi_{\mathbf{k}+\mathbf{Q}}}{2} \pm \sqrt{\frac{(\xi_{\mathbf{k}} - \xi_{\mathbf{k}+\mathbf{Q}})^2}{4} + \alpha^2 \Delta^2}. \quad (8.86)$$

这样可以得到两条能带，在 Δ 为零时它们可以交叉，在 Δ 不为零时能带交叉的部位就打开了一个能隙。

8.2.3 热点和它附近的低能有效理论

(8.85)是(8.78)在反铁磁序相的一个有效理论。反铁磁序破缺了晶格上的平移不变性，因为反铁磁序状态中每个晶格和临近的晶格的自旋差一个负号。反铁磁序还保留了一部分

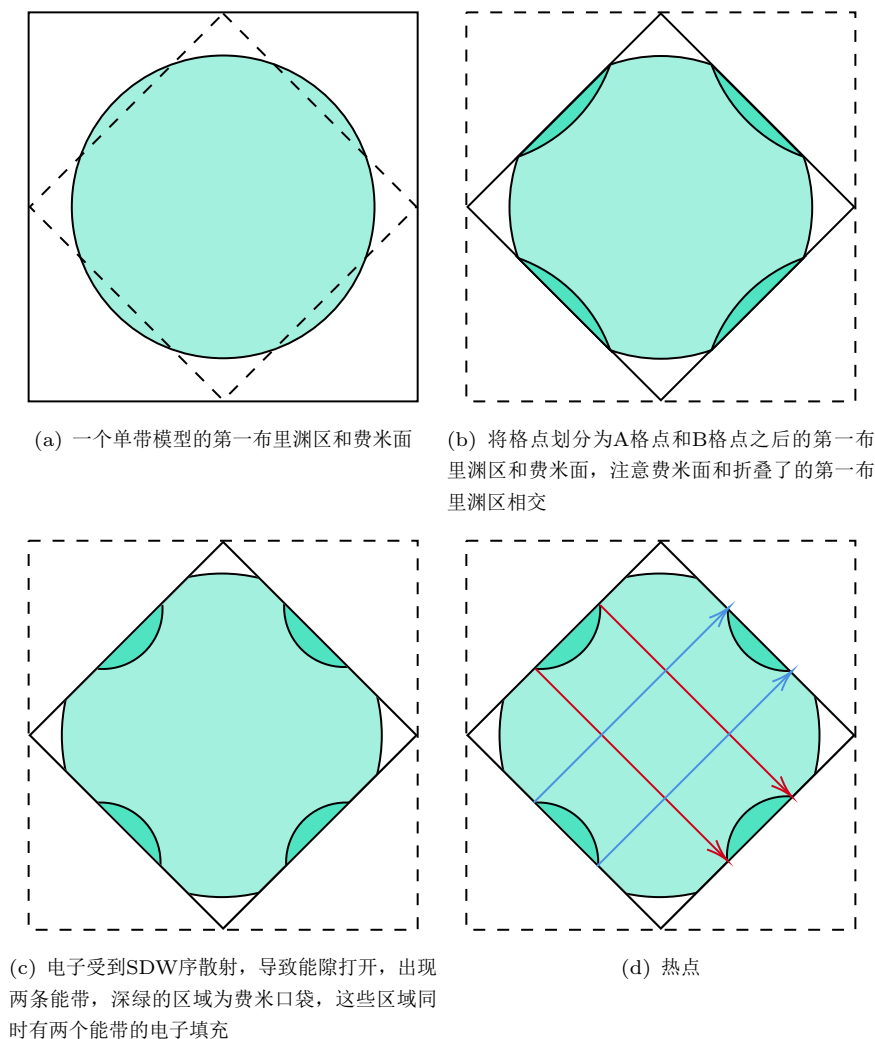


图 8.6: 布里渊区折叠和费米口袋形成

平移不变性：子格点A和B上仍有平移不变性，这两个子格点对应的倒格矢都是 \mathbf{Q} 。这样，(8.85)的第一布里渊区是

$$|k_x| + |k_y| \leq \pi,$$

这是(8.78)的第一布里渊区折叠之后得到的。布里渊区变小当然是因为原胞变大了——从单个格点变成了一个A格点加上一个B格点。（见第2.4.2节）

如果费米面和折叠后的第一布里渊区相交，费米面交叉处打开能隙，形成一系列小的费米口袋。热点附近的 \mathbf{k} 和 $\mathbf{k} + \mathbf{Q}$ 都在热点附近。

接下来讨论在热点附近的低能有效理论，换言之，我们开始讨论超越平均场的物理。需要注意的是考虑超越平均场的物理意味着将原本被忽略的涨落重新加入，这可能改变热点的位置或者甚至让热点消失（在本问题中不可能因为涨落不大，但确实有这样的体系，涨落会让平均场理论中出现的现象——如相变——消失，比如一维伊辛模型）。因此，下面讨论的热点附近的低能有效理论建立在三个基础上：

1. 费米面上存在热点，这是通过平均场理论算出来的，我们假定这个现象确实存在，而不是像一维伊辛模型一样只是幻象；

2. 热点附近的低能有效理论是(8.78)的低能有效理论，即相互作用项一定是自旋-自旋相互作用（这并非假设，而是必然的事实）；
3. 低能自由度是低能电子和自旋波模式这两种场（实际上这是最重要的一个假设，因为低能自由度是什么通常难以直接计算得到）。

(8.78)中的相互作用项是一个电子和一个自旋波模式（这是玻色子）发生散射，得到另一个电子，或者也可以说是自由电子的自旋和自旋波模式的相互作用。⁴因此低能有效理论的散射项形如⁵

$$\phi_{\mathbf{q}} c_{\mathbf{k}+\mathbf{q}\sigma}^{\dagger} c_{\mathbf{k}\sigma'},$$

由于是低能有效理论，我们认为形成了一个基本上稳定的反铁磁序，于是 \mathbf{q} 接近 \mathbf{Q} ，而电子能量 $\epsilon_{\mathbf{k}}$ 和 $\epsilon_{\mathbf{k}+\mathbf{q}}$ 都在费米面附近。这些条件只有在热点附近才能达到。当然由于热点是费米面交叠之后打开能隙的位置，在这附近有低能有效理论是非常合理的。这样，一个散射过程涉及两个热点。分别用1和2标记两个热点，记它们的动量为 \mathbf{K}_1 和 \mathbf{K}_2 ，则有

$$\mathbf{K}_2 = \mathbf{K}_1 + \mathbf{Q},$$

并重新定义 \mathbf{q} 和 \mathbf{k} 为它们偏离 \mathbf{Q} 和 \mathbf{K}_1 的大小。这样，相互作用哈密顿量就是

$$\phi_{\mathbf{q}} \cdot \left(\sum_{\alpha, \beta} c_{1(\mathbf{k}+\mathbf{q})\alpha}^{\dagger} \sigma_{\alpha\beta} c_{2\mathbf{k}\beta} + \text{h.c.} \right),$$

这里任何一个散射过程中入射电子为热点1附加，出射电子在热点2附加或相反的原因是要保证所有电子的动量都在热点附加。于是完整的有效理论的哈密顿量为

$$H_{\text{eff}} = \sum_{\mathbf{k}, \sigma} (\xi_{1\mathbf{k}} c_{1\mathbf{k}\sigma}^{\dagger} c_{1\mathbf{k}\sigma} + \xi_{2\mathbf{k}} c_{2\mathbf{k}\sigma}^{\dagger} c_{2\mathbf{k}\sigma}) + H[\phi] + \lambda \sum_{\mathbf{q}} \phi_{\mathbf{q}} \cdot \left(\sum_{\alpha, \beta} c_{1(\mathbf{k}+\mathbf{q})\alpha}^{\dagger} \sigma_{\alpha\beta} c_{2\mathbf{k}\beta} + \text{h.c.} \right). \quad (8.87)$$

上式中没有明确提玻色场——也就是自旋波场——的自由哈密顿量。这个自由哈密顿量通常是通过对称性分析得出的。进一步，由于是低能理论，将 $\xi_{\mathbf{k}}$ 在费米面附近做展开，仅保留一阶项，得到

$$\xi_{1\mathbf{k}} = \epsilon_{\mathbf{k}} - \mu = \mathbf{v}_1 \cdot \mathbf{k}, \quad \xi_{2\mathbf{k}} = \epsilon_{\mathbf{k}} - \mu = \mathbf{v}_2 \cdot \mathbf{k}. \quad (8.88)$$

使用 \mathbf{v}_1 和 \mathbf{v}_2 来标记 $\nabla_{\mathbf{k}} \xi_{\mathbf{k}}$ 是因为如果是自由电子气，那么它们就是费米速度。

8.2.4 朗道阻尼

现在我们尝试把电子完全积掉，只留下玻色型的自旋波模式。需要指出的是这个操作并不总是可行的：由于自旋波模式和电子都是低能自由度，简单地积掉其中一个自由度可能不能得到一个良定义的有效理论。实际上，对二维体系的确会有这种棘手的细节。下面的操作都是在默认确实可以积掉电子的前提下进行的。

从(8.87)可以看出，一个自旋波模式可以衰变成一对电子空穴对，或者说一个自旋波模式可以将动量转移给一个电子而得到另一个电子，而自身湮灭。因此自旋波模式是有限

⁴自旋波模式是一部分形成了自旋波模式的电子被积掉之后得到的，和未形成自旋波长程序的电子是两群不同的电子。

⁵一个可能的问题是，为什么一定保留低能电子自由度和自旋波自由度？为什么不能是自旋波自由度和密度波自由度？不考虑自旋波-密度波自由度是因为没有明确的物理机制让这两个模式发生耦合，保留低能电子自由度是因为自旋算符和电子数算符对易，但和单个费米子算符并不对易，因此可能出现费米子自由度积不掉的情况。这和BCS理论是不一样的，在BCS理论中大量电子参与配对，以至于哪怕电子自由度积不掉，它也不会产生太大作用。

的寿命的。由费米黄金法则，一个动量为 \mathbf{q} ，能量为大于零的 ω 的自旋波模式的寿命倒数为

$$\begin{aligned} \frac{1}{\tau} &\sim 2\lambda^2 \int \frac{d^2\mathbf{k}}{(2\pi)^2} \delta(\omega + \epsilon_{1\mathbf{k}} - \epsilon_{2(\mathbf{k}+\mathbf{q})}) \theta(-\epsilon_{1\mathbf{k}}) \theta(\epsilon_{2(\mathbf{k}+\mathbf{q})}) \\ &\sim \lambda^2 \int \frac{dp_1 dp_2}{(2\pi)^2 |\mathbf{v}_1 \times \mathbf{v}_2|} \delta(\omega + p_1 - p_2) \theta(-p_1) \theta(p_2), \end{aligned}$$

其中我们设

$$p_1 = \mathbf{v}_1 \cdot \mathbf{k}, \quad p_2 = \mathbf{v}_2 \cdot (\mathbf{k} + \mathbf{q}),$$

我们仅考虑低能理论，因此假定入射电子在费米面以下，出射电子在费米面以上。由几何关系，

$$\int \frac{dp_1 dp_2}{|\mathbf{v}_1 \times \mathbf{v}_2|} \delta(\omega + p_1 - p_2) \theta(-p_1) \theta(p_2) = \sqrt{2}\omega,$$

$\omega < 0$ 的情况也是一样的。总之最后自旋波模式的寿命为

$$\frac{1}{\tau} \sim \frac{\lambda^2}{|\mathbf{v}_1 \times \mathbf{v}_2|} |\omega| = \gamma |\omega|. \quad (8.89)$$

在格林函数中，衰变几率对应着自能修正，会直接反映在 ϕ 的格林函数——从而有效作用量——中，也即，自旋波模式的推迟格林函数形如

$$G_\phi^{-1} = i\gamma|\omega| + \cdots = i\gamma \operatorname{sgn}(\omega)\omega + \cdots,$$

相应的，松原格林函数形式为⁶

$$G_\phi^{-1} = \gamma|\omega_n| + \cdots.$$

这就意味着自旋波模式的有效热力学作用量为

$$\begin{aligned} S_{\text{eff}} &= \sum_{\mathbf{q}, \omega_n} \phi(-\mathbf{q}, -\omega_n) \cdot (\gamma|\omega_n| + \omega_n^2 + c^2 \mathbf{q}^2) \phi(\mathbf{q}, \omega_n) \\ &= \sum_{\mathbf{q}, \omega_n} \phi^*(\mathbf{q}, \omega_n) \cdot (\gamma|\omega_n| + \omega_n^2 + c^2 \mathbf{q}^2) \phi(\mathbf{q}, \omega_n). \end{aligned} \quad (8.90)$$

第二个等号是因为 ϕ 是实场，因为负的动量/频率等价于取复共轭。 ω_n^2 项和 \mathbf{q}^2 项都是对称性分析加入的项，除此以外的项在低能有效理论中并不重要。 ω_n^2 项和 \mathbf{k}^2 项可以容易地切换到实空间，它们分别对应着时间导数平方项 $(\partial_\tau \phi)^2$ 和梯度平方项 $(\nabla \phi)^2$ ，但 $|\omega_n|$ 在实空间中并没有简单的形式。不过，一个正比于 ω_n 的项意味着实空间中有某种阻尼，这也是正确的，因为自旋波模式如前所述会衰变。这种阻尼或者衰变称为朗道阻尼。朗道阻尼指的是没有粒子之间的相互碰撞，仅仅粒子和波强烈耦合也能够产生阻尼。

8.2.5 RPA近似计算朗道阻尼

还可以通过直接计算格林函数的方式来得到 ϕ 的有效理论。这个有效理论当然取

$$Z_{\text{eff}} = \int \mathcal{D}\phi e^{-\sum_{\mathbf{q}, \omega_n} \phi^*(\mathbf{q}, \omega_n) G_\phi^{-1}(\mathbf{q}, \omega_n) \phi(\mathbf{q}, \omega_n)}$$

的形式。

8.2.6 Hertz理论

最后我们讨论上一节得到的低能有效理论的

⁶这里的步骤是：先做Wick转动，即 $\omega = i\omega_n$ ，

8.2.7 RPA近似

我们尝试对平均场近似做一些修正。为此我们将不再直接处理自旋算符，而是把一切都转化到费米子算符上。做傅里叶变换

$$S_q = \frac{1}{\sqrt{N_{\text{site}}}} \sum_i S_i e^{i\mathbf{r}_i \cdot \mathbf{q}} = \frac{1}{\sqrt{N_{\text{site}}}} \sum_{\mathbf{k}} c_{\mathbf{k}+\mathbf{q},\alpha}^\dagger \boldsymbol{\sigma}_{\alpha\beta} c_{\mathbf{k}\beta},$$

这样相互作用项就是（请注意 $\boldsymbol{\sigma}$ ）

对相互作用项做正规序不会影响定性的结果，我们不需要动手算就知道正规序和原本的相互作用只会差一个单体项，而这个单体项是自旋旋转不变的，那么它只会对 $\epsilon_{\mathbf{k}} - \mu$ 做一个修正。于是我们将要处理以下相互作用哈密顿量：

$$H = \frac{1}{2N_{\text{site}}} \sum_{\mathbf{k}, \mathbf{k}', \mathbf{q}, \alpha, \beta} c_{\mathbf{k}-\mathbf{q},\alpha}^\dagger c_{\mathbf{k}'+\mathbf{q},\beta}^\dagger V(\mathbf{q}) c_{\mathbf{k}'} c_{\mathbf{k}}, \quad V(\mathbf{q}) = 2J(\cos(q_x) + \cos(q_y)). \quad (8.91)$$

$$Z = \int \mathcal{D}\phi \exp\left(-\int_0^\beta d\tau\right) \quad (8.92)$$

ϕ 就像驱动自旋的一个外场一样，在只取鞍点近似时它就是平均场序参量。

8.3 local repulsion导致的反铁磁长程序

前一节讨论的模型中的反铁磁序能够形成是因为我们塞入了一个能够形成反铁磁序的哈密顿量。实际上，如果模型中没有显式的自旋-自旋相互作用，最后可能也是能够形成一个反铁磁序的。本节研究二维正方格子上的如下模型：

$$H = -t \sum_{\langle i,j \rangle, \alpha} t(c_{i\alpha}^\dagger c_{j\alpha} + \text{h.c.}) - \mu N + U \sum_i n_i^2, \quad (8.93)$$

其自由理论部分就是一个典型的紧束缚模型，其能谱为

$$\xi_{\mathbf{k}} = -2t(\cos(k_x a) + \cos(k_y a)) - \mu. \quad (8.94)$$

这个模型基本上是Hubbard模型去掉自旋依赖的版本。它在低温下可能能够产生CDW序或是SDW序。它能够产生什么序可以通过仿照BCS超导，通过计算相应的序参量的关联函数得到。

第九章 无序与局域化

理想晶体是高度有序的，但是实际的固体多少具有一些无序性。使用格点

9.1 Anderson局域化

9.1.1 固体中的无序

本节将介绍一种现象：当固体中出现无序，即平移对称性破缺时，电子会倾向于留在原地而不发生长距离跃迁。这种现象称为Anderson局域化。无序可以来自掺杂，也可以来自离子实际上并没有构成晶体这一事实。

例如可以写出这样的一个带有格点无序的紧束缚模型：

$$H = -t \sum_{\langle i,j \rangle} (c_i^\dagger c_j + \text{h.c.}) + \sum_i \epsilon_i c_i^\dagger c_i, \quad (9.1)$$

其中 ϵ_i 是一个随机变量，它基本上就是一个在空间各点不一样的“化学势”，让一些地方比另一些地方更能够留住电子。接下来考虑这样的 ϵ_i 的分布特例：

$$\overline{\epsilon_i} = 0, \quad \overline{\epsilon_i \epsilon_j} = \delta_{ij} W^2, \quad W^2 \neq 0. \quad (9.2)$$

这样， W/t 就表征了无序的大小。这里我们使用 $\overline{\cdots}$ 表示对杂质的构型的平均，而用 $\langle \cdots \rangle$ 表示给定杂质的构型之后做量子力学的平均。如果 ϵ_i 的分布服从高斯分布，对 $\overline{\cdots}$ 同样有Wick定理成立，这将为后面的费曼图计算提供很大的方便。

在 $W/t = 0$ 时就得到一个普通的能带，而如果 $W/t \gg 1$ ，即无序很强，则体系的哈密顿量近似为

$$H = \sum_i \epsilon_i c_i^\dagger c_i,$$

这个哈密顿量在以瓦尼尔波函数为基底的表象下是对角化的，这样电子倾向于固定在原地，即出现局域化。另一方面，完全没有随机扰动时，电子可以自由地移动。这意味着随着 W/t 的增大，或是会出现一个金属-绝缘体相变，或是会出现一个crossover。

一些简单的论证可以说明这种模型的电子结构。局域态和（能够长距离移动，从而能够导电的）延展态应该不共享能量。否则，设一个局域态和一个延展态具有一样的能量，那么将这个局域态和延展态做线性组合，得到另一个延展态，也能具有一样的能量，因此最终我们会发现，同一个能量上有一个被大量延展态环绕着的局域态，这种情况下我们其实可以直接忽略这个局域态。 $W = 0$ 时我们有一个连续的延展态能带。通过微扰论可以发现，在这个能带两端的波函数将转化为局域态。因此含无序的金属的电子能带结构包括一个延展态能带和位于延展态能带两边的所谓的带尾局域态（见图9.1）。随着 W/t 的增大，会有越来越多的延展态转化为局域态。

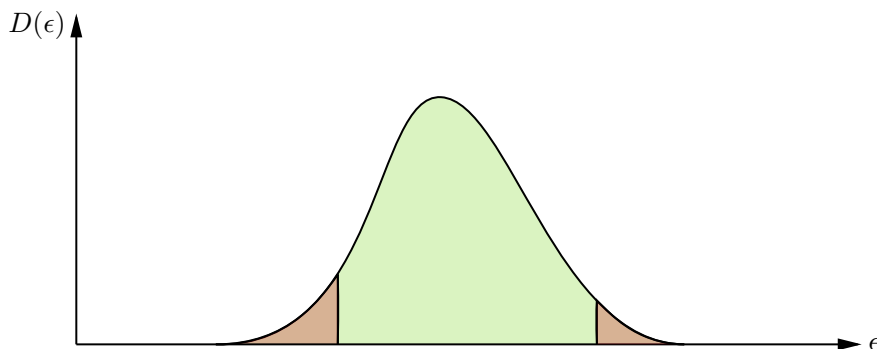


图 9.1: 一个无序系统的能带结构，延展态电子和局域态电子分别使用绿色和红色表示，两者的界限即为迁移率边

现在费米面的位置将决定系统的导电性。如果费米面的位置位于较低的局域态能带中，虽然此时有能带半填充，但是由于是局域态，系统是绝缘体。在实空间中，这等效于电子都被束缚在杂质周围。费米面进入延展态能带后，所有能够被占据的杂质轨道都被占据了，电子别无他法，必须有一部分是延展态的。在实空间中，这就好像大水漫灌将小水坑都灌满而仍然有不少剩余的可以四处流动。当费米面进入较高的局域态能带时，延展态能带被占满，系统又一次变成绝缘体。在实空间中，这就好像水缸被填满了，只留下几个气泡，然而气泡被定域在一些点附近无法移动。改变费米面高度，或者更加容易的，改变 W/t ，因而能够产生一个延展-局域相变，或者说是金属-绝缘体相变。这个相变通常称为Anderson转变。 W/t 增大只会让带尾局域态越来越多，最后延展态完全消失，因此无论费米面一开始在哪里， W/t 足够大时都会发现金属-绝缘体相变，这正是前面通过分析 $W/t = 0$ 和 $W/t \rightarrow \infty$ 而得出的结论。

9.1.2 弱局域化

上一节并没有解释为何当无序越来越强时到底发生了什么才导致局域化。本节将提供一些物理图像。

单杂质下格林函数的微扰计算

(9.1)是自由理论，因此是严格可解的。通常我们会计算所谓的无序平均，即对每个可能的 $\{\epsilon_i\}$ 构型，计算出某个物理量 O （此处不是指可观察量算符，而是真的实行观测测得的量，如 ϵ 给定后，某算符的期望值、方差，格林函数等），然后求出平均值 \bar{O} 。对一个热力学系统，如果 O 是关于系统全局的，则 O 和无序平均 \bar{O} 大几率是非常接近的，这是等概率原理的空间版本。

我们要计算两点格林函数。两点关联函数为

$$G_{ij}^{-1}(i\omega_n) = \omega_n + t_{ij} - \epsilon_i \delta_{ij}. \quad (9.3)$$

这里我们去掉了松原格林函数前面的负号，以便于书写。我们现在对 ϵ_i 做微扰展开（虽然是自由系统但微扰展开总是合法的），就有

$$\begin{aligned} G_{ij}(i\omega_n) &= \frac{1}{i\omega_n + t_{ij}} + \frac{1}{i\omega_n + t_{ik}} (-\epsilon_k \delta_{kl}) \frac{1}{i\omega_n + t_{lj}} + \cdots \\ &= \frac{1}{i\omega_n - \hat{T}} + \frac{1}{i\omega_n - \hat{T}} \hat{\epsilon} \frac{1}{i\omega_n - \hat{T}} + \cdots, \end{aligned} \quad (9.4)$$

这里我们已经引入了动能算符 \hat{T} 和微扰 $\hat{\epsilon}$ 。当然，这是非常直观的：电子和杂质散射了一次，两次，……用格林函数计算出各种可以观测的物理量，对 $\{\epsilon_i\}$ 求平均即可。(9.4)的第一项在杂质构型平均中保留，第二项消失，因为出现了奇数个 $\hat{\epsilon}$ ，第三项保留，为

$$\frac{1}{i\omega_n - \hat{T}} \hat{\epsilon} \frac{1}{i\omega_n - \hat{T}} \hat{\epsilon} \frac{1}{i\omega_n - \hat{T}} = \sum_j W^2 \frac{1}{i\omega_n - t_{ij}} \frac{1}{i\omega_n - t_{jj}} \frac{1}{i\omega_n - t_{jk}}. \quad (9.5)$$

这里，两个 $\hat{\epsilon}$ 同时出现，对它们求无序平均后就得到上式右边。考虑更高阶项，如果不做杂质平均，得到的项就是链状的费曼图：

$$\longrightarrow \times \longrightarrow \times \cdots \times \longrightarrow \times \longrightarrow,$$

由于杂质分布取高斯分布，Wick定理对杂质平均同样适用，因此将量子力学微扰论的每一项做杂质分布，图形地说就是将上图中的叉（ $\epsilon_i c_i^\dagger c_i$ 项的微扰）两两配对，并引入“相互作用”顶角（整个理论实际上当然还是自由的，并无真正的相互作用，但是 W^2 确实构成微扰）

$$\begin{array}{c} \text{---} \\ \text{---} \end{array} \begin{array}{c} \text{---} \\ \text{---} \end{array} = W^2 \delta_{jj'}. \quad (9.6)$$

直观地看，这是说电子运动路径如果经过它曾经经过的某一点，那么这个路径的权重就很大。这样(9.5)可以直观地写成

$$\begin{array}{c} \text{---} \\ \text{---} \end{array} \begin{array}{c} \text{---} \\ \text{---} \end{array} = W^2 \delta_{jj'} \frac{1}{i\omega_n - t_{ij}} \frac{1}{i\omega_n - t_{jj}} \frac{1}{i\omega_n - t_{j'k}}. \quad (9.7)$$

我们也可以定义动量空间中的费曼规则。计算两点关联函数的二阶微扰 $G_{ik}^{(2)}$ 的傅里叶变换，有

$$\begin{aligned} G^{(2)}(\mathbf{k}, i\omega_n) &= \frac{1}{N} \sum_{i,l} e^{-i\mathbf{k} \cdot (\mathbf{R}_i - \mathbf{R}_l)} \sum_j G_{ij}^0 G_{jj}^0 G_{jl}^0 W^2 \\ &= \frac{1}{N} W^2 \sum_{\mathbf{k}'} G_{\mathbf{k}}^0 G_{\mathbf{k}'}^0 G_{\mathbf{k}}^0, \end{aligned}$$

第一行的 $1/N$ 因子是我们同时对 i 和 j 求和所致；第二个等号来自

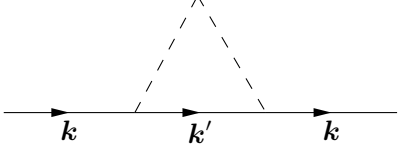
$$G_{jj}^0 = \frac{1}{N} \sum_{\mathbf{k}} G_{\mathbf{k}}^0.$$

因此动量空间的顶角就是

$$\begin{array}{c} \text{---} \\ \text{---} \end{array} \begin{array}{c} \text{---} \\ \text{---} \end{array} = \frac{1}{N} W^2 \delta_{\mathbf{k}\mathbf{k}'}. \quad (9.8)$$

可以看到该顶角函数要求入射、出射动量守恒，但是并不保证 \mathbf{k} 顶点和 \mathbf{k}' 顶点之间的传播子动量守恒。这是意料之中的，因为坐标空间中的费曼规则要求某两次散射发生在同一个地点，这个地点上的动量守恒条件实际上就是出入动量相等，但是不保证“内部”的动量和出入动量完全一样。虽然杂质破缺了空间平移对称性，但是因为杂质平均后空间平移对称性仍然成立，某种动量守恒还是成立的。此外也可以注意到，(9.8)中有一个因子 $1/N$ ，

虽然杂质造成的微扰只含有两个场算符，按理说切换到动量空间后不应该有任何含有 N 的因子；然而由于(9.8)实际上是两个 ϵ 乘在一起求平均的结果，它等效地相当于一个二粒子相互作用顶角，因此有因子 $1/N$ 是不奇怪的。于是动量空间中的(9.5)就是



$$= \frac{1}{N} W^2 \sum_{k'} G_{\mathbf{k}}^0 G_{\mathbf{k}'}^0 G_{\mathbf{k}}^0. \quad (9.9)$$

将无穷多个图加在一起实际上给出的就是自能修正。“一圈图”（虽然实际上是树图，因为虚线不提供传播子，也就没有圈图积分）的自能修正是

$$\overline{G(\mathbf{k}, i\omega_n)} = \frac{1}{i\omega_n - \xi_{\mathbf{k}} - \Sigma}, \quad (9.10)$$

其中自能就是

$$\Sigma = W^2 \frac{1}{N} \sum_{k'} \frac{1}{i\omega_n - \xi_{\mathbf{k}'}} \quad (9.11)$$

或者换成推迟格林函数中的自能，是

$$\overline{G^{\text{ret}}(\mathbf{k}, \omega)} = \frac{1}{\omega - \xi_{\mathbf{k}} - \Sigma + i0^+}, \quad \Sigma^{\text{ret}} = W^2 \frac{1}{N} \sum_{k'} \frac{1}{\omega - \xi_{\mathbf{k}'} + i0^+}. \quad (9.12)$$

这个自能修正的实部是非常平凡的，除了平移了一下化学势以外什么也没有做，而其虚部

$$\text{Im } \Sigma^{\text{ret}} = -\pi W^2 \frac{1}{N} \sum_{k'} \delta(\omega - \xi_{\mathbf{k}'}) = -\pi W^2 V_{\text{u.c.}} N(0). \quad (9.13)$$

则给出了电子的寿命，实际上最好说是电子的动量模式的寿命，即电子运动着运动着就被散射到别的动量模式上了。通过计算电子寿命可以得到体系的电阻。

随机游走图像

系统中总是有大量热涨落，因此电子在长距离运动时会出现退相干。设 l 为电子的相干长度，则参数

$$\gamma = \frac{1}{\pi k_F l} \quad (9.14)$$

表征了长距离相干性。在 γ 非常小时，两个杂质的间距远大于电子相关长度，因此可以使用经典模型处理问题。这样，我们认为电子以费米速度运动，并与杂质发生弹性碰撞。于是可以认为电子在做一个随机行走，两步之间没有任何关联。这是一个完全经典的过程，两条路径之间没有任何干涉。实际上这就是Drude模型。因此，在纯粹经典的情况下，杂质增多会缩短电子的平均自由程，但是它只会让电子线性上升，不会让系统一下子变成绝缘体。

现在假定虽然 γ 很小，但相干的散射还是可能的。这样两点之间的跃迁概率就会出现超越经典模型的修正。从一个点出发经过一系列点然后又回来的路径会因为相干叠加而获得更大的概率（经典概率的两倍）。这正是我们在费曼图计算中看到的： W 项的引入会导致(9.6)，它要求电子运动经过它曾经经过的某一点。从电子的能量本征态的角度，这意味着在杂质数量较多时将会形成电子在几个杂质之间打转的模式。这种模式的电子是不能够导电的，因为它是局域态而不是延展态。

电子运动情况的空间系综平均等价于一个随机行走

$$\langle r^2 \rangle(t) = D_0 t = \frac{l^2}{\tau} t = v_F^2 \tau t.$$

在时间 t 内, 电子经典情况下可以到达的体积为

$$V(t) \sim (\sqrt{D_0 t})^d = (D_0 t)^{d/2},$$

如果其体积在一个大小为

$$\sim \lambda_F^{d-1} v_F dt$$

的圆柱体中, 那么在 t 到 $t + dt$ 时间内它就可以回到原点。这样

$$P_{a \rightarrow a} \sim \int \frac{\lambda_F^{d-1} v_F dt}{(D_0 t)^{d/2}}$$

积分下限的数量级为 τ , 因为在比这更小的时间尺度上, 电子的运动只是布朗运动, 不是较连续的扩散。积分上限为退相干时间 τ_ϕ , 它也许来自电子和热化的声子的相互作用或是电子-电子相互作用, 无论如何都涉及一些有限温的机制。

$$\frac{\delta \sigma}{\sigma_0} = -\lambda_d \begin{cases} \sqrt{\frac{\tau_\phi}{\tau}}, & d = 1, \\ \ln\left(\frac{\tau_\phi}{\tau}\right), & d = 2, \\ \sqrt{\frac{\tau}{\tau_\phi}}, & d = 3. \end{cases}$$

当 $T \rightarrow 0$ 时, $d = 1, 2$ 的情况给出了缺乏物理意义的结果: 电导率趋向负无穷。这意味着此时微扰论失效。物理地分析, 此时所有电子能量本征态都只可能是局域态, 因此电导率为零, 系统变为绝缘体。 $d = 3$ 时随着温度降低, 电导降至零, 因此同样系统变为了绝缘体。

9.1.3 Thouless论证和重整化群

由于 $d = 1, 2$ 的情况是超越微扰论的, 需要使用重整化群的理论来分析问题。记归一化无量纲电导为

$$g = \frac{G}{e^2/\hbar}, \quad (9.15)$$

其中 G 为电导, 它和材料的长度尺度 L 有关。将 L 看成重整化群参数, L 增大表示往宏观性质方向移动。这样 β 函数定义为

$$\beta(g(L)) = \frac{d \ln g}{d \ln L}. \quad (9.16)$$

对 $g \gg 1$ 的情况, 即良导体极限, 由电阻定律

$$g = \sigma / (e^2/\hbar) L^{d-2},$$

则

$$\beta(g) = d - 2.$$

实际上, 无序的存在会让电阻变大, 电导变小, 即会让 $\ln g$ 增长得没有那么快, 于是对 g 比较大的情况, 我们做一阶展开:

$$\beta(g) = d - 2 - c \frac{1}{g}, \quad (9.17)$$

其中 c 是一个大于零的常数。

对 $g \ll 1$ 的情况即绝缘体的情况，电导会随着系统尺度快速衰减，通常是指指数衰减：

$$g = g_0 e^{-L/\xi},$$

于是

$$\beta(g) = \ln\left(\frac{g}{g_0}\right). \quad (9.18)$$

实际的 β 函数应该在(9.17)和(9.18)之间。绘图可知，对 $d = 1, 2$ 的情况， β 函数始终小于零，因此在 L 增大时 g 会衰减到零。换言之，这两个情况下含无序的固体只有一个地能有效理论，就是完全的局域化。三维系统的 β 函数有一个零点 g_c ，如果一开始 $g < g_c$ ，那么 β 函数始终是负的，于是重整化之后 $g = 0$ ，得到一个绝缘体相；如果一开始 $g > g_c$ ，那么 β 函数始终是正的，重整化之后得到一个导体相。换言之 g_c 是一个相变点。

对三维情况，在相变点附加将 β 函数线性化，有

$$\beta(g) = s \left(\frac{g - g_c}{g_c} \right), \quad (9.19)$$

计算得到

$$\frac{g(L) - g_c}{g_c} = \left(\frac{L}{\xi} \right) \quad (9.20)$$

第三部分

强关联电子系统

在传统凝聚态物理的框架以外的系统是强关联的。这包括：费米液体理论仍然在某种意义上成立，然而单粒子能谱和普通的金属或绝缘体已经非常不同，如强烈的相互作用导致能带劈裂；费米液体理论自身彻底失效，系统的基本自由度不再是电子型的激发；出现费米液体或是对称性自发破缺无法解释的相和相变。需注意强烈的库伦相互作用未必导致强关联系统（例如，库伦相互作用下系统中可能仍然有稳定的“电子占据数”，而只是能量有所修正，当然，这就是费米液体），而较弱的库伦相互作用也可能导致强关联系统（如Hubbard模型在能带很窄时的情况）。

这一部分主要讨论前两种情况，即强关联系统导致系统的基本自由度和费米液体不同。

第十章 强关联效应导致的金属-绝缘体转变

10.1 杂质导致的金属局域磁矩

10.1.1 Anderson杂质模型

考虑一个无相互作用的体系，我们在其中引入一个杂质，并且假定该杂质能够将电子紧密地约束在其周围。这样一来我们就有了两套能级：一套是原本的费米液体，还有一套是一个单独的能级，处于这个能级的电子被束缚在杂质周围。需要使用晶格动量和自旋标记前者（仅考虑能量最低的能带），称为巡游电子，因为它的波函数是布洛赫波函数，并不定域；后者是定域的，只需要使用自旋即可标记后者，称为d电子（因为很多时候这个轨道是杂质的d轨道）。前者和后者可以自然地转化，即两者之间有杂化。于是描述它们的模型就是以下单杂质的Anderson模型：

$$H = \sum_{\mathbf{k}, \sigma} (\epsilon_{\mathbf{k}} - \mu) c_{\mathbf{k}\sigma}^\dagger c_{\mathbf{k}\sigma} + \sum_{\sigma} \epsilon_d c_{d\sigma}^\dagger c_{d\sigma} + \sum_{\mathbf{k}, \sigma} V_{\mathbf{k}d} c_{\mathbf{k}\sigma}^\dagger c_{d\sigma} + \text{h.c.} + U n_{d\uparrow} n_{d\downarrow}. \quad (10.1)$$

不失一般性地认为 $V_{\mathbf{k}d}$ 都是实数，如果它不是实数，那总是可以通过重新定义 $c_{\mathbf{k}d}$ （乘上一个复数因子）来让它变成实数。最后一项是因为同处于d能级的两个电子之间会有库伦排斥作用，这一项实际上是唯一的真正的相互作用，因为前三项都是二次型。请注意这一项具有自选旋转不变性，这也是合理的。与d轨道上的自旋-自旋排斥相比，巡游电子相互作用不会有显著的影响。

d电子的相互作用项意味着d轨道上出现两个电子会大大增大能量，如果费米面位于d轨道出现一个电子和d轨道出现两个电子的能量之间，那么巡游电子总会填充d轨道，而且填充一个电子，其结果就是产生杂质附近的局域磁矩。实际上，这样会导致一个低能有效理论，见第10.2节。本节则主要观察什么时候会出现一个局域磁矩，即什么时候会出现对称性自发破缺。

在展开计算之前，首先尝试做一些定性的分析。记能谱展宽为 Δ ，则由费米黄金法则，

$$\Delta \propto \frac{1}{\tau} \propto \sum_{\mathbf{k}} |V_{\mathbf{k}d}|^2 N(\epsilon_d).$$

只有 U 很大时才能够产生局域磁矩，否则d轨道可以很容易地填满。当 $U \gg \epsilon_d \gg \Delta$ 时，d电子能谱发生弱展宽，但仍然有清晰的能级，并且会有一个良定义的局域磁矩。而当 $U \gg \Delta \gg \epsilon_d$ 时能级已经很不清楚，d电子可以和费米海中大范围的电子发生相互作用，因此虽然d轨道上只有一个电子，但它会频繁地发生自旋翻转，因此不会有局域磁矩。

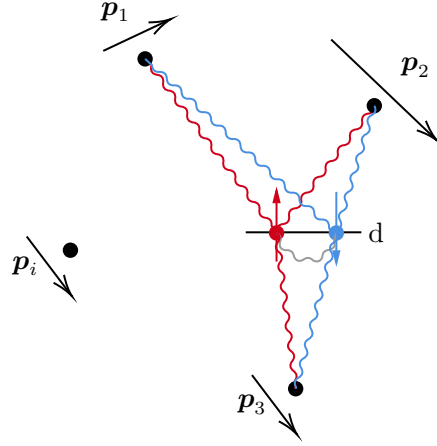


图 10.1: Anderson模型, d电子可以和巡游电子相互转换(使用红色和蓝色的波浪线表示), 不同自旋的d电子相互排斥(灰色波浪线)

10.1.2 平均场近似

相互作用项为

$$U n_{d\uparrow} n_{d\downarrow} = U c_{d\uparrow}^\dagger c_{d\uparrow} c_{d\downarrow}^\dagger c_{d\downarrow},$$

现在尝试应用平均场近似。假定体系近似自由, 我们有

$$\begin{aligned} \langle U n_{d\uparrow} n_{d\downarrow} \rangle &= U \langle c_{d\uparrow}^\dagger c_{d\uparrow} c_{d\downarrow}^\dagger c_{d\downarrow} \rangle \\ &= U (\langle c_{d\uparrow}^\dagger c_{d\uparrow} \rangle \langle c_{d\downarrow}^\dagger c_{d\downarrow} \rangle + \langle c_{d\uparrow}^\dagger c_{d\downarrow} \rangle \langle c_{d\uparrow} c_{d\downarrow}^\dagger \rangle), \end{aligned}$$

第二项如果有非零值, z 方向上的自旋旋转对称性就破缺了。确实有这样的可能, 就是系统基态有对称性自发破缺, 但这里暂时假定没有这种情况。¹这样就有

$$\langle U n_{d\uparrow} n_{d\downarrow} \rangle = U \langle c_{d\uparrow}^\dagger c_{d\uparrow} \rangle \langle c_{d\downarrow}^\dagger c_{d\downarrow} \rangle,$$

这又告诉我们, 我们有

$$\langle U n_{d\uparrow} n_{d\downarrow} \rangle = \left\langle U n_{d\uparrow} \langle n_{d\downarrow} \rangle + U \langle n | n_{d\uparrow} | n \rangle_{d\downarrow} - U \langle n_{d\uparrow} \rangle \langle n_{d\downarrow} \rangle \right\rangle,$$

那么如果相互作用哈密顿量适用平均场近似我们就有

$$U n_{d\uparrow} n_{d\downarrow} \approx U n_{d\uparrow} \langle n_{d\downarrow} \rangle + U \langle n | n_{d\uparrow} | n \rangle_{d\downarrow} - U \langle n_{d\uparrow} \rangle \langle n_{d\downarrow} \rangle. \quad (10.2)$$

当然, 这只是一种可能的平均场分解——没有理由认为这就是最理想的近似, 但实际上使用变分计算可以确定这确实是最理想的近似。忽略仅改变能量零点的常数项, 得到平均场哈密顿量

$$H_{\text{MF}} = \sum_{\mathbf{k}, \sigma} \xi_{\mathbf{k}} c_{\mathbf{k}\sigma}^\dagger c_{\mathbf{k}\sigma} + \sum_{\sigma} \epsilon_d c_{d\sigma}^\dagger c_{d\sigma} + \sum_{\mathbf{k}, \sigma} V_{\mathbf{k}d} c_{\mathbf{k}\sigma}^\dagger c_{d\sigma} + \text{h.c.} + U n_{d\uparrow} \langle n_{d\downarrow} \rangle + U \langle n | n_{d\uparrow} | n \rangle_{d\downarrow}. \quad (10.3)$$

这是一个二次型哈密顿量。(10.3)含有不确定的参数 $\langle n_{d\uparrow} \rangle$ 和 $\langle n_{d\downarrow} \rangle$ (这两个参数实际上是序参量, 它们的差给出了 z 轴上的磁矩), 但是可以将它们当成参数, 求解出 $n_{d\uparrow}$ 和 $n_{d\downarrow}$ 之后回代, 从而形成自洽方程。

¹从这里也可以看到平均场近似总是倾向于高估系统的对称性破缺, 因为我们完全可以假定 z 方向上的自旋旋转对称性真的破缺了, 从而得到一个 z 方向上自旋旋转对称性真的破缺的平均场理论。

求解(10.3), 在适当的条件上我们会看到 $\langle n_{d\uparrow} \rangle - \langle n_{d\downarrow} \rangle$ 不等于零, 即出现了一个自发磁矩。虽然出现了一个局域磁矩, 但这和我们的假定—— z 方向上自旋旋转不变——并不矛盾, 因为 z 方向上的自旋破缺的是 x 或 y 方向的自旋旋转不变性。换言之, 以上我们证明的结论是: 平均场理论下, x, y, z 三个方向上的自旋旋转不变性不可能全部保留, 保留 z 轴的自旋旋转不变性就必定破坏其它两个方向的自旋旋转不变性。

可以对(10.3)做对角化。定义

$$E_{d\sigma} = \epsilon_d + U \langle n_{d(-\sigma)} \rangle, \quad (10.4)$$

则

$$H_{MF} = \sum_{\mathbf{k}, \sigma} \xi_{\mathbf{k}} c_{\mathbf{k}\sigma}^\dagger c_{\mathbf{k}\sigma} + \sum_{\sigma} E_{d\sigma} c_{d\sigma}^\dagger c_{d\sigma} + \sum_{\mathbf{k}, \sigma} V_{\mathbf{k}d} c_{\mathbf{k}\sigma}^\dagger c_{d\sigma} + \text{h.c.} \quad (10.5)$$

下面要对角化该哈密顿量。设已有对角化形式

$$H_{MF} = \sum_{n, \sigma} \epsilon_{n\sigma} c_{n\sigma}^\dagger c_{n\sigma},$$

其中 n 是某个未知的量子数, $c_{n\sigma}$ 可以写成 $c_{\mathbf{k}\sigma}$ 和 $c_{d\sigma}$ 的么正变换, 通过要求对易关系一致(或者别的什么技巧)就得到方程组

$$\begin{aligned} \epsilon_{n\sigma} \langle \mathbf{k}\sigma | n\sigma \rangle &= \langle \mathbf{k}\sigma | n\sigma \rangle \xi_{\mathbf{k}} + \langle d\sigma | n\sigma \rangle V_{\mathbf{k}d}, \\ \epsilon_{n\sigma} \langle d\sigma | n\sigma \rangle &= \langle d\sigma | n\sigma \rangle E_{d\sigma} + \sum_{\mathbf{k}} \langle \mathbf{k}\sigma | n\sigma \rangle V_{\mathbf{k}\sigma}. \end{aligned} \quad (10.6)$$

求解此方程组, 并加入么正性条件, 就可以完成对角化。

10.1.3 平均场近似下的格林函数

由于系统自旋守恒, 两个自旋不同的算符的格林函数为零, 因此可以使用 $G_{dd, \sigma}$ 标记 d 轨道电子的格林函数, 用 $G_{\mathbf{k}d, \sigma}$ 标记从 \mathbf{k} 动量的巡游电子跃迁为 d 电子的格林函数。从一个巡游电子到另一个巡游电子的过程动量守恒(因为相当于积掉了 d 电子自由度), 用 $G_{\mathbf{k}\sigma}$ 标记这个过程的格林函数。我们并不会计算这些格林函数, 于是首先考虑 n - σ 表象下的格林函数, 然后再使用表象变换得到 d 电子或者别的什么东西的格林函数。要计算 d 点格林函数是为了计算 $\langle n_{d\uparrow} \rangle$ 和 $\langle n_{d\downarrow} \rangle$ 。

在 n - σ 表象下哈密顿量对角, 于是使用 $G_{n\sigma}$ 标记 n 表象电子的格林函数。首先考虑松原格林函数, 考虑到 n 表象下系统是自由的, 有

$$G_{n\sigma}(\omega_n) = \frac{1}{i\omega_n - \epsilon_n}, \quad (10.7)$$

请注意这里有两种 n : ω_n 中的 n 标记频率, ϵ_n 和 G 的下标中的 n 则是量子数。虚频单电子格林算符定义为

$$(i\omega_n - h_{MF})G(\omega_n) = 1,$$

使用 $G_{n\sigma}$, 单电子格林算符就是

$$G(\omega_n) = \sum_{n, \sigma} \frac{|n\sigma\rangle\langle n\sigma|}{i\omega_n - \epsilon_n} = \frac{1}{i\omega_n - h_{MF}}, \quad (10.8)$$

其中 h_{MF} 就是单体哈密顿量, 为

$$h_{MF} = \sum_{\mathbf{k}, \sigma} \xi_{\mathbf{k}} |\mathbf{k}\sigma\rangle\langle \mathbf{k}\sigma| + \sum_{\sigma} E_{d\sigma} |d\sigma\rangle\langle d\sigma| + \sum_{\mathbf{k}, \sigma} V_{\mathbf{k}d} (|\mathbf{k}\sigma\rangle\langle d\sigma| + \text{h.c.}). \quad (10.9)$$

格林算符在不同表象下的矩阵元就给出了全部的单电子格林函数（或者说二算符格林函数）。

计算(10.9)在 \mathbf{k}, d 表象下的不同矩阵元，可以得到 $G_{dd,\sigma}$ ， $G_{\mathbf{k}d,\sigma}$ 和 $G_{\mathbf{k},\sigma}$ 之间的关系。可以预期 $G_{\mathbf{k},\sigma}$ 和自由情况不会差太多，因为巡游电子远远多于d电子。d电子的格林函数相对自由电子格林函数则会有较大的修正，具体而言，是

$$(G_{dd,\sigma}(\omega_n))^{-1} = i\omega_n - E_{d\sigma} - \sum_{\mathbf{k}} \frac{V_{\mathbf{k}d}^2}{i\omega_n - \epsilon_{\mathbf{k}}}. \quad (10.10)$$

相应的推迟格林函数是

$$(G_{dd,\sigma}^{\text{ret}}(\omega))^{-1} = \omega - E_{d\sigma} - \sum_{\mathbf{k}} \frac{V_{\mathbf{k}d}^2}{\omega - \epsilon_{\mathbf{k}} + i0^+}. \quad (10.11)$$

自能修正为

$$\Sigma_{d\sigma}^{\text{ret}} = \sum_{\mathbf{k}} \frac{V_{\mathbf{k}d}^2}{\omega - \epsilon_{\mathbf{k}} + i0^+} = \frac{V}{(2\pi)^3} \int d^3\mathbf{k} \frac{V_{\mathbf{k}d}^2}{\omega - \epsilon_{\mathbf{k}} + i0^+}. \quad (10.12)$$

它来自d电子通过和巡游电子作用，间接地“自己和自己相互作用”。它的实部带来能级修正，它的虚部就是能级展宽的量级。实部是通常的柯西积分主值

$$\text{Re } \Sigma_{d\sigma}^{\text{ret}} = \frac{V}{(2\pi)^3} \text{P} \int d^3\mathbf{k} \frac{V_{\mathbf{k}d}^2}{\omega - \epsilon_{\mathbf{k}} + i0^+},$$

虚部为

$$\text{Im } \Sigma_{d\sigma}^{\text{ret}} = -\frac{V}{(2\pi)^3} \int d^3\mathbf{k} \pi \delta(\omega - \epsilon_{\mathbf{k}}) V_{\mathbf{k}d}^2.$$

我们在费米面附近工作，从而²

$$\begin{aligned} \text{Im } \Sigma_{d\sigma}^{\text{ret}} &= -\pi \langle V_{\mathbf{k}d}^2 \rangle_{\text{FS}} \frac{V}{(2\pi)^3} \int d^3\mathbf{k} \delta(\omega - \epsilon_{\mathbf{k}}) \\ &= -\pi \langle V_{\mathbf{k}d}^2 \rangle_{\text{FS}} N(\omega). \end{aligned}$$

费米子谱函数为

$$A_{d\sigma} = -\frac{1}{\pi} \text{Im } G_{d\sigma}^{\text{ret}},$$

它在 $\omega = E_{d\sigma}$ 附近有峰，其它位置接近零。由于通常能级展宽很小，在计算谱函数时假定自能中的 ω 始终取 $E_{d\sigma}$ ，同时将自能修正归入 $E_{d\sigma}$ 。（ $E_{d\sigma}$ 是做过重整化的，所以它到底取什么值根本就不知道，因此将自能实部也归入其中是合理的）以自能虚部绝对值为能级展宽量级，记作

$$\Delta = \pi \langle V_{\mathbf{k}d}^2 \rangle_{\text{FS}} N(\omega), \quad (10.13)$$

谱函数就是

$$A_{d\sigma} = \frac{1}{\pi} \frac{\Delta}{(\omega - E_{d\sigma})^2 + \Delta^2}. \quad (10.14)$$

²一种看起来更加舒服的记号是

$$\text{Im } \Sigma_{d\sigma}^{\text{ret}} = -\sum_{\mathbf{k}} \pi \delta(\omega - \epsilon_{\mathbf{k}}) V_{\mathbf{k}d}^2,$$

但是要注意，由于 $1/(\omega + i0^+)$ 的虚部是 $-\pi i \delta(\omega)$ 这件事其实只有在积分中成立，上式的求和号后的一串 δ 函数必须要做一定的平滑化处理才能让这个求和有定义，否则一串 δ 函数的求和是非常奇怪的。

10.1.4 平均场自洽计算

我们现在完成平均场计算的最后一步，即获得关于平均场序参量 $\langle n_{d\uparrow} \rangle$ 和 $\langle n_{d\downarrow} \rangle$ 的自洽方程。由谱函数的性质可以计算出

$$\langle n_{d\sigma} \rangle = \int d\omega A_{d\sigma} f(\omega),$$

上式与(10.4)和(10.14)联立，就得到自洽方程

$$\begin{aligned} \langle n_{d\sigma} \rangle &= \int d\omega \frac{1}{\pi} \frac{\Delta}{(\omega - E_{d\sigma})^2 + \Delta^2} f(\omega), \\ E_{d\sigma} &= \epsilon_d + U \langle n_{d(-\sigma)} \rangle. \end{aligned} \quad (10.15)$$

例如，在零温情况下，就得到自洽方程

$$\langle n_{d\sigma} \rangle = \frac{1}{\pi} \operatorname{arccot} \left(\frac{\epsilon_d + U \langle n_{d(-\sigma)} \rangle}{\Delta} \right). \quad (10.16)$$

这个三角超越方程难以写出解析解。为了观察它能否给出自发局域磁矩，考虑 $\epsilon_d = -U/2$ 的情况（此时实际上有电子-空穴对称性），且基态d轨道总占据数应该是1，这样

$$\langle n_{d\uparrow} \rangle + \langle n_{d\downarrow} \rangle = 1,$$

我们设

$$\langle n_{d\uparrow, \downarrow} \rangle = \frac{1}{2} \pm x,$$

若出现自发磁矩则 x 不为零。此时(10.16)为

$$x = \frac{1}{\pi} \arctan \left(\frac{U}{\Delta} x \right).$$

$x = 0$ 是平凡解；当

$$\frac{\pi \Delta}{U} < 1$$

时，有三个解，也即会有沿着 z 轴的局域磁矩。这是符合直觉的，因为当 Δ 比较大，即 V_{kd} 比较大时，巡游电子和d轨道电子不停发生相互作用，d轨道电子的自旋会不断上下翻转，因此不应该有局域磁矩。这和第10.1.1节中的分析一致。

10.2 Kondo效应

10.2.1 Anderson单杂质模型的低能有效理论

在(10.1)中 $U > |\epsilon_d| \gg V$ 时，d轨道上通常会有单个电子，从而导致一个局域磁矩。我们使用平均场近似得到了一些定性的结果，本节则讨论在此基础上的涨落。平均场使用相互作用的平均值代替它本身，但是在这个平均值上还有热涨落和量子涨落。 $U > |\epsilon_d|$ 意味着空的d轨道、半满的d轨道、全满的d轨道分得非常开，因此我们只讨论仅涉及单满的d轨道的一个低能有效模型，为此需要把空的d轨道、全满的d轨道这两个态积掉，而只保留低能子空间，即半满d轨道。

使用二阶微扰论处理这个问题，此时我们的任务是找到 H 在二阶微扰下的本征值（本征矢并不重要）。如下将Anderson模型分成两部分：

$$H = \underbrace{\sum_{\mathbf{k}, \sigma} (\epsilon_{\mathbf{k}} - \mu) c_{\mathbf{k}\sigma}^\dagger c_{\mathbf{k}\sigma} + \sum_{\sigma} \epsilon_d c_{d\sigma}^\dagger c_{d\sigma} + U n_{d\uparrow} n_{d\downarrow}}_{H_0} + \underbrace{\sum_{\mathbf{k}, \sigma} V_{kd} c_{\mathbf{k}\sigma}^\dagger c_{d\sigma} + \text{h.c.}}_{H_1}.$$

H_0 中巡游电子和d电子是完全解耦的。 H_1 会让半满的d轨道变成全满，或者让半满的d轨道变成全空，因此其一阶效应对低能有效模型没有影响。计算到二阶微扰，使用 n 标记高能自由度，使用希腊字母标记低能自由度，有

$$\langle \alpha | H_{\text{eff}}^{(2)} | \beta \rangle = \sum_n \langle \alpha | H_1 | n \rangle \langle n | H_1 | \beta \rangle \frac{1}{2} \left(\frac{1}{E_\alpha - E_n} + \frac{1}{E_\beta - E_n} \right),$$

其中等式左边的 $|\alpha\rangle$ 、 $|\beta\rangle$ 是微扰之后的本征态，右边的 $|\alpha\rangle$ 和 $|\beta\rangle$ 是微扰之前的。画费曼图可以得到两个初末态都在低能子空间中的二阶过程：³

1. 自旋为 σ' 的d电子转化为动量为 \mathbf{k}' 的巡游电子（此时d轨道空了），然后动量为 \mathbf{k} ，自旋为 σ 的巡游电子转化为d电子；
2. 动量为 \mathbf{k}' ，自旋为 σ' 的巡游电子转化为d电子（于是就有了两个d电子），自旋为 σ 的d电子转化为动量为 \mathbf{k} 的巡游电子。

实际上还有一些初末态完全一致的过程，但它们只会给哈密顿量加上一个常数，故略去。我们尝试写出这两个过程带来的修正。积掉高能自由度之后应该得到一个巡游电子-d电子之间的有效相互作用，这个相互作用的形式为⁴

$$H_{\text{eff int}} \sim c_d^\dagger c_{\mathbf{k}}^\dagger c_{\mathbf{k}} c_d,$$

于是通过计算微扰后能量修正，得到过程1对应的哈密顿量修正为

$$\sum_{\mathbf{k}, \mathbf{k}', \sigma, \sigma'} V_{\mathbf{k}'d} V_{\mathbf{k}d}^* c_{d\sigma}^\dagger c_{\mathbf{k}'\sigma'}^\dagger \frac{1}{2} \left(\frac{1}{(\epsilon_{\mathbf{k}} + \epsilon_d) - (\epsilon_{\mathbf{k}} + \epsilon_{\mathbf{k}'})} + \frac{1}{(\epsilon_{\mathbf{k}'} + \epsilon_d) - (\epsilon_{\mathbf{k}} + \epsilon_{\mathbf{k}'})} \right) c_{d\sigma'} c_{\mathbf{k}\sigma},$$

过程2对应的哈密顿量为

$$\sum_{\mathbf{k}, \mathbf{k}', \sigma, \sigma'} V_{\mathbf{k}'d} V_{\mathbf{k}d}^* c_{d\sigma'}^\dagger c_{\mathbf{k}\sigma}^\dagger \frac{1}{2} \left(\frac{1}{(\epsilon_d + \epsilon_{\mathbf{k}}) - (U + 2\epsilon_d)} + \frac{1}{(\epsilon_d + \epsilon_{\mathbf{k}'}) - (U + 2\epsilon_d)} \right) c_{d\sigma} c_{\mathbf{k}'\sigma'},$$

于是最后有效哈密顿量为

$$\begin{aligned} H_{\text{eff}} = & \sum_{\mathbf{k}, \sigma} \epsilon_{\mathbf{k}} c_{\mathbf{k}\sigma}^\dagger c_{\mathbf{k}\sigma} + \sum_{\sigma} \epsilon_d c_{d\sigma}^\dagger c_{d\sigma} \\ & + \sum_{\mathbf{k}, \mathbf{k}', \sigma, \sigma'} V_{\mathbf{k}'d} V_{\mathbf{k}d}^* c_{\mathbf{k}\sigma}^\dagger c_{\mathbf{k}'\sigma'}^\dagger c_{d\sigma'}^\dagger c_{d\sigma} \frac{1}{2} \left(\frac{1}{\epsilon_{\mathbf{k}} - \epsilon_d} + \frac{1}{\epsilon_{\mathbf{k}'} - \epsilon_d} + \frac{1}{U + \epsilon_d - \epsilon_{\mathbf{k}}} + \frac{1}{U + \epsilon_d - \epsilon_{\mathbf{k}'}} \right). \end{aligned} \quad (10.17)$$

(10.17)看起来非常复杂，但实际上通过对称性的论证可以发现它可以化简为非常简单的形式。首先d轨道电子的自旋角动量显然是

$$\mathbf{S}_d = \frac{1}{2} c_{d\alpha}^\dagger \boldsymbol{\sigma}_{\alpha\beta} c_{d\beta},$$

而巡游电子的自旋角动量在动量表象下，是

$$\mathbf{S}_{\mathbf{k}\mathbf{k}'} = \frac{1}{2} c_{\mathbf{k}\alpha}^\dagger \boldsymbol{\sigma}_{\alpha\alpha'} c_{\mathbf{k}'\alpha'}.$$

³ 由于 H_1 给出的都是二体散射，这是非连通图，但是由于这并不是在计算散射振幅，非连通图不能随意丢弃。

⁴ 需要强调会得到巡游电子-d电子相互作用，且这是二体相互作用，是因为使用微扰论时原则上应当考虑所有可能的初态 $|\alpha\rangle$ 和末态 $|\beta\rangle$ ，计算出能量修正，但这样非常繁琐。如果我们能够确定积掉高能自由度之后的有效哈密顿量中只会出现二体的巡游电子-d电子相互作用，就只需要讨论各含一个d电子和巡游电子的初末态就可以了，从而大大简化计算。

这里我们已经使用了爱因斯坦求和规则， α 和 β 标记了 \uparrow 和 \downarrow 两种自旋。有效哈密顿量(10.17)是自旋对称的，显然它的相互作用部分只能是下面的自旋标量的函数：

$$\sum_{\mathbf{k}, \mathbf{k}'} J_{\mathbf{k}\mathbf{k}'} \mathbf{S}_{\mathbf{k}\mathbf{k}'} \cdot \mathbf{S}_d,$$

而(10.17)中

$$H_I \sim c_{\mathbf{k}\alpha}^\dagger c_{\mathbf{d}\beta}^\dagger c_{\mathbf{k}'\beta} c_{\mathbf{d}\alpha},$$

因此必须有

$$H_I = \sum_{\mathbf{k}, \mathbf{k}'} J_{\mathbf{k}\mathbf{k}'} \mathbf{S}_{\mathbf{k}\mathbf{k}'} \cdot \mathbf{S}_d.$$

根据(4.101)可以计算出

$$\mathbf{S}_{\mathbf{k}\mathbf{k}'} \cdot \mathbf{S}_d = \frac{1}{2} c_{\mathbf{k}\alpha}^\dagger c_{\mathbf{k}'\beta}^\dagger c_{\mathbf{d}\beta} c_{\mathbf{d}\alpha} - \frac{1}{4} c_{\mathbf{k}\alpha}^\dagger c_{\mathbf{k}'\alpha} c_{\mathbf{d}\beta}^\dagger c_{\mathbf{d}\beta}.$$

由于我们要考虑的是低能有效理论，在其中 $c_{\mathbf{d}\beta}^\dagger c_{\mathbf{d}\beta}$ 几乎不变，因此它只会对 $\epsilon_{\mathbf{k}}$ 有一个常数修正。于是我们预期，费米面附近，有效哈密顿量应该是

$$H_{\text{eff}} = \sum_{\mathbf{k}, \sigma} \epsilon_{\mathbf{k}} c_{\mathbf{k}\sigma}^\dagger c_{\mathbf{k}\sigma} + \sum_{\sigma} \epsilon_d c_{\mathbf{d}\sigma}^\dagger c_{\mathbf{d}\sigma} + \frac{1}{2} \sum_{\mathbf{k}, \mathbf{k}'} J_{\mathbf{k}\mathbf{k}'} c_{\mathbf{k}\alpha}^\dagger c_{\mathbf{k}'\beta}^\dagger c_{\mathbf{d}\beta} c_{\mathbf{d}\alpha}. \quad (10.18)$$

事实上，在费米面附近， $\epsilon_{\mathbf{k}}$ 几乎就是化学势，于是(10.17)可以化成

$$H_{\text{eff}} = \sum_{\mathbf{k}, \sigma} \epsilon_{\mathbf{k}} c_{\mathbf{k}\sigma}^\dagger c_{\mathbf{k}\sigma} + \sum_{\sigma} \epsilon_d c_{\mathbf{d}\sigma}^\dagger c_{\mathbf{d}\sigma} + \sum_{\mathbf{k}, \mathbf{k}', \sigma, \sigma'} \left(\frac{1}{\mu - \epsilon_d} + \frac{1}{U + \epsilon_d - \mu} \right) V_{\mathbf{k}'d}^* V_{\mathbf{k}d} c_{\mathbf{k}\sigma}^\dagger c_{\mathbf{k}'\sigma'} c_{\mathbf{d}\sigma'}^\dagger c_{\mathbf{d}\sigma}.$$

这和我们通过对称性分析得到的(10.18)的形式完全一致。

实际上，在很多情况下，在坐标空间中只有两个电子足够接近才能发生相互作用，换言之坐标空间中的相互作用系数几乎是一个 δ 函数，于是动量空间中的相互作用系数随 \mathbf{k} 的变化不大，于是可以把它看成常数。这个常数是反比于系统体积 V 的，因为参与傅里叶变换的有两个 $c_{\mathbf{k}}$ ，于是会加入一个 $(1/\sqrt{V})^2$ 的因子。这样就得到一个非常简单有效理论：

$$H_{\text{eff}} = \sum_{\mathbf{k}, \sigma} \epsilon_{\mathbf{k}} c_{\mathbf{k}\sigma}^\dagger c_{\mathbf{k}\sigma} + \sum_{\sigma} \epsilon_d c_{\mathbf{d}\sigma}^\dagger c_{\mathbf{d}\sigma} + \frac{J_0}{2V} \sum_{\mathbf{k}, \mathbf{k}'} c_{\mathbf{k}\alpha}^\dagger \boldsymbol{\sigma}_{\alpha\beta} c_{\mathbf{k}'\beta} \cdot \mathbf{S}_d.$$

其中特意保留了一个 $1/2$ 因子是为了强调电子的自旋是 $1/2$ ；我们已经重新定义了单粒子能量，即不失一般性地取化学势为0，由于之后不会用到化学势，索性直接将 $\epsilon_{\mathbf{k}}$ 重定义为 $\xi_{\mathbf{k}}$ 。由于低能有效理论中d电子始终只有一个， ϵ_d 项是常数，故略去，就得到

$$H_{\text{eff}} = \sum_{\mathbf{k}, \sigma} \epsilon_{\mathbf{k}} c_{\mathbf{k}\sigma}^\dagger c_{\mathbf{k}\sigma} + \frac{J_0}{2V} \sum_{\mathbf{k}, \mathbf{k}'} c_{\mathbf{k}\alpha}^\dagger \boldsymbol{\sigma}_{\alpha\beta} c_{\mathbf{k}'\beta} \cdot \mathbf{S}_d. \quad (10.19)$$

这就是所谓的**Kondo模型**。注意Kondo模型中还是有d电子的，这个自由度没有被完全积掉（因为它提供的自旋和巡游电子的自旋之间有相互作用），但是始终只有一个d电子，它唯一可变的参数是自旋。

10.2.2 低温下电子的散射和电阻

在低温下，杂质对电子的散射会导致电阻随着温度下降而增加，这种反常性称为**Kondo效应**。正常情况下电子的散射主要来自电子热运动，因此本来应该是越热电阻率越大。照惯例将d电子在 x 和 y 方向上的自旋算符组合成升降算符：

$$S_d^\pm = S_d^x \pm iS_d^y, \quad (10.20)$$

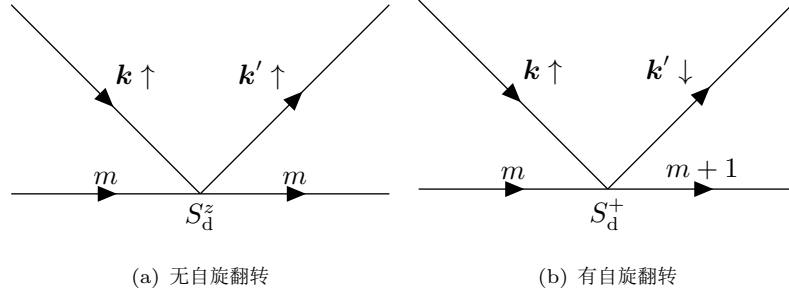


图 10.2: 一阶Kondo散射

则Kondo模型可以改写为

$$H_{\text{eff}} = \sum_{\mathbf{k}, \sigma} \epsilon_{\mathbf{k}} c_{\mathbf{k}\sigma}^\dagger c_{\mathbf{k}\sigma} + \frac{J_0}{2V} \sum_{\mathbf{k}, \mathbf{k}'} (S_d^z (c_{\mathbf{k}\uparrow}^\dagger c_{\mathbf{k}'\uparrow} - c_{\mathbf{k}\downarrow}^\dagger c_{\mathbf{k}'\downarrow}) + S_d^+ c_{\mathbf{k}\downarrow}^\dagger c_{\mathbf{k}'\uparrow} + S_d^- c_{\mathbf{k}\uparrow}^\dagger c_{\mathbf{k}'\downarrow}). \quad (10.21)$$

S_d^+ 和 S_d^- 项分别表示 z 方向上的自旋角动量从巡游电子传递给了 d 电子，或者从 d 电子传递给了巡游电子。

下面根据(10.21)计算一些散射振幅。我们需要的是跃迁概率，因此需要计算 T 矩阵的矩阵元。以下记 d 电子自旋为 m_s 。不失一般性地假定入射巡游电子的自旋为 \uparrow 。一阶过程包括两个图，一个是入射电子自旋没有发生翻转的，一个是入射电子自旋发生了翻转的，它们展示如图10.2。它们的跃迁矩阵元分别是

$$T_{\mathbf{k}\uparrow \rightarrow \mathbf{k}'\uparrow}^{(1)} = \frac{J_0 m_s}{2V}, \quad T_{\mathbf{k}\uparrow \rightarrow \mathbf{k}'\downarrow}^{(1)} = \frac{J_0}{2V} \sqrt{\frac{3}{4} - m_s(m_s + 1)}. \quad (10.22)$$

以上两式中没有出现任何温度依赖，因此仅考虑一阶过程得不到Kondo效应。这是合理的，因为一阶过程中没有传播子，而对温度的依赖是通过传播子引入的。于是仅考虑一阶过程， m_s 给定时，散射率为

$$\begin{aligned} \Gamma(m_s) &= 2\pi \sum_{\mathbf{k}'} \delta(\epsilon_{\mathbf{k}'} - \epsilon_{\mathbf{k}}) (|T_{\mathbf{k}\uparrow \rightarrow \mathbf{k}'\uparrow}^{(1)}|^2 + |T_{\mathbf{k}\uparrow \rightarrow \mathbf{k}'\downarrow}^{(1)}|^2) \\ &= \frac{\pi N(0)}{2V} J_0^2 \left(\frac{3}{4} - m_s \right). \end{aligned} \quad (10.23)$$

第二个等号实际上做了近似：我们假定 \mathbf{k} 总是出现在费米面附近，从而 $\epsilon_{\mathbf{k}}$ 就是化学势。总的散射率就是将不同的 m_s 对应的 Γ 加起来。

现在考虑二阶过程。容易验证初末自旋翻转的图有4个，不翻转的图也有4个。初末态自旋无翻转的图如图10.3所示，初末态自旋有翻转的图如图10.4所示。图10.3和图10.4中的每一个图都有一个巡游电子/空穴传播子（ d 电子传播子由于 d 电子已经被积掉，无需考虑），由于是有限温度，巡游电子传播子带有一个因子 $1 - f_q$ 而空穴传播子带有一个因子 f_q ，直观地说，就是中间过程产生的巡游电子必须被激发到一个尚未被占据的态，而中间过程产生

的空穴必须是一个已经被占据的态上的电子被打掉以后产生的。⁵ 图10.3对应的 T 矩阵元为⁶

$$T_{\mathbf{k}\uparrow \rightarrow \mathbf{k}'\uparrow}^{(2)} = \sum_{\mathbf{q}} \left(\frac{J_0}{2V} \right)^2 \frac{1}{2} \left(\frac{1}{\epsilon_{\mathbf{k}} - \epsilon_{\mathbf{q}}} + \frac{1}{\epsilon_{\mathbf{k}'} - \epsilon_{\mathbf{q}}} \right) \left(\frac{3}{4} - m_s + 2m_s f_{\mathbf{q}} \right). \quad (10.24)$$

注意到, 虽然图10.3的前两个图的因子 $f_{\mathbf{q}}$ 和 $1 - f_{\mathbf{q}}$ 相互抵消了, 后两个图由于自旋变化, 并不能抵消。这就引入了温度依赖。由于我们关注低温下的散射行为, 以下均假定 \mathbf{k} 和 \mathbf{k}' 在费米面附近, 且 $f_{\mathbf{q}}$ 是简单的阶跃函数。对 \mathbf{q} 的求和可以转化为积分, 于是

$$\begin{aligned} \frac{1}{V} \sum_{\mathbf{q}} \frac{1}{2} \left(\frac{1}{\epsilon_{\mathbf{k}} - \epsilon_{\mathbf{q}}} + \frac{1}{\epsilon_{\mathbf{k}'} - \epsilon_{\mathbf{q}}} \right) &= \int \frac{d^3 \mathbf{q}}{(2\pi)^3} m \left(\frac{1}{k^2 - q^2} + \frac{1}{k'^2 - q^2} \right) \\ &= \frac{m}{2\pi^2} \int q^2 dq \left(\frac{1}{k^2 - q^2} + \frac{1}{k'^2 - q^2} \right), \end{aligned}$$

然后我们会发现(10.24)右边最后一个括号中前两项是发散的, 但这是因为 J_0 实际上会随着 \mathbf{k} 变化而变化; 最后一项对 $T_{\mathbf{k}\uparrow \rightarrow \mathbf{k}'\uparrow}^{(2)}$ 的贡献就是

$$\begin{aligned} &\frac{J_0^2 m m_s}{4\pi^2 V} \int_0^{k_F} q^2 dq \left(\frac{1}{k^2 - q^2} + \frac{1}{k'^2 - q^2} \right) \\ &= \frac{J_0^2 m m_s}{4\pi^2 V} \left(-2k_F - \frac{k}{2} \ln \left| \frac{k - k_F}{k + k_F} \right| - \frac{k'}{2} \ln \left| \frac{k' - k_F}{k' + k_F} \right| \right), \end{aligned}$$

严格计算所有东西太过繁琐, 我们就做一个简单的数量级估计。由于是低温极限, $\epsilon_{\mathbf{k}}$ 和 $\epsilon_{\mathbf{k}'}$ 分布在费米面附近一个厚度大体上正比于 T 的区域内, 则

$$\left| \frac{k^2}{2m} - \frac{k_F^2}{2m} \right| \sim T,$$

于是(10.24)中的最后一项的贡献就是

$$\begin{aligned} &\sim -\frac{J_0^2 m m_s}{4\pi^2 V} k_F \ln \left(\frac{mT}{2k_F^2} \right) \\ &\sim N(0) \ln \left(\frac{T_F}{T} \right). \end{aligned}$$

在低温下上式对数发散, 因此除此以外的 $T_{\mathbf{k}\uparrow \rightarrow \mathbf{k}'\uparrow}^{(2)}$ 的部分都可以忽略。类似地可以计算出

$$T_{\mathbf{k}\uparrow \rightarrow \mathbf{k}'\downarrow}^{(2)} = \sum_{\mathbf{q}} \left(\frac{J_0}{2V} \right)^2 \frac{1}{2} \left(\frac{1}{\epsilon_{\mathbf{k}} - \epsilon_{\mathbf{q}}} + \frac{1}{\epsilon_{\mathbf{k}'} - \epsilon_{\mathbf{q}}} \right) (2m_s + 1) \sqrt{\frac{3}{4} - m_s(m_s + 1)}, \quad (10.25)$$

这里没有温度依赖。

这样, T 矩阵形如

$$T \sim T_0 \left(1 + J_0 N(0) \ln \left(\frac{T_F}{T} \right) + \dots \right),$$

从而散射率——它正比于 $|T|^2$ ——在低温下的数量级就是

$$\Gamma \sim \Gamma_0 \left(1 + 2J_0 N(0) \ln \frac{T_F}{T} + \dots \right). \quad (10.26)$$

⁵ 如果将这里的计算当作Goldstone图, 那么将电子传播子联系上 $1 - f_{\mathbf{q}}$ 而将空穴传播子联系上 $f_{\mathbf{q}}$ 简直就是匪夷所思的, 因为正确的做法应该是将电子数算符和空穴数算符的乘积放在有限温度态下面计算期望值, 而显然这个值不等于电子数的期望乘以空穴数的期望。实际上, 此处我们的处理应当在非平衡场论的框架下理解: 这里我们做了玻尔兹曼输运中的分子混沌性近似而忽视了高阶关联效应而将一个四点函数用两点函数的乘积表示。所谓的 T 矩阵实际上是在计算玻尔兹曼方程的碰撞项。

⁶ 虽然在这里费米子线组成了一个圈, 但我们并没有加上一个负号。这是因为我们使用的模型中, 在无相互作用时不存在d电子和巡游电子的转化, 因此d电子和巡游电子可以当成两种不同的粒子, 从而可以认为没有出现闭合的巡游电子线, 所以不用加入负号。

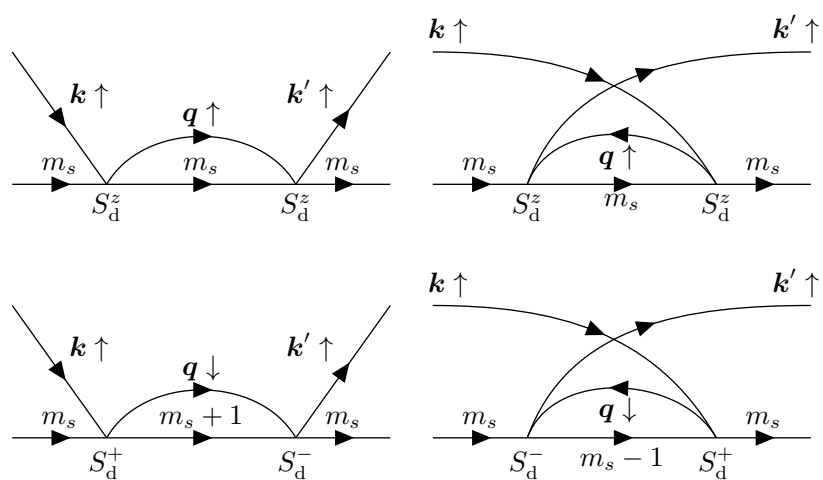


图 10.3: 初末态自旋无翻转的二阶过程

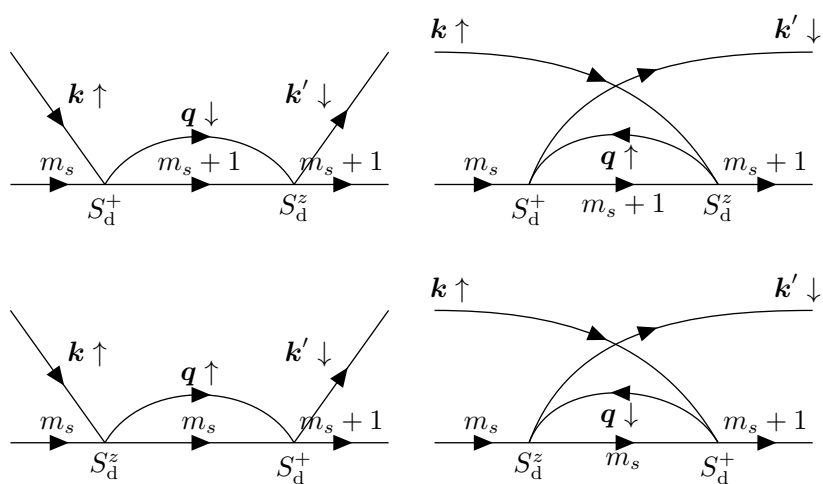


图 10.4: 初末态自旋有翻转的二阶过程

考虑到电阻率正比于散射率，我们这就导出了低温下电阻率反常对数发散的现象。需要注意的是，上式实际上意味着当温度非常低时算到二阶的微扰论是错误的，因为此时散射概率会超过1。会出现这种现象的原因在于当温度变小时有效的 J_0^{eff} 实际上会变大，从而破坏微扰论的基础。显然，只要

$$J_0 N(0) \ln \frac{T_F}{T} \sim 1, \quad (10.27)$$

那么微扰论就不再适用。对应的温度尺度

$$T_K = T_F e^{-\frac{4}{3J_0 N(0)}} \quad (10.28)$$

称为**Kondo温度**（3/4的因子是更严格计算导致的）。当温度大于Kondo温度时微扰论适用，而对小于Kondo温度的系统，d电子和巡游电子强烈的自旋-自旋相互作用会导致一个巡游电子和d电子紧密地结合起来，形成一个自旋单态，从而屏蔽d电子对巡游电子的扰动，这就是所谓的**Kondo屏蔽**。如果没有Kondo屏蔽，巡游电子和d电子可以有“自旋相同，空间回避”的效应，从而d电子对巡游电子也许没有什么散射，但是Kondo屏蔽发生后，这种效应不复存在，从而巡游电子被高效地散射。

10.2.3 重整化群

上一节中我们提到，随着温度下降，散射会增强。一个完整的包含温度的非平衡理论是不容易建立的。不过，由于温度决定了体系中的载流子（在Kondo模型中就是巡游电子）的动量分布，我们可以粗略地认为温度给出了我们关心的电子的能标。只要我们保证 $T \propto D$ ，或者更加确切地说，只要我们保证在对数意义下

$$\frac{T}{T_F} \sim \frac{D}{D_0}, \quad (10.29)$$

我们可以积掉高能电子而写下一个只关于满足 $-D < \epsilon - \epsilon_F < D$ 的电子的低能有效理论，而温度 T 下的散射率可以使用这个截断为 D 的低能有效理论中的散射率计算。在温度非常低从而 D 也非常低的情况下，这个低能有效理论应当不会特别复杂，从而粗略地说，低温下的散射率和这个低能有效理论中的有效相互作用常数的平方数量级是相当的。

所以我们要做的事情就是变动 D 而观察有效相互作用强度 J 的变动。我们并不做完整的重整化群计算，而是采用一个称为**Poor man's scaling**的方法。我们将始终假定相互作用非常弱（这其实是没有道理的，因为我们想要的是相互作用如何变强，那么肯定要涉及相互作用比较强的区域；这就是poor man一词的来历——它只是一个非常粗略的结果），从而可以做微扰的计算：我们将只计算具有如下形式的二阶费曼图：一个低能电子经过一次d电子散射，产生一个高能中间电子，然后再经过一次d电子散射产生一个低能电子。

(10.21)中，三个相互作用项都具有完全一样的耦合常数，这是因为我们导出(10.21)使用的Anderson模型是各向同性的。本节讨论一个稍微一般一些的模型，即如下的各向异性Kondo模型：

$$H_{\text{eff}} = \sum_{\mathbf{k}, \sigma} \epsilon_{\mathbf{k}} c_{\mathbf{k}\sigma}^\dagger c_{\mathbf{k}\sigma} + \frac{1}{2V} \sum_{\mathbf{k}, \mathbf{k}'} (J_z S_d^z (c_{\mathbf{k}\uparrow}^\dagger c_{\mathbf{k}'\uparrow} - c_{\mathbf{k}\downarrow}^\dagger c_{\mathbf{k}'\downarrow}) + J_+ S_d^+ c_{\mathbf{k}\downarrow}^\dagger c_{\mathbf{k}'\uparrow} + J_- S_d^- c_{\mathbf{k}\uparrow}^\dagger c_{\mathbf{k}'\downarrow}), \quad (10.30)$$

其中我们强行要求

$$J_+ = J_- = J_\pm, \quad (10.31)$$

从而保留有一个自旋的 \mathbb{Z}_2 对称性，即将自旋上下颠倒则什么也没有发生。于是在重整化群作用下， $c_{\uparrow}^\dagger c_{\uparrow}$ 项和 $c_{\downarrow}^\dagger c_{\downarrow}$ 项的系数始终只差了一个负号，从而在讨论耦合常数变化时可以只讨论入射电子自旋向上的情况。

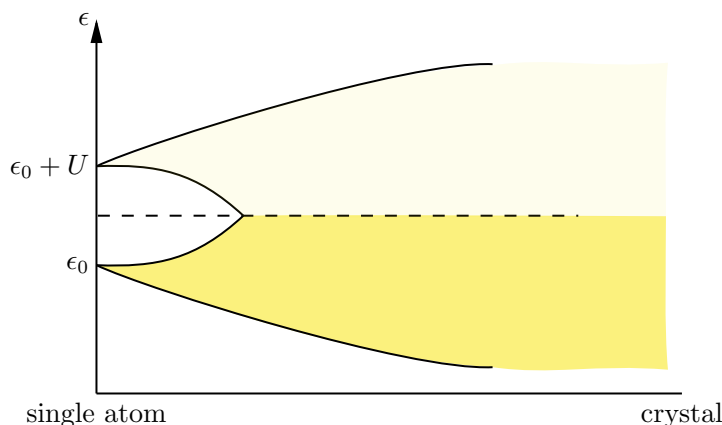


图 10.5: 电子-电子排斥导致的能带撕裂, 黄色区域表示有能级的区域

现在开始积掉从 $D - |\delta D|$ 到 D 这么多的能带中的电子自由度, 以及从 $-D$ 到 $-D + |\delta D|$ 这么多的能带中的电子自由度。我们考虑 J_z 的变化情况, δJ_z 对应所有入射、出射电子自旋均向上的图之和, 从而我们有

如果我们以(10.21)为起点做重整化群计算, 那么无论参数如何跑动, J_z , J_+ 和 J_- 总是相等的。于是参数跑动就是

$$\frac{dJ}{d \ln D} = N(0)J^2. \quad (10.32)$$

这意味着随着我们降低能标, 有效相互作用强度会变大。

我们先前不知道Kondo模型(10.19)中, 一个低能电子的行为是什么样的: 因为有大量的高能电子自由度, 即使体系中没有高能电子, 它们作为内线(所谓的虚粒子)也会对低能电子的行为产生很大的影响。重整化群方程(10.32)则告诉我们, 在一个只包含低能电子, 而这些低能电子的行为和(10.19)中的低能电子一致的理论中, 我们应当预期看到很大的电阻, 因为巡游电子和局域磁矩的相互作用在考虑虚的高能电子态的修正以后被大大增强了。如前所述, 低温时的散射率大体上可以用截断为 D 的低能有效理论中的相互作用强度的平方计算出来(它当然也可以用原始的Kondo模型计算出来, 可原始的Kondo模型因为包含了太多高能电子态而太复杂了), 所以我们确实应该预期低温下散射率大大增强, 即低温下电阻大大增强。

10.3 Mott绝缘体与Hubbard模型

能带较窄的系统与其说和金属比较相似, 不如说和孤立原子比较相似。设系统中电子浓度使得每个原子周围正好可以有一个电子, 则在能带很窄的情况下, 由于电子-电子排斥, 最稳定的状态显然是电子定域在各个原子周围。这种情况下系统是不导电的。如果还是相信库伦相互作用只是修正了能带云云, 那将会因为此时的能带是半满的而错误地以为系统导电。

以上物理图像意味着, 对局域电子, on-site interaction能够造成与巡游电子的能带理论非常不同的行为, 形象地说就是“能带撕裂”: 如图10.5所示, 在原子间隔较大, 单个原子可以看成孤立原子的极限下(图10.5最左边), 由于on-site repulsion, 可以认为同一个轨道上的两个电子的能量因为库伦排斥而取消简并, 表示为 ϵ_0 和 $\epsilon_0 + U$ 两个能级。由于晶体中有 N 个原子, 这两个能级均高度简并, 各有 N 重简并。随着原子间距拉近, 原子中的电子感受到由弱到强的周期势场, 能级取消简并而形成能带(图10.5中体现为 ϵ_0 和 $\epsilon_0 + U$ 均“变

宽了”)。在原子间距不很近时, 有两条能带, 中间隔了一个能隙。如果系统中共有 N 个电子, 那么从下往上填充能带, 会发现一条能带全满(图10.5中标为深黄色), 而另一条能带全空(图10.5中标为浅黄色), 从而系统是绝缘体。而当两条能带发生交叠时, 系统就成了导体。因此, 一个由一价原子形成的晶体电子跃迁能力相对于on-site repulsion小时为绝缘体而电子跃迁能力相对于on-site repulsion大时为导体。

10.3.1 Hubbard模型

哈密顿量定义

我们注意到Hubbard相互作用是on-site repulsion, 从而, 如下Hubbard模型

$$H = -t \sum_{\langle i,j \rangle, \alpha} c_{i\alpha}^\dagger c_{j\alpha} - \mu N + U \sum_i n_{i\uparrow} n_{i\downarrow} \quad (10.33)$$

很可能是Mott绝缘体的一个例子。

10.3.2 Hubbard模型的行列式量子蒙特卡洛(DQMC)模拟

背景知识 10.1: 费米子系统的行列式蒙特卡洛模拟

为了和时间演化算符的通常写法保持一致, 本文默认

$$\prod_{i=1}^n a_i = a_n a_{n-1} \cdots a_1.$$

如无特殊说明, 使用 τ 表示虚时间, n, m 等表示离散的虚时间点的编号; 使用 i, j 表示格点坐标, 使用 σ 表示自旋, 使用 x, y 等表示电子在坐标表象下的量子数, 即 $x = (i, \sigma)$ 。设 \hat{c}^\dagger 表示适当表象下的费米子产生算符排成的行向量, \mathbf{s}_n 和 \mathbf{s}_τ 表示第 n 个虚时间采样点(也即, $\tau = n\Delta\tau$)处的辅助场构型, 如无特殊说明 \mathbf{s} 就表示整个辅助场时间线。

由于费米子自由度难以直接在计算机中表示(格拉斯曼变量本质上是算符), 我们需要使用一个玻色场来等效处理它。一般来说费米子的哈密顿量可以写成自由哈密顿量(通常是某个动能项)加上四次型相互作用哈密顿量(即只有二体相互作用, 这是合理的, 因为基本上固体理论中的相互作用几乎总是来自库伦相互作用)的情况。

虚时间路径积分实际上就是要计算一个 e 指数矩阵的迹。

使用Trotter-Suzuki近似, 我们如下将虚时间路径积分写成自由部分和相互作用部分分离的形式:

$$e^{-\beta H} = \prod_{n=1}^m e^{-\Delta\tau H_1} e^{-\Delta\tau H_0}.$$

这里用到了Trotter-Suzuki近似, 这个近似的误差控制在 $\Delta\tau^2$ 量级。实际上也可以使用不同的Trotter分解顺序, 比如说把自由哈密顿量写在前面, 而把相互作用哈密顿量写在后面。如果没有相互作用哈密顿量, 则能够将上式写成

$$H = H_0 = \sum_{i,j} c_i^\dagger A_{ij} c_j$$

的形式, 那么这就比较容易, 因为

$$\text{tr}(e^{-\sum_{i,j} c_i^\dagger A_{ij} c_j}) = \det(1 + e^{-\mathbf{A}}). \quad (10.34)$$

很容易通过对角化验证上式。实际上, 更加一般的, 我们有

$$\text{tr}(e^{-\sum_{i,j} c_i^\dagger A_{ij} c_j} e^{-\sum_{i,j} c_i^\dagger B_{ij} c_j} \dots) = \det(1 + e^{-\mathbf{A}} e^{-\mathbf{B}} \dots), \quad (10.35)$$

甚至更一般的情况。总之, 自由费米子哈密顿量的路径积分可以很容易地将费米子算符积掉, 留下一个(可以使用标准的线性代数方法计算的)行列式。

实际的哈密顿量为

$$H = H_0 + H_1,$$

我们就需要设法将 H_1 转化为单粒子算符的形式, 也就是说要引入一个辅助场, 让费米子之间的相互作用等效为费米子和这个辅助场的相互作用。可以选择适当的Hubbard-Stratonovich参量, 使用离散H-S变换引入这个辅助场, 然后积掉费米子自由度而剩下辅助场自由度。辅助场自由度是一整个辅助场世界线, 也就是说如果有 d 个空间维度, 那么辅助场世界线就有 $d+1$ 个维度。这是量子统计的普遍特征: d 维的量子系统等价于 $d+1$ 维的经典系统, 多出来的一个维度是(有限大小的)虚时间。最后就使用关于一系列辅助场构型的行列式之和写出了虚时间路径积分。由于辅助场的形式往往需要按照相互作用哈密顿量的形式确定, 目前只是指出辅助场的存在性, 而不具体讨论怎么做H-S变换。

虚时间路径积分可以写成

$$e^{-\beta H} = \sum_{\mathbf{s}} C(\mathbf{s}) \prod_{n=1}^m e^{H_1(\mathbf{s}_n)} e^{-\Delta\tau H_0}, \quad (10.36)$$

应注意其中 $H_1(\mathbf{s}_n)$ 和之前定义的 H_1 未必相等, 它现在是积掉了费米子自由度之后的哈密顿量。由于 H_0 是二次型, 而由H-S变换的性质, H_1 也是二次型, 则可以设

$$H_0 = c^\dagger \mathbf{h}_0 c, \quad H_1 = c^\dagger \mathbf{h}_1 c, \quad (10.37)$$

其中 \mathbf{h}_1 和 \mathbf{h}_0 是系数矩阵。为了简写我们显式地引入虚时间演化算符(由于虚时间演化算符含有全部动力学, 我们这里采取的是虚时间的薛定谔绘景: 不需要考虑算符的变动)

$$U_{\mathbf{s}}(\tau_2, \tau_1) = \prod_{n=n_1+1}^{n_2} e^{H_1(\mathbf{s}_n)} e^{-\Delta\tau H_0}, \quad \mathbf{B}_{\mathbf{s}}(\tau_2, \tau_1) = \prod_{n=n_1+1}^{n_2} \underbrace{e^{\mathbf{h}_1(\mathbf{s}_n)} e^{-\Delta\tau \mathbf{h}_0}}_{\mathbf{B}_{\mathbf{s}}(\tau_1+\Delta\tau, \tau_1)}, \quad (10.38)$$

使用这些记号并考虑到(10.36),

$$\text{tr}(e^{-\beta H}) = \sum_{\mathbf{s}} C(\mathbf{s}) \text{tr} U_{\mathbf{s}}(\beta, 0) = \sum_{\mathbf{s}} C(\mathbf{s}) \det(1 + \mathbf{B}_{\mathbf{s}}(\beta, 0)),$$

而

$$\begin{aligned} \text{tr}(e^{-\beta H} O) &= \sum_{\mathbf{s}} C(\mathbf{s}) \text{tr}(O e^{-\beta H}) \\ &= \sum_{\mathbf{s}} C(\mathbf{s}) \text{tr} U_{\mathbf{s}}(\beta, 0) \frac{\text{tr}(O U_{\mathbf{s}}(\beta, 0))}{\text{tr} U_{\mathbf{s}}(\beta, 0)}, \end{aligned}$$

于是我们就得到

$$\langle O \rangle = \sum_{\mathbf{s}} C(\mathbf{s}) \det(1 + \mathbf{B}_{\mathbf{s}}(\beta, 0)) \frac{\text{tr}(OU_{\mathbf{s}}(\beta, 0))}{\text{tr} U_{\mathbf{s}}(\beta, 0)}. \quad (10.39)$$

为了略微增大一般性，例如，为了计算响应函数之类的东西，我们引入完整的虚时间轴上各点的期望值：

$$\langle O \rangle(\tau) = \frac{\text{tr}(U_{\mathbf{s}}(\beta, \tau)OU_{\mathbf{s}}(\tau, 0))}{\text{tr} U_{\mathbf{s}}(\beta, 0)}. \quad (10.40)$$

原则上，我们可以直接使用以上方法现在构造一个经典的随机现象，使得

$$p(\mathbf{s}) = \frac{C(\mathbf{s}) \det(1 + \mathbf{B}_{\mathbf{s}}(\beta, 0))}{\sum_{\mathbf{s}} C(\mathbf{s}) \det(1 + \mathbf{B}_{\mathbf{s}}(\beta, 0))}, \quad \langle O \rangle_{\mathbf{s}} = \frac{\text{tr}(OU_{\mathbf{s}}(\beta, 0))}{\text{tr} U_{\mathbf{s}}(\beta, 0)}, \quad (10.41)$$

并且使用标准的有限温度平衡态场论的方法，也就是

$$\langle O \rangle_{\mathbf{s}} = \left. \frac{\partial \ln \text{tr}(e^{\eta O} U_{\mathbf{s}}(\beta, 0))}{\partial \eta} \right|_{\eta=0}, \quad (10.42)$$

计算出 $\langle O \rangle_{\mathbf{s}}$ （由于此时费米子相互作用已经积掉了，而 \mathbf{s} 又固定死了，仅有的自由度就是费米子自由度，而且只有二次型，则Wick定理适用），我们就可以使用(7.33)得到各种物理量的期望值了。

例如在 O 是单体算符时，设

$$O = c^{\dagger} \mathbf{A} c, \quad (10.43)$$

则容易看出

$$\langle O \rangle_{\mathbf{s}}(\tau) = \text{tr}((1 - (1 + \mathbf{B}_{\mathbf{s}}(\tau, 0)\mathbf{B}_{\mathbf{s}}(\beta, \tau))^{-1})\mathbf{A}). \quad (10.44)$$

一旦单体算符的期望确定了，多体算符的期望可以使用单体算符的期望算出来。

从计算单体算符的期望的方式可以马上看出，格点坐标表象下的等时格林函数就是

$$\langle c_x(\tau)c_y^{\dagger}(\tau) \rangle_{\mathbf{s}} = (1 + \mathbf{B}_{\mathbf{s}}(\tau, 0)\mathbf{B}_{\mathbf{s}}(\beta, \tau))_{xy}^{-1}, \quad (10.45)$$

相应的记其系数矩阵为 $\mathbf{G}(\tau, \tau)$ ，或者简写为 $\mathbf{G}(\tau)$ ，则有

$$\mathbf{G}(\tau, \tau) = \mathbf{G}(\tau) = (1 + \mathbf{B}_{\mathbf{s}}(\tau, 0)\mathbf{B}_{\mathbf{s}}(\beta, \tau))^{-1}. \quad (10.46)$$

需要说明的是这里其实没有用到任何关于标签 x, y 的特殊性质——它们可以是 (i, σ) ，即格点坐标和自旋的组合，或者，如果能够保证自旋旋转不变性，那么不同的自旋实际上是解耦的，则可以取 x, y 为格点坐标，而单独附加标注自旋。容易看出等时格林函数满足以下递推关系：

$$\mathbf{G}_{\mathbf{s}}(\tau + 1) = \mathbf{B}_{\mathbf{s}}(\tau + 1, \tau)\mathbf{G}_{\mathbf{s}}(\tau)\mathbf{B}_{\mathbf{s}}(\tau + 1, \tau)^{-1}. \quad (10.47)$$

进一步，我们考虑不等时格林函数。这些格林函数实际上对蒙特卡洛更新非常重要（见下一节）。设 $\tau_1 > \tau_2$ ，则

$$\begin{aligned} G_{\mathbf{s}}(x, \tau_1; y, \tau_2) &= \langle c_x(\tau_1)c_y^{\dagger}(\tau_2) \rangle \\ &= \frac{\text{tr}(U_{\mathbf{s}}(\beta, \tau_2)U_{\mathbf{s}}^{-1}(\tau_1, \tau_2)c_x U_{\mathbf{s}}(\tau_1, \tau_2)c_y^{\dagger} U_{\mathbf{s}}(\tau_2, 0))}{\text{tr} U_{\mathbf{s}}(\beta, 0)}, \end{aligned}$$

因此问题的核心就是要计算 $U_s^{-1}(\tau_1, \tau_2)c_x U_s(\tau_1, \tau_2)$ 。我们首先容易证明恒等式

$$e^{\Delta\tau c^\dagger \mathbf{A} c} c e^{-\Delta\tau c^\dagger \mathbf{A} c} = e^{-\Delta\tau \mathbf{A}} c,$$

从而容易看出

$$U_s^{-1}(\tau_1, \tau_2)c U_s(\tau_1, \tau_2) = \mathbf{B}_s(\tau_1, \tau_2)c,$$

对上式取共轭转置，并重新定义 U ，就得到

$$U_s^{-1}(\tau_1, \tau_2)c^\dagger U_s(\tau_1, \tau_2) = c^\dagger \mathbf{B}^{-1}(\tau_1, \tau_2).$$

而 \mathbf{B} 矩阵中不显含任何电子自由度，于是可以把它提到迹运算的外面，并且使用矩阵形式，就得到

$$\mathbf{G}_s(\tau_1, \tau_2) = \mathbf{B}_s(\tau_1, \tau_2)\mathbf{G}_s(\tau_2), \quad \tau_1 > \tau_2.$$

在 $\tau_2 > \tau_1$ 时可以得到类似的结果，最后

$$\mathbf{G}_s(\tau_1, \tau_2) = \begin{cases} \mathbf{B}_s(\tau_1, \tau_2)\mathbf{G}_s(\tau_2), & \tau_1 > \tau_2, \\ -(1 - \mathbf{G}_s(\tau_1))\mathbf{B}_s^{-1}(\tau_1, \tau_2), & \tau_1 < \tau_2. \end{cases} \quad (10.48)$$

原则上(10.41)就足够计算接受率了，从而可以直接用于蒙特卡洛模拟，但是这样在计算上是非常不经济的，因为需要真的计算一系列矩阵的 e 指数的积，计算上非常耗时，并且由于每个矩阵都不大，乘起来会造成很大的误差。下面我们将讨论怎么高效、精确地计算更新率。

设我们一次只更新 (i, n) 位置的 \mathbf{s} 格点，也就是，一次只更新一个固定虚时间点上的辅助场的一个格点。显然这只会影响一个 \mathbf{B} 帧，于是设更新之后的 \mathbf{B} 矩阵是

$$\mathbf{B}_{s'}(\beta, 0) = \mathbf{B}_s(\beta, \tau)(1 + \Delta^{(i)})\mathbf{B}_s(\tau, 0), \quad \tau = n\Delta\tau. \quad (10.49)$$

这样更新前后的 \mathbf{B} 矩阵只需要动一帧就够了，大大简化了计算。同时可以计算出接受率为

$$\begin{aligned} R &= \frac{\det(1 + \mathbf{B}_s(\beta, \tau)(1 + \Delta^{(i)})\mathbf{B}_s(\tau, 0))}{\det(1 + \mathbf{B}_s(\beta, 0))} \\ &= \det(1 + \Delta^{(i)}(1 - \mathbf{G}_s(\tau))). \end{aligned} \quad (10.50)$$

可以看出，接受率实际上完全由 (i, n) 的位置以及等时格林函数确定。

实际上，我们还可以一并把更新之后的格林函数使用更新之前的格林函数计算出来，这样就避免了在更新后重新计算格林函数而需要连乘 \mathbf{B} 矩阵。我们需要用到公式

$$(\mathbf{A} + \mathbf{u}\mathbf{v}^\top)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1}\mathbf{u}\mathbf{v}^\top\mathbf{A}^{-1}}{1 + \mathbf{v}^\top\mathbf{A}^{-1}\mathbf{u}}.$$

按照(10.49)，我们有

然后再应用(10.47)。

然而，即使等时格林函数也不容易计算，因为它还是涉及 \mathbf{B} 矩阵的连续相乘和取逆。为了保证精度，我们必须分析什么地方可能出现数值不稳定性，并使用适当的矩阵算法规避这些不稳定性。请注意系统的能谱一般来说是比较宽的，即 H 最大的本征值和最小的本征值可以差很多，从而每一个 $\exp(-\Delta\tau\mathbf{h}_0)\exp(\mathbf{h}_1)$ 的最大和最小的特征值可以相差很多，它们连乘会产生一个条件数特别大的、几乎是奇异的矩阵，因

此 \mathbf{B} 矩阵是非常病态的，稍有误差就会产生很大影响。考虑到计算机的精度不可能无限制提高，我们应该尽可能避免计算真的显式计算 \mathbf{B} 矩阵，而应该使用一些不那么病态的对象代替它。至少我们应该把不同的能量尺度分开来考虑，来避免条件数没完没了地增大。

于是我们先来看一看怎么把“不同的能量尺度”（或者说不同的本征值）分开。设 \mathbf{M} 是这样一个非常病态的矩阵，记其维数为 $N_p \times N_p$ ，设

$$\mathbf{M} = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_{N_p} \end{pmatrix}, \quad (10.51)$$

对它们做Gram-Schmidt正交化，得到 $\{\mathbf{v}'_i\}$ ，并设两者之间的转换矩阵由下式给出：

$$\begin{pmatrix} \mathbf{v}'_1 & \mathbf{v}'_2 & \cdots & \mathbf{v}'_{N_p} \end{pmatrix} = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_{N_p} \end{pmatrix} \mathbf{V}_R^{-1}. \quad (10.52)$$

容易验证，存在对角矩阵 \mathbf{D}_R ，使得

$$\mathbf{M} = \underbrace{\begin{pmatrix} \mathbf{v}_1/|\mathbf{v}_1| & \mathbf{v}_2/|\mathbf{v}_2| & \cdots & \mathbf{v}_{N_p}/|\mathbf{v}_{N_p}| \end{pmatrix}}_{\mathbf{U}_R} \mathbf{D}_R \mathbf{V}_R, \quad (10.53)$$

可以估计， \mathbf{U}_R 和 \mathbf{V}_R 都不非常病态，则 \mathbf{D}_R 是比较病态的——这就是说，其对角线上的元素大小相差很大。于是我们就将 \mathbf{M} 中病态的部分收集到了一个对角矩阵中，以便妥善处理（例如要尽可能少用它们做任何计算，特别是矩阵连乘）。

以上整套方法就称为行列式量子蒙特卡洛（DQMC）。在行列式蒙特卡洛模拟中等时格林函数是核心。一方面它可以用于计算我们想要的所有物理量，一方面它决定了蒙特卡洛模拟的更新：更新前后的等时格林函数有闭形式的关系，并且接受率由等时格林函数确定。由于涉及大量矩阵运算，适当的优化对高精度、高效率的DQMC是非常关键的。Hubbard-Stratonovich参量的选择同样非常重要。

DQMC的主要弱点是量子蒙特卡洛方法常常具有的符号问题。

10.3.3 Hubbard模型的严格对角化

严格对角化是少数能够非常精确地处理Hubbard模型在任意的 U 和电子数下行为的数值方法。^[17]介绍了Hubbard模型的严格对角化。Hubbard模型保持电子数目守恒，因此我们只需要分析电子数目给定的系统即可——这也是严格对角化加速运行的常用手段，即尽可能利用守恒量来减小需要对角化的矩阵的大小。此外，总自旋

$$S_z = \sum_i \frac{1}{2} (n_{i\uparrow} - n_{i\downarrow}) \quad (10.54)$$

同样是一个守恒量——相互作用项 $n_{i\uparrow}n_{i\downarrow}$ 不改变格点 i 上的电子自旋，而跃迁项虽然能够让携带某种自旋的电子从一个地方移动到另一个地方，但是不翻转自旋。

希尔伯特空间的建立

下面我们尝试构建粒子数、总自旋均给定的希尔伯特空间上的一组基。为了避免重复和遗漏，我们采用坐标和 S^z 标记的粒子数表象，即将所有的格点排成一排：

$$\begin{bmatrix} i_1 & i_2 & \cdots & i_N \end{bmatrix},$$

然后将以上序列复制一遍以引入自旋：

$$\begin{bmatrix} i_1 & i_2 & \cdots & i_N \end{bmatrix}_{\uparrow} \begin{bmatrix} i_1 & i_2 & \cdots & i_N \end{bmatrix}_{\downarrow},$$

最后，在每个格点-自旋位置上放置或为0或为1的粒子数。这样就得到了粒子数表象中的全部系统构型，从而得到了一组基底。由于费米子产生算符反对易，需要指定这些粒子被加入的顺序。我们采纳如下顺序：

$$c_{i_1\uparrow}^{\dagger} c_{i_2\uparrow}^{\dagger} \cdots c_{i_N\uparrow}^{\dagger} c_{i_1\downarrow}^{\dagger} c_{i_2\downarrow}^{\dagger} \cdots c_{i_N\downarrow}^{\dagger} |0\rangle,$$

即将同种自旋的产生算符放在一起，并且让自旋向下的产生算符出现在右边，自旋向上的产生算符出现在左边，格点编号从左往右依次增大。例如，我们有

$$\left| \begin{bmatrix} 1 & 1 & 0 & \cdots & 0 \end{bmatrix}_{\uparrow} \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 1 & 1 \end{bmatrix}_{\downarrow} \right\rangle = c_{i_1\uparrow}^{\dagger} c_{i_2\uparrow}^{\dagger} c_{i_2\downarrow}^{\dagger} c_{i_{N-1}\downarrow}^{\dagger} c_{i_N\downarrow}^{\dagger} |0\rangle. \quad (10.55)$$

我们还需要将这些态编号，以便于构建Hubbard模型有关的算符的矩阵表示。通常使用的方法就是将态

$$\begin{bmatrix} i_1 & i_2 & \cdots & i_N \end{bmatrix}$$

的编号取为 i_1, i_2, \dots, i_N 的希尔伯特空间的基的编号的Kronecker积。我们将 $n = 0$ 编号为0而将 $n = 1$ 编号为1，那么

$$\begin{bmatrix} n_{i_1} \cdots n_{i_N} \end{bmatrix}$$

应当编号为

$$n_{i_N} + 2n_{i_{N-1}} + \cdots + 2^{N-1}n_{i_1},$$

实际上就是将数字串 $n_{i_1}n_{i_2}\cdots n_{i_N}$ 当成二进制得到的数字。我们记一个坐标和 S^z 标记的粒子数表象下的基矢量的自旋朝上的部分按照这种方式得到的编号为 I_{\uparrow} ，自旋朝下的部分按照这种方式得到的编号为 I_{\downarrow} ，则按照同样的思路，整个态应该编号为

$$I = 2^N I_{\uparrow} + I_{\downarrow}. \quad (10.56)$$

我们记编号为 I 的态为 $|\phi_I\rangle$ 。通过模算数可以立刻从 I 中读取 I_{\uparrow} 和 I_{\downarrow} ，而通过统计 I_{\uparrow} 和 I_{\downarrow} 的二进制表示中的1的个数（有现成的位运算方法可以快速做到这件事）即可得到 n_{\uparrow} 和 n_{\downarrow} 。

构建哈密顿量

哈密顿量中的on-site repulsion项是很容易构造的：它在 $\{|\phi_I\rangle\}$ 表象下是对角的，其矩阵元可以通过计算矢量点乘

$$\begin{bmatrix} n_{i_1\uparrow} & \cdots & n_{i_N\uparrow} \end{bmatrix} \begin{bmatrix} n_{i_1\downarrow} \\ \vdots \\ n_{i_N\downarrow} \end{bmatrix}$$

得到。同样，有高效的二进制位运算方法能够做到这件事。

下面需要计算出算符 $c_{i\sigma}^{\dagger} c_{j\sigma}$ 在 $\{|\phi_I\rangle\}$ 下的矩阵元以构建哈密顿量中的跃迁项。用 j_1 我们有

$$c_{i\uparrow}^{\dagger} c_{j\uparrow} |\phi_I\rangle = c_{i\uparrow}^{\dagger} c_{j\uparrow} (c_{i_1\uparrow}^{\dagger} \cdots c_{i_N\uparrow}^{\dagger}) (c_{i_1\downarrow}^{\dagger} \cdots c_{i_N\downarrow}^{\dagger}) |0\rangle,$$

显然计算结果和自旋向下的部分毫无关系： $c_{j\uparrow}$ 作用在右边的态上的结果无非这么两种：

- 如果原来 j 位置上有自旋向上的电子，那么将这个电子移除；此时 $c_{j\uparrow}$ 需要和位置 j 之前的所有产生算符交换，从而产生相应的因子；
- 如果原来 j 位置上没有自旋向上的电子，则得到0。

$c_{i\uparrow}^\dagger$ 的作用是类似的。于是我们发现， $c_{i\uparrow}^\dagger c_{j\uparrow}$ 作用在 $|\phi_I\rangle$ 上的结果是：

- 如果 j 上没有自旋向上的电子，给出0；
- 如果 i 上已有自旋向上的电子，给出0；
- 否则，设格点序列 $[i_1 \ i_2 \ \dots \ i_N]$ 中 i 格点和 j 格点之间有 $n_{ij\uparrow}$ 个自旋向上的电子，则

$$c_{i\uparrow}^\dagger c_{j\uparrow} |\phi_I\rangle = (-1)^{n_{ij\uparrow}} |\phi_{I'}\rangle, \quad (10.57)$$

其中 $|\phi_{I'}\rangle$ 为将 $|\phi_I\rangle$ 中格点 j 上自旋向上的电子移动到格点 i 上后得到的态。

$c_{i\downarrow}^\dagger c_{j\downarrow}$ 的作用是完全一样的，只需要将 \uparrow 换成 \downarrow 即可。。

利用空间对称性寻找哈密顿量中的不变子空间

10.3.4 演生海森堡模型和t-J模型

我们尚没有讨论Hubbard模型的Mott绝缘体基态附近的理论是什么样的。基本的自由度肯定不是能带导体中的载流子——实际上我们将看到，半满Hubbard模型（即 $\mu = 0$ 的情况）的Mott态附近的模型是海森堡模型，稍微偏离半满的情况对应t-J模型。

先分析半满Hubbard模型。 $U \gg t$ 的情况下哈密顿量为

$$H = U \sum_i n_{i\uparrow} n_{i\downarrow},$$

由于是半满Hubbard模型，系统基态是可以很容易确定的：每个格点上要么有一个自旋向上的电子，要么有一个自旋向下的电子。所有这些状态的能量都精确是0。由于 U 非常大，一个有一个格点上有两个电子的状态一定是非常高能的，因此我们可以有把握地说：系统的低能子空间是每个格点上一个电子的那部分态矢量组成的希尔伯特空间。

现在我们将跃迁项作为微扰加入回去。如图10.6所示，低能子空间中的每一个状态都是零，而两个电子挤在一个格点上的状态能量为 U ，于是我们有等效相互作用

$$\sum_{\langle i,j \rangle, \alpha, \beta} (-t c_{i\alpha}^\dagger c_{j\alpha}) \frac{1}{0 - U} (-t c_{j\beta}^\dagger c_{i\beta}) = -\frac{t^2}{U} \sum_{\langle i,j \rangle, \alpha, \beta} c_{i\alpha}^\dagger c_{j\alpha} c_{j\beta}^\dagger c_{i\beta}.$$

注意这个等效相互作用的形式已经自动地将泡利不相容原理考虑在内了，从而图10.6(a)展示的不可能性已经被满足了。根据(4.101)有

这就是说，在半满Hubbard模型中，低能激发中只有自旋自由度而没有电荷自由度。

从Hubbard模型产生出了海森堡模型是多少有些令人惊奇的，因为Hubbard模型中来自库伦排斥的 $n_\uparrow n_\downarrow$ 项表面上看不能提供交换两个电子的相互作用通道，似乎无法产生任何自旋-自旋相互作用。然而，跃迁项的存在让我们能够通过一些中间步骤产生电子交换和交换相互作用。

$$H = -t \sum_{i, \alpha} (c_{i\alpha}^\dagger c_{j\alpha} + \text{h.c.}) + \sum_{\langle i,j \rangle} J \mathbf{S}_i \cdot \mathbf{S}_j \quad (10.58)$$

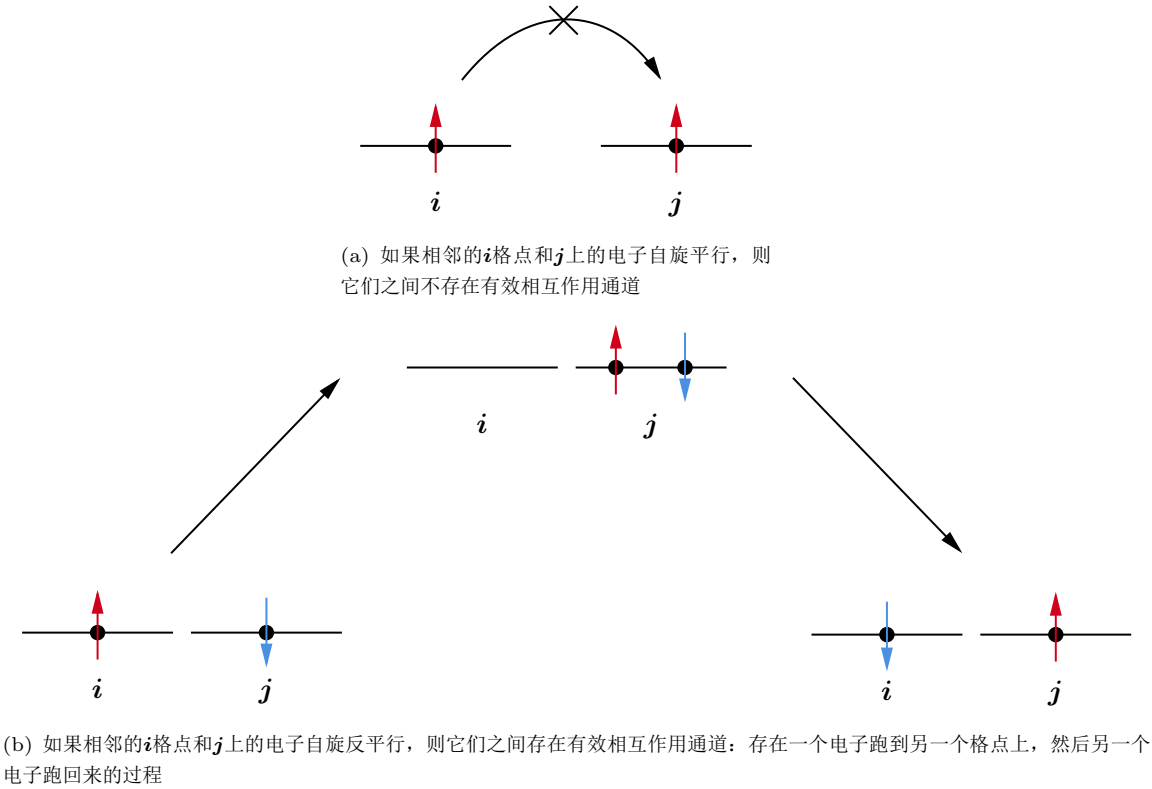


图 10.6: 紧束缚跃迁项导致的有效相互作用通道

Goodenough规则：如果电子跃迁是在两个半满轨道之间，那么就是反铁磁序，如果电子跃迁是从一个半满轨道到一个空轨道，或是从一个全满轨道到一个半满轨道，那么就是铁磁序。

10.3.5 Hubbard U 的第一性原理计算

通过constraint RPA方法可以得到等效Hubbard模型之类的东西。见^[3]。
<https://www.cond-mat.de/events/correl11/manuscripts/aryasetiawan.pdf>

10.4 Wigner晶体

第十一章 一维电子系统的玻色化

11.1 一维自由电子系统的玻色化

任何电子系统都可以展现出密度波，即 $\langle c^\dagger c \rangle$ 的长程序，这意味着电子系统的低能激发实际上未必总是“重整化之后的电子”——它甚至是玻色子。这件事我们实际上已经在第6.3.7节中提到过了，那里我们通过分析费米液体的动理学方程得到了玻色化的一种思路。

实际上，一维电子系统可以完全使用它的密度波和与之紧密相关的另一个场来描述。物理地说，这是因为一维电子系统中，一个电子不能“绕过”另一个电子，从而稍许相互作用就足够造成非常明显的密度波——或是别的什么波——的涨落。一维电子系统生来就是强关联的。

本节暂时忽略一维电子系统的自旋，或者，等价地说，假定不存在任何区分两个自旋的过程。

11.1.1 费米点附近的有效理论

考虑一个一维近独立电子系统。一维系统的动量只有一个可能的方向。因此，一维系统的费米面无非是两个**费米点**。由对称性这两个点距离动量原点的位置是相同的。动量空间的情况展示如图11.1，其中描黑的部分是一维费米“球”。

考虑费米点附近的能量，做线性近似，有

$$\xi_k = \pm v_F(k - k_F). \quad (11.1)$$

方程前面加正负号是因为色散关系是左右对称的，所以在两个费米点处的斜率互为相反数。从费米面以下到费米面以上能量总是增加的，于是斜率为正表示 $k > 0$ ，斜率为负表示 $k < 0$ 。

我们将讨论费米点附近的低能有效理论，则晶格上的布洛赫产生湮灭算符中只有晶格动量集中在 $|k_F|$ 附近的部分是有意义的。由于哈密顿量中不同动量的模式无耦合，可以直接弃去高动量模式，得到有效哈密顿量为

$$H_{\text{eff}} = \sum_{k \text{ near } \pm k_F} \xi_k c_{k\sigma}^\dagger c_{k\sigma} = \sum_{k \text{ near } k_F} v_F(k - k_F) c_{k\sigma}^\dagger c_{k\sigma} + \sum_{k \text{ near } -k_F} v_F(-k - k_F) c_{k\sigma}^\dagger c_{k\sigma},$$



图 11.1: 一维系统的动量空间，描黑的部分是一维费米球

为了简化, 记

$$p = \begin{cases} k - k_F, & k > 0, \\ k + k_F, & k < 0, \end{cases} \quad (11.2)$$

并使用 p 来标记布洛赫模式, 则

$$\begin{aligned} H_{\text{eff}} &= \sum_{|p| < \Lambda} (v_F |p| c_{Lp\sigma}^\dagger c_{Lp\sigma} + v_F |p| c_{Rp\sigma}^\dagger c_{Rp\sigma}) \\ &= \sum_{|p| < \Lambda} (v_F p c_{Lp\sigma}^\dagger c_{Lp\sigma} - v_F p c_{Rp\sigma}^\dagger c_{Rp\sigma}), \end{aligned} \quad (11.3)$$

其中 Λ 是一个截断参量, R和L分别表示对应的模式在 $k > 0$ 处 (称为右模式), 以及对应的模式在 $k < 0$ 处 (称为左模式)。费米海以外的右模式的 $p > 0$, 左模式的 $p < 0$ 。(下面会看到, 这么定义是为了让以 p 为动量的对 $c_{(L,R)\sigma}$ 的傅里叶逆变换能够给出物理意义明确的结果) 在自由哈密顿量中不同自旋之间完全解耦, 暂时只考虑一个自旋取值, 于是

$$H_{\text{eff}} = \sum_{|p| < \Lambda} (v_F p c_{Lp}^\dagger c_{Lp} - v_F p c_{Rp}^\dagger c_{Rp}). \quad (11.4)$$

若以 p 为动量, 则 c_{Rp} 在坐标空间中对应着什么? 容易发现, 它对应的坐标空间中的湮灭算符正是 $c_{n\sigma}$ 以频率 $\pm k_F$ 在空间上振荡的振幅, 即

$$c_{n\sigma} = e^{ik_F n a} c_{Rn\sigma} + e^{-ik_F n a} c_{Ln\sigma}, \quad (11.5)$$

它们在空间上是缓变的。

这里要多说一句: 我们可以看到, 在(11.4)中出现了两种彼此毫无瓜葛的电子运动模式 (左行和右行), 电子的色散关系 (近似) 是线性的。我们这里没有将自旋加入, 如果加入自旋, 我们发现系统中的电子部分具有如下的结构: 离散标签为两种自旋乘上左行/右行模式, 一共四种可能的标签; 色散关系线性, 或者说能谱是两个对顶的锥面, 费米面就在锥面顶点。当然, 这实际上就是狄拉克理论, 左行/右行模式对应两种涌现出来的手征。在我们讨论的一维电子系统是某个环的时候, “手征”一词尤其形象, 因为此时两种手征分别对应着顺时针转动的电子和逆时针转动的电子。

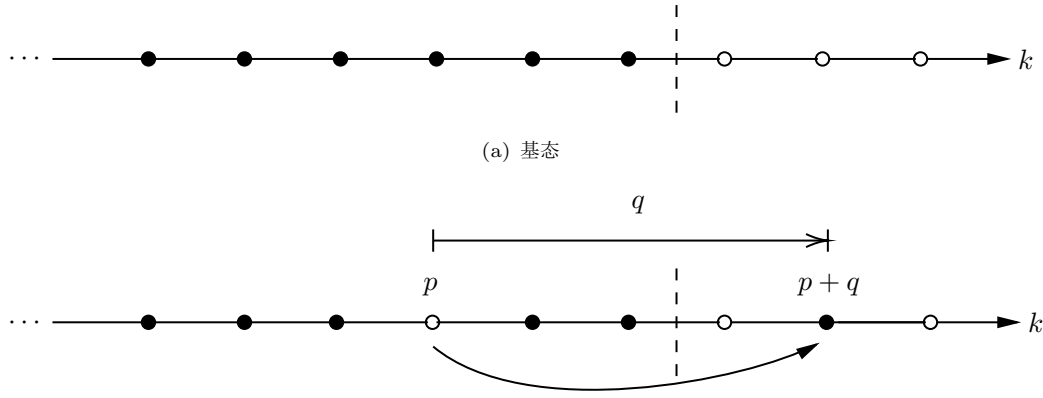
11.1.2 玻色化

首先考虑右模式的电子。按照(11.4), 对一个能量本征态, 将一个动量为 p 的右模式电子变为动量为 $p + q$ 的右模式电子, 则得到的结果仍然是一个能量本征态, 且能量上升了 $v_F q$ 。而由于粒子数守恒, 实际上任何一个能量本征态都可以通过“让费米海中的一些电子的动量增加”构造出来。这表明“将一个动量为 p 的右模式电子变为动量为 $p + q$ 的右模式电子”是一个准粒子模式, 其产生算符大体上如下:

$$b_{Rq}^\dagger \sim \sum_p c_{R(p+q)}^\dagger c_{Rp},$$

但是这样 p 可以取任何大的值, 而低能有效理论(11.4)仅仅对比较小的 p 成立。因此不出意外会产生一个发散的问题。例如, 按照(11.4)费米海中实际上有无限多的电子, 从而不受限制的对 p 求和会导致无穷大的结果。可以在 b_{Rq}^\dagger 的表达式中引入一个截断, 但是更好的做法实际上是使用正规排序, 从而将无穷大的电子数期望值去掉。我们接下来讨论怎么做到这一点。实空间的电子数密度算符为

$$\rho_R(x) = \psi_R^\dagger(x) \psi_R(x),$$



(b) 费米海中的一个动量为 p 的电子被激发到了费米海上方，动量为 $p+q$ ，让系统的能量和动量都有了一定增加，从而可以认为有一个动量为 q 的准粒子产生了

图 11.2: 右模式的玻色化

它在动量空间的形式为¹

$$\rho_R(p) = \frac{1}{\sqrt{L}} \int dx e^{-ipx} \rho_R(x) = \frac{1}{\sqrt{L}} \sum_k c_{R(k-p)}^\dagger c_{Rk}.$$

可以看到这个表达式和 b_{Rp}^\dagger 之间有线性关系，既然电子数在没有截断时发散， b_{Rp}^\dagger 必然也发散。为了消除掉发散，我们将 $\psi_R^\dagger(x)\psi_R(x)$ 的基态期望值减掉，得到下面的定义。这里使用了正规序的记号是因为这个步骤等价于以电子/空穴的产生算符为产生算符，换言之，对费米面以下的电子，要把电子湮灭算符放在左边（或者说把空穴产生算符放在左边）：

$$\rho_R(x) =: \psi_R^\dagger(x) \psi_R(x) :, \quad \rho_R(p) = \frac{1}{\sqrt{L}} \sum_k : c_{R(k-p)}^\dagger c_{Rk} :. \quad (11.6)$$

然后很容易验证，在基态（即所有电子都在费米海中的状态）下 $\rho_R(p)$ 为零，这样就避免了无穷大问题。计算对易关系得到

$$[\rho_R(-p), \rho_R(p)] = \frac{1}{L} \sum_k (c_{R(k+p)}^\dagger c_{R(k+p)} - c_{Rk}^\dagger c_{Rk}),$$

此时不能简单地做变量代换，因为求和实际上带有一个截断，做变量代换则截断也会发生变化。考虑到

$$: c_{Rk}^\dagger c_{Rk} : = c_{Rk}^\dagger c_{Rk} - \langle c_{Rk}^\dagger c_{Rk} \rangle_{\text{GS}},$$

而正规序无需截断（因为没有发散问题），可以推导出

$$[\rho_R(-p), \rho_R(p)] = \frac{1}{L} \sum_k (\langle c_{R(k+p)}^\dagger c_{R(k+p)} \rangle_{\text{GS}} - \langle c_{Rk}^\dagger c_{Rk} \rangle_{\text{GS}}),$$

等式右边的求和号中非零部分满足

$$k < 0 < k + p,$$

并做通常的求和化积分

$$\frac{1}{L} \sum_k \rightarrow \int \frac{dk}{2\pi},$$

¹它不一定是动量为 p 的电子数。

就得到

$$[\rho_{\text{R}}(-p), \rho_{\text{R}}(p)] = -\frac{p}{2\pi}.$$

而如果 $p \neq p'$, 应有

$$[\rho_{\text{R}}(-p), \rho_{\text{R}}(p')] = 0,$$

因为依照上面的步骤会得到的真空期望值一定是零。于是得到

$$[\rho_{\text{R}}(-p), \rho_{\text{R}}(p')] = -\frac{p}{2\pi} \delta_{pp'}. \quad (11.7)$$

这几乎就是一个玻色子对易关系了。缩放一下, 定义

$$b_{\text{Rp}} = \sqrt{\frac{2\pi}{p}} \rho_{\text{R}}(p), \quad b_{\text{Rp}}^\dagger = \sqrt{\frac{2\pi}{p}} \rho_{\text{R}}(-p), \quad (11.8)$$

就得到了一组玻色子产生湮灭算符, 它们描述的玻色子就是“将一个动量为 q 的右模式电子变为动量为 $p+q$ 的右模式电子”这种模式。

以上讨论的都是右模式。由于动量空间左右对称, 在变换

$$p \longrightarrow -p, \quad \text{R} \longrightarrow \text{L}$$

下物理规律不变, 于是只需要在每个关于右模式的公式中做以上代换就得到了关于左模式的公式。

容易检验可以从基态不断作用(11.8)中的产生算符而获得一组基矢量, 从而哈密顿量可以改写成

$$H = \sum_{p>0} v_{\text{F}} p \left(b_{\text{Rp}}^\dagger b_{\text{Rp}} + b_{\text{L}-p}^\dagger b_{\text{L}-p} \right), \quad (11.9)$$

其中我们忽略了一个无关紧要的能量零点变化。

11.1.3 玻色子场论

现在我们需要反过来做正则量子化: 在已经找到了一种准粒子的产生湮灭算符之后, 分析这个准粒子满足怎么样的一个场论。定义

$$\phi_{\text{R}}(p) = \frac{2\pi}{ip} \rho_{\text{R}}(p), \quad (11.10)$$

切换到坐标空间中就是

$$\rho_{\text{R}}(x) = \frac{1}{2\pi} \nabla \phi_{\text{R}}(x).$$

如果用傅里叶变换证明这个关系, 需要如下做傅里叶变换:

$$\phi_{\text{R}}(x) = \frac{1}{\sqrt{L}} \sum_p e^{-\alpha|p|/2} \phi_{\text{R}}(p) e^{ipx} = \frac{1}{\sqrt{L}} \sum_p e^{-\alpha|p|/2} \frac{2\pi}{ip} \rho_{\text{R}}(p) e^{ipx}, \quad (11.11)$$

其中 α 为一个软截断。一维系统中的傅里叶变换有时会发散, 需要手动压低。

定义

$$\begin{cases} \phi_{\text{R}} = \phi + \theta, \\ \phi_{\text{L}} = \phi - \theta, \end{cases} \quad (11.12)$$

通过傅里叶变换和 ρ 的对易关系可以得到

$$[\phi(x'), \nabla \theta(x)] = -i\pi \delta(x' - x). \quad (11.13)$$

因此 ϕ 和 $-\nabla\theta/\pi$ 是一对正则变量，可以尝试构造一个关于它们的场论。

回顾哈密顿量(11.9)，将(11.10)和(11.8)代入其中，得到

$$H = \pi v_F \sum_p (\rho_R(-p)\rho_R(p) + \rho_L(-p)\rho_L(p)).$$

将(11.11)代入上式，得到

$$H = \frac{v_F}{2\pi} \int dx ((\nabla\phi)^2 + (\nabla\theta)^2). \quad (11.14)$$

至此我们将一维自由系统的哈密顿量写成了关于连续玻色场 ϕ 和 θ 的形式，完成了玻色化。

实际上，一维系统中处理相互作用反而是比较容易的。事实上很大一类有相互作用的体系经过玻色化之后也会得到同样的表达式

$$H = \frac{v_F}{2\pi} \int dx \left(\frac{1}{K} (\nabla\phi)^2 + K (\nabla\theta)^2 \right), \quad (11.15)$$

其中 K 称为Luttinger参数。我们在分析费米液体时，曾经论证过一般来说密度-密度相互作用是重要的相互作用通道。考虑一个满足这一条件的一维电子系统，排斥性的密度-密度相互作用或者说前向散射为

$$H_{\text{int}} = \frac{U}{2} \int d^3\mathbf{r} (\rho_R(\mathbf{r}) + \rho_L(\mathbf{r}))(\rho_R(\mathbf{r}) + \rho_L(\mathbf{r})), \quad (11.16)$$

如果我们将这样一个相互作用加在二维及以上的电子系统中就得到了一个费米液体。在一维情况中，考虑到(11.10)，就得到

$$H_{\text{int}} = \frac{U}{2\pi^2} \int d^3\mathbf{r} (\nabla\phi)^2. \quad (11.17)$$

做修正

$$v'_F = v_F \sqrt{1 + \frac{U v_F}{\pi}}, \quad K = \left(1 + \frac{U v_F}{\pi} \right)^{-\frac{1}{2}}, \quad (11.18)$$

就得到了(11.15)。这也可以看成重整化导致参数跑动的一个平凡的例子。(11.15)描述的系统称为**Luttinger液体**，它不是费米液体。我们相信它不是费米液体的理由见下一节。

11.1.4 路径积分表述

对(11.15)做勒让德变换，在虚时间下做路径积分，得到

$$\begin{aligned} Z &= \int \mathcal{D}\theta \mathcal{D}\phi \exp \left(- \int d\tau \int dx \frac{1}{i} (-\nabla\theta/\pi) \partial_\tau \phi + \frac{v_F}{2\pi} \left(\frac{1}{K} (\nabla\phi)^2 + K (\nabla\theta)^2 \right) \right) \\ &= \int \mathcal{D}\theta \mathcal{D}\phi \exp \left(- \int d\tau \int dx \frac{i}{\pi} \nabla\theta \partial_\tau \phi + \frac{v_F}{2\pi} \left(\frac{1}{K} (\nabla\phi)^2 + K (\nabla\theta)^2 \right) \right). \end{aligned}$$

指数中的第一项实际上给出了一个Berry相。通过将指数中的内容配方，可以得到一个标准的高斯积分，将含 θ 的部分积掉，得到（略去了积掉 θ 之后剩下的一个常数因子）

$$\mathcal{L} = \frac{1}{2\pi v_F K} ((\partial_\tau \phi)^2 + v_F^2 (\nabla\phi)^2), \quad Z = \int \mathcal{D}\phi \exp \left(- \int d\tau \int dx \mathcal{L} \right). \quad (11.19)$$

做长度单位变换

$$\frac{x}{v_F} \longrightarrow x,$$

就得到

$$\mathcal{L} = \frac{1}{2\pi K} ((\partial_\tau \phi)^2 + (\nabla \phi)^2). \quad (11.20)$$

实际上, θ 和 ϕ 是对偶的, 因为在变换

$$\phi \longleftrightarrow \theta, \quad K \longleftrightarrow \frac{1}{K}$$

下系统不变。因此完全可以将 ϕ 积掉而留下 θ 。

考虑一个简单的例子来展示为什么Luttinger液体不是费米液体。在长波极限下, 有

$$\psi_L \sim e^{i\phi_L(x)}. \quad (11.21)$$

容易验证至少上式两边和 ψ_R^\dagger 的对易关系一致。实际上, 注意到(11.21)和KT相变中XY模型中的自旋和“涡旋场”的关系的类似之处, 我们发现在Luttinger液体中, 电子是密度涨落的(不稳定的)孤子。这样,

$$\langle \psi_R^\dagger(x) \psi_R(0) \rangle = \langle e^{i(\theta(0) + \phi(0) - \theta(x) - \phi(x))} \rangle.$$

由于是自由场, 通过Wick定理可以发现(这里的计算和研究XY模型的不同相时的操作也是类似的)

$$\langle e^A \rangle = e^{\frac{1}{2} \langle A^2 \rangle},$$

于是

$$\langle \psi_R^\dagger(x) \psi_R(0) \rangle = \exp \left(-\frac{1}{2} \langle (\theta(0) + \phi(0) - \theta(x) - \phi(x))^2 \rangle \right).$$

不失一般性地认为 $\tau = 0$, 并注意到系统在变换

$$\tau \longrightarrow -\tau, \quad \theta \longrightarrow -\theta$$

下不变, 于是

$$\langle \theta(x, 0) \phi(x, 0) \rangle = \langle -\theta(x, 0) \phi(x, 0) \rangle,$$

即

$$\langle \theta(x, 0) \phi(x, 0) \rangle = 0.$$

因此

$$\langle \psi_R^\dagger(x) \psi_R(0) \rangle = \exp \left(-\frac{1}{2} \langle (\theta(0) - \theta(x))^2 \rangle \right) \exp \left(-\frac{1}{2} \langle (\phi(0) - \phi(x))^2 \rangle \right).$$

我们考虑接近零温的情况, 此时(11.20)可以写成

$$\mathcal{L} = \sum_{\omega_n, k} \frac{K}{2\pi} \theta^*(k, \omega_n) (\omega_n^2 + k^2) \theta(k, \omega_n) = \sum_{\omega_n, k} \frac{1}{2\pi K} \phi^*(k, \omega_n) (\omega_n^2 + k^2) \phi(k, \omega_n),$$

于是

$$\langle \theta^\dagger(k, \omega_n) \theta(k, \omega_n) \rangle = \frac{\pi}{K(\omega_n^2 + k^2)}, \quad \langle \phi^\dagger(k, \omega_n) \phi(k, \omega_n) \rangle = \frac{\pi K}{\omega_n^2 + k^2}.$$

通过傅里叶逆变换得到

$$\frac{1}{2} \langle (\theta(0) - \theta(x))^2 \rangle = \frac{\pi}{k} \int \frac{dk d\theta}{(2\pi)^2} \frac{1 - e^{ikx \cos \theta}}{k} \sim \frac{1}{2K} \ln x, \quad |x| \gg 1.$$

最后就可以得到

$$\langle \psi_R^\dagger(x) \psi_R(0) \rangle \sim |x|^{-\frac{1}{2}(K + \frac{1}{K})}.$$

做傅里叶逆变换，得到 k 很小时的动量分布函数

$$\langle n_k \rangle \sim |k|^{\frac{1}{2}(K+\frac{1}{K})-1}. \quad (11.22)$$

如果 $K = 1$ ，上式右边有一个奇异性，这是费米液体的特征，因为费米面处电子数有一个跳变；但 $K \neq 1$ 时上式右边是光滑的。这就说明了有相互作用的一维电子系统不是费米液体。

11.1.5 自旋-电荷分离

做变换

$$\phi_{\text{charge}} = \frac{1}{\sqrt{2}}(\phi_{\uparrow} + \phi_{\downarrow}), \quad \phi_{\text{spin}} = \frac{1}{\sqrt{2}}(\phi_{\uparrow} - \phi_{\downarrow}), \quad (11.23)$$

注意到由于 $\phi \sim n$ ，如其下标所暗示的那样， ϕ_{charge} 和电荷密度涨落正相关而 ϕ_{spin} 和 z 方向上的自旋密度涨落正相关。因此实际上我们的体系中有两种彼此解耦的密度波，它们分别是自旋密度波和电荷密度波，两者的运行速度还不一样。这是自旋电荷分离，是分数化激发的一个典型例子，即强关联系统中作为基质的粒子携带一些标签，对称性保证了这些标签在系统的元激发中一定会保留下来，但是没什么保证这些标签会被同一种激发携带。在费米液体中自旋和电荷的标签都是被非常类似于裸的电子的准粒子携带着的，没有一个场代表“总自旋”，也没有一个场代表“总电荷”，但是在Luttinger液体中这些标签分开了，出现了 ϕ_{charge} 和 ϕ_{spin} 。我们称 ϕ_{charge} 为空穴子 (holon)，称 ϕ_{spin} 为自旋子 (spinon)。

11.2 一维t-J模型的玻色化

11.2.1 一维t-J模型和sine-Gordon模型

一维情况下，t-J模型中的格点可以使用一个整数标记，即

$$H = -t \sum_{i,\sigma} c_{i\sigma}^{\dagger} c_{i+1,\sigma} + \text{h.c.} + J \sum_i (\mathbf{S}_i \cdot \mathbf{S}_{i+1} - n_i n_{i+1}/4). \quad (11.24)$$

在做完玻色化之后，我们会发现其电荷密度波部分由一个 $U(1)$ 自由模型描述，而其自旋密度波部分由一个 $SU(2)$ 的sine-Gordon模型描述。

$$H = H_c + H_s + \frac{2g_1}{(2\pi\alpha)^2} \int dx \cos(\sqrt{8}\phi_s), \quad (11.25)$$

其中

$$H_{\nu} = \frac{1}{2\pi} \quad (11.26)$$

需注意自旋密度波部分的 $SU(2)$ 对称性并没有体现在哈密顿量的形式中，而是体现在其参数满足的约束中，只有

$$g_1 = K \quad (11.27)$$

时才具有 $SU(2)$ 对称性；这个约束正好是sine-Gordon模型的重整化群流中的两条斜对角的线。

sine-Gordon模型在 g_1 和 K 的不同安排下会落入两个相之一，其中之一的关联函数幂律衰减，对应无能隙的Luttinger液体，另一个的关联函数指数衰减，对应一个有能隙的系统。

11.2.2 超导

一维t-J模型既然是强关联系统，很容易想到，是否可能在其中寻找超导相。本节利用玻色化计算超导序参量。我们有

$$\Delta(x_i) = \psi_{i\uparrow}\psi_{i+1,\downarrow} - \psi_{i\downarrow}\psi_{i+1,\uparrow}, \quad (11.28)$$

用玻色化后得到的场表示，就有

$$\begin{aligned} \Delta(x_i) &= e^{i(\theta_{\uparrow}+\theta_{\downarrow}+\phi_{\uparrow}+\phi_{\downarrow}+2k_F x_i)} + e^{i(\theta_{\uparrow}+\theta_{\downarrow}-\phi_{\uparrow}-\phi_{\downarrow}-2k_F x_i)} + \\ &= \end{aligned} \quad (11.29)$$

通过自旋涨落解释铜基超导的高温超导仍然存在不少问题：首先，实验上，在高掺杂区域，超导相和反铁磁相并没有一起出现，使用自旋涨落解释高温超导必须首先解释这是为什么；其次，我们不能确定通过自旋涨落形成的电子配对产生的能隙是否真的超过了BCS能够提供的大小，我们也难以根据材料性质决定这种电子配对的强度，从而无法做出任何定量的预言；最后，由于材料中存在大量可能的、彼此竞争的序，自旋密度波序是不是总是能够形成实际上也是可疑的。

11.3 Hubbard模型的玻色化

第四部分

自旋系统和磁性

这一部分我们讨论自旋系统和定域磁性。自旋模型是指那些每个格点上放置了一个自旋自由度的模型（当然，这取决于格点的定义；自旋自由度当然可以定义在边上）。自旋是电子的内禀性质，似乎不应该被固定在格点上，然而在一些情况下，电子几乎总是定域在某些格点附近而不发生移动，此时系统中不存在电子位置的变化，主要的自由度就是各个格点上的自旋。自旋模型是非常常见的，例如它可以通过Hubbard模型演生出来：Hubbard模型在 U 非常大时，以动能项为微扰就能够得到一个海森堡模型。因此，自旋模型可以看成上一部分的一种特殊情况，即相互作用让系统的自由度不再是某种准电子。

对称性说明自旋-自旋相互作用通常可以取 $\mathbf{S}_i \cdot \mathbf{S}_j$ 的形式，或者也许是各向异性的 $\mathbf{S}_i \cdot \mathbf{T} \cdot \mathbf{S}_j$ 。

第十二章 铁磁性和反铁磁性的自旋模型

自旋模型是解释铁磁性和反铁磁性的重要模型。

12.1 海森堡模型

12.1.1 海森堡模型的哈密顿量

局域电子相互作用给出的海森堡模型

考虑一种晶体中的局域化电子，在一个晶胞上只有一个，一方面未配对，一方面不能远距离移动。在没有外加电磁场激励时它稳定地呆在自己的轨道上，不发生跃迁，因此作为一个低能有效理论，我们暂时只需要考虑它的位置。切换到Wannier表象下，由于这种电子是高度局域的，它自己的轨道波函数几乎就是Wannier波函数，跃迁能力很差，能带很窄。极端情况下动能项可以忽略，只需要考虑相互作用项。我们正在讨论一个单带模型，且电子跃迁能力差，则哈密顿量形式为(4.102)，且无序项不存在。由于假定每个晶胞上只有一个电子，化学势项、on-site repulsion项和密度-密度相互作用均可以直接略去，因为他们基本上是常数。因此唯一剩下的就是自旋-自旋相互作用，因此哈密顿量为

$$H = -J \sum_{\langle i,j \rangle} \mathbf{S}_i \cdot \mathbf{S}_j, \quad (12.1)$$

这里的 \mathbf{S}_i 和 \mathbf{S}_j 仍然是电子产生湮灭算符相乘得到的；然而，由于电子基本上没有跃迁，我们可以认为每个电子的 i 其实也是不怎么会变化的，因此我们可以彻底忽略电子的轨道自由度，只保留自旋自由度。这样就得到了自旋1/2的海森堡模型，如前所述，它是描述晶体中单占据、高度局域、跃迁能力差、绝缘的电子的模型。

回顾(4.102)的导出，我们发现导致海森堡模型中的自旋-自旋耦合的主要是电子之间的交换相互作用。我们后面会看到铁磁序和反铁磁序在海森堡模型中能够观察到，因此，这些磁性序的相当一部分是纯粹的量子效应的产物。

一般自旋的海森堡模型

(12.1)在两个方面可以推广：首先，可以有非最近邻的自旋-自旋相互作用；其次，实际上有相当一类体系不是自旋1/2的系统，但是仍然能够使用(12.1)形式的哈密顿量描述。前者的原因是显然的：交换相互作用（见(6.9)）不局限在最近邻电子。对后者，我们考虑某种磁性离子中有若干电子，这些电子任取两个都有交换相互作用。忽略电子的轨道跃迁，则系统状态可以使用系统中各个电子的自旋标记，通过自旋角动量的合成，等于是说可以使用各个原子的自旋来标记系统状态。因此，关于磁性离子的哈密顿量只包含一系列这些

离子的自旋算符。同一个原子上的电子彼此之间的交换相互作用无非导致哈密顿量中出现一个 \mathbf{S}_i^n 的多项式，而原子1和原子2上的电子彼此的交换相互作用形如

$$-J_{12} \sum_{i,j} \mathbf{S}_{1i} \cdot \mathbf{S}_{2j} = -J_{12} \mathbf{S}_1 \cdot \mathbf{S}_2,$$

这里 i 和 j 标记原子1和原子2上的不同电子。海森堡模型主要是关于磁性序形成的，因此我们忽略 \mathbf{S}_i^n 项。对简单晶格，我们现在写出最一般的海森堡模型：

$$H = - \sum_{i,j} J_{ij} \mathbf{S}_i \cdot \mathbf{S}_j, \quad (12.2)$$

其中 $J_{ij} = J_{ji}$ 。对复式晶格，不同子格上的 J 还可以不一样。

海森堡模型仍然有进一步扩充的空间。例如，没有什么保证哈密顿量只含有自旋的二次方；此外自旋相互作用也可以不是各向同性的。这些我们将在其它模型中引入。

12.1.2 简单晶格中磁性原子的铁磁序和自旋波

铁磁序基态

背景知识 12.1: 自旋自由度的一些性质

用 S_i^j 表示格点 i 上的自旋， $j = 1, 2, 3$ 或是 x, y, z 。通常习惯用 S^z 标记一个状态。我们有

$$[S_i^j, S_{i'}^{j'}] = i\epsilon_{jj'j''} S_i^{j''} \delta_{ii'}, \quad (12.3)$$

根据此关系可以对易所谓自旋升降算符，定义

$$S_i^+ = \frac{1}{\sqrt{2}}(S_i^x + iS_i^y), \quad S_i^- = \frac{1}{\sqrt{2}}(S_i^x - iS_i^y), \quad (12.4)$$

则有

$$[S_i^z, S_i^+] = S_i^+, \quad [S_i^z, S_i^-] = -S_i^-, \quad (12.5)$$

从而

$$\begin{aligned} S_i^+ |\cdots, m, \cdots\rangle &= \sqrt{\frac{(s+m+1)(s-m)}{2}} |m+1\rangle, \\ S_i^- |\cdots, m, \cdots\rangle &= \sqrt{\frac{(s+m)(s-m+1)}{2}} |m-1\rangle. \end{aligned} \quad (12.6)$$

特别的，对自旋1/2的情况，以 $|\uparrow\rangle$ 和 $|\downarrow\rangle$ 为基底，有

$$\mathbf{S}_i = \frac{1}{2} \boldsymbol{\sigma}_i, \quad S_i^+ = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad S_i^- = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}. \quad (12.7)$$

在 $J > 0$ 时相邻自旋的相互作用会让自旋倾向于形成铁磁态。不失一般性设其方向为 z 方向。我们将海森堡哈密顿量用自旋升降算符写出，为

$$H = -J \sum_{\langle i,j \rangle} (S_i^z S_j^z + S_i^+ S_j^- + S_i^- S_j^+). \quad (12.8)$$

铁磁态为

$$|\text{FM}\rangle = |S_1^z = S, S_2^z = S, \dots, S_N^z = S\rangle, \quad (12.9)$$

因为铁磁态中所有的 S_i^z 都是最大的，任何一个 S^- 算符作用于其上给出的都是0，因此 H 作用在 $|FM\rangle$ 上实际上就是一串 S^z 算符作用在 $|FM\rangle$ 上，因此 $|FM\rangle$ 是能量本征态，其能量为

$$E_{FM} = -J \sum_{\langle i,j \rangle} S^2 = -J \frac{N S^2 n_{\text{bond}}}{2}, \quad (12.10)$$

其中 n_{bond} 是一个格点的最近邻格点数目。由于 $J > 0$ ，这是一个非常小的值。其它用诸 S_i^z 标记的态不是能量本征态，并且它们的能量期望值要大于 E_{FM} （因为会对能量期望值有贡献的只有 $S_i^z S_j^z$ 一项）。因此， $|FM\rangle$ 实际上是基态。当然由于(12.1)具有自旋旋转不变性，自旋统一指向其它方向的态也是基态。这是为数不多能够严格求解出基态的凝聚态物理模型，虽然三维情况下整个海森堡模型的能谱是一个到现在还没有求出的问题。

自旋波

背景知识 12.2: Holstein-Primakoff变换

Holstein-Primakoff变换是一种将自旋自由度转化为玻色子自由度的局域变换。设某个自旋自由度的模长为 s ， S^z 为 m ，定义

$$n = s - m, \quad (12.11)$$

则有

$$S^+ |n\rangle = \sqrt{\frac{n(2s - n + 1)}{2}} |n - 1\rangle, \quad S^- |n\rangle = \sqrt{\frac{(n + 1)(2s - n)}{2}} |n + 1\rangle. \quad (12.12)$$

现在定义

$$S^+ = \sqrt{\frac{2s - a^\dagger a}{2}} a, \quad S^- = a^\dagger \sqrt{\frac{2s - a^\dagger a}{2}}, \quad S^z = s - a^\dagger a, \quad (12.13)$$

则能够验证，对易关系

$$[S^z, S^+] = S^+, \quad [S^z, S^-] = -S^-$$

等价于

$$[a, a^\dagger] = 1, \quad [a, a] = [a^\dagger, a^\dagger] = 0. \quad (12.14)$$

因此，可以认为

$$n = a^\dagger a, \quad (12.15)$$

即自旋偏离 n 的多少等价于某种玻色子的数目，但是要注意该玻色子的数量有上限。不同自旋自由度的自旋算符彼此对易，因此相应的 a 也对易，即 a 确实是玻色子算符。

我们现在要分析铁磁态附近的激发。由于是磁性原子，我们假定每个原子的 S^2 相比于激发态的自旋偏移足够大，从而做Holstein-Primakoff变换之后可以不考虑玻色子数目有上限这件事。对海森堡模型(12.8)做Holstein-Primakoff变换，得到

$$H = -J \sum_{\langle i,j \rangle} \left((S - a_i^\dagger a_i)(S - a_j^\dagger a_j) + \frac{1}{2} \sqrt{2S - a_i^\dagger a_i} a_i a_j^\dagger \sqrt{2S - a_j^\dagger a_j} + \frac{1}{2} a_i^\dagger \sqrt{2S - a_i^\dagger a_i} \sqrt{2S - a_j^\dagger a_j} a_j \right). \quad (12.16)$$

由于 a 玻色子对应于某格点上自旋偏离基态的多少，我们称它为自旋波。

由于只考虑低能激发态，对 $a_i^\dagger a_i$ 做小量展开，有

$$H = -\frac{1}{2}JS^2Nn_{\text{bond}} + JSn_{\text{bond}} \sum_i a_i^\dagger a_i - JS \sum_{\langle i,j \rangle} (a_i^\dagger a_j + a_j^\dagger a_i) - J \sum_{\langle i,j \rangle} \left(a_i^\dagger a_i a_j^\dagger a_j - \frac{1}{4} (a_i^\dagger a_i a_i a_j^\dagger + a_i a_j^\dagger a_j^\dagger a_j + a_i^\dagger a_i^\dagger a_i a_j + a_i^\dagger a_j^\dagger a_j a_j) \right). \quad (12.17)$$

上式中的第一项正是我们熟悉的基态能量，第二项是格点 i 上自旋偏离导致的局域能量，第三项给出不同格点的耦合。剩下的各项是自旋波模式之间的耦合。

我们先只考虑二次型部分。和求解电子紧束缚模型类似，可以直接做傅里叶变换

$$a_i = \frac{1}{\sqrt{N}} \sum_{\mathbf{k}} e^{i\mathbf{k} \cdot \mathbf{R}_i} a_{\mathbf{k}}, \quad (12.18)$$

得到

$$H = -\frac{1}{2}JS^2Nn_{\text{bond}} + \sum_{\mathbf{k}} JS \left(n_{\text{bond}} - \sum_{\delta} \cos(\mathbf{k} \cdot \delta) \right) a_{\mathbf{k}}^\dagger a_{\mathbf{k}}, \quad (12.19)$$

其中 δ 指的是任意一个连接两个最近邻格点的矢量。这样我们得到自旋波的色散关系

$$\omega_{\mathbf{k}} = JS \left(n_{\text{bond}} - \sum_{\delta} \cos(\mathbf{k} \cdot \delta) \right). \quad (12.20)$$

在 n_{bond} 较大时自旋波是有能隙的。

铁磁序可以看成自旋系统的一种“固态”，因为其中出现了对称性破缺序，该序的涨落即为自旋波，正如原子系统的固态中晶格序的涨落为声子一样。

12.1.3 反铁磁序

与铁磁的情况不同，反铁磁序不是海森堡模型的基态。直观地说，海森堡模型中同时有 S^z 和 S^x ， S^y 算符，因此实际上存在很明显的量子涨落；我们只是足够幸运，哈密顿量作用到 $|\text{FM}\rangle$ 上时有量子涨落的项正好给出零。然而，在反铁磁的情况下， E_{FM} 是非常大的能量，因此和模型的低能行为完全无关，这让我们无法利用 $|\text{FM}\rangle$ 是能量本征态这一事实。一些人猜测，海森堡模型中的量子涨落甚至可能在特定的晶格上完全破坏铁磁序！有关的情况见第十八章。

12.2 各向异性海森堡模型

12.3 伊辛模型

伊辛模型是一种极端各向异性的自旋模型，只有 z 方向的自旋存在耦合，哈密顿量为

$$H = -J \sum_{\langle i,j \rangle} S_i^z S_j^z, \quad (12.21)$$

在自旋1/2的情况下，很多时候我们会重新定义 J 而使用泡利矩阵给出哈密顿量：

$$H = -J \sum_{\langle i,j \rangle} \sigma_i^z \sigma_j^z. \quad (12.22)$$

12.3.1 经典伊辛模型

虽然各向异性通常会加大求解复杂程度，不过伊辛模型在没有外场或是外加磁场指向 z 方向时，即哈密顿量为

$$H = -J \sum_{\langle i,j \rangle} \sigma_i^z \sigma_j^z + h \sum_i \sigma_i^z \quad (12.23)$$

时是没有量子涨落的，因而称为经典伊辛模型。此时的伊辛模型是严格可解的，其基态、激发态非常清楚，甚至在一维和二维的时候，对应的配分函数都能解析计算。

12.3.2 横场伊辛模型

横场伊辛模型通过引入 x 方向的磁场，制造了量子涨落，其哈密顿量为

$$H = -J \sum_{\langle i,j \rangle} \sigma_i^z \sigma_j^x + h \sum_i \sigma_i^z. \quad (12.24)$$

第十三章 自旋链

一维的自旋模型——所谓自旋链——经常表现出一些独特的性质，因此值得单独拿出来讨论。

13.1 自旋1/2海森堡自旋链

13.1.1 Jordon-Wigner变换

自旋1/2的海森堡自旋链中实际上有费米型激发。当然，海森堡模型可以从Hubbard模型中推导出来而Hubbard模型是关于费米子的，但是海森堡自旋链中的费米型激发和电子并没有什么关系——例如，这些费米型激发并不携带电荷（当然，这是因为海森堡模型中根本就没有电荷）。

背景知识 13.1: Jordan-Wigner变换

一维自旋1/2自由度可以使用一种比较容易的方法转化为费米子自由度。回顾背景知识12.1，自旋有升降算符，且对自旋1/2自由度，每个点上的自旋不是向上就是向下，因此暗示我们可以将每个点上的自旋大小当成这一点的某种费米子的个数。这样我们尝试定义我们尝试指定

$$\begin{aligned}\sigma_i^z &= 2f_i^\dagger f_i - 1, \\ \sigma_i^x &= f_i^\dagger + f_i, \\ \sigma_i^y &= i(f_i - f_i^\dagger),\end{aligned}\tag{13.1}$$

则能够验证算符 f_i^\dagger 和 f_i 分别让 z 方向自旋加1/2和减1/2，或者说让 σ_i^z 加减1，或者说让取值范围是0和1的 $f_i^\dagger f_i$ 加减1。此外还可以验证

$$\{f_i, f_i^\dagger\} = 0,\tag{13.2}$$

但不同格点的 f_i 和 f_i^\dagger 均对易。

为了将 f 转化成真正的费米子算符，我们做一个Klein变换，通过一串 f 算符的乘积来让 f 被交换时相位正确。定义

$$c_i^\dagger = e^{i\pi \sum_{k<i} f_k^\dagger f_k} f_i^\dagger, \quad c_i = e^{-i\pi \sum_{k<i} f_k^\dagger f_k} f_i,\tag{13.3}$$

则可以验证 c_i 算符给出正确的费米子对易关系。这样我们就把自旋1/2自由度完全转化为了费米子自由度，具体的变换方式为

$$\begin{aligned}\sigma_i^- &= (\sigma_i^x - i\sigma_i^y)/2 = e^{i\pi \sum_{k<i} c_k^\dagger c_k} c_i, \\ \sigma_i^+ &= (\sigma_i^x + i\sigma_i^y)/2 = e^{-i\pi \sum_{k<i} c_k^\dagger c_k} c_i^\dagger, \\ \sigma_i^z &= 2c_i^\dagger c_i - 1.\end{aligned}\tag{13.4}$$

Jordan-Wigner变换在更高维的空间中^[25]和自旋模长更高的自旋模型中^[5]均可以定义。由于自旋系统是“局域地玻色”的（即不同格点上的自旋自由度总是对易的，虽然单个格点上会有非平庸的对易关系），这实际上暗示费米子自由度某种意义上是一种非常不局域的东西，产生或是湮灭一个费米子需要弦算符而不是局域的算符。

费米子哈密顿量

现在对自旋1/2反铁磁海森堡模型

$$H = J \sum_i \mathbf{S}_i \cdot \mathbf{S}_{i+1} = \frac{J}{4} \sum_i \boldsymbol{\sigma}_i \cdot \boldsymbol{\sigma}_{i+1} = \frac{J}{4} \sum_i (\sigma_i^z \sigma_{i+1}^z + 2\sigma_i^+ \sigma_{i+1}^- + 2\sigma_i^- \sigma_{i+1}^+) \quad (13.5)$$

做Jordan-Wigner变换，其中

$$\sigma_i^z \sigma_{i+1}^z = (2c_i^\dagger c_i - 1)(2c_{i+1}^\dagger c_{i+1} - 1),$$

而

$$\sigma^+ i \sigma_{i+1}^- = c_i^\dagger e^{i\pi c_i^\dagger c_i} c_{i+1} = c_i^\dagger (1 - 2c_i^\dagger c_i) c_{i+1} = c_i^\dagger c_{i+1},$$

以及

$$\sigma_i^- \sigma_{i+1}^+ = c_i (1 - 2c_i^\dagger c_i) c_{i+1}^\dagger = c_i c_{i+1}^\dagger - 2(1 - c_i^\dagger c_i) c_i c_{i+1}^\dagger = -c_i c_{i+1}^\dagger = c_{i+1}^\dagger c_i,$$

于是就得到

$$H = \frac{J}{2} \sum_i (c_i^\dagger c_{i+1} + \text{h.c.}) + J \sum_i (n_i - 1/2)(n_{i+1} - 1/2). \quad (13.6)$$

对XXZ模型可以做同样的操作，模型

$$H = \gamma J \sum_i S_i^z S_{i+1}^z + J \sum_i (S_i^x S_{i+1}^x + S_i^y S_{i+1}^y) \quad (13.7)$$

会转化为

$$H = \frac{J}{2} \sum_i (c_i^\dagger c_{i+1} + \text{h.c.}) + \gamma J \sum_i (n_i - 1/2)(n_{i+1} - 1/2). \quad (13.8)$$

边界条件

13.1.2 自旋1/2的XY自旋链

作为一个简单的例子，考虑 $\gamma = 0$ 的情况，此时不存在 z 方向上自旋的耦合，即我们实际上是在处理XY模型。此时的费米子哈密顿量为

$$H = \frac{J}{2} \sum_i (c_i^\dagger c_{i+1} + \text{h.c.}), \quad (13.9)$$

是一个典型的紧束缚模型，解之得到能谱

$$\epsilon_k = J \cos(ka). \quad (13.10)$$

费米点位于 $k = \pm\pi/2$ 处。

可以看到一维XY自旋链中存在无能隙的费米型激发，这和较高维德自旋波模式完全不同。

玻色化

海森堡模型是自旋模型而不是电子模型，然而通过Jordon-Wigner变换

13.1.3 张量网络模拟

背景知识 13.2: 格点模型的一种数值计算方法：张量网络

文献^[20]给出了对常用的张量网络的一些介绍。所谓**张量网络**是指通过缩并一些张量得到的格点系统波函数（称为**张量网络态**）的试探形式。使用张量网络方法能够自然地引入对称性，能够很容易地从最终计算结果中获得诸如纠缠熵等信息，能够用于分析各种边界条件的系统，处理无穷系统比较方便，图形化语言意义清晰等等（例如对局域的哈密顿量，相邻格点的纠缠总是比较大，通过分析张量网络态中的纠缠我们能够确定一个模型的基态的“有效格点结构”）。相较于量子蒙特卡罗方法，张量网络方法的不足之处则在于难以估计计算质量好坏。

需要保证使用的张量网络态真的是足够好的拟设——选择一个张量网络波函数拟设本质上是在将系统的希尔伯特空间按照纠缠程度做分类，选择一个张量网络拟设等价于选择纠缠程度适当的一个子空间，如对有能隙系统，低能能量本征态的一个空间部分的纠缠熵正比于该空间部分的表面积（所谓**area law**），因此我们应该选择一个服从area law的张量网络拟设。此外，高效地实施张量网络计算需要适当安排缩并顺序。设张量网络图中每条线代表的指标的取值范围大致在1到 d （这个 d 称为这条边的维数，但是这个维数和晶格的维数、每个格点上的标签的取值数目（所谓“本地维数”）这三者之间没有必然联系），则计算一个有 n_1 条外线， n_2 条内线的张量网络缩并的时间复杂度大致在 $\mathcal{O}(d^{n_1+n_2})$ 。例如，普通的矩阵乘法 $[A_{ij}B_{jk}]_{ij}$ 涉及三条线—— i, j, k ——因此其时间复杂度为 $\mathcal{O}(d^3)$ 。适当地安排缩并顺序可以大幅降低计算整体的时间复杂度。

张量网络和图形演算（见背景知识17.3）之间有比较紧密的关系，但是它们并不完全是一回事。如果说后者像是微分几何中的逆变、协变矢量，前者就像是矩阵论。一个很容易看出的区别是引脚位置：在图形演算中，有明确物理意义的量——量子态，算符等——的指标或者说引脚是分前后顺序的（因为需要区分逆变和协变），图形上，有的向左有的向右。张量网络中不存在这样的区分。不过，就像一个区分逆变协变的微分几何方程可以用一个只有矩阵运算的线性代数库做数值计算一样，一个图形演算中的图形原则上也总是可以画成张量网络——简单地将指标向左或是向右的意义忘掉，并且在适当的地方加一个复共轭即可。然而，一旦我们需要在某一组基矢量下展开态矢量等——这件事经常需要做，比如在我们谈论局域性的时候——没有更多结构的图形演算是很难用的。一个张量网络中很容易表示的MPS在一个没有附加结构的图形演算中要如何表示呢？不过这不是说两者在具体计算上没有任何关系。在一些数值计算问题中，我们需要自动地快速计算一些图形演算的图，如何将它们转化成特定基矢量下的、能够快速计算的张量网络图是非常非平凡的问题。还有一些时候张量网络无法展示图形演算中的结构，我们需要根据图形演算的一些数据来约束张量网络的形式。

可能最有名的张量网络方法就是**密度矩阵重整化群（DMRG）**了。这是一种主要用于分析一维格点系统的张量网络方法，其基态波函数拟设

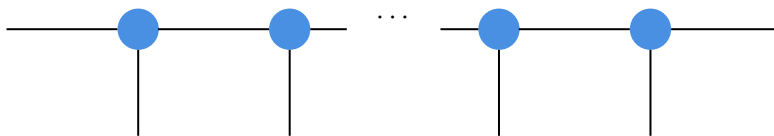


图 13.1: MPS拟设, 根据不同的边界条件可以调整最左边和最右边两个外线; 向下的线连接每个格点上的基矢量

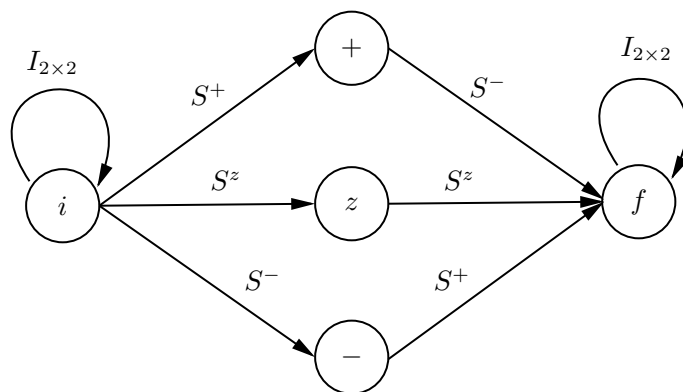


图 13.2: 海森堡模型的哈密顿量对应的自动机

为矩阵乘积态 (MPS)，绘制为图13.1。要施加开放边界条件只需移除左右两条外线，要施加周期性边界条件只需要将左右两个外线连接起来。在^[20]中介绍了MPS的一些性质：可以做到平移不变（只需让每个蓝点指向同一个张量即可），稠密（只要 d 足够大原则上可以构造出任意的一维格子上的波函数），纠缠熵服从area law，关联长度受控制等等。DMRG获得这个名字是因为MPS实际上是一种实空间重整化群（也称为DMRG）产生的基态波函数——实空间重整化群的主要难题在于， N 格点哈密顿量的最低几个能量本征态未必和 $N + 1$ 格点哈密顿量的最低几个能量本征态有足够大的重叠，从而朴素地“保留能量最低的几个能量本征态——加入新格点——对角化”的方法通常精度严重不足。DMRG重整化群根据约化密度矩阵来保留能量本征态，从而可以考虑到新加入格点和已有格点之间的量子纠缠信息，这就是它的名字的由来。

实空间重整化群由于种种问题（如根本不知道应该保留什么样的能量本征态），发展困难，而DMRG的张量网络形式却获得了快速发展，并成功地推广到更高维系统中。

旧式DMRG

我们首先来实现海森堡模型的旧式的、真的根据约化密度矩阵决定保留哪些能量本征态的DMRG。算法如下：

MPS方法

现在我们转而使用更加现代的方式理解海森堡模型的DMRG。

我们还需要将哈密顿量也写成张量网络的形式。一维的足够局域的系统哈密顿量总是可以写成矩阵乘积算符的形式。海森堡模型的哈密顿量可以看成一系列形如

$$\cdots \otimes I \otimes S^z \otimes S^z \otimes I \otimes \cdots$$

或

$$\cdots \otimes I \otimes S^+ \otimes S^- \otimes I \otimes \cdots$$

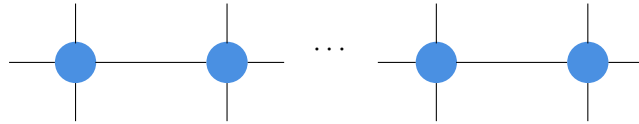
或

$$\cdots \otimes I \otimes S^- \otimes S^+ \otimes I \otimes \cdots$$

的项的和。这种连乘的序列可以使用图13.2所示的自动机描述。这个自动机的状态有五种，依次是 $f, -, +, z, i$ （排列顺序和转移矩阵(13.11)的矩阵元对应的基底的顺序一致）。海森堡模型中的任意一个直积序列都可以由这个自动机产生，而它也只产生海森堡模型哈密顿量中的直积序列。转移矩阵为

$$\begin{pmatrix} I_{2 \times 2} & 0 & 0 & 0 & 0 \\ S^+ & 0 & 0 & 0 & 0 \\ S^- & 0 & 0 & 0 & 0 \\ S^z & 0 & 0 & 0 & 0 \\ 0 & S^- & S^+ & S^z & I_{2 \times 2} \end{pmatrix} \quad (13.11)$$

的自动机产生。。将(13.11)连乘 N 次，其矩阵元的乘法视为直积，则最终得到的转移矩阵中对应 $i \rightarrow f$ 的矩阵元就是海森堡模型的哈密顿量。注意(13.11)实际上有四个指标，将两个(13.11)相乘无非是将不涉及 $I_{2 \times 2}, S^+, S^-, S^z$ 内部结构的两个指标缩并，于是以(13.11)为四阶张量构造如下张量网络



就得到了哈密顿量，或者说得准确一些，哈密顿量在 $\{S_i^z\}$ 表象下的矩阵元。

资料框 13.1: 张量网络库

Julia语言的iTensors库（见^[14]）提供了指标可命名（从而无需特别注意指标顺序）的张量，张量可以使用基于量子数的稀疏矩阵存储，可用于快速编写张量网络程序。iTensors也提供了许多有用的通用工具，使用很短的代码即可实现DMRG。APLS是一个含有各种计算物理方法的C++库。

13.2 横场伊辛模型

13.2.1 张量网络模拟

第十四章 自旋玻璃

自旋玻璃是一类自旋之间的相互作用随机的自旋系统，由于自旋间相互作用随机，无法形成自旋的长程序，从而，每个局部的自旋看起来都是稳定地冻结了的，即形成了短程序，然而无法定义任何长程的序参量，因为相邻的短程序的指向各不相同；相应的，系统存在一系列亚稳态，它们的能量彼此相差不多，从而如果系统构型落入了其中一个亚稳态，它可以在其中停留足够长的时间。这种大量实际上非常稳定的亚稳态存在，有短程序而没有长程序的物态和玻璃非常相似，这也就是我们称它为自旋玻璃的原因。

14.1 Edwards–Anderson模型

第五部分

短程纠缠的拓扑物态

在第二章中我们曾经问过：晶体的对称性确定了凝聚态系统的绝大多数性质，那么凝聚态系统的状态是否会有一些标签不是由通常意义上的对称性提供的？答案当然是肯定的，拓扑就是一个来源。

第十五章 拓扑能带论

本节要讨论的系统和第五章的哈密顿量基本上是差不多的：系统的基本自由度是某种电子，可以完全使用能带理论刻画它，相互作用可以忽略。然而，特殊的拓扑性质会让这些系统展现出和普通的金属、绝缘体非常不同的行为。

能带绝缘体中化学势位于两条能带中间，因此有明确的、彼此之间不连续的价带和导带，换言之，载流子存在能隙（见第5.3.3节）。另一方面，导体中载流子不存在能隙。单电子能谱是否存在能隙也决定了电子态具有或者不具有能隙，因为多电子态能够发生的最小偏移就是多一个或者少一个电子。

既然连续地、局域地调节哈密顿量不会改变能隙的有无，我们可以据此定义一个拓扑等价类。一种比较粗糙的方法是将所有的传统绝缘体（就是第五章中出现的那些）都可以归入一类，这个类别中也包括真空，因为真空中的电子同样可以认为有一个价带（电子）和一个满带（正电子），两者之间存在能隙。

有能隙系统的低能行为和费米能级以上的能带无关，或者说能带绝缘体中的电子态似乎可以认为是平凡的。然而，这不是必然的：具有能隙的电子态同样可以因为拓扑而有非平凡的行为。这就是所谓**拓扑绝缘体**。

背景知识 15.1: 哈密顿量的拓扑分类

哈密顿量变化，系统的基态和激发态也会随之变化。哈密顿量的局域、连续的变化**不能**让基态和激发态遍历所有的可能。例如，基态和激发态之间是否存在能隙这件事在哈密顿量的连续局域变化下是不会变的。这些诸如能隙有无之类的东西可以看成拓扑不变量，据此我们可以对哈密顿量做拓扑分类。

15.1 整数量子霍尔效应

整数量子霍尔效应是另一个拓扑导致的非常新奇的效应。经典霍尔效应预言的霍尔电导为(5.18)，于是霍尔电阻正比于 B 。然而在低温、强磁场下，观察到 ρ 和 B 之间的关系**存在平台**（所谓**电导平台**），稳定时的 ν 取值为

$$\nu = 1, 2, 3, \dots, \frac{1}{3}, \frac{2}{3}, \frac{2}{5}, \dots$$

ν 取整数的情况称为**整数量子霍尔效应**，取分数的情况为**分数量子霍尔效应**。

事实证明，整数量子霍尔效应仅仅在拓扑能带理论的框架下即可得到很好的解释，即它实质上还是一个短程纠缠系统。分数量子霍尔效应则涉及强关联效应，而且存在拓扑序，即有长程纠缠。

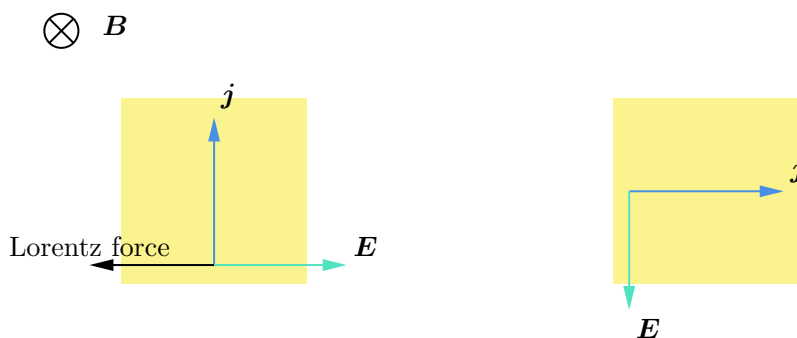


图 15.1: 电导率张量的不对称性: 在 x 方向加一个电场可以保持一个 y 方向的稳定的霍尔电流, 可是需要在 $-y$ 方向加一个电场才能够保持 x 方向的稳定霍尔电流。

15.1.1 整数量子霍尔效应的基本现象

在低温 ($\sim 1.5\text{ K}$)、强磁场 ($\sim 20\text{ T}$) 时对材料做霍尔测量, 可能观察到如下的现象: 霍尔电阻随着磁场的增强会有增长, 但是这个增长不是线性的, 而是会出现一系列平台; 在霍尔电阻的两个平台交界处, 纵向电阻出现高峰, 而在平台内部, 纵向电阻则是零。应当注意, 这里由于磁场的存在, 就电导率矩阵 σ 本身而言, 时间反演对称性是破缺的, 即它不是一个厄米的矩阵。事实上, 注意到我们有图15.1, 电导率张量应该写成

$$\sigma = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} \\ -\sigma_{xy} & \sigma_{xx} \end{pmatrix}. \quad (15.1)$$

电动力学的论证告诉我们, 体系中的能量损耗是

$$\mathbf{j} \cdot \mathbf{E} = \mathbf{j} \cdot \sigma^{-1} \cdot \mathbf{j},$$

于是我们可以定义电阻率张量

$$\rho = \sigma^{-1} = \begin{pmatrix} \rho_{xx} & \rho_{xy} \\ -\rho_{xy} & \rho_{xx} \end{pmatrix}, \quad \sigma_{xx} = \frac{\rho_{xx}}{\rho_{xx}^2 + \rho_{xy}^2}, \quad \sigma_{xy} = -\frac{\rho_{xy}}{\rho_{xx}^2 + \rho_{xy}^2}. \quad (15.2)$$

这里有一个看起来很奇怪的地方, 就是存在霍尔电阻的时候, $\sigma_{xx} = 0$ 对应着 $\rho_{xx} = 0$, 但是在普通的导体中, 电导率为零应该电阻率为无穷大才对。实际上, 我们可以看到, $\rho_{xx} \rightarrow \infty$ 时确实有 $\sigma_{xx} \rightarrow 0$ 。 $\sigma_{xx} = 0$ 和 $\rho_{xx} = 0$ 同时成立实际上对应着这样一种情况: 由于某些巧妙的安排, 材料里面是真的一点电流也没有, 从而也没有能耗。

15.1.2 整数电导平台的定性分析

朗道能级

在第5.4.1节中我们已经考虑过了朗道能级。设被填充的最高的朗道能级编号为 n , 且填充在朗道能级中的电子总数为 N_e , 则有

$$\frac{N_e}{2} = (n-1) \frac{\Phi}{\Phi_0} + n_{\text{high}}, \quad 0 < n_{\text{high}} \leq \frac{\Phi}{\Phi_0},$$

于是随着 $1/B$ 上升 (即 B 下降), n 能级上的电子轨道占据数 n_{high} 线性上升。如果我们能够保证电导大体上正比于电子填充数, 那么以上机制就给出了电导平台的一种可能成因。以下我们解释为什么。

如果材料是非常理想的，那么其实不应该出现电导平台，因为此时 n 朗道能级被填充之后，立刻会开始填充 $n+1$ 朗道能级，两个朗道能级之间由于不存在任何态，化学势可以发生不连续变化。然而实际的体系中有杂质，因此朗道能级会出现展宽，并且两个朗道能级之间的态基本上是局域化态，而朗道能级中的态的能量分布则舒展一些（见第9.1节）。虽然直观地看，朗道能级附近的电子似乎是原地打转，但加入一个电场之后，它们是可以自由移动的，能够形成延展态（比较没有磁场的晶体：价电子也只是出现在原子附近，原地打转，但是形成晶体后，晶格势场还是会让价电子形成延展态）。因此，化学势在它们附近时，它们可以贡献纵向电导，而两个朗道能级之间的电子高度定域，任何时候都无法贡献纵向电导。这解释了为什么纵向电导（从而，纵向电阻——见(15.2)附近的讨论）会在特定位置出现峰，因为只有在化学势在朗道能级附近时才可能有纵向电流。这就造成了一个非常矛盾的情况：要观察到量子霍尔效应，体系必须比较“脏”，这样才能够有明确的定域态，从而形成前述现象，否则，化学势可以从一个朗道能级快速跳转到另一个朗道能级。但实际上，在体系很脏时朗道能级附近的延展态也被破坏了，因此太脏的体系全然没有纵向电流，观察不到纵向电流随着磁场变化出现的峰。因此明显的量子霍尔效应需要体系有些脏但又不太脏。

实际上，如果我们暂且认可化学势以下的朗道能级上的电子全部能够对霍尔电导有贡献，我们还能够解释霍尔电阻的行为：在填充延展态时电导关于 $1/B$ 线性上升，而在填充定域态时电导没有变化，出现平台。然而我们马上会注意到一个不对头的地方，就是化学势都不在朗道能级上，而是在两个朗道能级之间，即化学势位于能隙之中时，体系是有能隙的，化学势以下的朗道能级被填满，要形成电流需要在这个朗道能级上形成空穴，因此需要消耗有限大的能量。因此，我们实际上是在要求有能隙的电子模式——朗道能级——产生电流。这真是不可理喻。但是实际上我们有一种方法绕过这个限制，那就是要求体态有能隙，但边界态没有能隙，从而边界态导电。如果真的能够形成能导电的边界态，体态有能隙倒不是坏事，因为这意味着体态内部的关联长度有限，因此边界态和体态相对独立。我们指望通过边界态电流来解释 σ_{xy} 的平台行为。

朗道能级确实能够形成稳定的边界态电流。可以从一个经典图像看到这一点：在体态中电子可以不停做圆周运动，而在边界附近电子做完半个圆周运动后被反弹，而又往前做半个圆周运动。因此，可以形成绕着整个边界运行的（手性的）电流。边界态上如果有杂质，电子可以潜到体态中，绕过杂质，形成一个新的边界态。因此边界态上的电流是无损耗的。由于磁场的存在，我们还会发现这个边界态电流是手征的（见(11.4)后面的讨论）。现在考虑我们给材料施加一个电场，如图15.2所示。如果材料在和电场垂直的方向上和外界导通，这会让和电场平行的两条边上的电子运动速度不同，由于电荷守恒，上边的电流要比下边的电流大，其结果就是外界能够测到一个从右往左的净电流。当然，这就是霍尔电流。

在磁场比较高时，只有几个朗道能级上面填充了电子，从而霍尔电导比较低，霍尔电阻比较高，从而能够容易地观察到量子霍尔效应。

我们应当指出，稍微仔细想想，以上说法有很大的问题，因为边界态上只有一种手征，而根据Nielsen–Ninomiya定理，狄拉克费米子不应该出现手征反常：左手费米子模式的数目和右手费米子模式的数目总是一样的，并且手征流应该是守恒的。我们在这里能够有手征反常，是因为我们在一个体态理论的边界上工作。具有特定的量子反常的理论能够良定义在更高维的理论的边界上，这是拓扑能带论中一个重要的想法。

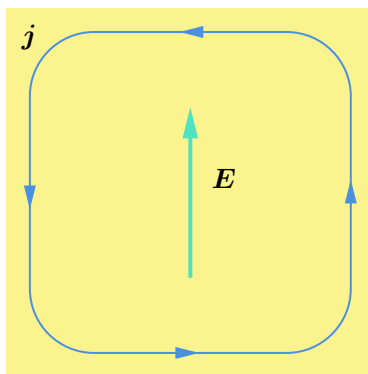


图 15.2: 霍尔电流如何从边界态电流中形成

霍尔电导的形式和Laughlin论证

然而以上论证无法解释一件事，就是为什么电导平台的值不多不少，就是几个基本物理常数确定的

$$\sigma_H = n \frac{e^2}{h}, \quad n = 0, 1, 2, \dots$$

通常的电导涉及关于材料的复杂性质，是不可能如此简洁的。**Laughlin论证**是一种直观理解关于为什么霍尔电导形式非常简洁的方法。设有一个半径为 L 的圆柱体，其体态有能隙而边界态导电。沿着它的轴向加入一个磁场，总磁通量为 Φ 。让 Φ 缓慢地发生变化，从零变化到 $\frac{h}{e}$ ，则 Φ 变化前后圆柱体均处于同样的状态，因为绕着大小为 h/e 的磁通量转一圈什么也不会发生。磁通量的变化会导致一个电场，在边界上，我们有

$$2\pi L E = -\frac{d\Phi}{dt},$$

从而

$$E(t) = \frac{1}{2\pi L} \frac{d\Phi}{dt}.$$

边界上的电场方向垂直轴向，从而产生一个平行于轴向的霍尔电流

$$I = 2\pi L j = 2\pi L \sigma_H E = \sigma_H \frac{d\Phi}{dt}.$$

这个电流造成的电荷量变化为

$$\Delta Q = \int dt I = \sigma_H \Delta \Phi = \sigma_H \frac{h}{e},$$

而由于电荷由电子携带，有

$$\Delta Q = m e, \quad m = 0, 1, 2, \dots,$$

于是

$$\sigma_H = m \frac{e^2}{h}, \quad m = 0, 1, 2, \dots$$

这就是整数阶量子霍尔效应的来源：它和体系的结构完全无关，只要体系体态有能隙，就能够得出存在整数霍尔效应的结论。

15.1.3 线性响应理论和拓扑不变量

以上的定性分析能够解释整数量子霍尔效应，不过我们当然想要一个更加一般的、能够适用于各种类似的系统的解释。本节我们将指出，二维系统的 σ_{xy} 实际上是一个拓扑不变量，它依赖于倒空间中的能带结构的拓扑分类，从而，

背景知识 15.2: Berry相位

设含时哈密顿量 H 依赖一系列连续参数 $\mathbf{R}(t) = (R_1(t), \dots, R_n(t))$, 用 $|n(\mathbf{R}(t))\rangle$ 标记该哈密顿量的本征态。假定 $\mathbf{R}(t)$ 变化得（相比于 H 的能谱的最小能隙）充分缓慢, 以至于如果系统以 $|n(\mathbf{R}(t))\rangle$ 为初态, 那么经过一段时间的演化之后不会跃迁到其他本征态上（**绝热演化**）。

设系统初态为 $|n(\mathbf{R}(t))\rangle$, 经过一段时间后系统状态为

$$|\psi_n(t)\rangle = e^{i\gamma_n(t)} \exp\left(-\frac{i}{\hbar} \int_0^t dt' E_n(\mathbf{R}(t'))\right) |n(\mathbf{R}(t))\rangle, \quad (15.3)$$

将上式代入

$$i\hbar \frac{d}{dt} |\psi_n(t)\rangle = H(\mathbf{R}(t)) |\psi_n(t)\rangle,$$

得到

$$\gamma_n(t) = i \int_0^t dt' \langle \psi_n(t) | \frac{d}{dt} | \psi_n(t) \rangle = i \int_{\mathbf{R}(0)}^{\mathbf{R}(t)} d\mathbf{R} \cdot \langle \psi_n(t) | \nabla_{\mathbf{R}} | \psi_n(t) \rangle. \quad (15.4)$$

注意到 $\gamma_n(t)$ 不含有任何 \hbar , 因此实际上这不是一个动力学相位, 而是所谓的几何相位。定义

$$\mathbf{A} = i \langle \psi_n(t) | \nabla_{\mathbf{R}} | \psi_n(t) \rangle \quad (15.5)$$

为**Berry联络**, 相应的相位称为**Berry相位**。这里的底流形是参数空间（当然也可以看成系统基态的空间）, 其上的协变导数定义要求相邻两个 \mathbf{R} 对应的态可以通过绝热演化连接。Berry联络就是满足这个条件的联络。从Berry联络自然可以得到Berry曲率张量

$$\Omega_{n,\mu\nu} = \partial_{R_\mu} \partial_{R_\nu} A_n - \partial_{R_\nu} \partial_{R_\mu} A_n. \quad (15.6)$$

Berry联络是规范可变的, 因为容易验证, 如果做变换

$$|n(\mathbf{R})\rangle \longrightarrow e^{i\zeta(\mathbf{R})} |n(\mathbf{R})\rangle, \quad (15.7)$$

就有

$$\mathbf{A}(\mathbf{R}) \longrightarrow \mathbf{A}(\mathbf{R}) - \nabla_{\mathbf{R}} \zeta(\mathbf{R}). \quad (15.8)$$

在时间从0演化到 T 的闭合的 $\mathbf{R}(t)$ 回路上, 由于单值性, 有

$$\zeta(\mathbf{R}(T)) - \zeta(\mathbf{R}(0)) = 2\pi n, \quad n \in \mathbb{Z},$$

从而积分

$$\oint d\mathbf{R} \cdot \mathbf{A} \quad (15.9)$$

在模 2π 的意义下是规范不变的。这实际上就是说, **Wilson loop**

$$e^{i \oint d\mathbf{R} \cdot \mathbf{A}} \quad (15.10)$$

是规范不变的。

以上讨论的问题中参数空间都是外加参数, 但是实际上也可以用系统内的物理量做参数。现在向系统中加入一个变化缓慢的外场, 用它“牵引”系统做绝热演化, 由于因此此时 $\langle g | \partial_t | g \rangle$ 就给出一个Berry相位, 参数空间是外场的参数空间, 而由于是绝

热演化，外场通常会让系统中的某一些物理量非常确定，那么我们就能够用系统中的物理量做参数了。例如，可以在空间中假象地引入一个势阱，用它让粒子每一时刻都几乎位于一个确定的位置，则这个势阱的位置——实际上就是粒子的位置——的取值范围（比如说一圆圈）就构成了一个参数空间，其上就可以定义Berry联络。A-B效应就是一个典型例子，在那里参数就是粒子位置 \mathbf{r} 。在拓扑能带论中，参数则可以是晶格动量 \mathbf{k} 。

Berry相位可以体现在路径积分中。回顾各种路径积分的推导方式，都可以找到类似于 $\langle g|\partial_t|g\rangle$ 的项，其中 $|g\rangle$ 通常是某种“相干态”。设相干态是关于某些变量 q 的，回顾路径积分的推导过程， $\langle g|\partial_t|g\rangle$ 项实际上就对应着经典的从哈密顿量到拉格朗日量的勒让德变换中 $p\dot{q}$ 这一项，因此如果有非平庸的Berry相位那么它一定会被收集在这一项中。因此拉氏量中的 $p\dot{q}$ 项常常被称为Berry相位项，虽然一般情况下它也会容纳一些动力学演化的信息。 q 高度确定的绝热演化过程中， p 中显含 \dot{q} 的项都可以丢弃，因为演化速度很慢（实际上 \dot{q} 通常和 q 是不对易的，不过无所谓，反正都是要丢弃的），于是 $\int dt p\dot{q}$ 可以写成 $\int p(q) dq$ ，此时它就是Berry相位。

可以看到上面的构造是非常一般的。比较平庸的情况是，系统中有一个根源上来自某种规范场（一般是电磁场）的外场，它引入了一个Berry相位。但是我们也可以把这个思路反过来：如果某个物理系统中有一些很像是Berry相位的東西，那么这个系统的有效理论就会很像一个规范场论。一个典型的例子见背景知识17.2。

15.1.4 整数量子霍尔效应的Chern-Simons场论描述

整数量子霍尔效应实际上可以使用Chern-Simons理论描述^[1]。实际上分数量子霍尔效应也和Chern-Simons理论有关，见第17.1.2节。这样做的好处在于，只要我们能够从得到的Chern-Simons理论中找到电子对应的自由度，那么就可以很容易地分析电子间相互作用比较重要时拓扑绝缘体的各种性质是否还能够维持。

我们要做的事情大体上是这样：既然整数量子霍尔效应中的体态电子有能隙而边界态能产生的物理效应无非是导电，实际上可以通过写出体态中的电磁场的低能有效理论来得到整数量子霍尔效应的低能有效理论。

整数量子霍尔效应还可以被赋予流体力学意义^[10]。

15.2 Su-Schrieffer-Heeger模型

Su-Schrieffer-Heeger (SSH) 模型是一个定义在一维格子上的紧束缚模型，在这个一维格子中有A和B两套子格^[4]。在一维导电分子链上这个模型很容易出现，因为Peierls不稳定性——一维晶格上的“声子凝聚”——意味着一维简单晶格是不可能稳定存在的。与之前不同，我们取开放边界条件，SSH模型为

$$H = v \sum_{m=1}^N (c_{mA}^\dagger c_{mB} + \text{h.c.}) + w \sum_{m=1}^{N-1} (c_{m+1,A}^\dagger c_{mB} + \text{h.c.}). \quad (15.11)$$

我们忽略了自旋指标，即只考虑一种自旋的电子，忽略自旋翻转的可能性；使用实际的分子实现SSH模型时，会得到两份SSH模型，我们只考虑其中一份。

在分析SSH模型时通常将电子在A、B子格上的哪一个看成内禀自由度而非坐标自由度或者说外部自由度，这实际上就是紧束缚模型中把能带标签当成抽象的一个指标而忽略它



图 15.3: SSH模型所在的格点，红色和蓝色圆圈分别代表A子格和B子格，只有相邻格点之间有跃迁

对电子空间分布的影响，而将 i 看成完全代表了电子空间位置的标签的做法。在这种约定下，(15.11)可以写成

$$H = v \sum_{m=1}^N |m\rangle\langle m| \otimes \sigma^x + w \sum_{m=1}^{N-1} |m+1\rangle\langle m| \otimes \frac{\sigma^x + i\sigma^y}{2} + \text{h.c.} \quad (15.12)$$

这里算符 $\sigma^{x,y,z}$ 作用于A, B两套子格张成的二维空间上——或者说对应的两条能带的能带编号张成的二维空间上（注意能带编号和子格编号未必是同样的表象）。

15.2.1 体态

能谱

一个有限大小的一维格子的体态是其内部，而边界则是两个点。本节分析SSH模型的体态。取热力学极限 $N \rightarrow \infty$ 并施加周期性边界条件，可以求解得到普通的能带。做傅里叶变换

$$c_{m\alpha}^\dagger = \frac{1}{\sqrt{N}} \sum_k e^{-ikma} c_{k\alpha}^\dagger, \quad \alpha = A, B, \quad (15.13)$$

得到

$$H = \sum_k \begin{pmatrix} c_{kA}^\dagger & c_{kB}^\dagger \end{pmatrix} \begin{pmatrix} 0 & v + we^{-ika} \\ v + we^{ika} & 0 \end{pmatrix} \begin{pmatrix} c_{kA} \\ c_{kB} \end{pmatrix}, \quad (15.14)$$

对角化给出

$$\epsilon_k = \pm \sqrt{v^2 + w^2 + 2vw \cos(ka)}. \quad (15.15)$$

两条能带之间的间距为

$$2\Delta = |u - v|. \quad (15.16)$$

容易看出在化学势为零时电子半填充，且电子和空穴的能隙（相对于费米面的距离）均为 Δ 。

在 $u = v$ 时我们得到通常的最为简单的紧束缚模型，系统是导体，因为载流子能隙为零；其余情况下，系统为绝缘体。换言之，跃迁系数的交错排列（staggering）让SSH模型的体态是绝缘体。

动量空间哈密顿量和卷绕数

在(15.12)中我们看到SSH模型的哈密顿量正比于作用于A, B两套子格张成的二维希尔伯特空间上的 σ 矩阵，从而动量空间中的哈密顿量可以写成

$$H_k = \sum_k (h_0(k)\sigma^0 + h_i(k)\sigma^i), \quad (15.17)$$

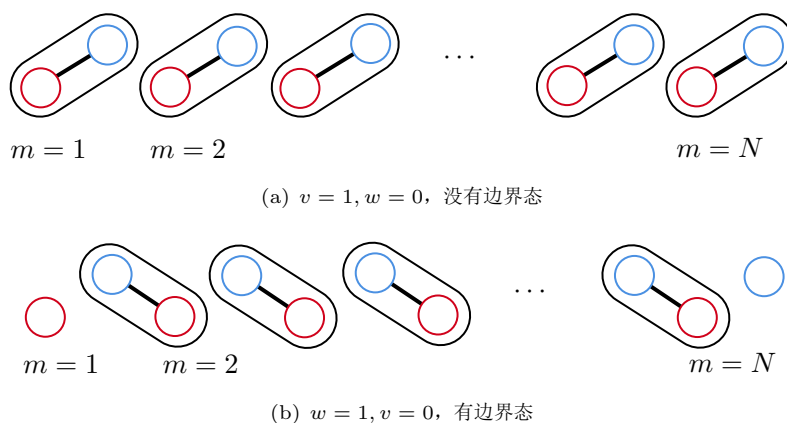


图 15.4: SSH模型的两个极限

这里我们以子格编号而不是能带编号为表象。显然，二能带模型都能够写成这种形式，因为 \mathbb{C}^2 中的算符一定能够写成四个 σ 矩阵的线性组合。在SSH模型中 $h_0 = 0$ ，将(15.11)的坐标部分对角化，可以计算得到

$$h_x(k) = v + w \cos(ka), \quad h_y(k) = w \sin(ka), \quad h_z(k) = 0. \quad (15.18)$$

对每个一维的二能带模型，我们都有一个这样的关系：

$$k \mapsto (h_x(k), h_y(k), h_z(k)) : \mathbb{R}/2\pi \rightarrow \mathbb{R}^3. \quad (15.19)$$

由于第一布里渊区的周期性， $k = 0$ 和 $k = 2\pi$ 处，各个 h 值应该是一样的，即(15.19)。如果我们在处理一个体态为绝缘体的系统，还不能有一个 k 点让所有 h 都是零，否则载流子不存在能隙，就是导体了。这样(15.19)就是一个 \mathbb{R}^3 中的闭合曲线，并且不经过原点。

对这条曲线我们很自然地会想到要计算卷绕数。对(15.18)， $w > v$ 时卷绕数为1，而 $w < v$ 时卷绕数为零。这两种参数选取的交界点满足 $w = v$ ，此时SSH模型体态为导体。更一般的，对曲线(15.19)的小幅变动不会改变卷绕数：要改变卷绕数，必须对 v 和 w 做有限大小的调整，让曲线(15.19)的其中一段通过原点。换言之， (v, w) 构成的参数空间中测度非零的区域都是绝缘体，但是这些区域中的卷绕数可能是不同的；不同卷绕数的区域的交界线一定是导体。这给出了一种新型相变的机制：拓扑量子相变。

然后我们马上会发现这样一个事实：宏观下我们可以同时有比较确定的坐标和动量，例如我们可以讨论坐标 \mathbf{r} 处的固体微团的能带结构，这样，实际上可以构造一个系统，其中每个固体微团都是一个SSH模型，其 v 和 w 是空间坐标 \mathbf{r} 的函数，如果一个区域中 $v < w$ 而另一个区域中 $v > w$ ，那么这两个区域的交界处——如果有的话——一定是导体。现实中我们就有这样的模型——一个有限大小的固体，如果其体态中有非平凡卷绕数，则它和空气接触的边界必定是导体！

这意味着对体态中的电子能带有非平凡的卷绕数的二能带绝缘体，其边界上一定有一些和体态中不同的无能隙电子模式。我们在下一节中将讨论，SSH模型中的这种边界态到底是怎么来的。

电子的二聚体

我们考虑两个极限, $w = 1, v = 0$ 和 $v = 1, w = 0$ 。 $v = 1, w = 0$ 时卷绕数为0, 哈密顿量为

$$H = v \sum_{m=1}^N |m\rangle\langle m| \otimes \sigma^x, \quad (15.20)$$

即不同 m 之间没有任何跃迁, 但是电子能够在同一个 m 的两个子格点之间跃迁。换言之, 每个 m 的两个子格点之间“成键”, 形成电子二聚体。对角化某个 m 就是要将 σ^x 在 σ^z 表象下对角化, 低能本征态为 $(|\uparrow\rangle - |\downarrow\rangle)/\sqrt{2}$, 或者用 A 和 B 来编号就是 $(|A\rangle - |B\rangle)/\sqrt{2}$, 从而 $v = 1, w = 0$ 的 SSH 模型的基态就是

$$|\text{ground}\rangle = \otimes_m \frac{1}{\sqrt{2}} (|mA\rangle - |mB\rangle), \quad (15.21)$$

画成图就是图 15.4(a)。可以看到这种情况下不存在任何非平庸的边界态: 位于边界上的格点被“吸收”进了体态中的电子二聚体中。

另一方面, $w = 1, v = 0$ 时哈密顿量为

$$H = \sum_{m=1}^{N-1} |m+1\rangle\langle m| \otimes \frac{\sigma^x + i\sigma^y}{2} + \text{h.c.}, \quad (15.22)$$

此时仔细观察, 会发现只有 mB 和 $m+1, A$ 之间能够出现跃迁, 因此同样会出现电子二聚体态, 但是将它画成图, 是图 15.4(b)。此时只能形成 $N-1$ 个二聚体, $m = 1, A$ 和 $m = N, B$ 两个态不属于任何二聚体。由于它们在边界上, 此时存在边界态。

通过(15.15), 我们发现无论是 $v = 1, w = 0$ 还是 $w = 1, v = 0$ 时电子能带都是平带, 两者的体态行为看上去完全一样。然而, $w = 1, v = 0$ 时存在边界态, 而 $v = 1, w = 0$ 时没有边界态。我们之间已经说明过, $w = 1, v = 0$ 时体态中有非零卷绕数, 从而的确应该有边界态, 因此所有东西都符合得很好。

15.2.2 边界态

15.2.3 对称性保护

下面我们要分析在什么扰动下 SSH 模型的拓扑性质会完全消失。

15.3 一维 Kitaev 链拓扑超导体

15.3.1 Kitaev 链及其解析解

以下一维模型称为 **Kitaev 链**:

$$H = -t \sum_i (c_i^\dagger c_{i+1} + \text{h.c.}) - \mu \sum_i c_i^\dagger c_i + \sum_i (\Delta c_i c_{i+1} + \text{h.c.}). \quad (15.23)$$

(15.23) 是一个 p 波超导模型, 这个模型通常是这么来的: 一个一维纳米线被放置在一个超导体上, 两者的相互作用诱导前者也发生 $U(1)$ 对称性破缺, 然后我们使用平均场理论分析问题而引入一个 Δ 参量。(15.23) 是一个紧束缚模型, 对(15.23)做傅里叶变换, 可以得到

$$H = \frac{1}{2} \sum_{\mathbf{k}} \underbrace{\begin{pmatrix} c_{\mathbf{k}}^\dagger & c_{-\mathbf{k}} \end{pmatrix}}_{\Psi^\dagger} \begin{pmatrix} \epsilon_{\mathbf{k}} - \mu & -2i\Delta^* \sin k \\ 2i\Delta \sin k & -\epsilon_{\mathbf{k}} + \mu \end{pmatrix} \underbrace{\begin{pmatrix} c_{\mathbf{k}} \\ c_{-\mathbf{k}}^\dagger \end{pmatrix}}_{\Psi}, \quad (15.24)$$

然后再做Bogoliubov变换，计算出以下能谱：

$$E_{\mathbf{k}} = \pm \sqrt{(2t \cos k - \mu)^2 + 4|\Delta|^2 \sin^2 k}. \quad (15.25)$$

(15.23)具有粒子-空穴对称性。总之就有一个约束就是设 P 为粒子空穴变换，我们有

$$P\Psi_{\mathbf{k}}P^{-1} = \tau^*\Psi_{-\mathbf{k}}^*$$

Kitaev链不存在对称性自发破缺，但能隙可开可闭。当

$$\mu = \pm 2t \quad (15.26)$$

时，能隙会关闭。除此之外任何参数的变动都只会引起连续的变化。因此，如果体系发生相变，那么只能是在(15.26)处发生一个和对称性无关的相变。在化学势很低时，即 μ 趋于负无穷时，根本就没有电子，因此从 $-\infty$ 到 $-2t$ 的部分肯定是平庸的。化学势非常高时（大于 $2t$ 时）电子全满，同样是平庸的。因此有趣的行为集中在 $-2t$ 到 $2t$ 之间。下面会看到，当 μ 越过(15.26)这两个点时，会发生一个拓扑相变。

$W = \pm 1$ ，这是一个定义在立体中的量？

15.3.2 Kitaev链中的拓扑不变量

下面定义一个能带的拓扑不变量。

总之，当 μ 扫过 $\mu = -2t$ 时，我们有

$$c_i = \frac{1}{2}(\gamma_{iA} + i\gamma_{iB}),$$

容易验证均为

$$\{\gamma_\alpha, \gamma_\beta^\dagger\} = 2\delta_{\alpha\beta}, \quad (15.27)$$

15.3.3 时间反演对称性保护的拓扑超导

刚才描述的拓扑超导和对称性没有特别明确的对称性。当然可以说它有粒子-空穴对称性，但是这完全是一个数学上的处理。

第十六章 自旋系统中的拓扑

16.1 一维海森堡自旋链

16.1.1 一维海森堡自旋链的路径积分

背景知识 16.1: 自旋的路径积分

自旋是可以使用路径积分描述的。这里的路径积分和通常的场论的路径积分不同，因为此时的基本自由度是自旋，没有形如 $[x, p] = i$ 这样的对易关系。虽然如此，我们仍然可以有路径积分，因为我们有“对自旋构型的积分”—— $SU(2)$ 上的Haar测度——并且实际上也有“自旋相干态”。首先，我们知道 $SU(2)$ 群的任何一个有限维不可约表示均可以使用如下的欧拉角表示出来：

$$g(\phi, \theta, \psi) = e^{-i\phi S_3} e^{-i\theta S_2} e^{-i\psi S_3}, \quad \phi, \psi \in [0, 2\pi], \quad \theta \in [0, \pi]. \quad (16.1)$$

设 $|\uparrow\rangle$ 为 S_3 的最高权本征态，即本征值最大的本征态，我们就有

$$e^{-i\psi S_3} |\uparrow\rangle = e^{-i\psi S} |\uparrow\rangle, \quad (16.2)$$

其中 S 是 S_3 的最大本征值，对自旋 $1/2$ 表示它是 $1/2$ ，对自旋 1 表示它是 1 ，等等。我们注意到

$$\begin{aligned} i \langle \uparrow | S_2 | \uparrow \rangle &= \langle \uparrow | [S_3, S_1] | \uparrow \rangle \\ &= S \langle \uparrow | S_1 | \uparrow \rangle - S \langle \uparrow | S_1 | \uparrow \rangle = 0, \end{aligned}$$

同理

$$\langle \uparrow | S_1 | \uparrow \rangle = 0.$$

定义

$$|g\rangle = g |\uparrow\rangle, \quad g \in SU(2), \quad (16.3)$$

称它为自旋相干态。这个说法的依据在于，根据Haar测度的定义，有

$$h \int dg |g\rangle \langle g| = \int dg |hg\rangle \langle g| = \int dg |g\rangle \langle h^{-1}g| = \int dg |g\rangle \langle g| h,$$

根据不可约表示的Schur引理，我们有

$$\int dg |g\rangle \langle g| = \text{const} \times \text{id}, \quad (16.4)$$

得到了定义路径积分需要的完备性关系，而 $|g\rangle$ 的地位和基于 \mathbf{x}, \mathbf{p} 的路径积分中的相干态类似。使用“将时间分片并插入完备性关系”的方法，就有

$$Z = \int \mathcal{D}g \exp \left(\int_0^\beta d\tau (\langle \partial_\tau g | g \rangle - \langle g | H | g \rangle) \right). \quad (16.5)$$

在这里我们可以看到，自旋的路径积分只涉及一类算符（ \mathbf{S} 的各个分量）而不是两类（ \mathbf{x} 和 \mathbf{p} ），从而相干态路径积分看起来会简单一些；但是自旋的路径积分中 \mathbf{S} 是在一个球上取值而不是在平直的坐标空间和动量空间中取值，并且彼此不对易，因此下面当我们把这些内积展开时又会有比坐标-动量路径积分更复杂的东西。

用欧拉角把 $|g\rangle$ 写出来就是

$$|g\rangle = e^{-i\psi S} e^{-i\phi S_3} e^{-i\theta S_2} |\uparrow\rangle.$$

我们首先处理(16.5)的第一项，我们有

$$\int_0^\beta d\tau \langle \partial_\tau g | g \rangle = \int_0^\beta d\tau (iS \partial_\tau \psi + \langle \uparrow | \partial_\tau (e^{i\theta S_2} e^{i\phi S_3}) e^{-i\phi S_3} e^{-i\theta S_2} |\uparrow\rangle),$$

其中的第一项是零，因为 ψ 在 $\tau = 0$ 和 $\tau = \beta$ 处是相同的。第二项是

$$\begin{aligned} \int_0^\beta d\tau \langle \uparrow | \partial_\tau (e^{i\theta S_2} e^{i\phi S_3}) e^{-i\phi S_3} e^{-i\theta S_2} |\uparrow\rangle &= \int_0^\beta d\tau i \partial_\tau \theta \langle \uparrow | S_2 e^{i\theta S_2} e^{i\phi S_3} e^{-i\phi S_3} e^{-i\theta S_2} |\uparrow\rangle \\ &+ \int_0^\beta d\tau \langle \uparrow | e^{i\theta S_2} i \partial_\tau \phi S_3 e^{i\phi S_3} e^{-i\phi S_3} e^{-i\theta S_2} |\uparrow\rangle, \end{aligned}$$

这里的第一项还是零，因为 $\langle \uparrow | S_2 | \uparrow \rangle$ 是零。第二项是

$$\begin{aligned} \int_0^\beta d\tau \langle \uparrow | e^{i\theta S_2} i \partial_\tau \phi S_3 e^{i\phi S_3} e^{-i\phi S_3} e^{-i\theta S_2} |\uparrow\rangle &= \int_0^\beta d\tau i \partial_\tau \phi \langle \uparrow | e^{i\theta S_2} S_3 e^{-i\theta S_2} |\uparrow\rangle \\ &= iS \int_0^\beta d\tau \partial_\tau \phi \cos \theta, \end{aligned}$$

这里我们用到了

$$\begin{aligned} \langle \uparrow | e^{i\theta S_2} S_3 e^{-i\theta S_2} |\uparrow\rangle &= \langle \uparrow | e^{i\theta [S_2, \cdot]} S_3 |\uparrow\rangle \\ &= \langle \uparrow | 1 + (i\theta) i S_1 + \frac{(i\theta)^2}{2} + \dots S_3 |\uparrow\rangle \\ &= \langle \uparrow | 1 - \frac{\theta^2}{2} S + \dots |\uparrow\rangle = S \cos \theta, \end{aligned}$$

因此我们有

$$\int_0^\beta d\tau \langle \partial_\tau g | g \rangle = iS \int_0^\beta d\tau \partial_\tau \phi \cos \theta. \quad (16.6)$$

然后我们计算 $\langle g | H | g \rangle$ ，它基本上是 \mathbf{S} 的线性函数加上某个常数项，因为自旋是矢量，而 H 是标量，因此 \mathbf{S} 出现在 H 中的方式不是和某个另外的矢量点乘就是和自己点乘，或者是以上两者的多项式。在一个不可约表示中 $\mathbf{S}^2 = S(S+1)$ ，是常数，因此 H 是 $\mathbf{B} \cdot \mathbf{S}$ 的函数，其中 \mathbf{B} 是某个矢量。

$$\langle g | S_i | g \rangle = \langle \uparrow | e^{i\theta S_2} e^{i\phi S_3} S_i e^{-i\phi S_3} e^{-i\theta S_2} |\uparrow\rangle$$

这里我们用到了更加一般的

$$S[\theta, \phi] = S \int_0^\beta d\tau (B \cos \theta + i(1 - \cos \theta) \partial_\tau \phi). \quad (16.7)$$

最终， ψ 没有出现在路径积分中。我们可以用单位矢量 \mathbf{n} 表示 (θ, ϕ) 在球坐标系中确定的 \mathbf{S} 指向。

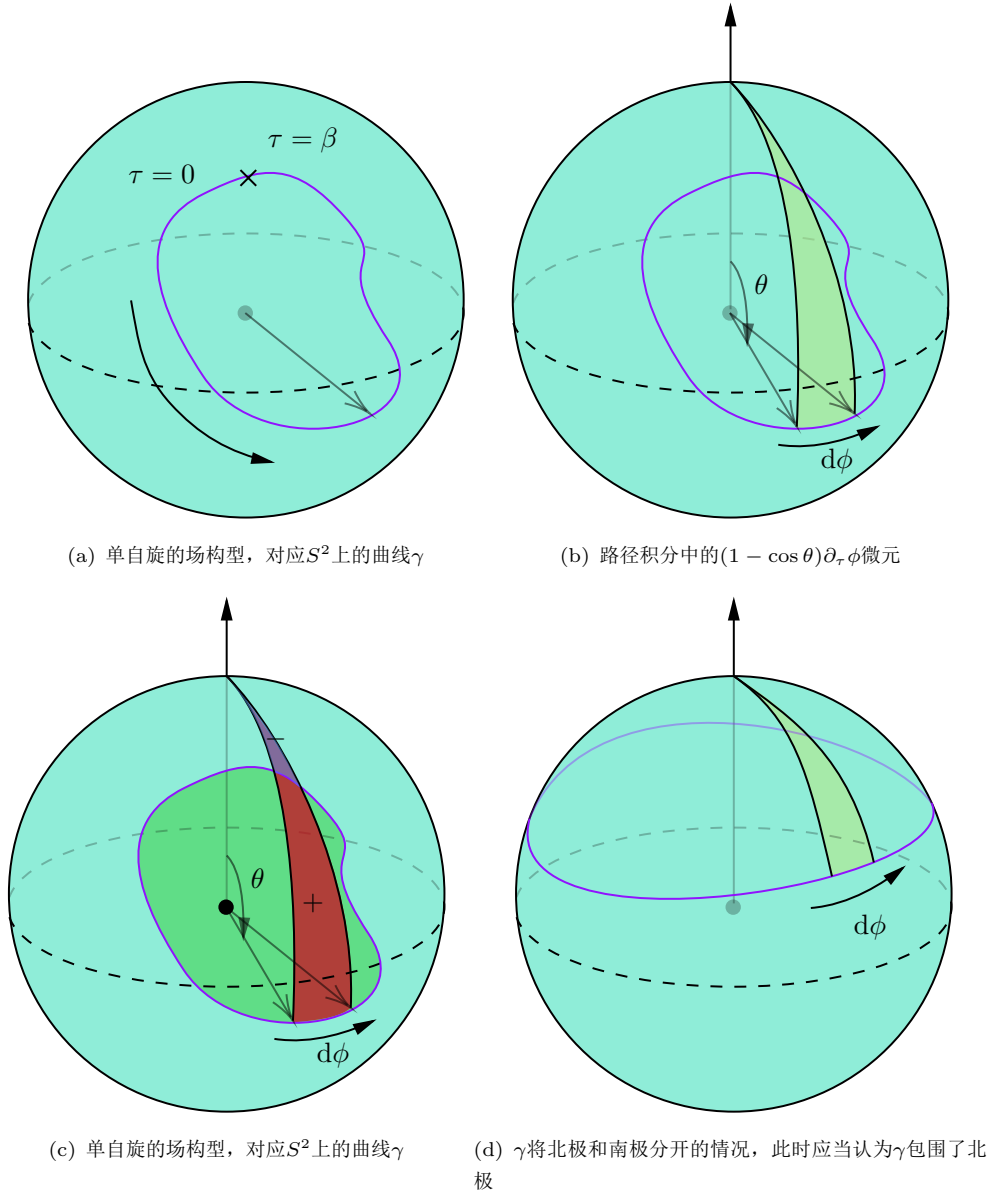


图 16.1: 单自旋的场构型和Berry相位

考虑一维海森堡自旋链

$$H = J \sum_i \mathbf{S}_i \cdot \mathbf{S}_{i+1}, \quad (16.8)$$

我们主要将考虑反铁磁的情况，反铁磁情况下 $J > 0$ ，否则 $J = -|J| < 0$ 。其自旋路径积分表述为（下面注意区分作用量 S 和自旋长度 S ）

$$Z = \int \prod_i \mathcal{D}\mathbf{n}_i e^{-S[\{\mathbf{n}_i\}]}, \quad S[\{\mathbf{n}_i\}] = \int d\tau \sum_i (iS(1 - \cos \theta_i) \partial_\tau \phi_i + JS^2 \mathbf{n}_i \cdot \mathbf{n}_{i+1}). \quad (16.9)$$

自旋系统的路径积分中的Berry相位项实际上具有特殊的意义。对单独的一个自旋，其场构型的取值范围就是 $\mathbb{R}/\beta \rightarrow S^2$ ， \mathbb{R}/β 指的是虚时间，而 S^2 则是一个特定的虚时间点的 \mathbf{n} 取值范围。因此，一个特定的场构型画出来就是Bloch球上的一个有方向的闭合曲线，可以记作 γ 。注意到立体角为

$$d\Omega = \sin \theta d\theta d\varphi,$$

半径为1的球面上，画出连接一小段路径 $d\mathbf{r}$ 上的某一点 \mathbf{r} 和北极的大圆弧，则 \mathbf{r} 运动时这段大圆弧扫过的面积是

$$(1 - \cos \theta) d\varphi = \int_0^\theta \sin \theta' d\theta' d\varphi,$$

$d\mathbf{r}$ 的方向顺时针为负，逆时针为正。上式左边乘以 S 正好是(16.7)中的Berry相位项的微元。将这些微元求和起来，我们会得到 S^2 上的这样一块面积，它的边界就是 γ （图16.1(c)）。在南极和北极的连线和 γ 有交叉时， γ 围成的面积存在歧义，由于我们是从 $\theta = 0$ 处开始积分，即从北极开始积分，应当认为 γ 围绕着北极。设 \mathbf{n}_0 是指向北极的单位矢量，我们会发现在 γ 围绕北极时我们有

$$A_\gamma = \int d\tau (1 - \cos \theta) \partial_\tau \varphi = \int_0^1 dt \int d\tau \mathbf{n}(\tau, t) \cdot (\partial_\tau \mathbf{n} \times \partial_t \mathbf{n}), \quad (16.10)$$

其中

$$\mathbf{n}(\tau, 0) = \mathbf{n}(\tau), \quad \mathbf{n}(\tau, 1) = \mathbf{n}_0, \quad \mathbf{n}(0, t) = \mathbf{n}(\beta, t). \quad (16.11)$$

我们常常将(16.7)中的Berry相位项称为“拓扑项”，即定义

$$S_{\text{topo}}[\mathbf{n}] = iS \int_0^\beta d\tau (1 - \cos \theta) \partial_\tau \phi = iS \int_0^1 dt \int d\tau \mathbf{n}(\tau, t) \cdot (\partial_\tau \mathbf{n} \times \partial_t \mathbf{n}). \quad (16.12)$$

这么说和它和拓扑量子场论的关系有关。首先它是一个Wess-Zumino项。以下设

$$\Gamma[\mathbf{n}_i] = \int_0^\beta d\tau (1 - \cos \theta_i) \partial_\tau \phi_i \quad (16.13)$$

以简化书写。

有量子涨落的自旋链的路径积分中的一个路径是一个二维底流形到 S^2 的映射，其中一维是虚时间，另一维是离散的格点坐标，即场构型的取值范围就是 $N \times (\mathbb{R}/\beta) \rightarrow S^2$ ，在自旋非常多时，就会得到 $\mathcal{M} \times (\mathbb{R}/\beta) \rightarrow S^2$ ，其中 \mathcal{M} 是系统占据的一维空间。不做任何粗粒化时 $\mathcal{M} \times (\mathbb{R}/\beta)$ 是一系列圆圈的并集，每个圆圈是固定一个格点， τ 从0演化到 β 的路径；将空间粗粒化， $\mathcal{M} \times (\mathbb{R}/\beta)$ 是一个圆柱面，其轴是空间坐标 x 轴。然而还应该注意，我们通常会认为空间无穷远处什么过程都不会发生，从而在 $x = \pm\infty$ 处可以认为没有时间演化， τ 恒定为零，从而实际上可以将圆柱的两端—— $x = \pm 1$ 处——“封起来”，将圆柱变成 S^2 。也即，一维自旋链的底流形实际上可以看成 S^2 。容易看出 (τ, x) 坐标系正是经纬度， τ 是经度而 x 是纬度。

为了方便起见——例如，如果形成了一个反铁磁序，那么显然 $(-1)^i \mathbf{n}_i$ 要比 \mathbf{n}_i 看起来光滑很多——我们有时候会重新定义 \mathbf{n} ，做代换

$$\mathbf{n}_i \longrightarrow (-1)^i \mathbf{n}_i, \quad (16.14)$$

则

$$\mathbf{n}_i = \phi_i + (-1)^i \mathbf{m}_i. \quad (16.15)$$

在变换(16.14)下

$$\sum_i JS^2 \mathbf{n}_i \cdot \mathbf{n}_{i+1} \longrightarrow -JS^2 \sum_i \mathbf{n}_i \cdot \mathbf{n}_{i+1},$$

而

$$-JS^2 \int d\tau \sum_i \mathbf{n}_i \cdot \mathbf{n}_{i+1} = \frac{JS^2}{2} \int d\tau \sum_i (\mathbf{n}_{i+1} - \mathbf{n}_i)^2 + \text{const.}$$

Berry相位项在(16.14)作用下则是

$$\sum_i iS\Gamma[\mathbf{n}_i] = iS \sum_{\text{odd } i} (\Gamma[\mathbf{n}_{i+1}] + \Gamma[\mathbf{n}_i]) \longrightarrow iS \sum_{\text{odd } i} (\Gamma[\mathbf{n}_{i+1}] + \Gamma[-\mathbf{n}_i]),$$

通过简单的几何分析可以看出,

$$\Gamma[-\mathbf{n}] = 4\pi - \Gamma[\mathbf{n}] = -\Gamma[\mathbf{n}] + \text{const}, \quad (16.16)$$

常数 $4\pi i$ 在配分函数中不会有任何贡献。于是Berry相位项之和为

$$iS \sum_{\text{odd } i} (\Gamma[\mathbf{n}_{i+1}] - \Gamma[\mathbf{n}_i]),$$

即它是Bloch球上的曲线 $\mathbf{n}_{i+1} : S^2 \rightarrow S^2$ 围绕Bloch球北极的面积减去 $\mathbf{n}_i : S^2 \rightarrow S^2$ 围绕Bloch球北极的面积。因此在变换(16.14)下, \mathbf{n} 的作用量为

$$S = iS \int d\tau \sum_{\text{odd } i} (\Gamma[\mathbf{n}_{i+1}] - \Gamma[\mathbf{n}_i]) + \frac{JS^2}{2} \int d\tau \sum_i (\mathbf{n}_{i+1} - \mathbf{n}_i)^2. \quad (16.17)$$

如果将 \mathbf{n} 场构型朴素地连续化, 就有

$$- \int d\tau \sum_i JS^2 \mathbf{n}_i \cdot \mathbf{n}_{i+1} = JS^2 a \int d\tau \int (\partial_x \mathbf{n})^2,$$

而根据几何意义, Berry相位项做连续化后得到

$$iS \sum_{\text{odd } i} (\Gamma[\mathbf{n}_{i+1}] - \Gamma[\mathbf{n}_i]) \longrightarrow iS \frac{1}{2a} \int dx \mathbf{n} \cdot (\partial_x \mathbf{n} \times \partial_\tau \mathbf{n}).$$

但是我们不能这样做, 因为如果有铁磁序, 那么 \mathbf{n} 中包含在空间维度上快速变化的铁磁序参量, 它的空间导数并不光滑, 因此不能直接做连续化。不过, 我们至少可以写出

$$iS \sum_{\text{odd } i} (\Gamma[\mathbf{n}_{i+1}] - \Gamma[\mathbf{n}_i]) \longrightarrow iS \sum_{\text{odd } i} \mathbf{n}_i \cdot ((\mathbf{n}_{i+1} - \mathbf{n}_i) \times \partial_\tau \mathbf{n}_i). \quad (16.18)$$

这个形式看起来非常像Wess-Zunimo-Witten (WSW) 项——我们在讨论后面的反铁磁自旋链的行为的时候会看到它的确就是, 这也就是它被称为拓扑项的原因。

16.1.2 铁磁情况

先考虑铁磁情况, 即

$$J = -|J| < 0 \quad (16.19)$$

的情况, 此时作用量为

$$S = \quad (16.20)$$

16.1.3 反铁磁情况

假定一个反铁磁序能够生成。我们可以做分解

$$\mathbf{n}_i = (-1)^i \phi_i + \mathbf{m}_i, \quad (16.21)$$

其中 ϕ_i 是反铁磁序参量而 \mathbf{m}_i 是铁磁序参量, 后者远远小于前者。请注意此时 ϕ_i 和 \mathbf{m}_i 的具体定义并没有唯一确定——我们只需要 ϕ_i 远大于 \mathbf{m}_i , 且两者在晶格常数 $a \rightarrow 0$ 时形成的

连续场足够光滑即可。由于 $|\mathbf{n}_i| = 1$ 总是成立而 \mathbf{n}_i 中 ϕ_i 占据压倒性地位，假定 $|\phi| = 1$ 恒成立总是可以的：此时 ϕ_i 确实远大于 \mathbf{m}_i 。

我们将(16.15)代入(16.17)，则

$$\begin{aligned} iS \sum_{\text{odd } i} (\Gamma[\mathbf{n}_{i+1}] - \Gamma[\mathbf{n}_i]) &\approx iS \sum_{\text{odd } i} \mathbf{n}_i \cdot ((\phi_{i+1} - \phi_i - 2\mathbf{m}_i) \times \partial_\tau \mathbf{n}_i) \\ &\approx iS \sum_{\text{odd } i} \phi_i \cdot ((\phi_{i+1} - \phi_i - 2\mathbf{m}_i) \times \partial_\tau \phi_i) \\ &\approx iS \frac{1}{2a} \int dx \phi \cdot ((\partial_x \phi - 2\mathbf{m}) \times \partial_\tau \phi), \end{aligned}$$

第一个约等号来自 \mathbf{m}_i 本身变化不很迅速的假设，第二个等号是小量近似，即 \mathbf{m}_i 和 $\partial_x \phi$ 都很小，从而只保留一阶小量，即丢弃 \mathbf{n} 和 $\partial_\tau \mathbf{n}$ 因子中的 \mathbf{m} 成分。另一方面，有

$$\sum_i (\mathbf{n}_{i+1} - \mathbf{n}_i)^2 =$$

综上，反铁磁一维海森堡自旋链的低能有效理论是

$$S = i \underbrace{\frac{S}{2} \int dx d\tau \phi \cdot (\partial_x \phi \times \partial_\tau \phi)}_{S_{\text{topo}}} + \underbrace{\int dx d\tau (\partial_x \phi)^2}_{S_0}, \quad (16.22)$$

其中 S_{topo} 称为**Wess-Zunimo-Witten (WSW) 项**，它就是(16.12)对所有自旋自由度求和的结果，且正好给出 $\phi: S^2 \rightarrow S^2$ “覆盖底流形 S^2 的面积”。注意到

$$\pi_2(S^2) = \mathbb{Z},$$

S_{topo} 实际上将不同的场构型按照第二同伦群划分成了不同的同伦等价类。这就是我们将(16.12)称为“拓扑项”的原因：它对所有自旋自由度求和之后真的就是场构型的一个拓扑不变量。设场构型 ϕ 在 $\pi_2(S^2)$ 中被分类到了 W 上，或者说它的卷绕数为 W ，或者说它覆盖底流形 S^2 的面积是 $4\pi W$ ，则对这样的构型

$$S_{\text{topo}} = 2\pi S W i. \quad (16.23)$$

由此可以看出，对整数自旋的反铁磁海森堡自旋链，拓扑项是平庸的：虽然 S_{topo} 能够给出场构型在 $\pi_2(S^2)$ 中的分类，但是它不会影响场构型的权重。但半整数自旋的反铁磁海森堡自旋链中的拓扑项是非平庸的，这类系统的低能有效理论是

$$Z = \int \mathcal{D}\phi (-1)^W e^{-S_0[\phi]}. \quad (16.24)$$

背景知识 16.2: 拓扑量子场论和同伦群

设 M 是底流形， T 是目标空间（即底流形上某一点处场值的取值范围），两者均为拓扑空间，一个场构型就是一个 $M \rightarrow T$ 的连续函数。一个物理理论可以写成这样的形式：

$$Z = \int_{\phi \in (M \rightarrow T)} \mathcal{D}\phi e^{i \int_{x \in M} dx \mathcal{L}[\phi]}.$$

大部分情况下， $S[\phi]$ 在场构型 ϕ 有小的局域变动时也会跟着变动，从而给出非平凡的运动方程。不过有一些特殊的 $S[\phi]$ 是场构型 ϕ 的拓扑不变量，场构型发生小的变动，它是不变的。这样的作用量无法给出非平庸的动力学，但是它能够给按照它划分的不同拓扑等价类赋予不同的权重，从而仍然有物理意义。这种实际上是在计算拓扑

不变量的量子场论称为**拓扑量子场论**。

最为“细致”的拓扑分类当然是同胚，但是使用一个形式很简洁的拓扑不变量大抵是无法做到这么细致的分类的。拓扑量子场论中的等价类通常都精确不到同胚的级别。一种常见的等价类划分方式为**同伦**。设 ϕ_1, ϕ_2 是两个 $M \rightarrow T$ 的场构型，如果存在连续映射

$$f : [0, 1] \times M \rightarrow T$$

使得 $f(0) = \phi_1, f(1) = \phi_2$ ，我们说 ϕ_1 和 ϕ_2 同伦。直观地看，如果能够有一个超然于 M 以外的连续时间演化过程能够将 ϕ_1 转化为 ϕ_2 ，那么两者是同伦的。如果 ϕ_1 和 ϕ_2 同伦，那么它们一定同胚，但是反之不然。例如，字母X和Y是同伦的，但是它们不同胚。同伦允许“高维对象缩成更低维的”等等，而同胚则有更多的限制。依据同伦的概念可以定义**同伦群**， M 的 n 阶同伦群 $\pi_n(M)$ 是 S^n 到 M 的保持某个基点不动的映射的同伦分类。其群乘法通过

$$(\phi_1 * \phi_2)(x_1, x_2, \dots, x_d) = \begin{cases} \phi_1(2x_1, x_2, \dots, x_d), & x_1 \in [0, 1/2], \\ \phi_2(2x_1 - 1, x_2, \dots, x_d), & x_1 \in [1/2, 1] \end{cases} \quad (16.25)$$

定义。

一个实际模型的有效场论中可能会有一个拓扑项。常见的情况是这个拓扑项可以分类场构型 ϕ 在某个同伦群中的同伦类，而这个同伦群是交换群，从而可以看成一些整数组成的，而群乘法为加法。设 W 为 ϕ 的同伦类，则能够找到某个函数使得

$$S_{\text{topo}}[\phi] = F(W). \quad (16.26)$$

特别的，如果 W 是某种局域的“拓扑激发”（如KT相变中的涡旋）的个数，那么必定有

$$F(W_1 + W_2) = F(W_1) + F(W_2), \quad (16.27)$$

因为设 ϕ_1 和 ϕ_2 携带了两个相隔甚远的拓扑激发，则

$$S_{\text{topo}}[\phi_1 * \phi_2] = F(W_1 + W_2),$$

而另一方面，由于两个拓扑激发相隔甚远，它们彼此没有相互作用，从而作用量具有可加性，于是我们有

$$S_{\text{topo}}[\phi_1 * \phi_2] = S_{\text{topo}}[\phi_1] + S_{\text{topo}}[\phi_2].$$

一个最为简单的拓扑激发的例子可以通过考虑约束在一个环上的粒子加于磁场中这样一个0+1维模型来得到。将该粒子热化，我们有

$$S = iA \int_0^\beta d\tau \dot{\phi}. \quad (16.28)$$

马上可以发现这个作用量计算的是粒子路径 ϕ 在虚时间演化中绕着 S^1 转的绕数。图16.2直观地展示了绕数对应的拓扑激发。在这个例子中，拓扑激发具有时间方向的结构，即这些拓扑激发和一般的使用场的空间构型即可完全定义的粒子并不一样。但是从路径积分的角度，这个区别并不重要。

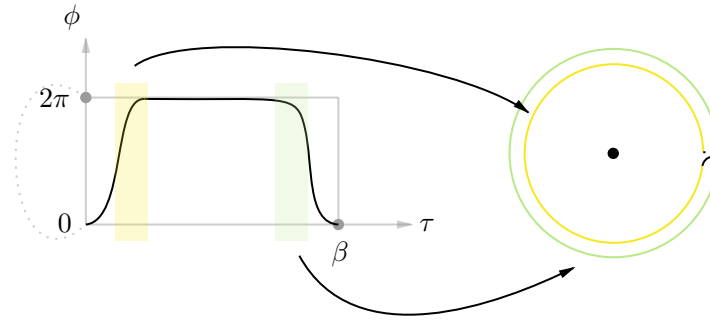


图 16.2: 磁场中的圆环上粒子的场构型中的拓扑激发。 $\phi = 0$ 和 $\phi = 2\pi$ 两点认同。图中, 黄色和绿色分别展示了绕数为1和-1的两个拓扑激发。绕数为1的拓扑激发指的就是 ϕ 快速地从0变动到 2π , 由于 $\phi = 0$ 和 $\phi = 2\pi$ 两点认同, 这相当于绕着粒子所在的圆环转了一圈(用右图中的黄色圆弧表示)。绕数为-1的拓扑激发正好相反。最终, 整个场构型的整体绕数为零。

可以看到, 一维海森堡自旋链中的量子涨落实际上是非常强烈的, 以至于在一维实际上根本无法形成反铁磁序。如果 S 是整数则拓扑项无贡献, 有能隙; Haldane猜想。

16.2 ALKT链

我们将两个相邻格点的自旋处在一个自旋单态中这件事称为一个**共价键** (valence bond)。有这个说法是因为, 虽然我们这里在研究自旋模型, 没有轨道自由度, 也没有成键、反键, 但是实际系统中, 如果两个电子形成化学键, 它们的轨道部分应该倾向于是对称的 (否则不存在电子云交叠), 从而自旋部分应该是反对称的, 即自旋形成自旋单态。因此, 如果一个自旋模型中的自旋自由度可以直接追溯到单个的电子上, 那么相邻格点处的自旋处在自旋单态中就意味着很可能实现这个自旋模型的实际体系中在这对格点处存在共价键。

The theorem and the boundary of a topological state look quite alike, and this is not coincidence.

spin-1/2 edge of a spin AKLT chain

第六部分

（内蕴）拓扑序和量子序

前面几章中的拓扑物态中不存在特别超出能带电子、自旋这些东西的“纯物理”。这些系统中均只存在短程纠缠，或者说总是可以作用一些局域的扰动，让系统基态回到没什么纠缠的状态。本节讨论的是一些由于强关联效应有长程纠缠的系统。具体来说，我们希望研究那些在零温下存在一个具有长程纠缠的量子相，其低能自由度仍然是点状激发的系统。长程纠缠意味着点状激发的产生、移动、湮灭可以以某种方式记录在系统状态中，且虽然低能自由度是点状激发，但是它们未必能够使用局域的算符单独地产生和消灭。

一般的金斯堡-朗道理论中演生出来的激发就是序参量的涨落，基本上是无能隙、玻色型、无自旋的激发，而构成凝聚态系统的是巡游电子和离子实。这提示我们一些有序态中能够演生出来与组成它的成分非常不同的激发。但是金斯堡-朗道理论中这种演生粒子的类型是很有限的：基本上都是电子配对或是自旋序的涨落形成的玻色子。然而我们会发现拓扑序中的演生粒子是非常丰富多样的。

第十七章 分数量子霍尔效应

17.1 $1/m$ 型分数量子霍尔效应

Laughlin论证——以及有关的各种基于单电子图像的计算——只能够得到整数量子霍尔效应。显然，唯一的可能是，分数量子霍尔效应是电子之间的相互作用产生的，并且只能发生在强关联系统中。本节讨论

$$\nu = \frac{1}{m}, \quad m = 1, 3, 5, \dots$$

型的分数量子霍尔效应，这是分数量子霍尔效应中比较简单的一种情况。

17.1.1 Laughlin波函数，Laughlin液体和其中的任意子

直接从库伦排斥出发做场论计算分析强关联效应显然是不现实的。在第十章中我们通常是通过将重要的相互作用通道挑选出来，构造特定的模型研究强关联效应，但分数量子霍尔效应的研究实际上走了一条相当不同的道路。Laughlin通过其天才的创造，一步到位地给出了能产生 $1/m$ 分数霍尔效应的系统的基态波函数：

$$\Psi_m(z_1, \dots, z_N) = \prod_{i < j} (z_i - z_j)^m e^{-\sum_i |z_i|^2 / 4l_0^2}, \quad m = 1, 3, 5, \dots \quad (17.1)$$

容易验证以上波函数满足交换反对称性；当 z_i 趋于 z_j 时波函数趋于零，这是多电子波函数的必然要求。如果 m 是偶数，那么上式就不再是一个好的电子多体波函数了，因为交换反对称性没有。这个波函数是怎么猜出来的还是有一些线索的：它实际上是推广了整数量子霍尔效应的多体波函数。

Laughlin波函数虽然是猜测出来的，不过将它和数值计算出来的波函数相对比，两者通常吻合得比较好。因此，作为讨论分数量子霍尔效应的起点它是够格的。相应的，基态为Laughlin波函数的分数量子霍尔效应系统可以称为Laughlin液体，它的名字称为液体是因为其中相互作用很重要，但是没有任何对称性自发破缺。例如，可以验证电子数密度是均一的。我们下面就来分析Laughlin波函数的一些性质。

与无动能的等离子体系统的对应

Laughlin波函数的一个相当值得注意的地方是它能够和一个无动能（从而，无量子涨落）的等离子体系统对应。考虑多粒子联合概率分布

$$p(z_1, z_2, \dots, z_N) = |\Psi(z_1, \dots, z_N)|^2, \quad (17.2)$$

则有

$$p(z_1, \dots, z_N) = e^{-\beta U(\{z_i\})}, \quad (17.3)$$

其中

$$\beta = \frac{2}{m}, \quad U(z_1, \dots, z_N) = -m^2 \sum_{i < j} \ln |z_i - z_j| + \frac{m}{4l_0^2} \sum_i |z_i|^2. \quad (17.4)$$

因此零温Laughlin波函数和一个逆温为 m ，能量为 U 的无动能等离子体系统对应。¹

U 的第一项的意义是显然的：二维平面上不同粒子之间的库伦排斥。其第二项

Laughlin液体中的准空穴和准粒子

由于我们是先有了Laughlin波函数基态波函数而并不知道怎么样的哈密顿量（或者说怎么样的晶格）才能够产生Laughlin波函数，实际上无法确切地知道Laughlin液体中的准粒子是什么样的。我们只能冒险猜测一些看起来还像回事的准粒子和准空穴定义确实是实际上的分数量子霍尔效应系统中的准粒子。因此，本节我们将“生造”出一些看起来像是粒子和空穴的构造，并且分析它们的性质。

在位置 ξ 处有一个准空穴的波函数定义为

$$\langle z_1, \dots, z_n | \Psi^h(\xi) \rangle = \sqrt{C(\xi)} \prod_i (\xi - z_i) \Psi_m. \quad (17.5)$$

这里 $C(\xi)$ 是归一化因子；我们相信这是正确的准空穴波函数，是因为在 ξ 处电荷密度为零，因此看起来这里确实存在一个空穴。

从 $|\Psi\rangle$ 中移走一个电子得到的波函数是什么样的？由于其它电子都没有任何变化，移走位置在 ξ 处的电子，得到的波函数应该是

$$\langle z_1, \dots, z_{N-1} | \Psi_m \text{ removed one electron at } \xi \rangle \propto \sum_{i=1}^{N-1} (\xi - z_i)^m \langle z_1, \dots, z_{N-1} | \Psi_m \text{ with } N-1 \text{ electrons} \rangle, \quad (17.6)$$

上式实际上就是将(17.1)中的一个电子坐标替换成 ξ 得到的结果。然后我们就发现了这里有一些不同寻常的地方：移走一个电子相当于创生了 m 个准空穴——换句话说，一个准空穴等于 $1/m$ 个电子空穴。

这真是匪夷所思。朗道费米液体中的准粒子就是经过修饰的电子；我们在第八章里面讨论的有序相中的最低能的准粒子都是两个或多个电子凝聚而成的；在相对“正常”一些的强大关联系统——如Luttinger液体中，也顶多存在自旋-电荷分离，怎么到了这里一个电荷还能够被拆分成几片？这显然是一种分数化现象，只不过和Luttinger液体相比更加不寻常：Luttinger液体中的分数化只是自旋标签和电荷标签分离，但是在这里，原本的单独的一个电荷标签也被拆开了。

背景知识 17.1: 任意子和任意子统计

费米子和玻色子的概念可以通过自由场的量子化直接得到。不过我们也不妨稍稍推广一下“点粒子”的定义：如果一个格点系统（连续空间可以先离散化）满足如下条件：

- 希尔伯特空间的某一组基矢量可以通过在每个格点上放置一个自然数 n_i （所谓粒子数）标记。
- 哈密顿量中有若干项跃迁项 t_{ij} ，其中 i 和 j 都是格点。跃迁项需要满足如下条件：

¹这不是一个很奇怪的操作——我们在DFT中使用分数占据时实际上就是这么处理。

- 如果状态 $|\psi\rangle$ 中 i 点没有粒子而 j 点有一个粒子，那么 t_{ij} 算符作用在 $|\psi\rangle$ 上，会产生一个 i 点有一个粒子而 j 点没有粒子的态，乘上一个相位因子；该相位因子未必只和 i 和 j 有关，而是可以依赖于全局的信息。
- 如果 j 点根本没有粒子，那么 t_{ij} 作用到系统状态上得到零，或者说 t_{ij} 消灭 j 点没有粒子的态。

- 如果 i, j, k, l 彼此不相同，则

$$[t_{ij}, t_{kl}] = 0. \quad (17.7)$$

这是为了保证哈密顿量的局域性。

那么就可以说这个系统是一个粒子系统。这几条假设捕捉了“粒子”的直觉概念的大部分。例如，弦算符

$$t_{i_1 i_2} t_{i_2 i_3} \cdots t_{i_{n-1} i_n}$$

将一个有一个粒子在 i_n 点，没有粒子在 i_1 点的状态转化为一个有一个粒子在 i_1 点而没有粒子在 i_n 点的状态乘上一个可能是全局性的复因子。

这里要注意一个地方：实际上一种粒子的统计是由系统哈密顿量确定的，而不是由希尔伯特空间。硬球玻色子同样有不相容性，而费米子大抵也可以通过粗粒化而近似地认为“多个费米子处于同一状态”。但是，产生湮灭算符的对易和反对易性是确定的，而系统哈密顿量由费米子或玻色子场算符组成，因此是哈密顿量决定了粒子的统计，或者更加准确地说是哈密顿量中的跃迁项决定了粒子的统计。这马上就会产生一个问题，就是如果哈密顿量可以拆分出两种不同的跃迁项怎么办——但是这实际上没有任何问题的，例如我们都知道产生超导的是一个电子系统（费米子系统），但是其低能激发是库伯对（玻色子系统）。

这些假设留下了很多变数。我们没有给出“粒子如何被产生”：实际上，对一些系统可能无法良定义何为“粒子产生”，也没有能够在局部激发出粒子的场算符。（这实际上是一件好事，因为这样一来，我们只需要通过某种方式确定系统的能量本征态可以使用某些“粒子”标记，就可以使用和普通的量子场论中非常类似的方法研究系统的行为，至于这些粒子是演生激发，拓扑缺陷还是别的什么很多时候不重要）我们也没有讨论粒子除了坐标以外还能有什么标签。在有多种粒子时，不同粒子的跃迁算符之间是什么关系也没有明确规定——没有什么规定它们一定要对易。

我们现在来分类可能的跃迁项，或者说分析粒子能够有怎样的统计。首先注意跃迁项不改变系统中的总粒子数，从而可以单独分析含有 n 个粒子的系统中粒子的统计。含有 n 个粒子的系统中所有粒子位置确定时，系统构型的取值范围为 $\otimes_n \mathbb{R}^d = \mathbb{R}^{nd}$ 。为了简化分析，我们假定每个空间点上同种类型的粒子只能有一个；如果有多个粒子，就讨论一个对偶的模型，在其中每个空间点被扩充为一个“停车场”，停车场内部可以有多个粒子，但是它们占据的空间点还是不一样的。这样，系统构型的取值范围为 \mathbb{R}^{nd} 去掉两个同种粒子占据同一个空间点的情况后剩下的 \mathbb{Y}^{nd} 。由于前述定义中粒子没有被编号，即 \mathbb{R}^{nd} 中的 i 号位置坐标为 \mathbf{r}_i ， j 号位置坐标为 \mathbf{r}_j 和 i 号位置坐标为 \mathbf{r}_j ， j 号位置坐标为 \mathbf{r}_i 完全一样，只要粒子 i 和 j 是同种粒子，最终的系统构型取值范围为 $X = \mathbb{Y}^{nd}/\mathcal{P}$ ，其中 \mathcal{P} 表示任意一个同种粒子的坐标重新排列。于是系统的希尔伯特空间的一组基矢量的标签可以取为 X ，希尔伯特空间为 \mathbb{C}^X 。

现在考虑一系列圈状的弦算符的乘积。经典的 X 中的构型在这些弦算符代表的粒子位移操作下保持不变，因此，量子的 \mathbb{C}^X 应当携带这种“绕了一圈又绕回来了”的操作——或者说“交换”（注意这和 \mathcal{P} 这种直接交换 \mathbf{r}_i 和 \mathbf{r}_j 不一样，交换必须要在 X 空间中真的绕上一圈）——的一个表示。或者更清楚的说， \mathbb{C}^X 应该携带 X 的基本群的一个表示。这个表示可以是一维的，此时两个任意子交换只会让态矢量多出来一个复因子，那么我们称涉及其中的任意子是阿贝尔的；这个表示也可以是高维的，此时涉及其中的任意子是非阿贝尔的。

$d \geq 3$ 时能够证明 \mathbb{Y}^{nd} 单连通，此时 X 的基本群同构于 \mathcal{P} ，是不同种类的粒子的置换群的直积。置换群的一维表示要么是平凡的，要么偶置换取1而奇置换取-1，于是 $d \geq 3$ 时阿贝尔任意子只有玻色子和费米子，或者有时候说没有阿贝尔任意子，虽然原则上可能有非阿贝尔任意子。

$d = 2$ 时 X 的基本群同构于辫子群，于是可以有非平凡的、但是也不是 ± 1 的阿贝尔任意子。这就是为什么我们时常在研究二维系统时看到任意子。我们前面都是假定粒子的坐标的取值范围为 \mathbb{R} 。但是实际上，粒子也可以在一个拓扑非平庸的空间中运行。因此任意子统计实际上能够“看到”空间的拓扑特性，因为它的可能选择由 X 的基本群决定。直觉上这是非常合理的，因为弦算符不是局域的算符，因此它当然能够探测空间的拓扑性质。

除了交换相位未必是 ± 1 以外，还应该注意到一个弦算符是闭环，它未必等价于恒等操作。这在背景知识17.3中很重要。

17.1.2 Chern-Simons理论

背景知识 17.2: 任意子和规范场论

任意子存在场论表述，具体方法是通过构造特定的规范场，通过A-B效应的方式引入额外的相位。这个规范场本身没有非平庸的动力学，然后通过相位 $\int d\mathbf{r} \cdot \mathbf{A}$ ，一个粒子可以“看到”被它的路径环绕的其它粒子，从而将这样一个规范场和某种费米子或是玻色子耦合，就得到了任意子。

描述一个自由任意子的规范场论不应该有自己的动力学，否则任意子交换产生的相位就包含动力学相位了。显然，该规范场自己的拉氏量通常是一个拓扑量子场论。该规范场最好也有能隙。会能隙与否会影响与该规范场耦合的粒子的行为。例如cyon是一种和电磁场耦合而带有非平庸交换相位的粒子，每个cyon带有和电磁场耦合而形成的电荷标签，同时也带有每个粒子产生的磁通量。由于电磁场无能隙，对cyon，flux看得见charge，从而两个cyon交换以后的相位是一个电荷绕着一块磁通量转动一圈的相位的两倍。但如果规范场有能隙，规范荷产生的场的小幅扰动在远处快速衰减，flux看不见charge，从而不会出现前述相位翻倍的现象。（这是^[27]中给出的一个例子）

背景知识 17.3: 任意子和张量范畴

二维系统中的任意子可以使用modular category描述。关于本节的各种定义的严格表述、交换图等、范畴论除了量子力学以外的其它例子等可以看^[6]，关于任意子的部

分^[12]中A categorical view of computing with anyons一章也提供了容易读懂的介绍。以彼此能够通过某些时间演化过程相互转化的量子系统为对象，以它们之间的时间演化过程为态射，可以组成一个范畴，具体来说是一个compact dagger monoidal category，其中monoidal代表某种范畴论意义下的直积存在，compact是确保投影算符 $|\psi\rangle\langle\psi|$ 可以良定义的，dagger的意思就是共轭转置的代数结构。需注意正如费米子和玻色子多体波函数所示的那样，该compact dagger monoidal category中的张量积——将两个粒子放在一起这个操作——未必是线性代数中的张量积。

通过数自由度我们承认系统 $A \otimes B$ 和 $B \otimes A$ 之间应该具有同构，即总是存在同构 $b_{AB} : A \otimes B \rightarrow B \otimes A$ ，但是一般来说，我们只能确信任意子系统是braided compact dagger monoidal category，而未必是symmetric compact dagger monoidal category。两者的区别在于，后者要求 $b_{AB} \otimes b_{BA} = \text{id}$ ，前者不需要；使用图形表示（所谓的图形演算），就是后者无需区分braiding操作 b_{AB} 的图形表示中， A 的轨迹和 B 的轨迹的上下叠放顺序，而前者需要区分。三维及以上的粒子由于是置换群的表示，因此由symmetric的范畴描述，没有非平庸的braiding结构；它们要么是费米子要么是玻色子。

描述任意子的范畴还必须将任意子的有自相交的路径和没有自相交的路径区分开——或者说，需要将闭弦算符并非恒等算符这件事考虑在内。有自相交的路径上多了一个“自己绕一圈”的操作，这个操作必须服从特定的相容性条件，直观地说即满足以上条件的范畴称为ribbon category，因为以上结构图形化地展示，就是一个丝带被扭曲的过程。

我们还要求任意两个对象之间的态射构成复数域上的向量空间，态射的张量积是双线性的等等。在compact dagger monoidal category中实际上已经能够定义一些态射的叠加等操作，不过我们要求它们构成复数域上的向量空间，从而确保任意子范畴是一个量子力学理论而不是别的什么框架下的理论。

由于种类确定的任意子除了坐标以外找不出更多的自由度，单任意子态是simple objects：它们的自态射一定是只能是 $k \text{id}, k \in \mathbb{C}$ 的形式，且两个不同的simple object之间没有态射。我们要求任意子范畴是semisimple的。这里semisimple的定义中的“直和”的物理意义略有些不清楚，不过可以做如下理解：设 ϕ_1 和 ϕ_2 是两种任意子，则 $\phi_1 \otimes \phi_2$ 同构于一个二维的希尔伯特空间，其中的波函数是两个单任意子态的线性叠加。另一方面， $\phi_1 \otimes \phi_1$ 同样对应一个二维的希尔伯特空间，其中的波函数中任意子数目确定只有一个，但是 ϕ_1 有“两种存在方式”：可能有一个额外的标签能够区分 $\phi_1 \otimes \phi_1$ 的两个基矢量。这样理解的话，任意子范畴是semisimple的意思实际上就是两个任意子放在一起，远看仍然是一个任意子，只不过这个任意子的类型是不确定的。这意味着我们可以有所谓的融合规则：对任意两种任意子 i 和 j ，存在一系列自然数（因为希尔伯特空间的维数总是正整数，而 N_{ij}^k 为零表示 i 粒子和 j 粒子融合产生不了 k 粒子） $\{N_{ij}^k\}$ 满足

$$i \otimes j = \sum_k N_{ij}^k k, \quad N_{ij}^k = \dim \text{Hom}(i \otimes j, k). \quad (17.8)$$

我们还需要引入迹的定义，这是比较直截了当的。

最后一个需要的元素是 S 矩阵——不是散射振幅的 S 矩阵而是CFT中的。

满足以上条件的范畴称为modular tensor category (MTC)。任意子体系中的

各种状态和它们之间的变换原则上都可以纳入这样一个范畴中。有有限个同构等价类的MTC也可以称为braided fusion category。需注意我们其实可以有并非braided的fusion category：有张量积、服从量子力学的基本假设（即前面所说的态射构成 C 上的向量空间等等）、semisimple的范畴就能够有fusion rules。

对称性破缺后的相残余的对称性之于金斯堡-朗道理论就好比张量范畴之于拓扑序。不过，给定了序参量的对称性，就能写下金斯堡-朗道理论，然而和金斯堡-朗道理论不同，单单知道了任意子的融合规则、braiding结构等信息并不能确定任意子的动力学。但是通常我们不试图得到任意子的动力学，一方面因为这很困难——在拓扑序中，怎么做出类似于金斯堡-朗道理论中“某某对称性被破缺了”的有序相分类本身已经非常非平庸了，一方面因为没有必要——我们写出任意子只是为了描述拓扑序的普适性质而暂时不关心相变时发生了什么又或者解禁闭相如何响应外加扰动。实际上，由于任意子激发态和基态之间有能隙，加入小的扰动并不会造成相变，从而，如果两个不同的模型均有有能隙的拓扑序，且具有一样的任意子交换相位、fusion rule等，那么可以缓慢地向哈密顿量加入扰动，让两个模型的拓扑序长得完全一样而不造成相变。因此，任意子的动力学其实并不那么重要。同样和金斯堡-朗道理论类比，任意子的动力学如同不相关算符，在重整化群（对金斯堡-朗道理论就是普通的Wilson重整化群，对拓扑序则更为自由，任何向哈密顿量引入小幅局域变化——从而将局域的细节抹平——的做法都是可以的）作用下最终会消失。

17.2 层级分数量子霍尔效应

第十八章 自旋液体和格点规范场

分数量子霍尔效应出现之后显然不出意外地引起了巨大的反响。很自然的想法是寻找更多具有拓扑序的系统，并且，通过研究“干净”一些的系统，或许可以分析出拓扑序的一些机制。自旋液体是一类展现了拓扑序的自旋系统，它们可能对理解拓扑序中的概念有帮助，或者是有可能展现拓扑序的系统的演生理论。

大部分自旋系统（如第十二章中展示的那些）在低温下都会落到铁磁序或反铁磁序中，因为零温时没有热涨落，而相互作用总是存在的，因此系统倾向于有序。然而，如果零温时有特别强的量子涨落，可能系统基态中不存在这种类型的序，其中没有出现任何对称性自发破缺，无法定义序参量。这样的自旋系统称为**自旋液体**，因为它们的基态是无序的，正如液体之于固体一样。自旋液体的出现可能是因为在零温下由于一些阻挫（frustration，即，让晶格取特定的形式使得反铁磁序无法形成）或者别的原因。阻挫未必导致自旋液体，而也有自旋液体完全没有阻挫（比如Toric-code模型；很多人实际上还更喜欢这种情况，因为往往可以对应到某种严格可解模型上）。自旋液体如何能够形成实际上仍然是不太确定的。需注意自旋液体中没有对称性自发破缺，但是这并不代表自旋液体中没有其它类型的序和相变，例如，它们完全可以有拓扑序。因此，自旋液体是一个非常有趣的状态，其制备方法以及性质都很引人注目。

虽然自旋液体似乎可以归结在磁性材料的范畴内，强烈的量子涨落实际上意味着自旋液体中有一些一般的磁性材料不会有的丰富行为。在真正的自旋液体中我们会观察到演生规范场和拓扑序——实际上，自旋液体是除了分数量子霍尔效应以外仅有的已知的在实验上有可能实现的拓扑序。分数量子霍尔效应确定有拓扑序，而目前没有确定无疑是自旋液体的材料，其它的拓扑序模型都是人为构造的。

资料框 18.1: 寻找自旋液体的尝试

目前没有人找到确定无疑是自旋液体的材料。1973年，P.W.Anderson考虑了一个三角晶格上的反铁磁模型，来给反铁磁序的形成制造一些阻挫，因为三角晶格上显然无法形成反铁磁序（见图18.1）。他猜测其基态为将晶格上最近邻自旋配对后将两个

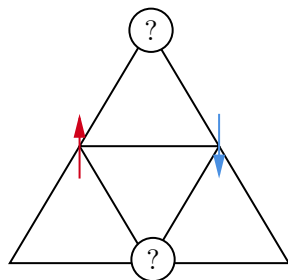


图 18.1: 三角晶格上的阻挫：无法适当安排自旋方向让相邻自旋反向

自旋自由度做分解

$$\frac{1}{2} \otimes \frac{1}{2} = 0 \oplus 1,$$

取所有可能的配对中的单态等权叠加的结果。这个状态称为**RVB(Resonance Valence Bond)**态。如果实际上基态真的是RVB态，那么显然基态上没有形成任何磁性序，并且会有一些和磁性序上的“扰动”截然不同的激发（这些后文会详述）。事实证明这个说法是错误的：三角晶格上的海森堡模型的基态是一种特殊的铁磁态。虽然如此，自旋液体仍然是一个非常有趣的状态，因为，这可能是由于，对称性破缺不能发生。一些有机盐被认为有可能产生自旋液体，因为它们做AMR实验观察不到任何磁性序，但始终没有定论；不少这种候选的自旋液体都被其它实验证实并非自旋液体了。

18.1 各向同性海森堡模型中的自旋液体态

各向同性海森堡模型是

$$H = J \sum_{\langle i,j \rangle} \mathbf{S}_i \cdot \mathbf{S}_j. \quad (18.1)$$

我们没有指定晶格是什么；本节将假定此模型在某个晶格上能够形成自旋液体，并且将处理自旋液体的标准手法作用于其上。

18.1.1 举例：三角晶格上的RVB态及其附近的spinon低能激发

RVB态

我们详细说明一下资料框18.1中的RVB态。所谓RVB态是指这样的基态：

$$|\text{ground}\rangle \propto \sum_{\text{all possible pair partitions}} \frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle)_{\text{pair 1}} \otimes \frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle)_{\text{pair 2}} \otimes \cdots, \quad (18.2)$$

即我们将三角晶格划分成许多不相交的相邻自旋对，然后让每个相邻自旋对上的两个自旋处于自旋单态，即 $(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle)/\sqrt{2}$ 上，将所有可能的这种态（图18.2(a)展示了一个这样的态，其中被同一块黄色区域覆盖的两个格点上的自旋处在一个自旋单态中）等权叠加起来，就得到了一个RVB态。resonance一词来自于(18.2)的各个组分都不是能量本征态，从而哈密顿量中，不同的自旋二聚体态之间存在跃迁，如果将一个自旋二聚体态设置为初态，那么一段时间后它会演化到另一个自旋二聚体态上，然后演化回来，如此不断振荡（如图18.3所示）。这和“苯在两种单双键排列的构型之间振荡”的说法很相似：实际上苯分子中的电子构成离域大 π 键，而单双键排列的构型不是能量本征态，如果强行从单双键排列的构型出发看问题，则苯的波函数是这些构型的叠加，随着时间演化，其中各种构型的分量周期性变化。

现在我们分析RVB态之上的元激发。将相邻的两个自旋当成一个自由度，则它可以分解为

$$\frac{1}{2} \otimes \frac{1}{2} = 0 \oplus 1, \quad (18.3)$$

RVB态中，像图18.2(a)中这样被同一块黄色区域覆盖的两个格点占据上式中的单态，那么可以给系统一个局域的激励，让某一对最近邻格点的状态变成一个自旋三重态。图18.2(b)展

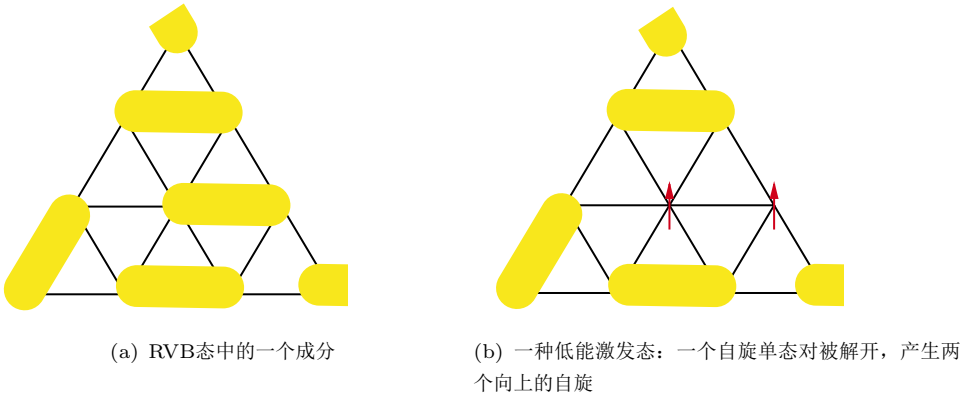


图 18.2: RVB态及其元激发

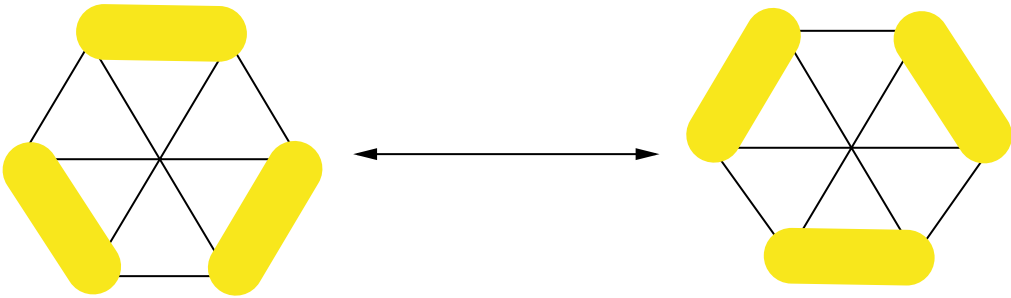


图 18.3: resonance: 图中的两个构型都不是能量本征态，它们可以彼此转化

示了一种这样的激发态：一个 $(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle)/\sqrt{2}$ 态被激发成 $|\uparrow\uparrow\rangle$ ，图形化地展示就是。组成自旋三态的两个自旋可以跑：我们总是可以重新安排自旋配对，于是我们可以认为自旋1的激发被拆成了两半，每一个都可以四处移动。这就得到了自旋子 (spinon)。可以验证这是费米子，这些自旋子全部都是自旋1/2的，并且总是有偶数个自旋子。

因此，至少有一类自旋系统——基态为RVB态的自旋液体——中可以演生出费米子激发，这让我们有理由相信，做费米型的部分子拆分至少对一些自旋系统是合理的。

然而，我们应当注意，图18.2(b)中展示的这种自旋子元激发实际上只有在纯粹的自旋二聚体态，即一个形如图18.2(a)的态上才是有明确定义的。以RVB为基态的体系中，自旋二聚体的量子涨落现在转移到了自旋子的量子涨落上面。因此，我们需要设法用自旋算符表示出自旋子，然后代入原哈密顿量中，去得到关于自旋子的有效理论。这里我们也可以看出为什么要构造RVB态：如果自旋二聚体态实际上接近体系的能量本征态，那么系统的低能自由度仍然是比较平凡的——我们可以将这里的spinon看成类似于XY模型中的涡旋的激发，它唯一略有不同寻常的地方就是总是成对出现。为了让体系处在一个无法用纯经典的图像理解的量子序当中（如拓扑序中存在交换相位等等），必须引入足够的量子涨落，这会让系统基态是大量原系统的低能态的叠加，具体到自旋系统中，就是大量自旋二聚体态的叠加，这就是要构造RVB态的原因。

此处从自旋二聚体出发构造分数化的spinon的方式可以推广到其它自由度中，如可以构造粒子二聚体，一个二聚体算符为

$$d_{ij} = \frac{1}{\sqrt{2}}(c_i^\dagger + c_j^\dagger), \quad (18.4)$$

然后我们会发现在一串二聚体链的边缘格点上粒子数期望为1/2，从而如果一个哈密顿量的低能态都是二聚体态的话，就出现了粒子一分为二的分数化。同样，这样得到的分数化激发本身尚不具有拓扑序，但是如果有前述拉格朗日乘子引入演生规范场等机制，那么这些分数化激发将会携带规范荷，在二维情况下就会形成拓扑序，当然也可以形成其它量子序。分数化本身和拓扑序没有直接关系（例如，AKLT链的边界态就是分数化产生的，可是其中并无拓扑序），演生规范场才是决定拓扑序行为的关键。

以上思路也可以扩展到经典的动力学理论（海森堡运动方程中取 $\hbar \rightarrow 0$ 极限的结果，没有物理量不对易，但是有时间演化，时间演化可以看成量子涨落的残余）中：我们可以考虑流体动力学中的涡旋在纳维-斯托克斯方程中如何演化，或许还能够将流体中的一些物理量诠释成演生电磁场。流体动力学中的涡旋有时间演化这件事的量子版本就是稳定的涡旋态是大量含有涡旋的流速场构型态的线性叠加。实际上，以流体动力学中的涡旋为粒子，流体密度可以看成磁场而流体速度可以看成电场，因此我们果真获得了演生规范场（虽然是完全经典的）！

最后，我们要大致讨论演生规范场及其物理意义。Fradkin 9.5指出了演生规范场可以从大量dimer state的线性叠加态附近得到，而dimer state需要线性组合才能够得到能量本征态本身就说明了它们有明显的量子涨落。在 \mathbb{Z}_2 规范理论中我们会看到是 σ^z 和 σ^x 的不对易导致了e激发和m激发的非平庸交换相位。因此可以看到，演生规范场是量子涨落的一种体现（并且，由于Wilson loop可以任意地大而取值不变，实际上演生规范场的出现意味着系统中存在长程纠缠；通过带约束的分数化构造演生规范场也体现了这一点）。在二维体系中，演生规范场将导致拓扑序，我们这里构造的自旋子等是拓扑序中的任意子（或者别的有趣的量子序中的点粒子）的某种前体，它可以看成对演生规范场而言的某种“外加测试电荷”。

资料框 18.2: XY模型中的拓扑序

在讨论KT相变时，一个可能的疑难是，似乎经典二维XY模型中的涡旋之间也有某种弦连接，从而这里的涡旋看起来很像是 \mathbb{Z}_2 规范理论中的vison或是 \mathbb{Z}_2 荷，因此，KT相变中的有序态离拓扑序有多远？Sachdev在^[23]和^[22]中（这个现象实际上在^[24]中已经研究过了），从一个三维的扩展的XY模型（可以看成二维的一个强关联的量子XY模型的零温路径积分，其中XY自旋可以发生各种时间演化，体现了强烈的量子涨落）出发，通过一个离散Hubbard-Stratonovich变换，引入了 \mathbb{Z}_2 规范场。注意引入Hubbard-Stratonovich变换之后，XY自旋自由度的哈密顿量将形如

$$-\sigma_{ij}^z \cos((\theta_i - \theta_j)/2), \quad (18.5)$$

其中1/2的因子是因为这样一来，积掉 \mathbb{Z}_2 规范场之后得到的理论中，XY自旋自由度的哈密顿量将是

$$\sim \cos^2(\theta_i - \theta_j)/2 \sim \cos(\theta_i - \theta_j),$$

正好给出引入规范场之前的XY模型。XY自旋的指向发生 2π 的剧变的格点连接成了一条线，其两端是两个涡旋，而由(18.5)，在这条线上 \mathbb{Z}_2 规范场的 σ^z 自由度应该取-1值，否则会有很大的能量消耗。因此，涡旋和 \mathbb{Z}_2 规范场中的m激发被粘合在了一起，连接XY涡旋的弦和 σ^z 弦被粘合在了一起，一个 \mathbb{Z}_2 拓扑序演生出来了。

以上推导明确地展示了XY模型中的涡旋和 \mathbb{Z}_2 规范场的vison（见第18.4.3节）如何冻结为一个自由度，从而直观地展示了强关联体系如何产生演生规范场和拓扑序，以及KT相变中的涡旋和连接两个涡旋的“弦”是拓扑序中的任意子的前体、在引入强烈的量子涨落之后能够形成拓扑序这一事实，另一方面也直观地展示了二维规范场和某种粒子（在这里是KT相变中的涡旋）耦合时，规范荷如何被粘到粒子上。

slave boson方法

背景知识 18.1: 自旋系统的部分子构造

部分子方法是指将某种自由度（如自旋自由度）写成一些费米子或是玻色子自由度（即所谓**部分子**）的组合，自旋算符是两个费米子或是玻色子产生湮灭算符的乘积，然后施加适当的约束来保证拆分后的物理和拆分前相同。这相当于说，我们使用一个费米子系统或是玻色子系统实现了一个自旋系统或是别的什么系统。这种方法有时也称为**投影构造**。这样做的好处在于，如果由此得到的费米子或玻色子理论中相互作用没有强到让这些费米子和玻色子又凝聚成对，那么实际上，这意味着自旋系统演生出了类似于费米子和玻色子的激发，即拆分得到的部分子正是自旋系统的低能自由度。

只要算符的代数关系不变，并且将部分子系统的希尔伯特空间固定为每个格点上只有一个部分子的那部分，不同的拆分方式不会改变物理。对自旋系统，我们手动给不同格点上的自旋算符贴上位置标签，而对费米子/玻色子系统，坐标标签是粒子产生算符自带的参数，因此表面上看起来，拆分之后得到的费米子/玻色子系统的希尔伯特空间由于对称化/反对称化的要求而受到限制，似乎部分子系统的希尔伯特空间要比自旋系统的希尔伯特空间小，但是实际上两者是同构的：自旋系统的希尔伯特

空间的基形如

$$|\sigma_1, \sigma_2, \dots, \sigma_n\rangle$$

而部分子系统的希尔伯特空间的基形如

$$c_{1\sigma_1}^\dagger c_{2\sigma_2}^\dagger \cdots c_{n\sigma_n}^\dagger |0\rangle,$$

后者中交换1,2等空间标签, 态矢量不变或者反号, 所以两种系统的希尔伯特空间是一样大的。不施加费米统计或是玻色统计, 粒子系统的希尔伯特空间实际上要比自旋系统的希尔伯特空间大, 反倒是施加了对称/反对称要求之后两者是同构的。

不过虽然部分子构造方式不改变物理, 一些拆分方式中的部分子更加接近自旋系统中实际出现的激发, 从而, 使用这些拆分方法得到的部分子的理论使用平均场之类的方法处理得到的结果相比于其它方案是更加可靠的。这和背景知识8.3中选择序参量很相似。

部分子构造中的部分子如果果真能够获得足够长的寿命, 则往往会是分数化现象的一种, 即系统中不同的“性质”似乎被不同的激发携带着。应当注意分数化并不必然意味着拓扑序的出现: 例如, Luttinger液体中就没有拓扑序。分数化的量子多体系统中出现拓扑序通常是因为正确的分数化方案会导致体系的希尔伯特空间扩张, 为了保证物理的希尔伯特空间不变, 必须向哈密顿量引入一个拉格朗日乘子以约束实际会出现的态仍然位于原有的希尔伯特空间中, 这个拉格朗日乘子可能就会成为演生规范场的来源。更加数学地看, 我们实际上是将扩张后的希尔伯特空间商掉多余的态, 这有可能导致规范对称性, 此时系统的低能有效理论必然是一个规范理论。

我们对海森堡模型使用**slave boson**方法, 这是指做分解

$$S_i = \frac{1}{2} f_{i\alpha}^\dagger \sigma_{\alpha\beta} f_{i\beta}, \quad (18.6)$$

其中 f 为费米子; 每个格点上的自旋的取值由这个格点上的费米型部分子的自旋携带。这个方法本来是用于分析携带电荷的费米子的, 但是后来被用在了分析自旋上, 相应的这种方法分解出来的部分子在自旋系统中就是费米子而不是玻色子。使用费米型部分子的好处在于可以避免玻色子陷入玻色-爱因斯坦凝聚态; 如果自旋系统中的元激发实际上并不处在玻色-爱因斯坦凝聚态, 这就意味着我们需要某种很强的玻色型部分子之间的相互作用让玻色-爱因斯坦凝聚态不稳定, 即需要很强的量子涨落, 于是我们无非是把一个强关联问题转化成了另一个强关联问题。但是, 如前所述, RVB态附近的激发真的好像一个费米子, 所以我们有理由相信, 如果在某个晶格上海森堡模型(18.1)真的演生出了基态是RVB态的自旋液体, 那么slave boson构造就是合理的。

slave boson拆分不是唯一的方法。还有一种拆分方式: slave fermion, 即Schwinger boson。

费米型spinon的探测

中子散射实验能够激发出magnon, 而magnon会“分裂”成两个spinon。因此普通的磁体做中子散射实验能够观察到清晰的峰, 而以spinon为基本自由度的自旋液体就是模糊一片的, 因为有两个spinon出射, 从而会得到连续谱。

18.1.2 spinon哈密顿量和演生 $U(1)$ 规范场

受约束的费米子哈密顿量

做了slave boson分解之后, 哈密顿量就成为一个四阶项, 描述了两个费米子的散射, 写出来是

$$H = \frac{J}{4} \sum_{\langle i,j \rangle} (2f_{i\alpha}^\dagger f_{i\beta} f_{j\beta}^\dagger f_{j\alpha} - f_{i\alpha}^\dagger f_{i\alpha} f_{j\beta}^\dagger f_{j\beta}),$$

第二项由于我们要求每个格点只有单占据而可以略去, 于是

$$H = -\frac{J}{2} \sum_{\langle i,j \rangle} f_{i\alpha}^\dagger f_{j\beta}^\dagger f_{i\beta} f_{j\alpha}. \quad (18.7)$$

这个哈密顿量所在的希尔伯特空间是每个格点上有且只有一个费米子的这部分空间, 写成公式就是

$$f_{i\alpha}^\dagger f_{j\alpha} = 1, \quad (18.8)$$

注意其中 i 不求和而 α 求和。这个约束在接下来的计算中必须始终被考虑到, 它的一个直接推论是

$$\epsilon_{\alpha\beta} f_{i\alpha} f_{i\beta} = 0. \quad (18.9)$$

(18.7)本身不会将一个满足(18.8)的态转化成一个不满足这一限制的态, 因此这里没有任何矛盾。

路径积分和演生规范场

下面我们使用路径积分方法引入约束(18.8)。使用一个逐点的拉格朗日乘子 λ 来施加约束条件, 我们有

$$S[f, \bar{f}, \lambda] = \int d\tau \left(\sum_i \bar{f}_{i\alpha} \partial_\tau f_{i\alpha} + \sum_i i\lambda_i (\bar{f}_{i\alpha} f_{i\alpha} - 1) - \frac{J}{2} \sum_{\langle i,j \rangle} \bar{f}_{i\alpha} \bar{f}_{j\beta} f_{j\alpha} f_{i\beta} \right). \quad (18.10)$$

积掉拉格朗日乘子 λ_i 就能产生硬约束(18.8)。

我们注意到, λ_i 出现的位置和电势基本上一模一样 (见(1.22)), 且显然具有局域 $U(1)$ 对称性。这让我们猜测, (18.10)实际是一个自旋1/2费米子和某个 $U(1)$ 规范场耦合的理论。实际上, 容易验证, 将理论

$$S[f, \bar{f}, \lambda, \chi, \bar{\chi}] = \int d\tau \left(\sum_i \bar{f}_{i\alpha} \partial_\tau f_{i\alpha} + \sum_i i\lambda_i (\bar{f}_{i\alpha} f_{i\alpha} - 1) + \frac{J}{2} \sum_{\langle i,j \rangle} (|\chi_{ij}|^2 - \chi_{ij} \bar{f}_{i\alpha} f_{j\alpha} + \text{h.c.}) \right) \quad (18.11)$$

积掉 χ , 得到的正是(18.10)。上式对应的哈密顿量是

$$H = \frac{J}{2} \sum_{\langle i,j \rangle} (|\chi_{ij}|^2 - \chi_{ij} f_{i\alpha}^\dagger f_{j\alpha} + \text{h.c.}) - \sum_i a_{0i} (f_{i\alpha}^\dagger f_{i\alpha} - 1). \quad (18.12)$$

注意作用量中不含有 χ 场的时间导数, 从而 χ 场没有Berry相位项; λ 场也是一样, 而且由于它是拉格朗日乘子, 在虚时间和实时间下它的定义还需要差一个 i 。这里我们为了避免混乱, 已经用 a_0 代替了 λ 。

χ_{ij} 看起来很像一个规范联络，不过由于它同时包含两种涨落——其振幅和相位——它实际上比规范联络稍微复杂一些。(18.12)中存在规范对称性：做变换

$$\chi_{ij} \longrightarrow \chi_{ij} e^{-i(\theta_i - \theta_j)}, \quad f_i \longrightarrow e^{i\theta_i} f_i, \quad (18.13)$$

那么哈密顿量不变。因此模型中存在演生 $U(1)$ 规范场，并且很容易看到，这个规范场正是 χ_{ij} 的相位。进一步，注意到路径积分形式中， χ_{ij} 的振幅的涨落带来的能量变化明显比其相位发生变化带来的能量变化高（前者直接会影响 $|\chi_{ij}|^2$ 项，后者没有什么明显的影响）。这让我们有信心认为 $|\chi_{ij}|$ 应该是一个有能隙激发而 a_{ij} 则是无能隙激发，或者至少后者比前者更低能。简单的费曼图图形分析会说明，积掉 χ_{ij} 能够给 a_{ij} 提供动力学，当然它也同时会产生一大堆含有 f_i 的项，如一个正比于 a_{ij} 两次方的Hubbard相互作用。既然积掉 χ_{ij} 后只剩下 f_i ， a_{0i} 和 a_{ij} 三种场，而 a_{0i} 不参与(18.13)变换，因此，积掉 $|\chi_{ij}|$ 之后得到的有效理论中仅包含 a_{ij} 的部分是一个 $U(1)$ 规范不变量，很明显就是 $U(1)$ 格点规范场自身的哈密顿量，而包含 f_i 的部分则形如

$$H_{\text{spinon}} = -\frac{J}{2} \sum_{\langle i,j \rangle} (e^{ia_{ij}} \bar{\chi}_{ij} f_{i\alpha}^\dagger f_{j\alpha} + \text{h.c.}) - \sum_i a_{0i} f_{i\alpha}^\dagger f_{i\alpha}. \quad (18.14)$$

总之，系统中演生出了一个 $U(1)$ 规范场和一个费米子场，后者给出spinon，前者传递spinon之间的相互作用。

当然，总的哈密顿量肯定不止(18.14)和 $U(1)$ 格点规范场自身的哈密顿量，例如，如前所述，还有spinon之间的正比于 a_{ij} 的Hubbard相互作用。然而如果我们相信spinon足够接近系统的能量本征态，那么至少spinon之间不应该有强关联效应（比如说，如果Hubbard相互作用很强，那么应该有反铁磁序，然而自旋液体里面不应该有磁性序），因此，至少是定性地，(18.14)加 $U(1)$ 格点规范场自身的哈密顿量应该构成对至少是一部分自旋液体的良好描述。

通常不计算 $U(1)$ 格点规范理论的哈密顿量中的具体参数，因为它显而易见的困难。原则上我们可以从平均场理论出发，通过一些方法——如大 N 展开——计算更高阶修正，但是这样非常繁琐，而仅计算有限阶可能效果也不好。

(18.14)相当于是做了一个比较“柔和”的平均场近似（忽略 χ_{ij} 的涨落）得到的。更为粗暴的平均场近似会直接忽略 a_{ij} ，给出自由spinon，而更加审慎的、包含了更多过程的分析则表明spinon被 $U(1)$ 规范场粘合在一起。图18.4图形化地展示了这一点。

$U(1)$ 演生规范场是否真的存在？

不同格点的基本自由度彼此对易的海森堡模型中有费米型spinon本身已经很离谱了，现在还多出来一个 $U(1)$ 格点规范场。由于我们并不知道受到阻挫的海森堡模型的基态到底是什么，因此实际上也无从推测在其附近能够有费米型spinon和演生规范场 a_{ij} ，而 $|\chi_{ij}|$ 的涨落受到压制的平均场基态是否真的和阻挫海森堡模型的基态足够接近。如果两者并不十分接近，通过以上平均场方法得到的“低能有效理论”(18.14)就是假象了。

我们知道规范理论有禁闭相，在其中规范场牢牢地将与之耦合的粒子固定在一起，从而系统的低能自由度中没有规范场也没有与之耦合的粒子，因为两者的能隙都非常大。QCD在低能标下就是这样。对海森堡模型，如果(18.14)耦合上 $U(1)$ 规范场后处于禁闭相，那么就没有低能的spinon——这正好说明系统中只存在**玻色型**的自旋涨落，如自旋波一类。海森堡模型没有形成spinon时就是这样的。因此此时演生规范理论(18.14)是无用的，但是**不是错的**——我们从它出发得到了合乎常理的结论，虽然绕了一个大弯子。

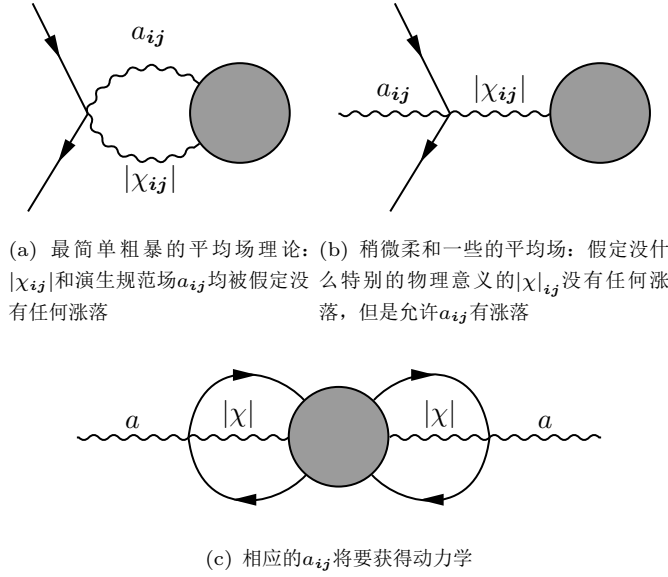


图 18.4: 两种平均场理论的对比

另一方面, 对于自旋液体态, 做完slave boson分解后直接对spinon做平均场是有问题的, 因为我们手动地破缺了每个格点上粒子数始终唯一的限制, 但是实际上, 由于这个费米子模型实际上是一个自旋模型, 这样的破缺是不正确的。为此在路径积分表述中, 我们加入一个拉格朗日乘子项来固定每个格点上的粒子数, 具体的拉格朗日乘子大小不能确定, 是路径积分中需要额外做积分的一个场变量。在正则量子化表述中可以通过将一个投影算符作用在普通的平均场波函数上来得到更加接近实际情况的波函数。

我们很关心演生规范场是不是禁闭的, 如果是的话, 它将不会有任何物理效应, 因为低能下规范场将会非常强, 从而spinon被紧密结合在一起, 不会出现任何spinon涨落 (而如果能量足够高, 那就不再能够“只考虑基态附近的涨落”, 而且, 在凝聚态物理的语境下, 能量足够高通常意味着温度足够高, 而热涨落足以破坏大量有趣的量子纠缠)。这种情况下不存在独立的spinon激发以及规范场涨落, 也就没有自旋液体态。

18.1.3 演生 $SU(2)$ 规范场

如果(18.10)有非平庸的鞍点解, 它附近的涨落肯定会破坏一些局域 $U(1)$ 对称性。现在我们假定在某些情况下可以有这样的鞍点解: Δ_{ij} 非零, 但是 χ_{ij} 为零。此时

$$S = \int_0^\beta d\tau \left(\sum_i \bar{f}_{i\alpha} (\partial_\tau - i\lambda_i) f_{i\alpha} - \tilde{J} \sum_{\langle i,j \rangle} \bar{\Delta}_{ij} \epsilon_{\alpha\beta} b_{i\alpha} b_{j\beta} + \text{h.c.} \right) \quad (18.15)$$

设

$$\Delta_{ij} = \epsilon_{\alpha\beta} f_{i\alpha} f_{j\beta}, \quad \chi = f_{i\alpha}^\dagger f_{j\alpha}, \quad (18.16)$$

有

$$H = -\frac{J}{2} \sum_{\langle i,j \rangle} \Delta_{ij}^\dagger \Delta_{ij} = -\frac{J}{2} \sum_{\langle i,j \rangle} \chi_{ij}^\dagger \chi_{ij} = -\frac{J}{4} \sum_{\langle i,j \rangle} (\Delta_{ij}^\dagger \Delta_{ij} + \chi_{ij}^\dagger \chi_{ij}), \quad (18.17)$$

于是可以对最后一个形式做平均场近似: 我们引入了两个参量, 为了增加变分计算的参数,

使之更加有效。直接做平均场分解，有（我们将算符的期望值去掉帽子）

$$H_{\text{MF}} = -\frac{J}{4} \sum_{\langle i,j \rangle} (\Delta_{ij}^* \Delta_{ij} + \Delta_{ij} \Delta_{ij} - |\Delta_{ij}|^2) - \frac{J}{4} \sum_{\langle i,j \rangle} (\chi_{ij}^* \chi_{ij} + \chi_{ij} \chi_{ij} - |\chi_{ij}|^2).$$

但是，实际上应该将系数设置成3/8，因为数值计算说明，直接做平均场近似效果非常糟糕。我们下面将会将系数统一称为 \tilde{J} 。实际上严格计算平均场理论并没有什么意义，因为平均场理论的定量结果向来是非常糟糕的，而使用变分蒙特卡洛方法不难得到较为精确的结果。

参考：RMP 51, 657, 1979 RMP 78 17 2006

flux-fusion anomaly test

LSM theorem: if in each unit cell of a system there are odd spin-1/2, there must be ground state degeneracy.

所以现在我们看到，磁体的相实际上是非常多样的，包括

1. 平凡的顺磁相，没有基态简并。
2. 对称性自发破缺相，存在基态简并。
3. 拓扑序带来的基态简并。
4. 无能隙激发带来的基态简并。

18.2 自旋液体模型的分类

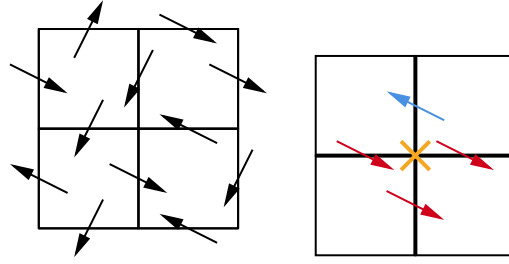
我们可以看到，要想确定一个给定的自旋模型中有没有解禁闭的自旋液体是非常难以判断的：我们得到自旋液体的低能有效理论的过程——对自旋自由度做一定的部分子构造，获得spinon，通过对希尔伯特空间的限制引入 χ_{ij} 场，发现规范结构，于是最终写出spinon和演生规范场耦合的有效理论——实际上并没有用到太多关于原理论的信息。例如，实际上从来没有人计算过演生规范场中的各种系数，自然也无法确定一个给定的自旋模型的演生规范场是禁闭的还是解禁闭的，从而也不知道它是否能够给出解禁闭的自旋液体态。

这意味着实际上研究自旋液体更好的办法是反过来：写下一个某种物质场和规范场耦合的理论，并要求该理论的禁闭相是一个自旋模型，我们就能够研究所有的自旋液体，虽然我们并不知道它们如何能够被演生出来。实际上，我们在研究寻常的能带电子态时使用的是一模一样的方法：我们只是假定相互作用较弱，从而只是修正了一下能带结构而已。能带电子态和自旋液体态的不同在于能带论中的参数可以被第一性原理地计算（如第七章），但是自旋液体态——以及大部分强关联物态——中的各种参数不能被第一性原理地计算。

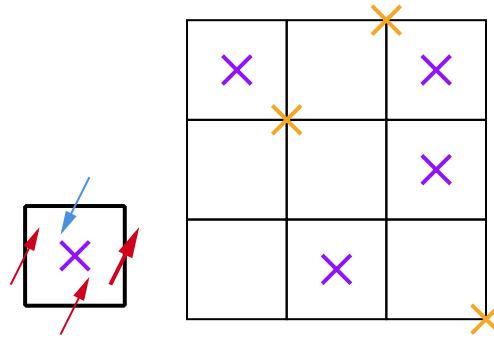
在二维自旋液体中如果出现了演生规范场，那么一般就意味着出现了拓扑序，因为规范场可以产生非平凡的Berry相位（见背景知识17.2）。这是寻找自旋液体的一个重要动机：它给出了产生分数量子霍尔效应以外的拓扑序的方法。

18.2.1 自旋液体中的对称性分数化

大部分对称性都有所谓的对称性反常，需要生存在高维SPT（??）的边界态上。直接从对称性分析出发可以得到大量的态，但是满足一定非常合理的条件的态是很少的，这些态均可以直接使用平均场构造出来。



(a) Toric-code模型的希尔伯特空间的一组基底由这样的态构成：上图是一个A激发的例子
每条边上都是有确定的 σ^z 或是有确定的 σ^x



(c) B激发定义在正方形方块上，上图是一个B激发的例子
(d) 虽然 σ^x 和 σ^z 之间有量子涨落，但是A和B对易，从而系统的能量本征态图是一个B激发的例子
可以用A激发和B激发标记；通过数自由度会发现也只需要这两个标记

图 18.5: Toric-code模型的系统构型和元激发

18.3 Toric-code模型

18.3.1 Toric-code哈密顿量与解析解

Kitaev最早提出了一种模型，作为一种可能的量子计算纠错编码，他发现这个模型放在一个环面上可以有非常有趣的结果。然而，事后发现这个模型实际上展现出了一个拓扑序。这个模型是一个严格可解模型，同时又是一个自旋液体（其基本自由度是自旋，并且没有磁性序），因此值得在这里深入研究。

考虑一个正方晶格，在每条边（不是每个格点！）上放有一个自旋1/2自由度。哈密顿量为

$$H = - \sum_s A_s - \sum_p B_p, \quad (18.18)$$

其中下标 s 表示格点， A_s 指的是格点 s 周围的四条边上的 x 方向上的自旋算符的乘积，即

$$A_s = \prod_{i \text{ near } s} \sigma_i^x, \quad (18.19)$$

而 p 表示格点中的一个最小正方形方块， B_p 指的是正方形 p 的四条边上的 z 方向上的自旋算符的乘积，即

$$B_p = \prod_{i \text{ of } p} \sigma_i^z. \quad (18.20)$$

(18.18)中显然有不小的量子涨落, 因为有大量彼此不对易的算符。然而, 我们将展示, 它其实是严格可解的。因此, Toric-code模型是一个很好的玩具模型, 能够向我们展示量子涨落强烈的自旋系统的行为。

首先可以验证 $\{A_s\}$ 和 $\{B_p\}$ 构成一组对易稳定子 (即平方为1的一组彼此对易的厄米算符), 这样就有

$$[A_s, H] = [B_p, H] = 0. \quad (18.21)$$

另一方面, 平方为1的厄米算符的本征值是 ± 1 , 于是我们就可以用它们的本征值 $A_s = \pm 1$ 和 $B_p = \pm 1$ 标记体系的能量本征态。实际上, 在热力学极限下只需要 $\{A_s\}$ 和 $\{B_p\}$ 就可以唯一地标记体系的能量本征态。这是因为设体系有 N 个格点, 那么有 $4N/2 = 2N$ 条边, 于是体系的希尔伯特空间的维数为 2^{2N} 。 s 和 p 均有 N 个, 于是所有可能的 $\{A_s\}$ 和 $\{B_p\}$ 的组合总数为 $2^N \cdot 2^N = 2^{2N}$ 。这样如果不考虑边界引入的微妙之处, 只需要 $\{A_s\}$ 和 $\{B_p\}$ 就可以唯一地标记体系的能量本征态。很容易看出体系的基态为所有 A_s 和 B_p 均为1的状态, 于是我们可以把 A_s 和 B_p 为 -1 的情况看成激发态。这样我们就得到了(18.18)的全部能量本征态, 从而完全求解出了它。

显然Toric-code模型确实是自旋液体, 因为其基态不具有任何经典意义上的序: 我们得到的是一大堆 σ^x 确定的态和一大堆 σ^z 确定的态的线性叠加。

18.3.2 环面上的情况

e激发和m激发

为了解析求解, 我们施加一个周期性边界条件, 这相当于把体系放在了一个二维环面上。此时诸 $\{A_s\}$ 和 $\{B_p\}$ 实际上不是彼此独立的, 因为此时显然有

$$\prod_s A_s = 1,$$

因为所有的 $\{A_s\}$ 乘起来, 每一条边被乘了两边, 所以一定会得到1。类似的有

$$\prod_p B_p = 1.$$

这两个方程要求

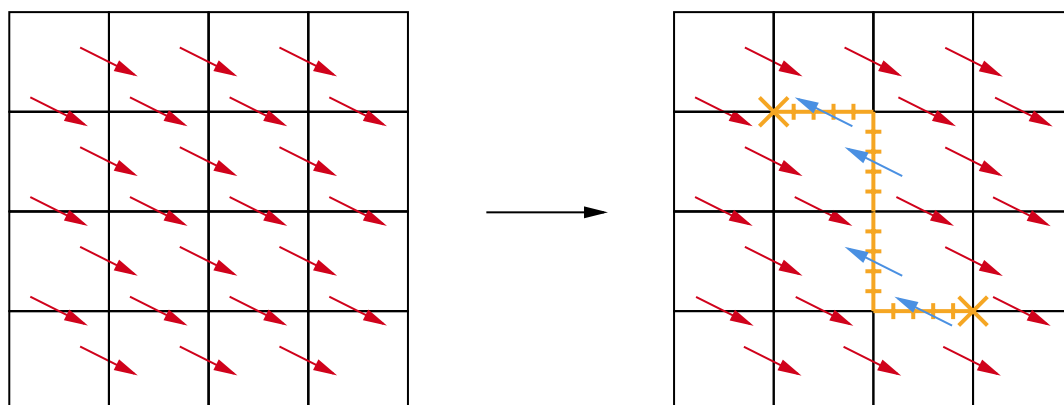
$$\prod_s A_s = \prod_p B_p = 1. \quad (18.22)$$

这就意味着 A_s 激发和 B_p 激发必须成对出现, 否则乘积将会是 -1 。我们将 A_s 激发称为e粒子, 而将 B_p 激发称为m粒子, 因为在某种意义上可以将 A_s 激发类比为电荷而将 B_p 激发理解为磁通量子。这两种粒子的性质和空间的拓扑结构显然关系很大, 因此称它们为拓扑激发。

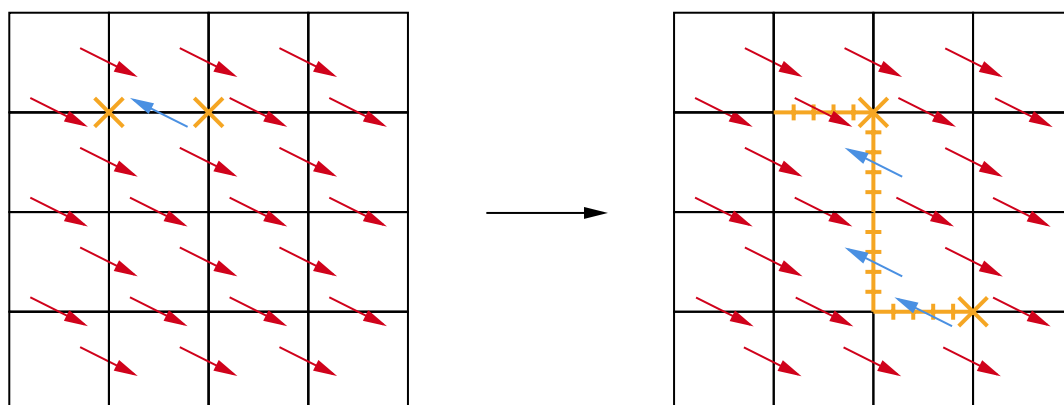
两种激发成对出现的事实意味着可以使用弦算符描述它们的产生和消灭。¹首先考虑由

¹一个可能的疑难是, 当任意子交换产生的相位差非常小, 接近0和 π 时, 系统应该平滑地过渡到玻色子系统或者费米子系统上, 但是费米子系统和玻色子系统可以使用局域的场算符描述, 似乎并没有方法能够平滑地从弦算符过渡到场算符上。这里的答案是, 有一些定理保证了有限个任意子交换后产生的相位差是有理数。一个比较直觉性的看法是, 如果任意子交换相位差是无理数, 则无法将有限个任意子做fusion而得到平凡的任意子, 因此系统中有无数种任意子。从而, 适当的情况下, 系统基态会有无数重简并, 这不是一个非常物理的结果。上面的论证在两个任意子做fusion以后是一系列任意子之和时不适用, 因此需要更加严格的证明。总而言之, 为了保证模型足够物理, 实际上不能够连续地调节相位差。因此任意子和玻色子、费米子之间是不能够平滑地过渡的。

任意子激发是弦算符产生的, 我们可以追踪每个任意子的位置, 或者按照量子场论的套路, 我们似乎应该讨论弦算符的动力学。不过应当注意, 由于很多弦算符 (如闭弦) 本身并不增加或者减少系统能量, 我们实际上不需要一般的关于弦的理论, 一定有一些更加平凡的理论等价于任意子模型。例如, 一些阿贝尔拓扑序可以使用Chern-Simons理论描述。通常是某个拓扑量子场论



(a) 将一个 O_e 开弦算符作用在一个元激发也没有的态上而得到的结果；弦两端出现了两个 e 激发



(b) 将一个 O_e 开弦算符作用在已有元激发的态上得到的结果：弦开头处的 e 激发被运送到了弦末尾，弦上其它各点上如果原本没有 e 激发，那么还是没有，如果原本有，那么现在还是有

图 18.6: Toric-code 模型的 O_e 开弦算符

一条边连接的两个格点，这条边上的 σ_i^z 算符可以将这条边上的 x 方向的自旋翻转，因此它可以做到以下三件事：

- 如果两个格点上原本没有 e 粒子，那么在两个格点上同时产生 e 粒子；
- 如果两个格点上原本都有 e 粒子，那么在两个格点上同时消灭 e 粒子；
- 如果两个格点一个有 e 粒子一个没有，那么该 e 粒子将被转移到原本没有 e 粒子的格点上。

这样设一系列首尾相连的边 $\{l\}$ 连接了两个格点，则弦算符

$$O_e = \prod_l \sigma_l^z \quad (18.23)$$

同样可以做到以上三件事。同样，将以上论述中的 σ^z 换成 σ^x ，“格点”换成“方块”，“连接两个格点的边”换成“方块共享的边”（我们可以在每个方块中间放置一个点，从而 m 粒子也定义在一个格点上），同样可以定义弦算符

$$O_m = \prod_i \sigma_i^x. \quad (18.24)$$

以上讨论的都是开放的弦，闭合的弦的行为需要具体分析，且对闭弦有

$$O_e |0\rangle = O_m |0\rangle = |0\rangle. \quad (18.25)$$

通过弦算符可以检查 e 粒子和 m 粒子绕对方转一圈（实际上就是使用一个闭合的弦算符作用在一个有 e 粒子或者 m 粒子的格点上），都会多出来一个 π 的相位，这是因为如果一个 m 粒子闭弦和一个 e 粒子开弦有单个交点，那么它们反对易（因为同一个边上的 σ^x 和 σ^z 反对易）。换言之， e 粒子和 m 粒子均为任意子激发：这是二维的特殊现象，因为二维的环路在二维平面上它围绕的区域被挖掉一个点之后就不可缩了，因此一个粒子转一圈之后可以有一个非零相位变化。本节涉及的激发尚为阿贝尔统计，即转一圈之后得到的量子态和转之前只差一个 $U(1)$ 变换；还有非阿贝尔统计，即转一圈可以转移到别的量子态上。

我们可以将 e 粒子称为 σ^x 弦的末端而将 m 粒子称为 σ^z 弦的末端。这里有一个定义上的模糊性，因为如果波函数中某一个路径上的 σ^x 均为 -1 ，那么它的两端就有两个 e 粒子，在此意义上我们称 e 粒子是 σ^x 弦的末端，然而能够让这种构型产生的算符却是 σ^z 算符的乘积，因此似乎也可以将 e 粒子称为 σ^z 弦的末端。

任意子表

现在的问题是，环面上的Toric-code模型中最多能够弄出来多少任意子？显然 e 粒子和 m 粒子都是任意子，虽然两者自己满足玻色统计，但它们之间有一个非平凡的相位。我们下面将以拓扑性质分类激发，即，拓扑性质相同的激发算作一种。可以用两个量来标记一种激发的拓扑性质：设 M_{ab} 为 b 绕着 a 转一圈导致的复数因子， θ_a 指的是交换两个 a 导致的复数因子（或者说一个 a 绕着另一个 a 转半圈导致的复数因子）。这样，有

$$\theta_e = \theta_m = 1, \quad M_{em} = -1. \quad (18.26)$$

除了 e 粒子和 m 粒子以外肯定还有一种 ϵ 粒子，它是一个 e 粒子和 m 粒子聚合²而成的粒子，即

$$\epsilon = e \otimes m. \quad (18.27)$$

²所谓聚合指的是将两个激发放得尽可能近，从而得到的复合激发。 e 粒子和 m 粒子定义在不同的格点上，因此一个 e 粒子和一个 m 粒子的聚合就是在一个正方方块中央放置一个 m 粒子，在它的某个角上放置一个 e 粒子之后得到的激发，从远处看这近似于一个粒子。

可以容易地验证

$$M_{ee} = M_{mm} = -1, \quad \theta_e = -1. \quad (18.28)$$

注意， e 粒子绕着自己转动时，粒子1的 e 部分和粒子2的 m 部分的交换、粒子2的 e 部分和粒子1的 m 部分各自贡献一个 π 相位，似乎是彼此抵消了，但是如果我们固定粒子1，那么粒子2中的 e 部分和 m 部分绕着粒子1转动时彼此又会有一个交换，所以最终的交换相位为 π ，即 e 粒子存在自统计。 e 粒子、 m 粒子和 e 粒子这三种拓扑激发都只能成对出现。除了这三种激发以外还有一些平凡的激发，比如声子之类，将它们全部记为 $\mathbb{1}$ 。

实际上， e 粒子、 m 粒子和 e 粒子和 $\mathbb{1}$ 就是全部拓扑激发。由于 $\mathbb{1}$ 无论如何绕圈都不会产生附加的相位，就有

$$\mathbb{1} \otimes a = a.$$

两个 e 粒子放在一起，得到的就是某个边上的 σ^x 发生了翻转，这是一个普通的激发； m 粒子和 e 粒子也是如此，于是

$$e \otimes e = m \otimes m = e \otimes e = \mathbb{1}.$$

上式实际上说明了一个非常重要的事实：封闭流形上无论有多少拓扑激发，这个态都可以通过对基态作用一些产生算符得到，或者等价地说改变基态上某些格点的值得到，那么如果将这些拓扑激发聚合到一起，得到的只是基态上局域的一些点被改变了，也即得到了一个平凡的激发。总之，封闭流形上所有的拓扑激发聚合在一起，只会得到平凡的激发。这就从另一个角度解释了为什么非平凡的拓扑激发一定成对出现。 e 和 e 粒子聚合，就相当于两个 e 粒子先聚合得到一个平凡的激发，剩下一个 m 粒子， e 粒子和 m 粒子聚合则会留下一个 e 粒子和一个平凡的激发，于是

$$e \otimes e = m, \quad e \otimes m = e.$$

因此， e 粒子、 m 粒子和 e 粒子和 $\mathbb{1}$ 在聚合运算 \otimes 下是封闭的。

在toric code模型中，我们从完全局域模型中演生出了任意子 e 和 m ，还演生出了具有非平庸的自统计的 e 粒子。 e 和 m 粒子的互统计是通过 σ^x 弦和 σ^z 弦的互统计引入的。表面上，似乎无法形成非平庸的自统计，但是通过组合具有非平庸互统计的任意子，我们是能够得到具有非平庸自统计的粒子的。如果我们适当调整toric code哈密顿量的形式，让 σ^x 弦和 σ^z 弦倾向于粘在一起，并且提高单独的 e 粒子和 m 粒子的能量而降低 e 粒子的能量，我们会发现这个修改后的模型的低能有效理论就是一些自由费米子。这样，我们可以在没有先验地引入任何费米统计的情况下，通过完全局域的算符代数来产生费米子。

Wen 10.3.3

暂时抛下 e 粒子不谈，toric code模型中存在 e 激发和 m 激发这件事也是令人惊奇的，因为实际上我们从一个完全局域模型演生出了 \mathbb{Z}_2 规范理论： m 激发代表一个 \mathbb{Z}_2 规范场中的磁通，而 e 激发是能够感受到这个磁通的 \mathbb{Z}_2 规范荷，整个理论有非平庸的基态简并。请注意，虽然事后诸葛亮地看，toric code模型具有局域 \mathbb{Z}_2 不变性，但是与一般的规范理论不同，我们无需特别注意这个规范不变性：我们不必从希尔伯特空间中商掉规范等价性，就能够获得一个自洽的量子理论，而这个量子理论的低能行为和一个规范理论完全相同。由于规范理论的希尔伯特空间商掉了规范等价性，按照文小刚的看法，不能认为这是一个真正局域的理论。相反，toric code模型却是一个局域的理论。由于 \mathbb{Z}_2 规范场没有类似于光子这样一旦产生就独立于规范荷或是磁通或是别的静态场构型的激发，toric code模型和 \mathbb{Z}_2 规范场的对应关系似乎仅仅只是将 σ 自由度换一个名字而已，但是我们将会看到，toric code模型的推广——弦网模型——可以通过弦密度波来产生规范玻色子。

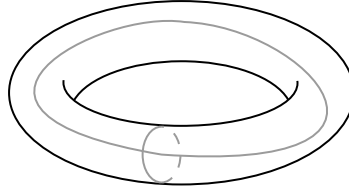
(a) 二维环面的同伦群是 \mathbb{Z}_2

图 18.7: 环面的拓扑性质

基态四重简并和Berry相

回忆一下，体系的希尔伯特空间维数为 2^{2N} 。当 $2N - 2$ 个边的自旋已经确定之后，系统的状态实际上已经确定了，因为约束条件(18.22)会确定剩下两条边的自旋。换言之，实际物理的希尔伯特空间维数只有 2^{2N-2} 。这就意味着总希尔伯特空间 2^{2N} 分裂成了4支，或者说每个状态都有四重简并。这个事实——环面上的Toric-code模型会出现基态四重简并——是(18.22)决定的，而(18.22)本身又来自环面的拓扑性质。如果我们在哈密顿量中引入一个局部的扰动，基态能量和基态波函数显然会发生扰动，但是由于 A 和 B 的定义没有变化，系统拓扑没有变化，(18.22)也是始终成立的。换句话说，环面上基态的四重简并是受到拓扑保护的，局域的扰动不能让它消失。

用什么标记这四重简并？容易想到，完全可以定义一种全局性的闭弦算符，它贯穿整个环面，而由周期性边界条件它是闭弦算符。（这些算符的定义本身和拓扑紧密相关，显然如果系统被放在一个平面上那么根本没法定义全局性的闭弦算符）分别沿着 x 轴和 y 轴定义

$$L_e^x = \prod_x \sigma_i^z, \quad L_m^x = \prod_x \sigma_i^x, \quad (18.29)$$

并可以验证它们和哈密顿量是对易的，且它们构成一对对易稳定子。这就意味着它们的本征值均为 ± 1 ，这就唯一地标记了四重简并。

以上两个弦算符标记的这种稳定的基态简并意味着Toric-code模型中显然有某种序，但这种序并不是使用对称性标记的，即和金斯堡-朗道理论中的那种局域的序是不同的。的确，Toric-code中存在拓扑序。

L_e^x 和 L_m^x 将 e 粒子绕着 x 轴转动一圈，因此它们的本征值实际上给出了 x 方向类似于磁通量的一个通量，这个通量导致了一个Berry相位。

类似地还可以定义 L_e^y 和 L_m^y ，并且

$$\{L_e^x, L_m^y\} = 0. \quad (18.30)$$

我们知道

$$|0\rangle = |L_e^y = 1, L_m^y = 1\rangle, \quad (18.31)$$

而使用这些关系可以证明，

$$\begin{aligned} L_e^x |0\rangle &= |L_e^y = 1, L_m^y = -1\rangle, \\ L_m^x |0\rangle &= |L_e^y = -1, L_m^y = 1\rangle, \\ L_m^x L_e^x |0\rangle &= |L_e^y = -1, L_m^y = -1\rangle. \end{aligned} \quad (18.32)$$

我们发现四重简并和四种基本的任意子正好能够对应上。这是拓扑序的一般特征：基态简并和任意子有对应，基态简并数目就是任意子数目的亏格次方。我们这里是在亏格（洞的

数目) 为1的环面上工作, 因此基态简并的数目为 $4^1 = 4$ 种。如果在亏格为0的球面上, 基态简并的数目就是 $4^0 = 1$ 种。还有另一种方法也可以推导出这个结果。设亏格为 g , 由欧拉公式

$$V - E + F = 2 - 2g,$$

于是

$$E - (V + F - 2) = 2g.$$

而 V 是 A_s 格点的数目, F 是 B_p 格点的数目, 再减去(18.22)造成的两个约束, 则 $V + F - 2$ 是一个二维表面Toric-code态的自由度个数。Toric-code模型总的自由度个数为 E , 因此有 $2g$ 个自由度用于标记简并态, 由于每个自由度有两个取值, 简并度为

$$2^{2g} = 4^g.$$

我们看到, 拓扑性质让基态简并出现, 而基态简并意味着基态中可以有持续存在的弦——基态可以不是空无一物的! 这个看起来非常神奇——但是完全在预料之中——的性质让Toric-code模型成为一类允许出现弦网凝聚的模型中比较简单的一个。

18.4 \mathbb{Z}_2 规范理论

Toric-code是一个严格可解模型; 实际上, 它的一些重要性质在哈密顿量形式更简单(但是不再严格可解)的所谓 \mathbb{Z}_2 规范理论中也会体现出来。历史上发生的事情其实是相反的: 先有了 \mathbb{Z}_2 规范理论, 然后再有了为了分析这一类的理论具有的性质而生造出来的Toric-code模型。

电动力学是一个 $U(1)$ 规范理论, 其中费米子场可以发生任意的局域相位转动, 而与之配套的规范场——电磁场矢势——发生一个局域平移。本文中我们不要 $U(1)$ 这么大的对称性, 而是只希望费米子场或者不发生相位转动, 或者相位就转动 π , 在这样的规范对称性——也就是 \mathbb{Z}_2 规范对称性下系统的动力学保持不变。如果我们还是在通常的四维时空中工作那么局域 \mathbb{Z}_2 变换就是不连续的: 因为0和 π 不能连续过渡。因此我们将在格点上工作, 即研究格点规范场论。

18.4.1 \mathbb{Z}_2 自旋液体的低能有效理论: 费米型spinon和 \mathbb{Z}_2 规范场耦合

费米子的规范不变的跃迁项

在二维格子上, 格点上的费米子的动能项无非是从一个点跃迁到另外一个点, 即

$$H_0 = - \sum_{i,j,\alpha} t_{ij} c_{i\alpha}^\dagger c_{j\alpha}. \quad (18.33)$$

这个哈密顿量在局域 \mathbb{Z}_2 变换下不是不变的。现在我们将尝试把这个哈密顿量改造成 \mathbb{Z}_2 规范理论。

显然, 费米子可以携带 \mathbb{Z}_2 群的表示: 我们只需要让 e 的作用是什么都不变, 而2的作用是将 c 变成 $-c$ 即可。在格点模型中规范联络应该放置在方块的边上而不是点上, 为此, 我们修改 t_{ij} 系数, 使之成为规范联络, 在 \mathbb{Z}_2 变换下能够吸收掉费米子场带来的变化。容易看到, 只需要指定

$$c_{i\alpha} \longrightarrow \eta_i c_{i\alpha}, \quad t_{ij} \longrightarrow \eta_i \eta_j t_{ij},$$

就能够让哈密顿量具有局域 \mathbb{Z}_2 对称性。由于 t_{ij} 只是在正负两种状态之间切换，替代它的规范联络 $\sigma_{ij} = \pm 1$ ，于是用哈密顿量

$$H = - \sum_{i,j,\alpha} t_{ij} \sigma_{ij} c_{i\alpha}^\dagger c_{j\alpha}$$

做路径积分，分别以 c, c^\dagger 和 σ_{ij} 为积分变量即可得到一个 \mathbb{Z}_2 规范理论。

现在我们回到正则量子化框架中， σ_{ij} 在每一个格点引入了 ± 1 两个状态，从而我们可以把它当成一个自旋1/2的自旋算符³，从而哈密顿量为

$$H = - \sum_{i,j,\alpha} t_{ij} \sigma_{ij}^z c_{i\alpha}^\dagger c_{j\alpha} \quad (18.34)$$

希尔伯特空间为费米子的态空间直积上每一点的自旋1/2空间。 \mathbb{Z}_2 规范变换为

$$c_{i\alpha} \longrightarrow \eta_i c_{i\alpha}, \quad \sigma_{ij}^z \longrightarrow \eta_i \eta_j \sigma_{ij}^z. \quad (18.35)$$

特别的，如果(18.33)实际上是一个紧束缚模型，(18.34)就成为

$$H = -t \sum_{\langle i,j \rangle} \sigma_{ij}^z c_{i\alpha}^\dagger c_{j\alpha} + \text{h.c.} \quad (18.36)$$

此时 σ_{ij}^z 实际上仅仅定义在方块的边上，即不需要对不相邻的点对 $\langle i, j \rangle$ 也对应应的 σ_{ij}^z 。

\mathbb{Z}_2 规范场自身的哈密顿量

现在我们讨论 \mathbb{Z}_2 规范场本身的哈密顿量，根据电动力学中的经验，这实际上就是将(18.36)中的费米子自由度积掉⁴后得到的仅仅关于 \mathbb{Z}_2 规范场（而没有任何物质场）的一个理论。严格做积掉费米子的计算是非常不现实的，不过也不必要——正如电动力学中那样，只需要保证哈密顿量本身是 \mathbb{Z}_2 规范不变的即可。我们首先分析 \mathbb{Z}_2 规范变换如何写成算符形式，然后分析积掉费米子自由度之后的哈密顿量会是什么形式的。

只有 σ 的理论中的规范变换是

$$\sigma_{ij}^z \longrightarrow \eta_i \eta_j \sigma_{ij}^z, \quad (18.37)$$

也就是说对每一条边上的 σ^z 本征态，规范变换或是不改变它，或是加一个负号。我们希望将 \mathbb{Z}_2 规范变换写成算符的形式，为此注意到在自旋1/2中，算符 σ^x 可以翻转 σ^z 的本征态，且 σ^x 是厄米算符，于是一条边上的规范场翻转就是

$$\sigma_{ij}^z \longrightarrow \sigma_{ij}^x \sigma_{ij}^z \sigma_{ij}^x.$$

³实际上，这个“自旋算符”未必来自某个体系的内禀旋转不变性。更加数学的说法是，由于每个格点都有两个状态，我们可以在每个格点引入一个 2×2 的厄米矩阵

$$\sigma = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

作为规范场对应的算符，而这正是泡利矩阵中的 σ^z 。后面引入 σ^x 等算符的目的也只是用于翻转规范场的状态。

⁴这当然要求费米子是有能隙的，不过这其实没什么问题，如果(18.36)中的费米子无能隙，我们就把它和一个普通的有能隙的紧束缚模型加起来再积掉费米子即可。与电动力学相类比，积掉费米子无非就是将介质中费米子对电磁场的响应等效为介电常数和磁化率的修正。

此外注意到，由于我们在正则量子化框架中工作，积掉费米子自由度后希尔伯特空间缩小，原本的希尔伯特空间中的纯态将对应一个混合态。但是实际上这无关紧要，因为我们只需要假装不知道费米子存在，分析 \mathbb{Z}_2 规范场的态空间，最后计算配分函数即可，并不需要真的处理自旋系统的密度矩阵。

任何一个 \mathbb{Z}_2 规范变换都可以拆解成一系列作用在格点上的规范变换相乘，而作用在格点 i 上的规范变换翻转和这个格点连接的四条边上的规范场，于是作用在格点 i 上的规范变换为

$$Q_i = \prod_{\langle i,j \rangle} \sigma_{ij}^x = \prod_{\langle j,i \rangle \in +i} \sigma_{ij}^x. \quad (18.38)$$

我们知道实际的系统都是定义在时空上的，所以似乎没有什么阻止我们在时间维上也做规范变换。实际上，(18.38)确实只是空间上的规范变换，但是由于处理时间上的规范变换需要使用并不常用的离散路径积分形式，一般的研究均只研究空间上的规范变换。

于是规范不变量就是和所有 Q_i 对易的算符。由于是低能有效理论，我们考虑最低阶的两个 \mathbb{Z}_2 规范不变量，得到

$$H = -K \sum_{\langle i,j \rangle} \sigma_{ij}^x - J \sum_I \prod_{l \in \square_I} \sigma_l^z. \quad (18.39)$$

这就是只含有 \mathbb{Z}_2 规范场的一个理论（当然，实际上还有很多其它的 \mathbb{Z}_2 规范理论，是取其它 \mathbb{Z}_2 规范不变量得到的）。这样spinon为费米子的 \mathbb{Z}_2 自旋液体的低能有效理论就是

$$H = -K \sum_{\langle i,j \rangle} \sigma_{ij}^x - J \sum_I \prod_{l \in \square_I} \sigma_l^z - t \sum_{\langle i,j \rangle} \sigma_{ij}^z c_i^\dagger c_j. \quad (18.40)$$

注意哈密顿量(18.40)和(18.38)不是对易的。物理地说这是因为(18.40)的第三项即费米子跃迁项的规范变换不仅仅关于 σ^z 也关于 c ，或者换一个角度，由于费米子此时携带 \mathbb{Z}_2 规范荷，单个格点上的 \mathbb{Z}_2 规范荷是不能守恒的。

关于为什么我们考虑了最低阶的两个 \mathbb{Z}_2 规范不变量而不是别的（特别是， σ^x 是怎么被牵扯进来的），可以从两个角度考虑。最显然的解释当然是，最低阶的 \mathbb{Z}_2 规范不变量具有最好的局域性，因此是更有可能出现的。

不过，实际上更加站得住脚的一个解释来自(18.39)的路径积分表述。(18.39)是格点模型，所以我们做一个离散时间路径积分。一个 \mathbb{Z}_2 规范场自由度在路径积分中要加上一个虚时间指标，即 $\{\sigma_{ij}^z(\tau)\}$ 。现在考虑以下理论：

$$Z = \sum_{\sigma^z} \exp \left(\sum_{\tau} (J_{xy} \sum_{\square} \prod_{l \in \square} \sigma_l^z(\tau)) \right), \quad (18.41)$$

由于没有 z 方向上的边上定义了 σ^z 自由度，我们可以直接沿用(18.37)作为三维经典统计理论的 \mathbb{Z}_2 规范变换。三维经典统计理论的形式应该是现在我们看到，(18.39)实际上是已经做了一定的规范固定的理论：我们将所有时间方向上的边上的 \mathbb{Z}_2 场都固定为1了。因此，(18.39)中残留的规范冗余性只能是空间上的，相应的规范变换由(18.38)给出。

18.4.2 \mathbb{Z}_2 规范理论的对偶理论

无论是(18.36)还是(18.39)都具有 \mathbb{Z}_2 规范不变性，如果我们认为规范自由度不具有物理含义（它实际上有没有物理含义取决于我们关心的物理量是不是只涉及规范不变量），那么这两个哈密顿量就含有额外的自由度。我们要设法把规范等价的构型全部映射到同一个构型上，而把规范不等价的构型映射到不同的构型上。为此，我们将每个方块赋予一个格点坐标 I ，从而诸 $\{i\}$ 和诸 $\{I\}$ 形成对偶格点坐标。设 \square_I 为 I 号方块（ I 标记了以所有的方块的中心为格点形成的新格子的格点坐标，称为对偶格子），我们定义

$$\tau_I^x = \prod_{l \in \square_I} \sigma_l^z, \quad (18.42)$$

上标 x 看起来很奇怪,不过我们很快会发现其作用。这样(18.39)中的第二项就可以很容易地写出了。注意到 τ_I^x 只有 ± 1 两种取值,我们可以把它看成某个表象下的 x 方向泡利矩阵。至于第一项,如果将 σ_{ij}^x 作用在某个 σ^z 表象下的基矢量上面,那么边 ij 上的 σ^z 反号,其余什么都不变,这就是说,设边 ij 由方格 I 和 J 共享,则由定义(18.42), I 和 J 对应的 τ^x 也反号,其余不变;另一方面,将 $\tau_I^z \tau_J^z$ 作用在一个态上,则 I 和 J 对应的 τ^x 均反号(同样依据泡利矩阵的性质,即 z 方向泡利矩阵可以翻转 x 方向泡利矩阵的本征态)。两个算符的作用效果完全一样,所以实际上

$$\tau_I^z \tau_J^z = \sigma_{ij}^x, \quad (18.43)$$

从而我们得到

$$H = -K \sum_{\langle I, J \rangle} \tau_I^z \tau_J^z - J \sum_I \tau_I^x. \quad (18.44)$$

现在没有规范冗余了—— τ_I^x 和 τ_I^z 都是规范不变量。要看出自由度减少了多少,注意到二维格子中一个方块有四条边,每条边由两个方块分享,因此如果有 N 个方块(从而有 N 个格点),那么有 $2N$ 条边。另一方面,只有 N 个方格。因此如果只以 τ_I^z 为动力学自由度,则我们将希尔伯特空间的维数从 2^{2N} 降到了 2^N 。丢自由度是正常的,因为在以上过程中我们抛弃了规范自由度,但是需要验证只以 τ_I^z 为动力学自由度是不是把一些并非规范自由度的自由度(它们没有出现在哈密顿量中)也抛弃了。如果规范不等价的态给出不同的 τ_I^x 取值,我们就可以确定没有丢掉真正含有信息的自由度。

在上述从 σ^z 到 τ^x 的过程中我们丢失了一半的自由度,也就是说,有一半的自由度是可以随意指定的;指定这些自由度就是选取了一个规范;不过要注意,这不是说我们指定一半 σ^z 的值就是选取了一个规范,因为(18.39)中 σ^z 确定的波函数随着时间演化, σ^z 会发生变化,这和规范选取需要满足的要求不符合——例如电动力学中如果选择 $\nabla \cdot \mathbf{A} = 0$,那么它不会在随后的时间演化中变得不是零。

(18.44)正是横场伊辛模型。因此(18.39)也经常称为**伊辛规范理论**,因为它实际上就是横场伊辛模型换了一个形式。横场伊辛模型是一个二维量子模型,其零温配分函数的精确形式对应一个三维经典统计模型,实际上这个三维经典统计模型就是一个各向异性的伊辛模型(在虚时间上的最近邻相互作用和空间方向上的最近邻相互作用不同)。我们知道三维伊辛模型一定会出现相变,有一个顺磁相和一个铁磁相,这来自其普适类⁵,因此零温下横场伊辛模型——从而 \mathbb{Z}_2 规范场——也会有一个相变,随着参数 K/J 的变化,从一个相切换到另一个相。

有一个看起来的佯谬:我们知道伊辛模型有一个全局 \mathbb{Z}_2 对称性,可以将全部自旋翻转过来,得到一个不同的态,而与此同时哈密顿量保持不变。实际上这个对称性在 \mathbb{Z}_2 规范场中就是不存在的——等效的横场伊辛模型的相变破缺的就是这个对称性,由横场伊辛模型和 \mathbb{Z}_2 规范场理论的等价性我们知道 \mathbb{Z}_2 规范场理论也有一个相变,并且这个相变没有破缺任何对称性。因此,在这里我们就已经知道了 \mathbb{Z}_2 规范场有一个与对称性无关的相变了,它称为**Ising*相变**。和此相变有关的激发见第18.4.3节。

Ising*相变的存在本身已经说明,去掉一个规范理论中的规范冗余性会导致一些局域的信息变成全局的,一些全局的信息变成局域的。实际上,考虑到 \mathbb{Z}_2 规范场和费米子的耦合是 $\sigma^z c^\dagger c$ 形式的,而使用(18.44)中出现的量不能局域地表示出 σ^z ,将 \mathbb{Z}_2 规范场部分的规范冗余性去除必然导致 \mathbb{Z}_2 规范场和费米子耦合部分出现非局域性。这正是规范理论的作用:允

⁵虽然(18.44)对应的经典统计模型是各向异性的,这并不改变其普适类,因为总是可以适当调节 β 的尺度让该经典统计模型变成各向同性的。

许规范冗余性存在意味着原本非局域的理论现在是局域的了。在高能物理中这件事尤其重要——实际上粒子物理标准模型就是这么建立的（见相对论性量子场论笔记）。

18.4.3 低能自由度和拓扑序

任意子：规范荷和磁通量

在二维的电动力学中，设通过一个方格的磁通量为 Φ ，则

$$e^{i\Phi} = \prod_{l \in \square} t_{ij}.$$

二维情况下一个区域内有磁通量是真的可以认为这里有一个粒子的，因为显然磁场在这个区域内有一个鼓包，因此量子化之后会得到粒子；至于这样的粒子是不是足够稳定（或者说，能够让“磁通量粒子”呈现为谐振子能级的哈密顿量和实际系统是不是足够接近）则又是另一回事。在本文涉及的 \mathbb{Z}_2 规范场中可以如法炮制地定义

$$e^{i\Phi_I} = \prod_{l \in \square_I} \sigma_{ij}^z = \tau_I^x, \quad (18.45)$$

也即，我们用费米子在方块上转一圈发生的相位改变来定义磁通量。与 $U(1)$ 的情况不同， \mathbb{Z}_2 规范场中磁通量只有0和 π 两种，因为四个 σ^z 相乘要么是1要么是-1。现在我们看到了 τ^x 的另一重意义：它标记了一个方块上的磁通量。与电磁场中的较为复杂的情况不同， \mathbb{Z}_2 规范场中的磁通量就是肉眼可见量子化的，而且只有两个状态。注意到 τ_I^x 取1时能量较低而取-1时能量较高，我们可以将某个方块的磁通量取 π 当成一种激发态，称为m激发，以体现它和磁通量的相似之处。一个空间区域内有m激发等于是说存在一条围绕着该空间区域的 $\sigma^z = -1$ 弦，实际上是一个涡旋（可以考虑一下电动力学中的磁通量是什么样的），也可以称它为vison激发（Vortex-Ising-son，三个词根分别表示此类激发形如涡旋、和伊辛场有关、是元激发）。

另一个可以模仿电磁场引入的概念是 \mathbb{Z}_2 规范荷，我们已经看到， \mathbb{Z}_2 规范变换对应的规范荷为(18.38)，这个量的取值只有 ± 1 （因为是四个 σ^x 的乘积）。由于(18.38)守恒，我们有如下 \mathbb{Z}_2 规范场的高斯定律：

$$\prod_{j \in +} \sigma_{ij}^x = \mathbb{Z}_2\text{-charge at } i = \text{const}, \quad (18.46)$$

这个常数可以取1也可以取-1，但是没有时间演化。这个额外的条件将希尔伯特空间划分成没有重叠的很多支，不同分支的 \mathbb{Z}_2 规范荷分布不同。虽然规范荷通常是通过规范场和物质场的耦合项引入的，在积掉物质场之后还是可以构造出规范荷的表达式，因为规范不变性的要求极大地限制了规范场荷物质场耦合的方式。正如在电磁场中，即使我们积掉了物质场，麦克斯韦方程中还是会有一个电荷守恒方程

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{j} = 0$$

一样——即使我们不知道电磁场实际上和一个物质场发生了耦合，我们还是可以将电荷当成电场线的某种特殊分布（源和汇），而以它们为某种激发。同理，在 \mathbb{Z}_2 规范场中， Q_i 取-1意味着更高的能量（计算一下能量期望值就知道），那么我们可以认为某个点*i*处 $Q_i = -1$ 意味着这里出现了某个激发，从而一个 \mathbb{Z}_2 规范荷被放置在了这里，无论其背后的机制是什么，

无论是不是真的有一个物质场和 \mathbb{Z}_2 规范场发生了耦合。我们称这种激发为e激发，以体现它和电荷的相似性。e激发是 σ^x 弦的末端。

实际上，从 \mathbb{Z}_2 规范变换的定义可以发现，在规范变换下 σ^z 算符变号而 σ^x 算符不变号，又注意到e激发和m激发的定义，我们会发现，大体上有

$$\sigma^z \sim e^{iA}, \quad \sigma^x \sim e^{iE}.$$

连接两个e激发的 σ^x 弦可以看成电场线；有许多可能的连接两个e激发的 σ^x 弦这件事对应于有多条电场线。注意，虽然一个经典电场经常被画成一族电场线，可是其中并没有良定义的单独一条电场线；有两个确定的e激发的多体波函数（经典地看，其中应该有确定的电场）是多个 σ^x 弦构型的线性叠加这件事就是前述“经典电场中无良定义的单独一条电场线”这个事实的量子版本。 σ^z 为负的bond连接成的线（注意，不是 σ^z 弦，而是处处和 σ^z 弦垂直）可以看成磁感线。

(18.46)和麦克斯韦方程导出的电荷守恒方程有一个重要的区别，就是前者要求规范荷在每一点都守恒，而后者允许规范荷的流动。但这个区别实际上并没有什么物理意义。麦克斯韦方程本身也不规定电荷应该如何流动（这是本构关系应该做的事情），因此，每一点的电荷密度算符和纯电磁场的哈密顿量也是对易的。麦克斯韦方程中会出现电流密度单纯是因为这本质上是一个洛伦兹协变的理论，从而如果一个四维矢量的时间分量出现了，其空间分量也会出现。类似的，协变密度泛函理论中用于标记系统基态的物理量不仅包括电子数密度也包括电子流密度。

(18.39)中没有引入任何真的携带 \mathbb{Z}_2 规范荷的物质场，如果引入了，本节中的 H 就只是 \mathbb{Z}_2 规范场的哈密顿量而不是完整的哈密顿量了，此时(18.46)的第一个等号当然仍然成立，但是每个点上的 \mathbb{Z}_2 规范荷就未必总是不变的了，虽然规范对称性要求规范荷总量保持不变；我们在(18.40)中看到的就是这种情况。我们将(18.39)中没有动力学的规范荷称为测试规范荷，这个名称的意味是显然的；它相比有动力学的规范荷更容易处理，后者的哈密顿量在坐标表象下基本上不会是对角化的。

还有一个表面上的不同是，麦克斯韦方程中显式地给出了电荷密度和电流密度，而此处的 \mathbb{Z}_2 规范场理论中似乎没有。但这个其实只是记号的问题，电磁场的哈密顿量中实际上也没有出现电荷密度和电流密度。麦克斯韦方程中会有电荷密度和电流密度是因为当我们从电磁场哈密顿量数学地导出运动方程时我们发现 \mathbf{E} 的散度和 \mathbf{B} 的旋度减去电场的时间导数似乎分别是某个东西的密度和流，从而暗示了“规范荷就在这里”，正如我们发现 \mathbb{Z}_2 规范场中似乎有某个守恒荷(18.38)一样。

最后一个疑难是，从无物质场的电动力学的拉氏量推导出的麦克斯韦方程显式地指明了空间中不存在电荷（因为 $\nabla \cdot \mathbf{E} = 0$ ）。然而这实际上是对“能量最低”的不同理解造成的：在没有有动力学的物质场的情况下， $\nabla \cdot \mathbf{E}$ 没有时间演化，即是守恒量；如果我们固定各点的 $\nabla \cdot \mathbf{E}$ 不动，对电动力学的拉氏量做有约束的最小化，就能够得到有测试电荷的麦克斯韦方程组；另一方面，如果不施加任何约束，那么只能得到没有任何电荷的麦克斯韦方程组。这里的关键在于，由于没有物质场，测试电荷——如果有的话——是完全不会变动的，初始场构型中测试电荷是多少，不管做怎样的时间演化，它都保持不变。因此，从一个有测试电荷的场构型出发，必须手动将“存在测试电荷”这一信息以约束的形式引入，才能够得到正确的时间演化方程，因为测试电荷分布不同的场构型彼此不可能从一个演化到另一个，而不施加约束地最小化作用量却是在假定测试电荷能够变动。从正则量子化的角度看，各点上的测试电荷都是守恒的，测试电荷分布不同的态之间不可能通过时间演化连接，即希尔伯特空间根据测试电荷的分布被划分为了不同分支。我们在研究 \mathbb{Z}_2 规范场时

没有遇到这个疑难，因为一直在使用正则量子化的语言。

有意**不使用**一个特定的场引入e激发（例如，在电动力学中，电子场引入携带着电荷的电子）而是使用上面的方式，通过规范理论内部的场构型引入e激发是有好处的。我们在二维平面内工作，所以（正如后面会看到的那样）很可能出现任意子。如果我们使用一个费米场引入e激发，那分析一个e激发和其它任意子的交换相位时就不必要地产生了额外的复杂性，例如，两个e激发交换同时需要考虑 \mathbb{Z}_2 规范场导致的交换相位和费米子的交换相位。相反，我们采取一种反过来的观点：认为 \mathbb{Z}_2 规范场自身的特殊构型已经产生了e激发，而 \mathbb{Z}_2 规范场和费米子耦合的哈密顿量（如(18.36)）的作用是“把e激发和费米子粘在一起”。⁶这样e激发是一个独立的单元，分析起来方便很多。e激发被粘到外界粒子上这个过程的一个直观的展示见资料框18.2。

可以看到e激发和m激发的定义和Toric-code模型中的定义完全一样——实际上，这就是为什么Toric-code模型中的两种激发被命名为e激发和m激发。相应的诸如弦算符、开弦、闭弦等的定义也和Toric-code模型中完全一样，e激发和m激发之间的交换相位也是一样的。

\mathbb{Z}_2 规范理论中的e激发是完全守恒的，没有任何时间演化：如果一开始某个格点上没有e激发，那么就是没有。最为简单的情况当然是系统中没有任何e激发，即对所有 i ，施加约束

$$Q_i = 1, \quad (18.47)$$

或者等价地说，往哈密顿量中引入一个罚函数

$$H = -g \sum_i Q_i, \quad g \gg 1, \quad (18.48)$$

这会让体系变成一个toric-code模型，其中e激发是一个能隙无限大的元激发。约束(18.47)可以看成是一个gauge sector，其基态为全体 σ_{ij}^z 有确定的值1，激发态只有m激发或者说vison。我们也有其它的gauge sector。做一个局部的规范变换会改变弦的走向，但是不会更动e激发和m激发的位置。

上述事实意味着从拓扑序的角度看，无物质场耦合， $K = 0$ 的(18.39)——所谓的Wegner模型——实际上是非常病态的一个东西。由于e激发完全没有出现在哈密顿量中，m激发分布相同的态全部是能量简并的，从而，何为能量本征态的合适标签——或者说我们要在哪一组表象下工作——实际上是不清楚的。在Wegner模型中，如果我们在 σ^z 的本征态附近工

⁶这个说法的合理性同样可以通过和电动力学相对比看出来。QED中量子化后的光子模式就只有两个，但是光子内线实际上可以理解成有三种光子模式，其中还有一个“有源无旋”的纵波“光子”模式，它对应静止电荷激发出的库伦场，正是它给出了库伦相互作用。这种模式不能在真空中远程传播，在没有源的时候无法传播，所以很自然，不可能出现在外线中。

但是，如果我们要抽象地分析“电荷激发出的场会有什么性质”，那么还是要把这个有源无旋的模式找回来。此时哈密顿量中电荷和电磁场耦合的项无非就是将电荷和这个有源无旋的模式结合在一起。

一个可能的问题是，e激发和m激发会出现在 \mathbb{Z}_2 理论的能谱中，但是将麦克斯韦理论做正则量子化得到的能谱中只有光子。实际上，这和我们将场算符展开成产生湮灭算符的傅里叶变换的步骤有关。设想我们希望不引入任何物质场，而是观察电场构型中有无库伦势场来判断一个地方有没有电荷，那么这就意味着一个库伦势场需要被量子化成一个e激发模式。可是，库伦势场根本无法良定义地写成一系列傅里叶分量的叠加：变换以后得到的 $2\pi/\mathbf{k}^2$ 在 $\mathbf{k} \rightarrow 0$ 处发散。在研究库仑相互作用电子气时，我们可以通过晶格的自相互作用引入的能量在哈密顿量中抵消这个发散，在这里则不行。因此，一般的通过傅里叶变换正则量子化电动力学以后，能谱中是不会出现电荷的。

或许一个更加值得问的问题是为什么离散规范场没有规范玻色子。直观地看，这是因为离散规范场不能产生“连续”的涨落。要什么样的离散格点模型能够演出规范玻色子呢？文小刚在他著名的量子多体书的最后一章给出了一种使用quantum rotor model演生 $U(1)$ 规范场的方法，其中导致了无能隙规范玻色子的哈密顿量为 $(S_{i+\mathbf{a},j}^z - S_{i,j-\mathbf{g}}^z)^2$ 。在转子模型中， S^z 可以取无限个值，在低能、大尺度理论中能够产生足够“连续”的变化，配合前述哈密顿量，能够形成弦密度波，从而形成（底层机制非常非局域的）规范玻色子。在自旋模型中无法做到这一点。使用经典图像思考的话，考虑百变尺和跳绳的行为的区别，就能够理解了。

作，那么态的规范不变性就没有了，然而此时我们仍然能够良定义m激发。一个 σ^z 的本征态是若干具有不同e激发分布的横跨多个gauge sector的态的叠加，其中观察不到清晰的e激发。由于 σ^x 弦在态中是完全缺席的，此时的vison激发和普通的涡旋毫无区别：没有任何非平庸交换相位。这样的模型似乎很难说有什么拓扑序，vison和二维经典XY模型中的涡旋或是一些磁体中的skyrmion没有什么本质区别。然而，我们这里研究 \mathbb{Z}_2 规范场是为了研究自旋液体的行为，因此讨论e激发仍然是必要的。讨论e激发的一个用处如下：如果在纯 \mathbb{Z}_2 规范场中引入两个e激发，发现系统能量正比于e激发间距，即e激发是禁闭的，那么(18.40)中的费米子——在自旋液体中就是spinon——也是禁闭的。存在清晰可辨、分布确定的e激发的某个态中， σ_i^z 不再是确定的了，于是实际上此时系统基态是一个弦网凝聚态，这样才能让 Q_i 有确定的值；在这样的态中仍然有确定的vison，因为它是Wegner模型的守恒量。因此，此时体系的激发态中存在清晰可辨的e激发和m激发，从而有交换相位等等，即系统具有拓扑序。

可见，在引入e激发之后（或者等价地说，向哈密顿量中引入(18.48)，但是让 g 很小，来制造量子涨落），我们就得到了拓扑序。因此，与资料框18.2类似，可以将使用 σ^z 的能量本征态定义的vison激发（可以说是**经典Wegner模型中的vison**）看成 \mathbb{Z}_2 拓扑序中的、使用弦网凝聚态定义的m激发的前体（因为我们有弦算符、连接在弦算符两端的“粒子”等结构，粒子的交换对应弦的交叉，可以引入非平庸相位）。正如资料框18.2通过向原本会产生被弦连接的点状粒子的XY模型引入强烈的量子涨落，能够得到演生规范场和拓扑序一样，此处我们通过引入 σ^x 项来引起量子涨落，让e激发变得重要，且e激发和m激发的弦的交叉引入非平庸相位，能够从原本平庸经典Wegner模型中的vison出发获得拓扑序。当然，资料框18.2中的拓扑序可以使用 \mathbb{Z}_2 规范理论充当有效理论，因为XY涡旋和 \mathbb{Z}_2 规范理论的m激发绑定在一起，而这里讨论的从Wegner模型出发得到的拓扑序中，有效理论就是 \mathbb{Z}_2 规范理论本身而已。

使用第19.1.1节中的语言说，(18.39)和(18.44)都可以看成弦网模型。看起来，(18.39)中的弦是 σ^x 弦，其第一项是弦张力项，第二项是弦动能项，而(18.44)中的弦是 τ^x 弦，其第二项是弦张力项而第一项是弦动能项。以 τ^x 弦为标准看待(18.39)，会发现 $\sum \sigma^x$ 项，也就是 σ^x 弦的弦张力项，实际上是 τ^x 弦的弦动能项，而 $\sum_{\square} \prod \sigma^z$ 项，也就是 σ^x 弦的弦动能项，实际上是 τ^x 弦的弦张力项，虽然看起来形式复杂。不过，既然我们关注的是 \mathbb{Z}_2 规范场中的 \mathbb{Z}_2 磁通量，实际上不应该将(18.44)看成一个弦网模型，而是应该将它看成粒子模型：横场伊辛模型中， $\tau^z \tau^z$ 项可以视为m激发的跃迁项，因为它可以让一个m激发跃迁到临近的格点上。

(18.39)中的 $\sum_{\square} \prod \sigma^z$ 。在Toric-code模型中十分幸运的是哈密顿量中的所有项彼此对易，并且哈密顿量在将 σ^x 和 σ^z 交换，格子和对偶格子交换的操作下保持不变，从而。

我们尚未讨论在引入横场项（或者说弦张力项）之后，定义e激发和m激发是否有意义。如果系统的低能自由度中实际上没有清晰可辨的这两种激发，那么引入这些概念毫无意义。e激发的意义是规范荷（当然，是根据 \mathbb{Z}_2 规范场构型而定义的，相当于电动力学中的纵波光子），m激发的意义是磁通量。在规范理论中一种常见的让它们不属于低能激发的机制是**禁闭**。一个规范理论的禁闭相中，规范荷之间存在强烈的相互吸引，或者说连接规范荷的弦有非常强的弦张力，从而规范荷被强烈地束缚在一起；另一方面，规范场模式本身期望不大而量子涨落很大，从而系统的低能自由度中同样没有规范场模式。因此如果(18.39)处在禁闭相，那么e激发和m激发都是没有意义的。

具体什么激发被禁闭什么被解禁闭当然由 J 和 K 调控。我们在第18.4.4节中分析 \mathbb{Z}_2 规范

场的量子相变。

电场和磁场

实际上(18.39)还能够写成 \mathbb{Z}_2 “电场”和“磁场”的函数。

Toric-code拓扑序

有一点需要注意：虽然(18.44)和(18.39)是对偶的，但在对称性、局域性等方面两者还是不一样的。就冗余性而言，(18.39)有 \mathbb{Z}_2 规范对称性，但(18.44)只有全局 \mathbb{Z}_2 对称性；这两个模型实际上都是有冗余的，后者的冗余意味着什么我们马上可以看到。同样，在(18.39)中显而易见的局域约束，在(18.44)中也会以一种非常不平凡的形式呈现出来。

vison激发一定成对出现，通过分析 \mathbb{Z}_2 规范场理论的希尔伯特空间可以直接确认这一点（我们在Toric-code模型中已经做过这件事了）。这就是说，vison激发不能够使用局域的算符单个地产生。表面上，与之对偶的横场伊辛模型(18.44)可以有奇数个vison激发，因为可以将奇数个方块上的 τ^x 设置为 -1 ，但实际上这来自横场伊辛模型的全局 \mathbb{Z}_2 冗余性。将 K 调大，则 τ^x 激发会越来越倾向出现，当出现得足够多时发生凝聚，此时就发生相变，从解禁闭相进入禁闭相。因此 \mathbb{Z}_2 规范场的解禁闭相中确实有一个序，这个序就体现在存在vison激发上，但是无法使用一个局域的序参量表示这个序。

总之，实际上 \mathbb{Z}_2 规范场的确有一个量子相变，然而这个量子相变和没有对称性任何关系：其有序相中特有的低能自由度是一些非局域的弦算符两端的任意子，而不是任何局域的序参量。 \mathbb{Z}_2 规范场理论(18.39)有零温量子相变，牵涉其中的两个相分别是任意子禁闭相和解禁闭相。从横场伊辛模型(18.44)中可以看到 m 激发是有能隙的，同样可以说明 e 激发也是有能隙的。 e 激发和 m 激发可能有吸引相互作用，使得它们聚合形成的激发无能隙，不过此时单独讨论 e 激发和 m 激发又变得没有意义了（与费米子类比，此时已经出现费米子配对了）。因此，解禁闭相中存在有能隙的任意子，这些任意子的任意子表等和Toric-code模型完全一样。于是这个 \mathbb{Z}_2 拓扑序称为**Toric-code拓扑序**，因为它和toric-code模型的拓扑序是一样的。此处“一样的”指的是阿贝尔拓扑序具有同样的任意子表和交换相位，即具有等价的张量范畴，并不代表其任意子动力学完全相同，不过任意子动力学在拓扑序研究中也并不是非常重要（见背景知识17.3）。我们可以认为Toric-code模型中的拓扑序构成了一个普适类，而Toric-code模型作为一个严格可解模型即为这个普适类中的重整化群不动点。

18.4.4 零温量子相变

对偶理论(18.44)的视角

根据 \mathbb{Z}_2 规范场和横场伊辛模型的对偶关系， \mathbb{Z}_2 规范场的确有一个零温量子相变。三维伊辛模型具有一个顺磁相和一个铁磁相，由于铁磁序的形成需要更多相互作用， J/K 超过某个点时会出现顺磁相，否则出现铁磁相。映射回二维横场伊辛模型，顺磁相意味着对偶格子上的各个 τ 基本指向 x 方向，即有确定的磁通量；铁磁相意味着对偶格子上的各个 τ 基本指向 z 方向，且要么几乎都为1要么几乎都为 -1 ，没有确定的磁通量。当 J 相对 K 很大时，对偶理论(18.44)处于顺磁相，此时的基态几乎就是每个 σ^z 都取1的态，此时直觉上看，费米子可以畅行无阻；而当 J 相对 K 很小时，对偶理论(18.44)处于铁磁相，此时的基态不是 σ^z 的本征态，投影在 σ^z 上有正有负，那么费米子似乎会被“迷惑”住（回忆一下， σ^z 实际上是跃迁矩阵元），不知道怎么走。那么，我们可以合乎情理地猜测，三维伊辛模型的顺磁相对

应着 \mathbb{Z}_2 规范场模型的一个解禁闭相，在其中 \mathbb{Z}_2 规范荷清晰可辨，而三维伊辛模型的铁磁相对应着 \mathbb{Z}_2 规范场模型的一个禁闭相，在其中携带有 \mathbb{Z}_2 规范荷的费米子不能够单个的自由移动。

这样的论证当然是不够的，所以下面做一些半定量的论证。考虑(18.39)的格点路径积分（即虚时间轴也是离散化的），也即它对应的三维模型，定义

$$W(C) = \left\langle \prod_C \sigma_l^z(\tau) \right\rangle, \quad (18.49)$$

其中期望值内部的算符乘积称为**Wilson环算符**， C 是一个在虚时间方向上有延展的环路。容易看出 $W(C)$ 给出了从某个虚时间点开始产生一对e激发，按照 C 指示的轨迹扩散，经过一段时间之后又湮灭的概率（在经典理论中这是良定义的，因为没有任何不确定关系；具体有没有手段可以用量子测量的标准方法测出这个概率则是另外一回事），随着 C 扩大它理所当然会衰减，如果随着 C 扩大它衰减得很快那么就说明e激发总是成对地被束缚在一起。我们本来可以使用两点格林函数来表征e激发被束缚的强度的，但是由于两点格林函数不是规范不变的，任何两点格林函数都是零。 $W(C)$ 的衰减方式有两种典型的形式：一种是**面积定律**，即

$$W(C) \sim e^{-A}, \quad (18.50)$$

其中 A 是 C 围成的面积，另一种是**周长定律**，即

$$W(C) \sim e^{-L}. \quad (18.51)$$

设 C 持续时间为 τ ，则按照虚时间演化，应该有

$$W(C) \sim e^{-\tau \Delta E},$$

其中 ΔE 是两个e激发同时出现造成的能量上升。我们马上可以看出，由于

$$A \sim \tau R,$$

其中 R 是两个e激发的距离，如果面积定律成立，必然有

$$\Delta E \sim R,$$

这是典型的禁闭效应：两个e激发离得越远，能量越高，当能量高到一定程度时涨落会导致新的e激发产生，从而产生两对离得非常近的e激发。另一方面，如果周长定律成立，则

$$\Delta E \sim 1 + \frac{R}{\tau},$$

而由于粒子通常不会跑太远，可以认为 ΔE 基本上是一个常数，因此没有禁闭效应。

那么，前述横场伊辛模型给出的两个相是不是分别对应面积定律和周长定律？我们尝试将 $W(C)$ 映射到横场伊辛模型中，并在二维平面上计算 $W(C)$ 。这实际上是有问题的，因为在使用 $W(C)$ 的行为判断是否有禁闭时，要求在时空上做Wilson回路，但此处我们单纯是在空间上做Wilson回路。在 \mathbb{Z}_2 规范场理论中这不会造成什么问题，因为它对应一个二维横场伊辛模型，后者对应一个三维经典伊辛模型，于是就有时间和空间的等价性。由于Wilson回路能够使用的大部分模型都具有类似的性质，使用二维平面上计算出的Wilson回路来判断是否禁闭不会有问题，否则需要寻找别的回路算符。因为 C 围成的区域内部的 σ^z 会被乘两次，于是就给出1，于是

$$\prod_{l \in C} \sigma_l^z = \prod_{l \in D} \sigma_l^z,$$

依照 τ^x 的定义即得到

$$W(C) = \left\langle \prod_{I \in D} \tau_I^x(\tau) \right\rangle, \quad (18.52)$$

其中 D 是 C 围成的区域。由于只有两个相，可以在 J/K 很小或很大时分别做微扰论，由此得到的关于相的结构的信息在整个相内部都是成立的。 $J \gg K$ 的情况对应顺磁相，系统基态形如

$$|\text{ground}\rangle = |\tau^x = \rightarrow \rightarrow \cdots \rightarrow\rangle + \frac{K}{J} \sum_{i,j} |\tau^x = \rightarrow \cdots \leftarrow_i \cdots \leftarrow_j \cdots \rightarrow\rangle + \cdots,$$

其中 i 和 j 相邻。上式看起来很奇怪，不过真的去算期望会发现翻转两个 τ^x 比翻转一个能量更低。（翻转相邻的两个 τ^x 的话， $\tau_i^z \tau_j^z$ 项的期望不为零）如果格点 i 和 j 在 C 内部，那么它不会对期望值有什么贡献，因为 -1 平方还是 1 ；格点 i 和 j 在 C 外部那肯定也不会有什么贡献。唯一会让期望值偏离 1 的激发是 i, j 中一个在 C 内部一个在 C 外部，于是我们有

$$\left\langle \prod_{I \in D} \tau_I^x \right\rangle \sim 1 - \frac{K}{J} L(C).$$

那么，合理的猜测是，在顺磁相中应该有周长定律成立，于是没有禁闭。铁磁相的讨论是类似的。我们仅仅讨论了两个极限情况，没有得到完整的 e 指数，但是一般的情况计算起来是非常困难的，此处略去。

总之，铁磁相（ $J/K \ll 1$ ）对应禁闭的 \mathbb{Z}_2 模型的状态，顺磁相则对应非禁闭的 \mathbb{Z}_2 模型的状态。禁闭相中，vison激发和 \mathbb{Z}_2 规范荷不再是有意义的低能自由度。由于禁闭相的基态非常接近 τ^z 的本征态，在其上讨论 τ^x 的排列模式——也就是 m 激发——并没有意义，在其中 m 激发有很强的量子涨落，然而没有任何比较稳定的低能激发可以使用 τ^x 标记。因此， m 激发或者说vison激发不是有意义的低能自由度。 e 激发不会被量子涨落抹去，然而注意到 $J = 0$ 时

$$H = -K \sum_{\langle i,j \rangle} \sigma_{ij}^x,$$

现在设一条弦两端各有一个 e 激发，则弦上各点都有 $\sigma^x = -1$ ，于是系统能量相比于基态能量的增量就是弦的长度乘以 K ，因此会有一个全局不衰减的力将两个 e 激发拉在一起，即携带 \mathbb{Z}_2 规范荷的费米子无法单个地自由移动。因此禁闭相中的“ \mathbb{Z}_2 激发”—— e 激发和 m 激发——都没有意义，即在这里 \mathbb{Z}_2 规范场的行为并不十分有趣。反之，在 $K = 0$ 时， m 激发没有量子涨落，且 e 激发之间没有相互作用，因此 e 激发和 m 激发都是能够存在的。与通常的情况不同，在 \mathbb{Z}_2 规范场模型中，顺磁相是非平凡的，铁磁相反而是平凡的。

禁闭相的性质

应注意vison激发被禁闭这件事并不意味着系统的低能能量本征态中真的没有vison激发，正如虽然夸克被禁闭了，但是显然有质子、中子这些粒子的地方就有夸克。vison被禁闭意味着，含有**清晰可辨**的vison的状态不会出现，但是低能能量本征态中有若干含有vison的态叠加在一起是非常正常的。实际上就是这样： $K \gg J$ 时基态波函数在 σ^z 表象下的不少分支都含有vison，然而vison激发也的确是被禁闭的。因此这里又有一个迷惑性的地方： \mathbb{Z}_2 解禁闭相的基态中 \mathbb{Z}_2 磁通量是被**排除**的，但是低能激发中确实有 m 激发； \mathbb{Z}_2 禁闭相的基态中确实是有（很多的） m 激发的，但是低能激发中没有稳定存在的 m 激发。因此， \mathbb{Z}_2 的禁闭相其实可以看成**任意子凝聚相**，而解禁闭相则是**弦网凝聚相**。

从禁闭相和解禁闭相均为有能隙相而它们的基本自由度完全不同这件事出发也可以看出零温量子相变的存在。

18.5 二维平面上格点 $U(1)$ 规范场论

我们现在来分析平面上的 $U(1)$ 格点规范场论的性质。与 \mathbb{Z}_2 规范理论不同，规范群为李群的规范理论中，规范场的元激发一般是无能隙的——电动力学就是一个典型例子。因此，如果一个自旋液体的演生规范场的规范群是李群，那么其行为和一般意义上的（有能隙）拓扑序会有一些不同之处。

在格点规范场中允许存在磁单极子；此时它称为紧致的规范场，因为规范群是紧致的。实际上，电动力学在格点上的规范群并不是 $U(1)$ 的，因为0和 2π 不等价。

第十九章 弦网凝聚

自旋液体的丰富演生行为——尤其是演生规范场——让我们

19.1 Levin-Wen模型

19.1.1 弦网系统的不动点和Levin-Wen模型

允许二分支的弦网系统

Levin-Wen模型是^[18]中提出的一类模型。我们在Toric-code模型中已经看到过，在晶格的每个边上放置特定的自由度，则它们的特定取值可以排列出“弦”，这些弦或者两头连接着演生粒子，或者能够将一个演生粒子转移到另一个地方。Levin-Wen模型不首先考虑晶格的边上的自由度应该是什么，而是以这些弦为起点，定义希尔伯特空间和哈密顿量。

\mathbb{Z}_2 理论中的弦是没有定向的，但是一些其它的理论——比如说 $U(1)$ 规范理论——中的弦就是有定向的。一条弦的弦算符做共轭转置之后显然会给出一条定向相反的弦，而如果构成弦的局域算符不是厄米的，那么弦的类型也会发生变化。此外，不妨稍微推广一下，允许弦“分支”，这样弦算符不仅能够转移粒子，还能够描述粒子的聚变或是裂变。这样，一个弦网系统中的自由度由以下数据描述：

1. 弦的类型：设共有 N 种弦，用 $i = 1, 2, \dots, N$ 标记。
2. 弦的共轭转置和定向： i 类型的正向弦做共轭转置，得到 i^* 类型的反向弦；显然 $i^{**} = i$ 。如果 $i = i^*$ ，那么这种弦实际上是没有定向的。
3. 分支规则：一系列三元组 $\{i, j, k\}$ ，描述怎样的弦分支是可能的；需要注意有可能三个确定的 i, j, k 有不只一种连接方式，因此实际上在顶点上有可能也需要放置一些自由度，用于确定连接到这个顶点上的三根弦的连接方式具体是哪一种。

给定弦的类型的集合 \mathcal{I} ， $*$ ： $i \rightarrow i^*$ 运算，以及分支规则 $\{\{i, j, k\}, n\}$ ，一个弦网系统中允许的弦网构型就确定了，希尔伯特空间为

$$\mathcal{H} = \otimes_s \mathcal{H}_s \otimes_v \mathcal{H}_v, \quad \mathcal{H}_s = \mathbb{C}^{\mathcal{I}}, \quad \mathcal{H}_v = \oplus_{i,j,k} \mathbb{C}^{\text{ways of connection of } i, j, k}, \quad (19.1)$$

其中 s 表示边而 v 表示格点。

我们通常认为没有弦经过的点被某种“零号弦”占据着。这样， i 类型的弦能够是开弦这件事就等价于分支 $\{i, 0, 0\}$ 是允许的，而一条 i 号弦上任意一点可以视为一个 $\{i, i, 0\}$ 分支。于是非常自然地，我们用 δ_{ijk} 表示允许的弦分支：

$$\delta_{ijk} = \begin{cases} 1, & \{i, j, k\} \text{ allowed}, \\ 0, & \text{otherwise}, \end{cases} \quad (19.2)$$

那么

$$\delta_{ij0} = \delta_{ij}, \quad (19.3)$$

这里右边的 δ_{ij} 就是通常的Kronecker δ 符号。

可以看出前述定义中，完整描述一个弦网系统的数据 $(\mathcal{I}, *, \{\delta_{ijk}\}_{ijk})$ 实际上就是一个unitary dagger category \mathcal{C} ，其对象为弦的末端的可能构型（单个点，两个点，多个点……），态射为弦，任意两个对象之间的态射是 \mathbb{C} 上的向量空间。显然 \mathcal{C} 中的对象也可以使用弦来标记：将 i 类型弦的末端也表示为 i 即可，如果某个对象是多种弦的末端，把它表示成 $i \otimes j \otimes k$ 即可。注意在弦网模型中，弦一方面是算符，一方面，一个弦网算符作用在一个“空”态上就给出了一个弦网态，即弦也标记了系统的某一组基底。这样就有

$$\mathcal{H}_v = \otimes_{i,j,k} \text{hom}(i \otimes j, k). \quad (19.4)$$

我们目前还没有确定该范畴是不是张量范畴，或者说没有确定如此定义的 \otimes 是不是一个好的张量积，因为尚未确定 $(i \otimes j) \otimes k \simeq i(j \otimes k)$ 之类的关系是不是成立。

以上对弦的作用的阐述——准粒子的产生、移动、聚变、裂变等——都是运动学； \mathcal{C} 中的态射是否能够诠释为这些物理过程取决于哈密顿量是否提供了这样的相互作用通道。弦网系统的哈密顿量大体上应该具有这样的形式：

$$H = tH_t + UH_U, \quad (19.5)$$

其中 H_U 是一个和（坐标表象下的；以下不注）弦网算符对易的厄米算符，可以看成弦网的势能，而 H_t 是一个和弦网算符不对易的算符，它的本征态是多个弦网态的线性组合，它是弦网的动能，因为它让弦网有量子涨落，也即，有时间演化。这和粒子模型的哈密顿量的构造的思路是完全一样的。 H_U 可以有很多取值，例如设 σ^x 是某种弦的局域算符，则

$$H_{\text{tension}} = \sum_{\langle i,j \rangle} \sigma_{ij}^x \quad (19.6)$$

赋予了弦一个张力，因为较长的弦具有较大的 H_{tension} 。另一种弦势能可以这么构造：注意到算符 $\prod_{p \in +i} \sigma_p^x$ 在弦的末端为1而在其它地方为零，哈密顿量

$$H_{\text{charge}} = \sum_i \prod_{p \in +i} \sigma_p^x \quad (19.7)$$

统计了全空间中弦末端的数目，实际上就是统计了和这种弦有关的准粒子的个数。 H_{charge} 的作用和连续介质电动力学中电荷的能量很相似：在弦网模型和连续介质电动力学的麦克斯韦方程部分中都没有出现电荷的动力学，但是弦或是规范场（我们之前已经看到这两者似乎非常相似，我们之后会看到它们为什么有关系）的构型会暗示某些地方按理来说应该有电荷，从而可以良定义“电荷自己贡献的能量”，虽然这种电荷实际上是规范场或是弦的特殊构型对应的元激发，并没有被某种规范场或是弦以外的物质场携带着。

我们已经见过一种这样的模型了： \mathbb{Z}_2 规范场。可以认为 \mathbb{Z}_2 规范场中只有一种弦，是 σ^x 构成的，哈密顿量中， $\sum \sigma_i^x$ 项就是一个弦张力项，而 $\sum_p \prod_i \sigma_i^z$ 项则是一个弦动能项。

$t \gg U$ 时系统基态是大量不同弦网态的线性组合，由于弦的非局域性，可以预期有系统拓扑保护的基态简并，而 $U \gg t$ 时系统基态中没有多少弦。因此弦网模型总是有一个从弦网凝聚相到平凡相的量子相变。弦网凝聚相显然是一个拓扑序：它有长程纠缠，并且有拓扑保护的基态简并。弦网系统的哈密顿量可以是完全局域的、时间反演不变的，但是弦网凝聚态中存在非局域的弦网，并且准粒子——即弦的末端——的移动是被弦网构型记录下来的。这暗示着弦网模型有着丰富的演生行为。

波函数重整化

要从一个特定的弦网模型中提取出基态波函数中的拓扑序信息，可以做所谓的**波函数重整化**：不断抹去局部的细节，直到达到一个不动点¹。这个操作和金斯堡-朗道理论中计算临界行为的步骤差不多，不同的是由于我们不需要保留动力学，操作的自由性可以大很多，但同时也缺乏系统的做重整化的步骤，因此下面我们将只分析不动点处的哈密顿量和基态波函数应该满足什么性质。

由于我们只希望研究拓扑序，不动点处的哈密顿量应该是自由的，由于我们只关心弦网凝聚态，不动点处的哈密顿量即应该是一系列局域的弦动能的和。特别的，由于缺乏量子涨落，它应该是无阻挫的，从而能够将它划分成很多空间片段上的哈密顿之和，且在基态，每一个空间片段上的哈密顿量均被最小化。设 $|\Phi_i\rangle$ 是系统整体的基态 $|\Phi\rangle$ 在 i 空间片段的投影，则只需要

$$H_i |\Phi_i\rangle = E_{\min} |\Phi_i\rangle \quad (19.8)$$

我们就找到了基态。具体的 E_{\min} 的数值等并不重要——重要的是，我们发现，系统基态可以通过施加一系列局域的约束而获得。(19.8)是线性的，因此这些局域约束也应该是线性的，具体来说，设 $\{X_n\}$ 是空间片段 i 上的弦网构型，则(19.8)实际上是关于 $\{\langle X_n | \Phi_i \rangle\}$ 的一个线性方程。我们设 $\Phi(X_n)$ 表示 $\langle X_n | \Phi_i \rangle$ ，则局域约束就是关于 $\{\Phi(X_n)\}$ 的线性约束，其中 $\{X_n\}$ 是一系列通过 H_i 可以互相变换的同一个空间片段上的弦网构型。指定这些局域约束实际上就是在定义怎样的基态波函数算是在不动点上的。

这些约束在^[18]中可以找到（方程(4)到(7)）。

我们可能会担心这些要求是不是彼此冲突，例如是不是总是能够找到又是自由的、无阻挫的，又是拓扑不变的哈密顿量呢？实际上是可以的，因为仅仅有这些条件是不足以确定 \mathcal{C} 是什么的，因此实际上，这些条件不仅没有彼此冲突，还留下了一定的调整的余地。我们还可以再施加一些条件。一种常见的选择是要求弦的末端真的就是良定义的准粒子，这样 \otimes 是合理的张量积，从而

$$(i \otimes j) \otimes k \simeq i(j \otimes k),$$

从而能够定义所谓的**F-symbol**：在没有指明其意义时，这个条件的形式看起来比较奇怪。实际上它也不是唯一的选择。

可能并非所有弦网模型的波函数重整化不动点都满足这些条件。然而，满足以上条件的系统的行为已经足够丰富了，所以我们姑且暂时研究满足以上条件的系统。给定一个unitary fusion category \mathcal{C} （无需是braided fusion category）就能够确定一个弦网模型在波函数重整化下的满足以上约束的不动点，且所有这类不动点都可以由某个unitary fusion category生成。

Levin-Wen哈密顿量

我们现在引入Levin-Wen模型的哈密顿量，它们是严格可解的，可以将一套指定的弦网系统实现出来，并且满足之前提出的不动点条件。为了比较自然地引入“弦分支”这一概念，Levin-Wen模型定义在每个格点周围有三个边的晶格上。最简单的这种格子当然就是平面六角格子，不过实际上三维空间中也能够构造出这种格子，虽然看起来会很不寻常。

¹需注意这和高能物理场论计算中的“波函数重整化”毫无关系，后者更确切的名字是“场重整化”。

19.1.2 Levin-Wen模型的低能有效理论

费米子、规范理论

规范理论

Chern-Simons理论

19.1.3 边界态

第二十章 速查

20.1 晶格

20.1.1 电子自由度计数

见(4.39)前后的论述。概述如下：

- 电子波函数的轨道部分的数目等于晶胞个数乘以每个晶胞内的电子轨道个数，每个晶胞内的电子轨道个数等于晶胞内每个原子的电子轨道个数之和。这里也可以看出改变晶胞的定义不会影响自由度计数。紧束缚模型中的 i, j 对应晶胞的自由度， m, n 对应每个晶胞内的电子轨道自由度。
- 或者也可以用 i, j 对应原子的自由度（同时囊括晶胞位置和晶胞内的原子选取）， m, n 对应原子周围电子轨道自由度。
- 如果用简约动量和能带编号标记电子，则简约动量对应晶胞的自由度，能带编号对应每个晶胞内的不同原子的自由度和每个原子周围的电子轨道的自由度。因此第一布里渊区内简约动量的个数就是晶体中晶胞的个数。
- 当然也可以用自由电子动量和每个原子周围的电子轨道来标记电子。

不要忘了还有自旋标签。

20.1.2 晶体中的元激发

晶体中的准粒子波函数是空间群的么正表示。空间群的平移部分导致Bloch定理（见）准粒子能谱需要满足的一般性质、为何第一布里渊区如此重要等内容见第2.4.2节。

20.2 归一化常数

20.3 电子结构和物态

固体中的电子，经过电子-电子库伦相互作用和晶格作用，可能形成以下电子结构：

- 产生看起来像是电子的激发：
 - 布洛赫电子和Wannier电子，紧束缚模型，各向同性电子气模型（凝胶模型）
 - 这些可以归结为费米液体中的准粒子

这样的材料是寻常的金属、绝缘体、半导体等，总之就是可以使用能带论解释。

- 低温下的配对：
 - 超导
 - SDW, CDW
- 分数化激发

20.4 电动力学相关

20.4.1 导电性

会导致电阻的因素：

- 电子-电子散射，和温度关系不大。
- 电子-声子散射，在高温下由于能量均分定理正比于 T ，低温下正比于 T^5 。
- 与杂质、缺陷的散射。
- Kondo效应，见第10.2节。

在比较纯的晶体中电子-声子散射是主要的电阻原因。

电阻随着温度的变化大体上可以分成这么几种：

- 随着温度降低，电阻先大体上线性下降然后平缓，“饱和”了。前面线性下降段是因为这是高温下，后面饱和是因为杂质散射导致的电阻和温度关系不大。
- 随着温度降低，电阻先大体上线性下降然后平缓，然后快速降到零，因为发生了超导。最常见的超导机制是第8.1节给出的BCS理论。
- 随着温度降低，电阻先大体上线性下降然后平缓，然后反常地增大，因为Kondo效应，见第10.2节。

20.4.2 光学性质

- 晶格衍射，由静态晶格导致
- 光子与固体中的非局域玻色子模式耦合：
 - 光子和光学声子耦合，见第3.3.1节。
- 等离子激元，由电子导致

20.5 力学和热学性质

热容：

- 晶格振动，随着温度降低，依次是无温度依赖的经典值，正比于 $e^{-\omega/T}/T^2$ 的爱因斯坦模型和正比于 T^3 的德拜模型。
- 电子的贡献，正比于 T 。

传热：

- 电子气传热。

金属的热传导主要来自电子气。声子气同样会导致传导。

20.6 著名的模型

Hubbard模型，海森堡模型，

参考文献

- [1] Topological insulators. In *Contemporary Concepts of Condensed Matter Science*, page i. Elsevier, 2013.
- [2] Alexander Altland and Ben D Simons. *Condensed matter field theory*. Cambridge university press, 2010.
- [3] F. Aryasetiawan, K. Karlsson, O. Jepsen, and U. Schönberger. Calculations of hubbardfrom first-principles. *Physical Review B*, 74(12), Sep 2006.
- [4] János K. Asbóth, László Oroszlány, and András Pályi. A short course on topological insulators. *Lecture Notes in Physics*, 2016.
- [5] C. D. Batista and G. Ortiz. Generalized jordan-wigner transformations. *Physical Review Letters*, 86(6):1082–1085, Feb 2001.
- [6] Kerstin Beer, Dmytro Bondarenko, Alexander Hahn, Maria Kalabakov, Nicole Knust, Laura Niermann, Tobias J. Osborne, Christin Schridde, Stefan Seckmeyer, Deniz E. Stiegemann, and Ramona Wolf. From categories to anyons: a travelogue, 2018.
- [7] C. P. Brock, Th. Hahn, H. Wondratschek, U. Müller, U. Shmueli, E. Prince, A. Authier, V. Kopský, D. B. Litvin, E. Arnold, D. M. Himmel, M. G. Rossmann, S. Hall, B. McMahon, M. I. Aroyo, C. J. Gilmore, J. A. Kaduk, H. Schenk, C. Chantler, F. Boscherini, and B. Bunker. International tables for crystallography.
- [8] K Capelle, NA Lima, MF Silva, and LN Oliveira. Density-functional theory for the hubbard model: numerical results for the luttinger liquid and the mott insulator. In *The Fundamentals of Electron Density, Density Matrix and Density Functional Theory in Atoms, Molecules and the Solid State*, pages 145–168. Springer, 2003.
- [9] A. H. Castro Neto and Eduardo Fradkin. Bosonization of fermi liquids. *Physical Review B*, 49(16):10877–10892, Apr 1994.
- [10] AtMa P. O. Chan, Thomas Kvorning, Shinsei Ryu, and Eduardo Fradkin. Effective hydrodynamic field theory and condensation picture of topological insulators. *Physical Review B*, 93(15), apr 2016.
- [11] Mani Chandra, Gitansh Kataria, Deshdeep Sahdev, and Ravishankar Sundararaman. Hydrodynamic and ballistic ac transport in two-dimensional fermi liquids. *Physical Review B*, 99(16), Apr 2019.

- [12] Bob Coecke, editor. *New Structures for Physics*. Springer Berlin Heidelberg, 2011.
- [13] Eberhard Engel and Reiner M. Dreizler. *Density Functional Theory*. Springer Berlin Heidelberg, 2011.
- [14] Matthew Fishman, Steven R. White, and E. Miles Stoudenmire. The ITensor software library for tensor network calculations, 2020.
- [15] F. D. M. Haldane. Luttinger’s theorem and bosonization of the fermi surface. 2005.
- [16] Robert C Helling. Solving classical field equations. <https://ncatlab.org/nlab/files/HellingClassicalQFT.pdf>.
- [17] S. A. Jafari. Introduction to hubbard model and exact diagonalization, 2008.
- [18] Michael A. Levin and Xiao-Gang Wen. String-net condensation: a physical mechanism for topological phases. *Physical Review B*, 71(4), Jan 2005.
- [19] N. David Mermin. Thermal properties of the inhomogeneous electron gas. *Phys. Rev.*, 137:A1441–A1443, Mar 1965.
- [20] Román Orús. A practical introduction to tensor networks: Matrix product states and projected entangled pair states. *Annals of Physics*, 349:117–158, Oct 2014.
- [21] Aurora Pribram-Jones, David A. Gross, and Kieron Burke. Dft: A theory full of holes? *Annual Review of Physical Chemistry*, 66(1):283–304, Apr 2015.
- [22] Subir Sachdev. Topological order, emergent gauge fields, and fermi surface reconstruction. *Reports on Progress in Physics*, 82(1):014001, Nov 2018.
- [23] Subir Sachdev. Topological order in xy models in $d = 3$. http://qpt.physics.harvard.edu/phys268b/Lec9_Topological_order_in_XY_model.pdf, 2018.
- [24] R. Sedgewick, D. Scalapino, and R. Sugar. Fractionalized phase in an xy-z2 gauge model. *Physical Review B*, 65(5), Jan 2002.
- [25] Wang Shaofeng. Jordan-wigner transformation in a higher-dimensional lattice. *Physical Review E*, 51(2):1004–1005, February 1995.
- [26] J Sykes and GA Brooker. The transport coefficients of a fermi liquid. *Annals of Physics*, 56(1):1–39, 1970.
- [27] Susanne F. Viefers. Field theory of anyons and the fractional quantum hall effect. https://inis.iaea.org/collection/NCLCollectionStore/_Public/30/061/30061237.pdf, 1997.
- [28] Xiao-Gang Wen. *Quantum Field Theory of Many-Body Systems*. 09 2007.
- [29] I. Yurkevich. Bosonisation as the hubbard-stratonovich transformation, 2001.