

物质的相与相变

陈童

August 30, 2025

Contents

1 相变简介	2
1.1 相变的概念	2
1.2 克劳修斯-克拉珀龙关系	5
1.3 吉布斯相律	8
2 范德瓦尔斯模型	9
3 朗道范式	16
3.1 朗道的平均场理论	16
3.2 朗道-金兹堡理论	21
3.3 连续对称性自发破缺	24
3.4 朗道范式总结	31
3.5 应用举例：固体的弹性理论	33
4 附录：化学反应的热力学	40
4.1 化学平衡条件	40
4.2 质量作用定律	41

人们常说物质有三种状态(三种相): 固态、液态和气态。中国古代哲学则有五行学说, 大致是说物质有五种状态: 金、木、水、火、土, 这在今天已经被认为是一种迷信。

稍微了解一点科普的人可能知道, 至少还有一种“等离子体”态, 它构成了恒星的大部分。最初的那三种状态, 如今已演变成了众多类型。从重要且精确的意义上说, 磁体是一种独特的物质形态。金属与绝缘体不同。超导体和超流体则是引人注目的新物质状态。你电脑的液晶屏幕是众多不同液晶物质状态中的一种(向列相、胆甾相、蓝相I、蓝相II和蓝雾相、近晶相A、B、C、C*、D、I等)。存在超过200种具有质的差异的晶体类型, 更不用说还有准晶体了。还有像自旋玻璃这样的无序物质状态, 以及像分数量子霍尔效应这样具有电荷为 $e/3$ 的激发态(类似夸克)的物质状态。

当只有三种物质状态时, 我们可以逐一了解它们。这时候人们关注的重点在于这三种相是如何相互转变的, 也就是相变。理解物质三态之间的相变是本章的第一个主要目标。

然而, 如今物质状态多种多样, 我们就必须建立一套体系, 它要能一般性地分类物质的不同相, 正确描述不同相的不同物理性质, 甚至还要能描述不同相之间的相变。这个理解物态的体系主要是由朗道建立起来的。介绍朗道的这个体系是本章的第二个主要目标。

不过, 朗道这个体系并不能解释诸如分数量子霍尔效应这样的新物态, 这些新物态的发现曾经令人们大伤脑筋, 虽然最终人们弄明白了它们为何会有那样的表现, 并且由此拓展了我们理解物态的体系。但是, 如果你想进一步理解这些新的拓展, 那就需要在本书之外进一步去接触科研的前沿了。

未来也许还不断有新的物质相被发现, 而这些新相可能又不符合原有的框架, 因此我们描述物态的框架也就需要不断拓展和完善。

1 相变简介

1.1 相变的概念

水可以凝固成冰, 也可以气化成水蒸气, 气、水、冰本质都是由 H_2O 分子组成的物质, 然而它们的物理性质截然不同。这种在环境变化无穷小时,

物质的物理性质发生的不连续的改变，就是相变。气、水、冰就是宏观数量的 H_2O 分子的三种不同形态，或者说不同的相(phase)。这里值得注意的是，相变不是组成物质的分子的微观改变，也不是少量分子的改变，而是宏观数量的分子宏观物理性质的改变。

通常有不止一个变量可以影响物质的相。例如可以通过降低温度或者增加压强来使得蒸气凝结。我们使用相图来显示平衡相与温度和压强的关系(注意压强的单位， $1\text{bar} = 10^5\text{Pa}$, Pa表示帕斯卡，1bar大约就是一个标准大气压(atm)，实际上1atm约为 1.01325bar)。

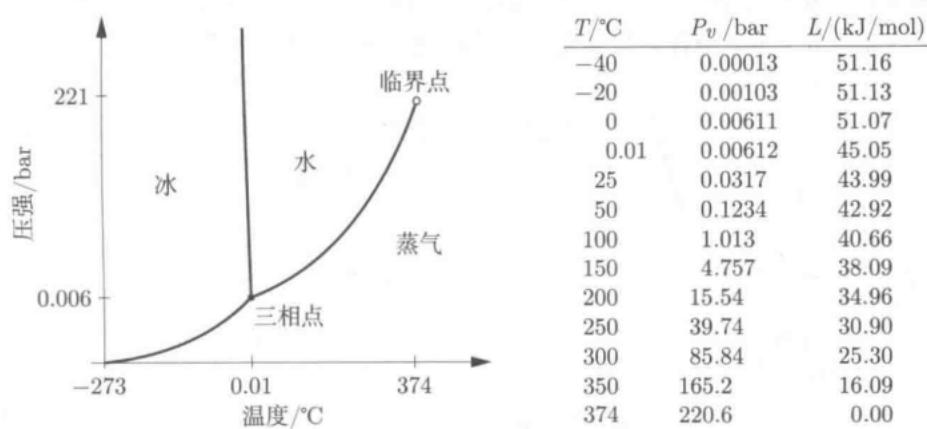


Figure 1: H_2O 的相图。右边的表格给出了蒸气压和摩尔潜热，前三行描述固气相变，余下几行描述液气相变。

图(1)显示了 H_2O 的定性相图，以及相变发生时的一些定量数据。从相图中可以清楚地看到三个区，对应气、水、冰三相，每个区的内部都是单独的一种相，两个区的交界线是两相共存的地方，例如冰和水可以在0摄氏度和1个标准大气压下稳定地共存，图中的三相点则是气、水、冰三相共存的地方，在三相点处 $T = 0.01$ 摄氏度， $P = 0.006\text{bar}$ 。在比三相点更低的压强下，液态水不能存在，此时冰会直接升华成蒸气。在高压下，冰事实上会展现出好多种不同的相，它们拥有不同的晶体结构和物理性质，不过，冰的这些不同相这幅图中并没有标示。值得注意的是，图中标示的这三相都是指稳定的相，实际上“亚稳”相也是可以持续存在的，例如，水可以“过冷”，即在冰点下依然保持液态并维持一段时间。

为了确定相图，19世纪人们对气液相变进行了系统的研究。其中一种办法是将一定量的液体(如水、乙醚)封在容器中，缓慢加热，测定压力随温度的变化曲线。只要容器中同时存在液体和它的蒸气，气液两相就在共存，因此这样测得的 $P-T$ 曲线，就是气液的相界线，它也叫气化线。

如图(2)所示，如果保持压强为 P_0 ，由液相开始加热，状态将从图中的 L 点沿水平线向右移动，与气化线交于 Q 点，一部分液体开始气化，这时尽管继续加热，温度却不会升高，一直保持在 Q 点的 T_0 值，加热的热量全部被吸收来用于气化了。这种在相变时才表现出来的气化热，是一种相变潜热。日常生活中，手上沾了水会有清凉感，就是因为水气化时从手里吸收了潜热。

液体全部气化之后，温度才继续沿着图中的 QG 线继续升高。值得注意的是，沿着 LQG 线做实验，为了保持压强不变，要不断改变系统的体积，因此这样的实验不能用前面提到的密封容器来做。

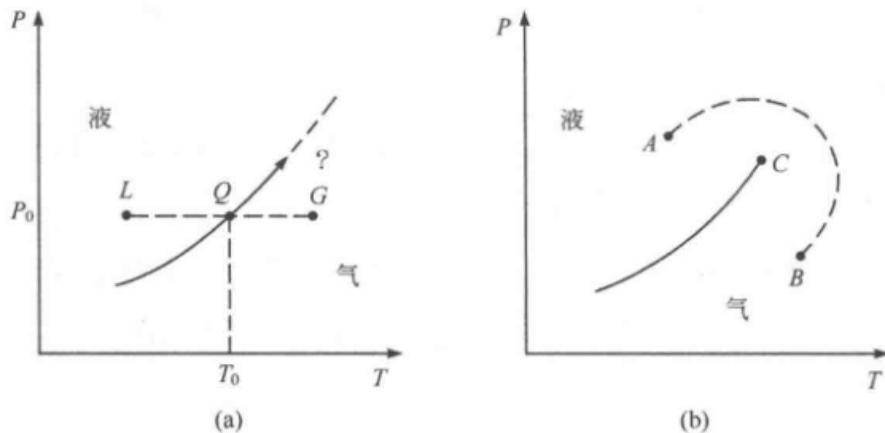


Figure 2: $P-T$ 相图上的气液相界线

一个自然的问题是，气化线是会随着温度和压强的不断增加而无限延伸，还是会有一个截止的端点？1869年英国物理学家安德鲁斯作了一个题为“论物质液态和气态的连续性”的报告，明确地回答了这个问题。他精确地测量了二氧化碳在液态与气态时的密度差，发现在31摄氏度附近，两者的差别消失了！这意味着气液相界线有一个明确的终点，安德鲁斯称之为“临界点”。其他物理学家也观察到这个现象，发现在这个点，相变潜

热等于零！

这其实不难理解，关键是要注意到液气相界线总是有正的斜率，因此对于共存的液相和气相，如果要提高温度，就必须同时施加更高的压强，然而，随着压强的增加，气体将变得更加稠密，因此液体和气体之间的密度差会逐渐缩小，直至消失，也就是达到临界点。对于 H_2O ，这个临界点在 $T_c = 374$ 摄氏度、 $P_c = 221$ bar。在临界点上，就不再有任何不连续的从液相到气相的变化了。

由于临界点的存在，液相其实可以连续地变成气相，而不需要经过任何不连续的相变。方法是沿着图(2)b中所示的虚线变化温度和压强。

那么，固液相界线有没有临界点呢？回答是没有，固液相界线会无限延伸。这是因为，固体和液体存在定性的差别，固态有晶格结构，而液态则是随机排列的分子，这两者有不同的对称性！而这个对称性的差异是无法消失的，它不是一个程度问题，而是类似于一种有和无之间的问题。

除了温度和压强，改变其他变量也会引起相变。图(3)显示了两个不同磁性系统的相图，它们都是改变外磁场和温度引起的相变。左图是第一类超导体的相图，超导相的电阻为零，仅在外磁场和温度都足够低时才出现。右图是诸如铁之类的铁磁体的相图。分为朝上和朝下两个磁化相，具体处于哪个相取决于外磁场的方向(是向上还是向下)。当施加的外磁场为零时，两个方向的磁化相可以共存。然而，随着温度的升高，两个相的磁化强度都会减小，最后，在居里温度时，磁化完全消失，因此两相的相界线终止于一个临界点。

1.2 克劳修斯-克拉珀龙关系

为了明晰起见，暂时限制于讨论 $P - T$ 相图，显然，这相当于等温等压的系统。问题是，给定一个压强和温度，如何判断系统会处于哪个相呢？

为了回答这个问题，考虑某个和温度为 T 的热库接触，体积可变，但是压强与热库一样为 P 的系统。假设经过某个过程，系统的体积发生了改变，同时从热库中吸收了热量 Q ，由于热库很大，其放热过程可以看作是可逆的，因此热库的熵变为 $dS_R = -Q/T$ 。假设系统的熵变为 dS ，则系统和热库一起的总熵变就是 $dS + dS_R$ ，由于这两者一起构成了一个大的孤立

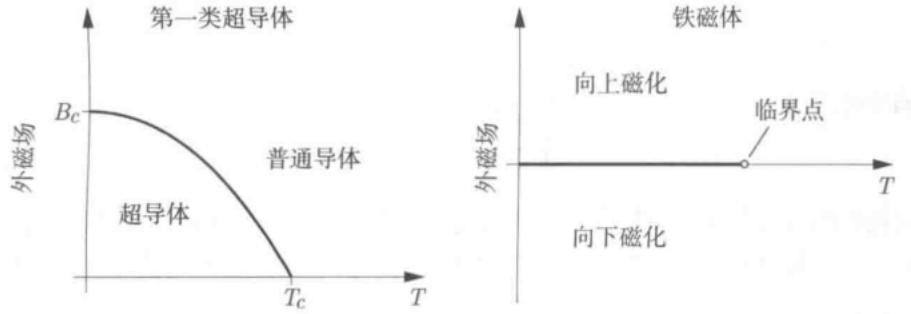


Figure 3: 左图为第一类超导体的相图。右图为铁磁体的相图，假设外磁场的方向和磁化的方向都是竖直方向。

系统，所以根据熵增加原理，必有

$$dS + dS_R \geq 0. \quad (1)$$

对系统应用热力学第一定律，即有 $Q = dE + PdV$, dE 是上述过程中系统的能量增加量, dV 是其体积增加量, 从而即有

$$dS + dS_R = dS - \frac{1}{T}(dE + PdV) = -\frac{1}{T}(dE - TdS + PdV) = -\frac{1}{T}dG \geq 0. \quad (2)$$

式中 $G = E - TS + PV$ 为系统的吉布斯自由能。因此熵增加原理告诉我们，等温等压系统的任何自发过程都必然会导致吉布斯自由能减少，即

$$dG \leq 0. \quad (3)$$

特别的，当系统达至热平衡时，吉布斯自由能取极小值。

将上述原理用到相变问题上，即有：在给定的温度和压强下，稳定的相永远是拥有更低吉布斯自由能的相！其自然推论是，在两相共存的相界线上，两个相的吉布斯自由能必定相等！

简单起见，这里只讨论液体和气体的相界线，类似的讨论当然也适用于任何相界线。我们考虑一定量的物质，比如1mol。我们已经知道，在相界线上，液体和气体的吉布斯自由能(分别记为 G_l 和 G_g)必定相等，即

$$G_l = G_g. \quad (\text{在相界线上}) \quad (4)$$

现在，设想沿着相界线，把温度升高 dT ，压强增加 dP ，由于是沿着相界线，吉布斯自由能将保持相等，从而即有

$$dG_l = dG_g. \quad (\text{沿着相界线}) \quad (5)$$

利用热力学关系式 $dG = -SdT + VdP + \mu dN$ ，即有（假设系统粒子数保持不变）

$$-S_ldT + V_ldP = -S_gdT + V_gdP. \quad (6)$$

进而即可以得到相界线的斜率为

$$\frac{dP}{dT} = \frac{S_g - S_l}{V_g - V_l}. \quad (7)$$

从这个式子可以看出，如果两相熵的差异巨大，则温度的微小变化就需要巨大的压强变化来补偿，从而相界线也就越陡峭。反过来，如果两相体积的差异巨大，那么压强的微小变化就需要巨大的温度变化来补偿，从而相界线也就越平。

通常会把两相熵的差 $S_g - S_l$ 写成 L/T ，其中 L 是把这些物质从液态转化为气态所需的潜热。这样一来，上式就可以重写为

$$\frac{dP}{dT} = \frac{L}{T\Delta V}, \quad \text{其中 } \Delta V = V_g - V_l. \quad (8)$$

这个结果就称作克劳修斯-克拉珀龙关系。它不仅适用于液体到气体的相变，也可以推广到 $P-T$ 图上的任何相界线。液体相变成气体时要吸收潜热，因此 $L > 0$ ，另外，气态的体积大于液态，从而 $\Delta V > 0$ ，因此对于液气相界线，其斜率 $\frac{dP}{dT}$ 总是正的。但是冰和水之间的相界线就不是这样，冰转化为水需要吸收潜热， $L > 0$ ，但是，水的体积比相应冰的体积要小，从而冰转化为水 $\Delta V < 0$ ，从而冰水相界线的斜率为负。

沿着相界线吉布斯自由能保持相等的结论也可以说成是：沿着相界线，两相的化学势相等。因为根据Gibbs-Duhem关系式有 $G = \mu N$ ，而不管物质处于哪一相，其粒子数肯定是一样的，所以吉布斯自由能相等就等价于化学势相等，即

$$\mu_l = \mu_g. \quad (9)$$

通常称这个条件为相平衡条件，它也可以用另一种方法来得到，即考虑共存的两相，把它们看成一个孤立系统的两个子系统，两相热平衡的必然推论就是：两相的温度相等；两相的压强相等；同时，两相的化学势也相等。

由于相变前后，吉布斯自由能保持相等，也就是吉布斯自由能在相变前后是连续的。但是，由于相变前后熵和体积都发生了不连续的突变，而熵和体积都是吉布斯自由能的一阶偏导(分别对 T 和 P)，所以这就说明，相变前后吉布斯自由能的一阶偏导是不连续的！像这样的相变就称之为一级相变，也就是吉布斯自由能连续但是其一阶偏导不连续的相变。

另外，对于临界点来说，沿着相界线从趋于临界点到越过临界点，系统当然也发生了相变。但是，前面说过，如果沿着相界线趋于临界点，那么相变潜热将趋于零，换言之，对于临界点处的这种相变，不仅吉布斯自由能是连续的，熵也是连续的，类似的体积也是连续的。像这样的，不仅吉布斯自由能连续，而且其一阶偏导也连续的相变就称作连续相变。因此，临界点处的相变就是连续相变。值得注意的是，连续相变也会产生物理量的不连续或者无穷尖峰的奇异行为，但是这些物理量都是吉布斯自由能的二阶以上的偏导，是诸如压缩率、磁化率这样的物理量。

1.3 吉布斯相律

前面考虑的是只有一种组元(如 H_2O)的物质，现在推广到有 k 种组元的物质。考虑一个 k 组元 σ 个不同相的系统，注意，相是一种宏观表现，同一相中可能有多种不同的物质组元。假设处于 α 相的物质中第 j 组元物质所占的百分比为 x_j^α ，很显然

$$\sum_{j=1}^k x_j^\alpha = 1. \quad (10)$$

由于不同相的温度相等，压强也相等，所以整个系统由以下状态参量刻画

$$T, \quad P, \quad x_1^\alpha, x_2^\alpha, \dots, x_k^\alpha, \quad (11)$$

总共 $k\sigma + 2$ 个参数，但是，其中 x_j^α 要满足约束条件(10)，很显然，这样的约束条件共 σ 个，所以，独立的状态参量个数为

$$k\sigma + 2 - \sigma = (k - 1)\sigma + 2. \quad (12)$$

但是，每一种物质组元还要满足相平衡条件(即化学势相等的条件，它对任意两不同相都要满足)，给定物质组元，这样的相平衡条件有 $\sigma - 1$ 个， k 种组元就有 $k(\sigma - 1)$ 个条件。所以，系统最后真正自由的状态参量个数 f 为

$$f = (k - 1)\sigma + 2 - k(\sigma - 1) = k - \sigma + 2. \quad (13)$$

这就是所谓的吉布斯相律。

对于单元系来说， $k = 1$ ，从而 $f = 3 - \sigma$ 。如果整个系统处于同一个相， $\sigma = 1$ ，则 $f = 2$ ，即系统仅有两个自由参量，比如理想气体仅有 T, P 两个自由参量。如果系统处于两相共存，则 $f = 1$ ，系统将仅由一个自由参量刻画，正因为如此，在单元系的相图中，两相共存的状态只能落在一条一维的曲线上，也就是落在相界线上。如果三相共存，那么 $f = 0$ ，这就是前面提过的三相共存点。单元系不可能出现4相共存的情况。

如果所考察的系统是二元系，则 $k = 2$ ，那么 $f = 4 - \sigma$ 。若这样的系统仅出现一个相，就需要3个自由参量来描述其状态，例如可以选 T, P 再加上其中一种组分的粒子数百分比 x 。若系统处于两相共存，则只能有两个自由参量。如果三相共存，那么依然有一个自由参量。由于最一般的单相状态需要3个自由参量刻画，所以二元系的相图严格来说要画成三维立体的。不过，实际应用中，人们常常会把压强先固定，然后只考虑剩余的两个参量 T 和 x 的相图，这样的相图称为定压相图。

2 范德瓦尔斯模型

1873年，荷兰青年范德瓦尔斯利用当时刚刚发展起来的分子运动论观念，对安德鲁斯关于临界点的实验结果提出了理论解释。他的论文题目与安德鲁斯的报告类似，也叫“论气态和液态的连续性”。

为了描写气液相变，范德瓦尔斯建议计入气体分子间的两种相互作用：第一，每个分子具有一定的体积 b (在小于 b 的范围之内，气体分子间会非常强烈地相互排斥)，因此气体活动的有效体积缩小为 $V - Nb$ 。第二，气体分子间在较远的距离上有微弱的相互吸引力，相当于补充了一点“内压力”，使 P 增加到 $P + a\left(\frac{N}{V}\right)^2$ 。这里“内压力”之所以能写成 $a\left(\frac{N}{V}\right)^2$ ，是因

为，首先它来自于分子之间的微弱吸引，既然是分子之间的，那就应该正比于 $N \times N = N^2$ 。其次，作为压强，“内压力”应该是强度量，所以就只能正比于 $(\frac{N}{V})^2$ 了。根据这些考虑，于是范德瓦尔斯把理想气体的物态方程 $PV = Nk_B T$ 修改为

$$(P + a(\frac{N}{V})^2)(V - Nb) = Nk_B T. \quad (14)$$

这就是本书第三章曾经推导过的著名的范德瓦尔斯方程，当然范德瓦尔斯可不是像我们在第三章中做的那样推导出来的，范德瓦尔斯的时代也没有那样成熟的统计物理理论可用，相反，他是猜出来的，本段介绍的就是范德瓦尔斯非常简单但是非常漂亮的思路。

我们可以在 $P - V$ 图上画出范德瓦尔斯方程的等温线，如图(4)所示。在足够高的温度下，减小体积会导致压强平稳地上升到无穷，同时体积逼近 Nb 。然而，在较低的温度下，等温线的行为要复杂得多，随着 V 的减小，等温线上升，下降，然后再次上升。似乎暗示着，对于某些状态来说，压缩液体会导致压强下降，即等温压缩系数 $\kappa_T \equiv -\frac{1}{V} \frac{\partial V}{\partial P}|_T < 0$ ，但是，前面第五章我们已经证明过了， κ_T 一定是大于零的。这就说明，为了正确地描写液气相变，范德瓦尔斯方程还需要一些修正。

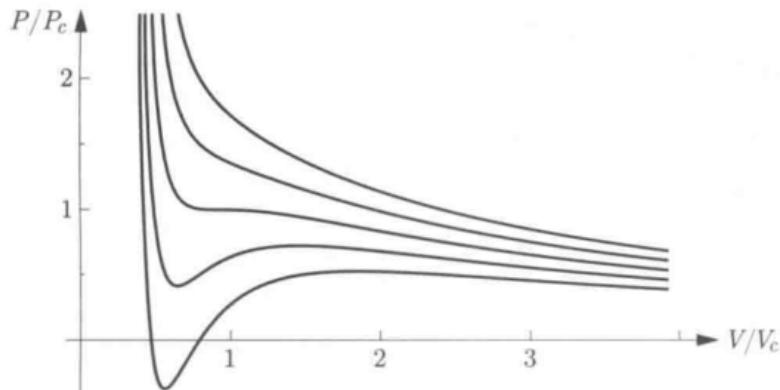


Figure 4: 范德瓦尔斯方程的等温线。

为了搞清楚该如何修正范德瓦尔斯方程，我们从热力学关系式 $dG = -SdT + VdP + \mu dN$ 出发。给定物质的数量不变，且固定温度，则有 $dG =$

VdP , 两边同时除以 dV , 可得

$$\frac{\partial G}{\partial V}|_{N,T} = V \frac{\partial P}{\partial V}|_{N,T}, \quad (15)$$

等式的右边可以直接从范德瓦尔斯方程计算出, 结果为

$$\frac{\partial G}{\partial V}|_{N,T} = -\frac{Nk_BTV}{(V-Nb)^2} + 2a\frac{N^2}{V^2}, \quad (16)$$

将等式右边第一项中的 V 写成 $(V - Nb) + (Nb)$, 然后对每一部分分别积分, 有

$$G = -Nk_B T \log(V - Nb) + \frac{(Nk_B T)(Nb)}{V - Nb} - 2a\frac{N^2}{V} + c(T), \quad (17)$$

式中 $c(T)$ 是积分常数, 对于我们当前的目的来说, 它其实无关紧要。这个公式使得我们可以画出给定温度下的吉布斯自由能曲线。

相较于将 G 绘制为 V 的函数, 我们将 V 当作参数, 把参数方程 $G(V)$ 绘制到纵轴、 $P(V)$ 绘制到横轴, 进而得到 G 作为 P 的曲线反而更有用。图(5)给出了一个示例, 其中等温线与 $G - P$ 图并排显示。虽然范德瓦尔斯方程在某些压强下存在多个对应的体积, 但是 $G - P$ 图中的这一部分三角形环对应的态是不稳定的(点2-3-4-5-6), 因为热力学稳定态一定具有最低的吉布斯自由能。当压强逐渐增加时, 系统将直接从点2跳到点6, 同时伴随体积突然减小, 也就是相变。在点2处, 我们应该将流体称为气体, 因为其体积随压强增加迅速减小。在点6处, 我们应该将流体称为液体, 因为它的体积在压强大幅增加的情况下, 仅略微下降。在这些点之间的体积, 热力学稳定态实际上是部分气体和部分液体的组合, 仍保持在转变压强, 如 $P - V$ 图上的水平直线所示。等温线被这条直线所截的弯曲部分并不描述真正的稳定态。

从 $G - P$ 图上, 我们可以读出2到6的相变对应的压强。但是有一种巧妙的方法可以直接从 $P - V$ 图中读取到它, 从而不需要画出吉布斯自由能的曲线。为此我们注意到, 当围绕三角形循环(2-3-4-5-6)一圈时, G 的变化为零:

$$0 = \int_{\text{一圈}} dG = \int_{\text{一圈}} \frac{\partial G}{\partial P}|_T dP = \int_{\text{一圈}} VdP, \quad (18)$$

最后一个等于号告诉我们, 这个积分可以从 $P - V$ 图中计算。我们可以把 $P - V$ 图旋转成 $V - P$ 图。从点2到点3的积分是 $V - P$ 曲线下这一部分的

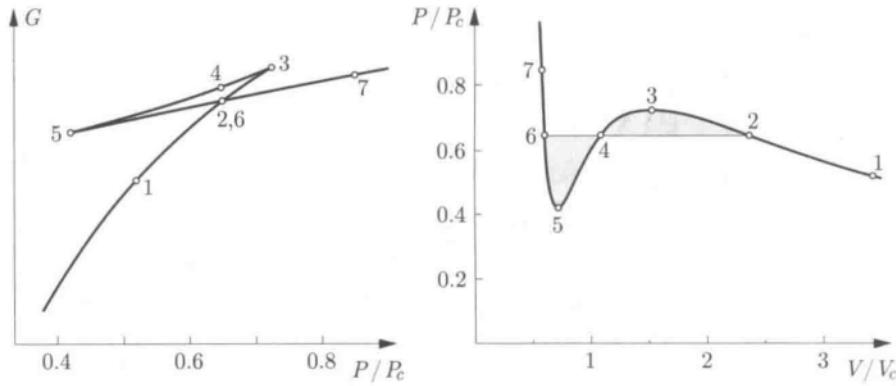


Figure 5: 在 $T = 0.9T_c$ 时，范德瓦尔斯流体的吉布斯自由能与压强的关系。此时对应的等温线如右图所示，在2-3-4-5-6之间的状态是不稳定的

面积，但是点3到点4的积分抵消了一部分，最后只剩下相应的阴影区域。同样，从点4到点6的积分也只剩下相应的阴影部分面积，而且是负的这个面积。所以最后一圈的积分就是这两块阴影部分面积相减。而结果等于零，就说明我们要合适地调节水平直线的位置，使得这两块阴影部分的面积相等！这就确定了水平直线的压强。这个方法叫做麦克斯韦构造法。

图(6)给出了在不同温度下重复麦克斯韦构造法得到的结果。对于每个温度，都有一个确定的压强(对应阴影部分的水平直线)，称为蒸气压，在该温度和压强下会发生液气转变。绘制这个压强和温度的关系就可以得到 $P - T$ 图上的液气相界线。

但是，为什么高温下的等温线随着 V 的减小而单调上升呢？对于这些温度，没有从低密度态到高密度态的突然转变，也即没有相变。因此，在一定温度以上，液气相边界会消失，这个温度就称作临界温度 T_c ，相应的蒸气压就称为临界压强 P_c ，其对应的体积称为临界体积 V_c 。这些临界值所定义的，正是临界点。

根据麦克斯韦构造法不难看出，临界点对应的情况是， $P - V$ 图上代表两相共存的水平直线刚好与等温线相切的情况。也即是等温线的拐点，即等温线的一、二阶导数都等于零的地方

$$\frac{\partial P}{\partial V}|_T = 0, \quad \frac{\partial^2 P}{\partial V^2}|_T = 0. \quad (19)$$

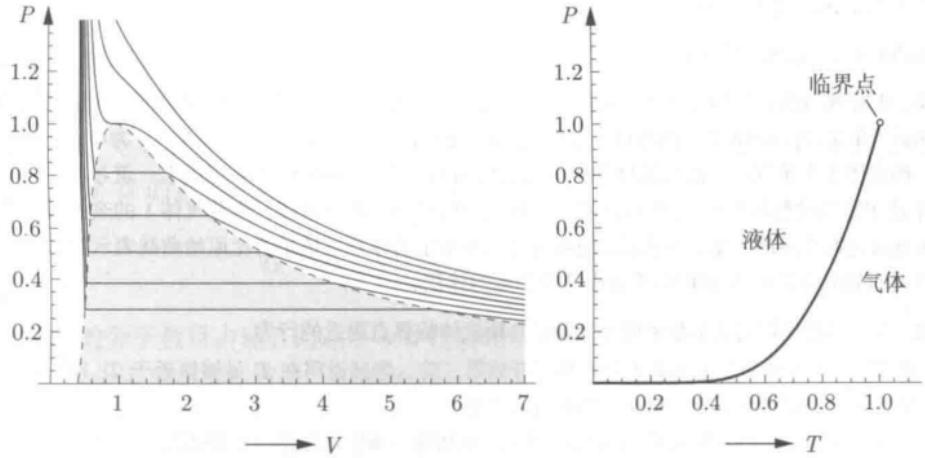


Figure 6: 范德瓦尔斯模型预测的相图。阴影部分的稳定态是两相共存。

代入范德瓦尔斯方程求出这些偏导，最后就能得到

$$T_c = \frac{8a}{27bk_B}, \quad P_c = \frac{a}{27b^2}, \quad V_c = 3Nb. \quad (20)$$

有趣的是，如果用 T_c, P_c, V_c 分别作为测量温度、压强和体积的单位，进而引入三个没有量纲的量(称为约化量)

$$t = \frac{T}{T_c}, \quad p = \frac{P}{P_c}, \quad v = \frac{V}{V_c}. \quad (21)$$

代回范德瓦尔斯方程，就可以得到

$$(p + \frac{3}{v^2})(3v - 1) = 8t. \quad (22)$$

通常称之为约化的范德瓦尔斯方程，注意，它不含有任何与具体物质有关的参数(如 a 和 b)，一切气体似乎都遵循这同一个方程(当然还要加上麦克斯韦构造法给出的修正)。这反映了临界点附近物理规律的普适性！

临界指数

现在我们来进一步研究物质在临界点附近的行为。我们主要关心当我们趋近临界点时各物理量是如何变化的？当然，可以研究的物理量很

多，这里只举比较有代表性的几个为例。注意，用约化量来看，临界点对应 $t_c = p_c = v_c = 1$ 。

我们首先关心沿着 $p - t$ 图上的液气共存线趋于临界点时，两相密度差的逆如何变化，也就是 $v_{\text{gas}} - v_{\text{liquid}}$ 如何变化？共存线当然满足 $t < 1$ ，此时给定共存线上的一个约化压强 p (也就是 $p - v$ 图上的一条水平直线)，它和约化范德瓦尔斯方程的等温线有两个交点，分别对应液态和气态，相应的约化体积记为 v_{liquid} 和 v_{gas} ，根据约化的范德瓦尔斯方程，有

$$p = \frac{8t}{3v_{\text{liquid}} - 1} - \frac{3}{v_{\text{liquid}}^2} = \frac{8t}{3v_{\text{gas}} - 1} - \frac{3}{v_{\text{gas}}^2}, \quad (23)$$

从中可以解出相应的温度 t 为

$$t = \frac{(3v_{\text{liquid}} - 1)(3v_{\text{gas}} - 1)(v_{\text{liquid}} + v_{\text{gas}})}{8v_{\text{liquid}}^2 v_{\text{gas}}^2}. \quad (24)$$

注意，趋于临界点时 $v_{\text{liquid}} \rightarrow 1$, $v_{\text{gas}} \rightarrow 1$, 而且上面这个 t 的表达式关于 v_{liquid} 和 v_{gas} 对称。所以我们令 $v_{\text{gas}} = 1 + \epsilon/2$, $v_{\text{liquid}} = 1 - \epsilon/2$ (从而 $v_{\text{gas}} - v_{\text{liquid}} = \epsilon$)，代入上面 t 的表达式，并保留到 ϵ 的领头阶，不难得得到

$$t \approx 1 - \frac{1}{16}(v_{\text{gas}} - v_{\text{liquid}})^2. \quad (25)$$

从而立即可以得到，当我们沿着液气相界线趋于临界点时

$$v_{\text{gas}} - v_{\text{liquid}} \sim (t_c - t)^{1/2}, \quad (26)$$

注意 $t_c = 1$ 。

我们的第二个问题是：当我们沿着临界温度所在的等温线趋于临界点时，压强如何随着体积变化？这个问题其实很好回答，因为前面我们说过，临界点处 $\frac{\partial p}{\partial v}|_{t_c} = \frac{\partial^2 p}{\partial v^2}|_{t_c} = 0$ ，而给定温度 t_c ，函数 $p(v, t_c)$ 是由约化的范德瓦尔斯方程唯一确定的，所以将这个函数泰勒展开到3阶项(泰勒展开的前两阶都是零)，即有

$$p - p_c \sim (v - v_c)^3. \quad (27)$$

最后，我们来关心约化的等温压缩率，定义是

$$\kappa_t = -\frac{1}{v} \frac{\partial v}{\partial p}|_t. \quad (28)$$

我们知道对于 $t < t_c$ 的两相共存区，范德瓦尔斯方程给出的等温压缩率常常是负的，从而是错的(所以才要进行麦克斯韦构造法的修正)，所以这次我们关心从 $t > t_c$ 的一侧趋向 t_c 时， κ_t 如何变化？回答也不难，注意在临界点附近，压强并没有不连续的突变，而且 $\frac{\partial p}{\partial v}|_{t_c} = 0$ ，所以在 t_c 附近将 $\frac{\partial p}{\partial v}|_{t,v=v_c}$ 对变量 t 泰勒展开到一阶，即有

$$\frac{\partial p}{\partial v}|_{t,v=v_c} \sim (t - t_c) + \dots \quad (29)$$

由此即知

$$\kappa_t \sim (t - t_c)^{-1}. \quad (30)$$

也即是说，趋于临界点时，等温压缩率是发散的！

以上我们得到了三个有趣的结果，(26)式、(27)式、以及(30)式。那么，它们与实验吻合吗？这个问题关系着我们关于相变的范德瓦尔斯方程是否能描写液气相变临界点附近的行为。结果表明，定性上是吻合的，定量上却不吻合。

的确，实验表明以上涉及到的三个量在趋于临界点时的行为和我们上面得出的结果一样有普适性，也就是不依赖于物质的具体组成。但是定量上，我们上面用范德瓦尔斯方程计算出来的标度行为却是错的。实验得出的结果是：当我们沿着液气共存线趋于临界点时

$$v_{\text{gas}} - v_{\text{liquid}} \sim (t_c - t)^\beta \quad \text{其中 } \beta \approx 0.32, \text{ 而不是 } 1/2. \quad (31)$$

而当我们沿着等温线趋于临界点时

$$p - p_c \sim (v - v_c)^\delta \quad \text{其中 } \delta \approx 4.8, \text{ 而不是 } 3. \quad (32)$$

最后，当我们从高温趋于临界点时

$$\kappa_t \sim (t - t_c)^{-\gamma} \quad \text{其中 } \gamma \approx 1.2, \text{ 而不是 } 1. \quad (33)$$

以上式子中的三个指数 β, δ, γ 是所谓的临界指数的例子。很显然实验测出来的这三个临界指数与我们用范德瓦尔斯方程算出来的都不同！

3 朗道范式

3.1 朗道的平均场理论

关于临界点处的连续相变，朗道发展了一个重要而又优雅的理论。为了讲清楚朗道的理论，我们还是举铁磁相变为例。铁磁体的相图如图(7)所示。这里为了简单起见，我们同样假设铁磁体只沿着竖直轴磁化，外磁场也只沿着竖直轴。

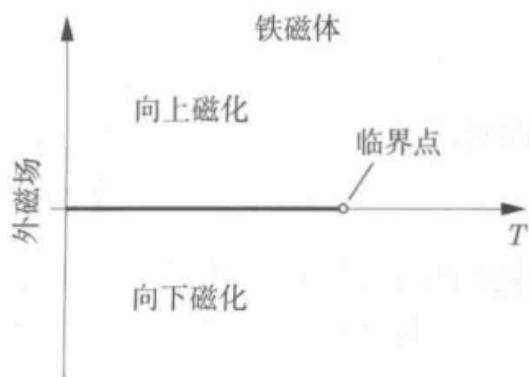


Figure 7: 铁磁体的相图

微观上看，铁磁体是由宏观数量相互作用的自旋规则地排布而成的，这些自旋排布成规则的晶格，每个格点上有一个自旋，相邻格点上的自旋会有相互作用，这种作用趋向于使得相邻自旋的取向相同，具体细节与我们关系不大。按照前面的简单假设，这些自旋只能取向上和向下两种朝向。如果没有外磁场，那么向上和向下对于系统来说是等价的，我们说系统的微观相互作用有 Z_2 对称性， Z_2 就是 $\{+1, -1\}$ 两种可能性(按照通常的乘法进行运算)，可以分别指代自旋向上和自旋向下。另外，这些自旋磁矩在宏观局部上的平均值，就是磁化强度，记作 $m(x)$ 。

假设外磁场为零，只调节系统的温度，也就是只关心相图(7)中水平轴上的情况。在高温时，热运动的能量远远超过相邻自旋之间相互作用的能量，因此，这些自旋各自为政，每个自旋的取向都在向上和向下之间疯狂而随机地改变，自旋向上和自旋向下完全对称，整个系统是向上和向下对

称而且无序的。这时候平均来说，每个自旋的自旋磁矩都是零，因此宏观上的磁化强度 $m(x)$ 也是零。我们说这时候系统处于具有 \mathbf{Z}_2 对称性的相，也就是系统的宏观表现有 \mathbf{Z}_2 对称性，它以之前微观相互作用的 \mathbf{Z}_2 对称性为基础，但不是一回事。高温相的对称性是微观相互作用所能允许在宏观上表现出来的最大对称性。

当温度降到临界点以下时，自旋之间的相互作用效应开始超过无规则热运动的效应，这时候会发生一件神奇的事情：这些自旋开始自发且协调一致地行动，形成一种要么都向上(总体上)、要么都向下(总体上)的宏观模式。换言之，系统从无序的热运动中自发产生了有序。这种“有序的程度”，正好可以用磁化强度 m 来刻画。由于各自旋协调一致地要么都向上、要么都向下，这时候磁化强度当然就不是零，而是一个非零常数，系统出现了自发磁化，这就是临界点处铁磁相变的物理本质。由于磁化强度为零对应无序的高温相，而磁化强度非零对应有序的铁磁相，所以朗道就把这里的磁化强度称之为序参量。

由于自发磁化选取了向上或者选取了向下，因此在铁磁相，原来高温相的 \mathbf{Z}_2 对称性就没有了，通常称这为 \mathbf{Z}_2 对称性破缺了。但是，由于向上磁化还是向下磁化原则上是机会均等的，系统只是随机地选取了其中一个方向，所以也称这种对称性破缺为对称性自发破缺。

为了给上述物理图像建立物理模型，朗道假设系统自由能是序参量的解析函数(假定序参量不依赖于空间点 x)，因此可以把自由能密度 $f = F/V$ (注意自由能是广延量)泰勒展开成如下形式

$$f(m) = a(t) + \frac{1}{2}b(t)m^2 + \frac{1}{4}c(t)m^4 + \dots \quad (34)$$

式中 $t = (T - T_c)/T_c$ 是约化温度(和之前范德瓦尔斯模型中的约化温度的定义略有差异)。由于微观相互作用有 \mathbf{Z}_2 对称性，自由能作为 m 的函数必定会保持这种对称性，因此我们已经要求 $f(m)$ 在 $m \rightarrow -m$ 的 \mathbf{Z}_2 变换下保持不变了，这也就相当于要求 $f(m)$ 是 m 的偶函数，所以上述展开式中没有 m 的奇次项。上述表达式忽略了高于四次的泰勒展开项，因为在临界点附近，自化磁化的 m 其实是小量，四次以上的项太小而可以忽略，而且，正如我们马上会看到的，保留到四次项已经足以解释连续相变的临界行为了。

在上述自由能泰勒展开的表达式中， m 是待定的变量。定 m 的办法也很

简单，只要注意到热平衡时自由能取极小值，根据极小值条件

$$\frac{\partial f}{\partial m} = 0 \Rightarrow [b(t) + c(t)m^2]m = 0. \quad (35)$$

注意到临界温度以上，即 $t > 0$ 时，没有自发磁化，物理上序参量只能取 $m = 0$ ，且这个 $m = 0$ 要使 $f(m)$ 取极小值，因此朗道假设 $t > 0$ 时， $b(t) > 0, c(t) > 0$ ，否则要么 $m = 0$ 不是极小，要么上述方程还存在 $m \neq 0$ 的解。同样， $t < 0$ 时，物理上要求 $m \neq 0$ ，因此上述 $m = 0$ 的解必定不是 $f(m)$ 的极小值，这就要求 $t < 0$ 时， $b(t) < 0$ ，进一步要求上述方程存在 $m \neq 0$ 的解，因此必定还有 $t < 0$ 时， $c(t) > 0$ 。此时这个非零的 m 解为

$$m = \pm m_0 \quad \text{其中 } m_0 = \sqrt{-\frac{b(t)}{c(t)}}, \quad (36)$$

\pm 号表示这时候向上磁化和向下磁化均有可能，虽然实际的物理系统只会随机取其中一种磁化。 $b(t) > 0$ 时，自由能密度 $f(m)$ 的函数图像如图(8)所示，而 $b(t) < 0$ 时，自由能密度 $f(m)$ 的函数图像如图(9)所示。

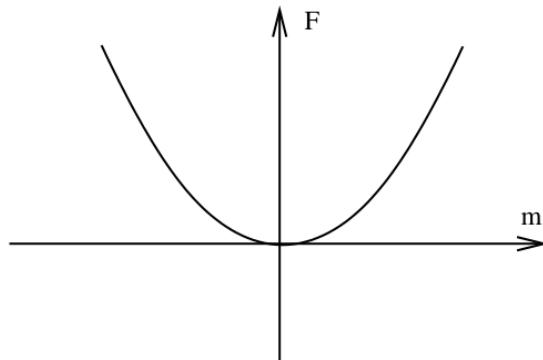


Figure 8: $b(t) > 0$ 时，自由能密度 $f(m)$ 的函数图像

朗道理论清楚地把序参量由零变成非零的过程与对称性自发破缺联系了起来。首先，理论本身具有 Z_2 对称性，因为自由能函数在 $m \rightarrow -m$ 的 Z_2 变换下保持不变。其次， $t > 0$ 时，自由能函数只有 $m = 0$ 这一个极小值，它本身就是 Z_2 不变的。但是， $t < 0$ 时， $m = 0$ 不再是极小值，而是局部极大值，相反，这时候极小值分裂成了两个，分别在 $\pm m_0$ 。虽然这两个极小值

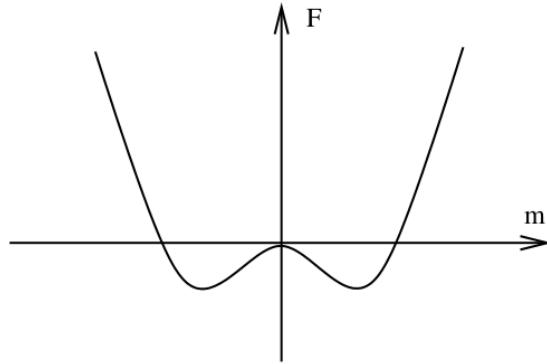


Figure 9: $b(t) < 0$ 时，自由能密度 $f(m)$ 的函数图像

是 \mathbf{Z}_2 对称的，但是，只看其中一个的话，每一个都没有 \mathbf{Z}_2 对称性，而真实的系统只能选取两个极小值中的一个，但只要选取了一个，那么 \mathbf{Z}_2 对称性就自发破缺了。

综合以上分析(特别要注意 $b(t)$ 在 $t > 0$ 和 $t < 0$ 时要改变正负号而 $c(t)$ 则不变号)，朗道假设在临界点附近，参量 $a(t), b(t), c(t)$ 可以泰勒展开成

$$\begin{aligned} a(t) &= a_0 + a_1 t + \dots \\ b(t) &= b_0 t + \dots \quad \text{其中 } b_0 > 0 \\ c(t) &= c_0 + c_1 t + \dots \quad \text{其中 } c_0 > 0 \end{aligned} \tag{37}$$

由此立即可以得到沿着相界线趋于临界点时磁化强度 $m(t)$ 的临界行为

$$m(t) = m_0 \sim (-t)^{1/2}. \tag{38}$$

注意这个临界行为与前面范德瓦尔斯模型中计算出来的临界行为(26)完全一样，指数都是 $1/2$ 。

请注意，我们并未在自由能中引入任何不连续性。函数 $f(m, t)$ 中的所有项都是光滑且良好的。当进行泰勒展开时，它仅包含 m 和 t 的整数次幂。然而，正如上面 $m(t)$ 的表达式所示，自由能的极小值却表现出非解析的行为。

为了进一步和范德瓦尔斯模型中计算出来的临界行为比。我们引入外磁

场 B , 由于外磁场和自旋磁矩的耦合, 此时自由能密度将成为

$$f(m) = a(t) - Bm + \frac{1}{2}b(t)m^2 + \frac{1}{4}c(t)m^4 + \dots \quad (39)$$

自由能极小的条件将告诉我们

$$B = [b(t) + c(t)m^2]m. \quad (40)$$

当我们从高温趋于临界点时, 前面已经看到, 磁化强度一直是零, 这一点当外磁场很弱时只有很弱的修正, 因此可以忽略(40)式中的 m^3 次方项, 进而可以计算磁化率 $\chi(t)$ 如下

$$B \approx b(t)m \Rightarrow \chi(t) \equiv \frac{\partial m}{\partial B} \sim 1/b(t) \sim t^{-1}. \quad (41)$$

很显然, 这个磁化率的临界行为也和前面范德瓦尔斯模型中计算的压缩率的临界行为(30)式一样, 在临界点附近两者都发散, 且指数都是 -1 。

为了研究沿着 $t = 0$ 的等温线趋于临界点时的行为, 我们注意到此时 $b(t) \approx 0$, 因此(40)式可以近似成

$$B \approx c_0 m^3. \quad (42)$$

很显然, 这个临界行为也和范德瓦尔斯模型中计算出来的临界行为(27)式相同, 指数都是 3 !

以上比较说明, 在临界点附近, 朗道的这个平均场论其实和范德瓦尔斯模型有完全一样的普适性, 两者的差别只在于由物理情境的不同而导致的物理变量名称不同, 而这些变量之间在临界点附近的关系则是一样的。

但是, 我们也说过, 范德瓦尔斯模型算出来的这些临界指数与实验并不吻合, 同样, 朗道理论给出来的这些临界指数也与实验不吻合。问题出在, 在朗道理论中, 我们是把序参量当成一个整体的均匀的平均场, 而忽略了在不同的空间局部上, 序参量可以有涨落。换言之, 实际上, 序参量不是常量, 而是随空间位置变化的动力学场 $m(\mathbf{x})$ (\mathbf{x} 表示空间位置矢量), 尤其在临界点附近, 序参量的涨落将变得非常强, 而且长程关联, 平均场近似此时是失效的。不过, 如何系统地考虑序参量的涨落并不是一件简单的事情。这个困难最终是威尔逊通过他的重整化群理论解决的, 不过, 这有点超出本书的范围了。

3.2 朗道-金兹堡理论

朗道平均场理论是物理学上罕见的优雅理论，优雅就优雅在它极为简单，和范德瓦尔斯的模型一样简单，但是，有点过于简单了。比方说，它没有考虑序参量的涨落。朝着这个方向走的第一步是朗道-金兹堡理论，它的想法也很简单，就是考虑到涨落，需要把序参量变成一个随空间位置改变的场 $m(\mathbf{x})$ 。

现在自由能 F 就变成了场 $m(\mathbf{x})$ 的泛函 $F[m(\mathbf{x})]$ 。朗道-金兹堡同样假设这个泛函可以解析地进行泰勒展开，不过，现在由于 $m(\mathbf{x})$ 可以随空间变化，所以我们要在展开中包含 $m(\mathbf{x})$ 的空间梯度项，忽略无关的常数项 $a(t)$ ，这个展开即是

$$F[m(\mathbf{x})] = \int d^3\mathbf{x} \left[\frac{1}{2}b(t)m^2(\mathbf{x}) + \frac{1}{4}c(t)m^4(\mathbf{x}) + \frac{1}{2}d(t)(\nabla m(\mathbf{x}))^2 + \dots \right] \quad (43)$$

同样，我们假设没有外磁场，并要求自由能具有 \mathbf{Z}_2 不变性，当然，此外还要求它具有空间平移和空间旋转不变性。

自由能极小的条件现在变成了泛函 $F[m(\mathbf{x})]$ 对 $m(\mathbf{x})$ 的变分求极小。不难求出

$$\begin{aligned} \delta F &= \int d^3\mathbf{x} \left[b(t)m(\mathbf{x})\delta m(\mathbf{x}) + c(t)m^3(\mathbf{x})\delta m(\mathbf{x}) + d(t)\nabla m(\mathbf{x}) \cdot \nabla \delta m(\mathbf{x}) + \dots \right] \\ &= \int d^3\mathbf{x} \left[b(t)m(\mathbf{x}) + c(t)m^3(\mathbf{x}) - d(t)\nabla^2 m(\mathbf{x}) + \dots \right] \delta m(\mathbf{x}). \end{aligned} \quad (44)$$

式中第二行是进行了分部积分，并丢掉了边界项(假设边界上的变分 $\delta m(\mathbf{x})$ 为零)。由 $\delta F = 0$ ，即可以得

$$b(t)m(\mathbf{x}) + c(t)m^3(\mathbf{x}) - d(t)\nabla^2 m(\mathbf{x}) = 0. \quad (45)$$

这个方程当然不止一个解，其中最简单的解是 $m(\mathbf{x})$ 取常数，这时候就退回了上一小节平均场理论中的方程。因此上一小节所有关于序参量如何从零变成非零，以及如何对称性自发破缺的所有讨论在这里依然成立。

而朗道-金兹堡理论最重要的应用之一，是理解系统在空间不同点上涨落之间的关联。具体来说，假设我们在空间某点——比如说 $\mathbf{x} = 0$ 的原点处——加上一个无限尖锐的外磁场 $B(\mathbf{x}) = \epsilon\delta^3(\mathbf{x})$ (式中 ϵ 为小量， $\delta^3(\mathbf{x})$ 为三维delta函数)来扰动它，以激发起当地的磁化强度，我们想知道这对别处

的磁化强度会产生什么影响？加了这样的一个外磁场扰动以后，原来的自由能泛函就要多加上一项，即 $\int d^3\mathbf{x}[-m(\mathbf{x})B(\mathbf{x})] = \int d^3\mathbf{x}[-m(\mathbf{x})\epsilon\delta^3(\mathbf{x})]$ 。进而变分求自由能极小时对应的方程就是

$$b(t)m(\mathbf{x}) + c(t)m^3(\mathbf{x}) - d(t)\nabla^2m(\mathbf{x}) = \epsilon\delta^3(\mathbf{x}). \quad (46)$$

假设温度低于临界值，即 $t < 0$ ，且扰动前系统原本处在 $m(\mathbf{x}) = m_0 = \sqrt{-\frac{b(t)}{c(t)}}$ 的极小值处，外磁场把系统扰动到了

$$m(\mathbf{x}) = m_0 + \delta m(\mathbf{x}), \quad (47)$$

式中扰动 $\delta m(\mathbf{x})$ 是小量。代入(46)式，并保留到扰动 $\delta m(\mathbf{x})$ 的线性阶，即有

$$\begin{aligned} & (b(t) + 3c(t)m_0^2)\delta m(\mathbf{x}) - d(t)\nabla^2\delta m(\mathbf{x}) = \epsilon\delta^3(\mathbf{x}) \\ & \Rightarrow -2b(t)\delta m(\mathbf{x}) - d(t)\nabla^2\delta m(\mathbf{x}) = \epsilon\delta^3(\mathbf{x}). \end{aligned} \quad (48)$$

这个方程的解是

$$\delta m(\mathbf{x}) = \frac{\epsilon}{4\pi d(t)} \frac{e^{-|\mathbf{x}|/\xi(t)}}{|\mathbf{x}|} \quad (49)$$

式中 $\xi(t)$ 为

$$\xi(t) = \sqrt{\frac{d(t)}{-2b(t)}}. \quad (50)$$

$\xi(t)$ 的含义是，原点处外磁场扰动的效应在离原点远大于 $\xi(t)$ 的距离上会迅速衰减为零，因此通常称 $\xi(t)$ 为关联长度。

朗道假设，在临界点附近 $d(t) \approx d_0 > 0$ ，从而不难看到，沿着相界线趋于临界点时，关联长度的临界行为是

$$\xi(t) \sim (-t)^{-1/2}, \quad (51)$$

特别的，这意味着，在临界点附近，关联长度会发散。换言之，在临界点附近，序参量的涨落会在长程上保持关联。正是因为这种长程关联，使得气液相变临界点时对光的散射会增强，从而使得原来透明的气体或者液体，在接近临界点时变得浑浊起来，呈现一片乳白色，这就是气液临界点附近的临界乳光现象。

不过，和朗道平均场理论一样，朗道-金兹堡理论给出的这个关联长度的临界指数也与实验值在定量上不吻合。更正确地计算临界指数需要用到前面提到过的威尔逊重整化群理论。

那么，和统计物理有什么关系？

允许序参量依赖于空间的主要动机是为了考虑涨落效应。要理解我们如何做到这一点，首先需要更深入地思考泛函 $F[m(\mathbf{x})]$ 的含义以及我们可以用它来做什么。为了理解这一点，最好还是回到最基本的概念。我们知道，系统的真实自由能可以等同于配分函数的对数。我们通常将系统的真实自由能称为 F 。然而，我们现在已经把朗道-金兹堡泛函称为 $F[m(\mathbf{x})]$ ，虽然它与真实自由能有密切关系，但并不完全相同，这一点我们很快就会看到。而系统的真实自由能永远是 $-k_B T \log \mathcal{Z}$ ，其中 \mathcal{Z} 为系统的配分函数，我们将其表示为：

$$\mathcal{Z} = \sum_n e^{-\beta E_n}. \quad (52)$$

我们希望在这个框架下正确地理解泛函 $F[m(\mathbf{x})]$ 的意义。这里我们按照David Tong讲义的讲解方式，给出一个启发式的、较为粗略的论证。更完整的处理方式会涉及到重整化群的思想。

其基本思想是：系统的每一个微观态 $|n\rangle$ 都可以与某个特定的、空间变化的序参量函数 $m(\mathbf{x})$ 相联系。为了说明这一点，我们将使用伊辛模型的语言进行讨论，尽管该讨论可以推广到任何系统。在伊辛模型中，我们可以考虑通过对某点附近一个物理小(所谓物理小就是宏观上看足够小，但是微观上看依然包含了许多的格点在内)的区域内的所有自旋取平均，来为这个物理小的区域赋予一个磁化强度 $m(\mathbf{x})$ 。显然，这样定义的磁化强度在宏观上可以近似看作是平滑的连续函数，这些函数在小于晶格间距的距离尺度上基本上是常数。通过这种方式，我们就得到了从微观态空间到磁化强度函数的一个映射： $|n\rangle \rightarrow m(\mathbf{x})$ 。但这个映射并不是一一对应的。例如，由于平均过程是在足够多的格点上进行的，那么仅仅翻转某一个格点的自旋，不太可能对平均值产生显著影响。因此，许多微观态会映射到同一个磁化强度函数上。仅对这些微观态求和，就可以从第一性原理构造

出 $F[m(\mathbf{x})]$:

$$e^{-\beta F[m(\mathbf{x})]} = \sum_{n|m(\mathbf{x})} e^{-\beta E_n}. \quad (53)$$

当然，我们实际上并没有通过这样的步骤得到朗道-金兹堡理论中具体展开出来的那个 $F[m(\mathbf{x})]$: 我们只是写下了临界点附近最一般的形式，并引入了一些未知系数 $b(t), c(t), d(t)$ 。但如果我们愿意接受挑战，上述过程告诉我们如何从第一性原理出发推导出这些函数。更重要的是，它也告诉我们应该如何处理朗道-金兹堡自由能。

因为在公式(53)中，我们只对那些对应于特定 $m(\mathbf{x})$ 的微观态进行了求和。而要计算完整的配分函数，我们需要对所有微观态求和。但我们可以通过对所有可能的 $m(\mathbf{x})$ 值求和来实现这一点。换句话说：

$$\mathcal{Z} = \int \mathcal{D}m(\mathbf{x}) e^{-\beta F[m(\mathbf{x})]}. \quad (54)$$

等式右边是一个棘手的对象：它是一个泛函积分。我们是在对所有可能的函数 $m(\mathbf{x})$ 进行积分，这相当于进行无限多重积分。当然，实际上，由于序参量 $m(\mathbf{x})$ 源自底层的晶格结构，并且在短距离尺度上是足够平滑的，所以这个问题有所缓解。

(54)式在物理上非常优美。它意味着我们应该将朗道-金兹堡自由能视为一个关于连续变量 $m(\mathbf{x})$ 的新有效哈密顿量。它是通过对配分函数的大部分微观信息进行求和而得到的，但最终仍然留下了一个对这些微观信息的平均量（即序参量）的涨落进行求和或积分的过程。

要完整解决这个问题，我们需要计算这个泛函积分(54)。这很难。这里的“难”意味着理论物理学中大多数未解决的问题都可以归结为计算这种类型的积分。然而，它很难这一点不应让我们气馁，因为这一泛函积分中隐藏着大量丰富而美妙的物理，包括“普适性”背后的深刻原因。为了理解这些，读者应该进一步去学习统计场论的相关知识。

3.3 连续对称性自发破缺

还是接着讲铁磁相变，为简单起见本节默认没有外磁场。不过，现在我们注意到真实的自旋并不是只有向上和向下两种取向，而是一个三维空间矢

量。在不加外磁场的时候，真实铁磁体的微观相互作用不是具有 Z_2 对称性，而是具有作用在自旋上的三维空间旋转对称性，通常称之为 $O(3)$ 对称性。在临界温度以上，格点上每一个自旋因为热运动都随机地取向，看起来并没有哪个方向显得比较特殊，系统是无序而对称的，我们称系统处于具有 $O(3)$ 对称性的相。再次强调一下，要区分相的对称性和微观相互作用的对称性，微观相互作用的对称性表现在系统的微观哈密顿量上，而相的对称性则表现在系统的具体宏观行为上。

当然，我们也可以将自旋磁矩在物理小的局部区域上取平均，进而定义磁化强度，现在它同样是三维矢量，记作 $\mathbf{m}(\mathbf{x})$ 。 $\mathbf{m}(\mathbf{x})$ 是空间上的矢量场，可以想象成是每个空间点上放上一个小箭头(表示 \mathbf{m} 的方向和大小)，只不过，每一点上的小箭头方向都可以进行任意的三维旋转。在那个高温的具有 $O(3)$ 对称性的相，由于每一个自旋都随机取向，因此 \mathbf{m} 的取值为零，没有哪个方向特殊。

在临界温度以下，每一个空间点附近的物理小区域内的自旋就会因为相互作用而协调起来，进而都指向某一个特定的方向，产生了宏观上的秩序，这种秩序就表现为磁化强度 $\mathbf{m}(\mathbf{x})$ 不再是零了，系统出现了自发磁化！因此 $\mathbf{m}(\mathbf{x})$ 从取值为零变成非零就标志着系统从高温无序相转变成了低温铁磁相，因此 $\mathbf{m}(\mathbf{x})$ 就是序参量。

人们完全可以仿照朗道的办法，将自由能对序参量进行泰勒展开，进行建立上述铁磁相变的物理模型。很显然，在 $T > T_c$ 的高温相，自由能的极小要求 $\mathbf{m} = 0$ 。而和前面 Z_2 对称模型一样的讨论可知，在 $T < T_c$ 的低温相，自由能的极小要求 $\mathbf{m}(\mathbf{x})$ 为常矢量，且

$$|\mathbf{m}(\mathbf{x})| = m_0, \quad (55)$$

式中 $m_0 > 0$ 是常数。也即是说，在临界温度以下，相互作用会协调空间每一点上的小箭头，使得它们长度都为 m_0 ，且方向都一样。

我们已经看到，如果不考虑序参量的涨落，那么临界温度以下，空间点上这些长度一样的小箭头就会同步地选取一个特定的空间指向 \mathbf{n} (这里 \mathbf{n} 是一个不依赖于空间点 \mathbf{x} 的单位矢量，满足 $|\mathbf{n}| = 1$)，进而破缺掉三维旋转对称性。换言之，在临界温度之下，序参量可以写成 $\mathbf{m}(\mathbf{x}) = m_0\mathbf{n}$ 。当然，小箭头们具体选哪个指向是随机的，任何一个方向都同样可能。因此在临界温度以下， $O(3)$ 对称性自发破缺了！不过，破缺之后其实还剩余一种对

称性，那就是，选定了空间指向的小箭头们还可以绕着各自的轴线旋转任意一个角度，通常称这为 $O(2)$ 对称性，也就是二维平面绕着某个竖直轴的旋转对称性。所以，在临界温度以下，铁磁体的 $O(3)$ 对称性自发破缺到了 $O(2)$ 对称性。

所有不同指向但是地位平等的单位向量 \mathbf{n} 的集合构成了三维空间中的二维单位球面 S^2 。在绕 \mathbf{n} 自身轴旋转的 $O(2)$ 对称性的作用下，这个 S^2 上的点是保持不动的，但在模去 $O(2)$ 的 $O(3)$ 旋转作用下， \mathbf{n} 会改变方向，从而会把 S^2 上的一点变换到另一点。正是在这个意义上，我们说这个 S^2 可以等同于破缺前的对称性 $O(3)$ 模去破缺后剩余的对称性 $O(2)$ ，记作

$$S^2 = O(3)/O(2). \quad (56)$$

由于空间旋转可以连续地进行，是所谓的连续对称性，不同于之前的 Z_2 对称性， Z_2 是离散的，所以现在的这种对称性自发破缺叫做**连续对称性的自发破缺**，与离散对称性自发破缺相比，它有一些新的物理效应。

依然假定在临界温度之下。不过，如果考虑到序参量的涨落，那方向 \mathbf{n} 就可以依赖于空间点 \mathbf{x} ，也即是说不同点的小箭头可以指向不同的方向了。这时候的序参量就是一个不均匀的场分布 $\mathbf{m}(\mathbf{x}) = m_0 \mathbf{n}(\mathbf{x})$ ， $|\mathbf{n}(\mathbf{x})| = 1$ 。很显然， $\mathbf{n}(\mathbf{x})$ 定义了一个三维空间到 S^2 的连续映射，

$$\mathbf{n}(\mathbf{x}) : \mathbf{R}^3 \rightarrow S^2, \quad (57)$$

这个映射决定了临界温度之下序参量场 $\mathbf{m}(\mathbf{x})$ 的空间分布。低温铁磁相的物理性质显然由这些序参量的涨落决定。

为了决定低温铁磁相的物理性质，我们需要弄清楚在临界温度之下序参量的涨落对自由能泛函(或者正如前面说过的，系统有效哈密顿量)具体有什么样的贡献。注意，这不同于朗道-金兹堡理论中考察的自由能泛函，那是在临界点附近的，而现在是在临界温度之下的，是有序的铁磁相的自由能泛函。不妨将这个自由能泛函记作 $H_e[\mathbf{m}(\mathbf{x})]$ 。我们可以设所有的小箭头都同步指向某个特定方向时 $H_e = 0$ 。但是，由于所有的方向都是等价的，所以如果把这些指向某个特定方向的小箭头同步地转过一个任意的方向，结果能量将依然为零，即依然有 $H_e = 0$ 。由此我们可以推断， H_e 的非零贡献只能来自于 $\mathbf{m}(\mathbf{x})$ 随不同空间点的变化(因为同步的方向变化不贡献

能量), 也就是依赖于 $\mathbf{m}(\mathbf{x})$ 的空间梯度, 进而也就是依赖于 $\mathbf{n}(\mathbf{x})$ 的空间梯度。

由于 $\mathbf{n}(\mathbf{x})$ 就是一个 $\mathbf{R}^3 \rightarrow S^2$ 的连续映射, 如果想把这个映射写得更清楚的话, 就需要在 S^2 上引入局部坐标系, 比如球坐标。这里的要点是, S^2 上的局部坐标完全是人为的, 因此不止可以取球坐标, 实际上是可以有无数多种局部坐标, 但是, 自由能泛函 $H_e[\mathbf{m}(\mathbf{x})]$ 当然不依赖于 S^2 上局部坐标的选取方式。因此, 为了写出自由能泛函, 我们首先得问, S^2 上什么东西是不依赖于局部坐标的, 回答可能很多, 但最简单的一定是, 两个无限靠近的点在 S^2 上的距离平方不依赖于局部坐标的选取! 这个距离平方很显然是 $ds^2 = (dn_x)^2 + (dn_y)^2 + (dn_z)^2$, 其中 $\mathbf{n} = (n_x, n_y, n_z)$, 且满足约束 $n_x^2 + n_y^2 + n_z^2 = 1$ 。

出于一些纯数学上的方便, 可以在 S^2 上引入复的局部坐标 $(\phi, \bar{\phi})$ ($\bar{\phi}$ 表示 ϕ 的复数共轭), 其定义如下

$$n_x + in_y = \frac{2\phi}{1 + |\phi|^2}, \quad n_x - in_y = \frac{2\bar{\phi}}{1 + |\phi|^2}, \quad n_z = \frac{1 - |\phi|^2}{1 + |\phi|^2}. \quad (58)$$

利用这个局部坐标, 经过一番计算即可以将 S^2 上的 ds^2 写成

$$ds^2 = \frac{4d\phi d\bar{\phi}}{(1 + |\phi|^2)^2}. \quad (59)$$

注意这个式子虽然是在特定局部坐标下写出来的, 但 ds^2 本身实际上并不依赖于局部坐标的选取。

$\mathbf{R}^3 \rightarrow S^2$ 的连续映射当然也可以看作是由 $\phi(\mathbf{x})$ 和 $\bar{\phi}(\mathbf{x})$ 决定。而自由能泛函 H_e 只依赖于 $\mathbf{n}(\mathbf{x})$ 的空间梯度也就是只依赖于 $\phi(\mathbf{x})$ 和 $\bar{\phi}(\mathbf{x})$ 的空间梯度。结合 H_e 不依赖于 S^2 上局部坐标的具体选取的要求(因此由 S^2 上的 ds^2 决定), 可知, 对 H_e 的最简单贡献必定可以写成

$$H_e = \int d^3x \rho_s \frac{4\nabla\phi \cdot \nabla\bar{\phi}}{(1 + |\phi|^2)^2} + \dots, \quad (60)$$

式中比例系数 ρ_s 为某个常数。式中这一项是 $(\nabla\phi)^2$ 阶项, 省略号表示诸如 $(\nabla\phi)^4$ 这样的具有更多个偏导的高阶修正项。这些更多偏导的项之所以能被忽略, 是因为我们主要考察随空间变化比较平缓的序参量涨落, 它的空间偏导是很小的, 因此某项中包含的偏导个数越多就越能够被忽略, 这个论证也适用于后文的相关地方, 之后我们不再重复。

很显然，这个有效哈密顿量说明，在对称性破缺的铁磁相，序参量涨落的能量可以任意地小，只需要让序参量场足够平滑，随空间的变化足够缓慢就可以了。这个结论，有时候人们称之为戈德斯通定理。

很显然，这个有效哈密顿量是场变量 $\phi(\mathbf{x})$ 和 $\bar{\phi}(\mathbf{x})$ 的泛函，而且不是平方类型的泛函，而是真正非线性的，所以它也称作非线性西格玛模型。注意，它描写的是低温铁磁相的物理。这样的非线性西格玛模型有效哈密顿量在离散对称性自发破缺中是不存在的，它是连续对称性自发破缺特有的物理。

自旋波

一个自然的问题是研究场 ϕ 和 $\bar{\phi}$ 如何随时间 t 演化的动力学(注意，这里 t 是时间，不是约化温度)。为此我们就需要一个有效作用量。注意到在经典力学中相空间作用量由 $S = \int dt[p\dot{q} - H]$ 给出，推广到场论是类似的。这里我们有了一个有效哈密顿量 H_e ，因此原则上可以类似地写出一个相空间作用量，关键是前面的 $\Theta = \int dt p\dot{q}$ 项该怎么写。需要注意的是，磁化强度来源于自旋，而自旋是量子的，因此在我们这里这样一项原则上是可以用量子力学方法推导出来的。不过，一般情况下的推导有点繁琐，也超出了本书的范围，下面我们仅就一种简化的但是物理上人们最关心的情况给出这一项。

我们关心的简化情况是，假设磁化强度的小箭头原本都指向 z 方向，即 $(n_x, n_y, n_z) = (0, 0, 1)$ ，如图(10)A所示，序参量是在这个方向附近涨落，也就是考虑 $n_z \approx 1$ ，同时 n_x, n_y 都很小的情形。从(58)式不难看出，这相应于 ϕ 和 $\bar{\phi}$ 是小量。由此立即有 H_e 可以近似成

$$H_e = \int d^3\mathbf{x} 4\rho_s \nabla\phi \cdot \nabla\bar{\phi} + \dots, \quad (61)$$

式中我们忽略了 $|\phi|^2$ 的高阶项。推导表明，这时候相应于经典力学 $\Theta = \int dt p\dot{q}$ 项的是 $\Theta = \kappa \int dt d^3\mathbf{x} i\bar{\phi}\dot{\phi}$ ，其中 κ 是某个比例常数，式中出现虚数单位 i 其实是为了保证这一项为实数，因为通过分部积分并丢掉边界项，我们有

$$\Theta = \kappa \int dt d^3\mathbf{x} i\bar{\phi}\dot{\phi} = -\kappa \int dt d^3\mathbf{x} i\dot{\bar{\phi}}\phi = \overline{\Theta}, \quad (62)$$

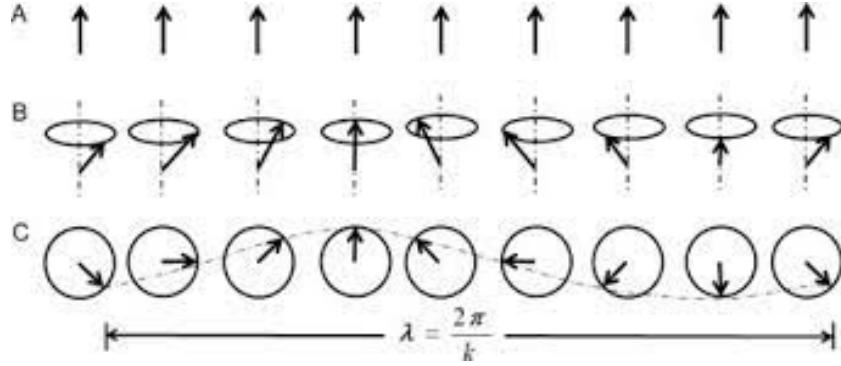


Figure 10: 自旋波, C是从上往下俯视

$\bar{\Theta}$ 表示 Θ 的复共轭。从而因为有了一个*i*, $\Theta = \int dt d^3\mathbf{x} \kappa i \bar{\phi} \dot{\phi}$ 的确是实的。最后归纳起来, 就可以得到如下有效作用量 $S_e[\phi, \bar{\phi}]$

$$S_e[\phi, \bar{\phi}] = \Theta - \int dt H_e = \int dt d^3\mathbf{x} [i\kappa \bar{\phi} \partial_t \phi - 4\rho_s \nabla \phi \cdot \nabla \bar{\phi} + \dots]. \quad (63)$$

将 $S_e[\phi, \bar{\phi}]$ 对 $\bar{\phi}$ 变分, 并利用最小作用量原理, 不难得到

$$i\kappa \partial_t \phi + 4\rho_s \nabla^2 \phi = 0. \quad (64)$$

这是一个波动方程, 它描写的是磁化强度的波动, 而磁化强度来源于自旋, 所以也称之为自旋波的波动方程。特别的, 假设这种自旋波是平面波 $\phi \sim e^{-i\omega t + i\mathbf{k} \cdot \mathbf{x}}$, 代入上述方程, 即有

$$\omega = 4 \frac{\rho_s}{\kappa} \mathbf{k}^2. \quad (65)$$

如图(10)B,C所示。可见, 铁磁体自旋波的色散关系与一个非相对论性自由粒子完全类似。将这种自旋波进行量子化得到的粒子称之为磁振子, 它是一种玻色子, 色散关系由上式给出。

很显然, 当温度很低时, 由于序参量的涨落, 铁磁体内部充满了这种磁振子的气体, 它们决定了铁磁体在低温时的热容。利用第七章对玻色气体进行的统计物理方法, 不难算出, 这些磁振子气体的热容 C_V 为

$$C_V \sim T^{3/2}. \quad (66)$$

注意这个热容不同于光子气和声子气在低温时的热容 $C_V \sim T^3$, 这是因为磁振子的色散关系是 $\omega \sim \mathbf{k}^2$, 而不是像光子气或声子气那样的 $\omega \sim |\mathbf{k}|$ 。

拓扑缺陷

在铁磁相, 序参量是 $\mathbf{R}^3 \rightarrow S^2$ 的映射。通常这个映射可能就如水面上的小涟漪, 比如自旋波就是这样。但是, 有时候也可以出现更奇特的映射。比如如图(11)所示的那样, 磁化强度从一点向四周发散出去, 形成类似于刺猬那样的序参量场位形。这样的场位形是所谓拓扑缺陷的一种。所谓拓扑缺陷, 是指在某些物理系统中, 由于整体结构的“拓扑性质”限制, 而无法通过平滑连续的微小变化消除的、稳定的不规则性或者“瑕疵”。

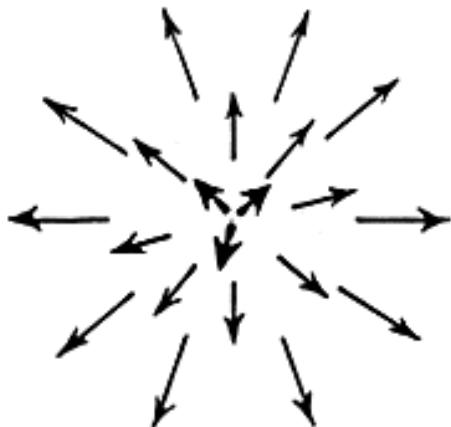


Figure 11: 铁磁相刺猬形的拓扑缺陷

那么限制如图(11)所示刺猬形拓扑缺陷的整体结构是什么呢? 为了回答这个问题, 不妨设想取一个二维球面包围刺猬中心的那一个点, 因为刺猬的每一根刺都对应序参量空间 S^2 上的一个点, 因此这球面上的那些“刺”就定义了一个 $S^2 \rightarrow S^2$ 的连续映射, 后一个 S^2 表示的是序参量空间, 也就是 $O(3)/O(2)$ 。而且, 具体对于图(11)中画的刺猬, 很显然, 这个映射正好是用前一个 S^2 包裹后一个 S^2 一次。这种包裹当然无法通过连续的微小变化消除, 因此是限制刺猬形拓扑缺陷的整体结构。

然而, 人们也可以想象映射 $S^2 \rightarrow S^2$ 不止包裹后者一次, 而是可以包裹任意整数次, 其中负整数表示内外翻转的包裹。很显然, 不同的包裹次数

之间不能连续地过渡，因此包裹次数的不同就表示不同的拓扑缺陷，包裹次数分类了铁磁相的刺猬形拓扑缺陷。包裹次数只能是整数，称作刺猬形拓扑缺陷的**拓扑量子数**。所有不同次数的包裹集合在一起就构成整数集**Z**，数学家通常称之为序参量空间 $O(3)/O(2)$ 的第二同伦群，记作

$$\pi_2(O(3)/O(2)) = \pi_2(S^2) = \mathbf{Z}. \quad (67)$$

也即是说，铁磁相的刺猬形拓扑缺陷由 $\pi_2(O(3)/O(2))$ 分类。

假设把两个拓扑量子数分别为 n_1 和 n_2 的刺猬形拓扑缺陷相互靠近，最终它们就可能结合成一个“更大”的刺猬形拓扑缺陷，这个“更大”的刺猬形拓扑缺陷的拓扑量子数必定是 $n_1 + n_2$ 。也即是说，在拓扑缺陷的结合过程前后，其拓扑量子数是守恒的。同样，一个“大”的刺猬形拓扑缺陷也可能分裂成两个“小”的拓扑缺陷，分裂前后拓扑量子数也是守恒的。特别的，假设取一个拓扑量子数为 n 的拓扑缺陷，和一个拓扑量子数为 $-n$ 的拓扑缺陷，让它们结合，结果就会得到拓扑量子数0，也就是不存在拓扑缺陷了，也即是说，这样的两个拓扑缺陷相互湮灭了！这就和正反粒子对会相互湮灭一样。

值得注意的是，虽然自旋波的能量可以任意小，但拓扑缺陷的能量一定是有有限的，不能任意小。因此，你用很小的能量就可以在铁磁相中激发出自旋波，但却必须耗费一定的有限的能量才能激发出拓扑缺陷。

3.4 朗道范式总结

朗道天才地引入了对称性自发破缺的概念，并把它和相变联系起来，这给了我们一个一般性地理解物质的相和相变的范式，称作朗道范式。下面对朗道范式进行一个总结。

给定一块宏观物质，它是由大量原子组成的，可以设想这些原子的微观相互作用具有某种对称性 G ，为了明确起见，不妨假设 G 是一种连续对称性，比如前面铁磁材料例子中的 $O(3)$ 对称性。高温时，各原子近乎独立地做热运动，系统处于无序相，这种无序相常常具有微观相互作用允许的最高对称性，也就是具有对称性 G 。

随着温度降低到临界温度以下，相互作用就会把大量原子的行为协调起来，表现出宏观上有序的运动模式，从而使得系统转变为有序相。根据

朗道的思想，为了描述这种无序到有序的相变，可以引入一种描述整个系统集体行为宏观量，称作序参量，它在高温无序相取零，在低温有序相非零。比如在铁磁相变中，磁化强度可以作为序参量，而在液气相变中，液态和气态的密度差可以作为序参量。值得强调的是，序参量和对称性是有关关系的，具体来说即是，在微观对称性 G 的作用下，序参量的值要跟着变换。比如铁磁体的序参量磁化强度是一个三维矢量，在微观对称性 $O(3)$ 的作用下，也就是在三维旋转之下，它的方向要跟着变。用数学的话来说即是，序参量构成了微观对称性 G 的线性表示。

在高温时，序参量的取值为零，因此它相对于微观对称性 G 完全对称，所以高温无序相具有最高对称性 G 。但是，在临界温度之下，序参量选取了一个确定的值和“方向”，原来的对称性 G 就破缺了。不过，这时候系统可能还会有剩余的对称性，也就是 G 中那些能够保持序参量的选取“方向”不变的对称性，不妨记这一剩余对称性为 H (注意，这不是哈密顿量)。比方说铁磁体中保持磁化方向不变的 $O(2)$ 旋转。当然，相变时序参量选取哪个“方向”完全是随机的，我们完全也可以用 G 的作用让它变一个“方向”，而这和原来的“方向”完全平等。相变以后，序参量所有可以选取的“方向”的集合就构成空间 G/H ，比方说铁磁相变中的 $O(3)/O(2) = S^2$ 。

临界温度之下，如果不考虑序参量的涨落，那它总要在全空间选取一个上面所述的“方向”，从而把对称性从高温相的 G 自发破缺到低温相的 H 。破缺以后，如果考虑到序参量的“方向”可以在整个空间 G/H 上涨落，那低温有序相的序参量实际上就是从三维空间到 G/H 上的映射，即

$$\mathbf{R}^3 \rightarrow G/H. \quad (68)$$

低温有序相的物理性质就取决于序参量的这种涨落。所以为了描述低温有序相，就需要写出低温时自由能作为序参量场的泛函，或者说有效哈密顿量，记作 H_e 。不过，由于序参量的“方向”作一个整体的改变结果在物理上完全等价，也即是说，序参量“方向”的整体改变完全没有物理效应，从而可以知道， H_e 必定只依赖于序参量场随空间的变化情况，也就是序参量场的空间梯度。这里的一个推论是，序参量场涨落的能量可以任意地小，只需要让序参量场足够平滑，随空间的变化足够缓慢。这个推论被称为戈德斯通定理。

另外， H_e 当然不依赖于空间 G/H 的局部坐标选取。因此对 H_e 最简单的贡献只能由 G/H 空间上两临近点的距离平方 ds^2 决定。但是，为了写出 ds^2 的表达式，我们需要在 G/H 上选取一个局部坐标，不妨记为 ϕ^i ，指标 $i = 1, 2, \dots, m$ ， m 为空间 G/H 的维数。假设在此局部坐标下，一般地有

$$ds^2 = \sum_{i,j} g_{ij}(\phi) d\phi^i d\phi^j. \quad (69)$$

$\phi^i(\mathbf{x})$ 作为序参量场就定义了映射 $\mathbf{R}^3 \rightarrow G/H$ 。那么结合这两段的讨论可以知道，有效哈密顿量的最低阶贡献必定可以写成

$$H_e = \int d^3\mathbf{x} \rho_s \sum_{i,j} g_{ij}(\phi) \nabla \phi^i \cdot \nabla \phi^j + \dots \quad (70)$$

这就是描写低温有序相物理的非线性西格玛模型。

利用低温时的这个有效哈密顿量，我们就可以计算低温时系统的配分函数，公式是

$$\mathcal{Z} = \int \mathcal{D}\phi(\mathbf{x}) e^{-\beta H_e[\phi(\mathbf{x})]}, \quad (71)$$

式中 $\phi(\mathbf{x})$ 示意性地指代所有 $\phi^i(\mathbf{x})$ 。当然，计算这个泛函积分并不是一件简单的事情，感兴趣的读者请自行阅读统计场论的相关书籍。

如果输入的能量足够，那么就有可能在低温有序相中激发出拓扑缺陷，比方说类似于铁磁体情形的那种刺猬状拓扑缺陷，类比于铁磁体情形的相关讨论可以知道，这种拓扑缺陷由 G/H 的二阶同伦群分类，也即由下式分类

$$\pi_2(G/H). \quad (72)$$

不过，在有些例子中，由于对称性 G 和 H 不够凑巧，导致 $\pi_2(G/H) = 0$ ，那就说明在这些例子中不存在这种刺猬状拓扑缺陷。

3.5 应用举例：固体的弹性理论

固体的弹性

作为朗道范式的一个应用，我们可以用它来理解固体的弹性。严格按照朗道范式来的话，这需要从固液相变开始讲起，虽然固液相变不是连续相变，但是朗道关于对称性自发破缺的物理图像在这里依然可以用。给定压强下，液态就属于高温相，固态就属于低温相。

在高温的液相，组成物质的分子位置并不固定，而是在作无规则的热运动，因此处于液相的系统可以进行任意的空间平移而不改变系统的宏观状态，也即使说，液相具有空间平移对称性，记作 $G = \mathbf{R}^3$ ，这里 \mathbf{R}^3 就代表所有空间平移的集合。

但是，在低温的固态，分子的位置会规则地排列，形成晶格，如图(12)所示，图中虚线所画的小圆圈就代表那些规则排列的分子，当然图中画的是二维晶格，但是其显示的物理图像是一般性的。为了简单起见，假设这种晶格是立方形晶格，记晶格的分子间距为 $a_i = (a_1, a_2, a_3)$ ，分别代表长、宽、高方向相邻分子的间距。

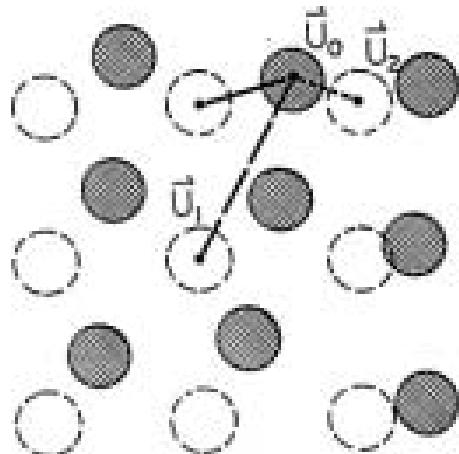


Figure 12: 固体晶格以及位移场

很显然，形成晶格以后，任意的空间平移就不再是系统的对称性了，换言之，平移对称性自发破缺了。不过，晶格依然有一些剩余的平移对称性，那就是它可以往三个方向任意平移整数个格间距而不改变固体的宏观状态。显然，任意这样的平移都对应于某个平移矢量 $(n_1 a_1, n_2 a_2, n_3 a_3)$ ，其中 n_i 为整数。换言之，晶格允许的平移与整数组 (n_1, n_2, n_3) 一一对应，因此

可以把所有这些晶格平移的集合记作 $H = \mathbf{Z}^3$, 它就是固体剩余的空间平移对称性。

那么平移对称性自发破缺对应的序参量涨落是什么呢？回答是，对应于晶格上的每个分子偏离其理想晶格位置的位移，通常可以记作 $u_i(\mathbf{x}) = (u_1(\mathbf{x}), u_2(\mathbf{x}), u_3(\mathbf{x}))$, $u_i(\mathbf{x})$ 表示在理想晶格位置 \mathbf{x} 处的分子发生的位移，如图(12)所示。不过，很显然，如果这个位移刚好是把晶格进行了一个允许的对称平移，那就相当于什么也没发生，因此，我们要对位移变量 u_i 进行如下周期性等同

$$u_i \sim u_i + n_i a_i. \quad (73)$$

不难想明白，等同之后序参量 u_i 的取值空间就变成了三维环面 \mathbf{T}^3 , 这也就是

$$\mathbf{T}^3 = \mathbf{R}^3 / \mathbf{Z}^3. \quad (74)$$

而序参量的涨落就由物理空间 \mathbf{R}^3 到 \mathbf{T}^3 上的一个映射来描述

$$u_i(\mathbf{x}) : \mathbf{R}^3 \rightarrow \mathbf{T}^3. \quad (75)$$

作为低温相的固体的物理性质就由这个序参量的涨落来描述。

和前面关于朗道范式的一般总结中论证的一样，固体的自由能泛函(或者说有效哈密顿量)只能依赖于序参量场随空间的变化情况，也就是只能依赖于 $\partial_i u_j$, 式中 ∂_i 表示对位置坐标的 i 分量进行偏导。通常人们会把 $\partial_i u_j$ 分解成对称部分和反对称部分，分别是

$$\frac{1}{2}(\partial_i u_j + \partial_j u_i), \quad \frac{1}{2}(\partial_i u_j - \partial_j u_i). \quad (76)$$

其中对称部分描写了固体的形变，拉伸挤压等等，而反对称部分描写的是固体的旋转。直观上我们知道，固体的旋转并不影响其内能(因为旋转并不会改变分子间的距离，从而也就不会影响固体的内能)。因此，固体的有效哈密顿量 H_e 只能依赖于对称部分 $\frac{1}{2}(\partial_i u_j + \partial_j u_i)$, 通常会把这个对称部分记作 ϵ_{ij} , 称作应变张量

$$\epsilon_{ij} = \frac{1}{2}(\partial_i u_j + \partial_j u_i). \quad (77)$$

ϵ_{ij} 当然可以看成是 3×3 的对称矩阵，通常就把这个矩阵抽象地记为 ϵ 。 H_e 只能取决于用 ϵ_{ij} 形成的标量，如果固体是各向同性的，那最简单的这样的标量只有两个，分别是

$$\text{Tr}(\epsilon^2), \quad (\text{Tr}\epsilon)^2 \quad (78)$$

后者之所以要平方，是因为要保证在 $u_i \rightarrow -u_i$ 的变换下也是标量。所以，我们可以把各向同性固体的有效哈密顿量写成

$$H_e[\epsilon] = \int d^3\mathbf{x} \left[\frac{1}{2} \lambda (\text{Tr}\epsilon)^2 + \mu \text{Tr}(\epsilon^2) \right], \quad (79)$$

式中 λ, μ 是所谓的拉梅常数，它们刻画的是固体的刚性，因为它们越大，那形变对应的能量就越高。

但是，以上分析默认了没有对固体施加外力！如果有一个应力施加在固体上，那就要在有效哈密顿量中增加一项

$$\int d^3\mathbf{x} - \sum_{i,j} \sigma_{ij} \epsilon_{ij}, \quad (80)$$

这一项类似于外磁场与磁化强度的耦合项，其中 σ_{ij} 是所谓的应力张量。当热平衡时，根据自由能泛函极小的条件将自由能泛函 H_e 对 ϵ_{ij} 进行变分，就能得到

$$\sigma_{ij} = \lambda(\text{Tr}\epsilon)\delta_{ij} + 2\mu\epsilon_{ij}, \quad (81)$$

这就是著名的胡克定律，也就是所谓的应力与应变成正比，通常称这样的固体为线弹性固体。

如果我们要研究位移场随时间的动力学演化，也就是研究固体中的弹性波，那么就需要注意到有效哈密顿量(79)写出来的只是弹性势能，动力学演化的位移场当然还会有一个动能，它是 $\int d^3\mathbf{x} \frac{1}{2} \rho \sum_i (\partial_t u_i)^2$ ，式中 ρ 为固体的质量密度，通常可以看作是常数。为了讨论位移场的动力学演化，我们需要写出其有效作用量，它是(注意拉格朗日量是动能减势能)

$$S_e[u_i] = \int dt d^3\mathbf{x} \left[\frac{1}{2} \rho \sum_i (\partial_t u_i)^2 - \frac{1}{2} \lambda (\text{Tr}\epsilon)^2 - \mu \text{Tr}(\epsilon^2) \right]. \quad (82)$$

对位移场 $u_i(\mathbf{x}, t)$ 进行变分，并利用最小作用量原理，可得

$$\rho \partial_t^2 u_i = \lambda \partial_i (\text{Tr} \epsilon) + 2\mu \sum_j \partial_j \epsilon_{ij}. \quad (83)$$

如果记 $\mathbf{u} = (u_1, u_2, u_3)$ ，那么这个方程也可以重写成

$$\rho \partial_t^2 \mathbf{u} = (\lambda + \mu) \nabla (\nabla \cdot \mathbf{u}) + \mu \nabla^2 \mathbf{u}. \quad (84)$$

这就是线弹性各向同性固体中弹性波的波动方程。

如果把弹性波分解成额外满足 $\nabla \cdot \mathbf{u} = 0$ 的剪切波和额外满足 $\nabla \times \mathbf{u} = 0$ 的压缩波，那么上述方程就可以分解成

$$\rho \partial_t^2 \mathbf{u} = \mu \nabla^2 \mathbf{u} \quad \text{对于剪切波} \quad (85)$$

$$\rho \partial_t^2 \mathbf{u} = (\lambda + 2\mu) \nabla^2 \mathbf{u} \quad \text{对于压缩波} \quad (86)$$

额外满足的条件告诉我们，剪切波是横波而压缩波是纵波，它们各自的波动方程又告诉我们，剪切波的波速是 $c_s = \sqrt{\mu/\rho}$ ，而压缩波的波速是 $c_p = \sqrt{(\lambda + 2\mu)/\rho}$ 。可见，作为纵波的压缩波比作为横波的剪切波传播速度快，这就是为什么地震时，人们总是先感觉到上下起伏然后才感觉到水平方向的左右晃动。

位错

固体同样有拓扑缺陷，但不是刺猬状拓扑缺陷。为了讲清楚固体的拓扑缺陷，我们先从二维固体开始，也就是从一张厚度可以忽略的固体薄片开始。假设相应的晶格为二维四方晶格，因此现在位移场 $u_i(\mathbf{x})$, $i = 1, 2$ 就应该是 $\mathbf{R}^2 \rightarrow \mathbf{T}^2$ 的映射，二维环面 \mathbf{T}^2 为序参量的取值空间。

缺陷是序参量场中的一道裂痕。拓扑缺陷则是无法被修复的裂痕。请看图(13)中所示的那片二维晶体：从图中所示区域的中心开始，多出了一排原子（这被称为位错）。但在远离中心的位置，晶体在局部看起来仍然是良好的：虽然略有畸变，但我们仍能清晰地看出正方形晶格，并可以定义一个序参量。那么，我们是否可以通过重新排列那排多出的原子起始处周围小区域内的原子，来“修补”这个缺陷呢？

不行。问题在于，我们不用靠近这个缺陷的中心区域，就能判断出存在一排多余的原子。传统的方法是沿着一条围绕该缺陷的大环路行进，并统

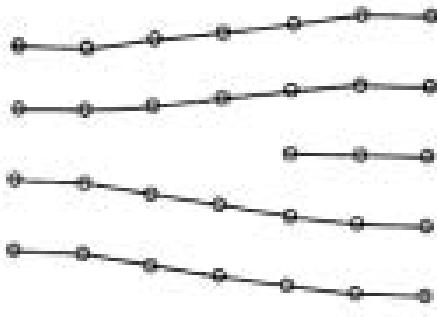


Figure 13: 位错

计路径上所穿过的净行数。在图(14)所示路径中，有两条晶格行向上走，而三条向下走：无论我们将路径设置得离中心多远，右边总会多出一排原子——这是固有的现象。

如何将这一基本思想推广到具有破缺对称性的一般问题中？回想一下，二维方形晶体的序参量空间是一个环面 \mathbf{T}^2 。某一点的序参量表示将理想方形晶格与该点变形晶格对齐所需的平移量。那么，在缺陷左侧远处（a点）与右侧远处（d点）的序参量有何不同？显然，右侧晶格在垂直方向上移动了半个晶格常数：序参量绕环面移动了半周。如图(14)所示，沿顺时针回路上半部分时，序参量（原胞内原子位置）向上移动，而下半部分同样向上移动。总体而言，序参量绕环面完整环绕一次。绕环面的卷绕数即为围绕缺陷一周时序参量环绕环面的净次数。

这就是它们被称为拓扑缺陷的原因。任何未环绕环面的序参量场，无论其如何扭曲，均可平滑恢复至均匀状态。但若沿某回路中序参量绕环面孔洞或穿过孔洞的净次数不为零，则该回路内必存在无法通过平滑连续变形消除的缺陷——因为整数卷绕数无法通过连续变化改变。

对于二维方形晶体的缺陷分类，存在两个整数：绕中心孔洞的次数与穿过孔洞的次数，分别记为 n_1, n_2 。传统描述中，这恰好对应我们经过的额外原子行和列的数量。也即是说，绕着环绕缺陷的回路L形走一周，我们有

$$\oint_L \sum_j dx^j \partial_j u_i = n_i a_i, \quad (87)$$

式中 $x^j = \mathbf{x} = (x^1, x^2)$ 为空间位置矢量， a_i 为两个晶格常数。我们称这两个

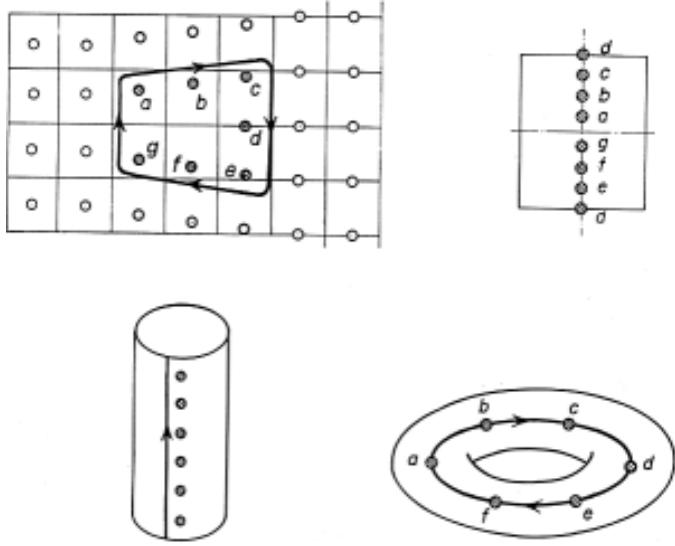


Figure 14: 用一条回路围绕位错, 其中第二幅图是将序参量空间的环面切开成一个长方形, 第三幅图是把这个长方形左右等同起来形成柱面, 第四幅图是进一步把上下边也等同起来, 以恢复成环面

卷绕数所构成的集合为环面的第一同伦群, 记为

$$\pi_1(\mathbf{T}^2) = \mathbf{Z}^2. \quad (88)$$

从而, 位错拓扑缺陷由 $\pi_1(\mathbf{T}^2)$ 分类。

推广到三维固体, 那位错中心就不是一个点了, 而是沿着垂直于屏幕方向的一条线。在真正的三维固体中, 这条位错线当然不必是直线, 它是可以弯曲的。大多数金属, 尤其是铜, 会表现出所谓的加工硬化现象。铜管很容易弯曲, 但要把它弯回原状却出奇地难。原始柔软的铜相对来说几乎没有缺陷。要使铜管弯曲, 晶体必须产生大量线位错, 这些位错移动从而产生弯曲。这些线缺陷会相互缠绕, 阻碍新缺陷的产生。所以, 当你试图把铜管再弯回去时, 金属就会变得硬得多。

最后, 我们要说, 类似于本小节的方法也可以用来构建液晶的弹性理论, 包括其拓扑缺陷。限于篇幅, 具体内容这里就不展开了。读者可以自己问一下DeepSeek等人工智能。更详细的内容可以参阅朗道理论物理教程第七卷《弹性理论》。

4 附录：化学反应的热力学

4.1 化学平衡条件

多元单相系的化学平衡条件

考虑一个含有 k 种不同物质粒子，每种粒子的数目为 N_i 的多元系。如果不同物质成分之间可以相互转化，就称这种转化为化学反应。对于多元系的平衡过程，我们可以把涉及吉布斯自由能的热力学关系式推广为

$$dG = -SdT + VdP + \sum_{i=1}^k \mu_i dN_i. \quad (89)$$

假设系统处于恒温恒压的条件下，那么就有 $dG = \sum_{i=1}^k \mu_i dN_i$ 。由于化学反应的粒子数变化属于系统内部的变化，因此化学平衡时对于整个系统必定有自由能取极小值，即 $dG = 0$ ，从而即可以得到化学平衡条件

$$\sum_{i=1}^k \mu_i dN_i = 0. \quad (90)$$

将系统中各种物质的分子式简记为 A_i ，化学反应的方程可以抽象地写为

$$\sum_i \nu_i A_i = 0. \quad (91)$$

其中， ν_i 表示反应方程中的配平系数， $\nu_i > 0$ 则表示相应的 i 物质为反应生成物， $\nu_i < 0$ 则表示是反应物。由于化学反应方程是齐次方程，系数 ν_i 的选择不是唯一的，可以相差一个共同的倍数。为了唯一确定 ν_i ，我们规定：在反应方程中，当我们主要关心某种物质成分的变化时，就将该物质成分分子式前的系数的绝对值取为1。

显然，在化学反应中，各物质成分粒子数的改变量 dN_i 应该正比于 ν_i ，因此就可以把化学平衡条件改写为

$$\sum_i \mu_i \nu_i = 0. \quad (92)$$

如果化学反应平衡被打破，这时系统就会朝着吉布斯自由能减少的方向演化，直至达成新的化学平衡。因此，化学反应进行的方向由下式决定

$$\sum_i \mu_i \nu_i \leq 0. \quad (93)$$

多元多相系的平衡条件

如果考虑的系统是多元多相系，那么在反应方程中，不仅要对不同的组元求和，同时还要对不同的相求和，因此多元多相系的相变-反应方程可以写成

$$\sum_{i,\alpha} \nu_{i\alpha} A_{i\alpha} = 0. \quad (94)$$

式中指标 α 是用来区分不同的相。很显然这时候平衡条件就应该写成

$$\sum_{i,\alpha} \mu_{i\alpha} \nu_{i\alpha} = 0. \quad (95)$$

平衡被打破时，系统的演化方向由下式决定

$$\sum_{i,\alpha} \mu_{i\alpha} \nu_{i\alpha} \leq 0. \quad (96)$$

4.2 质量作用定律

现在假设我们考虑的是单相化学反应，假定系统中各组元均处于气相，而且可以看作是理想气体。由于理想气体的各组元可以看成是相互独立的，因此可以对每一组元分别用第五章中谈到的微分Gibbs-Duhem关系式，对于等温过程有

$$d\mu_i = \frac{V}{N_i} dP_i, \quad (97)$$

式中 P_i 为组元*i*的分压(根据道尔顿分压定律，系统的总压强等于各组元分压之和)。代入理想气体物态方程 $P_i V = N_i k_B T$ ，即有

$$d\mu_i = k_B T \frac{dP_i}{P_i}, \quad (98)$$

对上式进行积分，即有

$$\mu_i = \mu_{i0} + k_B T \log \left(\frac{P_i}{P_{i0}} \right), \quad (99)$$

式中 μ_{i0} 为该组元在参考压强 P_{i0} 时的化学势。将这个式子代入(93)式，即有

$$\sum_i \nu_i \left[\frac{\mu_{i0}}{k_B T} + \log \left(\frac{P_i}{P_{i0}} \right) \right] \leq 0. \quad (100)$$

取指数，即有

$$\prod_i \left(\frac{P_i}{P_{i0}} \right)^{\nu_i} \leq K_P, \quad (101)$$

式中 $K_P = \prod_i [e^{-\frac{\mu_{i0}\nu_i}{k_B T}}]$ 是温度的函数。假设 i 组元粒子数所占的百分比为 x_i ，系统的总压强为 P ，那么即有

$$P_i = P x_i. \quad (102)$$

从而即可以将(101)式改写为

$$\prod_i (x_i)^{\nu_i} \leq K, \quad K = K_P \prod_i \left(\frac{P_{i0}}{P} \right)^{\nu_i}. \quad (103)$$

(101)式和(103)式统称为质量作用定律。通过改变外界条件来控制 K_P 和 K ，可以控制化学反应进行的方向和反应程度。例如，假设系统中的化学反应已经达到平衡，这时若通过改变环境温度使得 K 的值变大，那么系统中的化学反应将重新开始向正向演化，直至新的平衡被建立。相反，如果通过控制条件使得 K 变小，反向将向逆向进行。