Polymers and Complex Fluids

S. T. Milner

October 3, 2014

Contents

Chapter 1

Fundamentals

1.1 Themes of this course

[Reading: for general background (and an introduction to a textbook we will be consulting repeatedly throughout the course, Jones Ch. 1, Ch. 2 through section 2.2.]

Complex fluids

This course is concerned with a class of materials called "complex fluids", which most often contain polymers as an essential constituent. Complex fluids are so named both because they contain more ingredients than a simple pure small-mlecule fluid such as water or decane, and because as a consequence of their ingredients they often exhibit complex behavior in one or more physical properties.

Complex fluids abound in modern technological life as well as in nature. Common examples of complex fluids include emulsions (milk, makeup, ice cream), foams (shaving cream, firefighting foam, foam rubber), gels (contact lenses, Jello), colloidal suspensions (motor oil, liquid medicines, ink). Using properly designed complex fluids, we can often achieve some desirable set of properties, including flow behavior, mechanical properties, transport properties, or other functionality, that could not be achieved with a single pure material.

Self-assembly

A key strategy in the construction of complex fluids is self-assembly. In self-assembly, molecules or small particles are designed so that different portions attract or repel., in such a way that the molecules or small particles arrange themselves to form larger structures. Because the interactions between the constituents can be tuned to be strong or weak, a wide range of "strength" or "softness" of materials can be accessed. The materials found in living tissue are self-assembled soft materials in this sense. The length scale of self-assembled structure can be considerably larger than that of the constituents, which allows us to build in nanometer- and micrometer-sized structures with Angstrom-sized components. This can be much more practical than trying to create large structures by direct chemical routes (organic synthetic chemistry) or physical means (lithography).

If the interactions between constituents are taken to be sufficiently weak, then conditions of temperature, pH, and salt become relevant; we can design materials with properties that change significantly in the range of temperatures and ionic strengths relevant to living systems.

Polymers

This course does not assume any prior familiarity with polymers. So what are polymers? Perhaps the simplest definition is: polymers are big, primarily linear molecules, which can consist of one or more repeating units (monomers or mers). The simplest case is a linear homopolymer, in which one monomer is covalently bonded together in a linear molecule. The most common such polymer in the world today is polyethylene, in which the mer is CH_2. (Sometimes called – by chemists – polymethylene, in recognition of the repeat unit.)

Polymers can deviate from the simple linear homopolymer in two important ways. First, polymers can consist of more than one monomer, and the different monomers can be arranged in various ways along the polymer. Even the simplest case of two comonomers affords great variability in the sequence and resulting properties. Two extreme cases are random copolymers (in which the monomers appear randomly along the chain), and block copolymers, in which the two types of monomers appear in two or more blocks. We shall see that block copolymers afford a simple and potent route to self-assembled structures.

As the number of comonomers increases, the richness of behavior grows astoundingly. Proteins as found in nature are linear polymers with twenty possible comonomers (the twenty natural amino acids), and all of the structure and functionality of proteins ultimately arises from the particular comonomer (amino acid) sequence.

Two essential aspects of polymers are that they are "mostly linear", and that they are "large" molecules (typically consist of many monomers). The linearity of polymers is important in two respects: first, it makes polymer synthesis practical — the molecule is constructed serially (sequential monomer addition) or by joining together shorter sequences of monomers (condensation polymerization). One relies on self-assembly rather than unimaginably complex synthesis create three-dimensional mesoscale (bigger than small-molecule scale) structures.

Second, the linearity of polymers, together with their largeness (high molecular weight), implies that the molecules are big enough to be flexible. That is, polymers in the liquid state typically will adopt many configurations, not just one low-energy configuration. Thermal fluctuations and entropy are an essential aspect of polymer physical properties. Therefore, we will need the methods of statistical mechanics, which are designed to work out how the macroscopic properties of matter arise from a large collection of fluctuating molecules.

A great variety of monomers can be polymerized by various means, which would be the subject of a course in polymer chemistry. The selection of what monomers to use in a polymer is made based on a combination of factors, including the desired final properties, the practicality of the polymerization chemistry, and the cost and availability of the monomer. These are important considerations, which however we shall only touch on briefly at various points in this course.

Instead, in this course we shall focus on what one might call the "physical" rather than "chemical" control of polymer and material properties. That is, instead of selecting a new monomer every time we want to achieve a different result, we will examine how much variability there is in macroscopic properties that comes from changing things like how long a polymer is, how its monomers are arranged, how big a particle is, what the concentrations of various components are, and so forth. These sorts of variations have the great advantage that they are relatively easy to make for someone who is not a synthetic chemist, whereas to design a new monomer for polymerization is a big undertaking.

Random walks

A short hydrocarbon chain (butane, say) in the vapor phase or in solution will typically be found vibrating about its lowest-energy "all-trans" configuration. The interior carbon-carbon bond in butane has one trans and two gauche configurations, separated by a "dihedral rotation" of 120 degrees. From the viewpoint of the average direction of the all-trans zig-zag, a single gauche bond causes a deviation of 60 degrees. The trans-gauche energy difference is about 3kJ/mol or 1.75kT at room temperature, which means that typically 75 percent of the bonds will be found in the trans state, with 12.5 percent each in the gauche + and gauche - states (we will learn, or be reminded, how to make such estimates in the next few lectures).

Now consider a long polyethylene chain in solution (in hot xylene, say). Locally, the PE chain looks just like butane. All the C-C bonds have a dihedral rotation, and trans and gauche states. What will a typical configuration of the polymer look like, if 25 percent of the bonds are gauche?

The answer is: the polymer will have many configurations available, and a typical one will look like a random walk. We can avoid the local details by "coarse-graining" the polymer, by noting the position of only every nth monomer. Then enough bonds intervene that the direction of the chain from one step to the next becomes random. The displacement from one step to the next will be random as well, with some typical range of values — hence, the path of the chain will be a random walk, with a varying step length and random direction.

In fact, random walks appear in many contexts in complex fluids, so that it is useful to develop ways of describing their randomness. Examples of random walks include:

- $R(s)$, the configuration of a flexible polymer

- dislocations and other line defects in solids

- trajectories (paths in time) of diffusing molecules and colloidal particles

- trajectories of diffusing polymers in an entangled solution

- $T(s)$, the configuration of the tangent of a semiflexible polymer (DNA, e.g.)

- trajectories of light multiply scattered by an opaque solution (milk, e.g.)

We will encounter all of these, and more, throughout this course.

To summarize: the themes of this course are:

- polymers key to designing complex fluids

- statistical mechanics key to understanding polymers

- random walks (and ideal gases) key to statistical mechanics

In addition, throughout the course we will focus on the ability to estimate the rough size (order of magnitude) of various physical quantities. This is useful for design purposes, as well as for simplifying a complicated situation — we need to know what we can neglect, and what we need to focus on.

1.2   Mathematics of random walks

[Reading: Jones Appendix A for a quick summary, Reif Ch. 1 for more details]

In the last lecture, we remarked that random walks appear commonly, in many contexts in complex fluids. In each case, the details — what thing is doing the random walking, and precisely how — are different. Rather than focus on the details of a particular physical example first (such as the configurations of polyethylene), we begin with a "toy model" — simple enough to analyze, complicated enough to be interesting. We will then build on our understanding of the toy model as we consider more physical examples of random walks.

Tossing a coin

The simplest example of a random walk is this: toss a coin. If heads, take one step forward; if tails, take one step back. Repeat. Suppose we ask: where will a person playing this game be after n coin tosses? For a given walker, the answer cannot be known in advance, since it depends on the outcome of the coin tosses; all we can say is, the walker will be $n_+ - n_-$ steps forward, where n_+ and n_- are the number of heads and tails tossed.

But suppose we assign a large group of people to play this game; we start the game with everyone arranged side by side in a straight line. We can ask, what will be the number of people s steps from the starting point? (Equivalently, we could assign one person to repeat the game many times, and ask how often he ends up s steps ahead.) This question, it turns out, is better posed; although we cannot say in advance where a given walker will end up, we can predict roughly how many walkers will end up s steps ahead.

That is, we can predict the probability $P(s, n)$ that a walker will end up s steps ahead after n tosses. The meaning of $P(s, n)$ is: if we repeat the game N times where N is very large, $P(s, n)$ is the fraction of walks that end up at s.

Now we set about to compute $P(s, n)$. There are 2^n possible sequences of heads and tails, all of which are equally likely. How many sequences are there that lead to s? To wind up at s, a walker must have $n_+ - n_- = s$. Since $n_+ + n_- = n$, this means we must have exactly $n_+ = (n + s)/2$ heads and $n_- = (n - s)/2$ tails.

Combinatorics

Now we count how many sequences there are with the specified number of heads and tails. We must choose among the n tosses which will be the n_+ heads. Imagine we have n boxes in a row, and want to place $n_+ < n$ tokens in the boxes, with either 0 or 1 token in each box. How many ways are there to do this? We temporarily number the tokens 1 through n_+. We choose among all the empty boxes, where to put the first token; there are n possible choices. For the next token, there are $n - 1$ choices (one box is filled), and so on down to the last token, for which there are $n - (n_+ - 1)$ choices. Each of these choices is independent; the total number of arrangements is the product, $n \times (n - 1) \times \ldots n - (n_+ - 1)$, which is $n!/(n - n_+)!$.

But the numbers on the tokens were temporary; the tokens were actually identical. From the above process, for a given selection of boxes filled, the numbers on the tokens will appear once in every possible order. The number of ways to order the numbered tokens is $n_+!$ (again, think of selecting which numbered token will be first, for which there are n_+ choices, which second, and so on down to the last; the choices are independent, and the product is $n_+!$.)

Every distinct arrangement of the n_+ unnumbered tokens appears $n_+!$ times, one each for every ordering of the temporary numbers on the tokens. Hence the number of distinct ways of choosing

n_+ places to put a token among n boxes is $n!/(n - n_+)!n_+!$. This quantity is written $\binom{n}{n_+}$, read "n choose n_+", and is called the binomial coefficient.

The probability of any particular sequence of heads and tails being chosen is the same, namely $1/2^n$ (one sequence out of s^n possible). So the probability of a walk that winds up at s is

$$P(s, n) = 2^{-n} \frac{n!}{n_+!(n - n_+)!} = 2^{-n} \frac{n!}{(n + s)/2!(n - s)/2!} \equiv 2^{-n} \binom{n}{(n + s)/2} \tag{1.2.1}$$

Note that $P(s, n)$ is an even function of s, i.e., it is symmetric under interchange $s \to -s$; thus the probability of ending up s steps ahead equals the probability of ending up s steps behind.

Stirling formula

In dealing with the random motions of molecules rather than people, the number of steps taken can become very large indeed. Our result for the probability $P(s, n)$ to arrive at s after n steps is exact, but unwieldy. It is defined only for an integer number of steps, and we cannot easily see from the exact expression what the shape of the distribution will be. If we evaluate $P(s, n)$ for large n, we find that a bell-shaped histogram results. The exact expression looks increasingly like a smooth function, as n becomes large. It is therefore useful to develop approximations for $P(s, n)$ that hold when the number of steps is large, which represent $P(s, n)$ as a continuous function.

Consider how $n!$ changes when we increase n by one: we acquire one more factor, of $(n + 1)$. We would like to construct an approximation to the derivative $dn!/dn$; simpler is to take the log, so that

$$\frac{d \log n!}{dn} \approx \frac{\Delta \log n!}{\Delta n} = \log(n + 1) \approx \log n \tag{1.2.2}$$

This we can integrate, to obtain

$$\log n! \approx n \log n - n \tag{1.2.3}$$

More precise methods yield Stirling's formula:

$$n! = \sqrt{2\pi n} \left(\frac{n}{e}\right)^n (1 + O(1/n)) \tag{1.2.4}$$

Our simple approximation corresponds only to the factor $(n/e)^n$.

We use the simplified Stirling formula to approximate $P(s, n)$: after a bit of algebra, we have

$$\log P(s, n) \approx -(n/2) \left[(1 + s/n) \log(1 + s/n) + (1 - s/n) \log(1 - s/n)\right] \tag{1.2.5}$$

We expect $P(s, n)$ to have a maximum, which corresponds to the most likely outcome. For an unbiased coin, on symmetry grounds the most likely outcome must be $s = 0$. We expand about this point (s/n turns out to be a convenient expansion variable), to find the "spread" (variance) in the final locations.

Since $P(s, n)$ is even in s, its first derivative at $s = 0$ vanishes; we have

$$\log P(s, n) = -n/2 \left[(s/n)^2 + O(s/n)^4\right] \tag{1.2.6}$$

Thus we have an approximately Gaussian probability distribution for s

$$P(s, n) \propto e^{-s^2/2n} \tag{1.2.7}$$

We fix the normalization (which our Stirling approximation has been sloppy about) by requiring the integral under $P(s, n)$ be unity: this gives

$$P(s, n) = \frac{e^{-s^2/2n}}{(2\pi n)^{1/2}} \tag{1.2.8}$$

(In setting the normalization, we used the Gaussian integral result from Appendix B,

$$\int_{-\infty}^{\infty} dx\, e^{-\alpha x^2} = (\pi/\alpha)^{1/2}$$

A useful measure of the spread of possible values of s is the average $\langle s^2 \rangle$. This turns out to be

$$\langle s^2 \rangle = n \tag{1.2.9}$$

That is, there are "square root of n fluctuations" in the position of the walker after n steps.

(To compute this, note that we can get the integral of x^2 times a Gaussian by differentiating with respect to α (see Appendix B):

$$\begin{aligned}
\int dx\, x^2 e^{-\alpha x^2} &= -\partial_\alpha \int dx\, e^{-\alpha x^2} \\
&= -\partial_\alpha (\pi/\alpha)^{1/2} = \frac{(\pi/\alpha)^{1/2}}{2\alpha}
\end{aligned}$$

Then the average we seek is

$$\sigma^2 = \langle x^2 \rangle = \frac{\int dx\, x^2 e^{-\alpha x^2}}{\int dx\, e^{-\alpha x^2}} = \frac{1}{2\alpha}$$

In the present case, α is $1/2n$, whereupon $\langle s^2 \rangle = n$.)

We can invert this relation, to write $\alpha = 1/(2\sigma^2)$. So whenever we encounter a Gaussian distribution $e^{-\alpha(x-\bar{x})^2}$, we can identify the variance $\sigma^2 = \langle \Delta x^2 \rangle$ by expressing the coefficient α as $1/(2\sigma^2)$. This puts the Gaussian in a "standard form",

$$P(x) \propto e^{-x^2/(2\sigma^2)} \tag{1.2.10}$$

Difference equations and continuum description

In the previous section, we have approximated the discrete binomial distribution for a random walk by a continuous Gaussian distribution. Sometimes it is useful to describe a random walk process in continuum language to start with.

One route to get there is to start with a "difference equation" for the probability distribution $P(s,n)$. We focus on how $P(s,n)$ evolves as the number of steps increases. In the present case, to arrive at s after n steps, a walker must have come from either $s-1$ or $s+1$ on the previous step. And, for all walkers at $s-1$ on the previous step, half of them go to s (the half that toss heads); similarly for those at $s+1$ on the previous step, half go to s (the half that toss tails). That is, the number of walkers at s after $n+1$ steps must satisfy

$$P(s, n+1) = (1/2)\left(P(s-1, n) + P(s+1, n)\right) \tag{1.2.11}$$

We construct a finite difference for the n-dependence of $P(s,n)$:

$$P(s, n+1) - P(s, n) = (1/2)\left(P(s-1, n) - 2P(s, n) + P(s+1, n)\right) \tag{1.2.12}$$

Now Taylor expand everywhere about the point s, n, and keep only the lowest order terms; the result is

$$\partial_n P(s, n) = (1/2)\partial_{ss} P(s, n) \tag{1.2.13}$$

We have derived a diffusion equation for the probability distribution $P(s, n)$. The equation describes how $P(s, n)$ evolves in "time", as the number of steps n increases. The equation is deterministic: even though each individual walker moves randomly, the distribution evolves in a predictable fashion. As more steps are taken (as time passes), the "cloud" of random walkers broadens.

By direct substitution, we can verify that the continuum expression for $P(s, n)$ is a solution to the above diffusion equation. Note that for "early times" (small n), the width of the Gaussian becomes very narrow. The initial condition for the Gaussian solution is a "delta function", i.e., a very sharply defined function with a fixed (unit) normalization integral. This corresponds to the condition that the random walker at "time" zero (before any steps have been taken) starts out at location $s = 0$.

Biased random walk

Now suppose the coin toss is biased — the coin is weighted, or some other unspecified chicanery transpires, such that the chance of a head being tossed is $p > 1/2$. How will $P(s, n)$ change?

The number of walks that arrive at s is unchanged; it still corresponds to the number of walks with n_+ heads and n_- tails such that $n_+ - n_- = s$. However, the probability of any particular sequence of heads and tails changes. The probability of any one sequence with n_+ heads and n_- tails is $p^{n_+}(1 - p)^{n_-}$; hence

$$P(s, n) = p_+^{n_+} p_-^{n_-} \frac{n!}{n_+! n_-!} \equiv p^{(n+s)/2}(1 - p)^{(n-s)/2} \binom{n}{(n + s)/2} \tag{1.2.14}$$

in which we observe the expected symmetry under interchange of n_+ and n_- (with p_+ interchanged with p_-).

The probability distribution $P(s, n)$ is no longer an even function of s; indeed, we expect that the walkers will tend to drift forward at a steady rate of $p_+ - p_-$ or $2p - 1$ steps per coin toss.

Applying the simplified Stirling formula to the above biased random walk result, we find

$$\log P(s, n) \approx -(n/2) \left[(1 + s/n) \log \frac{(1 + s/n)}{2p} + (1 - s/n) \log \frac{(1 - s/n)}{2(1 - p)} \right] \tag{1.2.15}$$

To verify that the most-probable value of s (maximum of $\log P$) is $s = (2p-1)n$, we set $\partial_s \log P(s, n)$ to zero, to find

$$0 = \log \frac{(1 + s/n)}{2p} - \log \frac{(1 + s/n)}{2(1 - p)} \tag{1.2.16}$$

which after a bit of algebra gives the expected result.

Now we expand about the maximum to second order as before. Only linear terms in s resulted from p being different from $1/2$, so the second derivative expression is unchanged; as a result,

$$\log P(s, n) \approx \text{const} + \frac{n}{8p(1 - p)} (s - \bar{s})^2 \tag{1.2.17}$$

in which $\bar{s} = (2p - 1)n$.

Hence (after imposing normalization)

$$P(s, n) = \frac{e^{-(s-\bar{s})^2/(8p(1-p)n)}}{(8\pi p(1 - p)n)^{1/2}} \tag{1.2.18}$$

The probability distribution for this biased random walk is a displaced Gaussian, with a mean value $\bar{s}$ that drifts with a constant velocity $v = (2p - 1)$.

The variance $\langle (s - \bar{s})^2 \rangle$ is equal to $8p(1 - p)n$, which for $p = 1/2$ gives the previous unbiased result. For p different from $1/2$, we see that the variance is smaller, approaching zero when p either approaches 1 (all heads) or 0 (all tails). As the coin becomes more biased, the crowd of walkers spreads more slowly. In the limit that the coin is completely predictable ($p = 0$ or $p = 1$), the crowd does not spread at all, but simply moves along together.

The corresponding evolution equation for a biased walk begins with the discrete evolution equation,

$$P(s, n + 1) = pP(s - 1, n) + (1 - p)P(s + 1, n) \tag{1.2.19}$$

which simply says that to arrive at s in $n + 1$ steps, walkers come from $s - 1$ on the previous step with probability p, or from $s + 1$ with probability $1 - p$. Rearranging this leads to a finite difference equation

$$
\begin{aligned}
P(s, n + 1) - P(s, n) &= (1/2)\left[P(s - 1, n) - 2P(s, n) + P(s + 1, n)\right] \\
&\quad + (p - 1/2)\left[P(s - 1, n) - P(s + 1, n)\right]
\end{aligned}
\tag{1.2.20}
$$

Expanding this difference equation to arrive at a diffusion-type equation is slightly tricky; it turns out to be convenient to view the problem in the coordinate system that moves along with the mean drift of the walkers, i.e., to work in terms of $s' = s - (2p - 1)n$. After a bit more algebra (rewriting in terms of s', expanding as before, and converting back to s), we arrive at

$$\partial_n P(s, n) = 2p(1 - p)\partial_{ss}P(s, n) - (2p - 1)\partial_s P(s, n) \tag{1.2.21}$$

This is a biased diffusion equation. It can be written as a conservation law,

$$
\begin{aligned}
0 &= \partial_n P + \partial_s J \\
J &= -D\partial_s P + vP
\end{aligned}
\tag{1.2.22}
$$

in which the current J is the sum of diffusive and drift contributions, with diffusion constant $D = 2p(1 - p)$ and drift velocity $v = (2p - 1)$. In retrospect, we could have guessed this form, from the known results for $\bar{s} = \langle s \rangle$ and variance $\langle (s - \bar{s})^2 \rangle$.

Central limit theorem

It is usually the case that elementary steps in a physical process described by a random walk are more complicated than simply one step forward or one step back with equal likelihood. For example, there may be a range of possible elementary steps r, each taken with some probability $p(r)$. What can we say about the resulting probability distribution $P(s, n)$ for these more general random walks?

It turns out that under rather general conditions, the distribution after many steps of a random walk is still approximately Gaussian, of the form

$$P(s, n) = \frac{e^{-(s - n\langle r \rangle)^2 / 2n\langle \Delta r^2 \rangle}}{(2\pi n\langle \Delta r^2 \rangle)^{1/2}} \tag{1.2.23}$$

In the above, $\langle r \rangle$ is the average elementary step (zero for an unbiased walk), and $\langle \Delta r \rangle$ is the average variance in the elementary step ($\Delta r = r - \langle r \rangle$).

This result, called the Central Limit Theorem, is the reason for the ubiquity of the "bell curve" (Gaussian), that the sum of uncorrelated random numbers (here, the steps) has a distribution

that becomes Gaussian as the number of summands becomes large. The theorem holds whenever the elementary step distribution $p(r)$ has finite "moments" (averages of r^k for all positive k) — intuitively, the total number of steps taken must be large, and the probability of rare, giant-sized displacements must be very small, for a Gaussian distribution to emerge.

1.3   Distribution functions and phase space

[Reading: Reif, Chapter 2.1–2.6, Appendix A.13, Section 3.4]

In the last lecture, we introduced the probability distribution $P(s, n)$, which tells how likely it is for a random walker to be found at s after n steps. We can think of $P(s, n)$ as the number of possible paths the walker could have taken that start at $s = 0$ and arrive at s after n steps, normalized by the total number of all possible paths of n steps regardless of destination.

Such sums over configurations of a system that can freely adopt many configurations are a central feature of statistical mechanics. In this lecture, we introduce some basic concepts and machinery that are sufficiently general to describe not only the configurations of polymers, but essentially any collection of a large number of molecules.

Boltzmann factors

Consider a polyethylene (PE) chain. The dihedral angles between each successive pair of CH_2 groups can be rotated with a modest energy cost; as a result, the chain can adopt many conformations at finite temperature. The shortest hydrocarbon chain with this conformational freedom is butane, which has one such dihedral angle between the second and third carbons. At zero temperature, the chain adopts the lowest energy state, in which all bonds are in the trans state, resulting in a planar zig-zag conformation. At higher temperatures, the dihedral angles can take on different values. What governs the probabilities with which these states are found?

It turns out that the probabilities of the available states of a system at a given temperature T are given by the Boltzmann factor:

$$P_r \propto e^{-\beta E_r} \tag{1.3.1}$$

Here r denotes the "microstate", or configuration of the degrees of freedom of the problem, E_r the corresponding energy of that state, and β is $1/(kT)$, where k is the Boltzmann constant. Evidently kT has units of energy (so that the exponent is dimensionless). Roughly speaking, kT is the amount of energy typically available to a single degree of freedom for thermal fluctuations at a temperature T.

Because the system must be in some state or other, the probabilities P_r must sum to unity, which fixes the normalization:

$$P_r = \frac{e^{-\beta E_r}}{Z} \tag{1.3.2}$$

Here Z is the partition function,

$$Z = \sum_r e^{-\beta E_r} \tag{1.3.3}$$

so called because its summands describe how probability is partitioned among the different possible microstates.

We shall first give some examples of how Boltzmann factors behave, and later present the statistical mechanics arguments that lead to this remarkable and useful result.

To compute the probabilities for a given system, we must describe its microstates, and provide the corresponding energies. The choice of microstates is an opportunity for approximation. For butane with its single dihedral angle, the most straightforward choice would be to take the angle ϕ itself as the microstate. The dihedral potential $U(\phi)$ would supply the corresponding energies. However, we note that $U(\phi)$ has three well-defined minima, at angles of roughly $\phi = 0$ (trans), $\phi = 2\pi/3$ (gauche+), and $\phi = 2\pi/3$) (gauche-). The energies of the gauche+ and gauche- states lie about 3kJ/mol above the trans minimum. This suggests a simplified account of the microstates, in which we keep track only of whether the dihedral is in the t, g+, or g- microstate.

Using the Boltzmann factors, we can compute the probabilities of the trans and gauche states as

$$P_t = \frac{e^{-\beta E_t}}{e^{-\beta E_t} + 2e^{-\beta E_g}}$$

$$P_{g+} = P_{g-} = \frac{e^{-\beta E_g}}{e^{-\beta E_t} + 2e^{-\beta E_g}} \tag{1.3.4}$$

We note that only the trans-gauche energy difference $\Delta E = E_g - E_t$ actually enters the probabilities, which can be rewritten as

$$P_t = \frac{1}{1 + 2e^{-\beta \Delta E}}$$

$$P_{g+} = P_{g-} = \frac{e^{-\beta \Delta E}}{1 + 2e^{-\beta \Delta E}} \tag{1.3.5}$$

Furthermore, the energy difference ΔE only enters in the combination $\beta \Delta E$, which is to say $\Delta E/(kT)$, the ratio of ΔE to the "thermal energy" available at temperature T. If ΔE is large compared to kT, the exponential factor $e^{-\beta \Delta E}$ is very small, and we find $P_t \approx 1$, $P_{g+} = P_{g-} \approx 0$; the system is nearly always in the lowest energy, ground state. Whereas, if ΔE is small compared to kT, the exponential factor is very nearly unity, and we find $P_t \approx P_{g+} = P_{g-} = 1/3$; the system is found with equal probability in all accessible states, without regard to the small energy differences between them. Thus the Boltzmann factor adjudicates the contest between energy and entropy: at low temperatures, the tendency of a system to be found in the ground state predominates, while at high temperatures, the tendency of a system to be found randomly in all accessible states prevails.

For alkanes, the trans-gauche energy difference is about 3kJ/mol. The Boltzmann constant is such that at a temperature of $120K$, the thermal energy kT is about 1kJ/mol hence we may write $k = 1/120$kJ/(mol K)]. At room temperature ($T = 300K$ or so), kT is about $300/120 = 2.5$kJ/mol, so $\beta \Delta E = 1.2$; the exponential factors above are neither very small, nor nearly unity. With this value, we find $P_t = 0.62$ or so, and $P_{g+} = P_{g-} = (1 - P_t)/2 = 0.19$.

In the more precise view of the dihedral microstate as the angle itself, the microstates are a one-dimensional continuum, and the "sum" over microstates $\sum_r$ becomes an integral over ϕ. The probability of a given angle ϕ is written

$$P(\phi) = \frac{e^{-\beta U(\phi)}}{Z} \tag{1.3.6}$$

with the partition function given by

$$Z = \int_0^{2\pi} d\phi \, e^{-\beta U(\phi)} \tag{1.3.7}$$

The expressions look different, but fundamentally they are the same: ϕ is the microstate r, the integral $\int_0^{2\pi} d\phi$ is the sum $\sum_r$ over microstates, and the $e^{-\beta U(\phi)}$ is the Boltzmann factor $e^{-\beta E_r}$. The probability $P(\phi)$ is properly normalized by the denominator $\int_0^{2\pi} d\phi \, P(\phi)$ gives unity, as the denominator is designed to cancel the integral of the numerator.

Qualitatively, $P(\phi)$ is large where $U(\phi)$ is small, and nearly constant where $U(\phi)$ is constant to kT or so. The trans and gauche minima show up in $P(\phi)$ as three well-defined peaks. The width of these peaks is determined by the shape of the minima in $U(\phi)$. $P(\phi)$ will decrease from its peak value by a factor of $1/e^2$ (about 1/10) when $U(\phi)$ increases from its minimum by about $2kT$. For the alkane dihedral, the trans and gauche minima in $U(\phi)$ have similar widths, so that the peaks in $P(\phi)$ likewise have similar widths. The "barriers" in $U(\phi)$, i.e., the maxima between its various minima, give rise to minima in $P(\phi)$. When the barriers are many kT high, the corresponding

minima are very deep, with very small probability of the microstate being found near the top of the barrier.

Given the continuous probability distribution $P(\phi)$, we could define what it means to be in the trans, gauche+, or gauche- states by partitioning the range of ϕ into regions, perhaps between the minima of $P(\phi)$. We could then integrate (numerically) $P(\phi)$ in these regions to compute P_t and $P_{g\pm}$ (which, because $P(\phi)$ is properly normalized, would sum to unity). In such integrals, it is clear that the width as well as the height of the peaks in $P(\phi)$ matter in determining P_t and $P_{g\pm}$.

Indeed, we could roughly approximate P_t as the peak height times the width — the largest typical probability of the set of microstates near the trans minimum, times the number of states within kT or so of the minimum. A deep but narrow minimum would have very favorable energy (a tall peak), but few nearby states; a shallow by wide minimum would have a shorter but broader peak. The probabilities of being found near two distinct minima 1 and 2 will be about equal when

$$N_1 exp(-\beta E_1) \approx N_2 exp(-\beta E_2) \tag{1.3.8}$$

where N_1 and N_2 are the number of microstates (range of ϕ values, in our dihedral example) within kT or so of the corresponding minima.

Taking the log and multiplying by kT, this can be rearranged to

$$E_1 - kT \log N_1 = E_2 - kT \log N_2 \tag{1.3.9}$$

This result is very reminiscent of the thermodynamic relation between two phases in equilibrium, for which (neglecting contributions from volume changes) the Helmholtz free energies $F = E - TS$ are equal. This foreshadows the statistical definition of the entropy S as $S = k \log N$, where N is the number of accessible states, which we shall explore in detail in lectures to follow.

Isolated system in equilibrium

So why are microstate probabilities governed by the Boltzmann factor? To begin to answer this question, we investigate first an isolated system in equilibrium.

For example, consider a container of gas, rigid so the volume is fixed, sealed so no molecules can enter or leave, insulated so no energy can get in or out. When we say the system is in equilibrium, we mean that if the system is examined repeatedly, its average properties do not change with time. So if initially the top half of the system is hotter than bottom half, eventually energy will flow from top to bottom, and the temperature will become uniform. Likewise, if at first the top half has more molecules of a given type than the bottom half, eventually molecules will diffuse from top to bottom, and the concentration will even out. If the pressure is initially nonuniform, molecules will flow to equalize the pressure. In equilibrium, everything that can change, has changed.

In analyzing the probability distribution of random walkers, we introduced the idea of a large collection of walkers, all executing different random walks. This large collection of identical systems, which obey the same physical laws but have different detailed microstates, is called an ensemble. For the example of a container filled with gas, we may think of an ensemble as a large number of containers identically prepared. More practically, we can consider a large number of smaller but still macroscopic volumes within the container, each containing a large number of molecules. Finally, we can obtain an ensemble from observing the same system many times. Because the system is in equilibrium, its average properties do not change with time, so that the same system at a later time is in a practical sense a "different" system with the same properties.

In the isolated container of gas, many things can change over time. The particular arrangements of molecules are not permanent; the molecules diffuse about the system, rotate, vibrate, constantly

changing velocity as a result of collisions with each other. What properties of the isolated system are guaranteed to remain the same? Only those quantities that are conserved, and cannot change: the number and type of molecules (because the container is sealed, assuming no reactions take place), the volume (because the container is rigid), and the total energy (because the container insulated so no heat can flow in, and rigid so no work can be done). Everything else can change with time.

Motivated by this discussion, an inspired guess for the probabilities for different microstates of an isolated system in equilibrium is: every accessible microstate is equally likely. If they were not equally likely, and remained so over time, then "something else" would be conserved — we would know something else permanent about the system. (This assertion can be proved, for systems described by classical mechanics.) So for example, if an isolated container of dilute gas initially has all N molecules in the top half of the container, this is only a small fraction of all the accessible states (a fraction 2^{-N}, to be precise). Such a system is not in equilibrium, because the accessible states are not equally likely.

Microstates and phase space

Consider a system consisting of a large number of atoms or molecules, such as a container filled with gas or fluid. To fully describe the microstates of such a system, we must specify the position vectors $\{q_i\}$ and momentum vectors $\{p_i\}$ (for $i = 1, 2, \ldots N$) of all N particles in the system. For a system of macroscopic size, this will involve a truly enormous number of degrees of freedom, which we certainly will not ever explicitly specify.

Still, we can think of the microstate of the system as fully determined by the set of $\{q_i\}$ and $\{p_i\}$, because Newton's equations of motion are first-order differential equations that propagate the positions and momenta forward in time:

$$\dot{p}_i = -dU/dq_i \tag{1.3.10}$$
$$\dot{q}_i = p_i/m_i \tag{1.3.11}$$

Here $U(q_1, q_2, \ldots q_N)$ is the potential energy of the system; the derivative of the potential energy $-\partial U/\partial q_i$ with respect to the position vector q_i gives the force on the ith particle. (For mesoscopic systems consisting of thousands to millions of atoms, we can use computers to simulate the motion of the system, by solving the corresponding set of ODEs numerically.)

So the state of the entire system can be thought of as a point in "phase space", which is a $6N$-dimensional space of position and momentum coordinates. The dynamical trajectory of the system — how the entire set of positions and momenta vary — is a path of a point in phase space. Only one trajectory passes through a given point. Note also that the equations of motion are reversible; we can integrate backwards as well as forwards, and indeed we cannot tell from watching a movie of the motion of classical particles whether the movie is being run backwards or forwards.

The total energy of the system can be written

$$E = K + U = \sum_i \frac{p_i^2}{2m_i} + U(q_1, q_2, \ldots, q_N) \tag{1.3.12}$$

For an isolated system, the total energy is conserved, as can be seen directly by using the equations

of motion to compute $\dot{E}$,

$$
\begin{aligned}
\dot{E} &= \sum_i \left(\frac{\partial E}{\partial q_i} \dot{q}_i + \frac{\partial E}{\partial p_i} \dot{p}_i \right) \\
&= \sum_i \left(\frac{\partial U}{\partial q_i} \frac{p_i}{m_i} - \frac{p_i}{m_i} \frac{\partial U}{\partial q_i} \right) \\
&= 0
\end{aligned}
\tag{1.3.13}
$$

As a consequence of energy conservation, the constraint of fixed energy limits the accessible states for an isolated system. For a system such as a fluid, this constraint is very awkward to deal with, because a substantial part of the energy is potential, which depends on the relative locations of the particles. When particles in a fluid strongly overlap, for example, the energy greatly increases; if fluid particles are too dilute, attractive interactions are broken and the energy plummets. The constraint of fixed energy thus strongly affects the accessible microstates to those for which particles do not overlap too much nor too little, which is awkward to describe and enforce.

Density of states

It will turn out to be a useful thing to know "how many states there are" in phase space for a system with a given energy. This is usually unmanageable because of the complicated way fixed energy constrains the positions of particles; however, for an ideal gas, the fixed energy constraint acts in a simple way on the momenta.

For a dilute gas, interactions between molecules are very rare, so that the energy is essentially all kinetic. The constraint of fixed energy thus affects only the particle momenta. In the particular case of a dilute gas of featureless atoms (argon is a good example), for which the particles all have the same mass and move independently, energy conservation implies the sum of the squares of the momenta $\sum_i p_i^2$ is fixed:

$$
\sum_i p_i^2 = 2mE
\tag{1.3.14}
$$

So we consider an ideal gas of N particles of mass m in a volume V with fixed energy. Suppose we know that the system has energy between E and $E + \delta E$ (δE is a small uncertainty in our knowledge of the exact energy). We expect that the number of states $\Omega(E, \delta E)$ accessible to the system with energy in this range will be proportional to the narrow width δE. So we define $\omega(E)$, the "density of states", such that

$$
\Omega(E, \delta E) = \omega(E)\delta E
\tag{1.3.15}
$$

Briefly, the density of states is important because if a system is able to absorb some energy from its surroundings, there will be more states accessible as a result. If two systems are in thermal contact — able to exchange energy — they will do so, in such a way that the combined system can access the largest number of states in phase space. So to predict how much heat energy will pass between two systems, we need to know about the density of states.

To count the states accessible to the system, it is convenient to impose a "mesh size" in phase space, so that the set of states becomes "discrete". (Otherwise we must compute the *volume* of phase space accessible to the system.) Quantum mechanics provides a natural mesh size for phase space via the uncertainty relation,

$$
\Delta p \Delta q \geq \hbar
\tag{1.3.16}
$$

As we will see below, our main results will not depend on the size of the phase space mesh we impose. Thus we shall partition the total volume V of the system into many small volumes of

size V_0. Counting the number of spatial configurations of N particles of a dilute gas is then very simple: each particle can be placed in V/V_0 different "boxes". The energy does not constrain the particle locations, and because the gas is dilute we ignore the small probability of two particles ending up in the same box. To count the configurations, we temporarily affix a unique label on each particle. Then the number of spatial configurations of all N particles would be $(V/V_0)^N$. The labels are artificial, however; the same configuration of unlabeled identical particles will be generated $N!$ times, corresponding to the possible arrangements of the labels on the particles. So we divide by $N!$, to obtain the number of spatial configurations as

$$\Omega = \frac{(V/V_0)^N}{N!} \tag{1.3.17}$$

Energy conservation constrains the sum of the squares of the particle momenta to a fixed value. We can think of the set of particle momenta as forming a big $3N$-dimensional vector. The constraint forces this vector to lie on the surface of a sphere in $3N$-dimensional space. The radius p of the sphere is such that $p^2/(2m) = E$, or $p = \sqrt{2mE}$. The coordinates of the positions are all independent, and all lie somewhere within the volume V.

The volume of a sphere of radius R in n dimensions scales as R^n — the "volume" (area) of a "two-dimensional sphere" (circle) is πR^2, the volume of a three-dimensional sphere is $(4/3)\pi R^3$, and so forth. Correspondingly, the surface area of a sphere in d dimensions scales as R^{n-1}; the "area" (perimeter) of a "two-dimensional sphere" (circle) is $2\pi R$, the area of a three-dimensional sphere is $4\pi R^2$, and so on.

It turns out that the volume of a sphere in n dimensions of unit radius is

$$V_n = \frac{2\pi^{n/2}}{n\Gamma(n/2)} \tag{1.3.18}$$

in which $\Gamma(x) = (x-1)!$ is the gamma function.

We can find the phase space volume of accessible states with energy less than E, by the volume of a $3N$-dimensional hypersphere with radius $R = \sqrt{2mE}$, times V^N for the phase space volume explored by N independently moving particles:

$$\int^E \omega(E')dE' = V^N V_{3N}(2mE)^{3N/2} \tag{1.3.19}$$

Differentiating with respect to E, we have finally

$$\omega(E) = V^N V_{3N}(2mE)^{3N/2}(3N/2E) \tag{1.3.20}$$

The important feature of this result is that for a large number of particles N, $\omega(E)$ is an extremely rapidly growing function of E, growing as $E^{3N/2}$. This strong dependence on E arises because there are so many more ways to choose the momenta of a large number of particles, if there is more energy to go around. Although we have derived this only for the special case of an ideal gas, this is a general feature of the density of states, which we shall make use of in what follows.

1.4   Microcanonical ensemble

[Reading: Reif 3.3, 3.6–3.9, 3.12]

In the last lecture, we introduced the idea of an ensemble — a large set of copies of a system, with different microstates but the same constraints. (The ensemble in which the constraints are fixed energy, volume, and number of particles, appropriate to an isolated system, is called the microcanonical ensemble.) We argued that such a system, isolated and allowed to equilibrate, would be found with equal probability in any accessible state. Furthermore, that if a constraint were removed (like a separator in a vessel, preventing gas from accessing part of the volume) the system would over time uniformly populate the newly expanded phase space volume.

The assertion that an isolated system in equilibrium is found with equal probability in any accessible state, turns out to be very powerful; in fact, it is equivalent to the second law of thermodynamics. It turns out that we may identify the entropy S of an isolated system as

$$S = k \log \Omega \tag{1.4.1}$$

in which k is Boltzmann's constant and Ω is the number of accessible states. In the next two lectures, we shall show that this microscopic identification of entropy is consistent with macroscopic thermodynamics.

Intuitively, entropy should be a measure of the randomness of an ensemble. At first sight, one might have guessed that a good measure of randomness would be the number of accessible states itself. However, this would not be an extensive quantity, which doubles when we double the size of the system. If we have two weakly interacting systems A and B (two halves of a big volume of fluid), the number of states Ω_{A+B} of the combined system $A + B$ will be roughly Ω_A *times* Ω_B — a state of the combined system can be formed by independently choosing any state of A and any state of B. If we take the log, however, we have $\log \Omega_{A+B} = \log \Omega_A + \log \Omega_B$, which doubles when we double the system size.

Two systems exchanging energy

Consider two large systems A and A′ initially isolated and in equilibrium. Then, the systems are brought into thermal contact, in such a way that energy can be exchanged between them. The thermal contact is equivalent to the loss of a constraint (of each system at its initial energy).

After the combined system equilibrates, the total energy E_T will be partitioned between the two systems in various ways. We may ask, what is the probability $P(E)$ of finding the energy of system A equal to E? According to our postulate that all accessible states are equally likely, we must have

$$P(E) \propto \Omega_T(E) = \Omega_A(E)\Omega_{A'}(E_T - E) \tag{1.4.2}$$

in which Ω_T is the total number of accessible states in which system A has energy E.

Now both $\Omega_A(E)$ and $\Omega_{A'}(E')$ are very strongly increasing functions of their respective arguments. Because $E + E'$ is fixed, as we transfer energy from A to A', the entropy of A' increases sharply, while the entropy of A falls sharply. In somewhat colorful language, we may describe the two systems as competing for energy, each seeking to maximize its entropy at the expense of the other. As a result, the function $\Omega_T(E)$ will be a very strongly peaked function.

We can find the most likely value of E by maximizing $\Omega_T(E)$, or more conveniently the log of $\Omega_T(E)$:

$$0 = d \log \Omega_T = \left(\frac{\partial \log \Omega_A}{\partial E} - \frac{\partial \log \Omega_{A'}}{\partial E'} \right) dE \tag{1.4.3}$$

We introduce the definition

$$\beta(E) = \left. \frac{\partial \log \Omega}{\partial E} \right|_V \tag{1.4.4}$$

and we have

$$\beta_A(E) = \beta_{A'}(E') \tag{1.4.5}$$

Note that $k\beta$ is the derivative $(\partial S/\partial E)|_V$; and we recall from thermodynamics, the fundamental equation

$$dE = TdS - PdV + \mu dN \tag{1.4.6}$$

gives $(\partial S/\partial E)_{V,N}$ equal to $1/T$. So $\beta = 1/kT$, presuming our identification of the entropy as $k \log \Omega$ holds. What we have shown is, the energy will partition itself in equilibrium so that the temperature of the two systems is equal.

Suppose that before the thermal contact is made, the two systems do not have the same temperature (that is, their individual values of $\partial \log \Omega/\partial E$ are not equal). Then energy will flow from one system to the other. Suppose $\beta < \beta'$, which is to say $T > T'$. To reach the maximum $P(E)$, we need to increase β for the A$'$ system and decrease it for the A system. Remember that Ω grows like $E^{3N/2}$ (we showed this explicitly for the ideal gas), so that $\beta(E)$ goes roughly as $3N/2E$. So to equilibrate the two systems, we must move energy from the hotter system to the colder, as we expect.

Exchange of volume

In the same way we treated two systems able to exchange energy to obtain equilibrium conditions for systems in thermal contact, we can describe two systems that can exchange both energy and volume. For example, consider a cylinder with a movable, thermally conducting piston separating two volumes of gas.

The total number of accessible states Ω_T is the product of the number of accessible states for each volume of gas:

$$\Omega_T = \Omega(E, V)\Omega'(E', V') \tag{1.4.7}$$

Here E and E' are the energies of the two systems, and V and V' the volumes. The total energy $E_T = E + E'$ and total volume $V_T = V + V'$ are both fixed;.

In equilibrium, the energy and external parameter values will be such that the number of accessible states Ω_T is maximum for the two systems. Hence we have

$$0 = d \log \Omega_T = \left(\frac{\partial \log \Omega}{\partial E} - \frac{\partial \log \Omega'}{\partial E'} \right) dE + \left(\frac{\partial \log \Omega}{\partial V} - \frac{\partial \log \Omega'}{\partial V'} \right) dV \tag{1.4.8}$$

For $\log \Omega_T$ to be maximized, the partial derivatives with respect to E and V must vanish separately. The first of these conditions gives thermal equilibrium and equal temperatures, as we showed before. The second condition is

$$\frac{\partial \log \Omega}{\partial V} = \frac{\partial \log \Omega'}{\partial V'} \tag{1.4.9}$$

To see what this condition means, recall the fundamental equation of thermodynamics,

$$dE = TdS - PdV \tag{1.4.10}$$

from which we easily obtain

$$\left(\frac{\partial S}{\partial V} \right)_E = P/T \tag{1.4.11}$$

If we accept the identification of the thermodynamic entropy S as $k \log \Omega$, the derivative $\partial \log \Omega / \partial V$ becomes βP. Then the second equilibrium condition corresponds to the expected statement that the pressures P and P' of the two systems are equal in equilibrium.

Ideal gas properties.

We can determine the entropy of an ideal gas (up to an overall constant) from our previous result for the density of states $\omega(E)$, by taking the log. We retain only the largest terms, bearing in mind that N, V, and E all are proportional to the (large) size of the system. We obtain

$$S = k \log \Omega = kN \log V + \frac{3kN}{2} \log E + \text{smaller terms} \tag{1.4.12}$$

From this, we can compute the pressure and temperature of the ideal gas using the standard thermodynamic derivatives:

$$P = \frac{1}{\beta} \frac{\partial \log \Omega}{\partial V} = T \left(\frac{\partial S}{\partial V} \right)_E = \frac{NkT}{V} \tag{1.4.13}$$

which is the ideal gas equation of state.

Likewise, we find the temperature from

$$\beta = \frac{1}{kT} = \frac{\partial \log \Omega}{\partial E} = \frac{1}{k} \left(\frac{\partial S}{\partial E} \right)_V = \frac{3N}{2E} \tag{1.4.14}$$

This we rewrite to obtain

$$E = (3/2)kTN \tag{1.4.15}$$

The energy per molecule in our ideal gas of structureless molecules (argon, for example) is equal to $(3/2)kT$. Temperature is a measure of internal energy.

Taking the derivative to obtain the heat capacity, we have

$$C_V = \left(\frac{\partial E}{\partial T} \right)_V = (3/2)kN = (3/2)Rn \tag{1.4.16}$$

in which R is the molar gas constant and n the number of moles in the system. This is the usual thermodynamics result for the heat capacity of an ideal gas of structureless molecules.

1.5   Canonical ensemble

[Reading: Reif 6.2–6.10]

If statistical mechanics were only capable of reproducing the content of equilibrium thermodynamics from a microscopic perspective, it would be perhaps enlightening, but not especially useful. In fact, statistical mechanics can describe not only the mean values of thermodynamic quantities such as energy, temperature, and pressure, but fluctuations in these and other such quantities as well.

Reservoirs and small systems

Fluctuations are most important in small systems. We therefore begin by considering a small system in thermal contact with a large system. Physically, the large system serves as a thermal reservoir, fixing the temperature of the small system. A reservoir simply means a large enough system so that any amount of energy that may likely be exchanged with a smaller system in thermal contact, would be only a tiny fraction of the energy in the reservoir system.

We return to the probability $P(E)$ that system A has energy E, while system A$'$ (the reservoir) has energy $E' = E_0 - E$:

$$\log P(E) = \log \Omega_A(E) + \log \Omega_{A'}(E_0 - E) + c \tag{1.5.1}$$

Now, however, we assume that the system A$'$ is large; specifically, in expanding $\log P(E)$, we expand $\log \Omega_{A'}$ to first order only — i.e., we assume that the energy E is small compared to E_0. We obtain

$$\log P(E) \approx \log \Omega_A(E) - \beta E \tag{1.5.2}$$

in which β is the inverse temperature $\partial \log \Omega_{A'}/\partial E$ of the reservoir.

When the small system A is in a state with energy E, it has some number of microstates $\Omega_A(E)$ available to it, all equally probable. The probability $P(r)$ of each microstate individually, then, is

$$P(r) = P(E_r)/\Omega_A(E_r) \propto \exp(-\beta E_r) \tag{1.5.3}$$

This result is central: the probability of any single microstate of energy E_r in a system in contact with a thermal reservoir at inverse temperature β is proportional to $\exp(-\beta E_r)$. This factor is called a "Boltzmann factor". The reservoir "sets the price of energy", by the dependence of the reservoir entropy on E. The Boltzmann factors bias the microstates of the small system towards smaller energy. If the temperature is higher, high-energy states become more likely.

Partition function

In the ensemble appropriate to a system in equilibrium with a thermal reservoir, called the "canonical" ensemble, the probability of being in a microstate r is $\exp(-\beta E_r)$. The system energy is not strictly fixed; all states are summed over. The average energy is given by

$$\bar{E} = \frac{\sum_r E_r \exp(-\beta E_r)}{\sum_r \exp(-\beta E_r)} = \frac{\sum_E \Omega(E) \exp(-\beta E)}{\sum_E \Omega(E) \exp(-\beta E)} \tag{1.5.4}$$

Observe that this average can be written in the following way:

$$\bar{E} = -\frac{\partial \log Z}{\partial \beta} \tag{1.5.5}$$

in which we have defined Z, the partition function (so named, because it describes how probability is "partitioned" among different microstates), as

$$Z = \sum_r \exp(-\beta E_r) \qquad (1.5.6)$$

The derivative with respect to β "brings down" a factor of $-E$, and taking the derivative with respect to $\log Z$ puts a Z in the denominator for normalization.

Higher moments of E, such as the variance of the energy, written $\langle (E - \bar{E})^2 \rangle$, can be calculated by taking more derivatives of $\log Z$. In fact, we shall show if we know the partition function, we can calculate all other thermodynamic quantities. (Of course, it will turn out that calculating the partition function is not easy, but we will develop a number of approximate methods for this purpose in subsequent lectures.)

To begin with, how can we compute the entropy from the partition function? First of all, observe that we can write the partition function as

$$Z = \sum_E \Omega(E) \exp(-\beta E) \qquad (1.5.7)$$

since the density of states $\Omega(E)$ counts the number of states with a given energy. For a large system, for which the density of states is an extremely rapidly increasing function of E, the sum is very well approximated by expanding about the most likely value of energy.

To carry out this expansion, we write the summand as an "exponential of something" [namely, $-\beta E + \log \Omega(E)$], and then expand the exponent about its largest value. The corresponding peak value E^* is found from

$$0 = \frac{\partial \log \Omega(E^*)}{\partial E} - \beta \qquad (1.5.8)$$

This equation expresses the fact that the most likely value of the energy is the one satisfying the thermodynamic relation $1/T = (\partial S/\partial E)_V$.

Because we are expanding about the maximum of the exponent function, the linear term vanishes; to second order we have

$$-\beta E + \log \Omega(E) \approx -\beta E^* + \log \Omega(E^*) + (1/2)\frac{\partial^2 \log \Omega}{\partial E^2}(E - E^*)^2 \qquad (1.5.9)$$

We can identify the derivative coefficient in the second-order term as $\partial \beta/\partial E$, which is negative (average E increases as T increases, and $\beta = 1/(kT)$), and scales as the reciprocal of the system size (E is extensive and β is not).

Using the expansion, we approximate Z as

$$Z \approx \Omega(E^*)e{-}\beta E^* \sum_E e^{-(1/2)|\partial \beta/\partial E|(E-E^*)^2} \qquad (1.5.10)$$

The expansion results in a Gaussian distribution of energy fluctuations about the most probable value E^*, and a Gaussian integral to compute to obtain Z. We obtain

$$Z \approx \Omega(E^*)e{-}\beta E^* \left(2\pi/|\partial \beta/\partial E|\right)^{1/2} \qquad (1.5.11)$$

The variance of the energy fluctuations is given by the width of the Gaussian, $\sigma_E^2 = |\partial \beta/\partial E|^{-1}$. We can get an idea of the magnitude of energy fluctuations in a typical system by considering the simplest case of an ideal gas, for which $\beta = 3kN/(2E)$. Hence $\partial \beta/\partial E = -\beta/E$ for the ideal gas, and the variance of the energy is E^*kT.

Now the mean energy E^* itself is of order kT times the number of particles N, so the standard deviation of the energy compared the mean energy is

$$\frac{\sigma_E}{E^*} \sim \left(\frac{kT}{E^*}\right)^{1/2} \sim \frac{1}{\sqrt{N}} \tag{1.5.12}$$

For macroscopic systems (one mole or so, say about 10^{24} molecules), fluctuations in the system energy are extremely small, one part in 10^{12}. But for the number of molecules in a small simulation (10^4, say), or in a small local region of a fluid, fluctuations can be quite significant.

From the above, we see that a serviceable approximation to Z for macroscopic systems (in the sense of a good representation of the largest factors in Z, and hence a good approximation for $\log Z$) is simply $\Omega(E^*)\exp(-\beta E^*)$. Writing the mean value of energy $\bar{E} = E^*$, and recalling $S = k\log\Omega$, we have

$$Z = \exp(-\beta(\bar{E} - TS)) \equiv \exp(-\beta F) \tag{1.5.13}$$

in which F is the usual Helmholtz free energy from macroscopic thermodynamics. The log of the canonical partition function gives the Helmholtz free energy.

Ensemble averages

It turns out that the canonical ensemble, in which the system energy is summed over, is much more convenient for calculations of average quantities than the "microcanonical" ensemble, appropriate to an isolated system, in which the energy is strictly fixed.

In fact, if we are interested only in average quantities and not in the detailed probability of fluctuations away from the average, it is sometimes convenient to replace a microcanonical ensemble by a canonical one. We may think of this physically as placing an isolated system into a thermal bath adjusted so that the mean energy of the system is the same as its isolated energy. Or, mathematically, we can think of the value of β as being chosen so that the average energy is some specified value.

The average energy in the canonical ensemble (written $\bar{E}$, or $\langle E \rangle$) is given by

$$\langle E \rangle = \frac{\sum_r E_r e^{-\beta E_r}}{\sum_r e^{-\beta E_r}} = -\frac{1}{Z}\frac{\partial Z}{\partial \beta} = -\frac{\partial \log Z}{\partial \beta} \tag{1.5.14}$$

Note how taking the derivative of Z with respect to $-\beta$ "pulls down" a factor of E_r inside the sum, and how taking a derivative of $\log Z$ results in a factor of Z in the denominator, which we need to normalize the average.

Observe what happens when we take another derivative $-\partial/\partial\beta$ of $\log Z$:

$$\begin{aligned}
\frac{\partial^2 \log Z}{\partial \beta^2} &= -\frac{\partial}{\partial \beta}\left(\frac{\sum_r E_r e^{-\beta E_r}}{Z}\right) \\
&= \frac{\sum_r E_r^2 e^{-\beta E_r}}{Z} - \left(\frac{\sum_r E_r e^{-\beta E_r}}{Z}\right)^2 \\
&= \langle E^2 \rangle - \langle E \rangle^2 = \langle (E - \bar{E})^2 \rangle = \langle \Delta E^2 \rangle
\end{aligned} \tag{1.5.15}$$

One derivative of $\log Z$ with respect to β gives the average energy $\langle E \rangle$; two such derivatives give the variance $\langle \Delta E^2 \rangle$.

For systems well described by classical mechanics, the total energy E can be written as the sum of potential energy U and kinetic energy K, in which U is a function only of positions, and K is

the sum of kinetic energies for each independently moving particle:

$$K = \sum_i K_i = \sum_i p_i^2/(2m_i) \tag{1.5.16}$$

When K takes this form, it turns out to be easy to compute within the canonical ensemble the average kinetic energy of one moving particle (particle 1, say):

$$
\begin{aligned}
\langle K_1 \rangle &= \frac{\sum_r e^{-\beta E_r} p_1^2/(2m_1)}{Z} \\
&= \frac{\int dp_1 dp_2 \ldots dp_N \int dq_1 dq_2 \ldots dq_N \, e^{-\beta U(q_1,q_2,\ldots q_n)} e^{-\beta(p_1^2/(2m_1)+(p_2^2/(2m_2)+\ldots(p_N^2/(2m_N))} p_1^2/(2m_1)}{\int dp_1 dp_2 \ldots dp_N \int dq_1 dq_2 \ldots dq_N \, e^{-\beta U(q_1,q_2,\ldots q_n)} e^{-\beta p_1^2/(2m_1)+(p_2^2/(2m_2)+\ldots(p_N^2/(2m_N))} p_1^2/(2m_1)} \\
&= \frac{\int dp_1 \, e^{-\beta p_1^2/(2m_1)} p_1^2/(2m_1)}{\int dp_1 \, e^{-\beta p_1^2/(2m_1)}} \\
&= 3/(2\beta) = 3kT/2 \tag{1.5.17}
\end{aligned}
$$

In the above, the great simplification is that the sum over microstates, which corresponds to an integral over all the $\{p_i\}$ and $\{q_i\}$, factors into a product of Gaussian integrals over each p_i separately, times a big multidimensional integral over all the $\{q_i\}$ (which interact through $U(q_1,q_2,\ldots q_N)$. So the factors in the denominator Z cancel almost everything in the numerator, which leads to the second line above.

The ratio of Gaussian integrals is then computed using the same techniques as in the previous section (see Appendix B). Note that the momentum p_1 is a three-dimensional vector, so that $p_1^2 = p_{1,x}^2 + p_{1,y}^2 + p_{1,z}^2$; each component can be averaged separately, and each gives a contribution to $\langle K_1 \rangle$ of $kT/2$, for a total of $3kT/2$.

In fact, the entire velocity distribution for a single particle in the system can be obtained by asking for the probability that particle 1 has momentum p_1; since the sum over values of p_1 is independent of the rest of the partition function, we have

$$P(p_1) \propto e^{-\beta p_1^2/(2m_1)} \tag{1.5.18}$$

which is the Maxwell distribution.

A similar result holds for any "harmonic" degree of freedom in the system, i.e., any degree of freedom for which the contribution to the energy is of the form $(1/2)Kx^2$, with no coupling to other degrees of freedom. Vibrational modes, of isolated molecules or larger objects, can often be treated in this way. Applying the same approach as for particle momenta, we conclude that each such mode has a mean energy of $(1/2)kT$, and a corresponding mean-square amplitude $\langle x^2 \rangle$ of kT/K.

Often we are interested in the probability distribution for the speed of a particle in the system, independent of which direction the particle is moving. We can obtain this from $P(p)$ by summing over all momenta, subject to the constraint that the magnitude of the momentum corresponds to the desired speed. We enforce the constraint with a delta function, leading to

$$P(v) = \int d^3p \, P(p)\delta(|p| - mv) \propto 4\pi \int_0^\infty p^2 dp \, e^{-\beta p^2/(2m)} \delta(|p| - mv) \propto v^2 e^{-\beta m v^2/2} \tag{1.5.19}$$

in which we have evaluated the integral over p in spherical coordinates.

The factor v^2 biases the speed distribution to larger values; it results from the "measure factor" $4\pi p^2 dp$, which reflects the fact that the velocity vectors with a given speed lie on a spherical shell in phase space, and larger speeds correspond to a shell with larger area (hence more possible velocity values).

Information-theoretic entropy

Consider an arbitrary ensemble, defined by the set of probabilities $\{P_r\}$ for a system to be found in microstate r. We would like to define the entropy for such an ensemble. Our microcanonical definition $S = k \log \Omega$ where Ω is the number of accessible states is no longer suitable, because the accessible states are not equally probable. For example, if some fraction of the states are much less likely to occur than the rest, which are themselves equally probable, the entropy should approach that of a microcanonical ensemble of the more-likely states alone.

A robust definition of the entropy can be given by repeatedly sampling the ensemble, to create a long sequence of N microstates. If the ensemble were a single time-dependent system, we could repeatedly observe the microstate, generating a "slide show" of N successive observations of the microstate. Suppose we generate such slide shows repeatedly. Intuitively, if we get the same slide show every time, the entropy of the system is very small; if we get a great variety of slide shows, the entropy is large.

The entropy of the ensemble can then be defined in terms of the number of possible slide shows, i.e., the number of ways that a very large number N of distinct observations can be selected from the many microstates, consistent with the probabilities P_r. On the average, there will be $a_r = NP_r$ observations (slides) of microstate r. The slides come in random order. The number of ways Γ to arrange the slides is then the multinomial coefficient,

$$\Gamma = \frac{N!}{a_1! a_2! a_3! \dots} \tag{1.5.20}$$

As usual, we expand $\log \Gamma$ using the lowest-order Stirling approximation,

$$\begin{aligned} \log \Gamma &= \log N! - \sum_r \log a_r! \\ &\approx N \log N - \sum_r (NP_r) \log(NP_r) \\ &= -N \sum_r P_r \log P_r \end{aligned} \tag{1.5.21}$$

We see that $\log \Gamma$ grows linearly with N (the length of the slide show), times the information theoretic entropy of the ensemble S_i, given by

$$S_i = -\sum_r P_r \log P_r \tag{1.5.22}$$

Note that if all of the Ω microstates are equally likely, so that P_r are all equal to $1/\Omega$, S_i reduces to $\log \Omega$, our microcanonical definition of entropy. If some subset of the P_r approach zero, the corresponding terms $P_r \log P_r$ in S_i likewise vanish. If the remaining Ω' states are equally likely (with probability $1/\Omega'$), S_i again reduces to $\log \Omega'$, the microcanonical entropy of the accessible states.

In fact, we can show that taking all P_r equal maximizes S_i, subject to the constraint that the sum of the P_r must equal unity. To carry out this constrained minimization, we use the method of Lagrange multipliers, which instructs us to solve

$$0 = \frac{\partial}{\partial P_r} \left(-\sum_r P_r \log P_r - \alpha \sum_r P_r \right) \tag{1.5.23}$$

This leads after a bit of algebra to

$$\log P_r = -1 - \alpha \tag{1.5.24}$$

which is to say, $\log P_r$ is a constant, independent of r. Enforcing the normalization fixes the value of α, or more directly the value of P_r, equal to $1/\Omega$. For an isolated system, the accessible states must have fixed energy, volume, and particle number. So the microcanonical ensemble maximizes the information-theoretic entropy of an isolated system; the most random a system can be, is for every accessible microstate to be equally likely, so that we know absolutely nothing that can help predict what microstates will appear.

Suppose now that we have a system of fixed volume and particle number in contact with a reservoir, so that the energy is no longer fixed, but can fluctuate. Suppose further that we know the mean energy of the system, $\bar{E} = \sum_r E_r P_r$. What is the maximum entropy ensemble consistent with this additional piece of information? Again we use Lagrange multipliers, to maximize S_i with respect to the $\{P_r\}$, subject to the two constraints of 1) fixed $\bar{E}$, and 2) normalization of the probabilities, i.e., $\sum_r P_r = 1$.

This leads to

$$0 = \frac{\partial}{\partial P_r}\left((1/N)\log\Gamma - \beta\sum_r E_r P_r - \alpha\sum_r P_r\right) \tag{1.5.25}$$

After a bit of algebra, we have

$$\log P_r = \beta E_r + \text{const} \tag{1.5.26}$$

This equation can be solved subject to the constraints by

$$P_r = \frac{e^{-\beta E_r}}{\sum_r e^{-\beta E_r}} \tag{1.5.27}$$

The microstate probabilities that maximize the entropy subject to a fixed mean energy are Boltzmann weights, with the Lagrange multiplier β adjusted to give the correct mean energy, i.e., such that the sum $\sum_r P_r E_r$ is indeed equal to the desired average energy $\bar{E}$. The canonical ensemble is the ensemble of maximum entropy subject to the constraint of a fixed average energy.

Grand canonical ensemble

Imagine once more a small system in contact with a large system that acts as a reservoir, but now consider the different ways in which the small system can interact with the reservoir. Just as an isolated system has fixed values of volume V, energy E, and number of particles N for each molecular species present, we may have exchange of these otherwise fixed quantities between the small and large systems.

The canonical ensemble describes a small system in thermal contact with a large system, which means the two systems can exchange energy, but not volume or particles. We can define other ensembles appropriate to small systems able to exchange E and N, or E and V, or V and N with the large system. (We need to keep at least one of the three quantities E, V, or N fixed, or else there is no extensive quantity available to determine the "size" of the small system.)

For the canonical ensemble, the fact that the combined system in equilibrium has maximum entropy turns out to fix the temperature of the small system. We shall see that if the small system can exchange volume with the large system, the pressure of the small system is held fixed by the reservoir; if the two systems can exchange molecules, the chemical potentials of the small system are fixed.

The ensemble that describes a small system able to exchange energy and molecules with a large system, is called the grand canonical ensemble. To find the probability P_r of a microstate r in this ensemble, we use the same approach as for the canonical ensemble.

We first consider two large systems A and A' in contact, able to exchange energy and molecules, with energies E and E' and number of molecules N and N' respectively. The combined system has a fixed total energy $E_T = E + E'$ and total number of particles $N_T = N + N'$. As we have written twice before [see Eqs. (1.4.2) and (1.5.1)], the probability $P(E, N)$ of finding system A with energy E and number of particles N is proportional to the total number of accessible states $\Omega_T(E, N)$ for the combined system,

$$P(E, N) \propto \Omega_T(E, N) = \Omega_A(E, N)\Omega_{A'}(E_T - E, N_T - N) \qquad (1.5.28)$$

In equilibrium, the energy E and number of molecules N in system A must maximize $P(E, N)$, which leads to

$$0 = d\log\Omega_T = \left(\frac{\partial\log\Omega_A}{\partial E} - \frac{\partial\log\Omega_{A'}}{\partial E'}\right)dE + \left(\frac{\partial\log\Omega_A}{\partial N} - \frac{\partial\log\Omega_{A'}}{\partial N'}\right)dN \qquad (1.5.29)$$

In addition to the previous condition that the temperatures of systems A and A' must be equal, we find a new condition arising from the exchange of molecules,

$$\alpha_A(E, N) = \alpha_{A'}(E', N') \qquad (1.5.30)$$

in which we have defined

$$\alpha = \left.\frac{\partial\log\Omega}{\partial N}\right|_{V,E} \qquad (1.5.31)$$

The chemical potential μ is conventionally defined such that $\beta\mu = -\alpha$, so we have

$$\mu = -kT\left.\frac{\partial\log\Omega}{\partial N}\right|_{V,E} \qquad (1.5.32)$$

If the system A' is very large (a reservoir), its own energy E' and number of molecules N' will be approximately constant during any conceivable exchange of energy and molecules with the small system. Then, as in our development of the canonical ensemble, we can write the probability of system A having energy E and number of molecules N as

$$\log P(E, N) \approx \log\Omega_A(E) - \beta E + \beta\mu N + \text{const} \qquad (1.5.33)$$

in which β and μ are the inverse temperature and chemical potential of the reservoir.

The system A has many microstates r with the same value of E and N. In equilibrium, each of these microstates are equally probable. The probability of a single microstate r is then

$$P(r) = P(E_r, N_r)/\Omega_A(E_r, N_r) \propto \exp(-\beta(E_r - \mu N_r)) \qquad (1.5.34)$$

This probability defines the grand canonical ensemble. The corresponding grand partition function Z_g is

$$Z_g = \sum_r e^{-\beta(E_r - \mu N_r)} = \sum_{E,N}\Omega(E, N)e^{-\beta(E - \mu N)} \qquad (1.5.35)$$

In the grand canonical ensemble, both energy and number of particles fluctuate. It remains true that the entropy S is given in terms of the microstate probabilities $P_r = e^{-\beta(E_r - \mu N_r)}/Z_g$ by the information theoretic entropy Eq. (1.5.22) [the demonstration of which is left as an exercise to the reader].

The grand canonical ensemble conveniently describes "open systems" — for example, molecules adsorbing on a surface from solution, or equilibrium between a vapor phase and its coexisting liquid.

The grand canonical ensemble be also used to compute equilibrium average quantities for closed systems (fixed N) when it is convenient to relax the constraint of fixed N, much as the canonical ensemble is more convenient than the microcanonical ensemble even to describe average behavior of isolated systems.

We can evaluate Z_g approximately in the thermodynamic limit, following the same approach as for the canonical ensemble, with the result

$$Z_g \approx \Omega(E^*, N^*)e^{-\beta E^* + \beta \mu N^*} = e^{-\beta(\bar{E} - TS - \mu \bar{N})} \equiv e^{-\beta A} \tag{1.5.36}$$

in which A is the grand canonical free energy, given by

$$A = \bar{E} - TS - \mu \bar{N} \tag{1.5.37}$$

The succession of ensembles, from microcanonical to canonical to grand canonical, corresponds to a succession of relaxing constraints, first on energy, then on number of particles. The partition function goes from $\Omega(E, N)$ (all states of fixed energy equally probable) to $Z = \sum_E \Omega(E, N)e^{-\beta E}$ (states of fixed number of particles and different energies are Boltzmann weighted) to $Z_g = \sum_{E,N} \Omega(E, N)e^{-\beta E + \beta \mu N}$ (states of different numbers of particles and energies are Boltzmann weighted). The log of the successive partition functions, evaluated in the "thermodynamic limit" (large system, neglecting fluctuations) defines a succession of free energy functions, from $\log \Omega = S/k$ to $\log Z = -\beta F = S/k - \beta \bar{E}$ to $\log Z_g = -\beta A = S/k - \beta \bar{E} + \beta \mu \bar{N}$.

Gibbs ensemble

The Gibbs ensemble describes a system consisting of a fixed number N of particles, but able to exchange energy and volume with its surroundings, which act as a constant temperature and pressure reservoir.

Two systems A and A' able to exchange energy and volume can be treated as a single isolated system of fixed total energy and volume, in exactly the same way as in the discussion following Eq. (1.5.28). We may write

$$P(E, V) \propto \Omega_T(E, V) = \Omega_A(E, V)\Omega_{A'}(E_T - E, V_T - V) \tag{1.5.38}$$

and expand as before, imposing the condition that the entropy of the isolated system is maximal to obtain the new equilibrium condition that $\partial \log \Omega / \partial V$ must be the same in the two systems. Recalling Eq. (1.4.11), we find the expected result that the pressure in the two systems must be equal.

If we assume that A' is very large (a reservoir), the same approach we used in Eq. (1.5.33) for the grand canonical ensemble leads to the result analogous to Eq. (1.5.34), namely

$$P(r) \propto \exp(-\beta(E_r + PV_r)) \tag{1.5.39}$$

This probability defines the Gibbs ensemble. The corresponding partition function is

$$Z_g = \sum_r e^{-\beta(E_r + PV_r)} = \sum_{E,V} \Omega(E, V)e^{-\beta(E + PV)} \tag{1.5.40}$$

In the Gibbs ensemble, the energy and volume of the system fluctuate.

We can evaluate Z_G approximately in the thermodynamic limit, with the result

$$Z_G \approx \Omega(E^*, V^*)e^{-\beta(E^* + PV^*)} = e^{-\beta(\bar{E} - TS + P\bar{V})} \equiv e^{-\beta G} \tag{1.5.41}$$

in which G is the Gibbs free energy, given by

$$G = \bar{E} - TS + PV \tag{1.5.42}$$

1.6 Equilibrium and fluctuations

Ensembles of microstates, with probabilities given by their respective Boltzmann factors, give information not only about average quantities, such at the mean energy $\langle E \rangle$ of a system in contact with a temperature reservoir, but describe as well the fluctuations about the average, expressed for example in the variance $\langle \Delta E^2 \rangle$. The canonical, grand canonical, and Gibbs ensembles describe fluctuations of quantities other than the energy, particle number, and system volume. Indeed, any observable parameter that takes on different values in different microstates will fluctuate in the ensemble, and the corresponding Boltzmann weights suffice to predict how likely different values of the parameter will be observed.

Fluctuation probability

Consider a system in equilibrium with a temperature reservoir, thus described by canonical ensemble, in which some observable parameter y fluctuates. We can group the microstates that contribute the partition function according to those with the same value of the fluctuating parameter, as

$$
\begin{aligned}
Z = \sum_r e^{-\beta E_r} &= \sum_y \Omega(y) \frac{\sum_{r \in y} e^{-\beta E_r}}{\Omega(y)} \\
&= \sum_y \Omega(y) \langle e^{-\beta E_r} \rangle_y \\
\equiv \sum_y \Omega(y) e^{-\beta \bar{E}(y)}
\end{aligned}
\tag{1.6.1}
$$

Here $\Omega(y)$ is the number of microstates with a given value of y, and $\langle \cdot \rangle_y$ denotes an average over the microstates a given value of y. The quantity $\bar{E}(y)$ is a sort of "effective energy" of the canonical ensemble with a fixed value of y, defined above by the log of the average Boltzmann factor.

We can also define the restricted canonical partition function $Z(y)$ for a fixed value of y, and the corresponding restricted free energy $F(y)$, as

$$
Z(y) = \sum_{r \in y} e^{-\beta E_r} = e^{-\beta F(y)} = \Omega(y) e^{-\beta \bar{E}(y)}
\tag{1.6.2}
$$

If the system is large enough so that the energy at fixed y is relatively well defined, we can approximate the restricted partition function as we have previously, by identifying the most likely energy of the ensemble $E*(y)$ and the corresponding number of accessible states $\Omega(y)$. This leads to

$$
Z(y) \approx \Omega(y) e^{-\beta E*(y)}
\tag{1.6.3}
$$

from which we see that in this limit (for which the system is reasonably macroscopic) $\bar{E}(y)$ is the typical system energy.

In any case, the probability $P(y)$ of a given value of y in the canonical ensemble is given by

$$
P(y) \propto \Omega(y) e^{-\beta \bar{E}(y)} = e^{-\beta(\bar{E}(y) - T_0 S(y))} = e^{-\beta F_0(y)}
\tag{1.6.4}
$$

The most probable value y^* of the observable minimizes the effective free energy $F_0(y)$. If the system is prepared in a state with y different from y^*, as the system equilibrates it will tend to evolve to configurations with $y = y^*$. Equilibrium is the state in which no spontaneous changes are possible, hence the free energy F_0 is minimum at equilibrium.

Fluctuations about the most probable state are also governed by $P(y)$. The probability of the fluctuating value y relative to its most likely value y^* is governed by the corresponding change in free energy $F_0(y) - F_0(y^*) = \Delta F_0(y)$, and can be written

$$\frac{P(y)}{P(\bar{y})} = e^{-\beta \Delta F_0(y)} \tag{1.6.5}$$

If the change in free energy is small compared to kT, the fluctuation will be commonly observed.

Sometimes, the observable parameter y is coupled to some external force f, with a term $-fy$ in the energy. In this case, the probability $P(y)$ becomes

$$P(y) \propto e^{-\beta(F_0(y) - fy)} \tag{1.6.6}$$

The most probable value of y again maximizes the exponent, which leads to

$$f = \partial F_0(y)/\partial y \tag{1.6.7}$$

This relation tells how large a force is needed to achieve a desired value of y.

The work W done by the external force f, in biasing the parameter away from its unforced value (for simplicity, $y = 0$) to some desired value, can be computed from $dW = f dy$. Integrating from the unforced value ($y = 0$) to the final desired value, we obtain

$$W = F_0(y) - F_0(0) = \Delta F_0 \tag{1.6.8}$$

If we are driving the system to an unlikely value of y (away from the unforced minimum of $F_0(y)$), the process requires work; the minimum work to achieve the unlikely state is W — the change in free energy. Or, we can think of the reverse process: holding the system in an unlikely state to begin with, and slowly releasing the external force. As the force is released, the system does work W on the surroundings; the maximum work produced is again given by $W = -\Delta F_0$ — now negative, because we are computing the work done by the system, not on the system. (This is the origin of the term "free energy" for the function F_0.)

Analogous fluctuation results can be derived for a system in contact with a reservoir that supplies molecules at a fixed chemical potential as well as energy at a fixed temperature. We describe the system with the grand canonical ensemble, writing the partition function as a sum over the values of the fluctuating parameter, as

$$Z_g = \sum_r e^{-\beta(E_r - \mu_0 N_r)} = \sum_y \Omega(y) \langle e^{-\beta(E_r - \mu_0 N_r)} \rangle_y \equiv \sum_y \Omega(y) e^{-\beta(\bar{E}(y) - \mu_0 \bar{N}(y))} \tag{1.6.9}$$

Here $\bar{E}(y)$ are the effective energy and number of particles in the restricted grand canonical ensemble with a fixed value of y, and μ_0 is the chemical potential of the ensemble (fixed by the particle reservoir).

The probability $P(y)$ of a given value of y in the grand canonical ensemble is then

$$P(y) \propto \Omega(y) e^{-\beta(\bar{E}(y) - \mu_0 \bar{N}(y))} = e^{-\beta(\bar{E}(y) - T_0 S(y) - \mu_0 \bar{N}(y))} = e^{-\beta A_0(y)} \tag{1.6.10}$$

Here we have defined the grand canonical free energy A_0 by $A_0 = \bar{E} - T_0 S - \mu_0 \bar{N}$, analogous to the Helmholtz free energy F_0 introduced earlier. Fluctuations in the grand canonical ensemble are governed by the corresponding $P(y)$ in the same way as for the canonical ensemble treated above. Changes occur spontaneously in a system coupled to a thermal and particle reservoir if $\Delta A_0 \leq 0$; equilibrium holds if A_0 is at its minimum.

Finally, we use the Gibbs ensemble to describe fluctuations of a system in contact with a temperature and pressure reservoir. The partition function can be written as a sum over values of the fluctuating parameter y in the same way as Eq. (1.6.9),

$$Z_G = \sum_r e^{-\beta(E_r + P_0 V_r)} = \sum_y \Omega(y) \langle e^{-\beta(E_r + P_0 V_r)} \rangle_y \equiv \sum_y \Omega(y) e^{-\beta(\bar{E}(y) + P_0 \bar{V}(y))} \tag{1.6.11}$$

in which $\bar{E}(y)$ and $\bar{V}(y)$ are the effective energy and volume of the restricted ensemble with a fixed value of y, and P_0 is the pressure of the ensemble (fixed by the reservoir). The probability $P(y)$ of a given value of y in the Gibbs ensemble is then

$$P(y) \propto \Omega(y) e^{-\beta(\bar{E}(y) + P_0 \bar{V}(y))} = e^{-\beta(\bar{E}(y) - T_0 S(y) + P_0 \bar{V}(y))} = e^{-\beta G_0(y)} \tag{1.6.12}$$

Here we have defined the Gibbs free energy $G_0 = \bar{E} - T_0 S - P_0 \bar{V}$, analogous to F_0. Changes occur spontaneously in a system coupled to a thermal and pressure reservoir if $\Delta G_0 \leq 0$; equilibrium holds if G_0 is at its minimum.

Euler relation

Some important thermodynamic relations follow from the fact that E, V, and N are extensive quantities (i.e., are proportional to the total system size), while properties such as T, P, and μ_i are intensive quantities (i.e., do not depend on the absolute size of the system).

Suppose we regard S as a function of E, V, and N_i; this is in some sense natural, because the microcanonical ensemble fixes the value of these three quantities. Now suppose we increase the system size by a factor of λ; all four quantities are likewise multiplied by λ. This means that

$$S(\lambda E, \lambda V, \lambda N_i) = \lambda S(E, V, N_i) \tag{1.6.13}$$

Now we take the derivative $\partial/\partial \lambda$ on both sides, and set $\lambda = 1$; the result is

$$E \left(\frac{\partial S}{\partial E} \right)_{V,N} + V \left(\frac{\partial S}{\partial V} \right)_{E,N} + \sum_i N_i \left(\frac{\partial S}{\partial N_i} \right)_{V,E} = S \tag{1.6.14}$$

Using the results of the previous section to identify the derivatives, we have

$$E + PV - \sum_i \mu_i N_i = TS \tag{1.6.15}$$

This expression bears a resemblance to the fundamental equation of thermodynamics, but is remarkable in that it relates thermodynamic potentials and not their derivatives.

From the definitions $F = E - TS$, $G = E - TS + PV$, and $A = E - TS - \sum_i \mu_i N_i$, we can rewrite the above as

$$F = \sum_i \mu_i N_i - PV \tag{1.6.16}$$

$$G = \sum_i \mu_i N_i \tag{1.6.17}$$

$$A = -PV \tag{1.6.18}$$

Finally, if we take the total derivative of the extensivity relation above, and subtract from it the fundamental equation of thermodynamics Eqn. (1.4.6), we obtain after a bit of algebra

$$0 = S dT - V dP + \sum_i N_i d\mu_i \tag{1.6.19}$$

which is called the Gibbs-Duhem equation.

The variations of T, P, and μ_i cannot be independent, because of the Gibbs phase rule of thermodynamics, which gives the number N of possible coexisting phases with C components as $N \leq C + 2$. More generally, if there are less than N coexisting phases with $N < C + 2$, we can vary $C + 2 - N$ independent intensive parameters without disturbing the coexistence. If we have only one phase, we can vary at most $C + 1$ parameters. Thus the $C + 2$ intensive parameters T, P, $\{\mu_i\}$ cannot be independent.

Multiphase ensemble

Most of the time, partition functions describe single phases, with a single typical value of the energy and other observable parameters, and fluctuations about these typical values. We say "most of the time", because boundaries between phases, on which two or more phases coexist, are special conditions that must be reached by adjusting the pressure, temperature or other external parameters.

Coexisting phases appear differently in different ensembles. In a canonical ensemble, with fixed temperature, volume, and number of particles, coexistence between phases of different density (for example, coexisting liquid and vapor phases) necessarily involves "mixed states", in which the frequently observed microstate is not homogeneous, but consists of regions of each phase, present in the same microstate. This is because the ensemble fixes the average density, which may lie within the coexistence region of the phase diagram.

A particularly convenient ensemble for describing coexistence is the grand canonical ensemble, in which temperature, volume, and chemical potentials are fixed. As a result, the number of particles can be different in different microstates, allowing the same ensemble to contain microstates of homogeneous gas and liquid phases. When these two distinct sets of microstates become equally prevalent in the ensemble, we are at the coexistence condition. With this idea, we can develop the thermodynamic conditions for equilibrium between phases directly from the grand canonical partition function, written as

$$Z_g = \sum_{E, N_i} \Omega(E, N_i) e^{-\beta(E - \sum_i \mu_i N_i)} \tag{1.6.20}$$

in which β and μ_i are the fixed inverse temperature and chemical potentials for the ensemble.

In the single phase region of parameter space, a single typical value for the energy E^* and number $\{N_i^*\}$ of particles of species i dominates the partition sum. However, at coexistence between two phases, two distinct sets of values $E^{(A)}$ and $\{N_i^{(A)}\}$, and $E^{(B)}$ and $\{N_i^{(B)}\}$, give equally dominant contributions to Z_g. This implies the two phases have equal values of $\Omega e^{-\beta(E - \sum_i \mu_i N_i)}$, i.e., equal values of the number of states times the typical Boltzmann factor for those states. Note that this does not mean that the number of states is the same for the two phases, nor the Boltzmann factors; the number of states could be larger in one phase, with smaller Boltzmann factors to compensate.

Taking the log and rearranging, coexisting phases must have equal values of the grand potential A, given by $A = E - TS - \sum_i \mu_i N_i$. Evidently, the two phases also have equal values of T and μ_i (since these are fixed by the ensemble itself). Using the extensivity relation Eqn. (1.6.18), we see that equal values of the grand potential implies equal pressures (since the volume V is fixed in the ensemble), which gives the final equilibrium condition for the two coexisting phases.

Susceptibilities and fluctuations

In Section 1.5 we worked out expressions for the average energy $\langle E \rangle$ and the variance in the energy $\langle \Delta E^2 \rangle$ in the canonical ensemble. We observed there that successive derivatives of $\log Z$ with

respect to β led to successive energy averages. In particular, we found that

$$\frac{\partial \langle E \rangle}{\partial \beta}\bigg|_V = \frac{\partial^2 \log Z}{\partial \beta^2} = \langle \Delta E^2 \rangle \tag{1.6.21}$$

We can rewrite the left side of the above equation in terms of temperature derivatives, to obtain (after a bit of algebra) the form $kT^2(\partial \bar{E}/\partial T)_V$, which we recognize as kT^2 times the constant-volume heat capacity C_V. Hence we have

$$C_V = k\beta^2 \langle \Delta E^2 \rangle \tag{1.6.22}$$

We recall from Section 1.5 that the variance $\langle \Delta E^2 \rangle$ scales as the system size, hence we note that C_V scales as the system size as it should. We may also check the dimensions of our result; β has dimensions of 1/energy, C_V goes as k, which has dimensions of energy/temperature, as required.

Eqn. (1.6.22) is a remarkable result: it relates a "response function" (C_V, the linear response of the energy to a small change in temperature) to the variance of energy fluctuations. It turns out that this is a very general type of relation, which allows us to "measure" response functions by simply observing equilibrium fluctuations (in a simulation, for example).

Number fluctuations in the grand canonical ensemble provide another example of the relation between response functions and fluctuations. We see from the definition of Z_g that

$$\langle N \rangle = \frac{1}{\beta} \frac{\partial \log Z_g}{\partial \mu} \tag{1.6.23}$$

(This is of course also consistent with thermodynamic definitions, since $Z_g = e^{-\beta A}$ and $dA = -SdT - PdV - Nd\mu$.)

Note that μ appears linearly coupled to N in the exponent of the partition summand, just as β linearly couples to E. So taking one more derivative $\partial/\partial \mu$ leads to

$$\frac{1}{\beta}\left(\frac{\partial \bar{N}}{\partial \mu}\right)_{T,V} = \frac{1}{\beta^2} \frac{\partial^2 \log Z_g}{\partial \mu^2} = \langle \Delta N^2 \rangle \tag{1.6.24}$$

Thus we find another relation of the same form as our expression relating energy fluctuations to a response function.

Here, the response function $(\partial \bar{N}/\partial \mu)_{T,V}$ is not immediately recognizable as an interesting measurable quantity. However, with a bit of rearranging, we can relate the number fluctuations to a more familiar susceptibility. The number fluctuations $\langle \Delta N^2 \rangle_{T,V,\mu}$ in the grand canonical ensemble can be written in terms of concentration fluctuations,

$$\langle \Delta N^2 \rangle_{T,V,\mu} = V^2 \langle \Delta \rho^2 \rangle_{T,V,\mu} \tag{1.6.25}$$

in which we write the particle concentration N/V as ρ.

Our ensemble at fixed V and μ can be thought of as a portion of a larger fluid, defined by some fixed portion of the volume (a "viewing region"); the fluid serves as a reservoir of particles at fixed μ. Inside the viewing region, the number density fluctuates. But we could define the viewing region differently, to contain a fixed number of particles, rather than a fixed volume. The volume would then fluctuate, and the fluid would serve as a pressure reservoir. Intuitively, if the mean size of the two viewing regions is the same, the magnitude of the density fluctuations inside the region would be the same. Hence we have

$$\langle \Delta N^2 \rangle_{T,V,\mu} = V^2 \langle \Delta \rho^2 \rangle_{T,P,N} \tag{1.6.26}$$

In this latter ensemble (a Gibbs ensemble), N is constant and V fluctuates, so that $\Delta\rho$ can be replaced by $N\Delta(1/V)$, which we expand as $-(N/V^2)\Delta V$. This leads to

$$\langle\Delta N^2\rangle_{T,V,\mu} = (N/V)^2\langle\Delta V^2\rangle_{T,P,N} \tag{1.6.27}$$

The volume fluctuations themselves can be generated by taking a derivative with respect to P, since the Boltzmann factor in the Gibbs ensemble takes the form $e^{-\beta(E+PV)}$; thus we have

$$\langle\Delta V^2\rangle_{T,P,N} = -(1/\beta)\frac{\partial\langle V\rangle}{\partial P} = (1/\beta)\kappa_T \tag{1.6.28}$$

in which we recognize the derivative on the right-hand side as related to the isothermal compressibility κ_T. Using this, we have finally an expression relating the isothermal compressibility κ_T to the number fluctuations in a grand canonical ensemble,

$$\kappa = \frac{\beta V}{N^2}\langle\Delta N^2\rangle_{T,V,\mu} \tag{1.6.29}$$

which is useful for evaluating the compressibility from an equilibrium simulation in which we fix the chemical potentials rather than the number of particles.