Statistical Physics (526)

Daniel M. Sussman

December 26, 2020

Contents

Abstract and sources

This is a set of lecture notes prepared for PHYS 526: Statistical Physics (Emory, Spring 2020). It is somewhat more verbose than what I will actually write on the board, but far from a comprehensive textbook.

There are undoubtedly typos and errors in this document: please email any corrections corrections to:

daniel.m.sussman@emory.edu

There are large variations in how I wrote these notes as the semester progressed – this was my first time teaching, and what I needed out of a set of lecture notes on day 1 was... quite different from what I needed for recording zoom lectures by the end of the suddenly online semester.

Sources used

These notes are not original. They represent a merging of many of the sources that I learned stat mech from, as well as resources I've been reading over the course of the semester. As I said on the syllabus for the class, "Graduate-level statistical physics is a subject with many available textbooks and wide disagreements about which one(s) to use." For these notes I have particularly drawn from:

1. Pathria & Beale (*Statistical Mechanics, 3rd edition*; Primary source),
2. Kardar (lectures & *Statistical Physics of particles*; Primary source),
3. Goldenfeld (*Lectures on Phase Transitions and the Renormalization Group*; general secondary source, especially for chapter on phase transitions),
4. Preskill (Chapter 10 of his Quantum Information notes for discussion on information entropy and mutual information.)
5. David Tong (Chapter 2 of his lecture notes on Kinetic Theory, Chapter 1 of his notes on Statistical Physics for parts of Chapter 3 of this document)
6. Huang (Chapter 5 for some parts of hydrodynamics. Also, the structure of this book – which is, not surprisingly, echoed in Kardar – has inspired the progression of topics covered here)
7. Sethna (Entropy, Order Parameters, and Complexity; general source)
8. Kadanoff (book; general source)

Basic notation in the text

Triple lines, like so:

---

refer to estimated lecture breaks. These lost meaning once courses moved online (when I started recording individual sections or subsections as lectures – no need to stick to record in 75-minute blocks!).

Text that appears in blue in these documents are things that I probably won't write on the board, but will likely be discussed, or provide (hopefully) useful additional context, etc. My use of this command varies strongly by chapter at the moment, and is most present early on in the notes.

Text that appears in red in these documents are things that I intend to *not* go over in lectures, and which are perhaps not related to the core ideas of the course but are necessary to complete particular derivations. A first example are some elements of classical scattering theory that appears in Chapter 3: Calculating the differential cross sections that appear there are not particularly in the scope of the class, but the definitions help us get to the Boltzmann equation.

Chapter 0

Thermodynamics[1]: review and background

0.1 Thermodynamics: a phenomenological description of equilibrium properties of macroscopic systems

> "Suppose you've got theoretical physics cracked. Suppose you know all the fundamental laws of Nature, the properties of the elementary particles and the forces at play between them. How can you turn this knowledge into an understanding of the world around us? More concretely, if I give you a box containing 10^{23} particles and tell you their mass, their charge, their interactions, and so on, what can you tell me about the stuff in the box?
> There's one strategy that definitely won't work: writing down the Schrödinger equation for 10^{23} particles and solving it. That's typically not possible for 23 particles, let alone 10^{23}. What's more, even if you could find the wavefunction of the system, what would you do with it? The positions of individual particles are of little interest to anyone. We want answers to much more basic, almost childish, questions about the contents of the box. Is it wet? Is it hot? What colour is it? Is the box in danger of exploding? What happens if we squeeze it, pull it, heat it up? How can we begin to answer these kind of questions starting from the fundamental laws of physics?
> The purpose of this course is to introduce the dictionary that allows you translate from the microscopic world where the laws of Nature are written to the everyday macroscopic world that we're familiar with. This will allow us to begin to address very basic questions about how matter behaves." – David Tong, *Lecture notes on Statistical Physics*

We begin with a few phenomenological definitions:

---

[1] "Thermodynamics is a funny subject. The first time you go through it, you don't understand it at all. The second time you go through it, you think you understand it, except for one or two small points. The third time you go through it, you know you don't understand it, but by that time you are used to it, so it doesn't bother you any more." – Arnold Sommerfeld, As quoted in: J.Muller, Physical Chemistry in Depth (Springer Science and Business Media, 1992)

As phenomenology, it is based on empirical observations, summarized by the laws of thermodynamics; a consistent mathematical framework is then built on top of these observations.

Closed system

We will think about isolating a system thermally by "adiabatic walls" that do not allow heat exchange with the outside world. This is like a "point particle" approximation We will sometimes consider "diathermic walls" which do allow such exchange of heat.

Equilibrium

A state in which "properties" don't change over the *period of observation* – this dependence on observation time makes the definition a *subjective* one. You observer that you do something, the system goes through some transient behavior, and then settles down into a state which seems not to change.

Macroscopic properties

The systems under study will be characterized by *thermodynamic coordinates* or *state functions*, such as *mechanical* properties: $(V, P)_{\text{gas}}$, or $(L, F)_{\text{wire}}$, or $(M, B)_{\text{magnet}}$, etc..., *plus some thermal properties.*

Phenomenology: How do these thermodynamic coordinates depend on each other, or co-evolve? Rely on empirical observations and from them construct laws of thermodynamics.

0.2 0^{th} Law

The zeroth law is a statement of the transitivity of (thermal) equilibrium:

Observation

If two systems, A and B, are separately in equilibrium with system C, then they are in equilibrium with each other

Note that this implies the existence of "Temperature," some additional thermodynamic coordinate, we can use to describe a system.

Implications

We describe the state of each system, A, B, and C, by a set of thermodynamic coordinates, e.g. $\{A_1, A_2, \ldots\}$. The statement "A and C are in equilibrium" can be expressed by a constraint between these coordinates that is, a change in A_1 must be accompanied by some changes in $\{A_2, \ldots, C_1, C_2, \ldots\}$ to maintain the equilibrium between the states. There should be an additional coordinate (i.e., on top of the mechanical thermodynamic coordinates) to describe the system... we'll call that coordinate "temperature". Let's write that constrain via a function; what follows is physicists' math... hand-waving ensues.

$$f_{AC}\left(A_1, \ldots, C_1 \ldots\right) = 0. \tag{1}$$

Likewise

$$f_{BC}\left(B_1, \ldots, C_1 \ldots\right) = 0. \tag{2}$$

Each of the above can be, in general, written as a constraint on one of the coordinates of C:

$$F_{AC}\left(A_1, \ldots, C_2 \ldots\right) = C_1 = F_{BC}\left(B_1, \ldots, C_2 \ldots\right). \tag{3}$$

Is it obvious we can even *do* this, mathematically (i.e., go from Eq. 1 to 3)? No. But physically, yes.

The above is a statement about the first ("if") clause of the zeroth law. Great. But the zeroth law says that A and B are *also* in equilibrium, so there is some function

$$f_{AB}\left(A_1, \ldots, B_1 \ldots\right) = 0. \tag{4}$$

Furthermore, it must be possible to simplify Eq. 3 by cancelling the coordinates of system C. Thus,

$$F_A\left(A_1, \ldots\right) - F_B\left(, B_1 \ldots\right) = 0. \tag{5}$$

We'll denote this function, the *empirical temperature* by Θ, so

$$\Theta_A\left(A_1, \ldots\right) = \Theta_B\left(, B_1 \ldots\right). \tag{6}$$

Thus: equilibrium of systems can be cast as a function that depends only on the coordinates of one of the systems. Draw isotherm of ideal gas?

One can also say something along the lines of "The zeroth law is like an equivalence relation between mathematical sets – the equivalence relation partitions the space of all possible thermodynamic coordinates into mutually distinct subsets; let's *label* those subsets by something, and we'll call that something "temperature."

Ideal gas scale

The zeroth law states the existence of isotherms: e.g., $\Theta_A\left(A_1, \ldots\right) = \Theta$ Think of, e.g., ideal gas law, or Curie paramagnet, or Hooke's law for rubbers, or van der Waals gases, or.... To construct an actual temperature scale we need a well-defined reference systems. Empirical observation: the product of pressure times volume is constant along the isotherms of any gas that is sufficiently dilute. Think of, e.g., a piston in the ocean or something The ideal gas refers to the dilute limit of real gases (i.e., in the $P \to 0$ or $V \to \infty$ limits), and the constant of proportionality is determined by reference to the triple point of the ice-water-steam system, which was defined as 273.16°K By the 10th General Conference on Weights and Measures, 1954. Revisions in 1990? .

So define an empirical temperature by using a dilute gas as a thermometer:

$$T(^\circ K) \equiv 273.16\left(\lim_{P\to 0}(PV)_{\text{system}}\right) / \left(\lim_{P\to 0}(PV)_{\text{ice-water-steam}}\right) \tag{7}$$

0.3 1^{st} Law

The first law is a statement about the conservation of energy, adapted for thermal systems. We'll formulate it as:

Statement

If the state of an adiabatically isolated system is changed by work, the amount of work is only a function of the initial and final coordinates of the system. Draw, fake system with a spring, magnet, etc., a coordinate space representation with initial and final points, and many paths between them. ΔW doesn't depend on path

Consequences

We infer the existence of *another* state function, the internal energy $E(\boldsymbol{X})$. Think about how the path-independence of the work we have to do when pushing a ball up a frictionless hill in classical mechanics lets us deduce a potential energy.

$$\Delta W = E(\boldsymbol{X}_f) - E(\boldsymbol{X}_i) \tag{8}$$

Similarly, in the same sense that the zeroth law let us construct some function of coordinates that was relevant to equilbrium, the first law allows us to define another function, the internal energy.
Draw some squiggly paths on the board.

The real content of the first law is when we violate the condition. That is, allow walls that permit heat exchange, so that $\Delta W \neq E_f - E_i$. We, of course, still believe energy is a good, conserved quantity, so define *heat*:

$$\Delta Q = (E_f - E_i) - \Delta W. \tag{9}$$

Clearly, though, ΔQ and ΔW are not separate functions of state, so we will use notation like:

$$dE(\boldsymbol{X}) = đQ + đW, \tag{10}$$

Where d means a differential where the thing is a function of state, and $đ$ means that the thing is path-dependent. Note the sign convention, here, where work and heat *add* energy to the system.

Quasi-static transformation

A QS transformation is one which is done sufficiently slowly enough to maintain the system in equilibrium everywhere along the path. For such a transformation the work done *on* the system can thus be related to changes in the thermodynamic coordinates. Let's divide the state functions, $\boldsymbol{X}$, into *generalized displacements* $\boldsymbol{x}$ and *generalized forces* $\boldsymbol{J}$. Then, in a QS transformation

$$đW = \sum_i J_i dx_i \tag{11}$$

Common generalized coordinates

System	generalized force		generalized displacement	
Wire	tension	F	length	L
Film	surface tension	σ	area	A
Fluid	pressure	$-P$	volume	V
Magnet	field	B	magnetization	M

Note that the *displacements are generally extensive* and the *forces are generally intensive.*

Question: We've written

$$dE = \sum_i J_i dx_i + \boxed{?} \tag{12}$$

What is $đQ$? probably depends on T. What is it conjugate to?

Response functions

The usual way of characterizing the behavior of a system (measured from changes in thermodynamic coordinates with external probes). E.g.:

Force constants Measure the ratio of displacements to forces (think spring constants). For example, *isothermal compressibility of a gas*: $\kappa_T = \frac{1}{V}\frac{\partial V}{\partial P}\big|_T$

Thermal response Response to change in temperature, such as the expansivity of a gas $\alpha_P = \frac{1}{V}\frac{\partial V}{\partial T}\big|_P$

Heat capacity Changes in temperature upon adding heat. Note that heat is not a state function, so the path is important! Example: For an ideal gas we could calculate $C_V = \frac{đQ}{dT}\big|_V$ and $C_P = \frac{đQ}{dT}\big|_P$, and C_P has to be bigger since we use some heat to change the volume:

$$\begin{aligned} C_V &= \left.\frac{đQ}{dT}\right|_V = \left.\frac{dE - đW}{dT}\right|_V = \left.\frac{dE + PdV}{dT}\right|_V = \left.\frac{\partial E}{\partial T}\right|_V . \\ C_P &= \left.\frac{đQ}{dT}\right|_P = \left.\frac{dE - đW}{dT}\right|_P = \left.\frac{dE + PdV}{dT}\right|_P = \left.\frac{\partial E}{\partial T}\right|_P + P\left.\frac{\partial V}{\partial T}\right|_P . \end{aligned} \tag{13}$$

Joule's free expansion experiment

Take an adiabatically isolated gas, and let it expand (adiabatically, but we don't need QS. Draw on the board a two-chambered system) from V_i to V_f. Joule observed that the initial and final temperatures are the same! $T_f = T_i = T$.

So, $\Delta Q = 0$ and $\Delta W = 0$, so $\Delta E = 0$. We conclude that the internal energy actually depends only on temperature: $E(P, V, T) = E(T)$, i.e., a product of P and V. Note that since E depends only on T, $\frac{\partial E}{\partial T}\big|_V = \frac{\partial E}{\partial T}\big|_P$, and we can simplify the heat capacity expressions:

$$C_P - C_V = P\left.\frac{\partial V}{\partial T}\right|_P = \frac{PV}{T} = Nk_B. \tag{14}$$

That last equality is a statement of extensivity, that PV/T is proportional to an amount of stuff, and $k_B \approx 1.4 \times 10^{-23} J/K$.

0.4 2^{nd} Law

Why does heat flow from hot to cold? Why are there no perpetual motion machines that work by turning water into ice while doing work? There are many equivalent formulations of the 2nd law; in part because it is fun, we'll see how practical concerns about burning coal to do stuff leads directly to the idea of entropy and its inevitable increase!

Kelvin's statement

No process is possible whose sole result is the complete conversion of heat to work ("No ideal engines")

Clausius' statement

No process is possible whose sole result is the transfer of heat from cold to hot ("No ideal refrigerators")

Idealized work machines

We'll quantify these statements by defining "figures of merit" for an ideal engine and an ideal refrigerators.

The *efficiency* of an engine, a machine which takes Q_H of heat from a source, converts some of it to work W, and dumps some Q_C of it into a sink, is

$$\eta = \frac{W}{Q_H} = \frac{Q_H - Q_C}{Q_H} \leq 1. \tag{15}$$

The *performance* of a refrigerator, an engine running backwards, is

$$\omega = \frac{Q_C}{W} = \frac{Q_C}{Q_H - Q_C} \tag{16}$$

Of course, Kelvin and Clausius' formulations are equivalent! To see this, hook up an ideal engine to a fridge, and you get an ideal fridge (so, not Kelvin implies not Clausius). Additionally, run an ideal fridge and take the heat from the exhaust to power an engine and you get an ideal engine (so, not Clausius implies not Kelvin). Thus, Kelvin $\iff$ Clausius.

This seems trivial; with an excursion through Carnot Engines we'll see that it lets us answer a question posed in section 0.3, when we wrote:

$$dE = \sum_i J_i dx_i + \boxed{?}$$

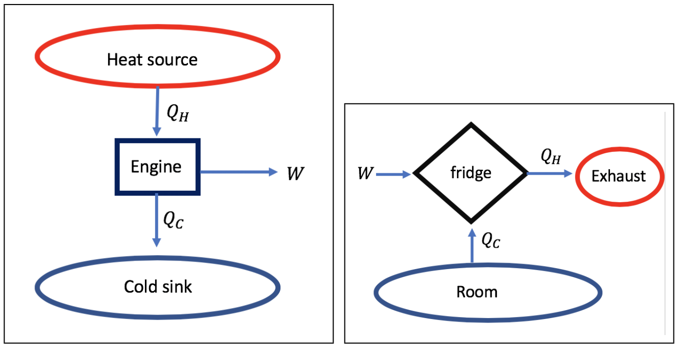


Figure 1: **Idealized engine (left) and refrigerators (right)**

0.5 Carnot Engines

A *Carnot Engine* (CE) is any engine that (1) is reversible, (2) runs in a cycle, and (3) operates by exchanging heat with a source temperature T_H and a sink temperature T_C. Note: (1) is like a generalization of "frictionless" condition in mechanics. Lets us go forward/backward by reversing inputs/outputs. (2) Start and end points are the same. (3) This is more precise than the figure we drew in 1; the sinks and sources have well-defined thermodynamic temperatures.

Ideal gas Carnot Cycle

We know from the 0th law that we can select two ideal-gas isotherms to be the two temperatures. For instance, we could go from A to B or from C to D in Fig. 2 by reversible, adiabatic paths where we maintain the temperature. But how to go *between* the isotherms adiabatically?

For an ideal gas, we know enough to compute the adiabatic curves. Let

$$E = \frac{3}{2} N k_B T = \frac{3}{2} PV. \tag{17}$$

Along a quasi-static path we have

$$\begin{aligned} \text{đ}Q &= 0 = dE - \text{đ}W = d(\frac{3}{2}PV) + pdV = \frac{5}{2}PdV + \frac{3}{2}VdP \\ \Rightarrow & \quad 0 = \frac{dP}{P} + \frac{5}{3}\frac{dV}{V} \Rightarrow PV^{5/3} = \text{constant}. \end{aligned} \tag{18}$$

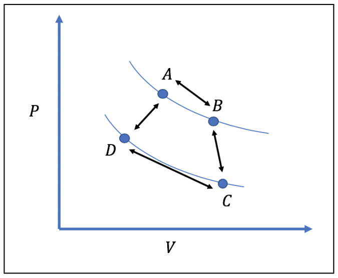


Figure 2: **Schematic of a Carnot Cycle for an ideal gas** Note that even schematically there is something obviously wrong with a path drawn here; do you see the error?

It's fun to see (i.e., will probably be a homework problem), that one can construct adiabatics for any two-parameter system with internal energy $E(J, x)$.

Carnot's Theorem

Off all engines operating between T_H and T_C, the Carnot engine is the most efficient!

Proof Take a Carnot Engine, and use a non-Carnot-Engine's output to run the CE as a refrigerator. Let primes refer to heat connected to the Carnot engine, and unprimes to the NCE. The net effect is to transfer heat $Q_H - Q'_H = Q_C - Q'_C$ from T_H to T_C. Clausius' formulation tells us you can't transfer negative heat, so $Q_H \geq Q'_H$. But the amount of work, W, was the same, so

$$\frac{W}{Q_H} \leq \frac{W}{Q'_H} \Rightarrow \eta_{CE} \geq \eta_{NCE}. \tag{19}$$

0.5.1 Thermodynamic Temperature Scale

We established (by finding the adiabatic paths) that we can (in theory) construct a Carnot Engine using an ideal gas. All Carnot engines operating between T_H and T_C have the same efficiency show by using 1 to run the other backwards, and vice verse, so $\eta_{CE_1} = \eta_{CE_2}$. Thus, the efficiency is independent of the engine; it must depend only on the temperatures, i.e. we have $\eta(T_H, T_C)$. So, already, *if* you can build a CE, it lets us define T independent of any material properties, just by knowing efficiencies of CE's at different T.

We'll make progress by running two engines in series; one between T_1 and T_2, and the other between T_2 and T_3, as in Fig. 3.

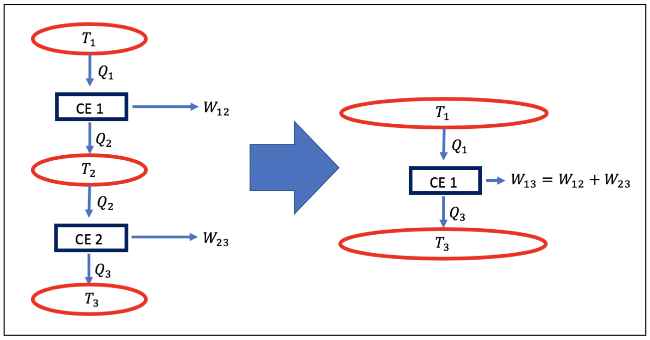


Figure 3: **Schematic of Carnot engines in series**

CE1 tells us

$$Q_2 = Q_1 - W_{12} = Q_1(1 - \eta(T_1, T_2)), \tag{20}$$

CE2 tells us

$$Q_3 = Q_2 - W_{23} = Q_2(1 - \eta(T_2, T_3)) = Q_1(1 - \eta(T_1, T_2))(1 - \eta(T_2, T_3)), \tag{21}$$

and the combined engine tells us

$$Q_3 = Q_1 - W_{13} = Q_1(1 - \eta(T_1, T_3)). \tag{22}$$

Comparing those last two expressions tells us

$$(1-\eta(T_1,T_3)) = (1-\eta(T_1,T_2))(1-\eta(T_2,T_3)), \tag{23}$$

which is a constraint on the functional form that η can take. We postulate that

$$(1-\eta(T_1,T_2)) = \frac{Q_2}{Q_1} \equiv \frac{f(T_2)}{f(T_1)}. \tag{24}$$

By convention, let $f(T) = T$. Thus

$$\eta(T_H,T_C) = \frac{T_H - T_C}{T_H}. \tag{25}$$

We've done it! Up to a constant of proportionality, Eq. 25 defines a thermodynamic temperature (and we'll again set the constant using the triple point of water-ice-steam). By running a Carnot cycle for an ideal gas you can show that the ideal gas scale and the thermodynamic temperature scale are identical. This is not useful, but rather conceptual in showing that temperature is not something that depends on the properties of a particular material. Fun note: thermodynamic temperatures must be positive, otherwise Kelvin's Formulation could be violated

0.5.2 Clausius' Theorem

Statement

For any cyclic process, with path parameterized by s

$$\oint \frac{đQ(s)}{T(s)} \leq 0, \tag{26}$$

where the heat $đQ(s)$ is an amount of heat delivered to the system at temperature $T(s)$ we need not be in equilibrium, so what is $T(s)$? The heat of the "machine" delivering the heat.

proof

We'll hook up the system to a Carnot engine note that we have not specified the sign of $đQ$; using a Carnot engine and thinking of the cycle as a series of infinitesimal cycles lets us make sure we're delivering whatever $đQ(s)$ needs to be. Let the Carnot engine be at a given reference temperature T_0, as in Fig. 4A. To prove the theorem, simply reinterpret things by pretending the Carnot engine and the system are a single "device." It looks like the setup in Fig. 4B. This looks dumb, but we're done! To deliver heat at a specified temperature, our efficiency functions from above tell us that $đQ_0 = T_0 đQ/T(s)$. So, From this view the net extracted heat is

$$\oint đQ_0(s) = \oint T_0 \frac{đQ(s)}{T(s)} \leq 0, \tag{27}$$

where the last inequality follows from Kelvin's formulation: We can't only convert Q to W!

This, as with so much Carnot engine manipulations, feels trivial. There are major consequences!

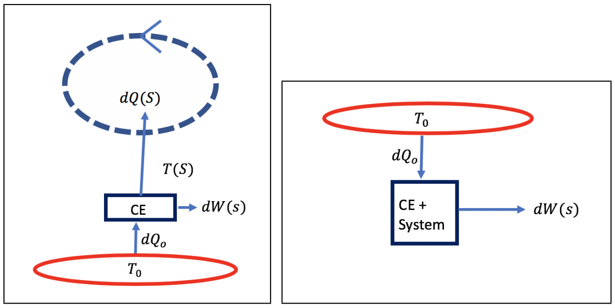


Figure 4: **Clausius theorem setup (left) and reinterpretation (right)**

1. Entropy! Let's apply Clausius' Theorem to a **reversible** transformation. Then we have

$$\pm \oint \frac{\text{đ}Q_{rev}(s)}{T(s)} \leq 0. \tag{28}$$

Since both the plus and the minus version are less than or equal to zero, the integral vanishes. Now, break the cycle into a path from A to B and from B to A "the other way." (draw picture). These integrals must be equal:

$$\int_A^B \frac{\text{đ}Q^1_{rev}(s)}{T^1(s)} = \int_A^B \frac{\text{đ}Q^2_{rev}(s)}{T^2(s)}, \tag{29}$$

so since the integrals depend only on their endpoints

$$\int_A^B \frac{\text{đ}Q_{rev}(s)}{T(s)} = S_B - S_A, \tag{30}$$

where we've just defined a quantity we'll call "entropy[2]" we have only defined it up to a constant of integration, of course. For a reversible process we can now compute the heat from $\text{đ}Q_{rev} = TdS$

2. The 1st Law, revisited For a reversible transformation that means we can now write, from $dE = \text{đ}W + \text{đ}Q_{rev}$,

$$dE = \sum_i J_i dx_i + TdS. \tag{31}$$

No big deal, just the most important expression in thermodynamics... note we've answered one of our questions: the thing we're calling entropy is an extensive generalized "displacement" with T as it's conjugate generalized force.

---

[2] "We might call S the *transformational content* of the body, just as we termed the magnitude U its *thermal and ergonal content*. But as I hold it to be better to borrow terms for important magnitudes from the ancient languages, so that they may be adopted unchanged in all modern languages, I propose to call the magnitude S the *entropy* of the body, from the Greek word $\tau\rho o\pi\eta$, transformation. I have intentionally formed the word *entropy* sa as to be as similar as possible to the word *energy*; for the two magnitudes to be denoted by these words are so nearly allied in their physical meanings, that a certain similarity in designation appears to be desirable." – Clausius, Ninth Memoir, *On several convenient forms of the fundamental equations of the mechanical theory of heat.* For all that talk of borrowing terms for important magnitudes from ancient languages, note that Clausius tried to name the unit of entropy "the Clausius," a calorie per degree Celsius.

3. Entropy increases for irreversible transformations Suppose we make an irreversible change as we go from state A to B, but then complete the cycle by making a reversible transformation from B back to A. Clausius tells us that

$$\int_A^B \frac{\text{đ}Q}{T} + \int_B^A \frac{\text{đ}Q_{rev}}{T} \leq 0 \Rightarrow \int_A^B \frac{\text{đ}Q}{T} \leq S_B - S_A, \tag{32}$$

which tells us that, in differential form, $\text{đ}Q \leq TdS$ for any transformation. For an adiabatic process, with $\text{đ}Q = 0$, we've just learned that $dS \geq 0$. As a system approaches equilibrium, apparently the arrow of time points in the direction of increasing entropy, since changes in a system's internal state can only increase S.

4. How many independent variables do I need to describe an equilibrium system? From Eq. 31 we see that if there n ways of doing mechanical work to a system (the n pairs $\{J_i, x_i\}$), then we need $n+1$ independent coordinates to describe equilibrium systems (i.e., if you know E then then J's and x's are connected. This gives us freedom we'll exploit later in defining different ensembles, etc. For example, suppose we choose as our coordinates E and all of the displacements x_i; Eq. 31 gives us the relations

$$\left.\frac{\partial S}{\partial E}\right|_{\{x_i\}} = \frac{1}{T}. \qquad \left.\frac{-\partial S}{\partial x_i}\right|_{\{E, x_{j\neq i}\}} = \frac{J_i}{T}. \tag{33}$$

0.6 3^{rd} Law

I said last class I wouldn't discuss this next – I've changed my mind, but we'll be very brief! We know from the second law how to compute the difference in entropy between two state points at the same temperature: we make sure we perform operations reversibly, and then calculate $\Delta S = \int \text{đ}Q_{rev}/T$.

0.6.1 Nernst-Simon statement of the third law

The change in entropy associated with a system undergoing a reversible, isothermal process approaches zero as the temperature approaches 0K:

$$\lim_{T\to 0} \Delta S(T) \to 0. \tag{34}$$

Stronger Nernst statement

The entropy of all systems at absolute zero is a universal constant, which we will define to be the zero point of the entropy scale: $\lim_{T\to 0} S(\boldsymbol{X}, \boldsymbol{T}) = 0$.

0.6.2 Consequences

1. Vanishing of entropy derivatives Since $\lim_{T\to 0} S(\vec{r}_{X,T}) = 0$ for all $\boldsymbol{x}$, we must have that

$$\lim_{T\to 0} \left.\frac{\partial S}{\partial x_i}\right|_T = 0 \tag{35}$$

2. Vanishing of thermal expansivities As $T \to 0$, we must have

$$\alpha_J = \frac{1}{x}\left.\frac{\partial x}{\partial T}\right|_J = \frac{1}{x}\left.\frac{\partial S}{\partial J}\right|_T . \tag{36}$$

That second equality follows straightforwardly from a Maxwell relation, which we will go through in the next section...

3. Vanishing of heat capacities As $T \to 0$, we can write

$$S(\boldsymbol{X}, T) - S(\boldsymbol{X}, 0) = \int_0^T \frac{C_{\boldsymbol{x}}(T')}{T'} dT', \tag{37}$$

but that integral diverges unless

$$\lim_{T\to 0} C_{\boldsymbol{x}}(T) \to 0. \tag{38}$$

Unattainability of absolute zero in a finite number of steps Loosely, suppose we are cooling a system by systematically reducing some conjugate force (e.g., cooling a gas by adiabatic reductions in pressure). The 3rd Law says that the functions $S(T)$ for different pressures must all merge at $T = 0$, so each step to lower T must involve progressively smaller changes.

0.6.3 Brief discussion

Note that the "vanishing heat capacity" consequence above was (basically) Nernst's original formulation of the 3rd law! Much disagreement ensued, and the third law's validity and proper framing was debated hotly contested. Why? Because the 3rd law is very different in character from the other laws of thermodynamics!

Microscopic origin of the laws We framed the last few lectures as "Thermodynamics is a phenomenological theory: treat various substances as black boxes and try to deduce a mathematical framework from observations." But *we know what's inside the black box (classical mechanics, quantum mechanics)!*

1. 1st law: Conservation of energy (and heat is a form of energy)
2. 0th and 2nd laws: "irreversible approach to equilibrium." Doesn't immediately seem to have an analog in microscopic equations of motions, but we'll try to derive it later as a consequence of $N \gg 1$.

3. 3rd law: We'll soon see statistical mechanical expressions like $S = k \ln g$, where g is a measure of the degeneracy of states: $S \to 0 \Rightarrow g = \mathcal{O}(1)$ as $T \to 0$. In classical mechanics, this is simply not true! Just think of an ideal gas!. But, as $T \to 0$, CM is not appropriate. It is hardly surprising, then, that a law whose validity actually rests on quantum mechanics was not well-understood or properly justified before QM itself was.

0.7 Various thermodynamic potentials (Appendix H of Pathria)

Mechanical equilibrium occurs at a minimum of a potential energy e.g., the mechanical equilibrium of a mass between springs, etc. etc.. Thermal equilibrium similarly occurs at the extremum of an appropriately defined thermodynamic potential. For example, in our discussion of Clausius' Theorem (Sec. 0.5.2) we found that the entropy of an adiabatically isolated system increases after any change until it reaches a maximum in equilibrium. But what about systems that are not adiabatically isolated? Or systems which are subject to mechanical work? In this section we will define a handful of thermodynamic potentials that are applicable.

Analogy with a mechanical system Briefly, suppose we have a mass on a spring connected to a fixed wall, and let x be the deviation of the mass's position away from the equilibrium rest length of the spring. We take the potential energy to be $U(x) = kx^2/2$, which is clearly minimized when $x = 0$. What if we apply an external force – what will be the new position of the mass? We could define a net potential energy which encompasses this external work, $H = kx^2/2 - Jx$, and set the variation of this with respect to x to be zero:

$$\frac{\partial H}{\partial x} = 0 \Rightarrow x_{eq} = J/k, \quad \text{and} \quad H_{eq} = \frac{-J^2}{2K}. \tag{39}$$

0.7.1 Enthalpy

What if the system is still adiabatically isolated ($đQ = 0$), but comes to equilibrium under a constant external force? We define *enthalpy*, by analogy with the mechanical example above, as

$$H = E - \boldsymbol{J} \cdot \boldsymbol{x}. \tag{40}$$

Variations in this quantity are given by

$$dH = dE - d(\boldsymbol{J} \cdot \boldsymbol{x}) = TdS + \boldsymbol{J} \cdot d\boldsymbol{x} - \boldsymbol{x} \cdot d\boldsymbol{J} - \boldsymbol{J} \cdot d\boldsymbol{x} = TdS - \boldsymbol{x} \cdot d\boldsymbol{J}. \tag{41}$$

Note that in general, at constant $\boldsymbol{J}$ the work added to the system is $đW \leq \boldsymbol{J} \cdot \delta\boldsymbol{x}$ (where equality occurs for reversible processes), so by the first law and making use of $đQ = 0$ we have $dE \leq \boldsymbol{J} \cdot d\boldsymbol{x}$, which means that $\delta H \leq 0$ as a system approaches equilibrium.

0.7.2 Helmholtz Free energy

What if the system is undergoing an isothermal (constant T) transformation in the absence of mechanical work ($đW = 0$)? We define the Helmholtz free energy

$$F = E - TS, \tag{42}$$

which has variations given by

$$dF = dE - d(TS) = TdS + \boldsymbol{J} \cdot d\boldsymbol{x} - SdT - TdS = -SdT + \boldsymbol{J} \cdot d\boldsymbol{x}. \tag{43}$$

Note that Clausius' theorem said that at constant T the heat added to the system is constrained by $đQ \leq TdS$, so making us of $đW = 0$ we have $dE = đQ \leq TdS$, so $\delta F \leq 0$.

0.7.3 Gibbs Free Energy

What if the system is undergoing an isothermal transformation in the presence of mechanical work done at constant external force? We define the Gibbs free energy by

$$G = E - TS - \boldsymbol{J} \cdot \boldsymbol{x}, \tag{44}$$

which has variations given by

$$dG = dE - d(TS) - d(\boldsymbol{J} \cdot \boldsymbol{x}) = \cdots = -SdT - \boldsymbol{x} \cdot d\boldsymbol{J}. \tag{45}$$

Note that in this case, we have both $đW \leq \boldsymbol{J} \cdot \delta\boldsymbol{x}$ and $đQ \leq TdS$, so $\delta G \leq 0$

0.7.4 Grand Potential

Traditionally "chemical work" is treated separately from mechanical work... for chemical equilibrium in the case of no mechanical work, we define the Grand potential by

$$\mathcal{G} = E - TS - \boldsymbol{\mu} \cdot \boldsymbol{N}, \tag{46}$$

where $\boldsymbol{N}$ refer to the number of particles of different chemical species, and μ refers to the chemical potential for each of them. Variations in $\mathcal{G}$ satisfy

$$d\mathcal{G} = -SdT + \boldsymbol{J} \cdot d\boldsymbol{x} - \boldsymbol{N} \cdot d\boldsymbol{\mu}. \tag{47}$$

0.7.5 Changing variables

In the last several subsections we've seen how to use Legendre transformations to move between different natural variables depending on the physical situation we find ourselves in. So, for instance, for adiabatically isolated systems with constant external force, we look at the enthalpy, which has natural variables $H(\boldsymbol{J}, S)$; for isothermal transformations with no external work we look at the Helmholtz free energy, which has natural variables $F(\boldsymbol{x}, T)$, etc. The equilibrium conjugate variables can then by found by partial differentiation. For instance, the equilibrium force and entropy can be found from F by

$$J_i = \left.\frac{\partial F}{\partial x_i}\right|_{T,x_{j\neq i}} \quad \text{and} \quad S = -\left.\frac{\partial F}{\partial T}\right|_{\boldsymbol{x}}. \tag{48}$$

Are there any limits on the manipulations we can perform here? For each set of conjugate force/displacement variables, can we always transform to choose whatever we want? *No.*. Let's see why

0.8 Two bits of math!

0.8.1 *Extensivity* (and Gibbs-Duhem)

Let's look at the differential for E, including chemical work:

$$dE = TdS + \boldsymbol{J}\cdot d\boldsymbol{x} + \mu\cdot d\boldsymbol{N}. \tag{49}$$

In general the extensive quantities are proportional to the size of the system, which we can write mathematically as

$$E(\lambda S, \lambda\boldsymbol{x}, \lambda\boldsymbol{N}) = \lambda(S, \boldsymbol{x}, \boldsymbol{N}). \tag{50}$$

Please note that this is not a requirement, nor does it have the same footing as the rest of the laws of thermodynamics; it is simply a statement about the behavior of "most things." Let's take the above and differentiate with respect to λ and then evaluate at $\lambda = 1$. This gives

$$\left.\frac{\partial E}{\partial S}\right|_{\boldsymbol{x},\boldsymbol{N}} S + \left.\frac{\partial E}{\partial x_i}\right|_{S,x_{j\neq i},\boldsymbol{N}} x_i + \left.\frac{\partial E}{\partial N_\alpha}\right|_{S,\boldsymbol{x},N_{\beta\neq\alpha}} N_\alpha = E(S,\boldsymbol{x},\boldsymbol{N}). \tag{51}$$

Note that the partial derivates here are (in order) T, J_i, and μ_α. This leads to what some people write as the fundamental equation of thermodynamics:

$$E = TS + \boldsymbol{J}\cdot\boldsymbol{x} + \mu\cdot\boldsymbol{N}. \tag{52}$$

Combining equations 49 and 52 lead to a constraint on allowed variations of the intensive coordinates:

$$SdT + \boldsymbol{x}\cdot d\boldsymbol{J} + \boldsymbol{N}\cdot d\mu = 0, \tag{53}$$

which is the *Gibbs-Duhem relation.* Again, this is valid for *extensive* systems, as defined by Eq. 50. Also, this answers the question at the end of the last section: you cannot (usefully) transform to a potential where the natural coordinates are all intensive, because these intensive coordinates are not all independent.

0.8.2 Maxwell relations

"Maxwell relations" follow from combining thermodynamic relationships with the basic properties of partial derivatives. If f, x, and y are mutually related, we can write

$$df(x,y) = \left.\frac{\partial f}{\partial x}\right|_y dx + \left.\frac{\partial f}{\partial y}\right|_x dy, \tag{54}$$

and we will then combine this with symmetry of second derivations,

$$\frac{\partial}{\partial x}\frac{\partial f}{\partial y} = \frac{\partial}{\partial y}\frac{\partial f}{\partial x}, \tag{55}$$

to related various thermodynamic derivatives.

Example

For instance, let's start with $dE = TdS + J_i dx_i$. We know, mathematically, that we can immediately write

$$T = \left.\frac{\partial E}{\partial S}\right|_{\boldsymbol{x}} \quad \text{and} \quad J_i = \left.\frac{\partial E}{\partial x_i}\right|_{S, x_{j\neq i}}. \tag{56}$$

We can take the mixed derivatives and discover a relationship:

$$\frac{\partial^2 E}{\partial S \partial x_i} = \left.\frac{\partial T}{\partial x_i}\right|_S = \left.\frac{\partial J_i}{\partial S}\right|_{\boldsymbol{x}}, \tag{57}$$

where we might call the latter equality a Maxwell relation.

Strategy for deriving Maxwell relations

There are several tricks to remembering how to rapidly find the Maxwell relation relevant to a particular expression. In the homework you will go through a method using Jacobian matrices.

Logically, though, it's not so hard to always construct them on the fly. Suppose someone asks you to find a Maxwell relation for

$$\left.\frac{\partial A}{\partial B}\right|_C. \tag{58}$$

We'll do the following: (1) write down the fundamental expression for dE, (2) transform it so that A will appear in a first derivative and B and C are differentials, and (3) profit.

Worked example:

I want to know $(\partial\mu/\partial P)|_T$ for an ideal gas.

Step 1 We write

$$dE = TdS - PdV + \mu dN. \tag{59}$$

Step 2 We note that μ is already in a position to appear in first derivatives. Moving on,

$$\begin{aligned} d(E - PV) &= TdS - VdP + \mu dN & (60) \\ d(E - PV - ST) &= -TdS - VdP + \mu dN. & (61) \end{aligned}$$

We did not really care what the name of $(E - PV - ST)$ was, let's just call it Y. Clearly

$$\mu = \left.\frac{\partial Y}{\partial N}\right|_{T,P} \quad \text{and} \quad V = \left.\frac{\partial Y}{\partial P}\right|_{T,N}, \tag{62}$$

so

$$\left.\frac{\partial \mu}{\partial P}\right|_{T} = \left.\frac{\partial V}{\partial N}\right|_{T}. \tag{63}$$

Step 3 We're done.

Chapter 1

Probability

In the last chapter we treated thermodynamics as a phenomenological theory, building up a consistent mathematical framework that expressed the consequences of experimental observations of various black box systems. Ultimately we will want to see how these properties arise from the microscopic rules that real systems evolve according to; to do so we will be expressing the statistical consequences of having large numbers of interacting units. In this brief chapter we will cover the parts of probability theory that we will be using. Since much of this is standard definitional stuff, we will also be sure to cover calculational methods and tricks that will serve us down the road.

1.1 A funny observation

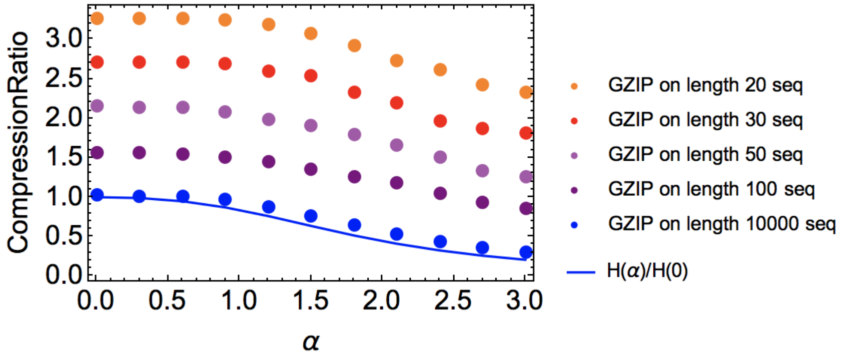


Figure 1.1: **Compressing randomly generated strings** The average ratio of the length of a compressed random string to its original length is bounded by the entropy of the probability distribution used to generate it.

The other day[1] I was generating random strings of characters in the English language... just things like "dynbaggdaaejfgoafkoadbdbadaenncadykabkfaaapkabpgabgciicecyktvaenoaeacgjc" or "tgxcabgljndjaankbafbohoewjmfvracaevwfmdmabbagtbbaabnhabfjvpdcbacfycalsjac", etc.

---

[1] for fun, you know?

I started compressing these strings using an off-the-shelf algorithm on my computer, *gzip*, and compared the length of the compressed string to the length of the original string. I repeated this a bunch of times, for sequences of different length, and for different probability distributions from which I was generating my random strings (if you're curious, they were power-law distributions parameterized by a decay strength α). The results are in Fig. 1.1, and I was shocked! Apparently, as I started compressing longer and longer strings, the amount I was able to compress them was bounded by what I've labeled $H(\alpha)$ in the figure – the entropy of the probability distribution! How is entropy – which so far seems like a thermodynamic concept – related to computation and information compression? Let's find out!

1.2 Basic Definitions

Random variable: A random variable x is a measurable variable described by a set of possible outcomes, S. This could be a discrete set, as for a coin $S_{coin} = \{\text{heads, tails}\}$, or a continuous range, as for a particle's velocity $S_{v_x} = \{-\infty \leq v_x \leq \infty\}$. We call each element of such a set as an *event* $E \subset S$

Probability We will assign a value, called the *probability* to each event, denoted $p(E)$, which has the following properties:

1. *positivity:* $p(E) \geq 0$
2. *additivity:* $p(A \text{ or } B) = p(A) + p(B)$ if A and B are distinct.
3. *normalization:* $p(S) = 1$.

We're not mathematicians, so we won't be starting from here and proving stuff. Rather, you may wonder, "how will we determine $p(E)$?" By one of two ways:

1. *Objectively:* (experimentally, frequentist) $p(E)$ is the frequency of outcome in many trials:
   $p(E) = \lim_{N\to\infty} N_E/N$
2. *Subjectively:* (theoretically, Bayesian) Based on our uncertainty among all outcomes. We'll see this is what we'll really use in stat phys! We'll formalize this later.

1.3 Properties of single random variables

Let's focus on random variables which are continuous and real-valued (the specialization to discrete ones is straightforward; we'll see an example in the next section).

Cumulative probability function: $P(x) = \text{probability}(E \subset [-\infty, x])$. This must be a monotonically increasing function, with $P(-\infty) = 0$ and $P(\infty) = 1$.

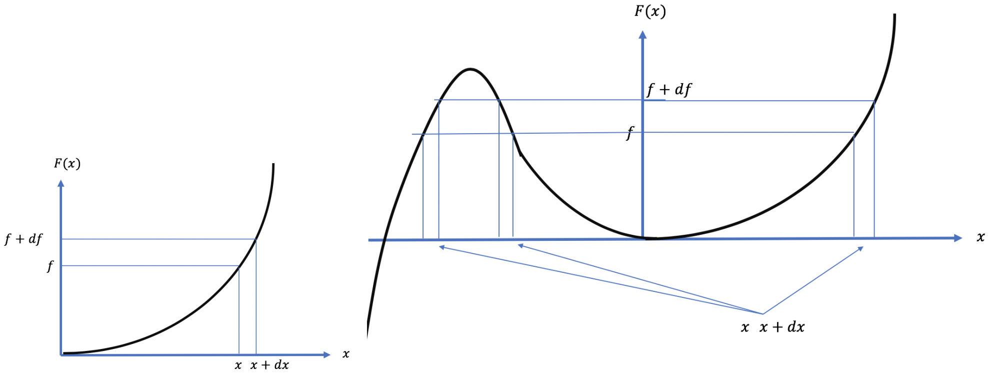


Figure 1.2: **Changing variables from x to $F(x)$**

Probability density function (pdf): $p(x) = dP(x)/dx$, so that $p(x)dx = \text{probability}(E \subset [x, x+dx])$. The pdf must be everywhere positive, and normalized so that $\int_{-\infty}^{\infty} p(x)dx = 1$. *Note:* $P(x)$ is dimensionless, which means $p(x)$ has dimensions of $[x^{-1}]$.

Expectation value $\langle F(x) \rangle = \int dx p(x) F(x)$

Change of variables Suppose x is a random variable and we have a function of it, $F(x)$. We can regard $F(x)$ itself as a new random variable. The probability that we find F in some interval between f and $f + df$ is written $p_F(f)df$, and it equals[2] $p_X(x)dx$ (see Fig. 1.2). Or, rather, we have that $|p_F(f)df| = |p_X(x)dx|$, since the probability in some differential *area* must should be invariant under a change of variables.

But we need to add up as many pieces of x contribute. Thus, we first solve $F(x) = f$, and call the set of solutions x_i. We can then write:

$$p_F(f) = \sum_i p(x_i) \left| \frac{dx}{df} \right|_{x_i} . \tag{1.1}$$

In standard texts you will usually see the $\frac{dx}{df}$ written in terms of inverse functions, but the meaning is clear (and it is the slope). Note that the factors of $\left|\frac{dx}{df}\right|$ are the *Jacobians* used to change variables.

Trivial example Suppose we have a random variable x with pdf $p_X(x) = 3x^2$ defined in the interval $0 < x < 1$, and we look at a new random variable $Y = X^2$.

---

[2] One can go back to the definition of the cumulative probability function to show this is true for monotonically increasing or decreasing functions; it is more work to show it for the sorts of piecewise monotonic functions we might want to consider

This is easily invertible in the range, and we can write $x(y) = \sqrt{y}$, and $dx/dy = y^{-1/2}/2$. Thus

$$p_Y(y) = p_X(x) \left| \frac{1}{2\sqrt{y}} \right| = \frac{3}{2} (\sqrt{y})^2 \frac{1}{\sqrt{y}} = \frac{3}{2}\sqrt{y}, \tag{1.2}$$

defined in the range $0 < y < 1$.

2-valued example Suppose instead that we have a random variable x where

$$p(x) = \frac{\lambda}{2} \exp\left(-\lambda |x|\right),$$

defined for any x on the real line. We want to know the probability density function for the random variable $F(x) = x^2$. There are, by inspection, two solutions to $F(x) = f$ (when f is positive!), and they are $x = \pm\sqrt{f}$. The derivatives we need are $|dx/df| = |\pm \frac{1}{2\sqrt{f}}|$. Thus, we have:

$$p_F(f) = \frac{\lambda}{2} \exp(-\lambda \left|\sqrt{f}\right|) \left|\frac{1}{2\sqrt{f}}\right| + \frac{\lambda}{2} \exp(-\lambda \left|-\sqrt{f}\right|) \left|\frac{-1}{2\sqrt{f}}\right| = \frac{\lambda \exp\left(-\lambda\sqrt{f}\right)}{2\sqrt{f}},$$

for any $f > 0$ (and $p_F(f) = 0$ for $f < 0$).

Moments We define the nth moment to be $m_n \equiv \langle x^n \rangle \int dx\ x^n p(x)$

Characteristic function The characteristic function is simply the Fourier transform of the pdf:

$$\tilde{p}(k) = \langle e^{-ikx} \rangle = \int dx\ p(x) e^{-ikx}. \tag{1.3}$$

Likewise, if you know the characteristic function, the pdf is the inverse FT:

$$p(x) = \frac{1}{2\pi} \int dk\ \tilde{p}(k) e^{ikx}. \tag{1.4}$$

We can use the characteristic function to generate the moments of the pdf (assuming the moments exist!). Making that assumption, let's expand the exponential in the definition, $\exp(-ikx) = \sum_{n=0}^{\infty} \frac{(-ik)^n}{n!} x^n$, so that

$$\tilde{p}(k) = \langle \sum_{n=0}^{\infty} \frac{(-ik)^n}{n!} x^n \rangle = \sum_{n=0}^{\infty} \frac{(-ik)^n}{n!} \langle x^n \rangle, \tag{1.5}$$

so that *if* you can expand the characteristic function in powers of k, the coefficients of the expansion give you the moments (up to some factor of $\pm n!$). Note that in general the moment generating function is defined differently from the characteristic function for precisely such a reason: the MGF may not always exist. We can do manipulations of this type of expression to get, e.g., relative moments pretty easily:

$$e^{ikx_0} \tilde{p}(k) = \langle \exp(-ik(x - x_0)) \rangle = \sum_{n=0}^{\infty} \frac{(-ik)^n}{n!} \langle (x - x_0)^n \rangle. \tag{1.6}$$

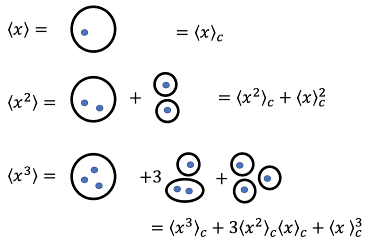


Figure 1.3: **Graphical expansion of the first three moments** s

Cumulant genenerating function Let's define the *cumulant generating function* to be the log of the characteristic function, and expand it in what we will define to be the *cumulants.* Note: some people would call this the "second characteristic function." The difference, again, is whether we end up with something which is always defined regardless of whether the moments themselves are well-defined.

$$\ln \tilde{p}(k) \equiv \sum_{n=1}^{\infty} \frac{(-ik)^n}{n!} \langle x^n \rangle_c. \tag{1.7}$$

$$\ln \tilde{p}(k) = \ln\left(1 + \sum_{n=1}^{\infty} \frac{(-ik)^n}{n!} \langle x^n \rangle\right) \tag{1.8}$$

$$= \left(\sum_{n=1}^{\infty} \frac{(-ik)^n}{n!} \langle x^n \rangle\right) - \frac{1}{2}\left(\sum_{n=1}^{\infty} \frac{(-ik)^n}{n!} \langle x^n \rangle\right)^2 + \frac{1}{3}\left(\sum_{n=1}^{\infty} \frac{(-ik)^n}{n!} \langle x^n \rangle\right)^3 + \cdots \tag{1.9}$$

We can then relate the cumulants to the moments by comparing Eqs. 1.7 and 1.8 and matching terms of order k^n. For example, the first two are pretty easy to see:

$$\langle x\rangle_c = \langle x\rangle \tag{1.10}$$

$$\langle x^2\rangle_c = \langle x^2\rangle - \langle x\rangle^2 = \left\langle (x - \langle x\rangle)^2 \right\rangle \tag{1.11}$$

$$\tag{1.12}$$

Why do we care? The cumulants are, in a sense, a collection of most important ways of describing a distribution (mean, variance, skewness, kurtosis, etc.).

Graphical connection between moments and cumulants The combinatorics of coefficient matching above might seem difficult to parse; there is a fun graphical way of remembering how to connect moments and cumulants. Notationally, let's represent the nth cumulant as a bag with n points inside of it (conveniently, since $\langle x\rangle_c = \langle x\rangle$, a bag with one point is the same as one point on its own). Then, the mth *moment* can be graphically expressed as the sum of all ways of distributing m points among bags. See Fig. 1.3.

Slightly more formally, one might say you represent the nth cumulant as a connected cluster of points, and obtain the mth moment by adding together all subdivisions of m points into groupings of connected or disconnected clusters. The contribution of each subdivision to the sum is then the product of the connected cumulants it represents. This graphical connection between moments and cumulants is the basis for several diagrammatic computations (in stat mech, in field theory,...) Will we see it again in this class? Stay tuned...

1.4 Important distributions

1.4.1 Binomial distribution

Given a discrete random variable with two outcomes, which occur with probability p_A and $p_B = 1 - p_A$, the binomial distribution gives the probability that event A occurs exactly N_A times out of N trials. It is equal to

$$P_N(N_A) = \binom{N}{N_A} p_A^{N_A} p_B^{N-N_A}, \qquad \binom{N}{N_A} = \frac{N!}{N_A!(N-N_A)!}. \tag{1.13}$$

The characteristic function for the discrete distribution is

$$\tilde{p}_N(k) = \langle e^{-ikN_A} \rangle = \sum_{N_A=0}^{N} \frac{N!}{N_A!(N-N_A)!} p_A^{N_A} p_B^{N-N_A} e^{-ikN_A} = \left(p_A e^{-ik} + p_B \right)^N. \tag{1.14}$$

This has the properties that we can easily relate the cumulant generating function for the N-trial case to that of the 1-trial case:

$$\ln \tilde{p}_N(k) = N \ln \left(P_A e^{-ik} + p_B \right) = N \ln \tilde{p}_1(k). \tag{1.15}$$

For a single trial, N_A can only be either zero or one, which means that we must have $\langle N_A^m \rangle = p_A$ for all powers m. Combining this property of the moments with the above feature of the cumulants, we learn that the cumulants for the N-trial case are

$$\langle N_A \rangle_c = N p_A, \quad \langle N_A^2 \rangle_c = N \left(p_A - p_A^2 \right) = N p_A p_B, \tag{1.16}$$

and higher order cumulants can be easily calculated. We'll see that this type of feature – where there is a trivial relation between an independent thing repeated N times and the case of an individual trial – will be of great use as we build up statistical mechanics.

1.4.2 Poisson distribution

We'll get at the Poisson distribution, a continuous pdf, relating it to the binomial distribution. Consider a process in time where two properties hold. First, the probability of observing (exactly) one event in the interval $[t, t+dt]$ is proportional to dt in the limit $dt \to 0$. Second, suppose the probability of observing an event in different intervals is uncorrelated. Example: radioactive decay. Then, the *Poisson distribution* is the probability of observing exactly M events in the interval T.

We get the details of the distribution by imagining dividing up the interval T into many segments of length dt, say $N = T/dt \gg 1$ such that dt is so small the probability of observing more than one event is negligible. So, in each segment we have an event occurring with probability $p = \alpha dt$ and no event occurring with probability $q = 1 - p$. From our expression for the binomial distribution, we immediately know the characteristic function for this process:

$$\tilde{p}(k) = \left(p e^{-ik} + q \right)^N = \lim_{dt \to 0} \left(1 + \alpha dt((e^{-ik} - 1) \right)^{T/dt} = \exp\left(\alpha (e^{-ik} - 1) T \right), \tag{1.17}$$

where the last equality is an example of the famous Euler limit formula. Knowing the characteristic function, we can take the inverse Fourier transform to get the pdf:

$$p(x) = \int_{-\infty}^{\infty} \frac{dk}{2\pi} \exp\left(\alpha(e^{-ik} - 1)T + ikx\right). \tag{1.18}$$

This can be solved You, the reader should verify this! by expanding the exponential, and using

$$\int_{-\infty}^{\infty} \frac{dk}{2\pi} e^{-ik(x-M)} = \delta(x - M)$$

to get the probability of M events in a time T for a process characterized by α as

$$p_{\alpha T}(M) = e^{-\alpha T} \frac{(\alpha T)^M}{M!}. \tag{1.19}$$

Additionally, the cumulants can be read off of the expansion of the log characteristic function

$$\ln \tilde{p}_{\alpha T}(k) = \alpha T(e^{-ik} - 1) = \alpha T \sum_{n=1}^{\infty} \frac{(-ik)^n}{n!} \Rightarrow \langle M^n \rangle_c = \alpha T. \tag{1.20}$$

That is, while the binomial distribution has the property that every moment is the same, the Poisson distribution has the property that every *cumulant* is the same.

1.4.3 Gaussian distribution

We will definitely, definitely use this, you know? Define the gaussian pdf as

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x - \lambda)^2}{2\sigma^2}\right). \tag{1.21}$$

The characteristic function is compute by the usual means of completing the square inside the integral, a trick I believe we all know:

$$\begin{aligned}
\tilde{p}(k) &= \int \frac{dx}{\sqrt{2\pi\sigma^2}} \exp\left(-ikx - \frac{(x-\lambda)^2}{2\sigma^2}\right) && (1.22)\\
&= e^{-ik\lambda} \int \frac{dy}{\sqrt{2\pi\sigma^2}} \exp\left(-iky - \frac{y^2}{2\sigma^2} + \frac{k^2\sigma^2}{2} - \frac{k^2\sigma^2}{2}\right), \quad \text{for } y = x - \lambda && (1.23)\\
&= e^{-ik\lambda - \frac{k^2\sigma^2}{2}} \int \frac{dz}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-z^2}{2\sigma^2}\right), \quad \text{for } z = y + ik\sigma^2 && (1.24)\\
&= \exp\left(-ik\lambda - \frac{k^2\sigma^2}{2}\right) && (1.25)
\end{aligned}$$

A manipulation that shows that the Fourier transform of a Gaussian is, itself, a Gaussian. The cumulants of this are easily identified:

$$\ln \tilde{p}(k) = -ik\lambda - \frac{k^2\sigma^2}{2}, \tag{1.26}$$

immediately showing that

$$\langle x\rangle_c = \lambda, \quad \langle x^2\rangle_c = \sigma^2, \quad \langle x^{n>2}\rangle_c = 0. \tag{1.27}$$

So, the Gaussian is completely specified by its first two cumulants, and all moments involve only products of one- and two-point clusters.

1.5 Properties of multiple random variables

Joint probability density function We define, by analogy, the joint pdf $p(x_1, x_2, \ldots, x_N)$ as

$$p(\boldsymbol{x}) = \lim_{\{dx_i \to 0\}} \frac{\text{prob. of outcome in } \{(x_1, x_1+dx_1), \ldots, (x_N, x_N+dx_N)\}}{dx_1 dx_2 \cdots dx_N}. \tag{1.28}$$

The normalization of the joint PDF is

$$p_{\boldsymbol{x}}(S) = 1 = \int d^N\boldsymbol{x} p(\boldsymbol{x}), \tag{1.29}$$

and *iff* the N random variables are *independent*, then the joint pdf simplifies to the product of the individual probability density functions:

$$p(\boldsymbol{x}) = \prod_{i=1}^{N} p_i(x_i)). \tag{1.30}$$

Joint characteristic function is just the N-dimensional Fourier transform:

$$\tilde{p}(\boldsymbol{k}) = \langle \exp(-i\boldsymbol{k}\cdot\boldsymbol{x}\rangle = \int \left(\prod_i dx_i e^{-ik_i x_i}\right) p(x_1, \ldots, x_N). \tag{1.31}$$

Joint moments and cumulants Are defined perfectly analogously with the moments and cumulants of single random variable distributions. Recall that last lecture we talked about moments as related to the coefficient of the relevant power of k... more generally we can express these as the following derivatives:

$$\langle x_1^{m_1} x_2^{m_2} \cdots x_N^{m_N}\rangle = \left[\frac{\partial}{\partial(-ik_1)}\right]^{m_1} \cdots \left[\frac{\partial}{\partial(-ik_N)}\right]^{m_N} \tilde{p}(\boldsymbol{k})|_{\boldsymbol{k}=0} \tag{1.32}$$

$$\langle x_1^{m_1} x_2^{m_2} \cdots x_N^{m_N}\rangle_c = \left[\frac{\partial}{\partial(-ik_1)}\right]^{m_1} \cdots \left[\frac{\partial}{\partial(-ik_N)}\right]^{m_N} \ln \tilde{p}(\boldsymbol{k})|_{\boldsymbol{k}=0}\,. \tag{1.33}$$

As a simple – but perhaps the most important – example, the "co-variance" between two random variables is

$$\langle x_1 x_2\rangle_c = \langle x_1 x_2\rangle - \langle x_1\rangle\langle x_2\rangle. \tag{1.34}$$

The graphical expansion we wrote earlier still applies; one just has to label the points by the corresponding variables. See Fig. 1.4.

Unconditional probability: The unconditional PDF describes the PDF for a subset of random variables independent of what the others are doing:

$$p(x_1, \ldots, x_m) = \int \left(\prod_{i=m+1}^{N} dx_i \right) p(x_1, \ldots, x_N). \tag{1.35}$$

For example, a gas particle would generically have a PDF over both its position and velocity, $p(\boldsymbol{x}, \boldsymbol{v})$, but we might only care about the distribution of positions, so we would integrate out the velocities:

$$p(\boldsymbol{x}) = \int d^3\boldsymbol{v} p(\boldsymbol{x}, \boldsymbol{v})$$

Conditional probability: The *conditional* PDF describes the behavior of a subset of the random variables *given specified values for the other random variables.* Following the above example, suppose we are interested in the conditional probability of a velocity *given* a position, denoted $p(\boldsymbol{v}|\boldsymbol{x})$. This should be proportional to the full joint PDF:

$$p(\boldsymbol{v}|\boldsymbol{x}) = \frac{p(\boldsymbol{x}, \boldsymbol{v})}{\mathcal{A}},$$

where the constant of proportionality is just the probability of having that value of position in the first place:

$$\mathcal{A} = \int d^3\boldsymbol{v} p(\boldsymbol{x}, \boldsymbol{v}) = p(\boldsymbol{x}).$$

This is given by **Bayes' Theorem**, and in general we write the connection between conditional and unconditional PDFs as

$$p(x_1, \ldots, x_s | x_{s+1}, \ldots, x_N) = \frac{p(x_1, \ldots, x_N)}{p(x_{s+1}, \ldots, x_N)}. \tag{1.36}$$

We see that in the case of *independent* random variables, the conditional probability is the same as the unconditional probability. Right?

Joint Gaussian distribution You might have thought the natural generalization of Eq. 1.21 was

$$p(\boldsymbol{x}) = \frac{1}{\sqrt{\prod_{n=1}^{N} 2\pi\sigma_n^2}} \exp\left(\frac{-1}{2} \sum_{n=1}^{N} \frac{-(x_n - \lambda_n)^2}{\sigma_n^2} \right). \tag{1.37}$$

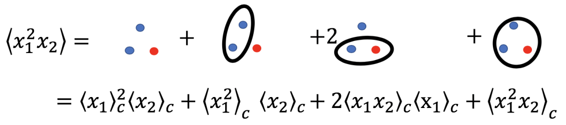


Figure 1.4: **Graphical expansion of a joint moment**

but this neglects the potential for cross-correlations! The most general form is, instead,

$$p(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^N \det \boldsymbol{C}}} \exp\left(\frac{-1}{2}\sum_{n,m=1}^{N}(x_n-\lambda_n)(x_m-\lambda_m)\left(\boldsymbol{C}\right)^{-1}_{nm}\right), \tag{1.38}$$

where the matrix $\boldsymbol{C}$ is symmetric, and for $p(\boldsymbol{x})$ to be a well-defined probability the matrix $\boldsymbol{C}$ must be positive definite. We can write this somewhat more compactly as

$$p(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^N \det \boldsymbol{C}}} \exp\left(\frac{-1}{2}(\boldsymbol{x}-\boldsymbol{\lambda})^T\boldsymbol{C}^{-1}(\boldsymbol{x}-\boldsymbol{\lambda})\right). \tag{1.39}$$

The matrix $\boldsymbol{C}$ is called the covariance matrix. If one goes through and performs the fourier transform on the above joint PDF, one finds

$$\tilde{p}(\boldsymbol{k}) = \exp\left(-i\boldsymbol{k}\cdot\boldsymbol{\lambda} - \frac{1}{2}\boldsymbol{k}^T\boldsymbol{C}\boldsymbol{k}\right), \tag{1.40}$$

or, in index notation,

$$\tilde{p}(\boldsymbol{k}) = \exp\left(-ik_m\lambda_m - \frac{1}{2}k_m C_{mn} k_n\right). \tag{1.41}$$

The latter re-writing lets us immediately read off the joint cumulants of the joint Gaussian distribution:

$$\langle x_m\rangle_c = \lambda_m, \quad \langle x_m x_n\rangle_c = C_{mn}, \tag{1.42}$$

with all higher-order cumulants vanishing.

> Note that there is an important special case when $\boldsymbol{\lambda} = \boldsymbol{0}$. Consider the joint cumulant
>
> $$\langle x_1^{n_1}x_2^{n_2}\cdots x_N^{n_N}\rangle, \tag{1.43}$$
>
> and think about the combinatorics of the graphical expansion we've been discussing.
>
> First, if the sum of the n_i is *odd*, then in the graphical expansion there is no way to avoid a term with an odd-power cumulant, and in this special case of the joint Gaussian distribution with $\boldsymbol{\lambda} = \boldsymbol{0}$, *all such terms are zero*!
>
> Second, if the sum is odd, we know that there will only be contributions from *combinations of covariances*: all even-power cumulants with power greater than two vanish because we are dealing with the joint Gaussian. Thus, the cumulant can be obtained by all ways of summing over pairs of the random variables. For example,
>
> $$\langle x_i x_j x_k x_l\rangle = C_{ij}C_{kl} + C_{ik}C_{jl} + C_{il}C_{jk}, \tag{1.44}$$
>
> where it didn't matter if the i, j, k, l were distinct. For instance:
>
> $$\langle x_1^2 x_2 x_3\rangle = C_{11}C_{23} + 2C_{12}C_{13}. \tag{1.45}$$
>
> This property of the joint Gaussian distribution is sometimes summarized as:
>
> $$\langle x_1^{n_1}x_2^{n_2}\cdots x_N^{n_N}\rangle = \begin{cases} 0 & \text{if } \sum_\alpha n_\alpha \text{ is odd} \\ \text{Sum over all pairwise contractions of covariances} & \text{else} \end{cases} \tag{1.46}$$
>
> In this formulation, we see the analogy of *Wick's Theorem* applied to fields.

1.6 Math of large numbers

We typically think of statistical mechanics being relevant when the number of microscopic degrees of freedom, N, becomes very large; indeed, in the *thermodynamic limit*, $N \to \infty$, a number of mathematical simplifications become available to our analysis of how systems behave.

1.6.1 The Central Limit Theorem

The central limit theorem, which I trust we have all encountered before, is a core engine in allowing us to make precise statements of the sort we encountered in thermodynamics. For instance, we observed that heat flows from hot to cold – not sometimes, or most of the time, but always. If we're going to make probabilistic arguments at the microscopic core, how do we end up with precise, essentially deterministic thermodynamic statements? We'll make the case for the *classical* CLT, by the way, not the Lyapunov or other versions with weaker conditions.

Let's start by considering the sum of N random variables, $X = \sum_{i=1}^{N} x_i$, where the random variables x_i have some joint PDF $p(\boldsymbol{x})$. What is the cumulant generating function of the sum, $\ln \tilde{p}_X(k)$? Well,

$$\ln \tilde{p}_X(k) = \ln \left\langle e^{-ikX} \right\rangle = \ln \left\langle \exp(-ik\sum_{i=1}^{N} x_i \right\rangle = \ln \tilde{p}_x \left(k_1 = k, k_2 = k, \ldots, k_N = k\right), \tag{1.47}$$

That is, it is the same as the log of the *joint characteristic function* of the x_i, but evaluated at the same k. Let's expand each side of the above equation, writing things so we can easily match powers of k:

$$\sum_{n=1}^{\infty} \frac{(-ik)^n}{n!} \langle X^n \rangle_c = (-ik) \sum_{i=1}^{N} \langle x_i \rangle_c + \frac{(-ik)^2}{2!} \sum_{i,j=1}^{N} \langle x_i x_j \rangle_c + \ldots \tag{1.48}$$

Matching terms of order k^n, we see that

$$\langle X \rangle_c = \sum_i \langle x_i \rangle_c, \qquad \langle X^2 \rangle_c = \sum_{i,j} \langle x_i x_j \rangle_c, \ldots \tag{1.49}$$

Now, we specialize to the case of the classical central limit theorem by supposing that the x_i are both *independent*, so that $p(x_1, \ldots, x_N) = p_1(x_1) p_2(x_2) \cdots p_N(x_N)$ and *identically distributed*, i.e., each of the labeled probability distributions $p_i(x_1)$ are the same, so that

$$p(x_1, \ldots, x_N) = \prod_{i=1}^{N} p(x_i). \tag{1.50}$$

This combination of conditions, independent and identically distributed, is often abbreviated *iid*. Now, the fact that the variables are independent means that the cross-correlations in

Eq. 1.49 vanish (do you see why the math tells us this?), so that in the double sum only terms with $i = j$ contribute. Thus, we get that

$$\langle X^m \rangle_c = \sum_{i=1}^{N} \langle x_i^m \rangle_c \tag{1.51}$$

. The condition of identically distributed takes us back to the case we looked at with the binomial distribution: the cumulants of N repeated but independent draws from the same distribution are easily related to the cumulants of the single-random-variable distribution:

$$\langle X^m \rangle_c = \sum_{i=1}^{N} \langle x_i^m \rangle_c = N \langle x^m \rangle_c . \tag{1.52}$$

The (classical) Central Limit Theorem follows directly. Define a new random variable to be

$$y = \frac{X - N \langle x \rangle_c}{\sqrt{N}}, \tag{1.53}$$

and then one computes its cumulants:

$$\langle y \rangle_c = 0, \quad \langle y^2 \rangle_c = \frac{\langle X^2 \rangle_c}{\sqrt{N}} = \langle x^2 \rangle_c , \quad \langle y^m \rangle = \frac{N \langle x^m \rangle_c}{N^{m/2}}. \tag{1.54}$$

In words: as N becomes large, distribution for a sum of random variables with mean μ and variance σ^2 converges to a distribution that itself has a finite mean, a variance that only grows as $\sqrt{N}$, and higher-order cumulants that all decay to zero as $N \to \infty$. Thus, sums of random variables converge to normal distributions, rather ignoring the details of what the original random variables looked like (up to some point). Note that the condition is really on the existence of the moments in question, and a condition on how correlated the variables are allowed to be:

$$\sum_{i_1,\ldots i_m}^{N} \langle x_{i_1} \ldots x_{i_m} \rangle_c \ll \mathcal{O}(N^{m/2}).$$

1.6.2 Adding up exponential quantities

In stat mech we tend to run into (1) *intensive variables* (like T, P, etc.), which are independent of system size ($\mathcal{O}(N^0)$), (2) *extensive variables* (like S, V, etc.), which scale linearly with system size ($\mathcal{O}(N^1)$), and (3) *exponential variables* (like volumes of phase space), which are independent of system size ($\mathcal{O}(V^N) = \mathcal{O}(e^{aN})$. Of course, polynomial dependences, etc., are also possible and sometimes arise (especially in interesting systems). The behavior of adding exponential quantities together makes calculating thermodynamic limits possible

Summing exponentials

Frist, suppose we have a sum of a large number of exponentially large numbers:

$$S = \sum_{i=1}^{\mathcal{N}} \mathcal{E}_i, \tag{1.55}$$

where the terms $0 \leq \mathcal{E}_1 \sim \mathcal{O}(\exp(Na_i))$ and we are summing up $\mathcal{N} \sim \mathcal{O}(N^P)$, a number of terms that grows at most polynomial in N.

Claim We can approximate the entire sum just by the largest term! That is,

$$S \approx \mathcal{E}_{max}. \tag{1.56}$$

proof We mean that claim in a specific sense, as follows. First, it is clear that the we can bound the sum by

$$\mathcal{E}_{max} \leq S \leq \mathcal{N}\mathcal{E}_{max}. \tag{1.57}$$

Now, let's switch to an intensive variable by first taking $\ln S$ and then dividing by N. This gives the bounds:

$$\frac{\mathcal{E}_{max}}{N} \leq \frac{\ln S}{N} \leq \frac{\mathcal{E}_{max}}{N} + \frac{\ln \mathcal{N}}{N}, \tag{1.58}$$

but

$$\frac{\ln \mathcal{N}}{N} = \frac{p \ln N}{N} \tag{1.59}$$

according to our assumption, and this goes to zero as $N \to \infty$. Thus,

$$\lim_{N\to\infty} \frac{\ln S}{N} = \frac{\ln \mathcal{E}_{max}}{N} = a_{max}. \tag{1.60}$$

So, even if the second-largest a_i is only slightly less than the maximum one, upon exponentiation N times it gets completely dominated by the larger term. Think about going from the microcanonical ensemble (where you specified precisely all of the energies, NVE, and are summing over energy levels), or the canonical ensemble (where you do not, NVT); you get the same result because the system behavior is dominated by the most likely energy.

Integrating exponentials – Saddle-point integrations

We generalize the above result to get a simple version of saddle-point integration[3]. We wish to make a similar claim about the integral, $\mathcal{I} = \int dx \exp(N\phi(x))$ being dominated by the place where the function $\phi(x)$ itself is maximized. Well, let's Taylor expand ϕ about its *maximum* x_m (which I emphasize so you remember the first derivative vanishes and the second derivative is negative):

$$\mathcal{I} = \int dx \exp\left(N\phi(x_m) - \frac{N}{2}|\phi''(x_m)|(x-x_m)^2 + \cdots\right) \tag{1.61}$$

---

[3] c.f. the section of Pathria in Chapter 3 which treats more general integrands with integration paths in the complex plane

This term has two types of corrections encoded in that set of $\cdots$. First, of course, there are the higher-order terms in the expansion of the function $\phi(x)$ about its maximum value; these terms lead to a power series in $1/N$. Second, there could be contributions to this sum from additional local maximum. But, by arguments similar to those made in the previous subsection, any such contribution will be completely subdominant! Thus, we truncate the series at quadratic order as above and write

$$\mathcal{I} = e^{N\phi(x_m)} \int dx \exp\left(-\frac{N}{2}|\phi''(x_m)|(x-x_m)^2\right), \tag{1.62}$$

which is just another Gaussian integral, but one missing its normalization factor, so

$$\mathcal{I} = e^{N\phi(x_m)} \sqrt{\frac{2\pi}{N|\phi''(x_m)|}} \quad \Rightarrow \quad \lim_{N\to\infty} \frac{\ln \mathcal{I}}{N} = \phi(x_m). \tag{1.63}$$

Note / example: Stirling's approximation The above machinery can be used to derive Stirling's approximation for the factorial. Start by noting that

$$N! = \int_0^\infty x^N e^{-x}, \tag{1.64}$$

which can be see by starting with $\int_0^\infty \exp(-\alpha x) = 1/\alpha$, and then taking N derivatives and setting $\alpha = 1$. Some rearrangements of the above equation (writing $\phi(x) = \ln x - x/N$), expanding about $x_m = N$, and doing the Gaussian integral gets you to

$$N! = N^N e^{-N}\sqrt{2\pi N}\left(1 + \mathcal{O}\left(\frac{1}{N}\right)\right), \tag{1.65}$$

the log of which is Stirling's formula. Filling in the missing steps should be straightforward, but also an excellent way to make sure you understand the machinery of this method.

1.7 Information Entropy

We end this chapter by thinking about an information-based view of what we mean by entropy, one introduced by Shannon in a groundbreaking 1948 paper[4]. We will discuss the connection between information and entropy, and by thinking about 'unbiased" ways of assigning probabilities, we will formalize the *subjective* procedure of assigning probabilities discussed at the beginning of this chapter.

---

[4] which you can read at this harvard site: http://people.math.harvard.edu/ ctm/home/text/others/shannon/entropy/entropy.pdf. Note that Shannon named the symbol of entropy H, after Boltzmann's H-theorem, which we'll encounter in the very next chapter.

1.7.1 Shannon entropy

We briefly change our focus from thermodynamics and statistical physics to a setting which seems very different: the problem of sending messages over a wire. We begin by imagining a source trying to send us a message from an "alphabet" of k characters, $\{a_1, \ldots, a_k\}$ that have an discrete associated probability distribution $p(a_i)$, $X = \{a_i, p(a_i)\}$. (Think, for instance, of the actual alphabet, where indeed some letters appear more frequently than others in real messages.), where we will assume that the characters are *iid* (In real messages there are, of course, correlations; we neglect them in this idealized setting). With this assumption, the probability that the source sends the n-character message $x = x_1 x_2 \cdots x_n$ is just

$$p(x) = \prod_{i=1}^{n} p(x_i). \tag{1.66}$$

Let's denote the entire ensemble of n-length messages chosen with the assumption that the x_i are iid by X^n.

Compressing messages Suppose the length of the message, n, grows very large. In this setting, is it possible to *compress* the message into a shorter string that conveys the same "information"? As long as $p(a_i)$ is not uniform, then yes! The total number of messages is k^n, but for large n, we expect each character to occur about $n_i = np(a_i) + \mathcal{O}(n^{-1/2})$ times. So the number of *typical* strings, g, is not k^n but rather

$$g = \frac{n!}{\prod_{i=1}^{k}(np(a_i))!}. \tag{1.67}$$

Applying Stirling's formula, we find that

$$\log_2 g \equiv nH(X) \approx -n \sum_{i=1}^{k} p(a_i) \log_2 p(a_i), \tag{1.68}$$

where $H(X)$ is the *Shannon entropy* of the ensemble $X = \{a_i, p(a_i)\}$. If we imagine adopting a code for messages of length n where integers label "typical" messages of such length, a typical n-letter string could be communicated using about $nH(X)$ bits. To be extra explicit, for discrete probability distributions with values $\{p_i\}$ we will be defining the entropy in this way:

$$S = H(X) = -\langle \ln p \rangle = -\sum_i p_i \ln p_i. \tag{1.69}$$

Compressing binary messages Let's briefly show this in more detail for a binary alphabet: each character is either zero with probability p or one with probability $1-p$, so the ensemble X is completely specified by the single value p. Well, for large values of n there are going to be about np zeros and $n(1-p)$

ones, and the number of distinct strings of this form is given by the binomial coefficient. So, using $\log x! = x\log x - x + \mathcal{O}(\log x)$, we have:

$$
\begin{aligned}
\log g &= \log \binom{n}{np} = \log\left(\frac{n!}{(np)!(n(1-p))!}\right) && (1.70)\\
&\approx n\log n - n - (np\log(np) - np + n(1-p)\log(n(1-p)) - n(1-p)) && (1.71)\\
&= nH(p), && (1.72)\\
\text{for} \quad & H(p) = -p\log p - (1-p)\log(1-p). && (1.73)
\end{aligned}
$$

What about actual compression? Again, we make up an integer code that labels every typical message. There are about $2^{nH(p)}$ messages, and a priori *typical* messages occur with equal frequency, so we need to specify a given message by a binary string whose length is about $nH(p)$. If $p = 1/2$ (and thus $H(p) = 1$ for $\log_2$) we haven't done anything: we need as many bits to communicate the message as there are in the message. But if the probability $p \neq 1/2$, our new code shortens typical messages. The insight here is that we don't need a codeword for every message, just typical ones, since the probability of atypical messages is negligible!

1.7.2 Information, conditional entropy, and mutual information

The Shannon entropy is a way of quantifying our ignorance (per letter) about the output of a source operating with X: if the source sends an n-character message, we need about $nH(X)$ bits to know the message. *Information* quantifies how much knowledge you gain by knowing the probability distribution the characters came from, i.e., "if you know the $\{p_i\}$, how many fewer bits do I need to transmit to tell you the source's (typical) message?". Well, the total reduction in the number of bits for a n length message from the alphabet of k characters is

$$n\log_2 k - (-n\sum_i p_i \log_2 p_i) = n\left(\log_2 k + \sum_i p_i \log_2 p_i\right). \tag{1.74}$$

Given a knowledge of the p_i, we define the *information per bit* as

$$I(X) = \log_2 k + \sum_i p_i \log_2 p_i, \tag{1.75}$$

so that information and entropy are the same (up to signs and constants).

information and entropy of the uniform distribution As a quick example, suppose we have a uniform distribution of k characters, $p_i = 1/k$. Well:

$$
\begin{aligned}
S &= k\left(\frac{1}{k}\log_2\frac{1}{k}\right) = \log_2\frac{1}{k} && (1.76)\\
I &= \log_2 k + \log_2\frac{1}{k} = 0. && (1.77)
\end{aligned}
$$

So, the entropy is the log of the number of equal-probability characters (sound familiar from the microcanonical ensemble?), and there is no information in the distribution.

information and entropy of a delta function distribution The opposite extreme is also trivial to work out. Suppose the distribution is such that a particular event definitely happens: $p_i = \delta_{\alpha,i}$. Well:

$$\begin{aligned} S &= 0 && (1.78)\\ I &= \log_2 k. && (1.79) \end{aligned}$$

By knowing the distribution you already know *everything* about the outcome of an n-length message, and the entropy (a quantification of ignorance) is zero.

Finally, suppose we have two *correlated* sources of information, X and Y (for uncorrelated sources we would have $p(x,y) = p_X(x)p_Y(y)$). Then, if I read a message in Y^n I can further reduce my ignorance about a message generated by X^n (if I know the correlations!), which means I should be able to further compress messages in X^n than I could without access to Y. This is captured by the *conditional entropy*,

$$H(X|Y) = H(XY) - H(Y) = \langle -\log(p(x,y) + \log p(y)\rangle = \langle -\log p(x|y)\rangle\,, \tag{1.80}$$

where we see the conditional probability distribution introduced earlier in this chapter. Unsurprisingly (given the connection between information and entropy above), the *mutual information* is closely connected: The information about X gained when you learn about Y (again, "the number of fewer bits per letter needed to specify X when Y is known) is

$$\begin{aligned} I(X;Y) &= H(X) - H(X|Y)\\ &= H(X) + H(Y) - H(XY) \qquad (1.81)\\ &= H(Y) - H(Y|X), \end{aligned}$$

which is a quantification of the degree to which X and Y are correlated, and is symmetric under the interchange of X and Y as we see above.

1.7.3 Unbiased estimation of probabilities

We can now *use* the entropy as a way to quantify subjective estimates/assignments of probabilities! To start, if we have no information, the *unbiased estimate* is that every outcome is equally likely. Indeed, we saw above that the uniform distribution encodes no information and maximizes the entropy. Given additional information, we obtain the unbiased estimate by maximizing entropy subject to the constraints.

To give an example, suppose we observe a random variable for a while and observe that it has a specific mean value, $\langle F(x)\rangle = f$. We want make an unbiased estimate of the probabilities p_i, and we will use Lagrange multipliers (α, β) to impose the two constraints we now know about the probability distribution (that it is normalized and has a specific mean):

$$S\left(\{p_i\}, \alpha, \beta\right) = -\sum_i p_i \ln p_i - \alpha\left(\sum_i p_i - 1\right) - \beta\left(\sum_i p_i F(x_i) - f\right). \tag{1.82}$$

Maximizing this with respect to the p_i:

$$\begin{aligned} 0 = \ \tfrac{dS}{dp_i} = &\ -\ln p_i - 1 - \alpha - \beta F(x_i) && (1.83) \\ \Rightarrow \quad p_i = &\ e^{-(1+\alpha)} e^{-\beta F(x_i)}. && (1.84) \end{aligned}$$

Given this form (which should remind you of Boltzmann weights!), we can then solve for α and β that satisfy the given constraints.

Chapter 2

Kinetic Theory: from Liouville to the H-theorem

Perspective and questions for this chapter

At the beginning of class we talked about thermodynamics as a phenomenological theory built on encoding "black-box" observations of material systems into the laws of thermodynamics. But, as emphasized at the time, "We know what's inside of the box! It's molecules interacting with each other (via either classical or quantum mechanics, as the scale of the problem demands)!" So, we should be able to derive the structure of thermodynamics from what we already know about equations of motion and the probability theory we learned in Chapter 2.

In this section we will explore the classical mechanics of a dilute gas. Major questions we want to think about (ref Fig. 2.1):

1. How do we even *define* the idea of "equilibrium" for a system made out of particles?

2. Do such systems evolve towards equilibrium? How could they?! Every microscopic equation we want to write down is *time reversible*, but if a system evolves from a non-equilibrium to an equilibrium state it is picking out a direction of time.

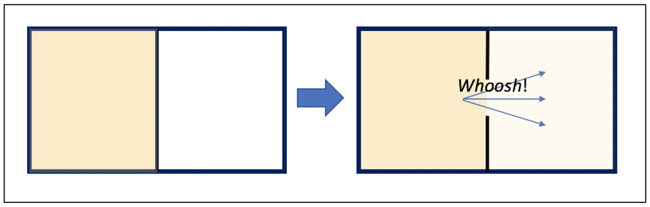


Figure 2.1: **An ideal gas rushes to fill the available space after a partition is removed** The gas is composed of microscopic degrees of freedom evolving according to completely time-reversible dynamics, but "Entropy increases." What is entropy and how is it time irreversible for such a process?

Sources for this chapter: The material here will cover most of Pathria, Chapter 2; thanks to the work we did in the lectures on Probability we will also either implicitly or explicitly cover Pathria Chapter 3.1 - 3.3. We will also be covering the BBGKY hierarchy and Boltzmann's H-theorem, which are not in Pathria.

2.1 Elements of ensemble theory

2.1.1 Phase space of a classical system[1]

We want to connect a thermodynamic description of a system – which is captured by only a handful of coordinates, like V, N, T, etc. – with a *classical microstate*, which is specified by the positions and momenta of each of my N particles: $\mu = \{\boldsymbol{q}_1, \boldsymbol{p}_1, \ldots, \boldsymbol{q}_N, \boldsymbol{p}_N\}$, which is a point in a $6N$-dimensional phase space (and which is, hence, hard to draw). Since we're starting with a classical mechanical description, we'll say the system is governed by some Hamiltonian, $\mathcal{H}$, so that the equations of motions we'll work with are

$$\begin{cases} \frac{d\boldsymbol{p}_i}{dt} &= -\frac{\partial \mathcal{H}}{\partial \boldsymbol{q}_i} &= \dot{\boldsymbol{p}}_i \\ \frac{d\boldsymbol{q}_i}{dt} &= \frac{\partial \mathcal{H}}{\partial \boldsymbol{p}_i} &= \dot{\boldsymbol{q}}_i, \end{cases} \tag{2.1}$$

and where time-reversal invariance means that if we reverse the direction of time we transform $\boldsymbol{p} \to -\boldsymbol{p}$ and $\boldsymbol{q}(t) \to \boldsymbol{q}(-t)$.

Now, given that there are a handful of thermodynamic coordinates describing the system and $\sim 10^{23}$ describing the classical microstate, it will not surprise us that there is a many-to-one mapping between equilibrium states and microstates. How do we formalize this? Let's start by imagining we have $\mathcal{N}$ *"copies"* of the same macrostate of our system, each corresponding to a different representative microstate, and we'll think about the *ensemble density function*,

$$\rho(\boldsymbol{p}, \boldsymbol{q}, t) = \lim_{\mathcal{N}\to\infty, d\Gamma\to 0} \frac{d\mathcal{N}(\boldsymbol{p}, \boldsymbol{q}, t)}{\mathcal{N} d\Gamma}, \tag{2.2}$$

where

$$d\Gamma = \prod_{i=1}^{N} d^3\boldsymbol{p}_i d^3\boldsymbol{q}_i$$

is how we'll be writing classical phase space volume differentials and $d\mathcal{N}$ is the number of microstates corresponding to the target macrostate in our differential box. Note that if we integrate ρ over all of phase space we get

$$\int d\Gamma \rho(\boldsymbol{p}, \boldsymbol{q}) = \int \frac{d\mathcal{N}}{\mathcal{N}} = 1, \tag{2.3}$$

so we see that between positivity (we're counting numbers of points, so it's not negative) and the above, the ensemble density is actually a probability density function. We already know things, then, like computing expectation values:

$$\langle A \rangle = \int d\Gamma \rho(\boldsymbol{p}, \boldsymbol{q}, t) A(\boldsymbol{p}, \boldsymbol{q})$$

---

[1] This section covers Pathria 2.1

.

Note that we've been writing time dependences in the above, but for equilibrium macrostates we believe that ensemble averages shouldn't depend on time. Microscopically, though, we know that if we take snapshots of the state of our system the microstates at time t and $t+\Delta t$ will look measurably different. Thus, in equilibrium we will look for *stationary* ensemble densities, for which

$$\frac{\partial \rho}{\partial t} = 0. \tag{2.4}$$

2.1.2 Liouville's theorem and its consequences[2]

Incompressibility of phase space volumes

Liouville's theorem, which I'm sure you encountered in classical mechanics, is a characterization of the evolution of ρ with time, and states that ρ behaves like an imcompressible fluid. So, first, how does the phase space density evolve? Let's look at Fig. 2.2, which shows two dimensions of phase space around some representative point p_α, q_α, as well as where those points have moved after a short time dt.

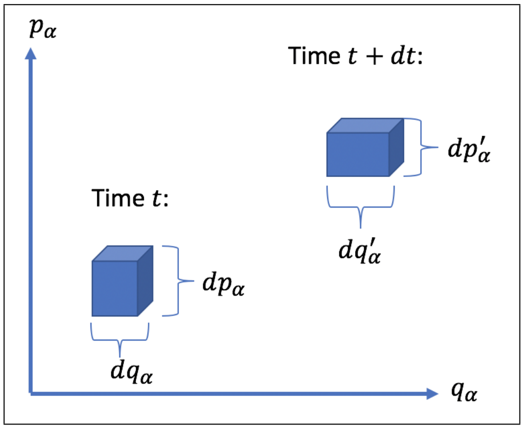


Figure 2.2: **Evolution of phase space volumes**

First, note that the representative point itself flows like

$$\begin{cases} q'_\alpha = q_\alpha + \dot{q}_\alpha dt + \mathcal{O}(dt^2) \\ p'_\alpha = p_\alpha + \dot{p}_\alpha dt + \mathcal{O}(dt^2), \end{cases} \tag{2.5}$$

and the nearby points in phase space also flow:

$$\begin{cases} dq'_\alpha = dq_\alpha + \frac{\partial \dot{q}_\alpha}{\partial q_\alpha} dq_\alpha dt + \cdots \\ dp'_\alpha = dp_\alpha + \frac{\partial \dot{p}_\alpha}{\partial p_\alpha} dp_\alpha dt + \cdots , \end{cases} \tag{2.6}$$

so for each pair of conjugate coordinates we see that

$$dq'_\alpha dp'_\alpha = dq_\alpha dp_\alpha \left[1 + dt(\frac{\partial \dot{q}_\alpha}{\partial q_\alpha} + \frac{\partial \dot{p}_\alpha}{\partial q_\alpha}) + \mathcal{O}(dt^2)\right] = dq_\alpha dp_\alpha, \tag{2.7}$$

---

[2] This section covers Pathria 2.2, but done slightly differently

where the last equality holds because the term proportional to dt vanishes by equality of mixed partials. This is the just the familiar statement that Hamiltonian dynamics preserves phase space volumes: $d\Gamma = d\Gamma'$.

Liouville's theorem

The above has consequences for our ensemble density. All of the states $d\mathcal{N}$ that were originally near $(\boldsymbol{p}, \boldsymbol{q})$ move to the neighborhood of $(\boldsymbol{p}', \boldsymbol{q}')$, but occupy the same phase space volume. Thus, $d\mathcal{N}/d\Gamma$ is unchanged, and ρ itself behaves like an incompressible fluid.Given how phase space transforms, we can write

$$\rho(\boldsymbol{p}, \boldsymbol{q}, t) = \rho(\boldsymbol{p}', \boldsymbol{q}', t + dt). \tag{2.8}$$

Expanding out this expression, we can write

$$\begin{aligned} \rho(\boldsymbol{p}, \boldsymbol{q}, t) &= \rho\left(\boldsymbol{p} + \dot{\boldsymbol{p}}dt, \boldsymbol{q} + \dot{\boldsymbol{q}}dt, t + dt\right) && (2.9) \\ &= \rho(\boldsymbol{p}, \boldsymbol{q}, t) + \left[\sum_\alpha \dot{p}_\alpha \frac{\partial \rho}{\partial p_\alpha} + \dot{q}_\alpha \frac{\partial \rho}{\partial q_\alpha} + \frac{\partial \rho}{\partial t}\right] dt + \mathcal{O}(dt^2). && (2.10) \end{aligned}$$

Let's define the total derivative (or "streamline derivative") of a function f as

$$\frac{d}{dt} f(\boldsymbol{p}, \boldsymbol{q}, t) = \frac{\partial f}{\partial t} + \sum_\alpha \frac{\partial f}{\partial p_\alpha} \dot{p}_\alpha + \frac{\partial f}{\partial q_\alpha} \dot{q}_\alpha, \tag{2.11}$$

where the interpretation is that d/dt is the derivative as you flow (following the evolution of the volume of fluid as it moves through phase space), whereas the partial derivative ∂/∂_t is like sitting a fixed position in space and watching the changes in f in time at that location. Anyway, the equation for ρ says that

$$\begin{aligned} \frac{d\rho}{dt} = 0 &= \frac{\partial \rho}{\partial t} + \sum_\alpha \left(\frac{-\partial \mathcal{H}}{\partial q_\alpha}\right) \frac{\partial \rho}{\partial p_\alpha} + \left(\frac{\partial \mathcal{H}}{\partial p_\alpha}\right) \frac{\partial \rho}{\partial q_\alpha} && (2.12) \\ \Rightarrow \qquad & \frac{\partial \rho}{\partial t} = \{\mathcal{H}, \rho\}, && (2.13) \end{aligned}$$

where we have defined the Poisson bracket[3]

$$\{A, B\} = \sum_i \frac{\partial A}{\partial q_i} \frac{\partial B}{\partial p_i} - \frac{\partial A}{\partial p_i} \frac{\partial B}{\partial q_i} \tag{2.14}$$

2.1.3 Equilibrium ensemble densities

We can combine Liouville's theorem above with our criteria that in equilibrium ρ is stationary to come up with a criteria for equilibrium:

$$\frac{\partial \rho_{eq}(\boldsymbol{q}, \boldsymbol{p})}{\partial t} = 0 = \{\mathcal{H}, \rho_{eq}\}. \tag{2.15}$$

---

[3] Using curly braces, not the weird straight bracket of Pathria. Shots fired. I am going to use Pathria's / Goldstein's sign convention here, though, instead of Landau's.

In principle, we now have to solve the system of $6N$ variables coming from setting the Poisson bracket to zero. In practice, we'll guess! We have already required that in equilibrium ρ has no explicit time derivative; we could also assume it has no dependence on $\boldsymbol{q}$ or $\boldsymbol{p}$ at all: $\rho_{eq} =$ const. is certainly a valid solution of the above (which is like saying the ensemble of systems corresponding to the equilibrium macrostate are uniformly distributed through phase space).

More generally, though allow implicit dependencies on the phase space coordinates and let ρ be an arbitrary function of $\mathcal{H}$:

$$\{\rho_{eq}(\mathcal{H}), \mathcal{H}\} = \sum_i \left(\frac{\partial \rho}{\partial \mathcal{H}} \frac{\partial \mathcal{H}}{\partial q_i}\right) \frac{\partial \mathcal{H}}{\partial p_i} - \left(\frac{\partial \rho}{\partial \mathcal{H}} \frac{\partial \mathcal{H}}{\partial p_i}\right) \frac{\partial \mathcal{H}}{\partial q_i} = 0. \tag{2.16}$$

For example, choosing

$$\rho(\mathcal{H}) = \delta(\mathcal{H} - E)$$

gives us the microcanonical ensemble, and choosing

$$\rho(\mathcal{H}) \propto \exp(-\beta \mathcal{H}(\boldsymbol{q}, \boldsymbol{p}))$$

gives us the canonical ensemble.

Even more generally, while we're in the business of simply guessing solutions to Eq. 2.15, we can assume that ρ is an arbitrary function of both $\mathcal{H}$ itself *and* of any conserved quantities. To see why, we first note that if some quantity A is conserved under the Hamiltonian, $\{A, \mathcal{H}\} = 0$. Then, using the same manipulations as above, we can write

$$\{\rho_{eq}(\mathcal{H}, A), \mathcal{H}\} = \frac{\partial \rho}{\partial \mathcal{H}}\{\mathcal{H}, \mathcal{H}\} + \frac{\partial \rho}{\partial A}\{A, \mathcal{H}\} = 0, \tag{2.17}$$

satisfying our condition for ρ_{eq} to be an equilibrium distribution.

Time dependence of observables

It is perhaps worth explicitly pointing out that to find the time dependence of some ensemble average of a quantity A, we do:

$$\frac{d}{dt}\langle A \rangle = \int d\Gamma \frac{\partial \rho(\boldsymbol{p}, \boldsymbol{q}, t)}{\partial t} A(\boldsymbol{p}, \boldsymbol{q}) \tag{2.18}$$

$$= \sum_{i=1}^{3N} \int d\Gamma A(\boldsymbol{p}, \boldsymbol{q}) \{\mathcal{H}, \rho\}, \tag{2.19}$$

where the key note is that you *cannot* just bring the total time derivative inside the integral sign. One can then write out the Poisson bracket and change the partial derivatives acting on ρ by integration by parts Make sure we're all okay with this! If you're reading this, actually do the manipulations to go from the above to the below. This leads to:

$$\frac{d}{dt}\langle A \rangle = \sum_{i=1}^{3N} \int d\Gamma \rho \left[\left(\frac{\partial A}{\partial p_i} \frac{\partial \mathcal{H}}{\partial q_i} - \frac{\partial A}{\partial q_i} \frac{\partial \mathcal{H}}{\partial p_i}\right) + A \left(\frac{\partial^2 \mathcal{H}}{\partial p_i \partial q_i} - \frac{\partial^2 \mathcal{H}}{\partial q_i \partial p_i}\right)\right] \tag{2.20}$$

$$= -\int d\Gamma \rho \{\mathcal{H}, A\} = \langle \{A, \mathcal{H}\} \rangle \tag{2.21}$$

2.2 BBGKY[4] hierarchy

Starting comments

We are about to embark on a sequence of formal manipulations, so it is worth emphasizing where we are going today (and why). In the last lecture we addressed the first question posed at the start of the chapter – "How do we define the idea of equilibrium for a system described microscopically as a collection of particles?" – by thinking about stationary ensemble density functions.

Departing from the sequence in Pathria, the rest of the chapter will be devoted to the second question: if we start from a non-equilibrium density, can we show it evolves towards equilibrium (a la Fig. 2.1 – an experiment which we know reproducibly leads from one equilibrium state to another)? Where does time irreversibility come from?

We have already seen "entropy" defined in two seemingly very different ways: entropy as a thermodynamic state function somehow related to heat and temperature, and entropy in the information theoretic sense as a quantification of our ignorance about the output of a probabilistic source. By the end of the chapter we will see yet another version of entropy: the Boltzmann version of entropy, capturing information about one-particle densities.

BBGKY

In the last lecture we encountered the ensemble density, $\rho(\boldsymbol{p}_1, \ldots, \boldsymbol{p}_N, \ldots, \boldsymbol{q}_N, t)$, but in general this contains *far* more microscopic information than we would ever need to compute or describe equilibrium properties (for instance – knowledge of the typical behavior of just a single particle would be sufficient to calculate the pressure of a gas. Let's define the *one-particle density* as the expectation value of finding any of the N particles at some particular location with some particular momentum:

$$f_1(\boldsymbol{p}, \boldsymbol{q}, t) = \left\langle \sum_{i=1}^{N} \delta^3(\boldsymbol{p} - \boldsymbol{p}_i)\delta^3(\boldsymbol{q} - \boldsymbol{q}_i) \right\rangle = N \int \prod_{i=2}^{N} dV_i \rho(\boldsymbol{p}_1 = \boldsymbol{p}, \boldsymbol{q}_1 = \boldsymbol{q}, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_N, \boldsymbol{q}_N), \tag{2.22}$$

where $dV_i = d^3\boldsymbol{p}_i d^3\boldsymbol{q}_i$. The general s-particle density is defined similarly, integrating over the $s+1$ through Nth variables:

$$f_s(\boldsymbol{p}_1, \ldots, \boldsymbol{q}_s, t) = \frac{N!}{(N-s)!}\rho_s(\boldsymbol{p}_1, \ldots, \boldsymbol{q}_s, t), \tag{2.23}$$

---

[4] Bogoliubov-Born-Green-Kirkwood-Yvon. Some aspects introduced by Yvon in 1935; hierarchy written out by Bogoliubov in 1945, Kirkwood did kinetic transport work in 1945/1946; and Born and Green used an analogous structure for the kinetic theory of liquids in 1946

where

$$\rho_s(\boldsymbol{p}_1,\ldots,\boldsymbol{q}_s,t)=\int\prod_{i=s+1}^{N}dV_i(\boldsymbol{p}_1,\ldots,\boldsymbol{q}_N,t) \tag{2.24}$$

is just the unconditional pdfs for the coordinates of s of the particles. These f_s functions differ from the unconditional pdfs by a simple normalization factor, and we'll see why each is vaguely preferable in different situations.

Now, this is a lovely set of definitions. What we really want is a way of expressing the time evolution of these s-particle densities (e.g., if we can tell the pressure from f_1, then knowing $f_1(t)$ will let us study how pressure equilibrates in time as we go from one equilibrium state to the other;again, c.f. Fig. 2.1). Well,

$$\frac{\partial f_s}{\partial t}=\frac{N!}{(N-s)!}\int\prod_{i=s+1}^{N}dV_i\frac{\partial\rho}{\partial t}=\frac{N!}{(N-s)!}\int\prod_{i=s+1}^{N}dV_i\{\mathcal{H},\rho\}. \tag{2.25}$$

It would be quite difficult to make any interesting headway here for a truly arbitrary Hamiltonian, so let's consider the case of up to two-body interactions:

$$\mathcal{H}=\sum_{i=1}^{N}\left[\frac{\boldsymbol{p}_i^2}{2m}+U(\boldsymbol{q}_i)\right]+\frac{1}{2}\sum_{i,j=1}^{N}V(\boldsymbol{q}_i-\boldsymbol{q}_j), \tag{2.26}$$

where U is some external potential and V is some pairwise inter-particle potential. What's our strategy going to be? Well, we have to integrate over a Poisson bracket, so integration by parts will be extremely useful whenever we're taking a derivative with respect to one of the variables we're integrating over. Because of this, let's partition the sums into dummy variables running over the first s particle coordinates, running over the last $N-s$, and cross terms:

$$\begin{aligned}
\mathcal{H} &= \mathcal{H}_s+\mathcal{H}_{N-s}+\mathcal{H}_x && (2.27)\\
\mathcal{H}_s &= \sum_{i=1}^{s}\left[\frac{\boldsymbol{p}_i^2}{2m}+U(\boldsymbol{q}_i)\right]+\frac{1}{2}\sum_{i,j=1}^{s}V(\boldsymbol{q}_i-\boldsymbol{q}_j) && (2.28)\\
\mathcal{H}_{N-s} &= \sum_{i=s+1}^{N}\left[\frac{\boldsymbol{p}_i^2}{2m}+U(\boldsymbol{q}_i)\right]+\frac{1}{2}\sum_{i,j=s+1}^{N}V(\boldsymbol{q}_i-\boldsymbol{q}_j) && (2.29)\\
\mathcal{H}_x &= \sum_{i=1}^{s}\sum_{j=s+1}^{N}V(\boldsymbol{q}_i-\boldsymbol{q}_j). && (2.30)
\end{aligned}$$

This lets us write

$$\frac{\partial\rho_s}{\partial t}=\int\prod_{i=s+1}^{N}dV_i\{\mathcal{H}_s+\mathcal{H}_{N-s}+\mathcal{H}_x,\rho\}, \tag{2.31}$$

and we'll take each part of the Poisson bracket in turn.

Self term, $\mathcal{H}_s$ Notice that the variables we are integrating over do not show up in ρ_s or $\mathcal{H}_s$, so we can interchange the order of integrations and differentiations and write

$$\int \prod_{i=s+1}^{N} dV_i \{\mathcal{H}_s, \rho\} = \{\mathcal{H}_s, \left(\int \prod_{i=s+1}^{N} dV_i \rho\right)\} = \{\mathcal{H}_s, \rho_s\}. \tag{2.32}$$

This has a natural interpretation / connection to what we already know: *if there are only s particles, we have just re-written the Liouville equation.*

Just the non-s particle terms, $\mathcal{H}_{N-s}$ We will handle all of these terms via integration by parts:

$$\int \prod_{i=s+1}^{N} dV_i \sum_{j=1}^{N} \left(\frac{\partial \rho}{\partial \boldsymbol{p}_j} \cdot \frac{\partial \mathcal{H}_{N-s}}{\partial \boldsymbol{q}_j} - \frac{\partial \rho}{\partial \boldsymbol{q}_j} \cdot \frac{\partial \mathcal{H}_{N-s}}{\partial \boldsymbol{p}_j}\right) \tag{2.33}$$

$$= [\text{surface terms}] + \int \prod_{i=s+1}^{N} dV_i \rho \left(\frac{\partial^2 H_{N-s}}{\partial \boldsymbol{p}_i \partial \boldsymbol{q}_i} - \frac{\partial^2 H_{N-s}}{\partial \boldsymbol{q}_i \partial \boldsymbol{p}_i}\right) \quad = \quad 0. \tag{2.34}$$

Cross terms, $\mathcal{H}_x$ Apparently if there is going to be anything interesting it will come from these cross terms. We're making progress, here:

$$\int \prod_{i=s+1}^{N} dV_i \sum_{j=1}^{N} \left[\frac{\partial \rho}{\partial \boldsymbol{p}_j} \cdot \frac{\partial \mathcal{H}_x}{\partial \boldsymbol{q}_j} - \frac{\partial \rho}{\partial \boldsymbol{q}_j} \cdot \frac{\partial \mathcal{H}_x}{\partial \boldsymbol{p}_j}\right] \tag{2.35}$$

$$= \int \prod_{i=s+1}^{N} dV_i \left(\left[\sum_{k=1}^{s} \frac{\partial \rho}{\partial \boldsymbol{p}_k} \cdot \sum_{j=s+1}^{N} \frac{\partial V(\boldsymbol{q}_k - \boldsymbol{q}_j)}{\partial \boldsymbol{q}_k}\right] + \left[\sum_{j=s+1}^{N} \frac{\partial \rho}{\partial \boldsymbol{p}_j} \cdot \sum_{k=1}^{s} \frac{\partial V(\boldsymbol{q}_j - \boldsymbol{q}_k)}{\partial \boldsymbol{q}_k}\right]\right) \tag{2.36}$$

$$= \sum_{k=1}^{s} \int \prod_{i=s+1}^{N} dV_i \frac{\partial \rho}{\partial \boldsymbol{p}_k} \cdot \sum_{j=s+1}^{N} \frac{\partial V(\boldsymbol{q}_k - \boldsymbol{q}_j)}{\partial \boldsymbol{q}_k} \tag{2.37}$$

where in the first step we have used the fact that $\mathcal{H}_x$ is independent of any $\boldsymbol{p}_i$, and in the second step yet another integration by parts shows that the second term above vanishes. Physically, we expect that we can treat all of the $j = s+1, \ldots, N$ particles equivalently, so we change the labels of our sums and replace the sum by $(N-s)$ equivalent integrations:

$$\int \prod_{i=s+1}^{N} dV_i \{\mathcal{H}_x, \rho\} = (N-s) \sum_{i=1}^{s} \int dV_{s+1} \frac{\partial V(\boldsymbol{q}_i - \boldsymbol{q}_{s+1})}{\partial \boldsymbol{q}_i} \cdot \frac{\partial}{\partial \boldsymbol{p}_i} \left(\int \prod_{s+2}^{N} dV_i \rho\right). \tag{2.38}$$

Notice that the final term in parentheses, above, is just ρ_{s+1}. This, at last, is the key connection we wanted to make.

Combining everything: the hierarchy Using either ρ_s or f_s, we have a connection between different s-body densities:

$$\boxed{\frac{\partial f_s}{\partial t} - \{\mathcal{H}_s, f_s\} = \sum_{i=1}^{s} \int dV_{s+1} \frac{\partial V(\boldsymbol{q}_i - \boldsymbol{q}_{s+1})}{\partial \boldsymbol{q}_i} \cdot \frac{\partial f_{s+1}}{\partial \boldsymbol{p}_i}}. \tag{2.39}$$

This, at last, is the BBGKY Hierarchy! But why have we gone through the trouble?? We now have a hierarchy of relations: the 1-body density is given by a self-term plus a sum over 2-body terms, the 2-body density is given by a self-term plus a sum over 3-body terms, and so on... You may think to yourself, "Great – we've replaced a very complicated function of $\mathcal{O}(10^{23})$ variables with set of $\mathcal{O}(10^{23})$ coupled equations... what a day!" But this is actually quite useful – the hierarchy has isolated the simplest variables, e.g. f_1, which are often the most physically important. Then, given a particular problem we want to solve, we can *inject physical approximations* to decide what terms in the hierarchy can be truncated, simplified, etc., as we are about to see.

2.3 The Boltzmann Equation – intuitive version[5]

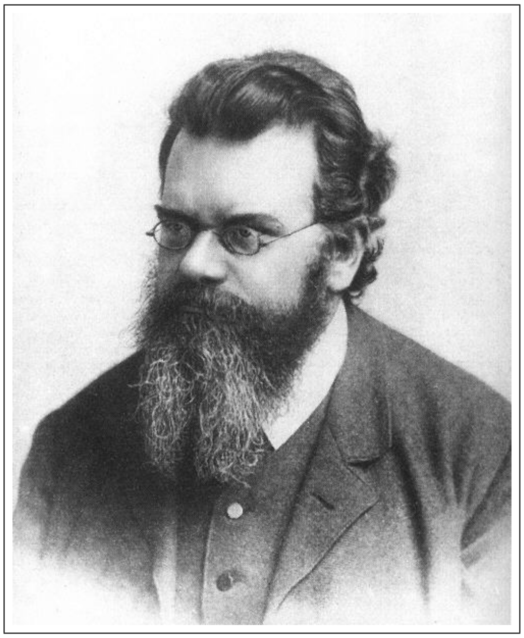


Figure 2.3: **"Eleganz sei die Sache der Schuster und Schneider"** is apparently a thing Boltzmann said. Does that inspire excitement for this section?

Going from the BBGKY hierarchy to the Boltzmann equation is fussy (cf. the quote in Fig. 2.3), and so we begin with a hand-wavy derivation in which we basically just guess the right answer first to provide some intuition. Let's write the first level of the hierarchy as

$$\frac{\partial f_1}{\partial t} = \{\mathcal{H}_1, f_1\} + \left(\frac{\partial f_1}{\partial t}\right)_{coll}, \tag{2.40}$$

where we know the thing we've labeled "coll" above,

$$\left(\frac{\partial f_1}{\partial t}\right)_{coll} = \int dV_2 \frac{\partial \Phi_{12}}{\partial \boldsymbol{q}_1} \cdot \frac{\partial}{\partial \boldsymbol{p}_1} f_2, \tag{2.41}$$

---

[5] This section follows in part the treatment in David Tong's notes, available here, from his course on Kinetic Theory, and in part the treatment in Kardar's Statistical Physics of Particles.

represents the way single particles change their momenta by collisions with a second particle. Well, let's assume that collisions are local, so a particle at $\boldsymbol{q}$ with momentum $\boldsymbol{p}$ collides with a second particle *also* at $\boldsymbol{q}$ but with momentum $\boldsymbol{p}_2$; after this collision the particles have new momenta $\boldsymbol{p}_1'$ and $\boldsymbol{p}_2'$, respectively. If we use the *scattering function* $\omega(\boldsymbol{p},\boldsymbol{p}_2|\boldsymbol{p}_1',\boldsymbol{p}_2')$ to keep track of the information about the dynamics of the collision process (we'll get into more detail later, but in general this is something you can just compute given the pairwise potential $V(\boldsymbol{q})$), we can write the rate at which the collision process happens as

$$\text{rate} = \omega(\boldsymbol{p},\boldsymbol{p}_2|\boldsymbol{p}_1',\boldsymbol{p}_2')f_2(\boldsymbol{q}_1=\boldsymbol{q},\boldsymbol{q}_2=\boldsymbol{q},\boldsymbol{p}_1=\boldsymbol{p},\boldsymbol{p}_2,t)d^3p_2d^3p_1'd^3p_2'. \tag{2.42}$$

That is, the rate is proportional to the details of the scattering process times f_2, which itself gives us the probability of of having particles with $(\boldsymbol{q},\boldsymbol{p})$ and $(\boldsymbol{q},\boldsymbol{p}_2)$ in the first place. Note that when using this in Eq. 2.41 we need to consider both scattering *out of* the state where the particle has momentum $\boldsymbol{p}$ but also *into* the state with momentum $\boldsymbol{p}$. This suggests a collision integral with two terms:

$$\left(\frac{\partial f_1}{\partial t}\right)_{coll} = \int d^3p_2d^3p_1'd^3p_2'\left[\omega(\boldsymbol{p}_1',\boldsymbol{p}_2'|\boldsymbol{p},\boldsymbol{p}_2)f_2(\boldsymbol{q},\boldsymbol{q},\boldsymbol{p}_1',\boldsymbol{p}_2') - \omega(\boldsymbol{p},\boldsymbol{p}_2|\boldsymbol{p}_1',\boldsymbol{p}_2')f_2(\boldsymbol{q},\boldsymbol{q},\boldsymbol{p},\boldsymbol{p}_2)\right]. \tag{2.43}$$

Let's put some generic constraints on the scattering function (without knowing much about the details of the interparticle interactions). Of course it should only by nonzero if momentum and energy are conserved, and first we assume that any external potential only varies on scales much larger than the interaction range relevant to the collisions. So, ω is only nonzero if

$$\boldsymbol{p}+\boldsymbol{p_2} = \boldsymbol{p}_1'+\boldsymbol{p}_2', \qquad p^2+p_2^2 = p_1'^2+p_2'^2.$$

We want to say that the scattering rate is unchanged if you simple exchange the ingoing and the outgoing momenta (and that it also doesn't depend on where the collisions take place), so that:

$$\omega(\boldsymbol{p},\boldsymbol{p}_2|\boldsymbol{p}_1',\boldsymbol{p}_2') = \omega(\boldsymbol{p}_1',\boldsymbol{p}_2'|\boldsymbol{p},\boldsymbol{p}_2).$$

We can use this to simplify the collision integral:

$$\left(\frac{\partial f_1}{\partial t}\right)_{coll} = \int d^3p_2d^3p_1'd^3p_2'\omega(\boldsymbol{p}_1',\boldsymbol{p}_2'|\boldsymbol{p},\boldsymbol{p}_2)\left[f_2(\boldsymbol{q},\boldsymbol{q},\boldsymbol{p}_1',\boldsymbol{p}_2') - f_2(\boldsymbol{q},\boldsymbol{q},\boldsymbol{p},\boldsymbol{p}_2)\right]. \tag{2.44}$$

To finish the "derivation" of the Boltzmann equation and write a closed equation for f_1, we make a final, big approximation (*"the assumption of molecular chaos"*) that momenta of the two particles are uncorrelated:

$$f_2(\boldsymbol{q},\boldsymbol{q},\boldsymbol{p},\boldsymbol{p}_2) = f_1(\boldsymbol{q},\boldsymbol{p})f_1(\boldsymbol{q},\boldsymbol{p}_2). \tag{2.45}$$

This, perhaps, doesn't look so strong an assumption on its surface, but looking at how f_2 enters the rate of collision expressions we've written down, we see that we are more explicitly assuming that the momenta are uncorrelated *before* the collision, and then after the collision the momenta follow from conservation properties during the scattering process. This has, quite sneakily, been a means to smuggle in an arrow of time. We will see the implications when we get to the H-theorem later in the chapter.

For now, let's just finish putting the pieces together by writing the Boltzmann equation itself:

$$\boxed{\frac{\partial f_1}{\partial t} - \{\mathcal{H}_1, f_1\} = \int d^3p_2 d^3p_1' d^3p_2' \omega(\boldsymbol{p}_1', \boldsymbol{p}_2' | \boldsymbol{p}, \boldsymbol{p}_2) \left[f_1(\boldsymbol{q}, \boldsymbol{p}_1') f_1(\boldsymbol{q}, \boldsymbol{p}_2') - f_1(\boldsymbol{q}, \boldsymbol{p}) f_1(\boldsymbol{q}, \boldsymbol{p}_2) \right]} \tag{2.46}$$

Shoemakers and tailors, indeed. The Boltzmann equation combines derivatives and integrals and nonlinearity all at the same time, and exact solutions to it are not so easy to come by. But in this framework we can show that systems do reach equilibrium if they start out of it, and we didn't have to add friction or some other means of dissipating energy; we just had to innocuously assume that momenta are uncorrelated before collisions. Let first do a somewhat more formal derivation, making use of the hierarchy we sketched out earlier.

2.4 Boltzmann a la BBGKY

Now that we know where we are heading, let's derive the Boltzmann equation with a bit more rigor. After having done so, we'll ask ourselves about the *consequences* of the Boltzmann equation. What do its solutions tell us about the behavior of equilibrium states? What does it say about the origin of irreversibility?

2.4.1 BBGKY for a dilute gas

You may have noticed that so far we have not used the fact that we are studying a dilute gas – that is about to change as we are finally in a position to make some physically motivated approximations to the BBGKY hierarchy. Let's start by explicitly writing the first two levels of the hierarchy, where for notational ease we'll write the derivative of the pairwise potential as a force: $\frac{\partial V(\boldsymbol{q}_i - \boldsymbol{q}_j)}{\partial \boldsymbol{q}_i} = \frac{\partial \Phi_{ij}}{\partial \boldsymbol{q}_i}$, which is the contribution to the force on i from j. Our first two levels are:

$$\left[\frac{\partial}{\partial t} - \frac{\partial U}{\partial \boldsymbol{q}_1} \cdot \frac{\partial}{\partial \boldsymbol{p}_1} + \frac{\boldsymbol{p}_1}{m} \cdot \frac{\partial}{\partial \boldsymbol{q}_1} \right] f_1 = \int dV_2 \frac{\partial \Phi_{12}}{\partial \boldsymbol{q}_1} \cdot \frac{\partial}{\partial \boldsymbol{p}_1} f_2, \tag{2.47}$$

$$\left[\frac{\partial}{\partial t} - \frac{\partial U}{\partial \boldsymbol{q}_1} \cdot \frac{\partial}{\partial \boldsymbol{p}_1} - \frac{\partial U}{\partial \boldsymbol{q}_2} \cdot \frac{\partial}{\partial \boldsymbol{p}_2} + \frac{\boldsymbol{p}_1}{m} \cdot \frac{\partial}{\partial \boldsymbol{q}_1} + \frac{\boldsymbol{p}_2}{m} \cdot \frac{\partial}{\partial \boldsymbol{q}_2} - \frac{\partial \Phi_{12}}{\partial \boldsymbol{q}_1} \cdot \left(\frac{\partial}{\partial \boldsymbol{p}_1} - \frac{\partial}{\partial \boldsymbol{p}_2} \right) \right] f_2 = \int dV_3 \left[\frac{\partial \Phi_{13}}{\partial \boldsymbol{q}_1} \cdot \frac{\partial}{\partial \boldsymbol{p}_1} + \frac{\partial \Phi_{23}}{\partial \boldsymbol{q}_2} \cdot \frac{\partial}{\partial \boldsymbol{p}_2} \right] f_3. \tag{2.48}$$

Relative importance of terms

Would you like to write/TEX the next level? Neither would I. Let's think physically about the terms in the above two levels: We've arranged things so that every term in the square brackets has dimensions of inverse time, so lets *estimate the typical magnitudes* of the various terms! We're studying a gas, and a reasonable speed for a gas particle at room temperature is on the order of $c = 100m/s$; to make a characteristic time, the typical length scale will depend on the nature of the term in question.

1. First, there is a characteristic time related to the external potential, like $\tau_U^{-1} \sim \frac{\partial U}{\partial \boldsymbol{q}} \cdot \frac{\partial}{\partial \boldsymbol{p}}$: these are spatial variations in the external potential, which we will typically think of taking place over basically *macroscopic* distances, where the characteristic length L is at least a millimeter. Very roughly, that would give us:
$$\tau_U \sim L/v \sim 10^{-5}s$$
2. Next there are terms that scale like a typical collision duration, like $\tau_c^{-1} \sim \frac{\partial \Phi}{\partial \boldsymbol{q}} \cdot \frac{\partial}{\partial \boldsymbol{p}}$; that is, these terms have a magnitude which should be commensurate with the duration over which two particles are within a characteristic effective range of the potential, d. If we restrict ourselves to reasonably short-range interaction potentials (van der Waals, or, say, Lennard-Jones interactions), this effective distance is on the scale of angstroms, $d \sim 10^{-10}m$. Very roughly, that would give us:
$$\tau_c \sim d/v \sim 10^{-12}s$$
3. Finally, there are collisional terms like
$$\tau_x^{-1} \sim \int dV \frac{\partial \Phi}{\partial \boldsymbol{q}} \cdot \frac{\partial}{\partial \boldsymbol{p}} N \frac{\rho_{s+1}}{\rho_s}$$
The integral has some non-zero contribution over a volume that scales like the characteristic volume of the potential, d^3, and the ratio $N\rho_{s+1}/\rho_s$ is like a probability of finding an additional particle in the vicinity of the s particles, which should be of the order of the particle number density, $n = N/V \sim 10^{26}m^{-3}$. Combining this and the above gives, very roughly:
$$\tau_x \sim \frac{\tau_c}{nd^3} \sim \frac{1}{nvd^2} \sim 10^{-8}s$$

What does all of this buy us? Well, we see that the second level (and all higher-s levels) of the hierarchy is balance between three competing terms: something like τ_U^{-1}, something like τ_c^{-1}, and something like τ_x^{-1}. Well, as long as we are in the *dilute limit*, we see that the terms on the RHS of Eq. 2.48 are orders of magnitude smaller than the terms on the LHS, and so we approximate these levels of the hierarchy just by the balance of terms that are like τ_U^{-1} and τ_c^{-1}.

In contrast, the first level is different. It has no terms that are like τ_c^{-1}, and so we have no choice but to keep all of the terms.

Basic closure

This illustrates a common strategy in kinetic (and other) theories...at first glance the hierarchy does not seem helpful, because it is not *closed* (i.e., to solve for ρ_s one needs to know

ρ_{s+1}). However, we can try to come up with a model/theory/approximation that governs the higher-order levels of the hierarchy; the quality of our predictions will then be related to the quality of our closure of the theory. In the present case, explicitly, we have

$$\left[\frac{\partial}{\partial t} - \frac{\partial U}{\partial \boldsymbol{q}_1}\cdot\frac{\partial}{\partial \boldsymbol{p}_1} + \frac{\boldsymbol{p}_1}{m}\cdot\frac{\partial}{\partial \boldsymbol{q}_1}\right] f_1 = \int dV_2 \frac{\partial \Phi_{12}}{\partial \boldsymbol{q}_1}\cdot\frac{\partial}{\partial \boldsymbol{p}_1} f_2, \tag{2.49}$$

$$\left[\frac{\partial}{\partial t} - \frac{\partial U}{\partial \boldsymbol{q}_1}\cdot\frac{\partial}{\partial \boldsymbol{p}_1} - \frac{\partial U}{\partial \boldsymbol{q}_2}\cdot\frac{\partial}{\partial \boldsymbol{p}_2} + \frac{\boldsymbol{p}_1}{m}\cdot\frac{\partial}{\partial \boldsymbol{q}_1} + \frac{\boldsymbol{p}_2}{m}\cdot\frac{\partial}{\partial \boldsymbol{q}_2} - \frac{\partial \Phi_{12}}{\partial \boldsymbol{q}_1}\cdot\left(\frac{\partial}{\partial \boldsymbol{p}_1} - \frac{\partial}{\partial \boldsymbol{p}_2}\right)\right] f_2 = 0. \tag{2.50}$$

You may wonder if this closure is sufficient to break the time-reversal symmetry of the underlying equations of motion. It is not! At this stage we have the collisions which will in principle allow us to relax to equilibrium, but everything is still time-reversible. So we proceed with a sequence of physically motivated approximations...

Continued simplifications

Let's focus on the evolution of the two-body term. In particular, we expect that most of the changes we are interested in are those that are due to the changes wrought by collisions, rather than the slower changes of evolutions under the effect of the external potential. So, Eqs. 2.49,2.50 we will ignore the terms related to $\frac{\partial U}{\partial \boldsymbol{q}}$. Additionally, we see that in Eq. 2.50 the collision term depends not on absolute positions but on *relative* positions (which makes sense! they're collisions!), so let's switch coordinates to the center of mass, relative position frame (and similarly for momenta):

$$\boldsymbol{R} = \frac{1}{2}(\boldsymbol{q}_1 + \boldsymbol{q}_2), \boldsymbol{r} = (\boldsymbol{q}_1 - \boldsymbol{q}_2), \boldsymbol{P} = (\boldsymbol{p}_1 + \boldsymbol{p}_2), \boldsymbol{p} = \frac{1}{2}(\boldsymbol{p}_1 - \boldsymbol{p}_2).$$

We now have a distribution function $f_2(\boldsymbol{R}, \boldsymbol{r}, \boldsymbol{P}, \boldsymbol{p}, t)$, where the distribution function depends on the center of mass variables, $\boldsymbol{R}, \boldsymbol{P}$ *"slowly"*, and has a much faster dependence on the the relative coordinates $\boldsymbol{r}, \boldsymbol{p}$ which vary over the small distance d and the time scale τ_c.

Since the relative distributions in f_2 vary so quickly, we assume that in a since f_2 reaches equilibrium and then enters the dynamics of f_1. That is, we focus in on time intervals that are long compared to τ_c (but perhaps short compared to τ_U), to get the "steady state" behavior of f_2 at small relative distances that are relevant to the collision term. Combining the approximations in the above paragraph, we have

$$\left(\frac{\boldsymbol{p}}{m}\cdot\frac{\partial}{\partial \boldsymbol{r}} - \frac{\partial \Phi(\boldsymbol{r})}{\partial \boldsymbol{r}}\cdot\frac{\partial}{\partial \boldsymbol{p}}\right) f_2 \approx 0. \tag{2.51}$$

This is the right form to allow us to start massaging the collision term in the RHS of the equation for f_1:

$$\begin{aligned}\left(\frac{\partial f_1}{\partial t}\right)_{coll} = \int dV_2 \frac{\partial \Phi_{12}}{\partial \boldsymbol{q}_1}\cdot\frac{\partial}{\partial \boldsymbol{p}_1} f_2 &= \int dV_2 \frac{\partial \Phi(\boldsymbol{r})}{\partial \boldsymbol{r}}\cdot\left[\frac{\partial}{\partial \boldsymbol{p}_1} - \frac{\partial}{\partial \boldsymbol{p}_2}\right] f_2 \\ &= \frac{1}{m}\int_{|\boldsymbol{r}|\leq d} dV_2 (\boldsymbol{p}_1 - \boldsymbol{p}_2)\cdot\frac{\partial f_2}{\partial \boldsymbol{r}}.\end{aligned} \tag{2.52}$$

In the first equality (where we put in an extra $\frac{\partial}{\partial \boldsymbol{p}_2}$, we're just noting that if we integrate by parts that term vanishes (we've added a derivative of something we're integrating over...), and in the next line we're plugging in the results of Eq. 2.51.

Scattering theory[6]

This part is not crucial to our conceptual discussion, but it is what allows us to massage the above expression into the Boltzmann equation form. Let's think more about classical two-particle collisions, which begin with momenta $\boldsymbol{p}_i = m\boldsymbol{v}_i$ and end with momenta $\boldsymbol{p}'_i = m\boldsymbol{v}'_i$. We proceed to transform into the rest from of the first particle, so that it is being bombarded with oncoming particles that have velocity $\boldsymbol{v}_2 - \boldsymbol{v}_1$, and these oncoming particles are uniformly distributed over the plan normal to that oncoming velocity. We define several relevant quantities in Fig. 2.4. Geometrically, we see that the solid angles are $d\sigma = b\ dbd\phi$ and $d\Omega = \sin\theta\ d\theta d\phi$. The number of particles scattered *into* $d\Omega$ per unit time is related to the flux of particles hitting the plane and the other solid angle, $Id\theta$, typically written $I\frac{d\sigma}{d\Omega}d\Omega = Ibdbd\phi$, where the *differential cross section* is

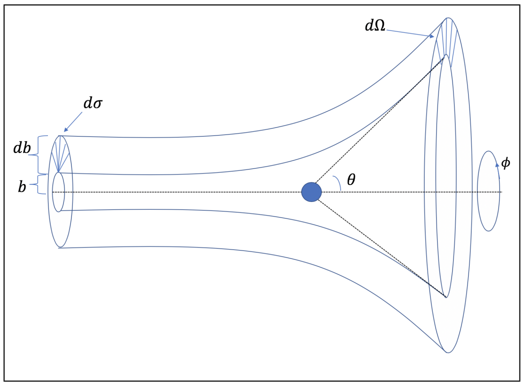


Figure 2.4: **Differential cross section for a scattering process** b is the impact parameter, i.e. the distance from the asymptotic trajectory to the central line, which denotes a head-on collision with the particle (shown as a blue disk here); b and the polar angle ϕ together parameterize the plane normal to the incoming particle. The scattering angle θ is the angle by which the incoming particle is deflected. The solid angles $d\sigma$ and $d\Omega$ are illustrated, with relations between them in the text.

$$\left|\frac{d\sigma}{d\Omega}\right| = \frac{b}{\sin\theta}\left|\frac{db}{d\theta}\right| = \frac{1}{2}\left|\frac{d(b^2)}{d\cos\theta}\right|.$$

What we are really saying here is that for a fixed relative incoming velocity there is a particular relationship between the impact parameter, b, and the scattering angle, θ, and this is something you can figure out for any particular classical pair potential Φ.

---

[6] *very* closely following Tong's notes

If we compare these types of scattering expressions to what we had in the "intuitive" version's expressions Eq. 2.42, we see that when we talked about the rate of scattering *into* some small area of momentum space we can express this in terms of the differential cross section:

$$\omega(\boldsymbol{p}, \boldsymbol{p}_2, \boldsymbol{p}_1', \boldsymbol{p}_2') d^3p_1' d^3p_2' = |\boldsymbol{v}_1 - \boldsymbol{v}_2| \left| \frac{d\sigma}{d\Omega} \right| d\Omega$$

.

Great. Let's go back to our collision integral:

$$\left(\frac{\partial f_1}{\partial t}\right)_{coll} = \frac{1}{m} \int_{|\boldsymbol{r}| \leq d} dV_2(\boldsymbol{p}_1 - \boldsymbol{p}_2) \cdot \frac{\partial f_2}{\partial \boldsymbol{r}}. \tag{2.53}$$

Let's transform to the coordinate system illustrated in Fig. 2.5: the direction parallel to the relative velocity is parameterized by x, we have our interaction range of the potential d, and the plane normal to the relative velocity is still parameterized by ϕ and b. Using all of the above we can write

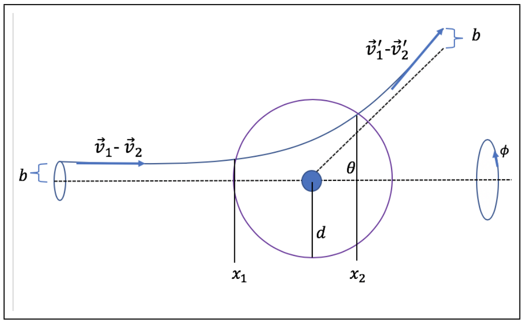


Figure 2.5: **Coordinate system for two-particle collision**

$$\begin{aligned}\left(\frac{\partial f_1}{\partial t}\right)_{coll} &= \int d^3p_2 |\boldsymbol{v}_1 - \boldsymbol{v}_2| \int d\phi db\, b \int_{x_1}^{x_2} \frac{\partial f_2}{\partial x} \\ &= \int d^3p_2 d^3p_1' d^3p_2' \omega(\boldsymbol{p}_1', \boldsymbol{p}_2', \boldsymbol{p}, \boldsymbol{p}_2) \left[f_2(x_2) - f_2(x_1)\right]. \end{aligned} \tag{2.54}$$

Getting to Boltzmann

Working out the classical scattering theory to massage the collision term into the form of Eq. 2.54, all that's left is to decide on the same simplifications for f_2 itself. We once again invoke the assumption of molecular chaos to say that the particles are uncorrelated just before the collision, and we imagine coarse graining over space (on the scale of d) so that we evaluate $f_2(x_1)$ and $f_2(x_2)$ at the same location, $\boldsymbol{q}$. We once again arrive at Eq. 2.46,

$$\frac{\partial f_1}{\partial t} - \{\mathcal{H}_1, f_1\} = \int d^3p_2 d^3p_1' d^3p_2'\ \omega(\boldsymbol{p}_1', \boldsymbol{p}_2' | \boldsymbol{p}, \boldsymbol{p}_2) \left[f_1(\boldsymbol{q}, \boldsymbol{p}_1') f_1(\boldsymbol{q}, \boldsymbol{p}_2') - f_1(\boldsymbol{q}, \boldsymbol{p}) f_1(\boldsymbol{q}, \boldsymbol{p}_2)\right]. \tag{2.55}$$

2.5 The H-Theorem

We are finally ready to begin addressing the second question we asked at the very beginning of the chapter. We have been dancing around the question of how is it that thermodynamics tells us that systems will eventually settle into equilibrium states – which involves an arrow of time that distinguishes past from future – even though the equations of motion are fundamentally invariant under the reversal of time. Specifically, we'll first show that within the framework of the Boltzmann equation, entropy does indeed increase[7] and systems do indeed equilibrate[8].

So what is the H-Theorem, exactly? Let's define a (possibly) time-dependent quantity, $H(t)$, defined by

$$H(t) = \int d^3q d^3p \; f_1(\boldsymbol{q}, \boldsymbol{p}, t) \log\left(f_1(\boldsymbol{q}, \boldsymbol{p}, t)\right). \tag{2.56}$$

Had we not already progressed through Chapters 1 and 2 this might seem like a somewhat strange-looking function, but we instantly recognize the form. f_1 is (possibly up to a factor of normalization) a probability density function, and we recognize H as something like $\langle \log f_1 \rangle$, which is intimately related to the Shannon entropy associated with the probability function (or, from the sign convention, really the *information content*).

H-Theorem: If f_1 satisfies the Boltzmann equation, then

$$\boxed{\frac{dH}{dt} \leq 0,} \tag{2.57}$$

where we have added the frame around the equation because it is a microscopic statement of the increase of entropy with time!

Proof:

We take the time derivative in the same way we took the time derivative of ensemble average quantities earlier in the chapter: H only has a time dependence through explicit time-dependences:

$$\frac{dH}{dt} = \int d^3q d^3p \left(\log f_1 + 1\right) \frac{\partial f_1}{\partial t} = \int d^3q d^3p \log f_1 \frac{\partial f_1}{\partial t}, \tag{2.58}$$

where we exploited the fact that $\int d^3r d^3p f_1 = N$ is independent of time, so $\int \partial_t f_1 = 0$. Using the fact that f_1 satisfies the Boltzmann equation, we can write the above as

$$\frac{dH}{dt} = \int d^3q d^3p \log f_1 \left(\frac{\partial U}{\partial \boldsymbol{q}_1} \cdot \frac{\partial f_1}{\partial \boldsymbol{p}_1} - \frac{\boldsymbol{p}_1}{m} \cdot \frac{\partial f_1}{\partial \boldsymbol{q}_1} + \left(\frac{\partial f_1}{\partial t}\right)_{coll} \right). \tag{2.59}$$

Actually, though, the first two terms in the above equation vanish: to see this we integrate by parts twice, first moving the derivative from f_1 onto the $\log f_1$ term, and then from the

---

[7] this section

[8] next section

$\log f_1$ term back onto the f_1. Thus, in fact, the time evolution of H is entirely governed by the collision terms:

$$\begin{aligned}\frac{dH}{dt} &= \int d^3q d^3p \log f_1 \left(\frac{\partial f_1}{\partial t}\right)_{coll} \\ &= \int d^3q d^3p_1 d^3p_2 d^3p_1' d^3p_2' \omega\left(\boldsymbol{p}_1', \boldsymbol{p}_2' | \boldsymbol{p}_1, \boldsymbol{p}_2\right) \log f_1(\boldsymbol{p}_1) \left[f_1(\boldsymbol{p}_1') f_1(\boldsymbol{p}_2') - f_1(\boldsymbol{p}_1) f_1(\boldsymbol{p}_2)\right]\end{aligned} \tag{2.60}$$

In the above I'm suppressing the $\boldsymbol{q}$ and t arguments, and named the dummy integration variable $\boldsymbol{p}_1$.

To make progress, we play with the dummy indices. First, let's relabel $1 \leftrightarrow 2$, which only changes the argument of the log. Adding the resulting (equivalent) expression and averaging it with the original expression gives a more symmetric expression:

$$\frac{dH}{dt} = \frac{1}{2} \int d^3q d^3p_1 d^3p_2 d^3p_1' d^3p_2' \omega\left(\boldsymbol{p}_1', \boldsymbol{p}_2' | \boldsymbol{p}_1, \boldsymbol{p}_2\right) \log\left[f_1(\boldsymbol{p}_1) f_1(\boldsymbol{p}_2)\right] \left[f_1(\boldsymbol{p}_1') f_1(\boldsymbol{p}_2') - f_1(\boldsymbol{p}_1) f_1(\boldsymbol{p}_2)\right]. \tag{2.61}$$

We can play the same trick with the incoming and outgoing momenta, swapping $\boldsymbol{p} \leftrightarrow \boldsymbol{p}'$, while simultaneously making use of the symmetry properties of the scattering processes[9]. This gives us

$$\frac{dH}{dt} = \frac{-1}{2} \int d^3q d^3p_1 d^3p_2 d^3p_1' d^3p_2' \omega\left(\boldsymbol{p}_1', \boldsymbol{p}_2' | \boldsymbol{p}_1, \boldsymbol{p}_2\right) \log\left[f_1(\boldsymbol{p}_1') f_1(\boldsymbol{p}_2')\right] \left[f_1(\boldsymbol{p}_1') f_1(\boldsymbol{p}_2') - f_1(\boldsymbol{p}_1) f_1(\boldsymbol{p}_2)\right]. \tag{2.62}$$

Finally, we average the above two numbered equations to get

$$\begin{aligned}\frac{dH}{dt} &= \frac{-1}{4} \int d^3q d^3p_1 d^3p_2 d^3p_1' d^3p_2' \omega\left(\boldsymbol{p}_1', \boldsymbol{p}_2' | \boldsymbol{p}_1, \boldsymbol{p}_2\right) \\ &\quad \times \left[\log\left[f_1(\boldsymbol{p}_1') f_1(\boldsymbol{p}_2')\right] - \log\left[f_1(\boldsymbol{p}_1) f_1(\boldsymbol{p}_2)\right]\right] \left[f_1(\boldsymbol{p}_1') f_1(\boldsymbol{p}_2') - f_1(\boldsymbol{p}_1) f_1(\boldsymbol{p}_2)\right].\end{aligned} \tag{2.63}$$

You may feel like this sequence of manipulations – averaging different versions of the same expression together to get a symmetric-looking expression – has done us little good, but we're actually done! Let's think about the terms in the integral. First, the scattering rate, ω, is definitionally a positive quantity. Second, the terms involving f_1 have been manipulated into the form

$$\left(\log a - \log b\right)\left(a - b\right),$$

which, given the positivity of f_1, is always a positive number! Thus,

$$\boxed{\frac{dH}{dt} \leq 0 \quad \Leftrightarrow \quad \frac{dS}{dt} \geq 0}$$

A few comments are in order:

1. The arrow of time, again, emerges from the assumption of molecular chaos. If we had decided that the rate of scattering was proportional to f_2 *after* the collision instead of before, and still kept $f_2 \sim f_1 f_1$, we would have found $\frac{dH}{dt} \geq 0$, suggesting entropy *decreases* as we move into the future. Clearly some real subtleties are in the assumptions we made!

---

[9] $\int d^3p_1' d^3p_2' \omega(p_1', p_2' | p_1, p_2) = \int d^3p_1' d^3p_2' \omega(p_1, p_2 | p_1', p_2')$

2. Note also that the H-theorem permits the time derivative of H to vanish (i.e., it's not a strict inequality). We will see (in the homework) that some distributions satisfy a notion of "local equilibrium" by satisfying a condition of *detailed balance*, making $\frac{dH}{dt}$ vanish by satisfying:
$$f_1(\boldsymbol{p}_1')f_1(\boldsymbol{p}_2') = f_1(\boldsymbol{p}_1)f_1(\boldsymbol{p}_2).$$
These distributions are not quite in equilibrium, as they do not satisfy the streaming terms, but they *do* make the collision terms vanish. These systems have things like densities, temperatures, drift velocities, etc., varying over space. We'll see more about this in the next section.

2.6 Introduction to hydrodynamics[10]

The equilibrium properties of a macroscopic system are governed by thermodynamics, but we said at the outset of this chapter that we also care about, e.g., the common situation shown in Fig. 2.1. What happens if you start we an equilibrium system and perturb it (perhaps in a large way, as by suddenly and radically expanding the volume available for a gas)? *Hydrodynamics* provides a systematic way to think about characteristically long-wavelength, low-energy excitations of a system. Phenomenologically one can write down hydrodynamic equations based on the symmetries of a system, but here (in the context of the Boltzmann equation) we'll see that you can also explicitly derive hydrodynamic descriptions by starting with the microscopic dynamics of a system.

To motivate a bit of what follows, let us think about the *equilibrium condition* in the context of the Boltzmann equation, which is that $\frac{dH}{dt} = 0$. This sets up the following tension, which we will resolve in this section: One way to satisfy this condition is to satisfy the constraints of local equilibrium, writing a candidate one-body distribution which takes the form
$$f(\boldsymbol{p},\boldsymbol{r}) = \exp\left(-\alpha(\boldsymbol{r}) - \beta(\boldsymbol{r})\frac{(\boldsymbol{p}-\boldsymbol{\pi}(\boldsymbol{r}))^2}{2m}\right), \tag{2.64}$$
where α, β, and $\boldsymbol{\pi}$ are functions of the spatial coordinates. A distribution of this form sets the time-derivative of H to zero, but *it does not* satisfy Boltzmann's equation itself! We saw quite generically when discussing the Liouville equation that the left hand side of the Boltzmann equation, even if there is no explicit time dependence, requires the Poisson bracket of the one-particle distribution and the one-particle Hamiltonian to vanish, $\{H_1, f_1\} = 0$; given the set of conserved quantities this tells us that in general f_1 in global equilibrium should just be a function of H_1:
$$f(\boldsymbol{p},\boldsymbol{r}) \propto \exp\left(\beta\left(\frac{p^2}{2m} + U(\boldsymbol{r})\right)\right). \tag{2.65}$$

---

[10] "Hydrodynamics is what you get if you take thermodynamics and splash it." – David Tong. Anyway, there are many references available if you want to go deeper into this subject. The treatment here is kind of a cross between the treatment in Kardar's "Statistical physics of fields" and David Tong's lecture notes on Kinetic Theory, with a tiny splash of Chapter 5 of Huang's "Statistical Mechanics", 2nd edition.

These, in general, are not the same. The key is that the Boltzmann equation is built on a separation of time scales that we can physically interpret. At the fastest, collisional time scale, we approximate $f_2 \sim f_1 f_1$, where there are correlations in these quantities. On time scales related to the mean time between collisions, τ_x, f_1 relaxes to a local equilibrium form, and quantities that are conserved in collisions reach this state of local equilibrium. Finally, there is a subsequent slow relaxation to the global equilibrium state, governed not by the collision terms or intergrals over the collision terms, but by the streaming terms on the LHS of the Boltzmann equation.

2.6.1 Collision-conserved quantities

Let's think about a function over the single-particle phase space, $A(\boldsymbol{r}, \boldsymbol{p})$ – this could be density, or kinetic energy, or... but importantly, we are thinking of quantities that do not have explicit time-dependences. Now, we want to think about the typical way that A varies with space, so we will integrate over momentum. This could be motivated by saying it is more common to experimentally measure spatial dependences than momentum dependences of the kind of systems we're studying, but more relevantly I would say the following: When we started with the Liouville equation there was complete symmetry between $\boldsymbol{p}$ and $\boldsymbol{q}$, but in the derivation of the Boltzmann equation we started treating the two inequivalently. To reflect the fact that $\boldsymbol{p}$ and $\boldsymbol{q}$ are no longer on the same footing, I'll switch from $\boldsymbol{q}$ to $\boldsymbol{r}$ to represent position. So, let's define averages of A as

$$\langle A(\boldsymbol{r}, t)\rangle = \frac{\int d^3p A(\boldsymbol{r}, \boldsymbol{p}) f_1(\boldsymbol{r}, \boldsymbol{p}, t)}{\int d^3p f_1(\boldsymbol{r}, \boldsymbol{p}, t)}.$$

Note that the denominator of the above expression is just a local number density of particles,

$$n(\boldsymbol{r}, t) = \int d^3p f_1(\boldsymbol{r}, \boldsymbol{p}, t), \tag{2.66}$$

so

$$\langle A(\boldsymbol{r}, t)\rangle = \frac{1}{n(\boldsymbol{r}, t)} \int d^3p A(\boldsymbol{r}, \boldsymbol{p}) f_1(\boldsymbol{r}, \boldsymbol{p}, t) \tag{2.67}$$

Importantly, the time-dependence of these average quantities only come through the fact that f_1 can evolve in time.

We're not going to be interested in arbitrary choices of A: ultimately we want to think about the kind of slowly-varying quantities that are relevant as we are approaching equilibrium, and we know from our discussion above that typically terms involved in the collision integral will vary over a fast time scale. So, we want quantities A that will vanish when integrated against the collision part of the Boltzmann equation. I.e., we want A's that have the property

$$\int d^3p A(\boldsymbol{r}, \boldsymbol{p}) \left(\frac{\partial f_1}{\partial t}\right)_{coll} = 0.$$

We can insert the expression for the collision term and go through precisely the same kind of manipulations we did in Sec. 2.5 to find that the A's that have this property obey

$$A(\boldsymbol{r}, \boldsymbol{p}_1) + A(\boldsymbol{r}, \boldsymbol{p}_2) = A(\boldsymbol{r}, \boldsymbol{p}_1') + A(\boldsymbol{r}, \boldsymbol{p}_2'). \tag{2.68}$$

That is, *they are properties that are conserved in the course of a collision.*

Time evolution of collision-conserved quantities: Before we investigate the particular A's of interest, let's write down the general way that A changes with time if f_1 satisfies the Boltzmann equation. To do this, we start with the Boltzmann equation, which we will from now on occasionally write as

$$\mathcal{L}f_1 = C[f_1, f_1], \tag{2.69}$$

where $\mathcal{L} = \partial_t + \frac{p_\alpha}{m}\partial_\alpha + F_\alpha \partial_{p_\alpha}$ and $C[f_1, f_1] = \left(\frac{\partial f_1}{\partial t}\right)_{coll}$. Let's multiply by a collision-invariant $A(\boldsymbol{r}, \boldsymbol{p})$ and integrate both sides $\int d^3p$. The RHS will vanish (by definition of how we're choosing the A), leaving us with

$$\int d^3p\ A(\boldsymbol{r}, \boldsymbol{p}) \left(\frac{\partial}{\partial t} + \frac{p_\alpha}{m}\partial_\alpha + F_\alpha \partial_{p_\alpha}\right) f_1(\boldsymbol{r}, \boldsymbol{p}, t) = 0, \tag{2.70}$$

where $\boldsymbol{F}$ is the external force, $\boldsymbol{F} = -\nabla U$ (and the rest of the notation should be clear... i.e. with the summation convention $\frac{\boldsymbol{p}}{m} \cdot \frac{\partial}{\partial \boldsymbol{r}} = \frac{p_\alpha}{m}\partial_\alpha$). We can simplify this expression (for instance by integrating the term involving the external potential by parts) and various simple manipulations. Making use of how we defined the angle brackets here to be averages over momenta, we can write the above expression as (suppressing dependencies, and with $\boldsymbol{v} = \boldsymbol{p}/m$)

$$\frac{\partial}{\partial t}\langle nA\rangle + \frac{\partial}{\partial \boldsymbol{r}} \cdot \langle n\boldsymbol{v}A\rangle - n\left\langle \boldsymbol{v} \cdot \frac{\partial A}{\partial \boldsymbol{r}}\right\rangle - n\left\langle \boldsymbol{F} \cdot \frac{\partial A}{\partial \boldsymbol{p}}\right\rangle = 0. \tag{2.71}$$

Specific collision-conserved quantities

Let's apply this general expression to quantities which we *know* are conserved in a collision.

Particle number The most trivial thing conserved in a collision between particles is simply *the number of particles*!. If we insert the choice $A = 1$ into Eq. 2.71, we simply get

$$\partial_t n + \partial_\alpha (n u_a) = 0. \tag{2.72}$$

This result is often written using the particle *current*, $\boldsymbol{J}(\boldsymbol{r}, t) = n(\boldsymbol{r}, t)\boldsymbol{u}(\boldsymbol{r}, t)$. It is the *continuity equation*, expressing the fact that if particle number is conserved, variations in local particle density are related to particle currents. Note that we have introduced a new quantity, $\boldsymbol{u} = \langle v\rangle$

Momentum Linear functions of the momentum are also conserved during collisions, and it is convenient to look not at $\boldsymbol{A} = \boldsymbol{p}$, but rather at the momentum relative to the mean local velocity we just defined. Choosing $\boldsymbol{A} = \boldsymbol{c} \equiv \boldsymbol{p}/m - \boldsymbol{u}$ and substituting it into Eq. 2.71, and exploiting the fact that we defined things so that $\langle c_\alpha\rangle = 0$, we find

$$\partial_t u_\alpha + u_\beta \partial_\beta u_\alpha = \frac{F_\alpha}{m} - \frac{1}{mn}\partial_\beta P_{\alpha\beta}, \tag{2.73}$$

where the *pressure tensor* for the fluid is

$$P_{\alpha\beta} \equiv mn\langle c_\alpha c_\beta\rangle. \tag{2.74}$$

This expression is like a generalized Newton's law, telling us that the fluid elements experience accelerations that come both from the external forces and also gradients in the pressure tensor.

Kinetic energy Finally, we look at the kinetic energy of the particles as our last collision-conserved quantity. As before, it is slightly easier to work with the relative kinetic energy, $A = \frac{m}{2}(\boldsymbol{v}-\boldsymbol{u})^2 = \frac{m}{2}\boldsymbol{c}^2$. We substitute this into Eq. 2.71, and go through some simplifications to find

$$\partial_t \epsilon + u_\alpha \partial_\alpha \epsilon = -\frac{1}{n}\partial_\alpha h_\alpha - \frac{1}{n}P_{\alpha\beta}u_{\alpha\beta}, \tag{2.75}$$

where I have just introduced the *average local kinetic energy*:

$$\epsilon \equiv \left\langle \frac{mc^2}{2} \right\rangle,$$

the *local heat flux*:

$$h_\alpha \equiv \frac{nm}{2}\left\langle c_\alpha c^2 \right\rangle,$$

and the *strain rate tensor*:

$$u_{\alpha\beta} = \frac{1}{2}\left(\partial_\alpha u_\beta + \partial_\beta u_\alpha\right).$$

Equations 2.72, 2.73, and 2.75 form a set of coupled equations for the time evolution of the particle density n the local average velocity $\boldsymbol{u}$, and the local average kinetic energy ϵ (which is itself going to be related to the temperature!). But the equations are not closed, as to calculate those three quantities we would need expressions for the pressure tensor and the heat flux. We next show how we can build up approximations, using the Boltzmann equation, to these quantities to finally get a simple, hydrodynamic description of how systems approach equilibrium.

2.6.2 Zeroth-order hydrodynamics

Let's make progress by straight-up guessing a functional form for the one-body distribution function f_1. Our logic in making this guess is that we know the collision term will induce fast relaxations, so if we want a distribution function which only varies slowly, a good place to start would be with a distribution function that satisfies $C[f_1, f_1] = 0$. Let's take one such solution which we already met in Eq. 2.64; getting the normalization correct and introducing a new variable that we'll suspiciously label T, we'll take our zeroth-order guess, denoted f_1^0, to be

$$f_1^0(\boldsymbol{p}, \boldsymbol{r}, t) = \frac{n(\boldsymbol{r}, t)}{\left(2\pi m k_B T(\boldsymbol{r}, t)\right)^{3/2}} \exp\left[-\frac{\left(\boldsymbol{p} - m\boldsymbol{u}(\boldsymbol{r}, t)\right)^2}{2mk_B T(\boldsymbol{r}, t)}\right] \tag{2.76}$$

This Gaussian form is clearly normalized so that, e.g., $\int d^3p f_1^0 = n$, $\langle \boldsymbol{p}/m \rangle^0 = \boldsymbol{u}$, and $\langle c_\alpha c_\beta \rangle^0 = \frac{k_B T}{m}\delta_{\alpha\beta}$. This lets us calculate the approximations for the pressure tensor, energy density, and heat flux:

$$P^0_{\alpha\beta} = nk_B T\delta_{\alpha\beta}, \quad \epsilon = \frac{3}{2}k_B T, \quad \boldsymbol{h}^0 = 0.$$

The equations we derived by thinking about collision conserved quantities are very simple in this approximation. Defining the *material derivative*

$$D_t \equiv \partial_t + u_\beta \partial_\beta$$

we get:

$$\begin{aligned} D_t n &= -n\partial_\alpha u_\alpha \\ mD_t u_\alpha &= F_\alpha - \frac{1}{n}\partial_\alpha(nk_BT) \\ D_tT &= -\frac{2}{3}T\partial_\alpha u_\alpha. \end{aligned} \tag{2.77}$$

The inadequacy of zeroth-order hydrodynamics: Sadly, these very simple equations do a terrible job of describing the relaxation of a system to equilibrium. Let's imagine starting with a system with initial $\boldsymbol{u}_0 = 0$ in the absence of external forces $\boldsymbol{F} = 0$, and making a small perturbation:

$$n(\boldsymbol{r},t) = \bar{n} + \delta n(\boldsymbol{r},t), \quad T(\boldsymbol{r},t) = \bar{T} + \delta T(\boldsymbol{r},t). \tag{2.78}$$

We want to study what happens to these small perturbations, so we expand Eqs. 2.77 to first order in $(\delta n, \delta T, \boldsymbol{u})$, where we note that to first order the material derivative is just $D_t = \partial_t + \mathcal{O}(u)$, so our linearized zeroth order equations become:

$$\begin{aligned} \partial_t\delta n &= -\bar{n}\partial_\alpha u_\alpha \\ m\partial_t u_\alpha &= -\frac{k_B\bar{T}}{\bar{n}}\partial_\alpha\delta n - k_B\partial_\alpha\delta T \\ \partial_t\delta T &= -\frac{2}{3}\bar{T}\partial_\alpha u_\alpha. \end{aligned} \tag{2.79}$$

The easiest way to investigate the effect of our perturbations is to take Fourier transforms,

$$\tilde{A}(\boldsymbol{k},\omega) = \int d^3q dt\ A(\boldsymbol{r},t)\exp\left[i(\boldsymbol{k}\cdot\boldsymbol{r} - \omega t)\right],$$

where A is any of $(\delta n, \delta T, \boldsymbol{u})$. This gives us the matrix equation

$$\omega\begin{pmatrix} \tilde{\delta n} \\ \tilde{u}_\alpha \\ \tilde{\delta T}\end{pmatrix} = \begin{pmatrix} 0 & \bar{n}k_\beta & 0 \\ \frac{k_B\bar{T}}{m\bar{n}}\delta_{\alpha\beta}k_\beta & 0 & \frac{k_B}{m}\delta_{\alpha\beta}k_\beta \\ 0 & \frac{2}{3}\bar{T}k_\beta & 0 \end{pmatrix}\begin{pmatrix} \tilde{\delta n} \\ \tilde{u}_\beta \\ \tilde{\delta T}\end{pmatrix}. \tag{2.80}$$

This matrix has the following modes: There are two modes describing a transverse shear flow in a uniform and isothermal gas ($n = \bar{n}, T = \bar{T}$), with the velocity field varying transverse to its orientation, e.g. $\boldsymbol{u} = f(x,t)\hat{y}$, and both have $\omega = 0$. There is another $\omega = 0$ mode describing a gas with uniform pressure nk_BT, where n and T may be spatially varying but with a constant product. Lastly, there is a mode with variations along the direction of $\boldsymbol{k}$; the eigenvector looks like:

$$\boldsymbol{v}_{longitudinal} = \begin{pmatrix} \bar{n}|\boldsymbol{k}| \\ \omega(\boldsymbol{k}) \\ \frac{2}{3}\bar{T}|\boldsymbol{k}| \end{pmatrix}, \qquad \omega(\boldsymbol{k}) = \pm\sqrt{\frac{5k_B\bar{T}}{3m}}|\boldsymbol{k}|.$$

Well, shoot. Apparently within this approximation *none* of our conserved quantities relax to equilibrium if we apply a little perturbation: shear flow persists forever, sound modes have undamped oscillations, etc.

2.6.3 First-order hydrodynamics

Perhaps this should not have surprised us: we picked something that satisfied local equilibrium, but it is straightforward to check that $\mathcal{L} f_1^0 \neq 0$. Actually, it's more straightforward to show that if n, T, u_α satisfy the zeroth-order hydrodynamic equations, then the effect of $\mathcal{L}$ on the log of the zeroth-order approximation for f_1 is

$$\mathcal{L}\left[\ln f_1^0\right] = \frac{m}{k_B T}\left(c_\alpha c_\beta - \frac{\delta_{\alpha\beta}}{3}c^2\right)u_{\alpha\beta} + \left(\frac{mc^2}{2k_B T} - \frac{5}{2}\right)\frac{c_\alpha}{T}\partial_\alpha T. \tag{2.81}$$

Our instinct for moving forward is to note that, although the RHS of the above is not zero, it depends on gradients of temperature and on the local velocity. In a sense, then, if we stick to long-wavelength variations in T and $\boldsymbol{u}$ we are "close" to a solution. Thus, we will try to construct first-order hydrodynamics by adding a little something extra to the distribution:

$$f_1^1 = f_1^0 + \delta f_1. \tag{2.82}$$

Relaxation time approximation What happens if we act on f_1^1 with our collision operator?

$$\begin{aligned} C[f_1^1, f_1^1] &= \int d^3p_2 d^3p_1' d^3p_2' \omega(\boldsymbol{p}_1', \boldsymbol{p}_2' | \boldsymbol{p}_1, \boldsymbol{p}_2)\left[f_1(\boldsymbol{p}_1')f_1(\boldsymbol{p}_2') - f_1(\boldsymbol{p}_1)f_1(\boldsymbol{p}_2)\right] \\ &= \int d^3p_2 d^3p_1' d^3p_2' \omega(\boldsymbol{p}_1', \boldsymbol{p}_2' | \boldsymbol{p}_1, \boldsymbol{p}_2)\left[f_1^0(\boldsymbol{p}_1')\delta f_1(\boldsymbol{p}_2') + \delta f_1(\boldsymbol{p}_1')f_1^0(\boldsymbol{p}_2') - f_1^0(\boldsymbol{p}_1)\delta f_1(\boldsymbol{p}_2) - \delta f_1(\boldsymbol{p}_1)f_1^0\right. \end{aligned}$$

where we have used the fact that f_1^0 vanishes in the collision integral and dropped any terms of order $(\delta f_1)^2$. We now have a linear function of δf_1, albeit a messy one to work with. At this point there is a proper way to proceed[11], and a physically just fine way to proceed, which has the great virtue of being much easier while still capturing the dissipational piece that was missing from our zeroth-order description. We simply approximate

$$C[f_1^1, f_1^1] \approx \frac{-\delta f_1}{\tau_x}, \tag{2.84}$$

which is called the *relaxation time approximation* or the *single collision time approximation* or using the *BGK operator*[12]. With this approximation the Boltzmann equation is

$$\frac{\partial f_1^0 + \delta f_1}{\partial t} - \{H_1, f_1^0 + \delta f_1\} = \frac{-\delta f_1}{\tau_x}, \tag{2.85}$$

but we assume that $\delta f_1 \ll f_1^0$, so we ignore the δf_1 on the LHS. We can then explicitly work out the additional contribution to our improved estimate:

$$\delta f_1 = -\tau_x\left[\left(\frac{mc^2}{2k_B T} - \frac{5}{2}\right)\frac{\boldsymbol{c}}{T}\cdot\frac{\partial T}{\partial \boldsymbol{r}} + \frac{m}{k_B T}\left(c_\alpha c_\beta - \frac{\delta_{\alpha\beta}}{3}c^2\right)u_{\alpha\beta}\right]f_1^0. \tag{2.86}$$

---

[11] The Chapman-Enskog expansion, doing a careful expansion in δf_1

[12] Bhatnagar-Gross-Krook, 1954

One can use this improved description to calculate corrections to various quantities. For example, the first-order estimate of the pressure tensor becomes

$$P^1_{\alpha\beta} = nk_BT\delta_{\alpha\beta} - 2nk_BT\tau_x\left(u_{\alpha\beta} - \frac{\delta_{\alpha\beta}}{3}u_{\gamma\gamma}\right), \tag{2.87}$$

and the heat flux acquires a dependence on spatial gradients in the temperature. These are important: they say that shear flows get opposed by off-diagonal terms in the pressure tensor, and that spatial variations in temperature generate heat flows that in turn smooth out those variations! These are the sorts of effects that cause the relaxation to equilibrium.

In case you're curious, if I've TEX'ed this correctly the matrix equation corresponding to the first-order hydrodynamic equations after a Fourier transformation look like

$$\omega\begin{pmatrix}\tilde{\delta n}\\ \tilde{u}_\alpha\\ \tilde{\delta T}\end{pmatrix} = \begin{pmatrix} 0 & \bar{n}\delta_{\alpha\beta}k_\beta & 0\\ \frac{k_B\bar{T}}{m\bar{n}}\delta_{\alpha\beta}k_\beta & -i\frac{\mu}{m\bar{n}}\left(k^2\delta_{\alpha\beta} + \frac{k_\alpha k_\beta}{3}\right) & \frac{k_B}{m}\delta_{\alpha\beta}k_\beta\\ 0 & \frac{2}{3}\bar{T}\delta_{\alpha\beta}k_\beta & -i\frac{2Kk^2}{3k_B\bar{n}}\end{pmatrix}\begin{pmatrix}\tilde{\delta n}\\ \tilde{u}_\beta\\ \tilde{\delta T}\end{pmatrix}, \tag{2.88}$$

where $K = (5\bar{n}k_B^2\bar{T}\tau_x)/(2m)$ and $\mu = \bar{n}k_B\bar{T}\tau_x$ is a viscosity coefficient. The important point of writing this is the ability to verify that now all of the modes have an imaginary component (either they are strictly imaginary eigenvalues (for variations in pressure and for transverse shear modes) or complex ones (for longitudinal sound modes), so that we know that over long time scales perturbations to the gas die away, and the gas eventually reaches its equilibrium state.

You'll go through a simple example of first-order hydrodynamics in the homework!

Chapter 3

Classical Statistical Mechanics[1]

Statistical mechanics is connected with the phenomenological, "thermodynamical" view of macroscopic properties we saw in Chapter 1 via a *probabilistic description of large numbers of degrees of freedom.* In this chapter we will focus not on microscopic theories for which we can study both equilibrium and approach-to-equilibrium dynamics as in Chapter 3, but rather on attempting to provide probability distributions that connect microstates to macrostates. We will use the idea of unbiased estimates of probability to assign these probability distributions for different equilibrium ensembles, and use the mathematics of the large-N limit to show that the ensembles are equivalent in the thermodynamic limit.

3.1 The microcanonical ensemble and the laws of thermodynamics

We begin, just as we did in the chapter on Thermodynamics, with a simplified version of our system of interest, taking an adiabatically isolated state. In the absence of adding heat or work to the system, the *macrostate* M is specified completely by the internal energy E, the set of generalized coordinates $\boldsymbol{x}$, and the number of particles N: $M(E, \boldsymbol{x}, N)$. In the absence of any other information, or any knowledge of other conserved quantities, we say that at a minimum the Hamiltonian evolution equations conserve the total energy of the system, so that the evaluation of the Hamiltonian on a microstate μ is $\mathcal{H}(\mu) = E$.

The central postulate of statistical mechanics is that the equilibrium probability distribution is

$$\boxed{p_{(E,\boldsymbol{x},N)}(\mu) = \frac{1}{\Omega(E,\boldsymbol{x},N)} \cdot \begin{cases} 1 & \text{if } \mathcal{H}(\mu) = E \\ 0 & \text{otherwise} \end{cases}} \tag{3.1}$$

This is often called "the assumption of equal a priori probabilities," and we see that it is the same as the unbiased estimate of probability given only a constraint of constant energy E. Certainly we saw that this is one of the allowed assignments consistent with Liouville's

---

[1] In this chapter we'll basically go through Pathria Chapters 2-4... because of the way we started this course, though, we've already done much of the work!

theorem, although it is not the only one! Although, having seen the chapter on probability, we are not surprised by this assignment of probability it is nevertheless a deep assumption,

There is a potential subtlety in defining the normalization factor[2] Ω here: as written, in order to make p a properly normalized probability density function (so that it integrates to unity), we want Ω to be the area of the relevant surface of fixed energy E over the microscopic phase space. You might be worried about defining probability densities that are non-zero only on a surface, and so we sometimes *define* a microstate in the microcanonical ensemble to be within Δ of the specified energy: $E - \frac{1}{2}\Delta \leq \mathcal{H}(\mu) \leq E + \frac{1}{2}\Delta$. The normalization Ω' is now the volume of a shell rather than the area of a surface, and $\Omega' \approx \Omega\Delta$. We will see that, since Ω is typically exponentially large in E, which is itself typically proportional to N, that the difference between the surface area and the shell volume is negligible, so we'll go back and forth between Ω and Ω' freely.

We define the entropy of the uniform probability distribution exactly as you by now expect:

$$S(E, \boldsymbol{x}, N) = k_B \ln \Omega(E, \boldsymbol{x}, N), \tag{3.2}$$

where we have introduce a factor called "k_B" so that entropy has units of energy divided by temperature. Note, by the way, that we know from Liouville that under *canonical transformations* volumes in phase space are invariant, and for the transformed probability distribution stays uniform on the transformed phase-space surface of constant energy. This tells us that both Ω and S are invariant under canonical coordinate changes.

To highlight the interconnections with the beginning of this class, we now show that Eq. 3.1 can be used to derive the laws of thermodynamics (with the exception of Nernst's theorem, which as we hinted at the time requires quantum mechanics).

3.1.1 0th Law

Let's think about bringing two previously isolated microcanonical systems, which originally had energies E_1 and E_2, into contact in a way that lets them exchange *heat*, but not *work*. Certainly the combined system has energy $E = E_1 + E_2$, and we assume (by assuming that the interactions between the systems are small) that the microstate of the combined system corresponds to a pair of microstates of the components. We'll write this as $\mu = \mu_1 \otimes \mu_2$, assuming $\mathcal{H}(\mu) = \mathcal{H}_1(\mu_1) + \mathcal{H}_2(\mu_2)$. We thus write the fundamental postulate, Eq. 3.1 as

$$p_E(\mu) = \frac{1}{\Omega(E)} \cdot \begin{cases} 1 & \text{if } \mathcal{H}_1(\mu_1) + \mathcal{H}_2(\mu_2) = E \\ 0 & \text{otherwise} \end{cases}. \tag{3.3}$$

We have a fixed total energy, so we can compute the normalization factor as

$$\Omega(E) = \int dE_1\ \Omega_1(E_1)\Omega_2(E - E_1) = \int dE_1\ \exp\left[\frac{S_1(E_1) + S_2(E_2)}{k_B}\right]. \tag{3.4}$$

We written the normalization factor for our probability this way to make contact with our "sums and integrals of exponentials" discussion in a previous chapter! We think (and

---

[2] "we must learn how to count the number of states it is possible for a system to have or, more precisely, how to avoid having to count that number" – David Goodstein, *States of Matter*

will see later) that entropy is extensive, so that S_1 and S_2 are proportional to the number of particles in the system. We use the simple (i.e., non-complex) saddle-point method to approximate the integral by the maximum of the integrand, giving us

$$S(E) = k_B \ln \Omega(E) \approx S_1(E_1^*) + S_2(E_2^*), \tag{3.5}$$

where the starred quantities are the values that maximize the value of the integrand. We can calculate this values of setting the first derivative (w/r/t E_1) of the exponent above to zero, which gives us:

$$\left.\frac{\partial S_1}{\partial E_1}\right|_{\boldsymbol{x}_1, N_1} = \left.\frac{\partial S_2}{\partial E_2}\right|_{\boldsymbol{x}_2, N_2}. \tag{3.6}$$

In words, there are many joint microstates which sum to a total energy of E, but there are exponentially *more* microstates sitting near (E_1^*, E_2^*), and so the system eventually flows from (E_1, E_2) to (E_1^*, E_2^*). We get no information about the time dynamics of this process; just the end result, which is a place where Eq. 3.6 is satisfied.

Also, ah ha!! We have found a condition satisfied by two systems that come into equilibrium with each other: the have equal partial derivatives of entropy with respect to energy (holding everything else fixed)! To be consistent with what we had from phenomenological thermodynamics, we'll choose a particular name for the state function corresponding to this condition:

$$\left.\frac{\partial S}{\partial E}\right|_{\boldsymbol{x}} = \frac{1}{T}. \tag{3.7}$$

3.1.2 1st Law

Let's think about what happens to S of a system in the microcanonical ensemble when we vary the generalized displacements[3] by $\delta \boldsymbol{x}$ This does a work on the system $đW = \boldsymbol{J} \cdot \delta \boldsymbol{x}$, and changes the internal energy to $E + đW$. To first order, the change in entropy of such a variation is

$$\delta S = S(E + \boldsymbol{J} \cdot \delta \boldsymbol{x}, \boldsymbol{x} + \delta \boldsymbol{x}) - S(E, \boldsymbol{x}) = \left(\left.\frac{\partial S}{\partial E}\right|_{\boldsymbol{x}} \boldsymbol{J} + \left.\frac{\partial S}{\partial \boldsymbol{x}}\right|_{E} \right) \cdot \delta \boldsymbol{x}. \tag{3.8}$$

Now, what does it mean to be in equilibrium? We say we have an equilibrium state at some value of E and some set of generalized displacements $\boldsymbol{x}$, and we have said that all consistent microstates are equally probable. Well, the above equation says that spontaneous changes in the system will occur, taking us into more likely states, *unless the terms in the parentheses vanish*! Thus, a condition to be in equilibrium is that

$$\left.\frac{\partial S}{\partial x_i}\right|_{E, x_{j \neq i}} = -\frac{J_i}{T}, \tag{3.9}$$

where we have used the relationship between temperature and $\partial S / \partial E$ from the zeroth law. With this constraint on the variations, we get that in general

$$dS(E, \boldsymbol{x}) = \frac{\partial S}{\partial E} dE + \frac{\partial S}{\partial \boldsymbol{x}} d\boldsymbol{x} = \frac{dE}{T} - \frac{\boldsymbol{J} \cdot d\boldsymbol{x}}{T} \Rightarrow dE = T dS + \boldsymbol{J} \cdot d\boldsymbol{x}. \tag{3.10}$$

---

[3] for the remainder of this section I'm going to be TEX-lazy and stop distinguishing mechanical work from chemical work.... it's all just $\boldsymbol{x}$ now.

Wow – it's the first law! And where we've, of course, identified $đQ = TdS$.

3.1.3 2nd Law[4]

The second law is almost obvious, almost by construction, given what we have set up so far. Indeed, you already know from the section of unbiased estimates that we are assigning probabilities in a way that *maximizes the entropy* subject to the constraint on the energy. For instance, consider our example from the zeroth law of bringing two equilibrium states at E_1 and E_2 into contact. Well, clearly

$$S(E) \equiv k_B \ln \Omega(E_1 + E_2) \geq S_1(E_1) + S_2(E_2),$$

which must be true since the two states of the original system are a *subset of the possible combined joint microstates.*

Note that we can make additional mathematical statements by considering variations of entropy. When two systems are first brought into contact but have not yet reached equilibrium, the equality in Eq. 3.6 does not yet hold. Instead we have

$$\delta S = \left(\left.\frac{\partial S_1}{\partial E_1}\right|_{\boldsymbol{x}_1} - \left.\frac{\partial S_2}{\partial E_2}\right|_{\boldsymbol{x}_2} \right) \delta E_1 = \left(\frac{1}{T_1} - \frac{1}{T_2} \right) \delta E_1 \geq 0. \tag{3.11}$$

Thus we recover Clausius' statement of the second law: we know the variations in S are positive as we move towards a new equilibirum state, and we see that heat flows from the hotter to the colder system.

Note that, in principle, these are all probabilistic statements: it is merely *much more likely* that a combined system ends up at (E_1^*, E_2^*) rather than it's initial and (E_1, E_2). This obscures just how much work "much" is doing in the previous sentence, though: the number of microstates available grows exponentially with system size, and so if we were to ask how long we would have to wait before seeing our combined system at (E_1, E_2) rather than (E_1^*, E_2^*), the answer would be related to an exponential of that exponential. Needing to wait this long to see something is a practical definition of saying "you will never see that."

3.1.4 The ideal gas in the microcanonical ensemble

We return to our favorite toy system for illustrating concepts, the ideal gas of N particles. We ignore any particle interactions, and say we have

$$\mathcal{H} = \sum_{i=1}^{N} \frac{\boldsymbol{p}_i^2}{2m} + U(\boldsymbol{q}_i),$$

---

[4] "The law that entropy always increases holds, I think, the supreme position among the laws of Nature. If someone points out to you that your pet theory of the universe is in disagreement with Maxwell's equations - then so much the worse for Maxwell's equations. If it is found to be contradicted by observation - well, these experimentalists do bungle things sometimes. But if your theory is found to be against the Second Law of Thermodynamics I can give you no hope; there is nothing for it to collapse in deepest humiliation." – Arthur Eddington, *New Pathways in Science*

where the potential U simply imposes strict confinement to a box of volume V (i.e., we take $U(\boldsymbol{q}_i) = 0$ if particle i is inside the box and ∞ if it is outside the box). Explicitly, the microcanonical ensemble has a probability density function

$$p(\mu) = \frac{1}{\Omega(E,V,N)} \cdot \begin{cases} 1 & \text{if } \sum_i (\boldsymbol{p}_i^2/(2m)) = E \pm \frac{1}{2}\Delta_E \text{ and } \boldsymbol{q}_i \in \text{box} \\ 0 & \text{otherwise} \end{cases} . \tag{3.12}$$

We can calculate Ω by the requirement that $p(\mu)$ is properly normalized, i.e., $\int \prod_i dV_i p(\mu) = 1$. The integrals over $\boldsymbol{q}_i$ are all trivial: each of those N integrals just gives a factor of V. The integral over the momenta is down by noting we are constraining the momenta to a (finite-thickness shell around a) surface of a hypersphere, given by $\sum_i^N \boldsymbol{p}_i^2 = 2mE$. We thus need to know the area of a $3N$-dimensional sphere of radius $R = \sqrt{2mE}$. The relevant formula for a d-dimensional sphere is

$$\mathcal{A}_d = \frac{2\pi^{d/2}}{(d/2-1)!} R^{d-1}, \tag{3.13}$$

where it is easy to check that this gives the right result in $2D$ (and $3D$, since $(1/2)! = \sqrt{\pi}/2$). Putting this together, with $d = 3N$, we calculate:

$$\Omega(E,V,N) = V^N \frac{2\pi^{3N/2}}{(3N/2-1)!} (2mE)^{(3N-1)/2} \Delta_E \tag{3.14}$$

The entropy of the ideal gas is then just the log of this normalization factor. Using Stirling's approximation:

$$\begin{aligned} S(E,V,N) &= k_B \left[N \ln V + \frac{3N}{2} \ln(2\pi m E) - \frac{3N}{2} \ln \frac{3N}{2} + \frac{3N}{2} \right] \\ &= N k_B \ln \left[V \left(\frac{4\pi e m E}{3N} \right)^{3/2} \right], \end{aligned} \tag{3.15}$$

where in the first line we have dropped terms of order 1, and of order $\ln E \propto \ln N$ by arguing the latter are small compared to these terms of order N in the thermodynamic limit.

With the entropy in hand, and writing $dS = \frac{1}{T}dE + \frac{P}{T}dV - \frac{\mu}{T}dN$ we can get the usual properties of the ideal gas by differentiating the entropy as appropriate. For instance,

$$\frac{1}{T} = \left.\frac{\partial S}{\partial E}\right|_{N,V} = \frac{3}{2}\frac{N k_B}{E} \Rightarrow E = \frac{3}{2} N k_B T,$$

the usual equipartition result in the absence of a potential. Similarly,

$$\frac{P}{T} = \left.\frac{\partial S}{\partial V}\right|_{N,E} = \frac{N k_B}{V} \Rightarrow PV = N k_B T,$$

the ideal gas equation of state.

As a final example of a simple calculation, what about the probability distribution for finding a particle with some momentum vector $\boldsymbol{p}_1$? Well, we can calculate the unconditional

PDF by integrating out everything else:

$$\begin{aligned} p(\boldsymbol{p}_1) &= \int d^3\boldsymbol{q}_1 \prod_{i=2}^{N} dV_i p(\mu) \\ &= V\frac{\Omega(E-\frac{\boldsymbol{p}_1}{2m}, V, N-1)}{\Omega(E,V,N)}. \end{aligned} \tag{3.16}$$

One can explicitly check (by plugging in the result of Eq. 3.14, using Stirling's formula, and noting that $\boldsymbol{p}_1^2/E$ is a number of order $1/N$) that this explicitly gives you a Maxwell-Boltzmann distribution for the momentum:

$$p(\boldsymbol{p}_1) = \frac{1}{(2\pi m k_B T)^{3/2}} \exp\left(\frac{-\boldsymbol{p}_1^2}{2mk_BT}\right)$$

We derived the Maxwell-Boltzmann distribution result for the velocities in a dilute gas of non-interacting particles, but it is actually much more general, and Maxwell's original derivation relies on nothing by rotational invariance. It's cool, so I reproduce it here: Consider the distribution of velocities in, say, the x direction, and call that distribution $p(v_x)$. By rotational symmetry we have the same distribution in the y and z directions. Rotational symmetry *also* guarantees that the full distribution cannot depend on the particular direction for the momentum, but only on the speed $c = \sqrt{\boldsymbol{v}\cdot\boldsymbol{v}}$. So, we want functions $p_c(c)$ and $p(v_x)$ that satisfy

$$p_c(c)dv_xdv_ydv_z = p(v_x)p(v_y)p(v_z)dv_xdv_ydv_z.$$

Remarkably, there is only one solution which satisfies this, and it is

$$p(v_x) = A\exp\left(-Bv_x^2\right),$$

for some constants A and B. Thus, the distribution of speeds must be

$$p_c(c)dv_xdv_ydv_z = 4\pi c^2 p_c(c)dc = 4\pi A^3 c^2 e^{-Bc^2}dc.$$

Nifty. Determining that the coefficients have specific values, like $B = \frac{m}{2k_BT}$ as in the ideal gas, requires a harder microscopic calculation of the sort we just did.

3.1.5 Gibbs' Paradox: What's up with mixing entropy?

You will notice that Eq. 3.15, giving the entropy of an ideal gas, has a major flaw: we expect entropy to be extensive, but under a transformation $(E,V,N)\rightarrow(\lambda E,\lambda V,\lambda N)$ the entropy actually changes to $\lambda(S+Nk_B\ln\lambda)$ rather than λS. There are these extra contributions that come from the integration over positions, like V^N, and this additional contribution is related to the entropy of mixing distinct gases. Gibbs' paradox, the fact that this expression for the entropy suggests an increase even if two *identical* gases are mixed, is subtle, with some to this day arguing that its resolution must be quantum mechanical[5] and others arguing that

---

[5] see, e.g., Kardar's "Statistical Physics of Particles"

the paradox is toothless and can be resolved classically[6]. We will come back to this issue when we talk about quantum statistical mechanics; I would be delighted if one of you wanted to discuss the classical issue as an in-class presentation (as described on the syllabus).

As our yet-to-be-understood resolution to the paradox, we will from now on modify our phase-space measure for identical particles to be

$$d\Gamma = \frac{1}{h^{3N}N!}\prod_{i=1}^{N} d^3\boldsymbol{q}_i d^3\boldsymbol{p}_i. \tag{3.17}$$

3.2 The canonical ensemble

In the last section we defined the microcanonical ensemble by considering a macrostate with a specified energy, E, and we were able to derive an expression for the temperate as $T^{-1} = \frac{\partial S}{\partial E}\big|_{\boldsymbol{x}}$.In the chapter on thermodynamics, though, we saw that E and T are both functions of state, and that it should be possible to instead *start* with statistical description in which the temperature of a macrostate is prescribed and an expression for the energy is derived.

Welcome to the canonical ensemble, in which we specify the macrostate $M(T, \boldsymbol{x})$. We think of prescribing the temperature of the system by putting it in thermal contact with *the reservoir*, which is another macroscopic system that is so large that it's own temperature does not change as a result of interacting with the system of interest (think, for instance, of tossing a warm pebble into the ocean – the temperature of the pebble will surely equilibrate with the temperature of the ocean, and no thermometer reading the temperature of the ocean will notice the difference).

The goal of the statistical mechanical formulation of thermodynamics is to write down microstate probabilities that we can associate with macrostates, $p(\mu)$ (and from there derive thermodynamic consequences), and *here, too, we rely on the central postulate of statistical mechanics, Eq. 3.1!* Let's define the system to be in microstate μ with energy $\mathcal{H}_S(\mu)$ and the reservoir to be in microstate μ_R with energy $\mathcal{H}_R(\mu_R)$, satisfying $E_{total} = \mathcal{H}_S(\mu)+\mathcal{H}_R(\mu_R)$.The joint probability of the microstates $\mu_S \otimes \mu_R$ is then assumed to be

$$\boxed{p(\mu_S \otimes \mu_R) = \frac{1}{\Omega_{S\oplus R}(E_{total})} \cdot \begin{cases} 1 & \text{if } \mathcal{H}_S(\mu) + \mathcal{H}_R(\mu_R) = E_{total} \\ 0 & \text{otherwise} \end{cases}} \tag{3.18}$$

Ultimately, though, we are not interested in this microcanonical joint-probability distribution for the combined system, we want the unconditional PRF for the microstates. But this we can obtain (as expected) by summing the joint PDF over the microstates of the reservoir:

$$p(\mu) = \sum_{\mu_R} p_{(}\mu_S \otimes \mu_R) \tag{3.19}$$

---

[6] see Jayne's discussion, or that of Frenkel's more recent article

We make progress, here, by doing something similar to how we calculated $p(\boldsymbol{p}_1)$ in the microcanonical ensemble in Eq. 3.16: we say that by looking at a specific system microstate μ, the sum over reservoir states in the above equation is restricted to run over microstates with $\mathcal{H}_R(\mu_R) = E_{total} - \mathcal{H}_S(\mu)$:

$$p(\mu_S \otimes \mu_R) = \frac{\Omega_R(E_{total} - \mathcal{H}_S(\mu))}{\Omega_{S\oplus R}(E_{total})}. \tag{3.20}$$

Let's focus on the numerator for now (noting that the denominator, which just provides an overall normalization, can always be effectively recovered by imposing a normalization condition of $p(\mu)$ when we're done). Well, Ω_R is related to the entropy of the reservoir:

$$\frac{\Omega_R(E_{total} - \mathcal{H}_S(\mu))}{\Omega_{S\oplus R}(E_{total})} \propto \exp\left(\frac{S_R(E_{total} - \mathcal{H}_S(\mu))}{k_B}\right). \tag{3.21}$$

Let's write that entropy expression as

$$S_R\left(E_{total}\left(1 - \frac{\mathcal{H}_s(\mu)}{E_{total}}\right)\right) \approx S_R(E_{total}) - \mathcal{H}_s(\mu)\frac{\partial S_R}{\partial E_R} = S_R(E_{total}) - \frac{\mathcal{H}_S(\mu)}{T},$$

where we have finally used our assumption that the *reservoir is (energetically) humongous relative to the system.* Combining this approximation with our evaluation of the unconditional PDF, we ultimately arrive at the canonical probability distribution for a microstate with a specified temperature and set of generalized displacements:

$$p(\mu) = \frac{1}{Z(T,\boldsymbol{x})}e^{-\beta\mathcal{H}(\mu)}, \tag{3.22}$$

where I've dropped subscripts, introduced $\beta = (k_BT)^{-1}$, and defined the *partition function*

$$\boxed{Z(T,\boldsymbol{x}) = \sum_\mu e^{-\beta\mathcal{H}(\mu)}}. \tag{3.23}$$

3.2.1 The partition function as a generator of moments

So, we've specified the temperature of the system, and it is exchanging energy back and forth with the reservoir to maintain that temperature. The energy of the system is now a random variable, which we'll call $\mathcal{E}$ so that the notation doesn't get too confusing. What is the probability distribution associated with $\mathcal{E}$? Glad you asked! We change variables from μ to $\mathcal{H}(\mu)$ and get

$$p(\mathcal{E}) = \sum_\mu p(\mu)\delta\left(\mathcal{H}(\mu) - \mathcal{E}\right) = \frac{1}{Z}e^{-\beta\mathcal{E}}\sum_\mu \delta\left(\mathcal{H}(\mu) - \mathcal{E}\right),$$

but the sum over delta functions just picks out the number of microstates with the appropriate energy $\Omega(\mathcal{E})$, which is related to the entropy, so

$$p(\mathcal{E}) = \frac{\Omega(\mathcal{E})e^{-\beta\mathcal{E}}}{Z} = \frac{1}{Z}\exp\left[\frac{S(\mathcal{E})}{k_B} - \frac{\mathcal{E}}{k_BT}\right] = \frac{1}{Z}\exp\left[-\beta F(\mathcal{E})\right], \tag{3.24}$$

where $F = \mathcal{E} - TS(\mathcal{E})$ is obviously going to be related to the Helmholtz free energy. We can further simplify by noting that both S and E should be extensive, so we should be able to approximate sums of exponentials by the dominant term... we expect the probability to be sharply peaked about some most probably energy, E^*, and so we approximate the partition function as

$$Z = \sum_{\mu} e^{-\beta\mathcal{H}(\mu)} = \sum_{\mathcal{E}} e^{-\beta F(\mathcal{E})} \approx e^{-\beta F(E^*)}. \tag{3.25}$$

So, the *method of most probable values* suggests a logarithmic relationship between F and Z. This is reinforced by the *method of mean values*, where we compute the average energy of the system as

$$\langle\mathcal{H}\rangle = \sum_{\mu} \mathcal{H}(\mu)\frac{\exp\left(-\beta H(\mu)\right)}{Z} = -\frac{1}{Z}\frac{\partial}{\partial\beta}\sum_{\mu} e^{-\beta\mathcal{H}} = -\frac{\partial \ln Z}{\partial\beta}. \tag{3.26}$$

This, too, suggests the identification of the Helmholtz free energy as

$$F(T, \boldsymbol{x}) = -k_B T \ln Z(T, \boldsymbol{x}). \tag{3.27}$$

Are the mean and most probable values close? They should be if the above identification is to make sense via the two routes we just described! We can address this question by looking at the variance of the energy and comparing it to the mean. Notice that the partition function, viewed as function of β, *is proportional to the characteristic function of* $\mathcal{H}$, with β standing in for the combination ik. So, we can easily generate moments by taking derivatives of Z with respect to β, e.g:

$$-\frac{\partial Z}{\partial\beta} = \sum_{\mu} \mathcal{H}e^{-\beta\mathcal{H}}, \quad \frac{\partial^2 Z}{\partial\beta^2} = \sum_{\mu} \mathcal{H}^2 e^{-\beta\mathcal{H}}. \tag{3.28}$$

From this, we see that the first moment is just

$$\langle\mathcal{H}\rangle = \langle\mathcal{H}\rangle_c = -\frac{1}{Z}\frac{\partial Z}{\partial\beta} = -\frac{\partial \ln Z}{\partial\beta} \tag{3.29}$$

With a moment generating function comes a cumulant generating function, here just $\ln Z$! Generally, taking care of the minus signs, we have

$$\langle\mathcal{H}^n\rangle_c = (-1)^n \frac{\partial^n \ln Z}{\partial\beta^n},$$

and explicitly

$$\left\langle\mathcal{H}^2\right\rangle_c = \frac{\partial^2 \ln Z}{\partial\beta^2} = -\frac{\partial\left\langle\mathcal{H}\right\rangle_c}{\partial\beta} = k_B T^2 \left.\frac{\partial\left\langle\mathcal{H}\right\rangle}{\partial T}\right|_{\boldsymbol{x}} \Rightarrow \left\langle\mathcal{H}^2\right\rangle_c = k_B T^2 C_{\boldsymbol{x}},$$

where have identified that the variance of the energy is related to the heat capacity. Compare this condition to the stability requirement in HW 1 for why the heat capacity had to be positive! Also, to answer the question posed just above, note that *every* cumulant of $\mathcal{H}$ is proportional to N, which itself tells you that the relative error $\sqrt{\langle\mathcal{H}^2\rangle_c}/\langle\mathcal{H}\rangle_c \sim N^{-1/2}$, so that it vanishes in the thermodynamic limit. Thus, the mean energy and most probable energy are identical as $N \to \infty$.

Note, by the way, that with all of the above identifications, we can define the entropy using Eq. 3.22 directly, as

$$S = -k_B \langle p(\mu)\rangle = k_B \langle(-\beta\mathcal{H} - \ln Z)\rangle = \frac{E-F}{T}, \tag{3.30}$$

recovering the familiar expression for the Helmholtz free energy.

3.2.2 The ideal gas in the canonical ensemble

Let's explicitly show that the canonical ensemble returns familiar relations for an ideal gas. The probability distribution is

$$p(\mu) = \frac{1}{Z}\exp\left(-\beta\sum_{i=1}^{N}\frac{p_i^2}{2m}\right)\cdot\begin{cases}1 & \text{if } \boldsymbol{q}_i \in \text{box}\\ 0 & \text{otherwise}\end{cases}, \tag{3.31}$$

and we compute the partition function (with the correction to the phase space measure from Eq. 3.17!) as

$$Z(T,V,N) = \int \frac{1}{N!}\prod_{i=1}^{N}\frac{d^3\boldsymbol{q}_i d^3\boldsymbol{p}_i}{h^3}\exp\left(-\beta\sum_{i=1}^{N}\frac{p_i^2}{2m}\right) = \frac{V^N}{N!}\left(\frac{2\pi m k_B T}{h^2}\right)^{3N/2}. \tag{3.32}$$

The free energy is

$$F = -k_B T\ln Z = -Nk_B T\left(\ln\left(\frac{Ve}{N}\right) + \frac{3}{2}\ln\left(\frac{2\pi m k_B T}{h^2}\right)\right), \tag{3.33}$$

and using $dF(T,V,N) = -SdT - PdV + \mu dN$ we can extract the usual properties of the ideal gas. For example, the entropy is $-S = \frac{\partial F}{\partial T}\big|_{V,N} = \frac{F-E}{T}$, and $P = -\frac{\partial F}{\partial V}\big|_{T,N} = \frac{Nk_BT}{V}$. Note that the very form of the probability distribution in Eq. 3.31 tells us the microcanonical result we had for $p(\boldsymbol{p}_1)$, that the momenta of the particles are drawn from independent Gaussian distributions.

3.3 Gibbs canonical ensemble

Briefly[7], it is sometimes more convenient to work in an ensemble where the internal energy can change by exchanging heat with a reservoir (as in the canonical ensemble) but *also* by exchanging work.The macrostate $M(T,\boldsymbol{J})$ is now specified by the temperature and forces acting on the system, and we view both the energy and the generalized coordinates $\boldsymbol{x}$ as random variables (but *not* including chemical work, which we will treat in the grand canonical

---

[7] i.e., entirely by analogy with the last section

ensemble shortly... thus, N is fixed). We proceed just as above, and find that the microstates of the system have a probability distribution

$$p(\mu) = \frac{\exp\left(-\beta\mathcal{H}(\mu) + \beta\boldsymbol{J}\cdot\boldsymbol{x}\right)}{\mathcal{Z}(T,N,\boldsymbol{J})}, \tag{3.34}$$

where the Gibbs partition function is

$$\mathcal{Z}(T,N,\boldsymbol{J}) = \sum_{\mu,\boldsymbol{x}} \exp\left(-\beta\mathcal{H}(\mu) + \beta\boldsymbol{J}\cdot\boldsymbol{x}\right). \tag{3.35}$$

We can use, again, either the "most probable value" or "mean value" method to relate the expectation values of the generalized displacements to the Gibbs partition function, like

$$\langle\boldsymbol{x}\rangle = k_B T \frac{\partial \ln \mathcal{Z}}{\partial \boldsymbol{J}}, \tag{3.36}$$

and we use the thermodynamic result that $\boldsymbol{x} = -\frac{\partial G}{\partial \boldsymbol{J}}$ to make the identification

$$G(T,N,\boldsymbol{J}) = -k_B T \ln \mathcal{Z}, \tag{3.37}$$

where we have again written the Gibbs free energy encountered in Chapter 1 as $G = E - TS - \boldsymbol{x}\cdot\boldsymbol{J}$. On can, for instance, extract the enthalpy $H = E - \boldsymbol{x}\cdot\boldsymbol{J} = -\frac{\partial \ln \mathcal{Z}}{\partial \beta}$, or the heat capacity at constant force as $C_{\boldsymbol{J}} = \frac{\partial H}{\partial T}$

3.4 The grand canonical ensemble

We now generalize once more from the canonical ensemble to the *grand* canonical ensemble. For the canonical ensemble we said that even though energy was a conserved quantity, it often makes more sense[8] to put a system in contact with a reservoir of heat so that *temperature* is the fixed quantity and E becomes a random variable as the system and the reservoir exchange heat. We measure temperature all the time, but when, after all, was the last time you precisely measured the energy of a macroscopically large system? In the same way, it often behooves us to generalize yet further and allow our system to also exchange *particle number* with the reservoir – now both N and E are variables and the expectation values, $\langle N\rangle$ and $\langle E\rangle$ are identified with thermodynamically interested quantities. For this system we require that the reservoir be at a fixed value of temperature T and chemical potential μ, and we now create a probability distribution corresponding for the grand canonical ensemble, in which the macrostates are functions of $M(T,\mu,\boldsymbol{x})$.

Notationally, since μ is the chemical potential, I'll be careful about using μ_S to refer to a microstate of the system we're interested in. The probability density function for μ_S is again

---

[8] and, as we will see in later chapters, often makes it easier to calculate

derived by invoking the fundamental postulate and summing over all states of the reservoir, ultimately giving

$$p(\mu_S) = \frac{1}{\mathcal{Q}(T,\mu,\boldsymbol{x})}\exp\left[-\beta\mathcal{H}(\mu_S)+\beta\mu N(\mu_S)\right], \tag{3.38}$$

where $N(\mu_S)$ is the number of particles in microstate μ_S, and where $\mathcal{Q}$ is the *grand partition function*:

$$\mathcal{Q}(T,\mu,\boldsymbol{x}) = \sum_{\mu_S}\exp\left[\beta\mu N(\mu_S)-\beta\mathcal{H}(\mu)\right]. \tag{3.39}$$

We can usefully rearrange the above summation by first grouping together all of the microstates with the same number of particles in them. Letting $\mathcal{H}_N$ refer to the Hamiltonian associated with the N-particle system, we write

$$\mathcal{Q}(T,\mu,\boldsymbol{x}) = \sum_{N=0}^{\infty} e^{\beta\mu N}\sum_{\mu_S} e^{-\beta\mathcal{H}_N(\mu_S)} = \sum_{N=0}^{\infty} z^N \sum_{\mu_S} e^{-\beta\mathcal{H}_N(\mu_S)}, \tag{3.40}$$

where we have additionally defined the *fugacity*[9] $z=\exp(\beta\mu)$. Notice, by the way, that the second sum – over microstates with a particular number of particles – are the canonical partition functions associated with an N-particle system, so we can additionally write

$$\mathcal{Q}(T,z,\boldsymbol{x}) = \sum_{N=0}^{\infty} z^N Z_N(T,\boldsymbol{x}), \quad \text{defining } Z_0 \equiv 1. \tag{3.41}$$

This form makes it seem that to calculate the grand canonical partition function we need to have already calculated the canonical partition function. In principle we indeed do, but in later chapters[10] we'll see how we can sometimes make great progress in evaluating $\mathcal{Q}$ even in situations where evaluating Z is very difficult.

3.4.1 Number fluctuations in the grand canonical ensemble

Earlier we showed the generic equivalence of the canonical and microcanonical ensembles as $N\to\infty$ by establishing that the average value of the energy and the most probable value of the energy became indistinguishable in the thermodynamic limit. Here we do the same thing by considering both the *mean* and *most typical* value of the number of particles. We note that from the definition of $\mathcal{Q}$ we can read off the total weight of the microstates with N particles, the probability for finding the system with N particles is

$$p(N) = \frac{e^{\beta\mu N}Z(T,N,\boldsymbol{x})}{\mathcal{Q}(T,\mu,\boldsymbol{x})}. \tag{3.42}$$

We write the average as

$$\langle N\rangle = \frac{1}{\mathcal{Q}}\frac{\partial\mathcal{Q}}{\partial(\beta\mu)} = \frac{\partial\ln\mathcal{Q}}{\partial(\beta\mu)}, \tag{3.43}$$

---

[9] Derived from *fugere*, to flee. The term was popularized in an early textbook by Gilbert Lewis and Merle Randall as an "escaping tendency," referring to the flow of matter between phases, and playing a similar role to temperature in the flow of heat.

[10] both on quantum stat mech and on interacting systems

and the variance as

$$\left\langle N^2\right\rangle_c = \left\langle N^2\right\rangle - \langle N\rangle^2 = \frac{1}{\mathcal{Q}}\frac{\partial^2\mathcal{Q}}{\partial(\beta\mu)^2} - \left(\frac{\partial\ln\mathcal{Q}}{\partial(\beta\mu)}\right)^2 = \frac{\partial^2\ln\mathcal{Q}}{\partial(\beta\mu)^2} = \frac{\partial\langle N\rangle}{\partial(\beta\mu)}. \tag{3.44}$$

We again see that the variance is proportional to $\langle N\rangle$, so we again get that the relative fluctuations, $\sigma_N/\langle N\rangle$ vanish in the thermodynamic limit. This suggests the equivalence of the grand canonical ensemble with the others, but have we been a little too glib, here? Let's helpfully re-write the above, using some thermodynamic relationships that are straightforward to derive, as

$$\frac{\langle N^2\rangle_c}{\langle N\rangle^2} = \frac{k_BT}{V}\kappa_T, \tag{3.45}$$

where κ_T is once again the isothermal compressibility. From this we see that the relative RMS fluctuations in particle density are ordinarily $\mathcal{O}(N^{-1/2})$, but we will sometimes see exceptions. In particular, interesting things happen near phase transitions, and one often sees the compressibility of a system grow very large. For instance, near a liquid-vapor phase transition at temperature T_c experiments suggest that that isothermal compressibility itself scales with system size, like $\kappa_T(T_c)\sim N^{0.63}$, suggesting unusually large fluctuations of particle density near the critical point. Such fluctuations can be seen in experiments[11].

So, under these sorts of circumstances the formalism associated with the grand canonical ensemble could in principle give non-identical answers as the formalism associated with the canonical ensemble. In these cases, we have no choice but to use the grand canonical ensemble. Also, you may be wondering where the unusual real-number exponent came from in the above experimental statement about density fluctuations at the liquid-vapor transition... we will see at the end of class where, in principle, such exponents come from.

3.4.2 Thermodynamics in the grand canonical ensemble

From the above, we now once again approximate the sum in Eq. 3.39 by its single largest term, which corresponds to the typical value of N (note that we feel free to go back and forth between fugacity and chemical potential representations as we see fit):

$$\mathcal{Q}(T,\mu,\boldsymbol{x}) = \lim_{N\to\infty}\sum_{N=0}^{\infty}\left[e^{\beta\mu N}Z(T,N,\boldsymbol{x})\right] \approx e^{\beta\mu N^*-\beta F} = e^{-\beta\mathcal{G}}, \tag{3.46}$$

where

$$\mathcal{G}(T,\mu,\boldsymbol{x}) = E - TS - \mu N = -k_BT\ln\mathcal{Q}$$

is the *grand potential* (which we first met in Chapter 1), which is up to a factor of $-k_BT$ what Pathria defines as the q-potential. We can recover typical thermodynamic relationships using $d\mathcal{G} = -SdT - Nd\mu + \boldsymbol{J}\cdot d\boldsymbol{x}$, and extract pressures or heat capacities by usual derivative manipulations.

---

[11] check out, for instance, some of the critical opalescence videos that are easy to find on youtube

3.4.3 The ideal gas in the grand canonical ensemble

We once again churn through our standard example, the ideal gas, for the grand canonical ensemble. We have a macrostate $M(T,\mu,V)$, where the corresponding microstates are over particle positions and momenta with an indefinite number of particles in the specified volume. Introducing the thermal de Broglie wavelength[12] $\lambda = h/\sqrt{2\pi m k_B T}$, the grand partition function is

$$\begin{aligned}\mathcal{Q}(T,\mu,V) &= \sum_{N=0}^{\infty} e^{\beta\mu N}\frac{1}{N!}\int\left(\prod_{i=1}^{N}\frac{dV_i}{h^3}\right)\exp\left[-\beta\sum_{i=1}^{N}\frac{p_i^2}{2m}\right] \\ &= \sum_{N=0}^{\infty}\frac{1}{N!}e^{\beta\mu N}\left(\frac{V}{\lambda^3}\right)^N \\ &= \exp\left(e^{\beta\mu}\frac{V}{\lambda^3}\right) \qquad (3.47) \\ \Rightarrow \mathcal{G}(T,\mu,V) &= -k_BT\ln\mathcal{Q} = -k_BTe^{\beta\mu}\frac{V}{\lambda^3} \qquad (3.48)\end{aligned}$$

We can immediately read off things like[13]

$$P = -\left.\frac{\partial\mathcal{G}}{\partial V}\right|_{T,\mu} = k_BT\frac{z}{\lambda^3}$$

or

$$N = -\left.\frac{\partial\mathcal{G}}{\partial\mu}\right|_{T,V} = \frac{zV}{\lambda^3}\Rightarrow PV = Nk_BT,$$

as we have come to expect the equation of state for the ideal gas to look.

Notice, also, that $\mathcal{G}$ for an ideal gas only depends on a single extensive variable: V. Since we Expect $\mathcal{G}$ to itself be extensive[14], *it must be that* $\mathcal{G}\propto V$. We have a name for the constant of proportionality: "pressure," so $\mathcal{G}(T,V,\mu) = -P(T,\mu)V$. This makes for an easy method of computing pressures of systems that depend only on one extensive variable[15].

It is worth saying a bit more about the chemical potential, here. Rearranging that last expression for N gives

$$\mu = k_BT\log\left(\frac{\lambda^3N}{V}\right), \qquad (3.49)$$

and if $\lambda^3 < V/N$, as we certainly expect if we are to be doing classical calculations in the first place, then the chemical potential is negative! Is it clear why this makes sense? By writing

---

[12] Anticipating future developments in this course, λ helps characterize the regime in which classical statistical mechanics is typically valid: if λ is roughly the same order of the typical separation between particles then quantum effects become important.

[13] switching back to the fugacity, largely to save on TEXtime

[14] i.e., satisfying $\mathcal{G}(T,\lambda V,\mu) = \lambda\mathcal{G}(T,V,\mu)$

[15] As we'll see in the chapter on Quantum statistical mechanics, for example, when we calculate the grand partition function of ideal quantum gases.

μ as conjugate to N, we intuitively think of it as the energy cost associated with adding an extra particle to the system, but we need to look more carefully at the definition:

$$\mu = \left.\frac{\partial E}{\partial N}\right|_{S,V}.$$

That is, μ is the energy cost of adding a particle *at fixed entropy and volume*. In general, adding a particle will in fact increase the entropy (since there are more ways of partitioning the available energy), so if we are holding *entropy* fixed then the system must be lowering its energy when adding a particle. Hence, $\mu < 0$ for the classical ideal gas.

3.5 Failures of classical statistical mechanics

Classical statistical mechanics is an incredibly powerful framework for understanding the physical properties of many systems, but it certainly has limitations. We have already seen the Gibbs paradox and our current, somewhat *ad hoc* solution of throwing in a factor of $N!$ to account for classically indistinguishable particles. We have also seen that in going from $S = -\sum_i p_i \ln p_i$ to $S = -\int dx\ p(x) \ln p(x)$ there is a problem of choosing units, and in the last few sections we have introduced a random factor of h to fix the unit problem without understanding why a particular scale of coarse-graining is appropriate or what the value of h actually is. In this section we also emphasize that there are observable, low-temperature phenomena for which classical mechanics makes incorrect predictions, further necessitating the introduction of quantum mechanical to our description. The following subsections provide a few concrete illustrations.

3.5.1 Dilute diatomic gases

Actually, classical statistical mechanics fails to predict heat capacities both in the gaseous phase *and* and in the solid phase... not a great track record. Let's see what happens for a dilute gas of diatomic molecules.

We'll take a simplified view and consider diatomic molecules consisting of two atoms in a bound state, and let's write down an idealized, classical model for such a molecule: two equal masses connected by a spring. So, in addition to the usual translational modes we've been considering, the molecule can also move via (a) rotations in which the molecule rigidly rotates about either of the two axes[16] normal to the symmetry axis, with moment of inertia I, and (b) vibrations in which the molecule oscillates along the axis of symmetry.

We first assume that the gas is sufficiently dilute that the molecules behave independently, so that the total partition function is

$$Z(N) = \frac{Z_1^N}{N!},$$

Where Z_1 is the partition function for a single diatomic molecule. We further assume that three modes of molecular motion are all independent, in which case the single-molecule

---

[16] we neglect the rotation about the axis of symmetry, arguing that it has a low moment of inertia compared to the other two. We'll see that this is hardly the problem.

partition function factorizes into the contribution due to each term:

$$Z_1 = Z_{trans} Z_{vib} Z_{rot}.$$

We already know what the translational partition function looks like, what about the other two terms?

From your classical mechanics course, the Lagrangian for the rotational degrees of freedom is

$$L_{rot} = \frac{1}{2} I \left(\dot{\theta}^2 + \dot{\phi}^2 \sin^2 \theta \right),$$

with conjugate momenta

$$p_\theta = \frac{\partial L_{rot}}{\partial \dot{\theta}} = I\dot{\theta}, \quad p_\phi = \frac{\partial L_{rot}}{\partial \dot{\phi}} = I\dot{\phi} \sin^2 \theta.$$

The hamiltonian for the rotational piece is therefore

$$\mathcal{H}_{rot} = \dot{\theta} p_\theta + \dot{\phi} p_\phi - L = \frac{p_\theta^2}{2I} + \frac{p_\phi^2}{2I \sin^2 \theta}.$$

So, the rotational contribution to the partition function is

$$Z_{rot} = \frac{1}{h^2} \int d\theta d\phi dp_\theta dp_\phi e^{-\beta \mathcal{H}_{rot}} = \frac{8\pi^2 I k_B T}{h^2}. \tag{3.50}$$

What about the vibrational mode? It's just a harmonic oscillator. Denoting the displacement away from the equilibrium position of the molecular "spring" by ζ and the vibrational frequency by ω, the Hamiltonian is

$$\mathcal{H}_{vib} = \frac{p_\zeta^2}{2m} + \frac{1}{2} m \omega^2 \zeta^2,$$

from which we find the partition function contribution to be

$$Z_{vib} = \frac{1}{h} \int d\zeta dp_\zeta \; e^{-\beta \mathcal{H}_{vib}} = \frac{2\pi k_B T}{h\omega}. \tag{3.51}$$

Putting together all of these ingredients (or, by your expectations from equipartition of energy among all of the quadratic modes in the combined Hamiltonian), we expect that the heat capacity and constant volume for our diatomic gas is

$$C_V = \frac{7}{2} N k_B, \tag{3.52}$$

an end result which does not depend on the precise value of I or the stiffness of the bond between the atoms. The only problem is that our prediction for the heat capacity is... not borne out in the experimental data. In Fig. 3.1 I've schematically[17] plotted the heat capacity of H_2 – the simplest diatomic gas – over a broad range of temperatures. At very high temperatures we do see the heat capacity we expected, but at the lowest temperatures the system seems to behave like a *monatomic* gas, so apparently the diatomic molecules are

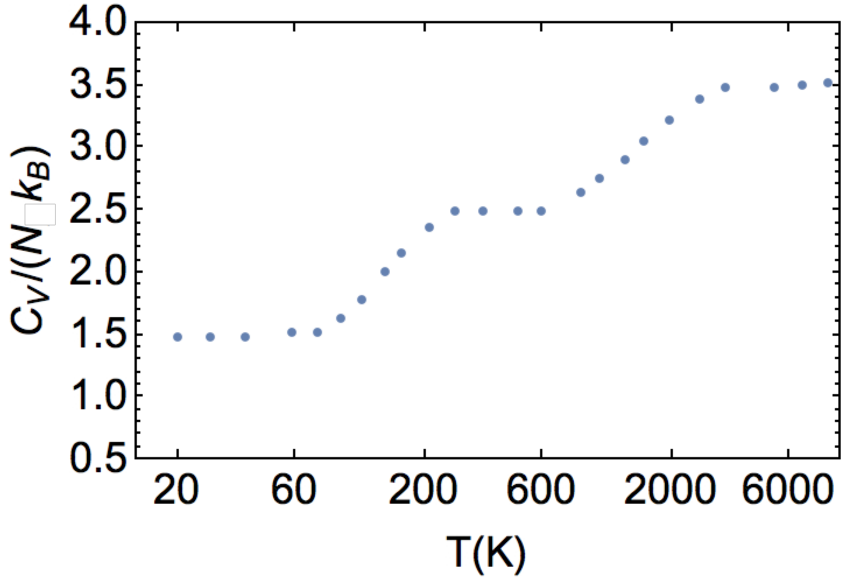


Figure 3.1: **(Schematic) heat capacity of hydrogen gas vs logarithmically scaled temperature.** That "N_A" is a typo, and should just be N, of course.

neither rotating nor vibrating. Even away from the typical "very low temperatures" we expect to see quantum effects in – that is, even at room temperature! – there is a large discrepancy between our prediction and the actual heat capacity, and apparently there are rotations but not vibrations (we'll see how we picked out that particular mode later).

This behavior was arguably the first time that quantum mechanics revealed itself in experiments, and scientists at the end of the 19th century were increasingly unsettled.

3.5.2 Black-body radiation[18]

The classical failure of the calculation of the black-body spectrum (i.e., what light is emitted from a source at a particular temperature) and its quantum resolution is a story that I suspect is familiar from previous courses in quantum mechanics. Very briefly, then, we consider our first quantum gas: a gas of photons. In principle we are interested in the emission spectrum of an idealized substance that can absorb photons of any wavelength and reflects none of them. In a real atomic system there would be a (potentially) interesting pattern of absorption and emission lines, but we ignore such details for now and consider our idealized substance which, at zero temperature, would appear black (hence the name).

So, we know a photon is characterized by its wavelength λ or its frequency $\omega = 2\pi c/\lambda = kc$ for speed of light c and wavevector k, with energy $E = \hbar\omega$. An important fact about photons is that they are *not conserved*: there is no reason that the walls of our black-body substance

---

[17] Actual data available from NIST, if you're interested

[18] "It was an act of desperation. For six years I had struggled with the blackbody theory. I knew the problem was fundamental and I knew the answer. I had to find a theoretical explanation at any cost, except for the inviolability of the two laws of thermodynamics." – Max Planck, in a letter to R. W. Wood, 1931

couldn't absorb a photon and then emit two[19]. Thus, when we calculate quantities in the canonical ensemble we need to make sure that we sum over possible states with different numbers of photons, since they are allowed states. Equivalently, we can imagine that we work in the grand canonical ensemble, but with chemical potential $\mu_{photon} = 0$.

To build up the partition function, let's first consider photons with a particular frequency ω; N such photons would have energy $E = N\hbar\omega$, and summing over the allowed N gives a partial partition function

$$Z_\omega = 1 + e^{-\beta\hbar\omega} + e^{-2\beta\hbar\omega} + e^{-3\beta\hbar\omega} + \cdots = \frac{1}{1 - e^{-\beta\hbar\omega}}. \tag{3.53}$$

We assume that the different frequencies are independent, and the total partition function is a product of independent partition functions, so we can write the logarithm as a sum:

$$\log Z = \int_0^\infty d\omega g(\omega) \log Z_\omega,$$

where $g(\omega)$ is the *density of states*: $g(\omega)d\omega$ counts the number of states in the frequency range between ω and $\omega + d\omega$. We can calculate this by, say, assuming periodic boundary conditions in a box of linear size L, which then permits wavevectors $\boldsymbol{k} = 2\pi\{n_x, n_y, n_z\}$, where $\{n_x, n_y, n_z\}$ are all integers. Planck suggested that the allowed values of energy must be quantized,

$$\mathcal{H}_{EM} = \sum_{\boldsymbol{k},\alpha} \hbar c k \left(n_\alpha(\boldsymbol{k}) + \frac{1}{2} \right), \quad n_\alpha(\boldsymbol{k}) = 0, 1, 2, \ldots,$$

where α refers to the polarization of the photon. One can use this to compute the associated density of states for the photon gas. For now I'll ignore terms relating to the factor of 1/2 above, since we usually only care about (or can detect!) energy differences. Taking into account the fact that photons can come in two polarization states, one eventually gets

$$g(\omega)d\omega = \frac{V\omega^2}{\pi^2 c^3} d\omega.$$

Combing these results, we get that

$$\ln Z = \int_0^\infty d\omega g(\omega) \log Z_\omega = -\frac{V}{\pi^2 c^3} \int_0^\infty d\omega \; \omega^2 \ln\left(1 - e^{-\beta\hbar\omega}\right). \tag{3.54}$$

From this we can get, e.g., the energy stored in the photon gas:

$$E = -\frac{\partial \ln Z}{\partial \beta} = \frac{V\hbar}{\pi^2 c^3} \int_0^\infty d\omega \frac{\omega^3}{e^{\beta\hbar\omega} - 1} = \frac{V}{\pi^2 c^3} \frac{(k_B T)^4}{\hbar^3} \int_0^\infty dx \frac{x^3}{e^x - 1}. \tag{3.55}$$

That last integral can be explicitly evaluated (with some work), with the end result

$$\frac{E}{V} = \frac{\pi^2 k_B^4}{15\hbar^3 c^3} T^4. \tag{3.56}$$

---

[19] Of course, you *know* this, as you demonstrate the non-conservation of photon number every time you change the state of a (functional) light switch

The free energy $F = -k_B T \ln Z$ can be calculated; integration by parts (to take care of the log inside the integral) gets us to

$$F = -\frac{V\pi^2}{45\hbar^3 c^3}(k_B T)^4, \tag{3.57}$$

from which we can calculate, e.g., the pressure due to electromagnetic radiation as

$$P = -\left.\frac{\partial F}{\partial V}\right|_T = \frac{E}{3V} = \frac{4\sigma}{3c}T^4,$$

where

$$\sigma = \frac{\pi^2 k_B^4}{60\hbar^3 c^2} = 5.67 \times 10^{-8}\frac{J}{sm^2K^4}$$

is the Stefan constant[20]

We didn't really emphasize the way the classical version of the above calculation fails, as I'm sure you've seen it before. Briefly, classically we should think that the Hamiltonian for the EM field can be written in terms of the normal modes characterized by $\boldsymbol{k}$, so that the energy looks like a collection of independent harmonic oscillators corresponding to photons of different wave-number and polarization. Classically, though, there is no limit on the size of $\boldsymbol{k}$, leading to the *ultraviolet catastrophe*: we assign $k_B T$ of equipartitioned energy to each independent quadratic mode of $\mathcal{H}$, leading to an infinite amount of energy stored in the high-frequency modes. This is... not physical.

By the way, all along we discarded the terms related to the extra 1/2 in $\mathcal{H}_{EM}$. Had we kept those terms around we would have found things like

$$P = -\left.\frac{\partial F}{\partial V}\right|_T = P_0 + \frac{E}{3V},$$

where P_0 is an infinite zero-point pressure. Usually, this infinite zero-point pressure cancels with some other part of your calculation, but not always! Differences in this quantity P_0 lead to Casimir forces between, e.g., conducting plates, which can be experimentally measured.

---

[20] the value of which in SI units is particularly easy to remember, since it is "five (point) six seven times (ten to the negative) eight."

Chapter 4

Quantum Statistical Mechanics[1]

Quantum statistical mechanics should be inherently *quantum* mechanical, *statistical* in nature, and *able to resolve* some of the issues with classical statistical mechanics that we encountered at the end of the last chapter[2]. When considering blackbody radiation we already introduced a simplistic version of quantum mechanical effects by forcing the energy levels of the photon gas to be quantized, and we assumed that the quantum mechanical microstates were specified by this quantized energy and governed by a probability distribution related to the Boltzmann weights.

Why were we allowed to make such assumptions? What is the traditional resolution of Gibbs' paradox? Why is Planck's constant floating around in classical partition functions? In this chapter we formalize a quantum description of statistical mechanics by following the steps that led us to the classical description (Chapters 3 and 4).

4.1 The classical limit of a quantum partition function

Let's jump right in and explain why there are factors of $1/h$ in the classical partition functions. Classically, we defined the partition function for a single classical particle as an integral over phase space:

$$Z_1 = \frac{1}{h^3}\int d^3\boldsymbol{q}d^3\boldsymbol{p}e^{-\beta H(p,q)},$$

where the $1/h$ was required to get the units right (i.e., so that Z is dimensionless), but where there was a particular value of h to use: Planck's constant $h = 2\pi\hbar \approx 6.6\times 10^{-34} Js$. Why is there this quantum-mechanical number in our classical formulas? We don't need to wave our hands: we can derive it.

To keep things simple, let's consider a single particle in one dimension, so that it's Hamiltonian is

$$\hat{H} = \frac{\hat{p}^2}{2m} + V(\hat{q}),$$

where $\hat{p}$ is the momentum operator and $\hat{q}$ is the position operator.

---

[1] Selections from Pathria Chapter 5-8; the majority of Chapter 5, 6.1, 6.2, and at the thermodynamics of ideal Bose and Fermi gases

[2] Future editions of these notes will maintain the parallelism.

Writing the eigenstates in the energy basis, with $|n\rangle$ associated with energy E_n, the quantum partition function[3] is

$$Z = \sum_n e^{-\beta E_n} = \sum_n \langle n|e^{-\beta \hat{H}}|n\rangle, \tag{4.1}$$

where the operator

$$e^{-\beta \hat{H}} = \sum_{n=0}^{\infty} (-1)^n \frac{(\beta \hat{H})^n}{n!}.$$

In general, by the way, we can think of functions of matrices in this way, taking $f(M) = \sum_{i=0}^{\infty} a_i M^i$ for a matrix M are where the a_i define the power series of the function. It is often, though, easier to write down the matrix in a basis which diagonalizes M. Let's write the diagonalized version of the matrix as $\tilde{M}$; in the diagonalizing basis any function can be written as

$$f(\tilde{M}) = \begin{pmatrix} f(\tilde{M}_{11}) & 0 & 0 & 0 & \cdots \\ 0 & f(\tilde{M}_{22}) & 0 & 0 & \cdots \\ \vdots & 0 & f(\tilde{M}_{33}) & 0 & \cdots \\ \vdots & \vdots & \ddots & \ddots & \ddots \end{pmatrix}. \tag{4.2}$$

One can then transform back into whichever basis you like. This procedure is a natural way to define the log, as in expressions for the entropy in terms of the density matrix ρ,

$$S = -k_B \mathrm{Tr}\left(\rho \log \rho\right). \tag{4.3}$$

As usual we are free to insert the identity operator, constructed by summing over any complete basis of states. We'll do this with both the position eigenvectors and the momentum eigenvectors:

$$\mathbf{1} = \int dq |q\rangle\langle q|, \quad \mathbf{1} = \int dp |p\rangle\langle p|.$$

We first add two copies of the position-eigenstate identity to the partition function, on either side of the $e^{-\beta \hat{H}}$:

$$\begin{aligned} Z &= \sum_n \langle n| \int dq |q\rangle\langle q| e^{-\beta \hat{H}} \int dq' |q'\rangle\langle q'|n\rangle \\ &= \int dq dq' \langle q|e^{-\beta \hat{H}}|q'\rangle \sum_n \langle q'|n\rangle\langle n|q\rangle \\ &= \int dq \langle q|e^{-\beta \hat{H}}|q\rangle, \end{aligned} \tag{4.4}$$

where in the last line we replaced $\sum_n |n\rangle\langle n|$ with the identity operator and used $\langle q'|q\rangle = \delta(q'-q)$ and integrated over q'. So far the result of this manipulation is to replace a sum over energy eigenstates with an integral over position eigenstates[4].

---

[3] We'll see this more explicitly soon, but the form should be familiar enough that you don't object...

[4] we can do this with any complete basis; as we'll see shortly we can write the partition function without referencing a specific basis by $Z = \mathrm{Tr}\left(e^{-\beta \hat{H}}\right)$

Let's take the classical limit of this partition function, by which we mean that we'll typically neglect terms that are of order h. We exploit this by trying to factorize $e^{-\beta\hat{H}}$ into a position and a momentum piece, remembering that[5]

$$e^{\hat{A}}e^{\hat{B}} = \exp\left(\hat{A}+\hat{B}+\frac{1}{2}[\hat{A},\hat{B}]+\cdots,\right) \tag{4.5}$$

where we additionally recall the commutation relation $[\hat{q},\hat{p}] = i\hbar$. Taking the classical limit means we neglect the corrections to the naive factorization, writing

$$e^{-\beta\hat{H}} = e^{-\beta\hat{p}^2/(2m)}e^{-\beta V(\hat{q})} + \mathcal{O}(\hbar). \tag{4.6}$$

To complete the derivation, let's carefully start transitioning exponentiated operators into ordinary functions:

$$\begin{aligned} Z &= \int dq\ \langle q|e^{-\beta\hat{p}^2/(2m)}e^{-\beta V(\hat{q})}|q\rangle \\ &= \int dq\ e^{-\beta V(q)}\langle q|e^{-\beta\hat{p}^2/(2m)}|q\rangle \\ &= \int dqdpdp'\ e^{-\beta V(q)}\langle q|p\rangle\langle p|e^{-\beta\hat{p}^2/(2m)}|p'\rangle\langle p'|q\rangle \\ &= \frac{1}{2\pi\hbar}\int dqdp\ e^{-\beta H(p,q)}, \qquad \text{using } \langle q|p\rangle = \frac{1}{\sqrt{2\pi\hbar}}e^{ipq/\hbar}. \end{aligned} \tag{4.7}$$

Thus, we see the natural consequence of the underlying quantum mechanical description even when we take the classical limit and do our best to ignore $\hbar$.

4.2 Microstates, observables, and dynamics

4.2.1 Quantum microstates

With that as an appetizer, let's write down a few definitions[6] to get us more properly started on quantum statistical mechanics. Classically, we started with microstates for N particles that were specified by a point in $6N$-dimensional phase space and governed by Hamiltonian evolution equations. Quantum mechanically, of course, positions and momenta are not independently observable, so this is a poor choice of microstate. Instead, a quantum system is specified by *a unit vector in a Hilbert space*[7], $|\psi\rangle$. Given a set of orthonormal basis vectors $|n\rangle$ we can write the microstate as

$$|\psi\rangle = \sum_n \langle n|\psi\rangle|n\rangle, \tag{4.8}$$

---

[5] Baker-Campbell-Hausdorff formula

[6] I will, naturally, assume you already know quantum mechanics, so here we're just dotting some i's

[7] A generalization of Euclidean space: a Hilbert space H is a vector space equipped with an inner product whose induced distance function makes H a complete metric space

where the $\langle n|\psi\rangle$ are complex numbers, where we keep in mind that $\langle\psi|n\rangle = \langle n|\psi\rangle^*$, the complex conjugate.

As a unit vector, ψ is normalized so that

$$\langle\psi|\psi\rangle = \sum_n \langle\psi|n\rangle\langle n|\psi\rangle = 1, \tag{4.9}$$

4.2.2 Quantum observables

Classically, in Chapter 3 we introduced observable functions of phase space, $A(\{\boldsymbol{p}, \boldsymbol{q}\})$. Quantum mechanically, observables get promoted to *operators* by substituting the position and momentum operators for the position and momentum variables in the classical expressions[8], e.g. $\hat{A}$, as we saw in the first part of this chapter. Just as classically we had the Poisson bracket, $\{\boldsymbol{p}_i, \boldsymbol{q}_j\} = \delta_{ij}$, here we have the commutation relation $[\hat{p}_a, \hat{q}_b] = \hat{p}_a\hat{q}_b - \hat{q}_b\hat{p}_a = \frac{\hbar}{i}\delta_{ab}$, and we can write our observables, if we wish, as functions of the position and momentum: $A(\hat{p}, \hat{q})$.

Unlike in classical mechanics, in addition to the (classical) probabilistic nature of our ensembles, quantum mechanically our *observables themselves* are matrices that don't have definite values; i.e., they are not uniquely determined for a particular microstate. This additional randomness means the observables are themselves random variables, so we must content ourselves with their expectation values, defined as

$$\langle A\rangle = \langle\psi|A|\psi\rangle = \sum_{m,n} \langle\psi|m\rangle\langle m|A|n\rangle\langle n|\psi\rangle. \tag{4.10}$$

Since we demand real observables, the operators A must be Hermitian[9]: $A^\dagger = A$.

4.2.3 Time evolution of states

Classically we got a lot of mileage out of the hamiltonian evolution of the phase space coordinates. The quantum mechanical state vector has a time evolution given by

$$i\hbar\frac{\partial}{\partial t}|\psi(t)\rangle = H|\psi(t)\rangle. \tag{4.11}$$

It is often convenient to work in the basis which diagonalizes the Hamiltonian (i.e., the basis formed by the energy eigenstates), satisfying $H|n\rangle = E_n|n\rangle$ where E_n are the "eigenenergies." In such a basis, exploiting the orthonormality of the basis $\langle m|n\rangle = \delta_{mn}$ lets us write the time evolution of the state as

$$i\hbar\frac{d}{dt}\langle n|\psi(t)\rangle = E_n\langle n|\psi(t)\rangle \Rightarrow \langle n|\psi(t)\rangle = \exp\left(\frac{-iE_nt}{\hbar}\right)\langle n|\psi(0)\rangle. \tag{4.12}$$

Probably the first basis we learn about is composed of the spatial coordinates, $|\{\boldsymbol{q}_i\}\rangle$, for which

$$\psi(\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_N) \equiv \langle\{\boldsymbol{q}_i\}|\psi\rangle$$

is the wavefunction.

---

[8] after, of course, properly symmetrizing products, for instance $pq \to (pq + qp)/2$

[9] which is why we worried about symmetrizing position and momentum when going from classical to quantum operators.

4.3 The density matrix and macroscopic observables

Classically, macrostates are specified by just a few thermodynamic coordinates, and we studied ensembles of large numbers of microstates μ_s, which were equipped with a probability $p_s \equiv p(\mu_s)$, which corresponded to a given macrostate. We often don't have precise knowledge of the microstate (i.e., the system is not a *pure state*); more generally we expect it to be a *mixed state*, existing as an incoherent mixtures of being in a variety of quantum states[10].

We similarly start out with a mixed state, an incoherent mixture of states $|\psi_\alpha\rangle$ with probabilities p_α. The ensemble average of the expectation value of an observable in such a mixed state is

$$\begin{aligned}\langle\bar{A}\rangle &= \sum_\alpha p_\alpha \langle\psi_\alpha|A|\psi_\alpha\rangle = \sum_{\alpha,m,n} p_\alpha\langle\psi_\alpha|m\rangle\langle n|\psi_\alpha\rangle\langle m|A|n\rangle \\ &= \sum_{m,n}\langle m|A|n\rangle\langle n|\rho|m\rangle = \mathrm{Tr}\,(\rho A)\,, \end{aligned} \tag{4.13}$$

where we have introduced the *density matrix* ρ, which in a given basis is

$$\rho = \sum_\alpha p_\alpha|\psi_\alpha\rangle\langle\psi_\alpha|, \tag{4.14}$$

and where the trace of an operator is the sum over the diagonal elements, $\mathrm{Tr}\,(M) = \sum_\alpha\langle\Phi_\alpha|M|\Phi_\alpha\rangle$, which is independent of which basis $\{\Phi_\alpha\}$ you use.

4.3.1 Basic properties of the density matrix

1. *Sufficiency:* All measurements in quantum mechanics involve expectation values of operators. Thus, the density matrix contains sufficient information for anything we might want to do.

2. *Pure states:* The density matrix corresponds to a pure state, a state with a definite wavefunction ψ, if and only if $\rho = |\psi\rangle\langle\psi|$, hence if and only if $\rho^2 = \rho$.

3. *Positive definite:* The eigenvalues of ρ are all positive, since for any state

$$\langle\phi|\rho|\phi\rangle = \sum_n p_n\langle\phi|\psi_n\rangle\langle\psi_n|\phi\rangle = \sum_n p_n|\langle\phi|\psi_n\rangle|^2 \geq 0.$$

4. *Normalization:* Since the ψ_α are themselves normalized, we have

$$\mathrm{Tr}\,(\rho) = \sum_n p_n\langle\psi_n|\psi_n\rangle = \sum_n p_n = 1.$$

5. Hermiticity: By inspection, the density matrix is Hermitian, with $\rho^\dagger = \rho$.

---

[10] Not in a superposition, of states, by the way. A quick example: take a spin in the up-down basis. The superposition state $2^{-1/2}(|\uparrow\rangle + |\downarrow\rangle)$ is a diagonally polarized state. An unpolarized spin is a mixture of half up and half down, described by a density matrix $\frac{1}{2}\left(|\uparrow\rangle\langle\uparrow| + |\downarrow\rangle\langle\downarrow|\right)$. The latter is what we mean.

Time evolution of the density matrix

Classically we had Liouville's theorem for the evolution of the density, $\frac{d\rho}{dt} = \frac{\partial \rho}{\partial t} - \{\mathcal{H}, \rho\}$; what do we have here? Again working in the energy eigenbasis, we can write

$$\begin{aligned} i\hbar\partial_t\rho &= \sum_n p_n i\hbar\partial_t \left(|\psi_n\rangle\langle\psi_n|\right) \\ &= \sum_n p_n \left[H|\psi_n\rangle\langle\psi_n| - |\psi_n\rangle\langle\psi_n|H\right] \\ &= H\rho - \rho H = [H, \rho]\,. \end{aligned} \tag{4.15}$$

4.4 Quantum ensembles

With this structure, we can follow the same logic that we did in the classical case: we define equilibrium by having none of the averages of the observables vary with time, which can be satisfied if we choose an equilibrium density matrix so that $\partial_t\rho = 0$. Just as when we were dealing with Poisson brackets, we accomplish this by having the density matrix be a function of the Hamiltonian itself, along with any conserved quantities A, $\rho(H, A_1, \ldots)$, that satisfy $[H, A_i] = 0$.

4.4.1 Quantum microcanonical ensemble

We define the microcanonical ensemble, specified by $(E, \boldsymbol{x}, N)$, but enforcing a fixed value for the ensemble average energy. We choose our density matrix

$$\rho(E) = \frac{\delta(H - E)}{\Omega(E)},$$

where in the energy eigen-basis we can write this as

$$\langle m|\rho|n\rangle = \sum_\alpha p_\alpha \langle m|\psi_\alpha\rangle\langle\psi_\alpha|n\rangle = \begin{cases} \Omega^{-1} & \text{if } E_n = E, \text{ and } m = n \\ 0 & \text{otherwise} \end{cases} \tag{4.16}$$

The first of those two conditions we recognize as the equivalent of the assumption of equal a priori probabilities The second, quantum mechanical condition is the *assumption of random a priori phases*, in which we don't get contributions from off-diagonal terms (even if they have degenerate and correct energies) because we assume the mixed state is in an *incoherent* superposition of the basis states. Finally, note that from the normalization condition on the density matrix, $\Omega(E)$ is again just counting the number of (eigen)states of H with the correct energy E.

4.4.2 Quantum canonical ensemble

You know what's coming: Now we're fixing temperature $T = \beta^{-1}$ by putting our quantum system in contact with a reservoir. Considering the above two assumptions for the combined system, we find that the density matrix for the system of interest is

$$\rho(\beta) = \frac{e^{-\beta H}}{Z(\beta)}, \tag{4.17}$$

where the normalization condition on the density matrix leads to the quantum canonical partition function for N particles,

$$Z_N(\beta) = \mathrm{Tr}\left(e^{-\beta H}\right) = \sum_n e^{-\beta E_n}. \tag{4.18}$$

As one would expect from the above formulas, the expectation value of a physical observable is given by

$$\langle A \rangle = \mathrm{Tr}\left(\hat{\rho}\hat{A}\right) = \frac{1}{Z(\beta)}\mathrm{Tr}\left(\hat{A}e^{-\beta\hat{H}}\right) = \frac{\mathrm{Tr}\left(\hat{A}e^{-\beta\hat{H}}\right)}{\mathrm{Tr}\left(e^{-\beta\hat{H}}\right)} \tag{4.19}$$

4.4.3 Quantum grand canonical ensemble

For completeness – and because we'll see it again soon – in the grand canonical ensemble we no longer fix the number of particles[11]. The density matrix is

$$\rho(\beta, \mu) = \frac{e^{-\beta H + \beta\mu N}}{\mathcal{Q}}, \tag{4.20}$$

where the grand canonical partition function is

$$\mathcal{Q}(\beta, \mu) = \mathrm{Tr}\left(e^{-\beta H + \beta\mu N}\right) = \sum_{N=0}^{\infty} e^{\beta\mu N} Z_N(\beta). \tag{4.21}$$

4.4.4 Example: Free particle in a box

Suppose we care about the quantum canonical ensemble version of a single particle in a box of volume V. Working in the coordinate basis, the Hamiltonian is

$$H = \frac{\boldsymbol{p}^2}{2m} = -\frac{\hbar^2}{2m}\nabla^2, \tag{4.22}$$

which has energy eigenstates $|\boldsymbol{k}\rangle$ specified by

$$\langle \boldsymbol{r}|\boldsymbol{k}\rangle = \frac{e^{-\boldsymbol{k}\cdot\boldsymbol{r}}}{\sqrt{V}}, \quad E_{\boldsymbol{k}} = \frac{\hbar^2 k^2}{2m}. \tag{4.23}$$

---

[11] Microstates with an indefinite number of particles span a *Fock space*, which is a set of Hilbert spaces associated with zero or more quantum particles

What are the allowed $\boldsymbol{k}$? Assuming for simplicity periodic boundary conditions for a cube of side length L, we can have $\boldsymbol{k} = \frac{2\pi}{L}(l_x, l_y, l_z)$, where the l_α are integers. So, the space of microstates is enormously larger than in the classical case: rather than 6 degrees of freedom per particle, one can have countably infinite numbers of states per particle. In the limit $L \to \infty$ the partition function becomes

$$\begin{aligned} Z &= \sum_{\boldsymbol{k}} e^{-\frac{\beta\hbar^2k^2}{2m}} = V\int \frac{d^3\boldsymbol{k}}{(2\pi)^3} e^{-\frac{\beta\hbar^2k^2}{2m}} \\ &= \frac{V}{(2\pi)^3}\left(\frac{\sqrt{2\pi m k_B T}}{\hbar}\right)^3 = \frac{V}{\lambda^3}, \end{aligned} \tag{4.24}$$

for $\lambda = h/\sqrt{2\pi m k_B T}$, which indeed coincides with our classical calculation when we use the right (adjusted) phase space measure.

What about the elements of the density matrix itself? We can compute

$$\begin{aligned} \langle \boldsymbol{r}'|\rho|\boldsymbol{r}\rangle &= \sum_{\boldsymbol{k}} \langle \boldsymbol{r}'|\boldsymbol{k}\rangle \frac{e^{-\beta E_{\boldsymbol{k}}}}{Z} \langle \boldsymbol{k}|\boldsymbol{r}\rangle \\ &= \frac{\lambda^3}{V}\int V\, \frac{d^3\boldsymbol{k}}{(2\pi)^3} \frac{e^{-i\boldsymbol{k}\cdot(\boldsymbol{r}-\boldsymbol{r}')}}{V} e^{-\frac{\beta\hbar^2k^2}{2m}} \\ &= \frac{1}{V}\exp\left(-\frac{\pi\left(\boldsymbol{r}-\boldsymbol{r}'\right)^2}{\lambda^2}\right). \end{aligned} \tag{4.25}$$

What does this mean? The diagonal elements are all $\langle \boldsymbol{r}|\rho|\boldsymbol{r}\rangle = V^{-1}$, the common expectation that the probability for finding the particle is uniform throughout the box. The off-diagonal terms are a quantum-mechanical effect, measuring the "spontaneous transition" between coordinates $\boldsymbol{r}$ and $\boldsymbol{r}'$, giving a measure of the "intensity" of the wave packet some distance from the center of the packet. Said another way, the spatial extent of the packet is a measure of the uncertainty involved in locating the particle position. This quantum mechanical effect vanishes in the $\beta \to 0$ limit, as the density matrix elements approach delta functions.

Finally, we can compute the expectation value of the Hamiltonian itself, $\langle H\rangle = \mathrm{Tr}\left(\hat{H}\hat{\rho}\right)$. I'll spare you the straightforward integrals in class – we've already calculated the partition function $Z = \mathrm{Tr}\left(e^{-\beta\hat{H}}\right)$, so the result is the last line of:

$$\begin{aligned} \langle H\rangle &= \mathrm{Tr}\left(\hat{H}\hat{\rho}\right) = \frac{\mathrm{Tr}\left(\hat{H}e^{-\beta\hat{H}}\right)}{\mathrm{Tr}\left(e^{-\beta\hat{H}}\right)} \\ &= -\frac{\partial}{\partial\beta}\ln \mathrm{Tr}\left(e^{-\beta\hat{H}}\right) \\ &= \frac{3}{2}k_B T. \end{aligned} \tag{4.26}$$

Utterly expected.

4.4.5 Example: An electron in a magnetic field

Suppose we care about the quantum canonical ensemble version of a single electron in a magnetic field. The election has spin $\hbar\hat{\sigma}/2$ and a magnetic moment $\mu_B = \frac{e\hbar}{2mc}$ (nothing to do with the chemical potential; it's just standard notation to use μ here, too), where $\hat{\sigma}$ is the Pauli spin operator.

When we apply a magnetic field, $\boldsymbol{B}$, the electron can have either spin up or spin down. If we take the applied field to be along $\hat{z}$, the configurational part of the Hamiltonian is

$$\hat{H} = -\mu_B \hat{\sigma} \cdot \boldsymbol{B}. \tag{4.27}$$

Life is easier when we work in the basis in which the Hamiltonian is diagonal, i.e.,

$$\hat{H} = -\mu_B B \hat{\sigma}_z, \tag{4.28}$$

where

$$\hat{\sigma}_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \hat{\sigma}_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \hat{\sigma}_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{4.29}$$

From this it is straightforward to calculate the density matrix:

$$\begin{aligned} \hat{\rho} &= \frac{e^{-\beta\hat{H}}}{\mathrm{Tr}\left(e^{-\beta\hat{H}}\right)} \\ &= \frac{1}{e^{\beta\mu_B B} + e^{-\beta\mu_B B}} \begin{pmatrix} e^{\beta\mu_B B} & 0 \\ 0 & e^{-\beta\mu_B B} \end{pmatrix}, \end{aligned} \tag{4.30}$$

from which we can calculate, e.g., the expectation value for σ_z:

$$\langle \sigma_z \rangle = \mathrm{Tr}\left(\hat{\rho}\hat{\sigma}_z\right) = \frac{e^{\beta\mu_B B} - e^{-\beta\mu_B B}}{e^{\beta\mu_B B} + e^{-\beta\mu_B B}} = \tanh\left(\beta\mu_B B\right), \tag{4.31}$$

an expression I'm quite sure you've seen in other classes.

4.5 Quantum indistinguishability

4.5.1 Two identical particles

Suppose we were to write down a simple two-particle Hamiltonian, for two particles of equal mass and with an interaction potential that depended only on the relative separation:

$$\mathcal{H}(1,2) = \frac{\boldsymbol{p}_1^2}{2m} + \frac{\boldsymbol{p}_2^2}{2m} + V(|\boldsymbol{r}_1 - \boldsymbol{r}_2|).$$

Clearly this Hamiltonian is symmetric under the exchange of particle label, $\mathcal{H}(1,2) = \mathcal{H}(2,1)$. Classically, our labeling of particles is meaningful (think, for instance, of running a computer simulation of a system of classical particles, in which there are Monte Carlo exchanges of particle identity that can take place), but for identical quantum mechanical particles these labels are arbitrary and convey no physical meaning.

For instance, the probability of finding two identical particles at positions $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ is given by $|\psi(\boldsymbol{r}_1, \boldsymbol{r}_2)|^2 = |\psi(\boldsymbol{r}_2, \boldsymbol{r}_1)|^2$. As long as the wavefunction is single-valued this leads to two distinct possibilities[12]:

$$\psi(\boldsymbol{r}_1, \boldsymbol{r}_2) = \begin{cases} \psi(\boldsymbol{r}_2, \boldsymbol{r}_1) & \text{for bosons} \\ -\psi(\boldsymbol{r}_2, \boldsymbol{r}_1) & \text{for fermions} \end{cases} . \tag{4.32}$$

4.5.2 N identical particles

Starting with a wavefunction for N particles, $\psi(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N)$, we generalize the above by introducing a *permutation operator* P, of which there are $N!$ possible permutation operators we might consider for our set of particles. We'll adopt the notation

$$P\psi(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N) = \begin{cases} \psi(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N) & \text{for bosons} \\ (-1)^P \psi(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_N) & \text{for fermions} \end{cases} \tag{4.33}$$

to represent the two classes of Hilbert spaces we might find ourselves in. Here we take $(-1)^P$ to represent the *parity*[13] of the permutation under question: if P can be represented by an even number of pairwise particle exchanges then $(-1)^P = 1$, and if it can be represented by an odd number of pairwise exchanges the $(-1)^P = -1$.

Note that the Hamiltonian for the particles, $\mathcal{H}$, must be symmetric: $P\mathcal{H} = \mathcal{H}$, but the Hamiltonian can admit eigenstates of different symmetries under the action of the permutation operator. The same Hamiltonian will thus allow eigenstates of either total symmetry or total anti-symmetry: the statistics one want to study must be specified independently of the Hamiltonian, and so one studies only a subspace (either the fermionic subspace, in which eigenstates are anti-symmetric, or the bosonic subspace, in which eigenstates are symmetric) of the total Hilbert space. Let's see a convenient way of representing these subspaces.

4.5.3 Product states for non-interacting particles

We consider an N-particle Hamiltonian which is just a collection of single-particle Hamiltonians[14] for free particles in a box of volume V:

$$\mathcal{H} = \sum_{\alpha=1}^{N} \frac{\boldsymbol{p}_i^2}{2m} = \sum_{\alpha=1}^{N} -\frac{\hbar^2}{2m}\nabla_\alpha^2. \tag{4.34}$$

Each of the single-particle Hamiltonians can be diagonalized by writing it in the energy basis, $|\boldsymbol{k}_\alpha\rangle$ with energy $\hbar^2\boldsymbol{k}_\alpha^2/(2m)$, and we will build our N-particle wavefunction out of these one-particle eigen-pieces.

---

[12] Because the square of the exchange operator must be the identity matrix. For single-valued functions this restriction means that there can only be a complex phase shift under the operation of a single application of the exchange operator, so the square being the identity means the phase shift can only be 0 (bosons) or π (fermions). You may have heard, though, of *anyons*! These have multi-valued wavefunctions, and in two dimensions one can find other allowed statistics without this constraint on the value of the phase shift.

[13] In Pathria's notation, what I will eventually write as η^P is denoted $\delta_P = (\pm 1)^{[P]}$

[14] "Hamiltonia"?

We define a *product state* as

$$|\boldsymbol{k}_1,\ldots,\boldsymbol{k}_N\rangle_\times \equiv |\boldsymbol{k}_1\rangle|\boldsymbol{k}_2\rangle\cdots|\boldsymbol{k}_N\rangle, \tag{4.35}$$

where in the coordinate representation the product state is

$$\langle \boldsymbol{r}_1,\ldots,\boldsymbol{r}_N|\boldsymbol{k}_1,\ldots,\boldsymbol{k}_N\rangle_\times = \frac{1}{V^{N/2}}\exp\left(-i\sum_\alpha \boldsymbol{k}_\alpha\cdot\boldsymbol{r}_\alpha\right), \tag{4.36}$$

and of course

$$\mathcal{H}|\boldsymbol{k}_1,\ldots,\boldsymbol{k}_N\rangle_\times = \left(\sum_\alpha \frac{\hbar^2\boldsymbol{k}_\alpha^2}{2m}\right)|\boldsymbol{k}_1,\ldots,\boldsymbol{k}_N\rangle_\times. \tag{4.37}$$

These product states are very convenient to work with, but they are too general! That is, they are appropriate for *distinguishable* particles, but for indistinguishable particles they do not have the correct symmetry for either bosons *or* fermions.

To show how we can compactly write either fermionic or bosonic states, let's start by defining a symbol

$$\eta = \begin{cases} 1 & \text{for bosons} \\ -1 & \text{for fermions} \end{cases}, \tag{4.38}$$

and we will write things like $|\{\boldsymbol{k}\}\rangle_+$ and $|\{\boldsymbol{k}\}\rangle_-$ for bosonic and fermionic states, respectively.

Fermionic eigenstates

We build the set of possible fermionic states by summing over all possible permutations of the product state, but including the appropriate anti-symmetrizing factor:

$$|\boldsymbol{k}_1,\ldots,\boldsymbol{k}_N\rangle_- = \frac{1}{\sqrt{N_-}}\sum_P (-1)^P P|\boldsymbol{k}_1,\ldots,\boldsymbol{k}_N\rangle_\times, \tag{4.39}$$

where $N_- = N!$ is a factor that ensures proper normalization of our fermionic eigenstate. Because of the anti-symmetrization, if there are any value $\boldsymbol{k}_\alpha$ appears more than once the whole eigenstate vanishes, and so anti-symmetrization is only possible if there are N distinct $\boldsymbol{k}_\alpha$. This is why we know there are as many distinct terms in the sum as there are particles, and thus why $N_- = N!$. For example, a three-particle anti-symmetrized state is[15]

$$|123\rangle_- = \frac{|123\rangle_\times + |231\rangle_\times + |312\rangle_\times - |213\rangle_\times - |321\rangle_\times - |132\rangle_\times}{\sqrt{6}} \tag{4.40}$$

Bosonic eigenstates

Formally, we write the bosonic states similarly, as sum of possible permutations of the product state with a factor of $(+1)^P = 1$ accounting for the parity of each permutation:

$$|\boldsymbol{k}_1,\ldots,\boldsymbol{k}_N\rangle_+ = \frac{1}{\sqrt{N_+}}\sum_P (+1)^P P|\boldsymbol{k}_1,\ldots,\boldsymbol{k}_N\rangle_\times. \tag{4.41}$$

---

[15] for convenience, let $\boldsymbol{k}_1 = 1$, etc.

Bosons, though, are allowed to have states in which the same $\boldsymbol{k}_\alpha$ appears multiple times, so computing the normalization factor is slightly more complicated. To see this, consider the bosonic state $|121\rangle_+$ (i.e., a state in which there are two identical $\boldsymbol{k}$ and one unlike $\boldsymbol{k}$. We see that

$$\begin{aligned} |121\rangle_+ &= \frac{1}{\sqrt{N_+}}\left(|112\rangle_\times + |121\rangle_\times + |211\rangle_\times + |112\rangle_\times + |121\rangle_\times + |211\rangle_\times\right) \\ &= \frac{2}{\sqrt{N_+}}\left(|112\rangle_\times + |121\rangle_\times + |211\rangle_\times\right), \end{aligned} \tag{4.42}$$

so for proper normalization of $|121\rangle_+$ we need $N_+ = 12$. The combinatorial generalization is that if each k is repeated n_k times in the N-particle bosonic state, then $N_+ = N!\prod_k n_k!$. We can see this by requiring

$$\begin{aligned} 1 &= {}_+\langle\{\boldsymbol{k}\}|\{\boldsymbol{k}\}\rangle_+ = \frac{1}{N_+}\sum_{P,P'} {}_\times\langle P'\{\boldsymbol{k}\}|P\{\boldsymbol{k}\}\rangle_\times \\ &= \frac{N!}{N_+}\sum_P {}_\times\langle\{\boldsymbol{k}\}|P\{\boldsymbol{k}\}\rangle_\times, \end{aligned} \tag{4.43}$$

but the $\langle\{\boldsymbol{k}\}|P\{\boldsymbol{k}\}\rangle$ vanish by orthogonality unless the permuted set of wavevectors matches the original, which happens $n_k!$ times for each repeated k. Thus,

$$1 = \frac{N!\prod_k n_k!}{N_+} \Rightarrow N_+ = N!\prod_k n_k!. \tag{4.44}$$

Compact notation

Actually, though, since for a fermionic state n_k can only be zero or one (again, the antisymmetrization gets rid of any states with multiply repeated k), we can combine the bosonic and fermionic notation above into:

$$|\boldsymbol{k}_1,\ldots,\boldsymbol{k}_N\rangle_\eta = \frac{1}{\sqrt{N_\eta}}\sum_P \eta^P P|\boldsymbol{k}_1,\ldots,\boldsymbol{k}_N\rangle_\times. \tag{4.45}$$

For both bosons and fermions $N_\eta = N!\prod_{\boldsymbol{k}} n_{\boldsymbol{k}}!$, and note that the states end up being uniquely specified by the set of occupation numbers, n_k, with the constraint

$$\sum_{\boldsymbol{k}} n_{\boldsymbol{k}} = N. \tag{4.46}$$

4.6 The canonical ensemble density matrix for non-interacting identical particles

We first write down the canonical density matrix for non-interacting sets of identical particles (and from it obtain the canonical partition function). This section is a bit of a technical calculation; at the end we will understand where the $1/N!$ in the classical partition function comes from, and we will see how the quantum statistics of non-interacting identical particles are approximately equivalent to introducing either attractive or repulsive classical interactions which are felt over distances comparable to the thermal de Broglie wavelength.

In the position basis this is $\langle\{\boldsymbol{r}'\}|\rho_N|\{\boldsymbol{r}\}\rangle_\eta$, where we know ρ_N will be diagonal in the energy basis:

$$\langle\boldsymbol{r}'_1,\ldots,\boldsymbol{r}'_N|\rho_N|\boldsymbol{r}_1,\ldots,\boldsymbol{r}_N\rangle_\eta=\sum_{\{\boldsymbol{k}\}_{\text{restricted}}}\left[\frac{1}{N_\eta}\sum_{P,P'}\eta^P\eta^{P'}\langle\{\boldsymbol{r}'\}|P'\{\boldsymbol{k}\}\rangle\rho(\{\boldsymbol{k}\})\langle P\{\boldsymbol{k}\}|\{\boldsymbol{r}\}\rangle\right].\tag{4.47}$$

The density matrix is (c.f. Eq. 4.17)

$$\rho_N(\{\boldsymbol{k}\})=\frac{\exp\left(-\beta\sum_{\alpha=1}^N\frac{\hbar^2k_\alpha^2}{2m}\right)}{Z_N},\tag{4.48}$$

and the "restricted" sum above makes sure that every unique indistinguishable particle state appears exactly once (correctly accounting for either bosonic or fermionic statistics). That restriction is, in fact, a bit cumbersome, so it is more convenient to sum over *all* $\{\boldsymbol{k}\}$ and then correct for any over-counting. Since the states are specified by the occupation numbers, and since (again) for fermions the $\eta^P\eta^{P'}$ cancels all contributions from terms with $n_k>1$, we can do this via

$$\sum_{\{\boldsymbol{k}\}_{\text{restricted}}}=\sum_{\{\boldsymbol{k}\}}\frac{\prod_{\boldsymbol{k}}n_{\boldsymbol{k}}!}{N!}\tag{4.49}$$

Making this change (combined with the factor of $1/N_\eta=1/(N!\prod_{\boldsymbol{k}}n_{\boldsymbol{k}}!)$), gives

$$\langle\{\boldsymbol{r}'\}|\rho_N|\{\boldsymbol{r}\}\rangle_\eta=\sum_{\{\boldsymbol{k}\}}\frac{1}{(N!)^2}\sum_{P,P'}\frac{\eta^P\eta^{P'}}{Z_N}e^{\left(-\beta\sum_{\alpha=1}^N\frac{\hbar^2k_\alpha^2}{2m}\right)}\langle\{\boldsymbol{r}'\}|P'\{\boldsymbol{k}\}\rangle\langle P\{\boldsymbol{k}\}|\{\boldsymbol{r}\}\rangle.\tag{4.50}$$

We reorder the sums and replace the sum over $\{\boldsymbol{k}\}$ with an integral to get

$$\langle\{\boldsymbol{r}'\}|\rho_N|\{\boldsymbol{r}\}\rangle_\eta=\frac{1}{(N!)^2Z_N}\sum_{P,P'}\eta^P\eta^{P'}\int\left(\prod_{\alpha=1}^N\frac{V\,d^3\boldsymbol{k}_\alpha}{(2\pi)^3}\right)e^{\left(-\beta\frac{\hbar^2k_\alpha^2}{2m}\right)}\frac{e^{-i\sum_\alpha^N(\boldsymbol{k}_{P\alpha}\cdot\boldsymbol{r}_\alpha-\boldsymbol{k}_{P'\alpha}\cdot\boldsymbol{r}'_\alpha)}}{V^N}.\tag{4.51}$$

Perhaps you feel that we have made things worse rather than better; fortunately, we are undeterred. Let's introduce a new label $\gamma=P\alpha,\ \alpha=P^{-1}\gamma$ to keep track of permutations. We'll make use of the fact that for functions / operators / variables f and g we can sum over indices $\sum_\alpha f(P\alpha)g(\alpha)=\sum_\gamma f(\gamma)g(P^{-1}\gamma)$; this allows us to focus on a particular wavevector:

$$\langle\{\boldsymbol{r}'\}|\rho_N|\{\boldsymbol{r}\}\rangle_\eta=\frac{1}{(N!)^2Z_N}\sum_{P,P'}\eta^P\eta^{P'}\prod_{\alpha=1}^N\int\frac{d^3\boldsymbol{k}_\alpha}{(2\pi)^3}e^{-i\boldsymbol{k}_\alpha\cdot\left(\boldsymbol{r}_{P^{-1}\alpha}-\boldsymbol{r}'_{(P')^{-1}\alpha}\right)-\beta\frac{\hbar^2k_\alpha^2}{2m}}.\tag{4.52}$$

The Gaussian integrals in this expression give

$$\int\frac{d^3\boldsymbol{k}_\alpha}{(2\pi)^3}e^{-i\boldsymbol{k}_\alpha\cdot\left(\boldsymbol{r}_{P^{-1}\alpha}-\boldsymbol{r}'_{(P')^{-1}\alpha}\right)-\beta\frac{\hbar^2k_\alpha^2}{2m}}=\frac{1}{\lambda^3}\exp\left(-\frac{\pi}{\lambda^2}\left(\boldsymbol{r}_{P^{-1}\alpha}-\boldsymbol{r}'_{(P')^{-1}\alpha}\right)^2\right).\tag{4.53}$$

Using this result and setting $\mu = P^{-1}\alpha$ we get

$$\langle\{\boldsymbol{r}'\}|\rho_N|\{\boldsymbol{r}\}\rangle_\eta = \frac{1}{Z_N\lambda^{3N}(N!)^2}\sum_{P,P'}\eta^P\eta^{P'}\exp\left(-\frac{\pi}{\lambda^2}\sum_{\mu=1}^{N}\left(\boldsymbol{r}_\mu - \boldsymbol{r}'_{(P')^{-1}P\mu}\right)^2\right). \tag{4.54}$$

The last step is to do one of the two sums over the permutations. We define $Q = (P')^{-1}P$, and since $\eta^P = \eta^{P^{-1}}$ we can write $\eta^P\eta^{P'} = \eta^{(P')^{-1}P} = \eta^Q$. With this, and summing over one set of $N!$ permutations, we arrive at

$$\langle\{\boldsymbol{r}'\}|\rho_N|\{\boldsymbol{r}\}\rangle_\eta = \frac{1}{Z_N\lambda^{3N}N!}\sum_{Q}\eta^Q\exp\left(-\frac{\pi}{\lambda^2}\sum_{\mu=1}^{N}\left(\boldsymbol{r}_\mu - \boldsymbol{r}'_{Q\mu}\right)^2\right). \tag{4.55}$$

We can *finally* get the canonical partition function by enforcing the normalization of the density matrix:

$$\mathrm{Tr}\left(\rho\right) = 1 \quad\Rightarrow\quad \int\prod_{\alpha=1}^{N}d^3\boldsymbol{r}_\alpha\langle\{\boldsymbol{r}\}|\rho_N|\{\boldsymbol{r}\}\rangle_\eta = 1 \tag{4.56}$$

$$\Rightarrow\quad Z_N = \frac{1}{\lambda^{3N}N!}\int\prod_{\alpha=1}^{N}d^3\boldsymbol{r}_\alpha\sum_{Q}\eta^Q\exp\left(-\frac{\pi}{\lambda^2}\sum_{\mu=1}^{N}\left(\boldsymbol{r}_\mu - \boldsymbol{r}_{Q\mu}\right)^2\right). \tag{4.57}$$

We see that the quantum mechanical partition function has within it a sum over the $N!$ permutations of identical particles, the classical result,

$$Z_N = \frac{1}{N!}\left(\frac{V}{\lambda^3}\right)^N,$$

corresponds to the term in which there are no exchanges, that is, where Q is the identity. We see that there are lots of other terms involving products of terms like

$$\exp\left(-\frac{\pi}{\lambda^2}\left(\boldsymbol{r}_1 - \boldsymbol{r}_2\right)^2\right),$$

but as $T \to \infty$, $\lambda \to 0$ and these quantum corrections vanish.

4.6.1 Statistical interparticle potential

Before we try to evaluate Eq. 4.57, let's pause to think about perturbations away from the infinite temperature limit. Clearly the lowest order correction from the classical canonical partition function involves permutations which just exchange two particles. So, let's consider the simplest possible (non-trivial) case, where $N = 2$ and $\sum_Q$ is a sum over the identity and either a symmetric or antisymmetric exchange of particles. The partition function is

$$Z_2 = \frac{1}{2!}\left(\frac{V}{\lambda^3}\right)^2\left[1 + \frac{\eta}{2^{3/2}}\frac{\lambda^3}{V}\right], \tag{4.58}$$

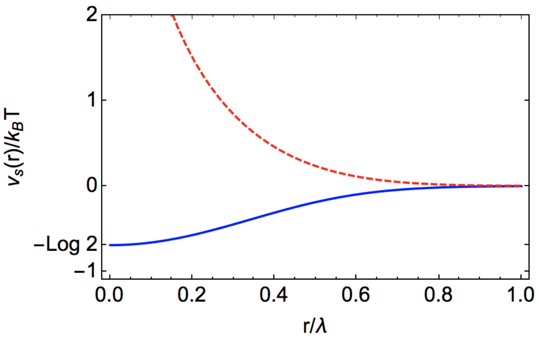


Figure 4.1: **The effective statistical interparticle potential is attractive (bosons) or repulsive (fermions)** The solid line corresponds to v_s in Eq. 4.60 with $\eta = +1$, as in the case of bosons, and the dashed line corresponds to the case of $\eta = -1$, as for particles with fermionic statistics.

and the diagonal elements of the density matrix are

$$\langle \boldsymbol{r}_1, \boldsymbol{r}_2 | \rho_2 | \boldsymbol{r}_1, \boldsymbol{r}_2 \rangle \approx \frac{1}{V^2}\left[1 + \eta \exp\left(-2\pi \frac{(\boldsymbol{r}_1 - \boldsymbol{r}_2)^2}{\lambda^2}\right)\right]. \tag{4.59}$$

This tells us that when the interparticle separation is comparable to λ the probability density for the two *non-interacting* identical particles is different from the classical result of $1/V^2$, on account of the statistics of identical particle permutations. For $\eta = 1$ there is a *greater* likelihood of finding the particles close together, and for $\eta = -1$ there is a reduction in the probability (all the way down to zero for $|\boldsymbol{r}_1 - \boldsymbol{r}|_2 \to 0$).

We can express these different statistical correlations by pretending that we have classical particles interacting according to a *statistical interparticle potential* $v_s(r)$. We'll see these kinds of statistical potentials again later in the course; the idea is that the Boltzmann weight associated with a particular separation should be exactly equal to change in the pair correlation away from the infinite temperature limit above:

$$\exp\left(-\beta v_s(r)\right) = 1 + \eta \exp\left(-2\pi \frac{r^2}{\lambda^2}\right). \tag{4.60}$$

This effective potential is shown in Fig. 4.1, again, it is a classical potential that mimics the effect of quantum correlations at high temperatures.

4.7 The grand canonical ensemble for non-interacting identical particles

Above we worked in the position basis; this let us see the relationship between the quantum and classical partition functions nicely, and demonstrated the traditional resolution to Gibbs' paradox for identical particles[16]. But explicitly calculating the sum over allowed permutations in Eq. 4.57 for either Bosonic or Fermionic states is... daunting, to say the least. We can first make our life a little bit easier by working with the canonical partition function in the basis for which the Hamiltonian is diagonal. Here

$$Z_N = \mathrm{Tr}\left(e^{-\beta H_N}\right) = \sum_{\{\boldsymbol{k}\}_{\text{restricted}}} {}_\eta\langle\{\boldsymbol{k}\}|e^{-\beta\sum_{\alpha=1}^{N} E_{\boldsymbol{k}_\alpha}}|\{\boldsymbol{k}\}\rangle_\eta. \tag{4.61}$$

We've switched from the particular single-particle Hamiltonian with only kinetic energy in the last section to arbitrary single-particle Hamiltonians that have some set of energy levels characterized by energies $E_{\boldsymbol{k}_\alpha}$ (so, these could be free particles in a box, or quantum harmonic oscillators, etc.). We still have this restriction on the $\{\boldsymbol{k}\}$, which we work around as follows.

We recall that the allowed states can be specified by the occupation numbers, $n_{\boldsymbol{k}}$, for each $\boldsymbol{k}$:

$$\begin{aligned} Z_N &= \sum_{\{\boldsymbol{k}\}_{\text{restricted}}} e^{-\beta\sum_{\alpha=1}^{N} E_{\boldsymbol{k}_\alpha}} \\ &= \sum_{\{n_{\boldsymbol{k}}\}_{\text{restricted}}} \exp\left(-\beta\sum_{\boldsymbol{k}} E_{\boldsymbol{k}_\alpha} n_{\boldsymbol{k}}\right). \end{aligned} \tag{4.62}$$

Furthermore, we have gone from sums over symmetry-restricted sets of $\boldsymbol{k}$ to restricted sums over the occupation numbers:

$$\sum_{\boldsymbol{k}} n_{\boldsymbol{k}} = N, \quad \text{and } n_{\boldsymbol{k}} = \begin{cases} 0 \text{ or } 1 & \text{fermions} \\ 0, 1, 2, \ldots & \text{bosons} \end{cases} \tag{4.63}$$

Performing this restricted sum over occupations numbers is *still* difficult, so we move to the grand canonical ensemble:

$$\mathcal{Q}_\eta = \sum_{N=0}^{\infty} z^N Z_N = \sum_{N=0}^{\infty} e^{\beta\mu N} \sum_{\{n_{\boldsymbol{k}}\}_{\text{restricted}}} \exp\left(-\beta\sum_{\boldsymbol{k}} E_{\boldsymbol{k}_\alpha} n_{\boldsymbol{k}}\right) \tag{4.64}$$

$$= \sum_{N=0} \left[\sum_{\{n_{\boldsymbol{k}}\}_{\text{restricted}}} \prod_{\boldsymbol{k}} \exp\left[-\beta\left(E_{\boldsymbol{k}} - \mu\right) n_{\boldsymbol{k}}\right]\right], \tag{4.65}$$

where the subscript η reminds us that the restriction on the sum implicitly depends on the quantum statistics in question. *At last* we see the utility of moving to the grand canonical ensemble: the double summation above – first over a restricted set of occupation numbers at fixed N, and then over all N – is equivalent to simply *summing over all values of the different occupation numbers independently for each* $\boldsymbol{k}$.

We now have

$$\mathcal{Q}_\eta = \sum_{\{n_{\boldsymbol{k}}\}_\eta} \prod_{\boldsymbol{k}} \exp\left[-\beta\left(E_{\boldsymbol{k}} - \mu\right) n_{\boldsymbol{k}}\right], \tag{4.66}$$

---

[16] i.e., now we have recovered both the factors of h and $N!$ we introduced in an ad-hoc way earlier

where $\{n_{\boldsymbol{k}}\}_\eta$ reminds us that the sum over occupation numbers is either $\sum_{n_{\boldsymbol{k}}=0}^{1}$ for fermions or $\sum_{n_{\boldsymbol{k}}=0}^{\infty}$ for bosons. We can evaluate the sums over the occupation numbers independently for each $\boldsymbol{k}$ above; explicitly, we can write

$$\begin{aligned}
\mathcal{Q}_\eta &= \sum_{n_{\boldsymbol{k}_1}, n_{\boldsymbol{k}_2}, \cdots} \left[\left(ze^{-\beta E_{\boldsymbol{k}_0}}\right)^{n_0}\left(ze^{-\beta E_{\boldsymbol{k}_1}}\right)^{n_1}\cdots\right] \\
&= \left[\sum_{n_0}\left(ze^{-\beta E_{\boldsymbol{k}_1}}\right)^{n_0}\right]\left[\sum_{n_1}\left(ze^{-\beta E_{\boldsymbol{k}_1}}\right)^{n_1}\right]\cdots \qquad (4.67)
\end{aligned}$$

for fermions this just gives us two terms per $\boldsymbol{k}$, and for bosons we get a simple geometric series *as long as that series converges*:

$$\begin{aligned}
\mathcal{Q}_- &= \prod_{\boldsymbol{k}}\left[1+\exp\left(\beta\mu-\beta E_{\boldsymbol{k}}\right)\right], \qquad (4.68) \\
\mathcal{Q}_+ &= \prod_{\boldsymbol{k}}\left[1-\exp\left(\beta\mu-\beta E_{\boldsymbol{k}}\right)\right]^{-1}, \quad \text{with } E_{\boldsymbol{k}}-\mu>0\ \forall\boldsymbol{k}. \qquad (4.69)
\end{aligned}$$

Thermodynamically we usually want to take derivatives of the log of the above expressions, so we combine them compactly as

$$\log\mathcal{Q}_\eta = -\eta\sum_{\boldsymbol{k}}\ln\left[1-\eta\exp\left(\beta\mu-\beta E_{\boldsymbol{k}}\right)\right]. \qquad (4.70)$$

From this we can calculate the usual suspects. For instance, we typically want to know how many particles we actually have for a given value of the chemical potential. Recall that classically we know that the unconditional probability of finding N particles in the system is

$$p(N) = \frac{e^{\beta\mu N}Z_N}{\mathcal{Q}},$$

and here we have products of independent single-particle states. So, we very similarly can write down the probability of having a particular set of occupation numbers:

$$p_\eta\left(\{n_{\boldsymbol{k}}\}\right) = \frac{1}{\mathcal{Q}_\eta}\prod_{\boldsymbol{k}}\exp\left[-\beta\left(E_{\boldsymbol{k}}-\mu\right)n_{\boldsymbol{k}}\right]. \qquad (4.71)$$

From this we can pull down the average occupation number of a particular state with energy $E_{\boldsymbol{k}}$, as

$$\langle n_{\boldsymbol{k}}\rangle_\eta = -\frac{\partial\ln\mathcal{Q}_\eta}{\partial\left(\beta E_{\boldsymbol{k}}\right)} = \frac{1}{\exp\left(\beta E_{\boldsymbol{k}}-\beta\mu\right)-\eta}. \qquad (4.72)$$

From this the average number of particles at fixed μ is

$$N_\eta = \sum_{\boldsymbol{k}}\langle n_{\boldsymbol{k}}\rangle_\eta = \sum_{\boldsymbol{k}}\frac{1}{z^{-1}e^{\beta E_{\boldsymbol{k}}}-\eta} \qquad (4.73)$$

and the average energy is

$$E_\eta = \sum_{\boldsymbol{k}}E_{\boldsymbol{k}}\langle n_{\boldsymbol{k}}\rangle_\eta = \sum_{\boldsymbol{k}}\frac{E_{\boldsymbol{k}}}{z^{-1}e^{\beta E_{\boldsymbol{k}}}-\eta} \qquad (4.74)$$

4.8 Ideal quantum gases

In the rest of the chapter we will specialize the above results to the case of ideal quantum gases, looking at some of the thermodynamic properties of both ideal Fermi and Bose gases. Before we dive into the details, it is worth saying a few words about what we will get out of this exploration. Classically, the ideal gas is a prime example we keep coming back to, and it provides a starting point for understanding real gases, but it's a pretty poor starting point for just about any other system (solids, dense liquids, etc).

The quantum ideal gas turns out to be much more applicable. Non-interacting systems of bosons are surprisingly accurate descriptors of photons (as one could anticipate from our earlier discussion of blackbody radiation), phonons (and giving insight into the heat capacity of solids), as well as actual dilute gases of bosons. You might be surprised to learn that the non-interacting approximation for fermions *also* apply to a range of systems! For instance, electrons are charged, and the interactions between electrons in an atom, or a material, always make a large contribution to the energy. Nevertheless, a gas of non-interacting fermions is a powerful description of atoms, metals, insulators, neutron stars, etc. Not a *free* gas of fermions: the trick is that collections of interacting fermions often act like collections of non-interacting fermions sitting in a modified external potential[17].

To specialize to (non-relativistic) idealized gases, we make the specific choice for the energies $E_{\boldsymbol{k}} = \hbar^2 k^2/(2m)$, where the energy levels have a *degeneracy* g associated with the spin s of the particles, $g = 2s+1$, and where $\sum_{\boldsymbol{k}} \to V\int d^3\boldsymbol{k}/(2\pi)^3$. The evaluation of the grand canonical partition function gives us the following results for the pressure, number density, and energy density:

$$\beta P_\eta = \frac{\ln \mathcal{Q}_\eta}{V} = -\eta g \int \frac{d^3\boldsymbol{k}}{(2\pi)^3} \ln\left[1 - \eta z e^{-\beta\frac{\hbar^2k^2}{2m}}\right], \tag{4.75}$$

$$n_\eta = \frac{N_\eta}{V} = g \int \frac{d^3\boldsymbol{k}}{(2\pi)^3} \left(z^{-1}e^{\beta\frac{\hbar^2k^2}{2m}} - \eta\right)^{-1}, \tag{4.76}$$

$$\varepsilon_\eta = \frac{E_\eta}{V} = g \int \frac{d^3\boldsymbol{k}}{(2\pi)^3} \frac{\hbar^2k^2}{2m} \left(z^{-1}e^{\beta\frac{\hbar^2k^2}{2m}} - \eta\right)^{-1}. \tag{4.77}$$

With a little more work we can manipulate these expressions into a more manageable form. For what's to come, let's define the following families of functions that are closely related to what are typically called Bose-Einstein and Fermi-Dirac integrals:

$$f_m^\eta(z) \equiv \frac{1}{(m-1)!}\int_0^\infty dx \frac{x^{m-1}}{z^{-1}e^x - \eta}; \tag{4.78}$$

these functions are closely related to the polylogarithm[18]. We're going to want m to take non-integer values, so recall that we really mean the gamma function above[19].

---

[17] See Sethna's "Statistical Mechanics: Entropy, Order Parameters, and Complexity" for a slightly expanded discussion of this point. If anyone want to talk about Landau-Fermi liquid theory that could be a cool presentation topic!

[18] The polylogarithm is defined by $\mathrm{Li}_m(z) = \sum_{\alpha=1}^{\infty} \frac{z^\alpha}{\alpha^m}$ (plus the analytic continuation in the complex plane). Note that $f_m^\eta(z) = \eta \mathrm{Li}_m(\eta z)$.

[19] $\Gamma(m) = (m-1)!$, with, e.g., $(1/2)! = \sqrt{\pi}/2$, etc.

With this definition in hand, let's make the obvious change of variables in the above thermodynamic expressions to $x = \beta\hbar^2 k^2/(2m)$, i.e., $k = \frac{2\sqrt{\pi}}{\lambda}x^{1/2}$. To show an explicit calculation, we substitute this into the equation for the pressure, integrate by parts, and end up with a very compact expression:

$$\begin{aligned}
\beta P_\eta = &= -\eta g \int \frac{d^3\boldsymbol{k}}{(2\pi)^3} \ln\left[1 - \eta z e^{-\beta\frac{\hbar^2 k^2}{2m}}\right] \\
&= -\frac{2\eta g}{\lambda^3\sqrt{\pi}} \int_0^\infty dx\; x^{1/2} \ln\left(1 - \eta z e^{-x}\right) \\
&= \frac{4g}{3\lambda^3\sqrt{\pi}} \int dx \frac{x^{3/2}}{z^{-1}e^x - \eta} \\
&= \frac{g}{\lambda^3} f^\eta_{5/2}(z).
\end{aligned} \tag{4.79}$$

Similarly, we can now compactly write the three thermodynamic expressions above as

$$\begin{aligned}
\beta P_\eta &= \frac{g}{\lambda^3} f^\eta_{5/2}(z) \\
n_\eta &= \frac{g}{\lambda^3} f^\eta_{3/2}(z) \\
\varepsilon_\eta &= \frac{3}{2} P_\eta.
\end{aligned} \tag{4.80}$$

Beautiful. These equations are a complete description of the thermodynamics of ideal Fermi and Bose gases, but of course we usually want equations of state where P is a function of the density and the temperature; here we have implicit relations between both P and n with the fugacity. We would like, therefore, to invert the middle equation[20] and know z in terms of the density, so we need to understand the behavior of the $f^\eta_m(z)$.

4.8.1 High-temperature and low-density limit of ideal quantum gases

The simplest limit to consider – in which we can continue to treat Fermions and Bosons simultaneously – is the high-T, small-n limit in which z is small. For small z one can perform a systematic expansion of the integral defining $f^\eta_m(z)$ (and you should step through this exercise!), but for brevity we will use the fact noted in footnote 18, by which we can write

$$f^\eta_m(z) = \eta \mathrm{Li}_m(\eta z) = \sum_{\alpha=1}^{\infty} \eta^{\alpha+1} \frac{z^\alpha}{\alpha^m} = z + \eta\frac{z^2}{2^m} + \frac{z^3}{3^m} + \cdots \tag{4.81}$$

We see that for $z \ll 1$, $f^\eta_m(z) \ll 1$ and hence n_η and P_η are all small, too – the calculation is nicely self-consistent. We are now in a position to find a relationship for z in terms of n (rather than the current n in terms of z in Eq. 4.80): we rearrange Eq. 4.80 as

$$z = d - \eta\frac{z^2}{2^{3/2}} - \frac{z^3}{3^{3/2}} - \cdots, \tag{4.82}$$

---

[20] you might say that in picking a name for a complicated integral we've just parameterized our ignorance, without gaining any understanding yet

where we define the *degeneracy factor* $d = n_\eta \lambda^3/g$, which is another way of characterizing the regime in which quantum effects become important (i.e., when $n_\eta \lambda^3 > g$, quantum mechanical effects become crucial to keep track of). From here, we can compute z as a power series to any order in n (plus corrections of higher order) by recursively substituting lower order solutions in. Explicitly, to lowest order $z \approx d$. To improve this to next order, we substitute it into the power series above, getting

$$z \approx d - \frac{\eta}{2^{3/2}} d^2.$$

To get the next order term, we substitute this improved approximation into the series (keeping all terms up to third order), giving

$$\begin{aligned} z &\approx d - \frac{\eta}{2^{3/2}}\left(d - \frac{\eta}{2^{3/2}} d^2\right)^2 - \frac{1}{3^{3/2}}\left(d - \frac{\eta}{2^{3/2}} d^2\right)^3 \\ &\approx d - \frac{\eta}{2^{3/2}} d^2 + \left(\frac{1}{4} - \frac{1}{3^{3/2}}\right) d^3 - \cdots, \end{aligned}$$

etc. The point is not (necessarily, unless you want to calculate certain precise quantities) to work out the numerical values of these prefactors, but rather than we can systematically and self-consistently rearrange $n_\eta(z)$ into $z(n_\eta)$ in the limit we're considering. We can then substitute this power series back into Eq. 4.80 and get the equation of state for our high-temperature, low-density quantum gas:

$$P_\eta = n_\eta k_B T \left(1 - \frac{\eta}{2^{5/2}}\left(\frac{n_\eta \lambda^3}{g}\right) + \left[\frac{1}{8} - \frac{2}{3^{5/2}}\right]\left(\frac{n_\eta \lambda^3}{g}\right)^2 + \cdots\right) \tag{4.83}$$

We'll see in the next chapter that this is our first look at a *virial expansion* of the equation of state.

4.9 Ideal Bose gases

At higher temperature we were able to work out a power series representation for the equation of state for a Bose gas; as the temperature is reduced and $d = n_+\lambda^3/g$ grows that approach is no longer useful and we must work directly with the $f_m^\eta(z)$ functions. Recall that the average occupation number of a particular energy eigenstate, Eq. 4.72, is

$$\langle n_{\boldsymbol{k}} \rangle_+ = \frac{1}{\exp\left[\beta\left(E_{\boldsymbol{k}} - \mu\right)\right] - 1},$$

which clearly cannot be a negative number. This means that $\mu < E_{\boldsymbol{k}}$ for any choice of k; given our investigation of $E_{\boldsymbol{k}} = \hbar^2 k^2/(2m)$ here that means that $\mu < 0$ and hence $0 \leq z \leq 1$.

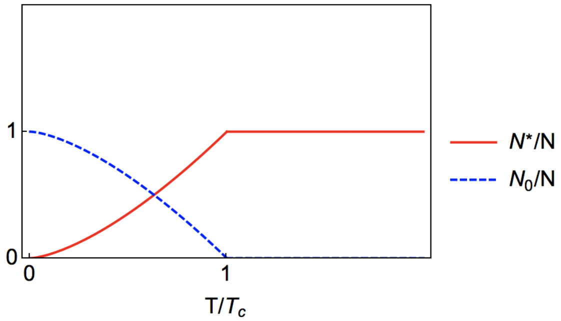


Figure 4.2: **Fraction of the normal phase and the condensed phase in an ideal Bose gas.** Since the density of excited states $n^* = gf^+_{3/2}(1)/\lambda^3 \propto T^{3/2}$, we can schematically draw the figure as shown.

Thus, since the $f^+_m(z)$ are monotonically increasing functions for $0 \leq z \leq 1$, we see that the density of excited states for the ideal Bose gas is bounded. From Eq. 4.80:

$$n_+ = \frac{g}{\lambda^3} f^+_{3/2}(z) \leq \frac{g}{\lambda^3} f^+_{3/2}(1), \tag{4.84}$$

where[21]

$$f^+_{3/2}(1) = 1 + \frac{1}{2^{3/2}} + \frac{1}{3^{3/2}} + \cdots \approx 2.612.$$

Since there is a bound on the number of excited states, what happens if we take a fixed number of particles in a fixed volume and start cooling them down? At high temperatures, the bound above is not relevant and the density of excited states is the same as the number density. Writing out the factors of λ above, though, we can see that there is a critical temperature at which the bound becomes relevant:

$$\begin{aligned} \frac{n_+\lambda^3}{g} = \frac{n_+}{g}\left(\frac{h}{\sqrt{2\pi m k_B T_c}}\right)^3 &= f^+_{3/2}(1) \\ \Rightarrow T_c(n) &= \frac{h^2}{2\pi m k_B}\left(\frac{n}{g f^+_{3/2}(1)}\right)^{2/3}. \end{aligned} \tag{4.85}$$

Below this temperature the fugacity is stuck[22] at $z = 1$; the limiting density of excited states, $n^* = gf^+_{3/2}(1)/\lambda^3$, is less than the total number density, and the rest of the particles are forced to occupy the $\boldsymbol{k} = 0$ zero-energy ground state. This is *Bose-Einstein condensation:* having a macroscopically large number of particles accumulating in just one single-particle state. The schematic growth of the number of particles in the ground state at low temperature is shown in Fig. 4.2.

---

[21] Note that $f^+_m(1) = \zeta(m)$, the Reimann zeta function

[22] Really, if N_0 is the number of particles in the ground state of $E = 0$, $z = N_0/(N_0 + 1)$

4.9.1 Pressure

We now turn to the pressure of the low-temperature phase of this system. For $T < T_c$ we have $z = 1$, and our thermodynamic expression given earlier gives us

$$\beta P_+ = \frac{g}{\lambda^3} f^+_{5/2}(1) \approx 1.341 \frac{g}{\lambda^3}, \tag{4.86}$$

a pressure which is *independent of* n and proportional to $T^{5/2}$. We can use our expression for the critical temperature to note, by the way, that *at* T_c we have

$$P(T_c)V = \frac{f^+_{5/2}(1)}{f^+_{3/2}(1)} (N k_B T_c) \approx 0.5134 N K_B T_c, \tag{4.87}$$

and we see that right at the transition the pressure of an ideal Bose gas is about half of what would be expected from a classical gas. More generally, for $T < T_c$ we can write

$$P(T) = \left(\frac{T}{T_c}\right)^{5/2} P(T_c) = \frac{f^+_{5/2}(1)}{f^+_{3/2}(1)} \frac{N - N_0}{V} k_B T, \tag{4.88}$$

showing that only the excited particles contribute to the pressure[23].

4.9.2 Heat capacity

Let's look a little closer at the transition from high- to low-temperature Bose gases by studying the heat capacity. Looking again at Eq. 4.80 to get an expression for the energy, we see the heat capacity (at constant volume and particle number) is

$$C_{V,N} = \left.\frac{\partial E_\eta}{\partial T}\right|_{V,N} = \frac{15 V g k_B}{4\lambda^3} f^+_{5/2}(z) + \frac{3 V g k_B T}{2\lambda^3} \frac{d f^+_{5/2}(z)}{dz} \frac{dz}{dT}. \tag{4.89}$$

Note that the first term contributes for the entire range of T, but the second term does not: z only appreciably varies above the critical temperature, so the second term only contributes for $T > T_c$. What are the limits here? At low temperatures we set $z = 1$, ignore the second term, and have

$$C_V = \frac{15 V g k_B}{4\lambda^4} f^+_{5/2}(1) \sim T^{3/2}, \quad \text{at low } T.$$

At high temperatures we know $z < 1$ and $dz/dT < 0$, and from monotonicity that $f^+_{5/2}(z) < f^+_{5/2}(1)$, so we expect that the heat capacity has a *maximum* at T_c. Let's see this explicitly by evaluating the derivatives in the second term above to get our full expression for the heat capacity. The first one is straightforward. Using the definition in footnote 18, we have

$$\frac{d}{dz} f^+_m(z) = \frac{1}{z} f^+_{m-1}(z),$$

---

[23] Which makes sense: only the excited fraction of the Bose gas has finite momentum

so our heat capacity is

$$C_{V,N} = \frac{3Vgk_BT}{2\lambda^3}\left(\frac{5}{2T}f^+_{5/2}(z) + \frac{f^+_{3/2}(z)}{z}\frac{dz}{dT}\right). \tag{4.90}$$

All that remains is to find how the fugacity changes with temperature. We straightforwardly do this by invoking our condition of fixed particle number in Eq. 4.80:

$$\begin{aligned} \left.\frac{dN}{dT}\right|_V = 0 \quad &\Rightarrow \quad 0 = \frac{gV}{\lambda^3}\left(\frac{3}{2T}f^+_{3/2}(z) + \frac{f^+_{1/2}(z)}{z}\frac{dz}{dT}\right) \\ &\Rightarrow \quad \frac{dz}{dT} = -\frac{3z}{2T}\frac{f^+_{3/2}(z)}{f^+_{1/2}(z)}. \end{aligned} \tag{4.91}$$

Substituting this in, we get

$$C_{V,N} = \frac{3Vgk_B}{2\lambda^3}\left[\frac{5f^+_{5/2}(z)}{2} - \frac{3}{2}\frac{\left(f^+_{3/2}(z)\right)^2}{f^+_{1/2}(z)}\right]. \tag{4.92}$$

Finally, we put all of this together. At low temperatures the heat capacity $C_v \sim T^{3/2}$. At high temperatures we can expand the power series in z and not that the heat capacity is larger than the classical value:

$$C_V/(Nk_B) = \frac{3}{2}\left(1 + \frac{n\lambda^3}{2^{7/2}} + \cdots\right).$$

What about at the transition, or close to it, when $z < 1$ but is not so small that the series expansion is helpful? We can rearrange our expression for the particle number, together with the expression for T_c, to find that to lowest order in $t = (T - T_c)/T_c$, when we *approach* T_c *from above*, that $z \approx 1 - Bt^2$, for some constant B that we could work out if we wanted to. Close to the critical temperature we can expand the polylog functions to get something of the form

$$C_{V,n} = \frac{15Vgk_B}{4\lambda^3}f^+_{5/2}(z) - b\frac{T - T_c}{T_c} \quad \text{at low } T, \text{ but } T > T_c. \tag{4.93}$$

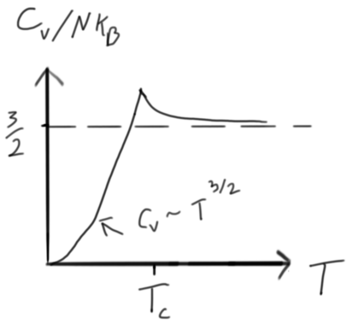


Figure 4.3: **Heat capacity of an ideal Bose Gase** (note the cusp at T_c.)

This is a heat capacity which is continuous at T_c (i.e., it will match the value from the $T < T_c$ calculation, from the first term), but which has a *discontinuous*[24] *derivative!* The behavior is schematically shown in Fig. 4.3. Now, in physics we're used to dealing with functions that are smooth and well-behaved; where do these sort of discontinuities come from? In the following chapter we'll more systematically look at interacting systems and phase transitions. First, we wrap up the chapter with a quick look at degenerate Fermi gases.

4.10 Ideal Fermi gases

Just as in the case of the ideal Bose gas, when $d = n_-\lambda^3/g$ approaches unity we can no longer usefully rely on the power series expansions of the $f_m^\eta(z)$ in Eq. 4.80, and we must start working with the full set of equations there. In the limit that $T \to 0$ we can look at the average fermi occupation number for states associated with $\boldsymbol{k}$:

$$\langle n_{\boldsymbol{k}}\rangle_- = \frac{1}{e^{\beta(E_{\boldsymbol{k}}-\mu)}+1} = \begin{cases} 1 & E_{\boldsymbol{k}} < \mu \\ 0 & \text{otherwise} \end{cases}. \tag{4.94}$$

At $T = 0$ this is just a step function, so at zero temperature all of the single-particle states up to $E_{\boldsymbol{k}} = \varepsilon_F$, the *fermi energy*, are completely filled, forming the so-called *fermi sea*. The corresponding wavenumber is[25] referred to as the *fermi wavenumber*, k_F. For an ideal gas with $E_{\boldsymbol{k}} = \hbar^2k^2/(2m)$, these are related by

$$N = \sum_{|\boldsymbol{k}|\leq k_F} g = gV\int_{k\leq k_f}\frac{d^3\boldsymbol{k}}{(2\pi)^3} = \frac{gV}{6\pi^2}k_F^3, \tag{4.95}$$

so

$$k_F = \left(\frac{6\pi^2 n}{g}\right)^{1/3}, \quad \varepsilon_F = \frac{\hbar^2}{2m}\left(\frac{6\pi^2 n}{g}\right)^{2/3} \tag{4.96}$$

Schematically the behavior of the occupation numbers are shown in Fig. 4.4.

We need to do some work to investigate the finite-temperature behavior (i.e., for large z). Here we follow Sommerfeld's approach, and first take Eq. 4.78 and perform an integration by parts:

$$f_m^-(z) = \frac{1}{m!}\int_0^\infty dx\; x^m \frac{d}{dx}\left(\frac{-1}{e^{x-\ln z}+1}\right), \tag{4.97}$$

where we've just written z^{-1} as $e^{-\ln z}$ for convenience. We then say that we know the fermi occupation itself changes very rapidly from 1 to zero across ε_F, so the derivative above

---

[24] Or does it? I've been a bit sloppy here, going back and forth about whether below T_c $z = 1$ or $z = (N_0)/(N_0+1)$. In fact, for any finite N the cusp will be smoothed out. Many more details on this to come!

[25] shockingly

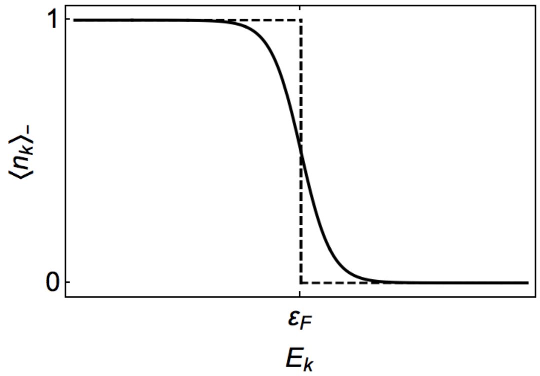


Figure 4.4: **Fermi occupation numbers** The dotted line shows the zero-temperature limit, and the solid curve shows the finite temperature result.

must be sharply peaked. We expand about this peak of the derivative (which occurs when $x = \ln z$), by setting $x = \ln z + t$ and taking the new integration variable $-\infty \leq t \leq \infty$. This trick gives us:

$$\begin{aligned} f_m^-(z) &\approx \frac{1}{m!}\int_{-\infty}^{\infty} dt\ \ (\ln z + t)^m \frac{d}{dt}\left(\frac{-1}{e^t+1}\right) \\ &= \frac{1}{m!}\int_{-\infty}^{\infty} dt\ \sum_{\alpha=0}^{\infty}\left[\binom{m}{\alpha} t^\alpha (\ln z)^{m-\alpha}\right]\frac{d}{dt}\left(\frac{-1}{e^t+1}\right) \\ &= \frac{(\ln z)^m}{m!}\sum_{\alpha=0}^{\infty}\frac{m!}{\alpha!(m-\alpha)!}(\ln z)^{-\alpha}\int_{-\infty}^{\infty} dt\ t^\alpha \frac{d}{dt}\left(\frac{-1}{e^t+1}\right). \end{aligned} \tag{4.98}$$

The last type of integral appearing above can be manipulated (exploiting the anti-symmetry of the integrand under exchange of sign of t, etc.) to give:

$$\frac{1}{\alpha!}\int_{-\infty}^{\infty} dt\ t^\alpha \frac{d}{dt}\left(\frac{-1}{e^t+1}\right) = \begin{cases} 0 & \text{if } \alpha \text{ is odd} \\ \frac{2}{(\alpha-1)!}\int_0^\infty dt \frac{t^{\alpha-1}}{e^t+1} & \text{if } \alpha \text{ is even} \end{cases}, \tag{4.99}$$

and in that last expression we recognize an expression which is just $2f_\alpha^-(1)$. So, we combine the above two equations, and exploit the fact that other people have computed the integrals associated with $f_m^-(1)$, to give the *Sommerfeld expansion*:

$$\begin{aligned} \lim_{z\to\infty} f_m^-(z) &= \frac{(\ln z)^m}{m!}\sum_{\alpha_{\text{even}}}^{\infty} 2f_\alpha^-(1)\frac{m!}{(m-\alpha)!}(\ln z)^{-\alpha} \\ &= \frac{(\ln z)^m}{m!}\left(1 + \frac{\pi^2}{6}\frac{m(m-1)}{(\ln z)^2} + \cdots\right). \end{aligned} \tag{4.100}$$

To first approximation, explicitly, we have

$$\begin{aligned}
f^-_{5/2}(z) &\approx \frac{8\left(\ln z\right)^{5/2}}{15\sqrt{\pi}}\left(1+\frac{5\pi^2}{8\left(\ln z\right)^2}+\cdots\right)\\
f^-_{3/2}(z) &\approx \frac{4\left(\ln z\right)^{3/2}}{3\sqrt{\pi}}\left(1+\frac{\pi^2}{8\left(\ln z\right)^2}+\cdots\right)\\
f^-_{1/2}(z) &\approx \frac{2\left(\ln z\right)^{1/2}}{\sqrt{\pi}}\left(1-\frac{\pi^2}{24\left(\ln z\right)^2}+\cdots\right)
\end{aligned} \tag{4.101}$$

We can now plug these results into Eq. 4.80. When $z \gg 1$ the degeneracy factor is

$$\frac{n_-\lambda^3}{g} = f^-_{3/2}(z) \approx \frac{4\left(\ln z\right)^{3/2}}{3\sqrt{\pi}}\left(1+\frac{\pi^2}{8\left(\ln z\right)^2}+\cdots\right). \tag{4.102}$$

The leading term reproduces our earlier result for the fermi energy:

$$\beta\varepsilon_F = \frac{\beta\hbar^2}{2m}\left(\frac{6\pi^2 n}{g}\right)^{2/3} = \left(\frac{3n\lambda^3}{4\sqrt{\pi}}\right)^{2/3} = \ln z \tag{4.103}$$

which gives a chemical potential of

$$\mu = k_B T\ln z \approx \varepsilon_F\left(1-\frac{\pi^2}{12}\left(\frac{k_B T}{\varepsilon_F}\right)^2\right). \tag{4.104}$$

Note that this is positive at low temperatures and negative at high temperatures, suggesting that the chemical potential changes sign at a fermi temperature $T_F \sim \varepsilon_F/k_B$. The energy density $E/V = 3P/2$, and the low-temperature pressure is

$$\beta P_- = \frac{g}{\lambda^3}\frac{8\left(\ln z\right)^{5/2}}{15\sqrt{\pi}}\left(1+\frac{5\pi^2}{8\left(\ln z\right)^2}+\cdots\right) \tag{4.105}$$

With the help of our previous expression (relating the fermi energy to $\ln z$, for instance), we can write this as a power series in temperature, as

$$P_- = \frac{2}{5}\varepsilon_F n_-\left(1+\frac{5\pi^2}{12}\left(\frac{k_B T}{\varepsilon_F}\right)^2+\cdots\right). \tag{4.106}$$

Unlike a classical gas, the degenerate fermi gas has finite pressure (and internal energy) even at zero temperature. Additionally, the heat capacity,

$$\frac{C_V}{Nk_B} = \frac{1}{Nk_B}\frac{\partial E}{\partial T} = \frac{\pi^2}{2}\frac{k_B T}{\varepsilon_F} \tag{4.107}$$

varies as the first power of temperature at low T (something which is quite general in fermi gases, regardless of the dimension). This reflects the fact that only a small fraction (of order T/T_F) of the particles are excited at temperature T, and most of the particles do not feel the effects of the finite temperatures. Each of those excited particles gains about $k_B T$ of energy, hence $C_V \sim (T/T_F)$. Additionally, we see that the heat capacity at low temperature is much smaller than the classical expectation of $3Nk_N/2$.

Chapter 5

Interacting systems

So far we have almost exclusively[1] focused on ideal systems in which the units composing the system did not interact with each other via interparticle potentials. This simplification helped us more clearly understand the structure of our statistical descriptions of macroscopic systems (and, helpfully, let us solve everything analytically) – and for quantum mechanical systems we even saw that non-interacting Bose and Fermi systems can exhibit rich/interesting behaviors.

However. Interactions are responsible for the amazing variety of phases of matter and material behaviors! Since most physical systems that we encounter cannot be described without considering interactions, in this chapter and the next we'll start building up ways of incorporating interactions into our statistical mechanical formalism. This chapter will focus on systematic expansions, in which an idealized, non-interacting system serves as a useful starting point. Implicitly, for instance, throughout the next sections you can imagine that we're trying to describe the properties of a dilute gas.

5.1 From cumulant expansions...

5.1.1 Moment expansion

Let's return to the idea of a classical system[2] and start with a general Hamiltonian in the absence of an external potential:

$$\mathcal{H} = \sum_{i=1}^{N} \frac{\boldsymbol{p}_i^2}{2m} + \mathcal{U}\left(\{\boldsymbol{r}\}\right),$$

where we've written a general interaction potential, $\mathcal{U}$, which could be an arbitrary function involving the spatial coordinates of the particles.

---

[1] With the exception of introducing some interactions in Chapter 3 to get to the Boltzmann equation

[2] With our corrected phase-space measure

In the canonical ensemble the partition function would be

$$\begin{aligned} Z(N,V,T) &= \frac{1}{N!h^{3N}}\int\left(\prod_i \mathrm{d}^3\boldsymbol{p}_i\mathrm{d}^3\boldsymbol{r}_i\right)\exp\left(-\beta\sum_{i=1}^{N}\frac{\boldsymbol{p}_i^2}{2m}\right)e^{-\beta\mathcal{U}(\{\boldsymbol{r}\})} \\ &= \frac{1}{N!}\left(\frac{V}{\lambda^3}\right)^N\int\prod_i\frac{\mathrm{d}^3\boldsymbol{r}_i}{V}e^{-\beta\mathcal{U}(\{\boldsymbol{r}\})}, \end{aligned} \tag{5.1}$$

where in the second line we've done the integral over momenta and taken the liberty of multiplying and dividing by N copies of the system volume V. We've done this because now we see terms directly related to ideal, non-interacting quantities. The prefactor is just the ideal gas partition function, and the integral is like doing an average over particle positions where there are no correlations between the particle positions – exactly as if the positions were those from an ideal gas with no interactions. Using the notation where a superscript (0) refers to these ideal-gas like quantities or averages, we can write the canonical partition function as

$$Z(N,V,T) \equiv Z^{(0)}(N,V,T)\left\langle e^{-\beta\mathcal{U}(\{\boldsymbol{r}\})}\right\rangle^{(0)} \tag{5.2}$$

$$= Z^{(0)}\sum_l\frac{(-\beta)^l}{l!}\left\langle\mathcal{U}^l\right\rangle^{(0)}. \tag{5.3}$$

This looks like a moment-based perturbative description of a system: when $\mathcal{U}=0$ we recover the ideal gas which we know how to solve, and when $\mathcal{U}$ is solve we can calculate corrections systematically. As we'll see shortly, this direct moment-based expansion is often not especially useful: at short ranges there are often strong repulsions (Pauli exclusion, or hard-core repulsion between particles, or...), so the moments of $\mathcal{U}$ need not be small. Nevertheless, working with this will lead us to an expansion which *is* useful.

5.1.2 Cumulant expansion

We know that what we often want to work with is the log of the partition function, so let's replace our moment expansion with a cumulant expansion: we see above that Z is acting like a generator of moments, so taking the log gives us a generator of cumulants:

$$\log Z = \log Z^{(0)} + \sum_{l=1}^{\infty}\frac{(-\beta)^l}{l!}\left\langle\mathcal{U}^l\right\rangle_c^{(0)}. \tag{5.4}$$

At this point we specialize the form of the interparticle potential away from complete generality, and study potentials, ϕ that are pairwise in nature,

$$\mathcal{U} = \sum_{i<j}\phi\left(\boldsymbol{r}_i-\boldsymbol{r}_j\right)\equiv\phi\left(\boldsymbol{r}_{ij}\right), \tag{5.5}$$

where we've introduce the notation $\boldsymbol{r}_{ij}$ to represent vector separations between pairs of particles, which we'll be seeing a lot of in the next sections.

Writing out the first few terms of this cumulant expansion, we have

$$\log Z = \log Z^{(0)} - \beta\left\langle\mathcal{U}\right\rangle^{(0)} + \frac{\beta^2}{2}\left(\left\langle\mathcal{U}^2\right\rangle^{(0)} - \left(\left\langle\mathcal{U}\right\rangle^{(0)}\right)^2\right) + \cdots \tag{5.6}$$

Let's evaluate these first few cumulants for our pairwise potential.

First cumulant

The first cumulant is quite straightforward, recognizing that for each pair of particles we pick as the interacting pair we're going to get copies of the same result:

$$\begin{aligned}
\langle \mathcal{U} \rangle^{(0)} &= \sum_{i<j} \int \left(\prod_{\alpha} \frac{\mathrm{d}^3 \boldsymbol{r}_\alpha}{V} \right) \phi(\boldsymbol{r}_i - \boldsymbol{r}_j) \\
&= \frac{N(N-1)}{2} \int \frac{\mathrm{d}^3 \boldsymbol{r}_1}{V} \frac{\mathrm{d}^3 \boldsymbol{r}_2}{V} \cdots \frac{\mathrm{d}^3 \boldsymbol{r}_N}{V} \phi(\boldsymbol{r}_1 - \boldsymbol{r}_2) \\
&= \frac{N(N-1)}{2V} \int \mathrm{d}^3 \boldsymbol{r} \phi(\boldsymbol{r}), \qquad (5.7)
\end{aligned}$$

where in the last line we've let $\boldsymbol{r} = \boldsymbol{r}_{12}$. There you go: someone hands you a particular interparticle potential (Lennard-Jones, or screened Coulomb, or...), and you go off, calculate an integral, and you've got the first term in the cumulant expansion.

Second cumulant

Let's write out the second cumulant as

$$\langle \mathcal{U}^2 \rangle_c^{(0)} = \sum_{\substack{i<j \\ k<l}} \left[\langle \phi(\boldsymbol{r}_{ij}) \phi(\boldsymbol{r}_{kl}) \rangle^{(0)} - \langle \phi(\boldsymbol{r}_{ij}) \rangle^{(0)} \langle \phi(\boldsymbol{r}_{kl}) \rangle^{(0)} \right]. \qquad (5.8)$$

This is a sum over $\left(\frac{N(N-1)}{2}\right)^2$ total terms, and it is helpful to divide those into three classes of terms:

All particle labels are distinct, i.e., i, j, k, l are all different indices. In this case we can look at the second moment and see

$$\begin{aligned}
\langle \phi(\boldsymbol{r}_{ij}) \phi(\boldsymbol{r}_{kl}) \rangle^{(0)} &= \int \left(\prod_{\alpha} \frac{\mathrm{d}^3 \boldsymbol{r}_\alpha}{V} \right) \phi(\boldsymbol{r}_{ij}) \phi(\boldsymbol{r}_{kl}) \\
&= \left(\int \frac{\mathrm{d}^3 \boldsymbol{r}_i}{V} \frac{\mathrm{d}^3 \boldsymbol{r}_j}{V} \phi(\boldsymbol{r}_{ij}) \right) \left(\int \frac{\mathrm{d}^3 \boldsymbol{r}_k}{V} \frac{\mathrm{d}^3 \boldsymbol{r}_l}{V} \phi(\boldsymbol{r}_{kl}) \right) \\
&= \langle \phi(\boldsymbol{r}_{ij}) \rangle^{(0)} \langle \phi(\boldsymbol{r}_{kl}) \rangle^{(0)}. \qquad (5.9)
\end{aligned}$$

Thus, these terms do not make any contribution to $\langle \mathcal{U}^2 \rangle^{(0)}$.

One particle label is shared i.e., we have $i, j, k = i, l$ as the four labels. We can make a similar argument for neglecting this class of terms, too. We first write

$$\langle \phi(\boldsymbol{r}_{ij}) \phi(\boldsymbol{r}_{il}) \rangle^{(0)} = \int \frac{\mathrm{d}^3 \boldsymbol{r}_i \mathrm{d}^3 \boldsymbol{r}_j \mathrm{d}^3 \boldsymbol{r}_l}{V^3} \phi(\boldsymbol{r}_{ij}) \phi(\boldsymbol{r}_{il}), \qquad (5.10)$$

and then change variables in the integration from $\boldsymbol{r}_i, \boldsymbol{r}_j, \boldsymbol{r}_l$ to $\boldsymbol{r}_i, \boldsymbol{r}_{ij}, \boldsymbol{r}_{il}$. This leaves us with

$$\langle \phi(\boldsymbol{r}_{ij}) \phi(\boldsymbol{r}_{il}) \rangle^{(0)} = \int \frac{\mathrm{d}^3 \boldsymbol{r}_i \mathrm{d}^3 \boldsymbol{r}_{ij} \mathrm{d}^3 \boldsymbol{r}_{il}}{V^3} \phi(\boldsymbol{r}_{ij}) \phi(\boldsymbol{r}_{il}) = \langle \phi(\boldsymbol{r}_{ij}) \rangle^{(0)} \langle \phi(\boldsymbol{r}_{il}) \rangle^{(0)}, \qquad (5.11)$$

which again means these terms do not contribute to $\langle \mathcal{U}^2 \rangle^{(0)}$.

The same pair is considered twice i.e., $i = k, j = l$. Well, the second cumulant doesn't vanish, and these are the remaining terms which we do need to keep track of. There are $\frac{N(N-1)}{2}$ terms of this form, and they all contribute identically. Thus, our complete expression for the second cumulant is

$$\langle \mathcal{U}^2 \rangle_c^{(0)} = \frac{N(N-1)}{2} \left[\int \frac{\mathrm{d}^3 \boldsymbol{r}_{ij}}{V} \phi^2(\boldsymbol{r}_{ij}) - \left(\int \frac{\mathrm{d}^3 \boldsymbol{r}_{ij}}{V} \phi(\boldsymbol{r}_{ij}) \right)^2 \right]. \tag{5.12}$$

So far, we have the following expansion for the log of the partition function:

$$\log Z = \log Z^{(0)} + \frac{N(N-1)}{2} \left[-\beta \int \frac{\mathrm{d}^3 \boldsymbol{r}}{V} \phi(\boldsymbol{r}) + \frac{\beta^2}{2} \left(\int \frac{\mathrm{d}^3 \boldsymbol{r}}{V} \phi^2(\boldsymbol{r}) - \left(\int \frac{\mathrm{d}^3 \boldsymbol{r}}{V} \phi(\boldsymbol{r}) \right)^2 \right) + \cdots \right]. \tag{5.13}$$

We now consider the thermodynamic limit, $V, N \to \infty$, and we find

$$\log Z \approx N \log \left(\frac{Ve}{N\lambda^3} \right) + \frac{N^2}{2V} \left[-\beta \int \mathrm{d}^3 \boldsymbol{r} \phi(\boldsymbol{r}) + \frac{\beta^2}{2} \int \mathrm{d}^3 \boldsymbol{r} \phi^2(\boldsymbol{r}) + \cdots \right]. \tag{5.14}$$

If we now, for instance, want to know the pressure of our system, we take the appropriate derivative and find

$$\begin{aligned} \beta P &= \frac{\partial \log Z}{\partial V} && (5.15) \\ &= \frac{N}{V} - \frac{1}{2} \left(\frac{N}{V} \right)^2 \left[-\beta \int \mathrm{d}^3 \boldsymbol{r} \phi(\boldsymbol{r}) + \frac{\beta^2}{2} \int \mathrm{d}^3 \boldsymbol{r} \phi^2(\boldsymbol{r}) + \cdots \right] + \mathcal{O}\left(\left(\frac{N}{V} \right)^3 \right). \end{aligned}$$

What have we done? We've basically written the equation of state of the system as a perturbation organized in powers of density, where the contribution at each order in density is a sum over a series of terms involving integrals of powers of the pairwise potential.

As written, this sort of cumulant expansion is still not very helpful. Why? Because for typical interactions there are large forces keeping molecules apart. For instance, one common potential is the Lennard-Jones potential:

$$\phi(\boldsymbol{r}) = 4\varepsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right], \tag{5.16}$$

where ε is related to the depth of the potential well and σ to the range over which the repulsion is felt. Where do these sorts of potential come from? The attractive r^{-6} part comes from fluctuating dipoles of electrically neutral atoms. Recall that if there were two interacting permanent dipole moments, p_1 and p_2, the potential energy would scale as $p_1 p_2 / r^3$. There are no permanent dipoles for neutral atoms, but atoms can acquire transient dipoles through quantum fluctuations. If the first atom has a transient dipole p_1 it will induce an electric field, which will in turn induce a dipole in the second atom $p_2 \sim E \sim p_1/r^3$. The resulting energy thus scales as $p_1 p_2 / r^3 \sim r^{-6}$; this is typically called the van der Waals attraction. The r^{-12} term is meant to reflect the rapid transition to strong repulsion as the atoms get very close. The exact form is not so important, and the common choice of a term like r^{-12} is simply a mathematical convenience.

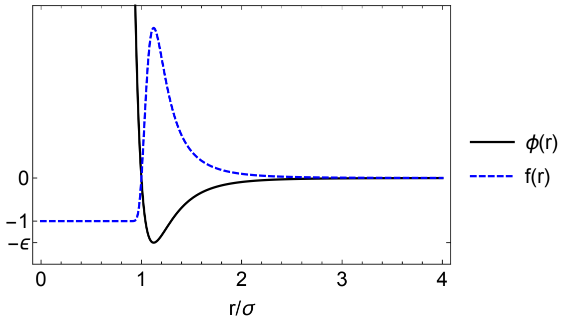


Figure 5.1: **Lennard-Jones potential and the corresponding Mayer f-function** The potential, ϕ, is shown as a solid black line and the corresponding $f = \exp(-\beta\phi) - 1$ is shown as a dashed blue line.

With such a potential, we see that *each integral in the series of series we wrote about diverges*! This is, on its face, not such a good perturbation theory. Before we integrate each term and despair, though, let's do two things. First, let's assume that the perturbative series in *density* is still okay; for a dilute gas we'll go ahead and truncate at order $\left(\frac{N}{V}\right)^2$. Second, lets sum the series first:

$$\begin{aligned} -\beta\int d^3\boldsymbol{r}\phi(\boldsymbol{r}) + \frac{\beta^2}{2}\int d^3\boldsymbol{r}\phi^2(\boldsymbol{r}) + \cdots &= \int d^3\boldsymbol{r}\left[-\beta\phi + \frac{\beta^2}{2}\phi^2 - \frac{\beta^3}{3!}\phi^3 + \cdots\right] \\ &= \int d^3\boldsymbol{r}\left[e^{-\beta\phi(\boldsymbol{r})} - 1\right] \\ &\equiv \int d^3\boldsymbol{r} f(\boldsymbol{r}), \end{aligned} \tag{5.17}$$

where in the final line we defined a new function (the *Mayer f function*) to stand for the combination $e^{-\beta\phi(\boldsymbol{r})} - 1$. In Figure 5.1 I show a plot of the Lennard-Jones potential together with the associated $f(\boldsymbol{r})$: At short distances (where the potential is diverging) the f functions converges to a value of -1, and at large distances (where the potential is vanishing) the f function converges to 0. Indeed, for reasonable potentials integrals over these f functions are perfectly well behaved, and we see that in our cumulant expansion we wrote down a series in which every term individually diverges, but the sum of the series is something we can evaluate!

Surveying our work in this section, we see that we tried to write down *cumulant expansion* – a perturbative expansion in the potential – but that for reasonable potentials we ended up having to re-express our result in terms of these Mayer f functions. It seems lie we would have been better off – and could perhaps have made more systematic progress – had we been able to *expand in powers of f*, rather than in powers of ϕ. The *cluster expansion* introduced in the next section allows us to do exactly that!

5.2 ...to cluster expansions!

Now that we have a *reason* that we might want to find an expansion of the partition function in powers of these f functions, let's do so. We continue working with our Hamiltonian with only pairwise interparticle potential,

$$\mathcal{H} = \sum_{i=1}^{N} \frac{\boldsymbol{p}_i^2}{2m} + \sum_{i<j} \phi(\boldsymbol{r}_{ij}),$$

for which the canonical partition function is

$$Z(N,V,T) = \frac{1}{N!\lambda^{3N}} \int \left(\prod_{\alpha} \mathrm{d}^3\boldsymbol{r}_\alpha\right) \prod_{i<j} e^{-\beta\phi(\boldsymbol{r}_{ij})}. \tag{5.18}$$

In order to understand how to manipulate this expression into a useful expansion, our initial goal of this section is going to be to transcribe Eq. 5.18 from math to pictures. We begin by introducing the additional bit of notation, writing the Mayer f functions as

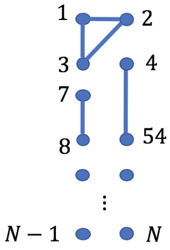
$$f_{ij} \equiv f(\boldsymbol{r}_{ij}) = e^{-\beta\phi(\boldsymbol{r}_i - \boldsymbol{r}_j)} - 1.$$

Using this notation, let's organize the terms in Z by how many powers of f they contain:

$$\begin{aligned} Z(N,V,T) &= \frac{1}{N!\lambda^{3N}} \int \left(\prod_{\alpha} \mathrm{d}^3\boldsymbol{r}_\alpha\right) \prod_{i<j} (1 + f_{ij}) \\ &= \frac{1}{N!\lambda^{3N}} \int \left(\prod_{\alpha} \mathrm{d}^3\boldsymbol{r}_\alpha\right) \left[1 + \sum_{i<j} f_{ij} + \sum_{\substack{i<j \\ k<l}} f_{ij} f_{kl} + \cdots\right] \end{aligned} \tag{5.19}$$

5.2.1 Diagrammatic representation of the canonical partition function

To help us organize the many terms that appear in Eq. 5.19, let's start representing integrals with diagrams[3]. The way we'll draw these diagrams is by (1) drawing N points, and then



$$= \int d^3\vec{r}_1 \cdots d^3\vec{r}_N \left(f_{1,2} f_{2,3} f_{1,3} f_{4,54} f_{7,8}\right)$$

$$= V^{N-7}\left(\int d^3\vec{r}_1 d^3\vec{r}_2 d^3\vec{r}_3 \left(f_{1,2} f_{2,3} f_{1,3}\right)\right)\left(\int d^3\vec{r}_4 d^3\vec{r}_{54}\, f_{4,54}\right)\left(\int d^3\vec{r}_7 d^3\vec{r}_8\, f_{7,8}\right)$$

Figure 5.2: **The contribution of a graph is a product of the linked clusters,** where each unlinked point gives a factor of the volume.

---

[3] Recall, from the section on probability, how we used a similar approach to deal with the combinatorics of relating cumulants to moments.

(2) representing f_{ij} by a line connecting points i and j. According to this prescription, an nth-order term in f corresponds to diagrams with n lines drawn.

An example of a diagram with 5 copies of f is shown in Fig. 5.2. An important observation is that the contribution of a particular diagram can be written as the product of the linked clusters it contains (and where each point without a connecting line contributes one power of the volume, V). Another important observation is that every diagram with the same structure contributes identically, i.e., independently of the *labels*; we only need to care about the number of 1-clusters, 2-clusters, triangles, etc., in a diagram.

We now have a prescription for writing the terms in Eq. 5.19 as pictures, and now we will *organize* those pictures in a particular way. We define *cluster integrals*,

$$b_l \equiv \text{Sum over contributions of } \textit{all} \text{ linked clusters of } l \text{ points.} \tag{5.20}$$

Ultimately, these b_l are the namesake of the *cluster expansion* we will construct by the end of this section. The first few cluster integrals are shown in Fig. 5.3:

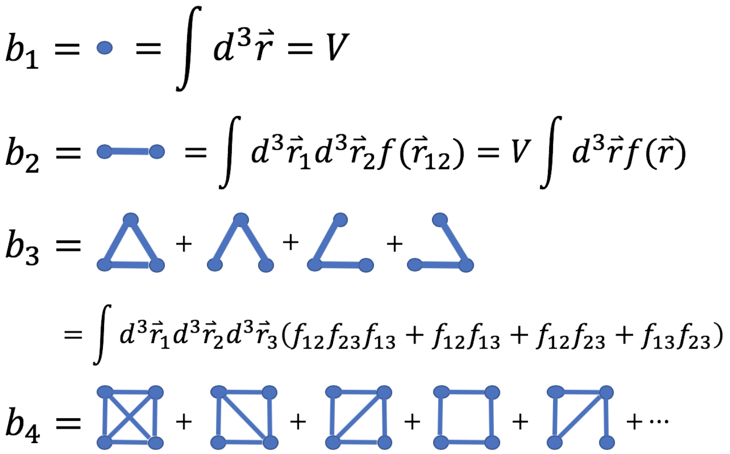


Figure 5.3: **Graphical representations of the first few cluster integrals**

5.2.2 The cluster expansion

Let's see why we care about these b_l. First, consider a diagram in which we take N points and partition it into n_1 clusters of size 1, n_2 clusters of size 2, and so forth (so, in the example in Fig. 5.2, we would have $n_1 = N - 7$, $n_2 = 2$, $n_3 = 1$, and $n_4 = n_5 = \cdots = n_N = 0$). This defines a set of numbers of clusters of different sizes, $\{n_l\}$. We noted above, though, that every diagram that looks the same contributes the same amount, regardless of the labels on the f's, so we are interested in the number of ways of choosing these sets of clusters of various sizes[4]. Let's call this number $W(\{n_l\})$, which we can compute as

$$W(\{n_l\}) = \frac{N!}{\prod_l n_l!(l!)^{n_l}}. \tag{5.21}$$

---

[4] Note that this problem of partitioning a large number, N, into sets of integers is an interesting math problem. Look up, for instance, the Hardy-Ramanujan Asymptotic Partition Formula if you're interested.

How did we arrive at this expression? Our strategy was essentially to first pick any particular partitioning into $\{n_l\}$ l-cluster. We first *overcount everything* by considering every possible permutation of the labels, giving us the $N!$ in the numerator above. Having overcounted, we now divide out to account for diagram symmetries. First, for each individual l-cluster, we note that the permutations of labels within that cluster give the same diagram (e.g., a line from point 1 to point 2 is unchanged when the labels 1 and 2 are permuted); this gives us a factor of $(l!)$ for each of the n_l clusters of size l. Additionally, we have to consider permutations of labels that replace all of the labels of one cluster of size l with a different cluster of size l (e.g., a line from point 1 to 2 and from 5 to 6 is unchanged when labels (1 and 5) and (2 and 6) are permuted); this gives us the additional factor of $n_l!$ in the denominator.

I highly encourage you to get a feel for this by playing around. Try drawing, say, 5 or 6 points, break them up into clusters of different sizes, and get a feel for the combinatorics.

With this counting of the number of different types of diagrams, we can write the canonical partition function as

$$Z = \frac{1}{N!\lambda^{3N}} \sum_{\{n_l\}_{\text{restricted}}} W(\{n_l\}) \prod_l b_l^{n_l}, \tag{5.22}$$

where the restriction is that every point has to be in a cluster, $\sum_l l n_l = N$. From the last chapter, this situation should feel quite familiar: we have another case where a constrained sum in the canonical ensemble is making our life difficult. In exactly the same way as before, this difficulty is lifted by moving to the grand canonical ensemble.

So, we write down the grand partition function, noting that the sum over all possible N of a restricted sum on the $\{n_l\}$ is equivalent to performing an *unrestricted sum* over the $\{n_l\}$:

$$\begin{aligned}
\mathcal{Q} &= \sum_{N=0}^{\infty} e^{\beta\mu N} Z_N \\
&= \sum_{\{n_l\}} \frac{1}{N!\lambda^{3N}} \left(e^{\beta\mu}\right)^{\sum_l l n_l} N! \prod_l \frac{b_l^{n_l}}{n_l!(l!)^{n_l}} \\
&= \prod_l \sum_{n_l=0}^{\infty} \left(\frac{e^{\beta\mu}}{\lambda^3}\right)^{l n_l} \frac{b_l^{n_l}}{n_l!(l!)^{n_l}}.
\end{aligned} \tag{5.23}$$

In the last line we recognize that we have written something like

$$\prod_l \sum_{n_l} \frac{1}{n_l!} [\cdots]^{n_l},$$

which is just the expansion of an exponential. Thus, we find[5]

$$\mathcal{Q} = \prod_l \exp\left[\left(\frac{e^{\beta\mu}}{\lambda^3}\right)^l \left(\frac{b_l}{l!}\right)\right], \tag{5.24}$$

---

[5] if you're comparing these expression with Pathria, or other texts, please note that different authors use slightly different conventions in the definitions of the cluster integrals; in particular, sometimes the b_l account for the factors of $l!$ in that divide it in the expressions I'm writing.

i.e.,

$$\boxed{\Rightarrow \log \mathcal{Q} = \sum_{l=1}^{\infty} \left(\frac{e^{\beta\mu}}{\lambda^3}\right)^l \left(\frac{b_l}{l!}\right)} \tag{5.25}$$

Before we make use of this expression for the grand canonical partition function, it is worth thinking for a moment about the graphical expressions we wrote for cumulants and moments. The graphical interpretation of the above result is that the log of the sum over all graphs is equivalent to the sum over only the connected clusters.

5.3 Virial expansion for a dilute gas

We consider, now, a dilute gas, and our goal for this section is to write a *virial expansion* for the equation of state of the dilute gas, i.e., something of the form

$$\beta P = B_1(T)\frac{N}{V} + B_2(T)\left(\frac{N}{V}\right)^2 + B_3(T)\left(\frac{N}{V}\right)^3 + \cdots , \tag{5.26}$$

where we've written temperature-dependent coefficients B_i for the powers of density. For a dilute gas we can certainly exploit the extensivity condition to write the grand potential as

$$-\beta\mathcal{G} = \log\mathcal{Q} = -\beta\left(E - TS - \mu N\right) = \beta PV. \tag{5.27}$$

Just as in Chapter 4 when we discussed the ideal gas in the grand canonical ensemble, we see that the grand potential is proportional to the volume; comparing with Eq. 5.25 we see that in fact each cluster integral $b_l \propto V$. We thus define a version of the cluster integrals with this volume dependence removed,

$$\bar{b}_l \equiv \frac{b_l}{V},$$

in terms of which we can write the pressure as

$$\beta P = \sum_l \left(\frac{e^{\beta\mu}}{\lambda^3}\right)^l \frac{\bar{b}_l}{l!}. \tag{5.28}$$

At the same time, we know that, specifying the chemical potential μ in the grand canonical ensemble we can compute the expected number density of particles as

$$n = \frac{N}{V} = \frac{1}{V}\frac{\partial \log\mathcal{Q}}{\partial\left(\beta\mu\right)} = \sum_l l\left(\frac{e^{\beta\mu}}{\lambda^3}\right)^l \frac{\bar{b}_l}{l!}. \tag{5.29}$$

Simplifying the notation by writing a dimensional version of the fugacity, $x \equiv e^{\beta\mu}/\lambda^3$, we arrive at the equation of state for a dilute gas:

$$\begin{aligned} \beta P &= \sum_l x^l \frac{\bar{b}_l}{l!} \\ n &= \sum_l l x^l \frac{\bar{b}_l}{l!}. \end{aligned} \tag{5.30}$$

This is... not exactly what we typically want. Just as in the chapter on quantum statistical mechanics, what we want is the pressure as a function of density, but what we have is the pressure as a function of fugacity together with a separate equation relating the number density to the fugacity. We encountered this exact situation in looking at ideal quantum gases, and we resolve it in the same way.

So, we write out the series expansion for n,

$$n = \bar{b}_1 x + \bar{b}_2 x^2 + \frac{\bar{b}_3}{2} x^3 + \cdots \tag{5.31}$$

Rearranging this (and noting that $\bar{b}_1 = 1$), we have

$$x = n - \bar{b}_2 x^2 - \frac{\bar{b}_3}{2} x^3 - \cdots , \tag{5.32}$$

which we can self-consistently solve to any desired order by substituting the ith-order solution for x into the $(i+1)$th-order expression. Starting with the first order approximation $x_1 \approx n$, we get

$$\begin{aligned} x_1 &= n + \mathcal{O}\left(n^2\right) \\ x_2 &= n - \bar{b}_2 n^2 + \mathcal{O}\left(n^3\right) \\ x_3 &= n - \bar{b}_2 \left(n - \bar{b}_2 n^2\right)^2 - \frac{\bar{b}_3}{2}\left(n - \bar{b}_2 n^2\right)^3 + \mathcal{O}\left(n^4\right) \\ &= n - \bar{b}_2 n^2 + \left(2\bar{b}_2^2 - \frac{\bar{b}_3}{2}\right) n^3 + \mathcal{O}\left(n^4\right), \end{aligned} \tag{5.33}$$

and so on. We can now substitute this into the equation for the pressure; to get an expansion correct to order i we use the x_i expression obtained by the above procedure. So, for instance, to third order we write (collecting all terms up to third order in the number density)

$$\begin{aligned} \beta P &= \bar{b}_1 x + \frac{\bar{b}_2}{2} x^2 + \frac{\bar{b}_3}{3!} x^3 \\ &= n - \bar{b}_2 n^2 + \left(\bar{b}_2^2 - \frac{\bar{b}_3}{3}\right) n^3 + \mathcal{O}\left(n^4\right). \end{aligned} \tag{5.34}$$

In general, we have a result of the form

$$\beta P = n + \sum_{l=2}^{\infty} B_l(T) n^l, \tag{5.35}$$

as desired! The first term, $B_1 = 1$, reproduces the ideal gas result. The second term is

$$B_2(T) = -\frac{\bar{b}_2}{2} = -\frac{1}{2}\int \mathrm{d}^3\boldsymbol{r}\left[e^{-\beta\phi(\boldsymbol{r})} - 1\right], \tag{5.36}$$

which diagrammatically is $-1/2$ times what we drew as a line between two points. We calculate the third term diagrammatically, as illustrated in Fig. 5.4, and find

$$B_3(T) = -\frac{1}{3}\int \mathrm{d}^3\boldsymbol{r}_{12}\mathrm{d}^3\boldsymbol{r}_{13} f(\boldsymbol{r}_{12}) f(\boldsymbol{r}_{13}) f(\boldsymbol{r}_{12} - \boldsymbol{r}_{13}) \tag{5.37}$$

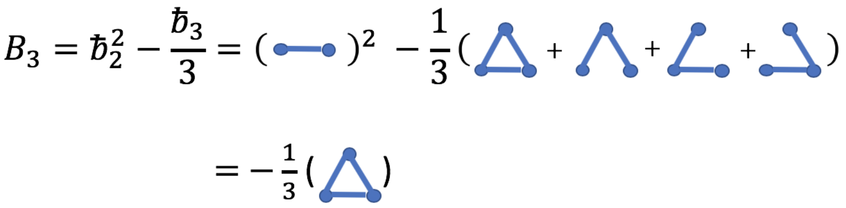


Figure 5.4: **Diagrammatic calculation of the third virial coefficient**

That calculation shows the general result, that in these terms all of the “1-particle reducible diagrams” – in which the removal of one point leads to disjoint clusters in the diagram – cancel; the general result (consider working through the B_4 term to verify your understanding) is

$$B_l(T) = -\frac{l-1}{l!}\bar{d}_l, \tag{5.38}$$

where $\bar{d}_l$ represents a sum over all 1-particle irreducible clusters of size l. In the next section we'll evaluate these sorts of virial coefficients for a “typical” gas, i.e., one governed by a potential similar to a Lennard-Jones form.

5.4 The van der Waals equation

To see how (or if!) the sort of virial expansion we just worked out is useful, let's consider just the first correction to the ideal gas equation of state – the term involving $B_2(T)$ – for the Lennard-Jones potential,

$$\phi(\boldsymbol{r}) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right], \tag{5.39}$$

which, again, is a simple but reasonable approximation to the interaction between two neutral molecules separated by some distance r.

5.4.1 The second virial coefficient for a Lennard-Jones interaction

The most straightforward approach is just to jump from the potential to the definition of the second virial coefficient that we derived in the last section,

$$\begin{aligned} B_2(T) &= \frac{-1}{2}\int \mathrm{d}^3\boldsymbol{r}\left(e^{-\beta\phi(r)} - 1\right) \\ &= -2\pi\int \mathrm{d}r\; r^2\left(e^{-\beta\phi(r)} - 1\right) \end{aligned} \tag{5.40}$$

That integral looks unfamiliar, but one can express the answer in terms of the Kummer confluent hypergeometric function[6] ${}_1F_1(a;b;z)$, and the Gamma function:

$$B_2(T) = \frac{-\pi\sigma^3\left(\beta\varepsilon\right)^{1/4}}{3\sqrt{2}}\left[\Gamma\left(\frac{-1}{4}\right){}_1F_1\left(\frac{-1}{4};\frac{1}{2};\beta\varepsilon\right) + 2\sqrt{\beta\varepsilon}\,\Gamma\left(\frac{1}{4}\right){}_1F_1\left(\frac{1}{4};\frac{3}{2};\beta\varepsilon\right)\right] \tag{5.41}$$

Is this helpful? This is, indeed, what you get by plugging things into, e.g., Mathematica, and you could plot the behavior of this second virial coefficient as you vary the temperature for various values of the Lennard-Jones parameters. This "plug-and-chug" approach, while technically correct, obscures the physical picture of how we expect a dilute gas to behave.

5.4.2 Approximate but physical treatment of B_2

Let's do a better job of capturing the physics by making a *rougher* approximation to the molecular interactions. In particular, let's replace the Lennard-Jones potential with the function

$$\phi(r) = \begin{cases} \infty & r < \sigma \\ -\varepsilon\left(\frac{\sigma}{r}\right)^6 & r > \sigma \end{cases}. \tag{5.42}$$

This approximate potential captures the following main features of the Lennard-Jones potential: it has the van der Waals r^{-6} at large distances, it has an attractive well of depth ε at short distances, and it is strongly (in fact, *infinitely*!) repulsive within a characteristic size of the molecular core. Note that this simpler form makes it clear that σ is like the characteristic *diameter* of one of the interacting molecules.

With this simpler potential we can compute the second virial coefficient as

$$\begin{aligned} B_2(T) &= -\frac{1}{2}\int \mathrm{d}^3\boldsymbol{r}\,\left(e^{-\beta\phi(r)} - 1\right) \\ &= -\frac{1}{2}\left[-\frac{4\pi\sigma^3}{3} + 4\pi\int_\sigma^\infty \mathrm{d}r\; r^2\left(e^{\beta\varepsilon\left(\frac{\sigma}{r}\right)^6} - 1\right)\right]. \end{aligned} \tag{5.43}$$

The first term above reflects an *excluded volume* in the potential we have written down. Keeping in mind that we are trying to describe a dilute gas at reasonably high temperatures, let's further approximate the second term by assuming $\beta\varepsilon \ll 1$, that is, that we are at high T relative to the well depth. We will use this approximation to write a series expansion for the exponential of the potential:

$$e^{\beta\phi} \approx \beta\varepsilon\left(\frac{\sigma}{r}\right)^6 + \mathcal{O}\left(\beta\varepsilon\right)^2.$$

With this approximation, the integral can be easily carried out:

$$\begin{aligned} B_2(T) &= -\frac{1}{2}\left[\left(-\frac{4\pi\sigma^3}{3}\right) + \frac{4\pi\sigma^3}{3}\beta\varepsilon\right] \\ &= \frac{\Omega}{2}\left[1 - \beta\varepsilon\right], \quad \text{where } \Omega = \frac{4\pi\sigma^3}{3}. \end{aligned} \tag{5.44}$$

---

[6] As a series expansion, ${}_1F_1(a;b;z) = \sum_{n=0}^{\infty}\frac{a^{(n)}z^n}{b^{(n)}n!}$, where $a^{(n)}$ is the "rising factorial," $a^{(0)} = 1$, $a^{(1)} = a$, $a^{(n)} = a(a+1)\cdots(a+n-1)$, etc.

This is a nice, compact expression for the second virial coefficient: we see a small perturbative correction that comes in at high temperatures, with an overall scale set by the excluded volume of the molecules.

A few comments are in order, directed at the question of "*when should this approximation be reasonable?*" Two obvious conditions stand out.

First, this is clearly meant to be a low-density expansion, but what we do mean by "low" in this context? As with any series expansion, we want the ratio of consecutive terms to be small, and the ratio of the correction to the ideal term is $\sim B_2 \frac{\rho^2}{\rho} \sim \frac{\Omega}{\rho^{-1}}$. Note that Ω is close to an atomic volume, so one can think of the above ratio as $\sim \frac{\rho_{gas}}{\rho_{liquid}}$. That is, when the density of the system is such that the gas is close to liquifying, we should assume our series approximation is insufficient.

Second, in our expansion of the attractive tail, we assumed that $\beta\varepsilon \ll 1$. If $\beta\varepsilon \gtrsim 1$ for these attractive potentials, we could not have done a reasonable series expansion in the first place. In general, for low temperatures the ground state of a system with attractive interactions is not a gas, but a dense collection of molecules sticking together.

There is an additional consideration, which speaks to our ability to factor out the Ω term in the series expansion we wrote. The fact that we could do this speaks to the sense that the *short-ranged* part of the potential dominates the integral we had to do. In this context there is a sense in which potentials which fall of faster than r^{-3} are "short ranged;" for potentials like $\phi \sim r^{-3}$ there are terms which are logarithmic in the size of the system, and for even longer-ranged potentials the whole expansion fails.

5.4.3 The van der Waals equation

Finally, we can write down the equation of state based on our approximate treatment of the second virial coefficient:

$$\begin{aligned} \beta P &= n + \frac{n^2}{2}\Omega\left(1-\beta\varepsilon\right) \\ \Rightarrow \beta\left(P + \frac{n^2\Omega}{2}\varepsilon\right) &= n\left(1 + \frac{n\Omega}{2} + \cdots\right). \end{aligned} \tag{5.45}$$

Since we're only being accurate to order n^2 in the above expression, we're free to re-write the right-hand side as

$$n\left(1+\frac{n\Omega}{2}+\cdots\right) \approx \frac{n}{1-\frac{n\Omega}{2}} + \cdots .$$

Using this rearranging, we arrive at the van der Waals equation of state:

$$\boxed{P = \frac{Nk_BT}{V - \frac{N\Omega}{2}} - \frac{\varepsilon\Omega}{2}\left(\frac{N}{V}\right)^2,} \tag{5.46}$$

which is often presented as

$$\left(P + a\left(\frac{N}{V}\right)^2\right)(V - bN) = Nk_BT.$$

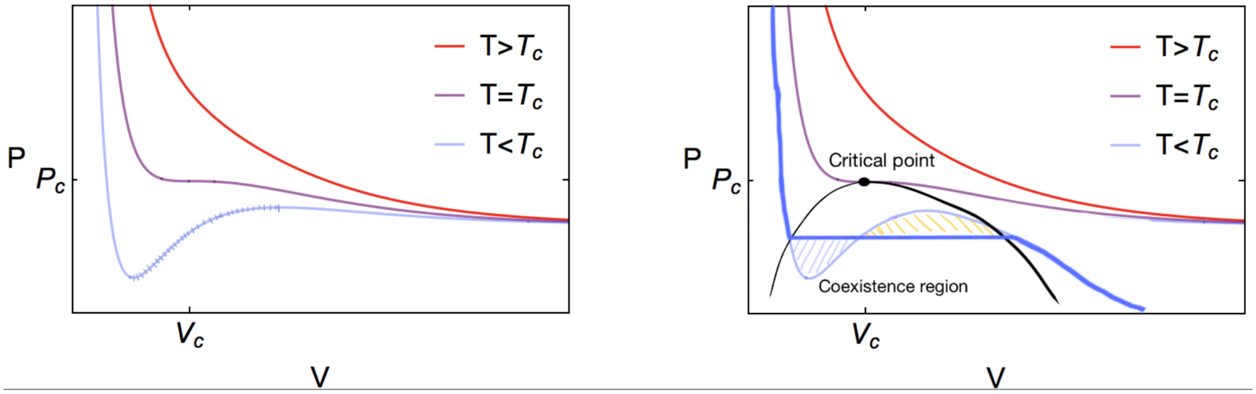


Figure 5.5: **van der Waals equation of state** (Left) The van der Waals equation of state for a dilute gas above, at, and below the critical temperature. (Right) The Maxwell equal-area construction for finding the coexistence region.

The van der Waals equation does a reasonable job of describing dilute gases, which is certainly how we derived it, but ultimately we would like to be able to describe *dense* phases of matter (going from weakly interacting systems to strongly interacting ones). To that end, and as an introduction to the phase transition material to come, we spend a bit more time thinking about what the van der Waals equation says.

We start with the left side of Fig. 5.5, which shows the isotherms predicted by the van der Waals equation for three different temperatures. For high temperatures we see curves that are basically the same as the ideal gas isotherms we encountered back when we were thinking about thermodynamics as a phenomenological theory. The pressure diverges not at $V = 0$ but at some finite value of V, thanks to the excluded volume effects, but otherwise this looks quite familiar. On the other hand, at lower temperatures the isotherms look very strange, developing a minimum at small volumes; precisely at a critical temperature there is part of the corresponding isotherm where the curve is flat.

Clearly, the low-temperature curves, with $T < T_c$, are unphysical (and they are, after all, in a parameter regime we did not expect our theory to work), which we can see in two ways. Early on, we argued in a problem set that thermodynamic stability requires $\delta P \delta V < 0$, which is violated on portions of an isotherm where $\frac{\partial P}{\partial V} > 0$. One can also see this using our our statistical mechanical relations. We recall that

$$\langle N \rangle = \frac{\partial \log \mathcal{Q}}{\partial(\beta\mu)} = V \left.\frac{\partial P}{\partial \mu}\right|_{T,v},$$

and

$$\left\langle N^2 \right\rangle = \frac{\partial^2 \log \mathcal{Q}}{\partial(\beta\mu)^2} = \frac{\partial \langle N \rangle}{\partial(\beta\mu)} = k_B T \left.\frac{\partial N}{\partial \mu}\right|_{T,v}.$$

Dividing these expressions we get the ratio of the number fluctuations to the mean and relate that ratio to the isothermal compressibility:

$$\frac{\langle N^2 \rangle - \langle N \rangle^2}{\langle N \rangle} = \frac{1}{\beta V} \left.\frac{\partial N}{\partial P}\right|_{T,V} = \frac{-1}{\beta V} \left.\frac{\partial N}{\partial V}\right|_{P,T} \left.\frac{\partial V}{\partial P}\right|_{N,T} = \frac{N}{V} k_B T \kappa_T. \tag{5.47}$$

For a stable system the compressibility should be positive, and since $\kappa_T = \frac{-1}{V}\frac{\partial V}{\partial P}\Big|_T$, we see again that the van der Waals curves for $T < T_c$ are unphysical.

Maxwell construction

So, our equation contains obviously unphysical features, but it also schematically *does* remind us of the phase behavior of real fluids, schematically illustrated on the right of Fig. 5.5. Can we make some ad hoc changes to the theory to get at the right answer? One reason for hope is that qualitatively our theory both gives the right behavior for $T \geq T_c$ *and* gives us a clear indication that it needs help for $T < T_c$ – theories that have this self-consistent limitation on their regime of validity often contain important clues to system behavior even outside of where they technically work.

To motivate the Maxwell construction which attempts to correct the van der Waals picture, let's consider changes in the chemical potential as we move along an isotherm below the critical temperature. From the above we know $N = V\frac{\partial P}{\partial \mu}\Big|_T$, giving us $d\mu = \frac{V}{N}dP$ at constant T. Integrating this up, starting from some reference point A, we have

$$\mu(T,P) = \mu(T,P_A) + \int_{P_A}^{P} dP' \frac{V(T,P')}{N}.$$

The qualitative behavior of the chemical potential as the isotherm is traversed is schematically illustrated in Fig. 5.6. There are some portions where the curves is well-behaved, but there are a range (determined by an *equal-area consideration*) over which the function μ is multi-valued!

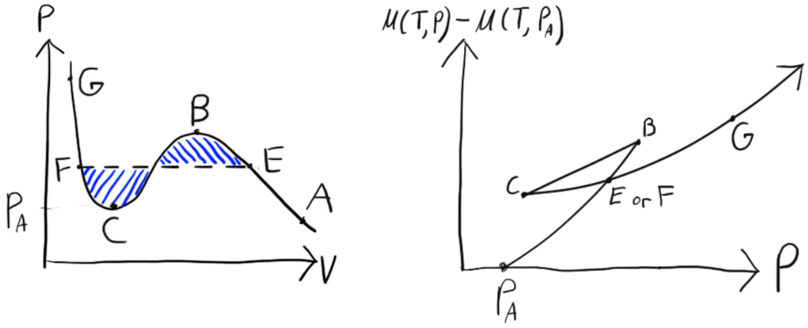


Figure 5.6: **Moving along an isotherm** from a reference point at state point A through the sequence $A \to B \to C \to G$, we calculate the chemical potential relative to that at the starting point. In addition to $B \to C$ being thermodynamically unstable, we see the portions $E \to B$ and $C \to F$ are metastable. In equilibrium, the system will have the minimum μ, and follow the path $A \to E \to F \to G$.

In equilibrium the number of particles N_α in phase α will adjust so as to minimize the Gibbs free energy $\sum_\alpha \mu_\alpha N_\alpha$, and so when confronted with a multi-valued set of possible chemical potentials we *choose the lowest one*. In the figure, the isotherm $A \to B \to C \to G$ gets replaced by the isotherm $A \to E \to F \to G$, corresponding to the straight line we drew in the coexistence region in Fig. 5.5. Notice, by the way, that the part of the original isotherm which gets replaced is *not* just the part where the thermodynamic stability condition was violated, but is actually a larger region! In these additional parts of the isotherm the system is *metastable*, rather than being strictly stable or unstable.

This Maxwell construction qualitatively succeeds, in that it now gives us largely the correct physical behavior, but it certainly feels weird! It required us to start with an approximate theory, consider regions of that theory where the behavior was clearly unphysical, and then try to sweep any unpleasantness under the rug. This was all done, additionally, in a regime where it is not at all clear that simply doing a better job with our perturbative expansion would be of any help. In the next chapter we'll take a more principled approach to these issues.

Chapter 6

Phase transitions

In the last few chapters we have been preoccupied with various methods of calculating *partition functions*, both for ideal and for interacting systems. In principle, once we know a partition function we can calculate any thermodynamic properties that we might be interested in. As we discussed a particular perturbative approach to calculating the partition function for a dilute gas – and then had to patch it up via Maxwell's construction in order to describe both a gas and a liquid state – you may be wondering: if it is even clear that a single partition function can actually describe multiple phases?

In fact, perhaps you have an even deeper concern. Phase transitions are characterized by discontinuities in the free energy or its derivatives, and the free energy can be written as

$$Z = e^{-\beta F} = \mathrm{Tr}\left(e^{-\beta H}\right),$$

where I'm using the trace notation to indicated a schematic sum over all of the degrees of freedom in the Hamiltonian, H. But most H that we study are non-singular functions of their degrees of freedom, so the partition function is just a sum of a large number of terms, each of which is exp (some analytic function). How could such a sum be non-analytic, and can a single partition function even describe a phase transition (let alone multiple phases)?

A deep understanding of the answer to these questions, together with an understanding of how near critical points whole classes of disparate systems behave in the same way, has only been understood for, say, the last 50 years[1]. In this chapter we will discuss mean field models of phase transitions, critical points, Landau theory for phase transitions, the relationship between fluctuations and correlations, and the scaling hypothesis for behavior near critical points. This will lead us right up to the precipice of the renormalization group[2]; time permitting I will try to throw together a few bonus lectures on this.

6.1 Mean field condensation

As a way of sliding from the last chapter to this one we'll return again to the van der Waals equation, but now with a somewhat different approach. In the section on interacting systems

---

[1] Perhaps making the subject the most "modern" of the graduate physics curriculum? Hmm...

[2] Or, since an anomalously large number of the professors I took graduate classes used British English, the "renormalisation group."

we were confronted with the problem of computing, say, the canonical partition function

$$Z(V,T,N) = \frac{1}{N!h^{3N}} \int \prod_i \mathrm{d}^3\boldsymbol{p}\mathrm{d}^3\boldsymbol{q}\ \exp\left(-\beta \sum_i \frac{p_i^2}{2m}\right) \exp\left(-\beta \sum_{i<j} \phi(\boldsymbol{r}_{ij})\right), \tag{6.1}$$

and with a great deal of effort we constructed a perturbative expansion to do just that in the high-temperature / low-density regime. We then muddled along and argued we could work with the resulting equations even when the system was cooled below the T_c associated with the gas transitioning to a liquid. It felt dubious.

Our goal is still to approximate the partition function, but let's consider instead a *mean field* calculation, in which we assume that the potential energy contribution of a particle is computed based on its interactions with an average – and uniform – background of particles. First, rather than specifying that we are working with a Lennard-Jones pairwise potential, let's just think of any potential which (1) has a "hard-core" component for $r < \sigma$, (2) is sufficiently short-ranged, and (3) has some attractive part of the potential, where the potential well has integrated area $-u$. In our approximation for the Lennard-Jones potential, for instance, we would have $-u = -\varepsilon\Omega$. Next, we first consider the actual (fluctuating) distribution of particle density,

$$n(\boldsymbol{r}) = \sum_i \delta\left(\boldsymbol{r} - \boldsymbol{r}_i\right),$$

in terms of which we could write the potential energy of our system as

$$U\left(\{\boldsymbol{r}_i\}\right) = \frac{1}{2} \int \mathrm{d}^3\boldsymbol{r}\mathrm{d}^3\boldsymbol{r}'\ n(\boldsymbol{r})n(\boldsymbol{r}')\phi(\boldsymbol{r} - \boldsymbol{r}'). \tag{6.2}$$

We now assume that the system has uniform density, and approximate $n(\boldsymbol{r}) \approx n = N/V$. Then the typical energy, $\tilde{U}$, is

$$\begin{aligned} \tilde{U} &= \frac{n^2}{2} \int \mathrm{d}^3\boldsymbol{r}\mathrm{d}^3\boldsymbol{r}'\phi(\boldsymbol{r} - \boldsymbol{r}') \\ &= \frac{n^2V}{2} \int \mathrm{d}^3\boldsymbol{r}\phi(r) \\ &= \frac{-un^2V}{2}. \end{aligned} \tag{6.3}$$

We substitute this average energy into Eq. 6.1 for the canonical partition function, and we get

$$Z = \frac{1}{N!\lambda^{3N}} \exp\left(\frac{\beta u N^2}{2V}\right) \int \prod_i \mathrm{d}^3\boldsymbol{r}_i. \tag{6.4}$$

We make one last approximation before we do the remaining integral over spatial coordinates, and that is a Ω/V expansion in how we treat the excluded volume effects. To first order in

(Ω/V) we write

$$\begin{aligned}
\int \prod_i \mathrm{d}^3\boldsymbol{r}_i &= V\,(V-\Omega)\,(V-2\Omega)\cdots(V-(N-1)\Omega) \\
&= V^N\left[1-\frac{1}{V}\left(\Omega+2\Omega+\cdots+(N-1)\Omega\right)+\cdots\right] \\
&= V^N\left[1-\frac{1}{V}\frac{N(N-1)}{2}\Omega+\cdots\right] \\
&\approx \left(V-\frac{N\Omega}{2}\right)^N . \qquad (6.5)
\end{aligned}$$

Together with our mean-field estimate of the effect of the attractions, then, our partition function is

$$Z(V,T,N)=\frac{\left(V-\frac{N\Omega}{2}\right)^N}{N!\lambda^{3N}}\exp\left(\frac{\beta u N^2}{2V}\right), \qquad (6.6)$$

from which we can compute the pressure in the canonical ensemble as

$$\begin{aligned}
\beta P_{\text{canonical}} = \frac{\partial \log Z}{\partial V} &= \frac{N}{V-\frac{N\Omega}{2}}-\beta\frac{u}{2}\frac{N^2}{V^2} \\
\Rightarrow P_{\text{canonical}} &= \frac{Nk_BT}{V-\frac{N\Omega}{2}}-\frac{u}{2}\left(\frac{N}{V}\right)^2 . \qquad (6.7)
\end{aligned}$$

This is precisely the van der Waals equation again! But here, rather than derive it in the context of a perturbative expansion in the density and the temperature, we have derived it in the context of an *approximation of uniform density.* And certainly a uniform density approximation is the sort of thing we expect to be equally valid in both the gas and liquid states, hence why we expect that we ought to be able to describe multiple phases with the same Hamiltonian and the same partition function.

6.1.1 Maxwell construction, once again

Let's pause to better understand how to connect this picture with the liquid-gas phase transition. Recall that one of the problems in the canonical picture is understanding what happens in the coexistence region. We've already seen several examples of how moving from the canonical to the grand canonical ensemble can make our life easier, so for practice (and variety), let's see an example of another Legendre transform and work in the Gibb's canonical ensemble, i.e., the isobaric ensemble, in which we imagine our system coupled to a piston which keeps the pressure controlled. Our partition function in this ensemble becomes

$$\mathcal{Z}(P,T,N)=\int_0^\infty dVe^{-\beta PV}Z(V,T,N)=\int_0^\infty dVe^{\psi}(V), \qquad (6.8)$$

where $\psi(V)=-\beta PV+\log Z(V,T,N)$. We now approximate this expression by the saddle point method we saw back in Chapter 2, and say there is some particular value of V which

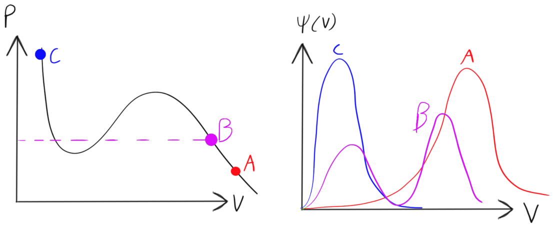


Figure 6.1: **Saddle-point picture for the condensation isotherm** Left shows a $T < T_c$ isotherm, right schematically sketches the function $\psi(V)$ for different points along the isotherm. The critical pressure is when the heights of the two purple peaks are precisely equal.

maximizes $\psi(V)$, which we will call V_{max}. Thus

$$\mathcal{Z}(P,T,N) \approx e^{\psi(V_{max})}.$$

We can find this V_{max} by extremizing ψ:

$$\frac{\partial \psi}{\partial V} = 0 \Rightarrow 0 = -\beta P + \frac{\partial Z}{\partial V} = -\beta(P - P_{\text{canonical}}(V)). \tag{6.9}$$

If $T > T_c$ this is fine, but we know that for $T < T_c$ sometimes there are three possible V to check for the canonical pressure. In this regime one V will correspond to a local *minimum* of ψ, so there are two candidate V_{max}, as sketched in Fig. 6.1. We would say

$$\mathcal{Z} \approx e^{\psi(V_{gas})} + e^{\psi(V_{liquid})},$$

and we again say that $\mathcal{Z}$ is dominated by exponential with the largest argument. Thus, as the temperature changes and the relative ordering of $\psi(V_{gas})$ and $\psi(V_{liquid})$ switch we can get a discontinuity in our system[3].

Notice, by the way, that by extensivity we can write

$$\exp\left[-\beta PV - \beta E + \beta TS\right] = \exp\left(-\beta\mu N\right),$$

so as we anticipated last chapter, choosing V to minimize ψ at fixed N is like choosing the lowest value of μ. Additionally, what is the critical value of pressure at which the discontinuity we sketched above can be found? It must be when $\psi(V_{gas}) = \psi(V_{liquid}) \Rightarrow \psi(V_{liquid}) - \psi(V_{gas}) = 0$. Writing the difference as an integral of a derivative:

$$\begin{aligned} 0 = \psi(V_{liquid}) - \psi(V_{gas}) &= \int_{V_{gas}}^{V_{liquid}} dV \frac{d\psi}{dV} \\ &= \beta \int_{V_{gas}}^{V_{liquid}} dV \left(P_{\text{canonical}} - P_c\right); \end{aligned} \tag{6.10}$$

In that last integral we recognize precisely the Maxwell equal-area construction for finding the isotherm.

---

[3] Note, again, that it is only a discontinuity in the assumption that $N \to \infty$; otherwise everything is smooth

6.2 The law of corresponding states

Having seen two different derivations of the van der Waals equation, let's write it in a simplified form,

$$P = \frac{k_B T}{v-b} - \frac{a}{v^2}, \tag{6.11}$$

where $v = V/N$ is just the volume per particle, and a and b are parameters that depend on the microscopic details of the systems under study (generically related to the characteristic size of the interacting particles, and how strong the attractive interactions are).

We have seen that this equation goes from being monotonic to non-monotonic at some critical value of temperature, and that at T_c there is a point where the critical isotherm is flat (with both the first and second derivative vanishing. Where does is this critical point? We find it by[4] writing

$$\begin{aligned} \frac{\partial P}{\partial v} = 0 &= -\frac{k_B T}{(v-b)^2} + \frac{2a}{v^3} && (6.12) \\ \frac{\partial^2 P}{\partial v^2} = 0 &= \frac{2k_B T}{(v-b)^3} - \frac{6a}{v^4}. && (6.13) \end{aligned}$$

Dividing the first equation by the second equation gives us $\frac{v_c - b}{2} = \frac{v_c}{3} \Rightarrow v_c = 3b$, and we can then plug this critical volume per particle into the above expressions to find the critical point on the critical isotherm:

$$v_c = 3b, \quad k_B T_c = \frac{8a}{27b}, \quad P_c = \frac{a}{27b^2}. \tag{6.14}$$

Suppose we wanted to know if the van der Waals equation was actually any good, and we convinced someone to do an experiment. In it's basic form, comparing experimental data to the van der Waals equation would simultaneously test both the quality of the van der Waals equation *and the ability to measure the microscopic parameters* that enter the van der Waals equation, a and b. But, in general, measuring the microscopic Hamiltonian is really hard! Better would be to try to locate the critical point of the fluid in question, noting that the van der Waals equation predicts that the ratio

$$\frac{P_c v_c}{k_B T_c} = \frac{3}{8} = 0.375 \tag{6.15}$$

is independent of the microscopic details, and should be the same for all fluids! Experimentally, this estimate is okay[5] Many real substances have a range for this ratio between 0.28 and 0.33 (carbon tetrachloride is about 0.27, argon is about 0.29, He^4 is about 0.31, water is

---

[4] A more elegant approach is to note that the van der Waals equation can be written as a cubic polynomial in v: $pv^3 - (pb + k_B T)v^2 + av - ab = 0$. For $T > T_c$ there is just one real root of this equation, and for $T < T_c$ there are three real roots; precisely at the critical point all three roots must be real and identical, so we must be able to write the equation as $p_c(v - v_c)^3 = 0$. Comparing the coefficients of this to the more general cubic expression immediately gives us the critical pressure, volume, and temperature.

[5] Different textbooks will compare 3/8 to the experimental numbers I'm about to quote and declare the agreement "good" or "bad" or "a little high." Agreement with experiment is in the eye of the beholder.

pretty low at 0.23) but similar substances seem to differ from the 3/8 ratio in similar ways; this is encouraging when we think about the types of approximations we made in deriving the van der Waals equation in the first place. For instance, we assumed the attractive potential was isotropic, so the fact that diatomic gases have a particular ratio while monatomic gases a different one feels consistent with the way we constructed our theory.

Even better, this suggests that we write down a *dimensionless version* of the van der Waals equation, in which we measure everything relative to the value at the critical point. Defining $P_r = P/P_c$, $v_r = v/v_c$, and $T_r = T/T_c$, we can write the equation of state as

$$P_r = \frac{8}{3}\frac{T_r}{v_r - \frac{1}{3}} - \frac{3}{v_r^2}, \tag{6.16}$$

a universal equation of state that all fluids obeying the van der Waals equation of state would be expected to obey. Remarkable! With no other parameters, all fluids are expected to have the same equation of state, and similarly all thermodynamic properties which are derived from the equation of state should also be in correspondence. So, if two different systems are in states that are characterized by the same values of v_r and T_r, one would predict their P_r would match as well: this is the *law of corresponding states.*

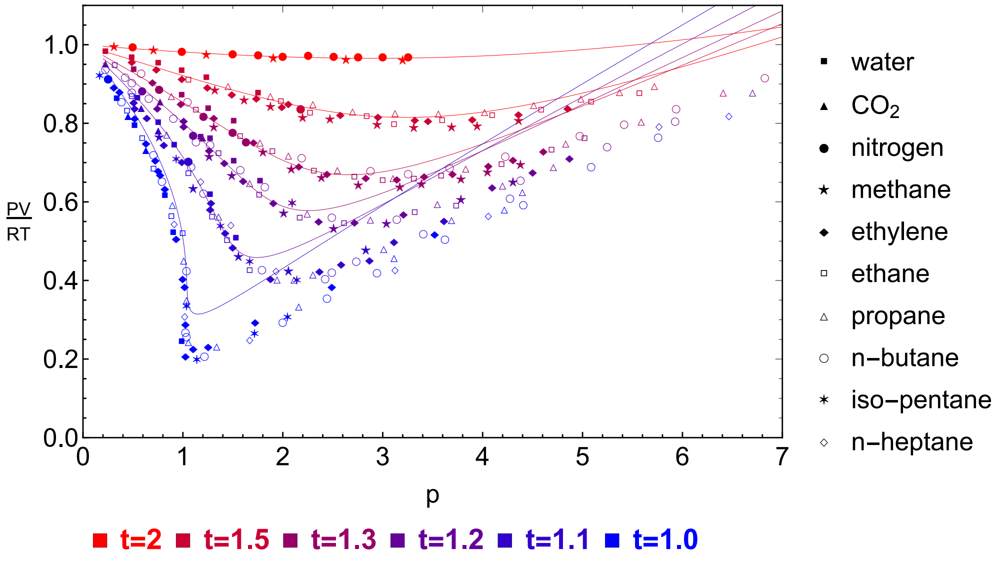


Figure 6.2: **Compressibility vs reduced pressure for several substances** Thin lines are curves generated by the van der Waals equation of state at various reduced temperatures; reduced temperature for the actual measurements is indicated by color. Data taken from H. E. Stanley, *Introduction to Phase Transitions and Critical Phenomena*, Oxford University Press, 1971

This is our first look at a type of *universality*, albeit of a different character than we will see when we look at universality near a critical point. The law of corresponding states is expected to hold not only close to the critical point, but in fact *everywhere in the phase diagram.* Experimentally the law of corresponding states is often well satisfied, even by fluids that do not obey the van der Waals equation! Generally, the expectation set forth by the corresponding states hypothesis is that we should be able to write the equation of state as

$$\frac{P}{P_c} = f\left(\frac{T}{T_c}, \frac{v}{v_c}\right) \tag{6.17}$$

for some function f which might be related to the van der Waals derivation, but might be completely different. A demonstration of this correspondence is shown in Fig. 6.2, showing the ratio PV/T vs reduced pressure for several substances. None of them follow the van der Waals equation particularly well, but do demonstrate the expectation that, for instance, this ratio when plotted against reduced pressure at a fixed reduced temperature should be independent of the fluid being measured.

6.3 Critical point behavior of a van der Waals fluid

To see why we might have expected the law of corresponding states, *at least near a critical point*, let's imagine expanding the equation of state close to the critical point[6]. We define the reduced variables

$$\pi = P_r - 1 = \frac{P - P_c}{P_c}, \quad \nu = v_r - 1, \quad \tau = T_r - 1,$$

in terms of which the reduced equation of state becomes

$$3\nu^3 + \pi(1+\nu)^2(2+3\nu) = 8\tau(1+\nu)^2. \tag{6.18}$$

We'll use this simplified form to study the relationship between our reduced variables and various thermodynamic quantities, and the compare those relationships to experimentally obtained results.

Critical isotherm: Along the critical isotherm, for which $\tau = 0$, we can easily expand Eq. 6.18 for close to the critical point (i.e., for small π and ν), and we find

$$\pi \approx -\frac{3}{2}\nu^3 \Rightarrow (P - P_c) \sim (v - v_c)^3. \tag{6.19}$$

Volume differences: We next look at how ν depends on τ as the critical point is approached from low temperatures. Rewriting Eq. 6.18 as

$$3\nu^3 + 8(\pi - 1)\nu^2 + (7\pi - 16\tau)\nu + 2(\pi - 4\tau) = 0,$$

we next recognize that near the critical point the coexistence curve is symmetric. This means that near the critical point, when we consider the polynomial in ν, we should have two roots that are approximately equal in magnitude (with opposite signs), and a third root which is very close to zero. Comparing this consideration to the above equation means that

$$\pi - 4\tau \approx 0 \Rightarrow \pi \approx 4\tau, \tag{6.20}$$

and so for the other possible values of ν we need to solve $\nu^2 + 8\tau\nu + 4\tau = 0$. Doing so gives

$$\lim_{T\to T_c^-} \nu \approx \pm 2|\tau|^{1/2} \Rightarrow \lim_{T\to T_c^-} (v_{gas} - v_{liquid}) \sim (T_c - T)^{1/2}. \tag{6.21}$$

---

[6] afterwards, we'll see why this approach has made a subtle but tremendously important mistake!

Critical isochore: Finally, for this comparison, we look at the isothermal compressibility, which is essentially determined by

$$\kappa_\tau \sim -\left.\frac{\partial \nu}{\partial \pi}\right|_\tau \approx \frac{2}{7\pi + 9\nu^2 - 16\tau}.$$

If we approach the critical point along the critical isochore ($\nu = 0$) from the high-temperature side, we get

$$\lim_{T\to T_c^+} -\left.\frac{\partial \nu}{\partial \pi}\right|_\tau \approx \frac{1}{6\tau} \Rightarrow \lim_{T\to T_c^+} \kappa_T(v_c) \sim (T - T_c)^{-1} \tag{6.22}$$

We did these calculations in the context of the van der Waals equation of state, but actually all we have really assumed is that our system is (a) mechanically stable and (b) analytic, in that close to the critical point we could expand the pressure like

$$P(T,v) = P_c + \alpha(T - T_c) - a(T - T_c)(v - v_c) + \frac{b}{2}(T - T_c)(v - v_c)^2 - \frac{c}{6}(v - v_c)^3 + \cdots,$$

where mechanical stability tells us that the coefficient $a > 0$ above T_c and $c > 0$ at T_c. What we arrive at is a prediction of *universal singular behavior* near the critical point, with various exponents characterizing the strength of the singularities.

Writing the above results using conventional names for the exponents, we can write

$$\begin{aligned} \lim_{T\to T_c^+} \kappa_T(v_c) &\sim (T - T_c)^{-\gamma} \\ \lim_{T\to T_c^-} (v_{gas} - v_{liquid}) &\sim (T_c - T)^{\beta} \\ (P - P_c) &\sim (v - v_c)^{\delta}. \end{aligned} \tag{6.23}$$

In our *mean field* calculation we had

$$\gamma = 1, \quad \delta = 3, \quad \beta = 1/2.$$

Experimentally, it is found that there *are* singularities in all of these quantities near the critical point, but the exponents are

$$\gamma \approx 1.2, \quad \delta \approx 4.8, \quad \beta \approx 0.32$$

On the one hand, that's not too bad! The δ exponent doesn't look great, but β and γ aren't so far off. On the other hand, usually in physics we are willing to make "spherical cow" approximations as we build our theories – i.e. capturing the basic, often qualitative behavior of some phenomenon of interest – as long as we are convinced that systematically improving our approximations will lead to systematically closer quantitative agreement with reality. The (perhaps surprising!) fact is that until the development of the theory of modern critical phenomena is was not possible *even in principle* to account for the difference between, say, $\gamma = 1$ in the theory and $\gamma \approx 1.2$ in real life. After all, as emphasized a few paragraphs above, all we have really done to come up with our theoretical critical exponents was assume stability and analyticity near the critical point! The physical systems are indeed stable, so apparently we made some subtly but mightily wrong assumption when we treated the equation of state as analytic near the critical point.

As a hint at what went wrong, recall from the last chapter that the ratio $\langle N^2\rangle / \langle N\rangle \propto \kappa_T$, but then we showed that near the critical point that the isothermal compressibility *diverges* near the critical point. That is, the fluctuations in particle number are enormous relative to the mean number: in short, our "mean field" assumption of uniform density *cannot* be correct close to the critical point. Somehow, by neglecting these fluctuations we were led to qualitatively correct beliefs about the behavior of fluids (corresponding states, divergences near the critical point), but were left in a quantitative pickle, with no apparent systematic way to improve our understanding of the exponents. The final few sections of this course will introduce a few of the big ideas in critical phenomenon; for those interested in more details I particularly like Goldenfeld's *Lectures on Phase Transitions and the Renormalization Group* and Kardar's *Statistical Physics of Fields.*

6.4 Another mean field theory, more critical exponents

6.4.1 Mean-field Ising model

So far we have been focusing our set of examples heavily on fluids (ideal gases, dilute gases, van der Waals fluids below T_c...), and here we will briefly discuss a[7] mean field theory for the Ising model.

We consider a nearest-neighbor Ising model on a d-dimensional hypercubic lattice composed of spins $\{s\}$; the Hamiltonian for the spins, with coupling constant J in an external field B, is

$$\mathcal{H} = -J\sum_{\langle ij\rangle} s_i s_j - B\sum_i s_i, \tag{6.24}$$

and we are interested in the magnetization

$$m = \frac{1}{N}\langle s_i\rangle = \frac{1}{N\beta}\frac{\partial(\log Z)}{\partial B}. \tag{6.25}$$

A mean field approximation for this problem follows by writing the interactions between neighboring spins *assuming the fluctuations of neighboring spins away from the average is small.* That is, we approximate

$$\begin{aligned} s_i s_j &= ([s_i - m] + m)([s_j - m] + m) \\ &= (s_i - m)(s_j - m) + m\left((s_i - m) + (s_j - m)\right) + m^2 \\ &\approx m(s_i + s_j) - m^2, \end{aligned} \tag{6.26}$$

and our mean field Hamiltonian is just

$$\mathcal{H} \approx \frac{JNqm^2}{2} - (Jqm + B)\sum_i s_i, \tag{6.27}$$

---

[7] Just a comment: mean field theories are not unique descriptions of a system! There are often many ways to generate a mean field theory and the results need not be identical. In general, all mean field theories for the same system will share the same scaling near a critical point, but will typically differ in their calculations of of non-universal features.

where q is the number of nearest neighbors ($q(d) = 2d$ for the hypercubic lattice). We have transformed our problem into a *non-interacting* Ising model in an effective external field, where

$$B_{eff} = B + Jqm.$$

The partition function is now easily expressible as N copies of the non-interacting case:

$$Z = e^{-\beta JNqm^2/2}\left(e^{-\beta B_{eff}} + e^{\beta B_{eff}}\right)^N = 2^N e^{-\beta JNqm^2/2}\cosh^N\left(\beta B_{eff}\right). \tag{6.28}$$

Finally, we self-consistently determine what the value of m is using $m = \frac{1}{N}\langle s_i\rangle = \frac{1}{N\beta}\frac{\partial(\log Z)}{\partial B}$; the result is

$$m = \tanh\left(\beta B + \beta Jqm\right). \tag{6.29}$$

6.4.2 Critical point behavior

First, in the absence of an external field, our mean field equation is $m = \tanh(\beta Jqm)$, and since $\tanh x \approx x - x^3/3 + \cdots$ the slope of m near the origin is βJq. If $\beta Jq < 1$, the only solution to the mean field equation is $m = 0$; if, however, $\beta Jq > 1$ there are three solutions: $m = 0$ or $m = \pm m_0$ (and the $m = 0$ solution turns out to be unstable, just like the putative middle solution for the volume at a particular pressure when $T < T_c$ in the van der Waals equation).

Given this zero-field critical temperature, $k_B T_c = Jq$, let's define the reduced *inverse* temperature $\tau = T_c/T$, let's apply the hyperbolic trig identity

$$\tanh\left(a+b\right) = \frac{\tanh a + \tanh b}{1 + \tanh a \tanh b}$$

to Eq. 6.29 to obtain

$$\tanh\left(\beta B\right) = \frac{m - \tanh\left(m\tau\right)}{1 - m\tanh\left(m\tau\right)}. \tag{6.30}$$

Close to the critical point, where we expect both the field and the magnetization to be small, we can expand this to obtain

$$\beta B \approx m(1-\tau) + m^3\left(\tau - \tau^2 + \frac{\tau^3}{3} + \cdots\right) + \cdots \tag{6.31}$$

From this, we can read off some critical exponents. In the absence of an external field we can approach T_c from below, and we see

$$m^2 \sim \frac{T_c - T}{T_c} + \cdots \Rightarrow m \sim \pm\left(T_c - T\right)^{1/2}. \tag{6.32}$$

Conventionally, this critical exponent is called β, and here $\beta = 1/2$.

We can also look at how the external field and the magnetization are related along the crtical isotherm (here, denoted $\tau = 1$). We immediately see

$$B \sim m^3; \tag{6.33}$$

this critical exponent is conventionally called δ, and here $\delta = 3$.

We can also easily look at the isothermal magnetic susceptibility, $\chi_T = \frac{\partial m}{\partial B}\big|_T$, and study how this changes as we vary T. Differentiating our series expansion of the equation of state with respect to B gives

$$\frac{1}{k_B T} = \chi_T(1-\tau) + 3m^2\chi_T\left(\tau - \tau^2 + \frac{\tau^3}{3}\right). \tag{6.34}$$

For $T > T_c$ the only solution is $m = 0$, so immediately

$$\chi_T = \frac{1}{k_B\,(T - Tc)}. \tag{6.35}$$

For $T < T_c$, we substitute in our result that $m \sim (T_c - T)^{1/2}$, and similarly find $\chi_T \sim |T - T_c|^{-1}$. This critical exponent is conventionally called $-\gamma$, and here $\gamma = 1$.

Comparison with exact results

You'll notice that the critical exponents we've just computed do not depend on the dimension, d, of the lattice. In $d = 1$ the mean field theory is disastrously wrong, as in 1D it turns out that there isn't even a phase transition! In higher dimensions the qualitative features of our calculation are correct (there is a phase transition; there are power-law divergences of the quantities we've studied; etc.). In $d \geq 4$, the mean field calculation turns out to give the *correct* critical exponents, so that's neat!

What about $d = 2$ and $d = 3$? Here's a table[8] (in $d = 2$ there is an exact solution for the Ising model; in $d = 3$ they have been determined via a great deal of numerical effort):

	Mean field	Exact result ($d = 2$)	Numerical result ($d = 3$)
β	1/2	1/8	≈ 0.32
δ	3	15	≈ 4.8
γ	1	7/4	≈ 1.2

What is going on, here?! The mean field results look the same as the mean field theory for the van der Waals equation, but perhaps you suspect that's just because I've shuffled variable names around to make things look good. But by comparing the $d = 3$ results with the experimental results for real liquid-gas critical behavior you should be convinced that we really are talking about the same critical exponents. Apparently, then, mean field models of ferromagnetism and mean field models of fluids give the same critical point behavior, *and get the answer wrong in the same non-obvious way!*

This is evidence for *universality at the critical point*, where apparently there is a single theory which describes the essential physics at the critical point for magnets, for the liquid-gas transition, for the 3D Ising model, and many other seemingly unrelated systems. We're living the dream! In physics we're always trying to strip away as much of the unnecessary

---

[8] There are other critical exponents, which I'm not telling you about yet but which fit the pattern. There's α (for the divergence of the heat capacity), there's ν (for how the correlation length depends on temperature close to the critical point), and there's η (which describes the long-range behavior of the two-point correlation function *precisely at* the critical point).

detail from a system as we can, and apparently near a critical point "Nature" does all the hard work for us!

As an aside, the pattern we saw above as we varied the dimension of our model under study is pretty generic: when you write down a mean field theory there is a dimension at or below which which the theory fails completely (called the *lower critical dimension*, d_l), a dimension at or above which the theory gives the right answers (called the *upper critical dimension*, d_c). For $d_l < d < d_c$ mean field theory often returns crudely correct phase diagrams that are wrong or, worse, misleading near critical points. Sadly for mean field theory, life tends to happen in between d_l and d_c.

6.5 Landau's phenomenological theory

If seemingly very different mean field theories – for very different physical systems! – give rise to the same essential behavior near their respective critical points, your instinct is that there ought to be a unified way of looking at phase transitions that reveals *why* this should be the case. Landau's phenomenological theory[9] of phase transitions serves exactly this role, focusing on understanding the universal behavior of physical systems based on two general considerations: analyticity and symmetry. The Landau approach to phase transitions is typically only qualitatively correct (as we are about to see, it gives exactly the same critical exponents as mean field theory), but it (a) let's us understand universality and (b) is extremely straightforward, letting you compute mean field critical exponents for systems in different universality classes with typically great rapidity.

Landau theory postulates the existence of an object, $\mathcal{L}$, called the Landau free energy density (it is not a thermodynamic free energy density, and it need not be convex). The theory starts by identifying an *order parameter* of the system, η, which is a quantity that is typically zero in a high-temperature or disordered phase, and non-zero in an ordered phase: For the liquid-gas transition one can take the difference in densities between the phases ($\eta = v_{gas} - v_{liquid}$), for the Ising model magnet we can take the magnetization ($\eta = m$), for a superconductor it is related to off-diagonal long-range order in the one-particle density matrix, etc. Order parameters for a particular system need not be unique, and depending on the nature of the problem they can be scalars, vectors, etc., although here we will assume it is a scalar.

Once we have identified an order parameter, we want to construct an $\mathcal{L}$ that will act like a free energy density (and has dimensions of energy per volume), insofar as we will compute thermodynamic quantities by taking appropriate derivatives of $\mathcal{L}$. We construct $\mathcal{L}$ via the following constraints:

1. The state of the system is specified by the *global minimum* of $\mathcal{L}$ with respect to η.
2. $\mathcal{L}$ *must be* consistent with the symmetries of the system.
3. Near the critical point, $\mathcal{L}$ is an analytic function of η and any coupling constants, which

---

[9] To be distinguished from a version of Landau theory which is motivated by systematic calculation starting with a microscopic Hamiltonian. This more complicated version is typically not more insightful.

I'll denote as $\{K\}$ here. Thus, for instance, for a spatially uniform system we can write

$$\mathcal{L} = \sum_{n=0} a_n(\{K\})\eta^n \tag{6.36}$$

4. In the disordered phase the order parameter should be $\eta = 0$, while it should be small and non-zero in the ordered phase.

The Landau free energy density for a magnetic system

Let's see how these constraints let us build $\mathcal{L}$ for a particular example: we'll choose the kind of Ising model we just discussed in the last section. By the third and the fourth constraint, near T_c we can expand $\mathcal{L}$ as a Taylor series, and since we expect η to be small, we don't need to go to very high order. We'll write

$$\mathcal{L} = \sum_{n=0}^{4} a_n\eta^n, \tag{6.37}$$

where the coefficients a_n could depend on the Ising model coupling term J, the external field B, the temperature T, or whatever else happens to be in our model.

Additionally, from the first constraint we have that $\mathcal{L}$ is extremized by solving

$$\frac{\partial \mathcal{L}}{\partial \eta} = a_1 + 2a_2\eta + 3a_3\eta^2 + 4a_4\eta^3 = 0. \tag{6.38}$$

Since we want $\eta = 0$ to be the solution for $T > T_c$, we have $a_1 = 0$.

Finally, what about the symmetry constraint? Consider our Ising model in the absence of any external field, $B = 0$: certainly the Hamiltonian here is invariant under the flipping of every spin[10], and we expect that the probability of finding the system with a particular value of the magnetization has the property $P(\eta) = P(-\eta)$. We expect, since we want $\mathcal{L}$ to behave like a free energy, that $P \sim \exp(-\beta\mathcal{L})$, so *we require* for this model that $\mathcal{L}(\eta) = \mathcal{L}(-\eta)$. The fact that $\mathcal{L}$ is even implies that in our Taylor series $a_i = 0$ for every odd value of i. Thus, we have

$$\mathcal{L} = a_0 + a_2\eta^2 + a_4\eta^4 + \mathcal{O}(\eta^6), \tag{6.39}$$

with the additional constraint that $a_4 > 0$ (if it is not, then $\mathcal{L}$ above would be minimized by $\eta \to \infty$, and we want η to be finite and small... if a_4 is negative for a particular system, one needs to include higher order terms to stabilize the system.

What remains is to ask, for the Ising ferromagnet, about the temperature dependence of the coefficients a_i. First[11], a_0 is the value of $\mathcal{L}$ in the high temperature phase, and in general we expect this coefficient to vary smoothly (i.e., without divergence) through the critical temperature. It represents, in a sense, degrees of freedom which are not described by (and are certainly not coupled to) the order parameter; it may be important for some detailed calculations, but we will typically set it to zero.

---

[10] Including the field, the Ising model has $\mathbb{Z}_2$ symmetry associated with $\mathcal{H}(B, J, \{s_i\}) = \mathcal{H}(-B, J, -\{s_i\})$

[11] or zeroth

Next, we expand the fourth-order coefficient as

$$a_4 = a_4^{(0)} + \frac{T - T_c}{T_c} a_4^{(1)} + \cdots , \tag{6.40}$$

where the notation indicates the part of the coefficient associated with a particular order of this series expansion. As it turns out, the temperature dependence in a_4 does not control the overall behavior of the system near T_c, so we will just assume that a_4 is some positive constant.

That leaves only a_2, which we similarly expand:

$$a_2 = a_2^{(0)} + \frac{T - T_c}{T_c} a_2^{(1)} + \cdots .$$

Once again, though, we want to find $\eta = 0$ for $T > T_c$ and $\eta \neq 0$ for $T < T_c$. Comparing with Eq. 6.38, whose solution (for $a_1 = a_3 = 0$) is

$$\eta = 0 \quad \text{ or } \eta = \pm\sqrt{\frac{-a_2}{2a_4}}, \tag{6.41}$$

we see that we want to set $a_2^{(0)} = 0$ and $a_2^{(1)}$ to some positive constant to ensure that the order parameter is non-zero below T_c. Thus

$$a_2 = \frac{T - T_c}{T_c} a_2^{(1)} + \cdots , \tag{6.42}$$

and as with a_4, it is this lowest order term which dominates behavior near the critical point.

Finally, we now reintroduce the possibility of an external field. This breaks the even symmetry of the system, and we know from the Ising model Hamiltonian that it adds an $a_1 = -B$ term. Dropping some of the cumbersome notation and introducing a and b as phenomenological constants, we have our final expression for the Landau free energy density for the Ising model universality class in the absence of spatial variations:

$$\mathcal{L} = -B\eta + at\eta^2 + \frac{b}{2}\eta^4, \qquad \text{where } t = \frac{T - T_c}{T_c} \tag{6.43}$$

In principle we are allowed by symmetry to now also add an a_3 term; a calculation reveals that it is not a leading term near the critical point, so we've neglected it for now. In general, $\mathcal{L}$ is constructed by writing down *all possible scalar terms which are powers of the order parameter* (or the order parameter components, if the order parameter itself is more complicated than a scalar *that are consistent with the symmetry of the system.*

Critical exponents

For the above $\mathcal{L}$, we now compute a few critical exponents. The β exponent is the one that characterizes the divergence of m with t below T_c. From the above, we already know that

$$\eta = \sqrt{\frac{-at}{2b}}, \tag{6.44}$$

so we read off $\beta = 1/2$.

We next differentiate $\mathcal{L}$ with respect to η to find the magnetic equation of state:

$$B = 2at\eta + 2b\eta^3. \tag{6.45}$$

On the critical isotherm ($t = 0$) we immediately see $B \sim \eta^3$, and this is our $\delta = 3$ exponent.

The isothermal magnetic susceptibility is

$$\chi_T = \left.\frac{\partial \eta(B)}{\partial B}\right|_T = \frac{1}{2\left(at + 3b\eta(B)^2\right)}, \tag{6.46}$$

where $\eta(B)$ is the value of the order parameter in the presence of the external field, i.e., the solution to Eq. 6.45. The exponent γ characterizes the divergence of the susceptibility at zero field. For $t > 0$ we know $\eta = 0$ so $\chi_T = (2at)^{-1}$. For $t < 0$ we know $\eta = (-at/b)^{1/2}$ so $\chi_T = (-4at)^{-1}$. In either case, $\gamma = 1$.

Ignoring fluctuations... not variations

We very briefly note that Landau theory neglects the (important!) effects of fluctuations, but it does not only apply to homogeneous systems. It is straightforward to generalize the ideas above to the case where the order parameter can be a spatially varying one: $\eta = \eta(\boldsymbol{r})$, and if this were a class on statistical field theory we would spend a lot of time thinking about these cases (this would also let us write down critical exponents related to the divergence of *correlation lengths* in the system). When we do treat spatially inhomogeneous systems, we need to add the constraint that

5. $\mathcal{L}$ should be a *local* function, depending only on a finite number of spatial gradients of the order parameter.

As with the earlier construction, the gradient terms we write down must be consistent with the symmetry of the system; for the Ising model where we keep only even terms, the lowest order term is the square of the gradient, which would give the Landau free energy of

$$L = \int \mathrm{d}\boldsymbol{r} \left[\mathcal{L}\left(\eta(\boldsymbol{r})\right) + \zeta\left(\nabla\eta(\boldsymbol{r})\right)^2\right], \tag{6.47}$$

where ζ is some new positive constant and $\mathcal{L}$ is the homogenous $\mathcal{L}$ of the preceding notes. In general for this symmetry there are also terms like $(\nabla\eta)^4$ and $(\nabla^2\eta)^2$), etc., and here we've just written the lowest order gradient term[12].

Just a note about statistical field theories.

---

[12] "Why isn't there a term like $\eta\nabla^2\eta$, which is of the same order as the term we did use, it's perfectly isotropic, and it has the right $m \to -m$ symmetry?" I hear you asking. The above form is customary because we know the identity

$$\nabla \cdot (\eta\nabla\eta) = \eta\nabla^2\eta + (\nabla\eta)^2,$$

which implies

$$\int d\boldsymbol{r}\ \eta\nabla^2\eta = -\int d\boldsymbol{r}\ (\nabla\eta)^2 + \int d\boldsymbol{S}\cdot\eta\nabla\eta.$$

In the thermodynamic limit we neglect the surface term, so in general we pick either $\left(\nabla\eta(\boldsymbol{r})\right)^2$ or $\eta\nabla^2\eta$, but not both.

6.6 Correlations and fluctuations

Having introduced the idea that the order parameter might have spatial variations, we now introduce a dimensionless two-point correlation function:

$$G(\boldsymbol{r}-\boldsymbol{r}') = \frac{1}{\bar{\eta}^2}\left[\langle\eta(\boldsymbol{r})\eta(\boldsymbol{r}')\rangle - \bar{\eta}^2\right] = \frac{1}{\bar{\eta}^2}\left\langle\left(\eta(\boldsymbol{r})-\bar{\eta}\right)\left(\eta(\boldsymbol{r}')-\bar{\eta}\right)\right\rangle. \tag{6.48}$$

This quantity measures correlations between the fluctuations of the order parameter at different distances[13], and it is deeply linked to some of the thermodynamic properties of the system.

For instance, suppose as the order parameter we chose the spatially varying density field, $\rho(\boldsymbol{r})$. The total number of particles is $N = \int d^d\boldsymbol{r}\rho(\boldsymbol{r})$, and the integral over G is

$$\int d^d\boldsymbol{r}d^d\boldsymbol{r}'G(\boldsymbol{r}-\boldsymbol{r}') = \frac{1}{\rho^2}\int d^d\boldsymbol{r}d^d\boldsymbol{r}'\left[\langle\rho(\boldsymbol{r})\rho(\boldsymbol{r}')\rangle - \rho^2\right] = \frac{1}{\rho^2}\left[\left\langle N^2\right\rangle - \langle N\rangle^2\right]. \tag{6.49}$$

Translational symmetry gives us one of the integrals over G for free, and we recognize a connection between the number fluctuations and the isothermal compressibility of the system. Combing everything, then, gives us

$$\int d^d\boldsymbol{r}G(\boldsymbol{r}) = k_BT\kappa_T. \tag{6.50}$$

This is an example of a *fluctuation-susceptibility* relation, and is the equilibrium limit of the more general *fluctuation-dissipation relation.*

6.6.1 Correlation function for a specific model

To be definite, let's calculate the two-point correlation function for a system described by the Ising universality class Landau theory. The spatially varying order parameter will be $\eta(\boldsymbol{r})$, the external field will be $B(\boldsymbol{r})$, and we'll write the Landau free energy as

$$L = \int d^d\boldsymbol{r}\left[(at)\eta^2 + \frac{1}{2}b\eta^4 - B\eta + \frac{c}{2}\left(\nabla\eta\right)^2\right], \tag{6.51}$$

where $a,\ b,\ c$ are phenomenological parameters and t is the reduced distance to the critical point. If the system was uniform we would relate the typical value of the order parameter to the appropriate derivative of the free energy: $\langle\eta\rangle = -\frac{\partial L}{\partial B}$. In the presence of spatial variations, these partial derivatives get replaced by *functional derivatives.*

Functional differentiation

Suppose $F\left[\eta(\boldsymbol{r})\right]$ is a functional[14] of $\eta(\boldsymbol{r})$. The functional derivative of F with respect to the function η is defined as

$$\frac{\delta F}{\delta\eta(\boldsymbol{r}')} = \lim_{\epsilon\to 0}\frac{F\left[\eta(\boldsymbol{r})+\epsilon\delta(\boldsymbol{r}-\boldsymbol{r}')\right] - F\left[\eta(\boldsymbol{r})\right]}{\epsilon}. \tag{6.52}$$

---

[13] We have assumed translational symmetry in writing $G(\boldsymbol{r}-\boldsymbol{r}')$; more generally we would have $G(\boldsymbol{r},\boldsymbol{r}')$

[14] Just as a function is a map – accepting, say, a set of numbers and returning an output number – a functional is a map that accepts a function of a set of numbers and returns an output number

This generalizes the definition of the usual derivative, and the operation satisfies properties like

$$\begin{aligned} \frac{\delta}{\delta\eta(\boldsymbol{r})}\int \mathrm{d}^d\boldsymbol{r}'\eta(\boldsymbol{r}') &= 1 \\ \frac{\delta}{\delta\eta(\boldsymbol{r})}\eta(\boldsymbol{r}') &= \delta(\boldsymbol{r}-\boldsymbol{r}') \qquad (6.53) \\ \frac{\delta}{\delta\eta(\boldsymbol{r})}\int \mathrm{d}^d\boldsymbol{r}'\frac{1}{2}\left(\nabla\eta(\boldsymbol{r}')\right)^2 &= -\nabla^2\eta(\boldsymbol{r}). \end{aligned}$$

That last expression, which you probably can see that we're about to use, involves an integration by parts that neglects the surface term.

Linear response

With those definitions, we compute the expectation value of the order parameter via functional differentiation:

$$\langle\eta(\boldsymbol{r})\rangle = -\frac{\delta L}{\delta B(\boldsymbol{r})}. \qquad (6.54)$$

This implies that a small change in the external field would cause a small change in the Landau energy

$$\delta L = -\int \mathrm{d}^d\boldsymbol{r}'\,\langle\eta(\boldsymbol{r}')\rangle\,\delta B(\boldsymbol{r}'). \qquad (6.55)$$

Since the susceptibility is a measure of how the order parameter changes when the field changes, we have

$$\chi_T(\boldsymbol{r},\boldsymbol{r}') = -\frac{\delta}{\delta B(\boldsymbol{r}')}\left(\frac{\delta L}{\delta B(\boldsymbol{r})}\right). \qquad (6.56)$$

Treating L as if it is related to a partition function, $L = -k_BT\log Z$ (for the purposes of remembering which thermodynamic derivatives we want to take), we re-write this as

$$\begin{aligned} \chi_T(\boldsymbol{r},\boldsymbol{r}') &= k_BT\left[\frac{1}{Z}\frac{\delta^2 Z}{\delta B(\boldsymbol{r})\delta B(\boldsymbol{r}')} - \frac{1}{Z}\frac{\delta Z}{\delta B(\boldsymbol{r})}\frac{1}{Z}\frac{\delta Z}{\delta B(\boldsymbol{r}')}\right] \\ &= \beta\left[\langle\eta(\boldsymbol{r})\eta(\boldsymbol{r}')\rangle - \langle\eta(\boldsymbol{r})\rangle\langle\eta(\boldsymbol{r}')\rangle\right] \\ &= \beta G(\boldsymbol{r}-\boldsymbol{r}'), \qquad (6.57) \end{aligned}$$

again connecting response functions with correlation functions. Note that, using tildes to denote Fourier transforms, one connects isothermal susceptibility with the wave-vector dependent susceptibility $\tilde{\chi}(\boldsymbol{k}) = \beta\tilde{G}(\boldsymbol{k})$ as

$$\chi_T \equiv \lim_{\boldsymbol{k}\to 0}\tilde{\chi}(\boldsymbol{k}) = \beta\,\tilde{G}(\boldsymbol{k})\Big|_{\boldsymbol{k}=0} = \beta\int \mathrm{d}^d\boldsymbol{r}G(\boldsymbol{r}), \qquad (6.58)$$

as in Eq. 6.50.

The correlation function

Enough dithering, let's calculate the two-point correlation function from Eq. 6.51. To do this, we extremize the Landau free energy to find the spatially varying field that the system will adopt in equilibrium. I.e., we set $\frac{\delta L}{\delta\eta(\boldsymbol{r})} = 0$. This tells us that $\eta(\boldsymbol{r})$ must satisfy

$$2at\eta(\boldsymbol{r}) + 2b\eta^3(\boldsymbol{r}) - B(\boldsymbol{r}) - c\nabla^2\eta(\boldsymbol{r}) = 0. \tag{6.59}$$

We then (functionally) differentiate this expression with respect to the spatially varying field:

$$\begin{aligned} \frac{\delta}{\delta B(\boldsymbol{r}')}\left[2at\eta(\boldsymbol{r}) + 2b\eta^3(\boldsymbol{r}) - B(\boldsymbol{r}) - c\nabla^2\eta(\boldsymbol{r})\right] &= 0 \\ \Rightarrow \beta\left[-c\nabla^2 + 2at + 6b\eta^2(\boldsymbol{r})\right]G(\boldsymbol{r}-\boldsymbol{r}') &= \delta(\boldsymbol{r}-\boldsymbol{r}'). \end{aligned} \tag{6.60}$$

Well, would you look at that! The correlation function is a Green's function; seems like a retroactively good reason to have used G for it...

We can simplify the above expression by noting that for translationally invariant systems the order parameter η we want is just the equilibrium value from the homogenous Landau theory calculation we did earlier: $\eta = 0$ for $t > 0$ and $\eta = \pm\sqrt{-at/b}$ for $t < 0$. We now introduce the *correlation length*, ξ, noting that $G(\boldsymbol{r}-\boldsymbol{r}')$ satisfies

$$\left(-\nabla^2 + \xi^{-2}(t)\right)G(\boldsymbol{r}-\boldsymbol{r}') = \frac{k_BT}{c}\delta(\boldsymbol{r}-\boldsymbol{r}'), \tag{6.61}$$

where

$$\xi(t) = \left\{ \begin{array}{ll} \sqrt{\frac{c}{2at}} & \text{for } t > 0 \\ \sqrt{-\frac{c}{4at}} & \text{for } t < 0 \end{array} \right\} \sim |t|^{-1/2}. \tag{6.62}$$

Great, given that $G(\boldsymbol{r}-\boldsymbol{r}')$ satisfies such an equation, what does it look like? The Fourier transform of Eq. 6.61 gives

$$\tilde{G}(\boldsymbol{k}) = \frac{k_BT}{c}\frac{1}{k^2+\xi^{-2}}. \tag{6.63}$$

Evaluated at $\boldsymbol{k} = 0$, this gives us the isothermal susceptibility (a *measurable* quantity for the system) in terms of the microscopic parameter c and the correlation length:

$$\chi_T = \beta\tilde{G}(0) = \frac{\xi^2}{c},$$

with which we can write the two-point correlation function as

$$\tilde{G}(\boldsymbol{k}) = \frac{k_BT\chi_T(t)}{1+k^2\xi^2}. \tag{6.64}$$

In real space one can take the inverse Fourier transform (i.e. "look it up in a table"), or one can solve the real-space differential equation in polar coordinates. I'll spare you the details for now. Using the correlation length as our unit of length and defining $\rho = r/\xi$, the result is:

$$\frac{c}{k_BT\xi^{d-2}}G(\rho) = \left\{ \begin{array}{ll} e^{-r/\xi} & \text{for } d = 1 \\ \frac{K_{\frac{d-2}{2}}(r/\xi)}{(2\pi)^{d/2}(r/\xi)^{(d-2)/2}} & \text{for } d \geq 2 \end{array} \right., \tag{6.65}$$

where the K_n are modified spherical Bessel functions of the third kind. We really are most interested in the short- and long-range behavior of these functions, which are

$$\begin{aligned} K_n(x) &\sim \left(\tfrac{\pi}{2x}\right)^{1/2} e^{-x}, \quad \text{for } x \to \infty \\ K_n(x) &\sim \tfrac{\Gamma(n)}{2}\left(\tfrac{x}{2}\right)^{-n}, \quad \text{for } x \to 0 \\ K_0(x) &\sim -\log x \quad , \quad \text{for } x \to 0. \end{aligned} \tag{6.66}$$

Combining everything, very close to T_c the correlation length has diverged, so we just need the $r \ll \xi$ limit of the K_n. For $d = 2$ we use the special log form appropriate to K_0, and for $d > 2$ all of the powers of ξ conveniently cancel out, giving us

$$G(r) \sim \begin{cases} \log(\xi/r) & \text{for } d = 2 \\ r^{2-d} & \text{for } d > 2 \end{cases}. \tag{6.67}$$

Far from the critical point, though, ξ is small; taking the $r \gg \xi$ limit of the K_n gives (for $d \geq 2$)

$$G(r) \sim \frac{k_B T}{c} \frac{\exp(-r/\xi)}{r^{(d-1)/2}} \frac{1}{\xi^{(d-3)/2}}, \tag{6.68}$$

where I've dropped prefactors related to 2's and π's for simplicity.

Comments on the correlation length

From the preceding analysis, and from the definition of $G(\boldsymbol{r})$, we see that ξ is a measure of the spatial extent over which correlations extend. In mean field models of the Ising type we predicted $\xi \sim |t|^{-1/2}$; far from the critical temperature the correlation length will be on the same scale as something microscopic (distance between spins, or the range of the attractive part of the Lennard-Jones interaction, or...), and any fluctuations in the local order parameter away from the average value will quickly wash out.

As $T \to T_c$, though, $\xi \to \infty$, and correlations can extend over the entire system. Even though the actual interactions in the system are short-ranged, *long-range order* can propagate. Since the correlations are essentially macroscopic, the microscopic differences that distinguish one system from another likely are irrelevant at such macroscopic scales; this is an important clue for building a general understanding of critical phenomena.

6.7 Critical exponents

We pause to briefly summarize (and comment on) the most important critical exponents that characterize various systems. In all of these cases we are investigating the idea that, close to the critical point, some thermodynamic quantity has the limiting form of a power law. Letting $t = (T - T_c)/T_c$ and looking a quantity f, when we write

$$f(t) \sim t^{\lambda}$$

what we *mean* is

$$\lambda = \lim_{t \to 0} \frac{\log f(t)}{\log t}. \tag{6.69}$$

This is a particularly relevant point when remembering (a) there may be other, non-dominant power law behavior near the exponent; we are capturing just the leading order term, and (b) sometimes we quote an exponent as having the value zero. This can either mean that the thermodynamic quantity has a discontinuity *or* that it has a logarithmic divergence rather than a power-law one. This last possibility comes from comes from using the identity

$$\log t = \lim_{\lambda \to 0} \left[\frac{1 - e^{-\lambda \log t}}{\lambda} \right] = \lim_{\lambda \to 0} \left[\frac{1 - t^{-\lambda}}{\lambda} \right]. \tag{6.70}$$

Suppose we are considering a system with Landau free energy L, order parameter m, and ordering field B.

$\boldsymbol{\delta}$: All critical exponents *except* for this one are evaluated at $B = 0$. In the presence of the field, though, δ characterizes the relationship between the field and the order parameter:

$$m \sim B^{1/\delta}. \tag{6.71}$$

$\boldsymbol{\alpha}$: The divergence of the heat capacity is measured via α. In principle, the divergence could be different on the two sides of the transition, which is commonly denoted by writing α for the divergence above T_c and α' for the divergence below T_c.

$$C = -T \frac{\partial^2 \mathcal{L}}{\partial T^2} \sim |t|^{-\alpha}. \tag{6.72}$$

In mean field $\alpha = 0$ (in the form of a discontinuity); the Ising model has $\alpha = 0$ (in the log divergence form) for $d = 2$ and $\alpha \approx 0.11$ for $d = 3$.

$\boldsymbol{\beta}$: Below the critical temperature, where the order parameter is non-zero, it diverges like

$$m \sim |t|^{\beta}. \tag{6.73}$$

$\boldsymbol{\gamma}$: The divergence of the low-field susceptibility is measured by γ; it, too, can in principle have different values above and below the transition:

$$\chi = \left. \frac{\partial m}{\partial B} \right|_{B=0} \sim |t|^{-\gamma}. \tag{6.74}$$

$\boldsymbol{\nu}$: The last two critical exponents on this list are related to the behavior of the two point correlation function. We did a mean-field calculation that suggested the correlation length diverged near $t = 0$ like $\xi \sim |t|^{-1/2}$. There's no reason to expect for real systems that the mean field prediction here is correct, so we introduce ν:

$$\xi \sim |t|^{-\nu}. \tag{6.75}$$

Mean field theory for the Ising universality class has $\nu = 1/2$, $\nu = 1,\ 0.63$ in two and three dimensional Ising models, respectively.

$\boldsymbol{\eta}$: Finally, we want to characterize how the two-point correlation function behaves precisely at the critical point ($t = 0$). Our mean field theory (using the $r \ll \xi$ limit of the K_n) predicted that $G(r) \sim r^{-(d-2)}$, and again, we expect real experiments could differ from this. We introduce the exponent η to measure how wrong mean field is at $t = 0$:

$$G(r) \sim r^{-(d-2+\eta)}. \tag{6.76}$$

Mean field theory for the Ising universality class has $\eta = 0$, $\eta = 1/4,\ 0.032$ in the two and three dimensional Ising models, respectively. Experimentally these small exponents are hard to measure

6.7.1 Dimensional analysis and mean field theory

To emphasize just how surprising it should be that mean field theory gets the critical exponents wrong, and the surprising sense in which a diverging correlation length *does not* mean that the system has completely forgotten about microscopic length scales, let's write a dimensionless version of the Landau free energy of the Ising type in the absence of an external field:

$$\beta L = \int \mathrm{d}^d\boldsymbol{r} \frac{1}{2}(\nabla\phi)^2 + \frac{r_0}{2}\phi^2 + \frac{1}{4}u_0\phi^4. \tag{6.77}$$

In terms of our earlier expressions, this is just writing

$$\phi = \eta\sqrt{\beta c}, \quad r_0 = \frac{2at}{c}, \quad u_0 = \frac{2b}{\beta c^2}. \tag{6.78}$$

Let's think about the correlation function in terms of basic dimensional analysis. Using the bracket notation to denote the dimension of various quantities, we have $[\beta L] = 1$ (i.e., is dimensionless), so each separate term in the integrand must be dimensionless, too. Thus:

$$\left[\int \mathrm{d}^d\boldsymbol{r}(\nabla\phi)^2\right] = 1 \Rightarrow (x^d)(x^{-2})\,[\phi]^2 = 1, \tag{6.79}$$

where x denotes units of length. From this, we immediately get

$$[\phi] = x^{1-d/2}, \quad [r_0] = x^{-2}, \quad [u_0] = x^{d-4}. \tag{6.80}$$

So, what does dimensional analysis say about the correlation length? Clearly $[\xi] = x$, but from the above equation *the only independent quantity* with units of length is r_0, with dimensions $[r_0] = x^{-2}$. So, apparently, we're done: by dimensional analysis we get that

$$\xi \sim r_0^{-1/2} \sim |t|^{-1/2},$$

where in the last line we remember that $r_0 \propto t$. In some sense, then, the deep mystery is *how could this have been wrong?*

The answer is that *we ignored one other source of an independent length scale* in our dimensional analysis! In particular, there is a *microscopically* small length scale embedded in the problem – say, the spacing between lattice sites in our Ising model – an apparently if mean field is wrong we need to include this length scale.

How does this solve our problem? Let's call this microscopic length λ, with $[\lambda] = x$, of course. By dimensional analysis we can conclude that

$$\xi = r_0^{-1/2} f(r_0 \lambda^2), \tag{6.81}$$

where f is some function we know nothing about (yet). We've written it this way, though, because $r_0 \propto t$ and λ is independent of t, so near the critical point we are interested in $\lim_{z\to 0} f(z)$. If, *for whatever reason*, it so happens that

$$f(z) \sim z^{\theta} \tag{6.82}$$

as $z \to 0$, then as the critical temperature is approached we get

$$\xi \sim t^{-1/2+\theta} \lambda^{2\theta}. \tag{6.83}$$

The exponent θ, characterizing the difference between an observed divergence of the correlation length and the prediction from Landau theory, is called the anomalous dimension.

Notice something remarkable that's happened here: we have been emphasizing that near a critical point the diverging correlation length sweeps away any microscopic details, and we are used to assuming that when trying to explain phenomena at one scale we can disregard phenomena at much shorter scales[15]. Apparently near the critical point this idea is not quite right: the *very existence* of a microscopic length scale allows for an anomalous dimension / departure from mean field theory. In general, although not a priori guaranteed, the *value* of the anomalous dimension does not depend on the microscopic length scale itself. Thus, near a critical point we expect that the *scaling* of various quantities, the way they diverge near the critical point, to be universal, but there is no reason to expect the prefactors of the scaling relations to share that universal character.

6.8 Scaling hypothesis

We just described a set of critical exponents that characterize the behavior of systems near their critical point; let's stare at a table[16] for a moment:

	Ising (MF)	Ising ($d = 2$)	Ising ($d = 3$)	Percolation ($d = 3$)	4-State Potts ($d = 2$)	Directed percolation (MF)	XY model ($d = 3$)
α	0	0	0.11	-0.625	2/3	-1	-0.015
β	1/2	1/8	0.327	0.418	1/12	1	0.349
δ	3	15	4.79	5.3	15	2	4.78
γ	1	7/4	1.24	1.793	7/6	1	1.32
ν	1/2	1	0.63	0.88	2/3	1/2	0.67
η	0	1/4	0.036	0.046	1/4	0	0.038

---

[15] "Don't model bulldozers with quarks" – Goldenfeld and Kadanoff, Science 284, 87 (1999)

[16] Anything that is not an integer should be assumed to be approximate. Sources are wikipedia (where there are tables of critical exponents for many more models), plus *https://arxiv.org/pdf/cond-mat/9701018.pdf* for the Potts model

In the early 60s, as mean field results were known and as some finite-dimensional estimates were obtained by numerical estimates, people started noticing that these exponents did not seem to be independent. For instance, the exponent values seem to satisfy

$$\alpha + 2\beta + \gamma = 2 \tag{6.84}$$
$$\gamma - \beta(\delta - 1) = 0 \tag{6.85}$$
$$\gamma - \nu(2 - \eta) = 0, \tag{6.86}$$

which are often called the Rushbrooke, Widom, and Fisher identities, respectively. Since then, there have been various proofs that thermodynamics requires (via, say, the convexity of the free energy, or the relative sizes of C_P and C_V certain *inequalities* to be satisfied, for instance

$$\alpha + 2\beta + \gamma \geq 2,$$

but that in the actual data these thermodynamic inequalities are *saturated.*

6.8.1 The static scaling hypothesis

The static scaling hypothesis[17] is an attempt to encode multiple features of the behavior of a system near the critical point in a single expression; as we will see, assuming this hypothesis allows one to strengthen the thermodynamic inequalities mentioned above for the relationships between exponents into *equalities* – which is good, because the equalities are satisfied.

Let's focus on a magnetic system, where just to keep you on your toes I'll write m for the magnetization and h for the external field (in units of k_BT); t will still stand for the reduced distance to the critical point. We want to simultaneously encode two things we have already seen:

$$m(t=0,h) = \pm C_1 |h|^{1/\delta} \quad \text{and} \quad m(t, h=0) = \begin{cases} 0 & t > 0 \\ \pm C_2 |t|^{\beta} & t < 0 \end{cases}. \tag{6.87}$$

One of Widom's many insights was that, staying in a regime where $|t| \ll 1$ and $|h| \ll 1$, both of these results can be expressed as

$$m(t,h) = \begin{cases} t^{\beta} F^{+}(h/t^{\Delta}) & t > 0 \\ (-t)^{\beta} F^{-}(h/(-t)^{\Delta}) & t < 0 \end{cases}, \tag{6.88}$$

where we assume that β and the *gap exponent* Δ are universal, as are the *scaling functions* above and below the critical temperature, F^+ and F^-. At first glance we don't know very much about these scaling functions, but the requirement that Eq. 6.88 reproduce the results in Eq. 6.87 will let us both put some constraints on F^+ and F^- and also derive relationships between the critical exponents.

First, we'll show that the gap exponent is actually not a new critical exponent at all! Consider the susceptibility in the low-field limit, $\chi_T \propto \frac{\partial m}{\partial h}\big|_{h=0}$. Differentiating Eq. 6.88 with respect to h gives

$$\chi_T \propto |t|^{\beta} \left.\frac{\partial F^{\pm}(h/|t|^{\Delta})}{\partial h}\right|_{h=0} \propto \frac{|t|^{\beta}}{|t|^{\Delta}} \left.\frac{\partial F^{\pm}(x)}{\partial x}\right|_{x=0} \sim |t|^{\beta-\delta} F'^{\pm}(0). \tag{6.89}$$

---

[17] B. Widom, (1963)

So, as long as the derivative of these scaling functions doesn't either diverge or vanish as the argument goes to zero, we can connect the gap exponent to a critical exponent we've already encountered:

$$\beta - \Delta = -\gamma. \tag{6.90}$$

Next, let's see what happens by requiring that Eq. 6.88 reproduces the results in Eq. 6.87. First, we take the *small field* limit of Eq. 6.88 and match it to the zero field result

$$m(t,h) = \begin{cases} t^{\beta}F^{+}(0) & t > 0 \\ (-t)^{\beta}F^{-}(0) & t < 0 \end{cases} = \begin{cases} 0 & t > 0 \\ \pm C_2(-t)^{\beta} & t < 0 \end{cases} \tag{6.91}$$

This already gives us some constraints on the scaling functions, namely

$$F^{+}(0) = 0 \quad \text{and} \quad F^{-}(0) = (\text{some finite constant}). \tag{6.92}$$

We also want to reproduce the result for the magnetization along the isotherm in the presence of a field ($m \sim h^{1/\delta}$), so we look at the limit $t \to 0$ while keeping h small and finite. This amounts to looking at the scaling functions in the limit that their argument diverges, but we also know that in this limit m is well-behaved. How do we reconcile the situation? We assume that both scaling functions adopt a power-law form for large argument:

$$F^{\pm}(x) \sim x^{\lambda}. \tag{6.93}$$

Making this assumption, the magnetization as the critical isotherm is approached is

$$m(t \to 0, h) \sim |t|^{\beta}\left(\frac{h}{|t|^{\Delta}}\right)^{\lambda} \sim |t|^{\beta - \lambda\Delta}h^{\lambda}. \tag{6.94}$$

On the critical isotherm, though, we need the t-dependence of this expression to cancel out (otherwise we would get the incorrect result that either $m = 0$ or $m = \infty$, depending on the sign of the exponent of $|t|$). We therefor have the simultaneous requirements

$$\beta = \lambda\Delta \quad \text{and} \quad \lambda = 1/\delta. \tag{6.95}$$

This both tells us that the scaling functions behave like

$$F^{\pm}(x \to \infty) \sim x^{1/\delta} \tag{6.96}$$

and

$$\Delta = \beta/\lambda = \beta\delta. \tag{6.97}$$

Combining the above result with Eq. 6.90 gives the Widom identity:

$$\beta\delta = \beta + \gamma. \tag{6.98}$$

The static scaling hypothesis, if true, suggests two powerful things. First, by derivations like the above, it places constraints on the critical exponent that should be measured in experiments or in exact theories (and, thus, also provides targets for approximate theories,

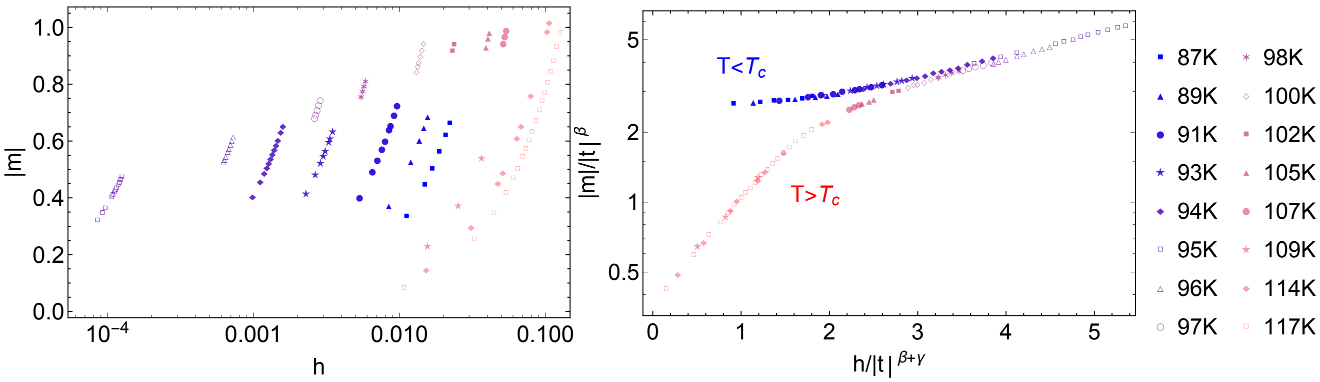


Figure 6.3: **Scaling plot for magnetization vs field near the critical point** for the "anhydrous compound with the formula $\left\{\left[\mathrm{Mn}_2^{\mathrm{II}}(\mathrm{pydz})\right]\left[\mathrm{Nb}^{\mathrm{IV}}\left(\mathrm{CN}\right)_8\right]\right\}_n$," taken at different temperatures, with estimates of $T_c \approx 95.3$, $\beta \approx 0.41$, and $\gamma \approx 1.39$. Data roughly extracted from Pełka *et al.*. Units on the plot have been suppressed.

if you're into that kind of thing). Second, the assumption that the scaling functions are also universal *suggests an important way of plotting data*!

For instance: suppose you were to measure the magnetization of a system as a function of temperature and external field and then make a plot of m vs h, you'd get different curves for each value of the field t. The static scaling hypothesis says that *even if you have no idea* what the functions $F^{\pm}$ are, if you plot $m/|t|^{\beta}$ vs $|h|/|t|^{\beta+\gamma}$ then you should observe *data collapse*, with all of your data falling onto two branches. To observe this collapse you would need to know the right values for T_c, β, and γ (which you usually don't know ahead of time). An example of this kind of data collapse is shown in Fig. 6.3, where I pulled data from an arXiv paper studying magnetization near the critical point of... let's just say "some material." Manipulating data in order to observe scaling collapse is a way of estimating both the location of the critical point and the values of the critical exponents, although it must be done with care.

Before closing, we briefly mention that in addition to scaling laws relating the critical exponents, there are also *hyperscaling laws* which involve both the critical exponents *and* the dimension of space. These can be derived from, e.g., assuming that the correlation function has a scaling form:

$$G(\boldsymbol{r},t,h) = \frac{1}{r^{d-2+\eta}} F_G^{\pm}(r|t|^{\nu}, \frac{h}{|t|^{\Delta}}),$$

where we've added the subscript to F_G since there is no reason that the scaling function for the two-point correlations should be the same as the scaling function for the magnetization. This scaling hypothesis leads to laws like the Josephson relation,

$$2-\alpha = d\nu.$$

Hyperscaling laws are on a different footing than the scaling laws – they must be, and clearly something must be afoot in mean field models, where the critical exponents stay the same for all $d > d_c$. We also mention that just as there are *static scaling hypotheses* there are *dynamic scaling hypotheses*. These relate the *relaxation time* – the way that fluctuations from the typical order parameter decay away – to the correlation length, and bring with them both new critical exponents and new relationships between those exponents.

Finally, you may be wondering where these scaling laws in fact derive from. The main physical idea is the diverging correlation length, ξ, is responsible for all of the singular behavior near the critical point. We've already indicated that this vague statement cannot be completely true, but the idea is the following: consider just the part of the free energy density which has the singularity, and by dimensional analysis write it as

$$\frac{\mathcal{L}_s}{k_B T} \sim \xi^{-d}\left(c_0 + c_1\left(\frac{\lambda_1}{\xi}\right)^{\pi_1} + c_2\left(\frac{\lambda_2}{\xi}\right)^{\pi_2} + \cdots\right). \tag{6.99}$$

Here the λ_i are a list of any microscopic length scales in the problem, the c_i are coefficients that depend at most weakly on temperature, and the $\pi_i > 0$.

With those *assumptions*, in the $t \to 0$ limit the leading behavior is dominated by

$$\mathcal{L}_s \sim \xi^{-d} \sim |t|^{d\nu},$$

and this immediately leads to the Josephson relation:

$$t^{-\alpha} \sim C = -T\frac{\partial^2 \mathcal{L}_s}{\partial T^2} \sim |t|^{d\nu - 2} \Rightarrow d\nu = 2 - \alpha. \tag{6.100}$$

This, perhaps, feels a bit unsatisfying: we do not really know a priori that, given a quantity A with $[A] = x^y$ we should always take the dimensionless quantity $A\xi^{-y}$ in our scaling theory – why might it not be the case that we should take $A\xi^{z-y}\lambda^{-z}$, mixing and matching the diverging length scale with one of the microscopic ones? Ultimately, the answers to all of these questions rely on the theory of the renormalization group: RG lets us see where scaling hypotheses come from, it gives criteria for which measurables acquire anomalous dimensions, allows us to *calculate* those anomalous dimensions, and (with a little bit of help) lets us derive the particular forms of scaling functions.