A selective review of methods of statistical mechanics

Kris Innanen

ABSTRACT

Elsewhere in this report we make some suggestions about the use of statistical - mechanical models to assess geophysical inverse problems. Computational difficulties appear in general when the Boltzmann theory is applied to real-world problems, with only a relatively small number of problems giving rise to summable series and closed-form solutions. So, our approach has been to start from well-known models that have been shown to be mathematically tractable, and press these into geophysical service. (If this proves to be a valuable exercise, expending the resources to solve more computationally involved versions may then be justifiable.) In this review, we set up some of the ideas and concepts of statistical mechanics, as well as the basics from which we establish and analyze models. Then, we review several relevant models, some with the aim of exemplifying the rules, and others because of their possible use in geophysics. The latter class includes especially (or so it seems at present) the molecular zipper and the law of atmospheres.

INTRODUCTION

Geophysical inversion, and optimization/estimation problems generally, continually bump up against the methods and ideas of statistical mechanics. An obvious instance of this is simulated annealing (e.g., Velis and Ulrych, 1996), in which model updates that increase the value of an objective function are permitted probabilistically, and decreasingly often, as the system is "cooled", in a process that is akin to cooling and quenching in metallurgy — or more simply, "shaking the bubble out of the crystal" (Gleick, 1992). However, Bayesian estimation methods involving sampling of a posterior probability density function (e.g., Buland and Omre, 2003) are rooted in statistical mechanics also. For instance, sampling with Metropolis-Hastings type acceptance criteria (e.g., de Figueiredo et al., 2019), and related sampling approaches, wherein a probability density function whose normalization factor is unknown is estimated, were initially developed because of the difficulties of determining the partition function of statistical models, such as the Ising model, which we will discuss in this report.

Less obvious relations between geophysical inversion and statistical mechanics are also suggested by the recent literature. For instance, the increased use of seismic full waveform inversion (FWI) methods has motivated research into quantifying FWI model uncertainties. Uncertainty estimates do not naturally emerge from descent-based optimization methods, and in the case of FWI they are badly needed. The method of *null-space shuttles* has been put forward as a partial solution to uncertainty assessment (Fichtner et al., 2021; Keating and Innanen, 2021). In the approach, model realizations are treated as shuttles, or point particles, that during updating move around in model space, like a space shuttle might in physical space. The objective function, meanwhile, acts like a gravitational potential field, attracting shuttles towards the nearest local minimum. In the same sense that the space shuttle in orbit around the Earth possesses a fixed total energy (kinetic from its motion and potential from the Earth), a model shuttle with a fixed objective function explores regions of model space which satisfy the data equally well. This exploration of what is referred to as the model null-space, then provides a means of making computable uncertainty predictions. A plausible next step, it would appear to the author, would be to increase the statistical explanatory power by treating the shuttle as an element in a 1 or many particle gas. To do this, in the context of the Hamiltonian approach of Fichtner et al. (2021), would be to bring to bear on the problem of uncertainty quantification in FWI the methods of statistical mechanics.

Problems of assessing non-uniqueness, and the importance of contributions from different grid cells in geophysical inversion, seem to be generally accessible to analysis via statistical-mechanical models. The full rationale for this will be left for a separate report. However, we can be motivated by the fact that doing, for instance, brute force calculations of the exact number of discretized slowness models that produce the same traveltime, requires the same combinatorical and statistical machinery underlying Boltzmann theory.

With all that in mind, and with the sense that research in these areas could contribute meaningfully to new geophysical inversion methods and appraisal schemes, a selective review of statistical-mechanical methods and models is presented here. At the broadest scale, the review is patterned after Susskind (2013), but in detail it involves quite a lot of excursions and additions. The early parts follow Susskind closely, especially the logical development of the Boltzmann distribution and the partition function, however the possibility of a chemical potential has been added in at the outset, as it seems to be more smoothly delivered this way. Furthermore, in the harmonic oscillator example, Susskind leaves a ground state term out of the energy levels of a quantum oscillator, producing nonstandard results, so that has been changed in order that results recognizable from other sources are produced. Several models from other sources have also been added, including the two that appear to be good starting points for geophysical applications. The barometric equation, or “law of atmospheres”, is re-derived, following the methods of Sears (1959) but translating the terms so they match with the rest of the review. Also, the “molecular zipper” model of Kittel (1969) is set out, first without and then with phase transitions as possibilities. The end of the review involves a full derivation of distributions and the partition function of the Grand Canonical Ensemble, which Susskind uses but does not develop. That derivation follows the logic presented in a different text, that of Kestin and Dorfman (1971), again with language and terminology adapted to fit in the review.

STATISTICS OF SYSTEMS THAT EXCHANGE ENERGY

Systems and ensembles

We will use the term *system* to mean some portion of space that we are paying particular attention to. “Space” here can be quite a general thing, however, so the system might be a box of gas, or a region of the atmosphere, but it might also be a strand of DNA or a grid cell in an Earth model. The system contains particles (or some general set of elements), often identical, whose motion (or behaviour in general) ultimately gives rise to the main results of the analysis. If the system is defined to be a box containing a monatomic gas, the elements would then be point-like and massive particles which are in motion in the box. The system is in contact with its *environment* (i.e., all of, or some portion of, the rest of space), and we assume that it is able to in some well-defined ways interact with it. Finally, whatever the exact nature of the exchanges between the system and its environment is, we will assume that they occur at *equilibrium*, meaning that as time evolves, they are steady and free of transient phenomena.

Each element contributes energy to the system, and changing the character of the element can be thought of as a process that costs energy to carry out. A precise specification of these costs, i.e., how an element contributes energy must be provided in order to produce a well-defined statistical-mechanical model. For instance, in a box of monatomic gas, the particles have mass and are in motion, so each one contributes kinetic energy to the system. If the particles interact with each other, the potential energy between pairs of particles also contributes. The number of degrees of freedom needed to describe the motion (and energy) of an element of the system can be very large — particles might also vibrate internally, or rotate and have angular momentum; they might interact not only in pairs but in triplets, etc. For the moment, however, it is sufficient to say that each element contributes energy, and those contributions combine to produce a well-defined total energy of the system as a whole.

A full accounting of the motion and behaviour of each element in the system at a given instant of time gives its *microstate*. We assume that while the behaviour of an individual element in the system may be quite simple to describe, the number of elements in the system is large enough that the microstate is not measureable or knowable. Instead, we envision a large number, say $\mathcal{N}$, of replicas of the system, each essentially similar, but each in a different precise microstate. Predictions about the system from statistical mechanics come from computing averages across this set of imaginary replicas, which are called *ensembles*.

It is helpful to cultivate a mental picture of systems and ensembles. One useful picture of an ensemble involves a set of systems, all arranged in a line, and each connected to its neighbours such that exchanges can occur. In this picture, one of the systems in the ensemble is the "actual" system, and the rest constitute the environment. An example is illustrated in Figure 1. $\mathcal{N}$ boxes, each containing some set of elements (e.g., atoms, or molecules), are arranged in a line. The boxes are represented as square wells plotted with energy on the vertical axis, to emphasize that for a particle to leave a vessel through the wall it must have more than a certain threshold of energy.

The boxes are connected such that two types of exchange are possible. They are thermally connected, meaning they can exchange energy, and they are chemically connected, meaning they can exchange particles. Energy exchange is represented notionally in Figure 1 by the pipes between the wells (the "E path"), and chemical exchange is represented notionally by the arrows going over the potential walls (the "N path"). This notional picture helps emphasize that for a particle or element to leave, energy greater than μ is needed.

We consider that the system and the environment are in equilibrium, meaning that transient behaviour does not occur. However, since energy and particles can be exchanged, any box at any time can have one of a variety of values of energy, and one of a variety of populations (i.e., numbers of particles). We consider that there is a list of allowable values the energy and population of a box can take on. For instance, the box labelled '3' in Figure 1 is in

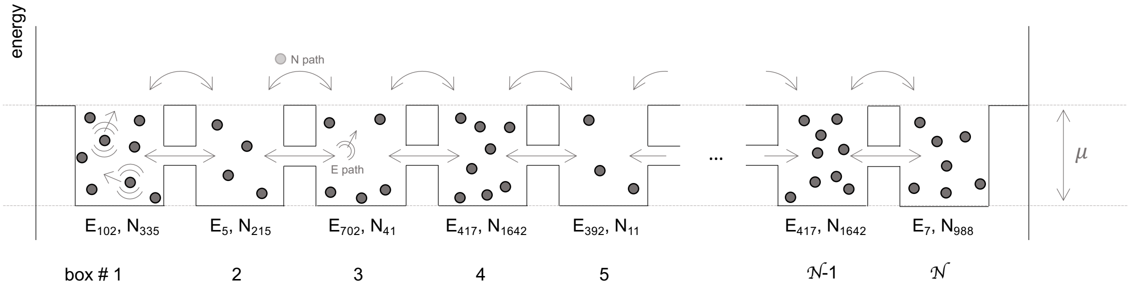

FIG. 1. Statistical-mechanical ensembles. A system (e.g., a box of gas) and its environment are envisioned as a suite of replica systems all of the same nature and in equilibrium – the system is one of the boxes and the rest constitute its environment. The boxes contain variable numbers of particles N_i and total energies E_i. In the Grand Canonical Ensemble, the system exchanges particles and energies with its environment (i.e., there are two paths for exchange, an N (number) path, over the energy barrier, and an E (energy) path, through thermal contact; in the Canonical Ensemble, the system exchanges only energy with its environment; in the Microcanonical Ensemble, the system is isolated from its environment. Averages in statistical mechanics are computed across these ensembles.

a microstate that has energy E_{702}. That is, the energy of the box is the 702nd entry on a list of available and allowable energy levels. This list has to be present at the outset, i.e., it is an input, not a prediction, of the statistical theory. The box, furthermore, contains N_{41} particles, which is, similarly, the 41st entry on a list of allowable numbers of particles[*].

This is called the grand canonical ensemble, and we will return to it near the end of this review. To begin our analysis, we will restrict the setup. We fix the number of particles to be the same in each box, and then increase μ until essentially no particle exchange is possible. Doing so we reduce the number of types of exchange between boxes to one, namely energy exchange. This is referred to as the canonical ensemble.

The canonical ensemble

Eliminating the number of particles per box as a variable, we focus our attention on energy. In general, many different microstates will have the same total energy. If we look at the example boxes in Figure 1, for instance, we see that two of the boxes have energy E_{417}. Suppose of the $\mathcal{N}$ boxes, 24 have this energy. We would then say that energy E_{417} had an *occupation number* of $n_{417} = 24$. We can imagine counting up all the n_i for $i = 1, ..,$ and retaining them in a list. This list of energies is not necessarily finite — in many models there is no real maximum number of allowable energy values. However, since there will tend to exist threshold energy values, beyond which no microstates are ever actually found, in almost any situation a practical maximum, say M, can be set without producing incorrect results.

We can use the occupation numbers $\{n_1, n_2, ..., n_M\}$ to formally define probabilities. If, of the $\mathcal{N}$ boxes, n_i have energy E_i, then we can define the probability of randomly choosing

---

[*] If the list of allowable numbers of particles is just the set of all integer numbers up to some maximum, this would just mean 41 particles.

a box with energy E_i from the ensemble as being

$$P_i = P(E_i) = \frac{n_i}{\mathcal{N}}, \tag{1}$$

which, since $\sum_{i=1}^{M} n_i = \mathcal{N}$, can be quickly confirmed to obey the rule for probabilities:

$$\sum_{i=1}^{M} P_i = 1. \tag{2}$$

So, by normalizing the list of occupation numbers, we produce a probability distribution. Given probabilities, we can also compute averages. For instance, the average energy per box, which we denote either with angle brackets, $\langle E \rangle$ or with the un-indexed letter E, is:

$$\langle E \rangle = E = \sum_{i=1}^{M} P_i E_i. \tag{3}$$

The fundamental question underlying statistical mechanics is, what probability distribution actually obtains for a box of gas? If boxes can find themselves with virtually any particular energy (apart from those above the maximum as discussed above), almost every arrangement of occupation numbers $\{n_1, n_2, ..., n_M\}$, and hence probabilities, could be discovered. Which one actually will be discovered in practice?

Boltzmann's (and Gibbs') answer to this is to observe that some sets of occupation numbers $\{n_1, n_2, ..., n_M\}$ are produced by many different microstates, and some are produced by relatively few microstates. If all microstates are equally probable, then the set of occupation numbers $\{n_1, n_2, ..., n_M\}$ we should expect to observe is that which is produced by the largest number of microstates.

Enumeration of microstates

After making this assertion, the problem becomes one of counting. We need to know how many microstates produce any given $\{n_1, n_2, ..., n_M\}$, or, equivalently, how many ways can we distribute $\mathcal{N}$ boxes into this specific set of M categories. The answer is a well-known result of combinatorics:

$$W\big(\{n_1, n_2, ..., n_M\}\big) = \frac{\mathcal{N}!}{n_1! n_2! ... n_M!}. \tag{4}$$

To motivate it, let us drop the technical use of energy for a moment and imagine instead a simpler labelling. Instead, let the $\mathcal{N}$ boxes be painted $\mathcal{N}$ different colours, and then ask how many different ways those differently-coloured boxes can be arranged. Start with $\mathcal{N} = 2$ case, and let all the boxes be different colours (say blue, b, and red, r). The boxes can be arranged in two ways:

$$\{b, r\},\ \{r, b\}. \tag{5}$$

Then go to the next higher case, $\mathcal{N} = 3$ (say with the new colour being green, g). Consider the first element in the above list, $\{b, r\}$. The new green box can appear at the front, the middle, or the back of this list:

$$\{g, b, r\},\ \{b, g, r\},\ \{b, r, g\}. \tag{6}$$

Same with the second element in the list:

$$\{g, r, b\},\ \{r, g, b\},\ \{r, b, g\}. \tag{7}$$

This process of dropping a g into every space between the r and b entries produces a complete list of the possible arrangements for $\mathcal{N} = 3$. Clearly this new list has 3×2 elements. For the $\mathcal{N} = 4$ list, we repeat this process again: we take each of the 3×2 elements, and drop another box (say orange o) at the front, in the back, and in the two internal slots:

$$\begin{aligned}
&\{o, g, b, r\},\ \{g, o, b, r\},\ \{g, b, o, r\},\ \{g, b, r, o\},\\
&\{o, b, g, r\},\ \{b, o, g, r\},\ \{b, g, o, r\},\ \{b, g, r, o\},\\
&\{o, b, r, g\},\ \{b, o, r, g\},\ \{b, r, o, g\},\ \{b, r, g, o\},\\
&\{o, g, r, b\},\ \{g, o, r, b\},\ \{g, r, o, b\},\ \{g, r, b, o\},\\
&\{o, r, g, b\},\ \{r, o, g, b\},\ \{r, g, o, b\},\ \{r, g, b, o\},\\
&\{o, r, b, g\},\ \{r, o, b, g\},\ \{r, b, o, g\},\ \{r, b, g, o\}.
\end{aligned} \tag{8}$$

This creates 4 arrangements for each of the 3×2 previous elements, for a total of $4 \times 3 \times 2$. This is enough to guess the pattern: $\mathcal{N}$ distinguishable boxes (i.e., boxes each with its own colour) can be arranged in $\mathcal{N}!$ different ways.

Now suppose the boxes are not necessarily all differently-coloured. To help adapt the formula, let us write the $\mathcal{N} = 4$ number again, this time as a fraction, with 4 instances of $1!$ formally introduced into the denominator:

$$\frac{4!}{1!1!1!1!} = \frac{(\text{total \# boxes})!}{(\text{\# red boxes})!(\text{\# blue boxes})!(\text{\# orange boxes})!(\text{\# green boxes})!}. \tag{9}$$

Suppose that instead of orange, the fourth box was blue, just like the first. To model this, we can swap all o values in (8) with b:

$$\begin{aligned}
&\{b, g, b, r\},\ \{g, b, b, r\},\ \{g, b, b, r\},\ \{g, b, r, b\},\\
&\{b, b, g, r\},\ \{b, b, g, r\},\ \{b, g, b, r\},\ \{b, g, r, b\},\\
&\{b, b, r, g\},\ \{b, b, r, g\},\ \{b, r, b, g\},\ \{b, r, g, b\},\\
&\{b, g, r, b\},\ \{g, b, r, b\},\ \{g, r, b, b\},\ \{g, r, b, b\},\\
&\{b, r, g, b\},\ \{r, b, g, b\},\ \{r, g, b, b\},\ \{r, g, b, b\},\\
&\{b, r, b, g\},\ \{r, b, b, g\},\ \{r, b, b, g\},\ \{r, b, g, b\}.
\end{aligned} \tag{10}$$

This is no longer a viable list of possibilities, because not all of the elements are distinct. In fact, hunting through the list, you can easily confirm that every element now has an identical partner (e.g., top row first column and second row third column). The number of distinct possibilities has, in other words, been cut in half:

$$\frac{1}{2}\frac{4!}{1!1!1!1!}=\frac{4!}{1!2!1!}=\frac{(\text{total \# boxes})!}{(\text{\# red boxes})!(\text{\# blue boxes})!(\text{\# green boxes})!}. \tag{11}$$

We observe that the right number can be gotten by removing one of the $1!$ from the denominator of the original formula, and changing another to $2!$. There are now only 3 groups of possible colours, red, green and blue, and there are 2 blues and 1 each of red and green. It suggests a more general pattern: divide $\mathcal{N}!$, the factorial of the total number of boxes, by the factorials of the numbers of boxes of each possible colour. We won't include bigger examples here, but a few of these should be enough to convince. Equation (4), of course, follows just that pattern, with energy re-introduced instead of colour.

Stirling's approximation and entropy

Now that we know how many microstates contribute to a given set of occupation numbers (or probabilities), we can start to ask the question we wanted to ask, namely, which set of occupation numbers is produced by the largest number of microstates. We will accomplish this in a manner very familiar to geophysicists, by solving a constrained optimization problem, in which we ask for the set $\{n_1, n_2, ..., n_M\}$, or the associated $\{P_1, P_2, ..., P_M\}$, for which both (1) W is a maximum and (2) the P_i are well-behaved probabilities.

Optimization problems are more convenient to solve if they do not involve unknowns occurring in a product, or factorials of unknowns. Let us start by adjusting the factorial aspect of the formula. Suppose we wish to evaluate $\mathcal{N}!$:

$$\mathcal{N}! = \mathcal{N}\times(\mathcal{N}-1)\times(\mathcal{N}-2)...\times 2\times 1. \tag{12}$$

Taking logarithms,

$$\log\mathcal{N}! = \log 1+\log 2+...+\log(\mathcal{N}-1)+\log\mathcal{N} = \sum_{x=1}^{\mathcal{N}}\log x. \tag{13}$$

As $\mathcal{N}$ becomes large, the distance between integers shrinks relative to it. If we treat the space between integers as infinitesimal, the sum in (13) becomes an integral that is easy to solve:

$$\sum_{x=1}^{\mathcal{N}}\log x \approx \int_1^{\mathcal{N}} dx\log x \approx \Big[x\log x - x\Big]_1^{\mathcal{N}} \approx \mathcal{N}\log\mathcal{N}-\mathcal{N}. \tag{14}$$

If you work through this integral, you will see that in forming the approximation we have also assumed that $\mathcal{N}$ is large enough that we can neglect the influence of an extra 1 that appears in the integral. Re-exponentiating, we have

$$\mathcal{N}! \approx e^{\mathcal{N}\log\mathcal{N}}e^{-\mathcal{N}} \approx \left(e^{\log\mathcal{N}}\right)^{\mathcal{N}} e^{-\mathcal{N}} \approx \mathcal{N}^{\mathcal{N}}e^{-\mathcal{N}}. \tag{15}$$

This is called Stirling's approximation (*approximation* because the space between integers is not infinitesimal). It dramatically simplifies the calculation of a factorial without much cost in terms of accuracy. For instance, with $\mathcal{N} = 100$, Stirling's approximation is roughly 4×10^{156}, and the actual factorial is about 25 times larger; the error in the logarithms of these numbers is therefore on the order of only about a percent.

We can then re-express W using Stirling's approximation:

$$W = \frac{\mathcal{N}!}{n_1!n_2!...n_M!} \approx \frac{\mathcal{N}^{\mathcal{N}} e^{\mathcal{N}}}{n_1^{n_1} e^{n_1} n_2^{n_2} e^{n_2} ... n_M^{n_M} e^{n_M}} \approx \frac{\mathcal{N}^{\mathcal{N}}}{n_1^{n_1} n_2^{n_2} ... n_M^{n_M}}. \tag{16}$$

Now the factorials are gone, which is convenient when we contemplate optimization. The multiplicative operations are still here, but we can drop them also by taking logarithms — fortunately, the argument that minimizes a scalar f is the same argument that minimizes $\log f$ in general, so it does not change our problem to consider $\log W$ instead of W. Recalling the rules $\log x^a = a \log x$ and $\log x/y = \log x - \log y$, we have

$$\log W = \mathcal{N} \log \mathcal{N} - n_1 \log n_1 - ... - n_M \log n_M = \mathcal{N} \log \mathcal{N} - \sum_{i=1}^{M} n_i \log n_i. \tag{17}$$

Let us lastly transform from the occupation numbers n_i to probabilities $P_i = n_i/\mathcal{N}$. Then

$$\begin{aligned} \log W &= \mathcal{N} \log \mathcal{N} - \sum_{i=1}^{M} (\mathcal{N} P_i) \log(\mathcal{N} P_i) \\ &= \mathcal{N} \log \mathcal{N} - \mathcal{N} \sum_{i=1}^{M} P_i \Big(\log \mathcal{N} + \log P_i \Big) \\ &= \mathcal{N} \log \mathcal{N} - \mathcal{N} \log \mathcal{N} \sum_{i=1}^{M} P_i - \mathcal{N} \sum_{i=1}^{M} P_i \log P_i \\ &= -\mathcal{N} \sum_{i=1}^{M} P_i \log P_i, \end{aligned} \tag{18}$$

where in the last line we have used (2). From this, we define $\log W = \mathcal{N} S$, where S is referred to as the entropy:

$$S = -\sum_{i=1}^{M} P_i \log P_i. \tag{19}$$

Since $\mathcal{N}$ is fixed, the problem of finding the $\{n_1, n_2, ..., n_M\}$ for which W is a maximum is seen to be identical to the problem of finding the P_i values for which S is a maximum, or (as we will frame it), the P_i values for which $(-S)$ is a minimum.

The method of Lagrange multipliers

At this stage the problem is nearly formulated – we set the derivatives of $(-S)$ with respect to all $\{P_1, P_2, ..., P_M\}$ to zero and solve. However, if we do this directly we will get into trouble, because we will likely arrive at an illicit set of P_i values — ones which sum to a number different from 1, or which fail to produce correct average energy values. In order to properly solve this problem, we have to enforce (2) and (3). The full problem is in fact to minimize

$$\sum_{i=1}^{M} P_i \log P_i \tag{20}$$

with respect to the P_i, subject to

$$\sum_{i=1}^{M} P_i - 1 = 0 \;\text{ and }\; \sum_{i=1}^{M} P_i E_i - \langle E \rangle = 0. \tag{21}$$

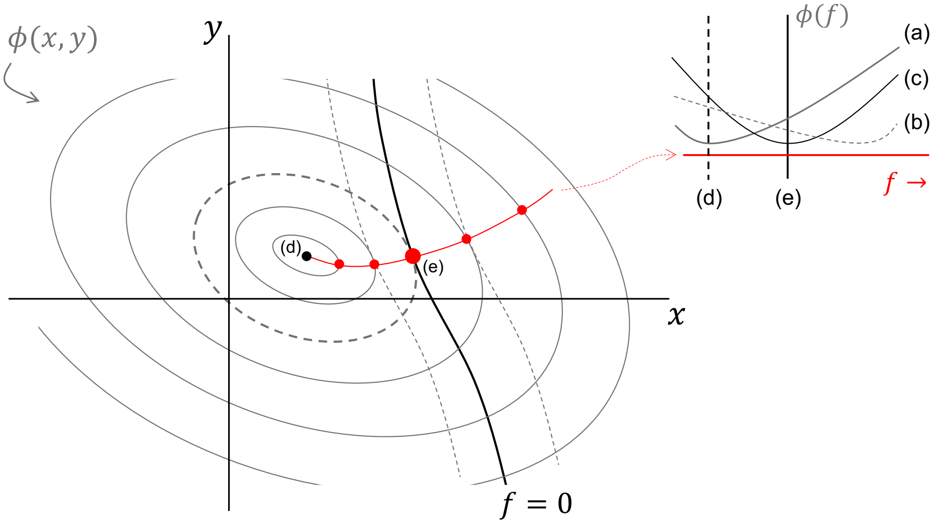

FIG. 2. A toy two-variable constrained minimization problem. Contours of a merit or objective function ϕ are plotted against two variables x and y. The minimum is at point (d). We are interested not in this (d), the global minimum of ϕ, but rather the minimum (e) that lies on the curve $f(x,y) = 0$, which is a curve embodying a constraint. We consider ϕ varying along a path passing through (d) and (e) which is everywhere perpendicular to the contours (red curve). We plot this $\phi(f)$ versus f in the upper right corner. It appears as the grey curve (a). The minimum of ϕ can be shifted arbitrarily, e.g., to (b), by adding a term that is linear in the independent variable. The method of Lagrange multipliers involves minimizing an adjusted version of ϕ, whose minimum has been shifted such that it coincides with $f = 0$. This is (c).

To solve this we will employ the method of Lagrange multipliers. The method is used quite regularly in geophysical inversion, but it is normally introduced very formally, and without intuition, so we will expand on the idea here for a moment. Suppose we have a two-variable problem, i.e., we wish to solve for the x and y that minimizes a function $\phi(x,y)$. The contours of ϕ, which are illustrated in Figure 2, make clear more or less where the minimum is – we mark it with a black dot and label it (d). Now we add to the problem the constraint $f(x,y) = 0$. Such an equation describes a curve on the x-y plane†. An

---

† For instance, if $f(x,y) = y - mx - b$, the curve is a line with slope m and intercept b.

example constraint curve is illustrated as a bold line in Figure 2. The constrained problem:

$$\min_{x,y} \phi, \quad \text{subject to} \quad f(x,y) = 0, \tag{22}$$

amounts to finding the minimum of ϕ lying along the $f = 0$ curve (the large red dot labelled (e) in Figure 2). The approach of Lagrange is to ask if a new function $\mathcal{L}$ can be created, whose minimum is guaranteed to be on the large red dot. The answer is yes. To find it, the only extra piece of information we have to fold in is a geometric fact: at (e), the curve $f = 0$ is tangent to the contour of ϕ.

Let us imagine that we will be able to adjust ϕ such that its minimum follows a path starting at (d) and ending at (e). This adjusted quantity can then be minimized to give us the answer we seek. Let us define the path to be that which starts at (d), ends at (e), and is everywhere perpendicular to the contours of ϕ. This path is sketched in red in Figure 2. Since $f = 0$ is tangent to the contour, the new path is perpendicular to it at (e), by definition. In fact, it defines a family of curves $f = ..., -1, 0, 1, 2, ...$, amongst which the actual constraint $f = 0$ is just one example. The trick is to allow this set of points to become a new independent coordinate axis, along which we can examine ϕ, i.e., we examine $\phi(f)$. The $\phi(f)$ curve is illustrated in the top right of Figure 2 (the grey curve labelled a). The minimum (d) is visible in the region of negative f, since it can be seen to be to the left of the constraint $f = 0$; the value of $\phi(f = 0)$ is also marked (e). The plan, then, is to adjust ϕ such that, after adjustment, the minimum has moved along this red curve so that it sits at (e), satisfying the constraint.

Any univariate curve with an independent coordinate (say z) can have its minimum shifted by adding a term proportional to z[‡]. In our case, the independent variable is f, so we accomplish the shift by adjusting ϕ as follows:

$$\mathcal{L}(f) = \phi(f) + \lambda f. \tag{23}$$

By selecting λ, we can shift the minimum arbitrarily (e.g., (b) in the top right of the Figure). Amongst these possibilities for λ, the useful one is the one that moves the minimum to $f = 0$, because this makes $\mathcal{L}$ a minimum on the constraint line (i.e., (c) in the top right of the Figure).

Assuming that this special λ can be found, we can now proceed by minimizing $\mathcal{L}$ instead of ϕ. In this example that produces two equations:

$$\frac{\partial \mathcal{L}}{\partial x} = 0, \quad \frac{\partial \mathcal{L}}{\partial y} = 0, \tag{24}$$

[‡] Take the parabola $h(z) = (z - z_0)^2$, whose minimum is at z_0. Add a term that is proportional to z:

$$h'(z) = h(z) + 2az = \left[z - (z_0 - a)\right]^2 + c.$$

Completing the square we find that the new curve has had its vertex moved by a distance a. Notice that amongst the possible a values is one, $a = z_0$, which moves the vertex to $z = 0$.

or

$$\frac{\partial \phi(x,y)}{\partial x} + \lambda \frac{\partial f}{\partial x} = 0, \quad \frac{\partial \phi(x,y)}{\partial y} + \lambda \frac{\partial f}{\partial y} = 0. \tag{25}$$

Now we just need to make sure that the λ in these equations is chosen such that the constraint is satisfied, making sure choice (c) is made rather than, say, (b), in the top right of Figure 2. The final bit of elegance in Lagrange's method is that, because λ appears in the adjusted function $\mathcal{L}$ as a factor multiplying the constraint equation, by adding the auxiliary equation $\partial\mathcal{L}/\partial\lambda = 0$, we automatically force λ to be chosen such that the constraint holds:

$$\frac{\partial \mathcal{L}}{\partial \lambda} = \frac{\partial}{\partial \lambda}\Big(\phi(f) + \lambda f\Big) = f = 0. \tag{26}$$

The Boltzmann distribution

The method of Lagrange multipliers can be applied to the problem of determining the probabilities associated with the canonical ensemble as follows. We set up the augmented functional

$$\mathcal{L} = \sum_{i=1}^{M} P_i \log P_i + \alpha\Big(\sum_{i=1}^{M} P_i - 1\Big) + \beta\Big(\sum_{i=1}^{M} E_i P_i - \langle E\rangle\Big), \tag{27}$$

having introduced two multipliers α and β. The primary equations for the P_i are then

$$\frac{\partial \mathcal{L}}{\partial P_i} = \log P_i + 1 + \alpha + \beta E_i = 0. \tag{28}$$

Solving for P_i, we obtain $P_i = e^{-1-\alpha} e^{-\beta E_i}$, or

$$P_i = \frac{1}{Z} e^{-\beta E_i}, \tag{29}$$

for convenience setting $Z = e^{1+\alpha}$ rather than using the more cumbersome exponential. We then must determine β and Z (i.e., α), having seen that the constrained minimum will have been properly reached as long as they are computed by enforcing the constraints. Z is determined by enforcing $\sum_{i=1}^{M} P_i = 1$:

$$\sum_{i=1}^{M} P_i = \frac{1}{Z} \sum_{i=1}^{M} e^{-\beta E_i} = 1, \tag{30}$$

therefore

$$Z(\beta) = \sum_{i=1}^{M} e^{-\beta E_i}. \tag{31}$$

The form we have found for P_i in equation (29) is called the *Boltzmann distribution*, or *Boltzmann factor*; the normalization $Z(\beta)$ in (31) is called the *partition function*; it can be seen in (31) to be a function of the second multiplier β, which we still need to determine.

Calculations with Z and the meaning of β

We will approach the interpretation of β slowly. First, let us substitute the Boltzmann factor into the second constraint equation:

$$\langle E\rangle = \sum_{i=1}^{M} E_i P_i = \frac{1}{Z}\sum_{i=1}^{M} E_i e^{-\beta E_i}. \tag{32}$$

Notice that if we take the derivative of Z with respect to β, we obtain

$$\frac{\partial Z(\beta)}{\partial \beta} = \sum_{i=1}^{M} \frac{\partial}{\partial \beta} e^{-\beta E_i} = -\sum_{i=1}^{M} E_i e^{-\beta E_i}, \tag{33}$$

which means the average energy in (32) can be written in terms of Z in two useful ways:

$$\langle E\rangle = -\frac{1}{Z(\beta)}\frac{\partial Z(\beta)}{\partial \beta} = -\frac{\partial \log Z(\beta)}{\partial \beta}. \tag{34}$$

It is likewise straightforward to develop further the form of the entropy S in terms of Z:

$$\begin{aligned} S &= -\sum_{i=1}^{M} P_i \log P_i \\ &= -\sum_{i=1}^{M} \frac{e^{-\beta E_i}}{Z(\beta)} \log\left(\frac{e^{-\beta E_i}}{Z(\beta)}\right) \\ &= \frac{\beta}{Z(\beta)} \sum_{i=1}^{m} e^{-\beta E_i} E_i + \frac{\log Z(\beta)}{Z(\beta)} \sum_{i=1}^{m} e^{-\beta E_i} \\ &= \beta\langle E\rangle + \log Z(\beta), \end{aligned} \tag{35}$$

where in the last line we have recognized the average energy in the first term, and the fact that the sum in the second term reproduces Z, which is also in the denominator.

To interpret β, let us compare (35) to an important result from classical thermodynamics. Switching (as we will regularly in this review) from the angle-bracket form $\langle E\rangle$ to the un-indexed form for the average energy E, the entropy then has the form

$$S = \beta E + \log Z(\beta). \tag{36}$$

Since S is a function of both E and β, its differential is

$$\begin{aligned} dS &= \left(\frac{\partial S}{\partial E}\right) dE + \left(\frac{\partial S}{\partial \beta}\right) d\beta \\ &= \beta dE + \left(E + \frac{\log Z(\beta)}{Z(\beta)}\right) d\beta, \end{aligned} \tag{37}$$

using (36) in the second line. But, the term in brackets vanishes, because of (34), leaving

$$dS = \beta dE. \tag{38}$$

This is a mathematical form that permits interpretation of β. We have not said anything yet about what kinds of behaviours give our systems this energy value or that energy value, or even what we want energy to be. Whatever the energy is, dE is a representation of a small change in its average value. The multiplier β is thus seen to be a coefficient telling how large of a change dS is enacted on the entropy, per (19), when the average energy is incremented by dE. In the abstract this is as far as we can go.

Classical thermodynamics can help us take the interpretation further, because it involves more concretely defined thermal systems, in which energy and entropy are pre-defined. Also, it describes specific experiments, involving heating and cooling gases with flames, etc. In these classical descriptions, heat flow into system via for instance a flame produces very concrete changes in energy, dE, and in those descriptions the thermodynamic entropy S changes by dS in the presence of such a dE by the amount

$$dS = \frac{dE}{T}, \tag{39}$$

i.e., it defines the change in entropy as the change in energy divided by the temperature. Comparison of (38) and (39) gives rise to the interpretation of β as the *reciprocal temperature*. This $T = 1/\beta$ deriving from the statistical description will, of course, correspond exactly to the classical temperature in the end, but only after the energies we have so far been discussing abstractly are properly specified. This abstractness is again important for us to maintain, because our later use of statistical mechanics will involve very different definitions of the energy; this in turn will change the interpretation of the temperature in (38).

Fluctuations

Any average quantity in the ensemble can be computed by

$$\langle Q \rangle = \sum_{i=1}^{m} P_i Q_i. \tag{40}$$

We used the formula to compute the average energy $\langle E \rangle$, but we can use it to compute average values of functions of E_i as well, for instance

$$\left\langle E^2 \right\rangle = \sum_{i=1}^{m} P_i E_i^2, \tag{41}$$

or

$$\left\langle \big(E - \langle E \rangle\big)^2 \right\rangle = \sum_{i=1}^{m} P_i \big(E_i - \langle E \rangle\big)^2. \tag{42}$$

This last expression is the square of the average of the difference between the E_i and $\langle E \rangle$, which is recognizable to us as the variance. In statistical mechanics it is denoted ΔE^2. The angle brackets comprise a linear operator, which means we can use simple algebra on quantites involving them. For instance, expanding $(E_i - \langle E \rangle)^2 = E_i^2 - 2E_i \langle E \rangle + \langle E \rangle^2$, we obtain

$$\Delta E^2 = \sum_{i=1}^{m} P_i E_i^2 - 2 \langle E \rangle \sum_{i=1}^{m} P_i E_i + \langle E \rangle^2 \sum_{i=1}^{m} P_i = \left\langle E^2 \right\rangle - \langle E \rangle^2 , \tag{43}$$

again using (2) and (3). The term *fluctuation* is often employed in statistical mechanics to discuss variances within the ensemble. Fluctuations are another example of a physical quantity that can be derived when the partition function $Z(\beta)$ is available. Returning to our results in (33), we observe that they can be extended to higher derivatives:

$$\frac{\partial^n Z(\beta)}{\partial \beta^n} = \sum_{i=1}^{m} e^{-\beta E_i} (-E_i)^n, \tag{44}$$

so

$$\frac{\partial Z(\beta)}{\partial \beta} = - \sum_{i=1}^{m} e^{-\beta E_i} E_i, \ \ \frac{\partial^2 Z(\beta)}{\partial \beta^2} = \sum_{i=1}^{m} e^{-\beta E_i} E_i^2, \tag{45}$$

from which ΔE^2 becomes

$$\Delta E^2 = \left\langle E^2 \right\rangle - \langle E \rangle^2 = \frac{1}{Z(\beta)} \frac{\partial^2 Z(\beta)}{\partial \beta^2} - \left(\frac{1}{Z(\beta)} \frac{\partial Z(\beta)}{\partial \beta} \right)^2 . \tag{46}$$

But

$$\begin{aligned} \frac{\partial^2 \log Z(\beta)}{\partial \beta^2} &= \frac{\partial}{\partial \beta} \left(\frac{1}{Z(\beta)} \frac{\partial Z(\beta)}{\partial \beta} \right) \\ &= \left(- \frac{1}{Z^2(\beta)} \frac{\partial Z(\beta)}{\partial \beta} \right) \frac{\partial Z(\beta)}{\partial \beta} + \frac{1}{Z(\beta)} \frac{\partial^2 Z(\beta)}{\partial \beta^2}, \end{aligned} \tag{47}$$

so in fact

$$\Delta E^2 = \left\langle E^2 \right\rangle - \langle E \rangle^2 = \frac{\partial^2 \log Z(\beta)}{\partial \beta^2} = - \frac{\partial}{\partial \beta} \langle E \rangle , \tag{48}$$

where in the last step we have recognized the original expression for the average energy.

MODELS BASED ON THE CANONICAL ENSEMBLE

The general calculation scheme in the previous section produces specific and interpretable results when the energies E_i are determined, at which point the partition function can be evaluated, and the behaviour of the system be predicted. In this section we will review some of the famous and rare cases in which the energies in an important problem produce a partition function that is straightforward to compute.

Ideal gas

Let the system under study be a rectangular box with volume $V = l_1 \times l_2 \times l_3$, filled with a dilute, monatomic gas of N atoms, each of mass m. Let the first of these atoms (Figure 3a) be labelled with position coordinates (x_1, x_2, x_3) and momentum components (p_1, p_2, p_3). Then let the second atom have position components (x_4, x_5, x_6) and momentum components (p_4, p_5, p_6). Continuing similarly, full lists of the position and momentum components of the system, $\{x_1, x_2, ..., x_{3N}\}$ and $\{p_1, p_2, ..., p_{3N}\}$, are produced, where the triplets $(x_{3k-2}, x_{3k-1}, x_{3k})$ and $(p_{3k-2}, p_{3k-1}, p_{3k})$ respectively represent the position and momentum components of the kth particle.

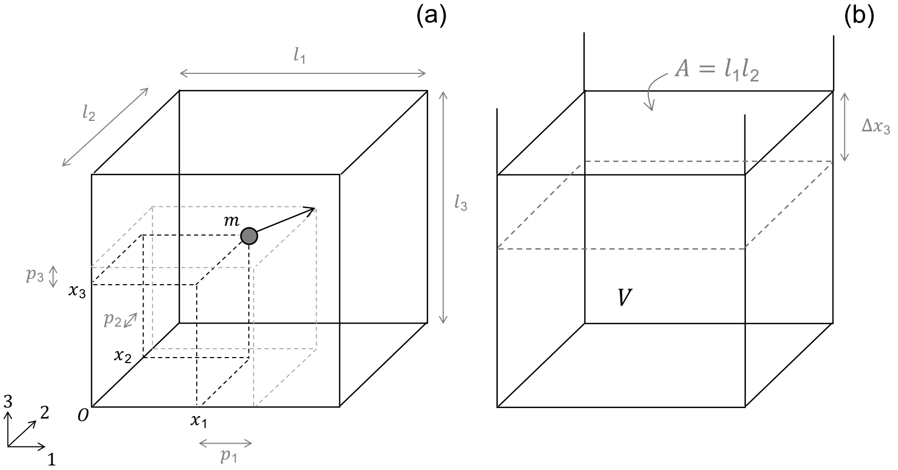

FIG. 3. (a) One of N atoms making up a dilute, monatomic gas occupying a box of volume $V = l_1 l_2 l_3$, and its position and momentum components. (b) The box as a column with piston; the top side of the box can move up and down when subject to a force, changing the box volume.

Atoms contribute to the total energy of a system through the kinetic energy of their motion, and any potential energy they produce or experience. The contribution of the kth atom to the total kinetic energy of the gas includes that due to its three momentum components, and any other generalized momenta associated with its internal vibrations, rotations, etc. Here, because we are assuming a monatomic gas, the atoms are point-like, meaning there are no such extra degrees of freedom, and the kinetic energy of the kth atom is[§]

$$\frac{p_{3k-2}^2}{2m} + \frac{p_{3k-1}^2}{2m} + \frac{p_{3k}^2}{2m}. \tag{49}$$

The kth atom can also contribute potential energy to the total energy of the gas. The energy may originate externally, e.g., if the gas is in a gravitational field, or it may originate from interactions between the atoms, e.g., when they collide, or otherwise exert forces on each

---

[§] We will use the momentum form of kinetic energy. Since $p = mv$,

$$\frac{1}{2}mv^2 = \frac{1}{2}m\left(\frac{p}{m}\right)^2 = \frac{p^2}{2m}.$$

The velocity form $mv^2/2$ can be used as well, in order to determine velocity distributions, results which are associated with Maxwell. We will not examine those in this review.

other. The potential energy contributed by an atom is generally a function of its position and those of all the others:

$$V(..., x_{3k-2}, x_{3k-1}, x_{3k}, ...). \tag{50}$$

We can now obtain the partition function for this problem. Formally, it is

$$Z(\beta) = \sum_i e^{-\beta E_i}, \tag{51}$$

where the sum is over all possible energy values accessible to the system. The problem then is to realize this sum given the specifics of the gas. In principle each different value of each of the $3N$ coordinates describing the gas produces its own energy value, so the sum when realized must be over every possible configuration of the gas. This suggests that the sum is an integral over the ranges of possible values of the $3N$ position values of the atoms and their $3N$ momenta:

$$\sum_i \rightarrow \frac{1}{N!} \int_0^{l_1} dx_1 ... \int_0^{l_3} dx_{3N} \int_{-\infty}^{\infty} dp_1 ... \int_{-\infty}^{\infty} dp_{3N}, \tag{52}$$

in which the spatial integrals are limited by the dimensions of the box; the momenta are not restricted.

The factor $1/N!$ is included because the integrals alone are only valid if each of the particles is distinct from all others. For example, somewhere amongst all of the configurations being counted in (52), there is a configuration in which some particle A has momentum/position p_A/x_A and some particle B has momentum/position p_B/x_B; but, elsewhere in the counting there is also a configuration in which particle B has momentum/position p_A/x_A and particle A has momentum/position p_B/x_B. If the system is indifferent to the exchange of particles A and B, then (52) evidently accounts for the same configuration twice. If only particles A and B are indistinguishable, we would correct for this by dividing the integral in (52) by 2. But all N particles are interchangeable, so the integral in (52) in fact over-counts by a factor of $N!$, i.e., the number of possible exchanges of particles. We compensate for this by dividing by $N!$.

The energy of a particular configuration is the sum of all the contributions to the kinetic and potential energy from each particle:

$$E(x_1, ..., x_{3N}, p_1, ..., p_{3N}) = \frac{p_1^2}{2m} + ... + \frac{p_{3N}^2}{2m} + ... + V(..., x_{3k-2}, x_{3k-1}, x_{3k}, ...) + ...,$$

where the V contributions include a term for each of the N particles that itself depends on the positions of all other particles. These terms can make calculations quite complicated, but fortunately important cases can be explored in which they are neglected. It is straightforward to set up a problem in which there are no external forces[¶], but all particles in nature

---

[¶] We simply consider gases made up of inert particles. The only force we cannot reasonably eliminate this way is gravity, since we have given the particles mass, but gravity can be neglected on the basis of its relative weakness.

interact with each other, at least at close range when they collide, which makes neglecting these a possible problem. However, since the gas we are currently describing is dilute, the fact that particles do interact at close range can itself be neglected, since it is rare for a pair of particles to be close enough to contribute significantly. With this assumption in place, the energy becomes

$$E(x_1, ..., x_{3N}, p_1, ..., p_{3N}) = \frac{p_1^2}{2m} + ... + \frac{p_{3N}^2}{2m}. \tag{53}$$

The partition function for this box of gas is therefore

$$Z(\beta) = \frac{1}{N!} \int_0^{l_1} dx_1 ... \int_0^{l_3} dx_{3N} \int_{-\infty}^{\infty} dp_1 ... \int_{-\infty}^{\infty} dp_{3N} \exp\left[-\beta\left(\frac{p_1^2}{2m} + ... + \frac{p_{3N}^2}{2m}\right)\right].$$

We observe that because the potential energy is neglected in this model, the integrand is not a function of any of the spatial integration variables. Each triplet of spatial integrals therefore gives the volume of the box. For instance, the first triplet gives:

$$\int_0^{l_1} dx_1 \int_0^{l_2} dx_2 \int_0^{l_3} dx_3 = V. \tag{54}$$

There are N of these triplets, so the space integrals in total contribute a factor of V^N. We further observe that each term in the argument of the exponential pairs with one of the momentum integral signs:

$$Z(\beta) = \frac{V^N}{N!} \left(\int_{-\infty}^{\infty} dp_1 e^{-\beta p_1^2/2m}\right) \times ... \times \left(\int_{-\infty}^{\infty} dp_{3N} e^{-\beta p_{3N}^2/2m}\right). \tag{55}$$

The momentum integrals therefore amount to $3N$ independent and identical integrals. Using $\int_{-\infty}^{\infty} e^{-ax^2} = \sqrt{\pi/a}$, we obtain

$$Z(\beta) = \frac{V^N}{N!} \left(\frac{2\pi m}{\beta}\right)^{3N/2}. \tag{56}$$

With a definite form for the partition function we may then rapidly compute features of the box of gas. The average energy, using (34), and again reverting to the non-bracketed form $E = \langle E \rangle$, is

$$E = -\frac{\partial Z(\beta)}{\partial \beta} = \frac{3N}{2} T, \tag{57}$$

or $T/2$ for each of the $3N$ degrees of freedom in the gas. (To derive this result, use the rules for logarithms of products and quotients, and separate out the term in β.) The entropy, being

$$S = \frac{E}{T} + \log Z(\beta), \tag{58}$$

is also now immediately computable:

$$S = \frac{3N}{2}\big(1 - \log\beta\big) + C, \tag{59}$$

where C is independent of β.

Probabilities and single-particle gases

Analysis of Z alone gives us statistical features of the gas as a whole, but we have to return to the full distribution to learn about how the atoms are arranged in space. We do not know, for instance, what the "number density" of particles as a function of x_1, x_2, and x_3 is for the gas. We can guess: this is an ideal gas, subject to no external forces, so a fairly trivial uniform distribution is expected.

To extract this quantity from the formalism, let us re-consider the box in Figure 3a. Let us ask on the backdrop of the Boltzmann theory what the probability $P(x_1, x_2, x_3, p_1, p_2, p_3, \beta)$ is — i.e., the probability of finding an atom at (x_1, x_2, x_3) and with momentum (p_1, p_2, p_3)[‖]. According to the Boltzmann theory, this probability is

$$P(x_1, x_2, x_3, p_1, p_2, p_3, \beta) = \frac{1}{Z(\beta)} \exp\left[-\beta\left(\frac{p_1^2}{2m} + \frac{p_2^2}{2m} + \frac{p_2^2}{2m}\right)\right], \tag{60}$$

where in this case the partition function is the sum over all possible energies of the individual particle:

$$Z(\beta) = \int_0^{l_1}\int_0^{l_2}\int_0^{l_3} dx_1 dx_2 dx_3 \int\int\int dp_1 dp_2 dp_3 \exp\left[-\beta\left(\frac{p_1^2}{2m} + \frac{p_2^2}{2m} + \frac{p_2^2}{2m}\right)\right].$$

Already we observe "uniformity" emerging – because the energy is not a function of position, the probability has no dependence on the spatial variables. Let us keep the erstwhile spatial dependence on the left hand side of (60) in any case, because this lack of dependence is emerging from a calculation that would have given the dependence if it had been there. In the partition function, we recognize the spatial integrals as again determining the volume of the box, and the momentum integrals as three uncoupled Gaussian integrals, we obtain

$$Z(\beta) = V\left(\frac{2\pi m}{\beta}\right)^{3/2}, \tag{61}$$

so in total we have

$$P(x_1, x_2, x_3, p_1, p_2, p_3, \beta) = \frac{1}{V}\left(\frac{2\pi m}{\beta}\right)^{-3/2} \exp\left[-\beta\left(\frac{p_1^2}{2m} + \frac{p_2^2}{2m} + \frac{p_2^2}{2m}\right)\right]. \tag{62}$$

From this we can obtain the probability associated with a particle at (x_1, x_2, x_3) for *any* momentum, by integrating over all (p_1, p_2, p_3). This means evaluating the same three momentum integrals once again, producing

$$\begin{aligned} P(x_1, x_2, x_3, \beta) &= \int\int\int dp_1 dp_2 dp_3 P(x_1, x_2, x_3, p_1, p_2, p_3, \beta) \\ &= \frac{1}{V}\left(\frac{2\pi m}{\beta}\right)^{-3/2}\left(\frac{2\pi m}{\beta}\right)^{3/2} = \frac{1}{V}. \end{aligned} \tag{63}$$

[‖] This sounds wrong at first blush — how can we determine statistics for *one* particle? However, we recall that the statistics are built up over the *ensemble*, not over a physical group of atoms. An ensemble consisting of many repetitions of a 1-particle system is just as good for statistics as an ensemble of systems of many particles.

We can then apply the results produced by considering a single particle back to the larger set. For a box with N particles, we set the number of particles per unit volume, n, at (x_1, x_2, x_3), as (N, the number of particles overall)$\times P$, or

$$NP(x_1, x_2, x_3, \beta) = \frac{N}{V}. \tag{64}$$

This is a constant corresponding to the ratio of N to the total volume of the box V, which confirms our expectations. To be sure, this is a simple — almost trivial — result. However, it sets out an approach that can be applied in more complex situations.

Helmholtz free energy, pressure, and the equation of state

In classical thermodynamics, the difference between the average energy and the product of the entropy and the temperature is called the *Helmholtz free energy* F. It is a measure of the energy available to the system to do work. To see this, first consider how work affects the box (warning: we have to introduce some thermodynamical quantities here, which we are not reviewing completely). We imagine that the top of the box is actually a piston, which can move up or down in a rectangular column (Figure 3b). Work is done when this happens. Specifically, an element of work dW is given by the formula**

$$dW = pdV = -dE, \tag{65}$$

where in the rightmost term we have equated dW with a negative change in the average energy of the gas — conventionally, if the gas does positive work, it does so at the expense of its total energy. Equation (65) holds only if *all* of the work went into changing the energy of the gas, or conversely if the whole change in the energy of the gas went into the work of moving the piston. If some energy was lost in the process, say through dissipation of heat from the box into its environment, $dW = -dE$ no longer holds. So, to use it, we have to add to (65) that the change dV occurred *isentropically*, with no change in entropy, which is another way of saying it happened without such losses.

This and the righthand equality in (65) suggest we express pressure as a derivative based on (65):

$$p = -\left.\frac{\partial E}{\partial V}\right|_S, \tag{66}$$

where the $\cdot|_S$ enforces the isentropic aspect of the process. In these circumstances E is determined fully by temperature T and volume V, so another way of writing dE presents

---

**Work is the product of distance an object moves and the net force it experiences as it moves, i.e, $dW = Fdx$. In our box, a force exerted by the gas moves the piston in the $+x_3$ direction. So if $A = l_1 l_2$ is the area of the top of the box, the work done by the gas is

$$dW = Fdx_3 = \frac{F}{A} \times Adx_3 = p \times dV,$$

where $p = F/A$ is the pressure exerted by the gas, and $dV = Adx_3$ is the change in volume.

itself, via the rule for differentials:

$$\begin{aligned} dE &= \left.\frac{\partial E}{\partial V}\right|_T dV + \left.\frac{\partial E}{\partial T}\right|_V dT \\ &= \left.\frac{\partial E}{\partial V}\right|_T dV + \left.\frac{\partial E}{\partial S}\right|_V \left.\frac{\partial S}{\partial T}\right|_V dT, \end{aligned} \tag{67}$$

or, dividing through by dV:

$$\frac{dE}{dV} = \left.\frac{\partial E}{\partial V}\right|_T + \left.\frac{\partial E}{\partial S}\right|_V \left.\frac{\partial S}{\partial T}\right|_V \frac{dT}{dV} = \left.\frac{\partial E}{\partial V}\right|_S . \tag{68}$$

The left side is an allowable form for dE/dV since the process is already constrained to occur at fixed S. This can be simplified, because in an isentropic process $dS = 0$, so, if it too is a function of T and V only,

$$dS = \left.\frac{\partial S}{\partial V}\right|_T dV + \left.\frac{\partial S}{\partial T}\right|_V dT = 0, \tag{69}$$

or

$$\left.\frac{\partial S}{\partial V}\right|_T dV = -\left.\frac{\partial S}{\partial T}\right|_V dT. \tag{70}$$

Dividing through by dV,

$$\left.\frac{\partial S}{\partial V}\right|_T = -\left.\frac{\partial S}{\partial T}\right|_V \frac{dT}{dV}, \tag{71}$$

and (68) becomes

$$\left.\frac{\partial E}{\partial V}\right|_S = \left.\frac{\partial E}{\partial V}\right|_T - \left.\frac{\partial E}{\partial S}\right|_V \left.\frac{\partial S}{\partial V}\right|_T . \tag{72}$$

We now identify several terms in (72). Because $dE = TdS$, $\partial E/\partial S|_V = T$. Also, the left hand side evidently expresses the pressure as it has been set up in (66). So,

$$p = -\left.\frac{\partial E}{\partial V}\right|_T + T\left.\frac{\partial S}{\partial V}\right|_T = \frac{\partial}{\partial V}\Big[E - TS\Big]_T . \tag{73}$$

The term in square brackets $[\cdot]$ is known as the Helmholtz free energy, and is usually given the letter F. But by our expression for the entropy in (58)

$$E - TS = -T \log Z(\beta), \tag{74}$$

so

$$p = -\left.\frac{\partial F}{\partial V}\right|_T = T\left.\frac{\partial \log Z}{\partial V}\right|_T . \tag{75}$$

Now we can bring in the Z associated with the ideal gas from (56):

$$Z(\beta) = \frac{V^N}{N!}\left(\frac{2\pi m}{\beta}\right)^{3N/2} = V^N C(\beta),$$

where everything unconnected to the volume is packed into a term C. Taking logs we get

$$\log Z(\beta) = N \log V + C(T), \tag{76}$$

so

$$p = T\left.\frac{\partial \log Z}{\partial V}\right|_T = T\frac{N}{V}. \tag{77}$$

This gives us the relationship

$$pV = NT, \tag{78}$$

which is the well known equation of state for an ideal gas.

Harmonic oscillator in a heat bath and diagnosis of incorrect physics

Consider a gas of particles that oscillate classically, as if it were built of two masses joined by a spring, or equivalently a mass M joined by a spring with constant k_M to a block that is massive enough not to itself perceptibly oscillate (Figure 4). Let the distance of the mass from its unperturbed position be x_M.

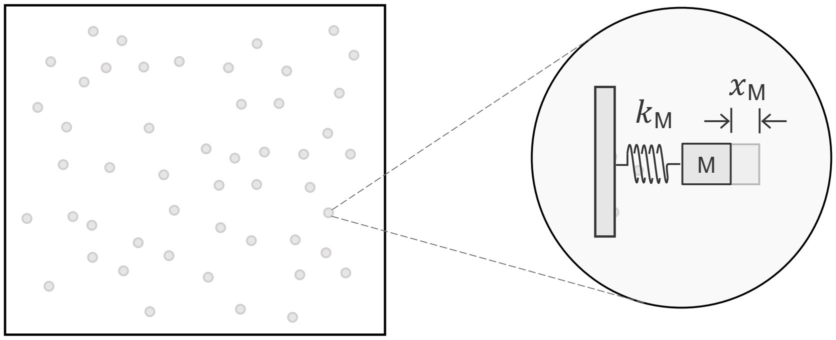

FIG. 4. A gas of harmonic oscillators.

We will analyze a single-particle gas, which, as a classical oscillator, contributes to the system the energy

$$\frac{p_M^2}{2M} + \frac{1}{2}k_M x_M^2, \text{ with } x_M \text{ and } p_M \text{ ranging from } (-\infty, \infty). \tag{79}$$

The first term represents the kinetic energy of the moving mass, and the second represents the potential energy stored by the spring. We recognize that this oscillator will have a characteristic frequency of oscillation of $\omega = \sqrt{k_M/M}$. The sum over states is

$$\sum_i \to \int_{-\infty}^{\infty} dp_M \int_{-\infty}^{\infty} dx_M, \tag{80}$$

and so the partition function is

$$Z(\beta) = \int_{-\infty}^{\infty} dp_M e^{-\beta p_M^2/2M} \int_{-\infty}^{\infty} dx_M e^{-\beta k_M x_M^2/2} = \frac{2\pi}{\beta\omega}. \tag{81}$$

If we included M independent harmonic oscillators, this would produce in turn M of these pairs of new integrals, each pair independent, producing not the Z in (81) but rather

$$Z(\beta) = \left(\frac{2\pi}{\beta\omega}\right)^M. \tag{82}$$

Equation (82) can be analyzed to produce equations of state, average energies, etc., and these would be appropriate for dilute gases involving diatomic particles, each of which acted as a classical oscillator. For instance, the average energy of the single-oscillator gas is

$$E = -\frac{\partial \log Z}{\partial \beta} = -\frac{\partial}{\partial \beta}\big(\log 2\pi - \log\omega - \log\beta\big) = \frac{1}{\beta} = T, \tag{83}$$

which suggests that each harmonic oscillator in a diatomic gas should contribute on average an amount T to the energy of the system; since this has included an incorporation of both kinetic and potential energy, we might further suggest that each of these aspects of the oscillator individually contributes $T/2$ to the system.

Diagnosis of wrong physics

The problem is, dilute diatomic gases at relatively low temperatures *do not behave like this*, i.e., according to the predictions of this updated partition function. The equation of state $pV = NT$, for instance, still holds for diatomic gases to a high degree of approximation — the altered phenomena predicted by the new partition function do not appear. More importantly, as T decreases, the system does not drop to zero $\langle E\rangle$ as this formula predicts; we see instead the famous zero-point energy being reached.

This disagreement is not the "fault" of statistical mechanics. In fact, the statistical-mechanical "answer" in (81) is correct — given the physical rules that we imposed. The origin of the problem is that the physical rules are wrong: harmonic oscillators at the atomic scale do not contribute energy in the manner of classical oscillators, they behave quantum mechanically. So equation (79) is the culprit. The oscillator cannot take on any energy value, as this equation suggests. It is restricted to taking on one of a discrete range of possible energies, namely multiples of $\hbar\omega = \hbar\sqrt{k_M/M}$, plus a minimum ground state energy of $\hbar\omega/2$.

This changes the partition function for the oscillator dramatically, to a discrete sum over all multiples of this basic amount:

$$Z(\beta) = e^{-\beta\hbar\omega/2} \sum_{n=0}^{\infty} e^{-\beta n\hbar\omega}, \tag{84}$$

rather than Z_2. As in the continuous/integral case, we are lucky here in that this sum can actually be evaluated. Setting $X = e^{-\beta\hbar\omega}$, (84) becomes

$$Z(\beta) = X^{1/2} \sum_{n=0}^{\infty} X^n = X^{1/2}\big(1 + X + X^2 + ...\big) = \frac{X^{1/2}}{1-X}, \tag{85}$$

that is

$$Z(\beta) = \frac{e^{-\beta\hbar\omega/2}}{1 - e^{-\beta\hbar\omega}}. \tag{86}$$

Checking the average energy contributed to the gas by the oscillator, as predicted by this alternative partition function, we obtain

$$\langle E \rangle = -\frac{\partial \log Z(\beta)}{\partial \beta} = \left(\frac{\hbar\omega}{2}\right) \frac{1 + e^{-\beta\hbar\omega}}{1 - e^{-\beta\hbar\omega}}. \tag{87}$$

Some examples of this average energy are plotted in Figure 5, with panel (a) showing some differences in the frequency dependence of $\langle E \rangle$ for different temperatures. In panel (b), $\langle E \rangle$ is instead plotted as a function of temperature for a fixed frequency. Scaled units are chosen such that $\omega = \hbar = 1$. Notice that as the temperature increases, $\langle E \rangle$ takes on a roughly linear dependence, with a slope of 1, i.e., $\langle E \rangle \approx T$ — the classical result in other words is arrived at for sufficiently high temperature. However, this behaviour does not persist as the temperature drops, in fact it flattens out and intersects the $\langle E \rangle$ axis at 1/2, or rather, recalling that we have set $\omega = \hbar = 1$, at $\hbar\omega/2$. This is the zero-point energy, a minimum that cannot be dropped below in the quantum theory.

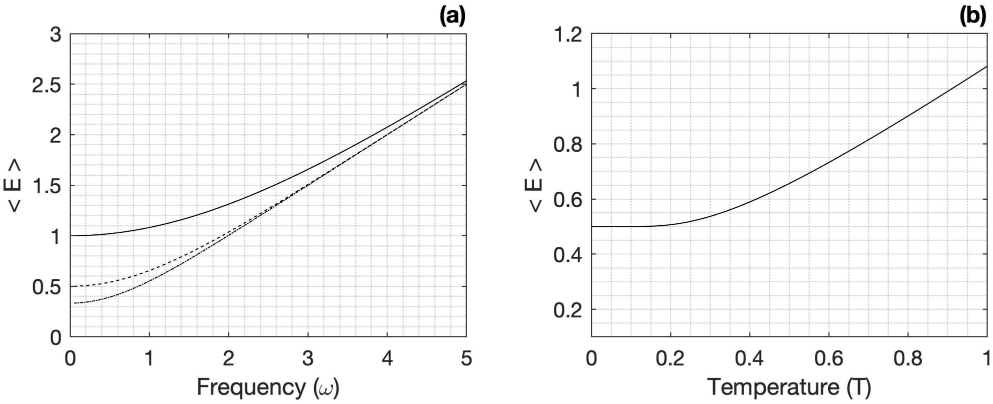

FIG. 5. Average behaviour of the quantum oscillator. (a) $\langle E \rangle$ versus frequency for three different temperature values; (b) $\langle E \rangle$ versus temperature for $\omega = \hbar = 1$.

This is a famous example of the use of a statistical-mechanical model to identify and adapt incorrect physics, and in fact was an early argument for the adoption of the quantum theory.

The molecular zipper

The “zipper model” of the unraveling of a DNA molecule in a cell (Kittel, 1969) behaves itself so well mathematically that it appears in many textbooks on the subject. In the model, one or many simple DNA molecules are assumed to have the structure of a set of parallel links resembling a zipper, which may be found in any state of openness or closedness (Figure 6a). Like a zipper, links can be open, but in the interior of the strand, a link can only be open if all previous links to one side are also open.

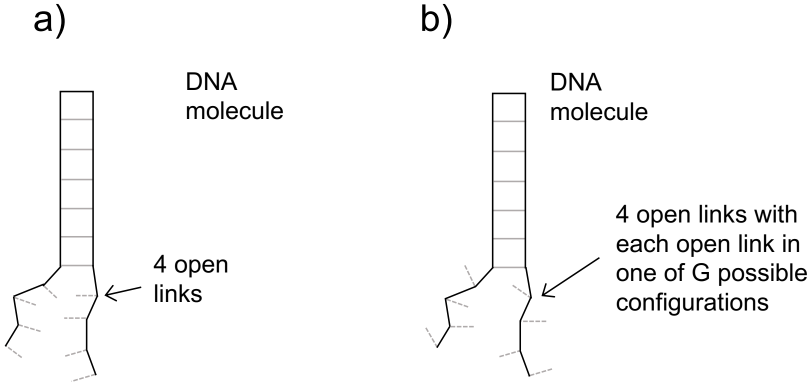

FIG. 6. The molecular zipper, a ladder-like structure with breakable links. Two variant models are considered, (a) the simple model in which broken links may take on only one possible configuration, and (b) a more complicated model in which broken links may take on one of G possible configurations.

In the model, it costs energy to open a link. If a closed link has an open link on one side of it, we set the energy needed to open it to ε. If a closed link has closed links on either side of it, the energy to open it will be defined to be infinite – meaning it cannot open. A zipper with n open links is thus in a state with energy $n\epsilon$.

One further element was added to the problem by Kittel (1969). Two types of zipper are possible. In the basic type of zipper, an open link is in a single configuration – “open” (Figure 6a). In the more complicated type, once a link is open, it can find itself in one of many different configurations, i.e., in which the broken link points either this way or that (Figure 6b). The statistical description of the second type is more complex, and more interesting, because each of these configurations has to be accounted for in the partition function. As links open, the number of possible configurations grows exponentially, and this explosion of accessible states has a profound effect on the predictions of the model, as we will see later in this review.

For now, however, let us develop the basics of the model by analyzing the simpler zipper type in panel (a). If we insist that the end of the molecule must remain closed, i.e., the molecule cannot completely separate, then the possible energy values of a molecule with N links is

$$0,\ \varepsilon,\ 2\varepsilon,\ \dots,n\varepsilon,\ \dots,\ \dots,(N-1)\varepsilon. \tag{88}$$

The partition function, which sums over all possible energy values the system can take on, is, then, for the single molecule,

$$Z(\beta) = \sum_{n=0}^{N-1} e^{-\beta n \varepsilon}. \tag{89}$$

This is a discrete sum, like it was in the quantum oscillator model, and we can use a similar strategy to analyze it. Setting $X = e^{-\beta\varepsilon}$, and using the identity

$$\sum_{m=0}^{M} q^m = \frac{1 - q^{M+1}}{1 - q}, \tag{90}$$

we can re-write Z as

$$Z(\beta) = \frac{1 - X^N(\beta)}{1 - X(\beta)}, \tag{91}$$

and subsequently the average energy is

$$\langle E \rangle = -\frac{\partial \log Z}{\partial \beta} = \varepsilon \frac{X}{1 - X} - N\varepsilon \frac{X^N}{1 - X^N} = \varepsilon \langle n \rangle, \tag{92}$$

where

$$\langle n \rangle = \frac{X}{1 - X} - N \frac{X^N}{1 - X^N} \tag{93}$$

is the average number of open links. We observe that $\langle n \rangle$ is a function of the temperature, via β. This makes sense: presumably as the temperature of the heat bath increases, the probability that a molecule has access to sufficient energy to open more links grows. In fact, we can plot the $\langle n \rangle$ as a function of temperature to see how it should be expected to vary. For simplicity, let us set $\varepsilon = 1$, and, rather than plotting $\langle n \rangle$ versus T, let us instead for convenience plot it against $X = e^{-\beta} = e^{-1/T}$. Let us plot three versions, to see the effect of increasing the number of links in the molecule (normalizing each time). See Figure 7a.

It appears that there is some potentially interesting behaviour predicted here. Consider a DNA molecule in a bath (of cellular fluid, presumably), whose temperature is being changed such that we move across Figure 7a from left to right, i.e., with increasing X. By Figure 7b, we see that we accomplish this by raising the temperature, or warming the bath. As we approach $X = 1$, we predict a very particular rate of un-ravelling of the molecules – noting that the number of links in the molecule is a strong parameter in deciding the rate of change.

Especially for cases involving large numbers of links (which is likely the case for an actual strand of DNA), further interesting behaviour is predicted as we pass $X = 1$. At this point, the average number of open links suddenly flips from a relatively small number to a very large number (nearly all of the links in the molecule). A discontinuous phenomenon of this kind, occurring as the temperature continuously varies, is called a *phase transition*;

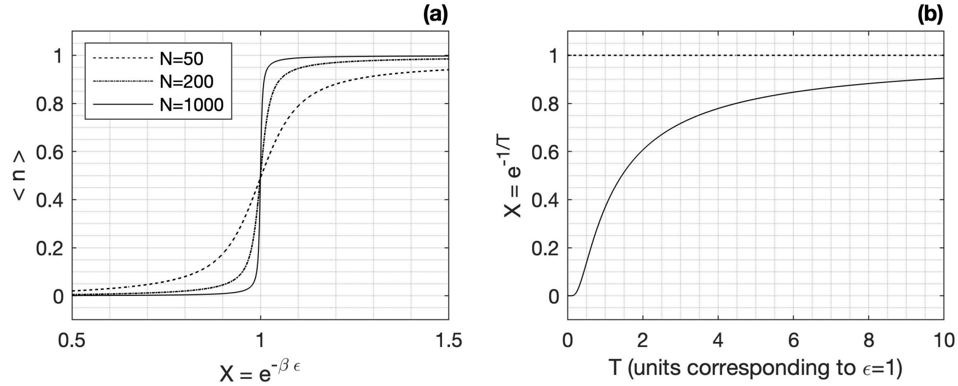

FIG. 7. Visualizing the consequences of the zipper model. (a) The average number of open links, normalized to 1, for various numbers of molecules, plotted versus $X = e^{-\beta\varepsilon}$. The plot is suggestive of a phase transition at a critical temperature T_c corresponding to $X = 1$. (b) The relationship between the independent variable X in plot (a) and the temperature. Although growing X corresponds to growing T, we observe that no temperature can be found at which $X = 1$, and the phase transition occurs. We would say then that no solution for a critical temperature T_c exists in this model, and no phase transition is possible.

the temperature at which this takes place is referred to as a *critical temperature*. In later parts of this review we will dig further in to such behaviour. For now, we unfortunately have to recognize this particular phase transition as a bit of a mirage, because X never passes through 1. To see this, we consult Figure 7b; note that as the temperature grows, X approaches but never touches 1. So, the toy model we have set up here speaks of (in some sense) a phase transition, but can never experience it. In spite of this, it is useful to see the way in which such transitions emerge mathematically in statistical-mechanical models.

The law of atmospheres

Earlier we saw how to determine the probability density and number density of particles in an ideal gas as functions of space, in the absence of an external field. To extend this somewhat, and create a less trivial result, suppose now that the box of gas in Figure 3a has no top, i.e., $l_3 \to \infty$, to simulate a column of atmosphere, and suppose further that in this column the particles feel the force of gravity. The difference appears in the energy, which then impacts the Boltzmann factor. Take the probability density function characterizing a particle of gas (mass m) at (x_1, x_2, x_3) with momentum components (p_1, p_2, p_3):

$$P(x_1, x_2, x_3, p_1, p_2, p_3, \beta) = \frac{1}{Z(\beta)} e^{-\beta E(x_1,x_2,x_3,p_1,p_2,p_3)}. \tag{94}$$

The particle has a kinetic energy determined by its momentum components, as before, but it now also experiences a force due to gravity, which confers on it a potential energy mgz, or, in the current coordinate system,

$$E(x_1, x_2, x_3, p_1, p_2, p_3) = E(x_3, p_1, p_2, p_3) = \frac{p_1^2}{2m} + \frac{p_2^2}{2m} + \frac{p_3^2}{2m} + mgx_3. \tag{95}$$

The partition function is the same in its momentum integrals as it was in the basic ideal gas case, but its spatial integrals now must be split up:

$$\begin{aligned} Z(\beta) &= \left(\int_0^{l_1} dx_1 \int_0^{l_2} dx_2\right)\left(\int_{-\infty}^{\infty} dp_1 e^{-\beta p_1^2/2m}\right)^3 \int_0^{\infty} dx_3 e^{-\beta m g x_3} \\ &= A\left(\frac{2\pi m}{\beta}\right)^{3/2}\left(\frac{1}{\beta m g}\right), \end{aligned} \tag{96}$$

where A is the area of the column in cross-section, and the rightmost bracketed term was produced using $\int_0^{\infty} e^{-sy}dy = {}^1\!/\!_s$. Substituting this Z into (94), and integrating over momentum to find the probability density function for position only, we obtain

$$P(x_3,\beta) = \left(\frac{\beta m g}{A}\right) e^{-\beta m g x_3}. \tag{97}$$

Replacing $\beta = 1/T$, the particle number density of the gas at height $z = x_3$ is therefore

$$n(z,T) = NP(z,T) = \left(\frac{Nmg}{AT}\right) e^{-mgz/T}. \tag{98}$$

We can also transition to an expression in terms of the pressure. This is still an ideal gas, external force notwithstanding, so the equation of state $pV = NT = nVT$ still holds. Thus $n = p/T$, allowing us to write (98) as

$$p(z,T) = \left(\frac{Nmg}{A}\right) e^{-mgz/T}. \tag{99}$$

Simplifying this, and recognizing the factor Nmg/A as a quantity in units of pressure that can be considered a reference value p_0, giving

$$p(z,T) = p_0 e^{-mgz/T}. \tag{100}$$

This is the well-known *law of atmospheres*, which approximates atmospheric pressure variations with temperature and elevation.

Magnetization

Another environment in which statistical mechanics has generated famous (and tractable) models is in the description of magnetism and magnetic phenomena. We will start with a 1D model, and then afterward extend to multiple dimensions. The system consists of a 1D lattice of atoms, each of which has a magnetic moment $\pm\mu$ (not to be confused with the chemical potential μ), whose amplitude is a fixed number, but which might either positive or negative. In Figure 8a the configuration is illustrated. In the figure, we have also introduced a further variable $\sigma = \pm 1$, which keeps track of whether the magnetic moment is positive or negative. The ± is the basis for our view of the orientation of the atoms as little magnets, so we illustrate this using arrows that point either up or down.

Energy enters this discussion through the inclusion of a background magnetic field H, also pictured. Suppose that the system is in a configuration in which, of the N atoms, n have a positive moment $+\mu$, and $m = N - n$ have a negative moment $-\mu$. The energy of a system with n upward-pointing moments is given by

$$E_n = (n - m)\mu H = MN\mu H, \tag{101}$$

where $M = (n - m)/N$ is the *magnetization* of the system. The energy contribution of a moment $+\mu$ is positive because it has the same sign as H. This means that it costs energy to point a magnet up. If the background magnetic field is reversed, i.e., $H \to -H$, the contribution to the energy of the downward pointing moment $-\mu$ becomes positive.

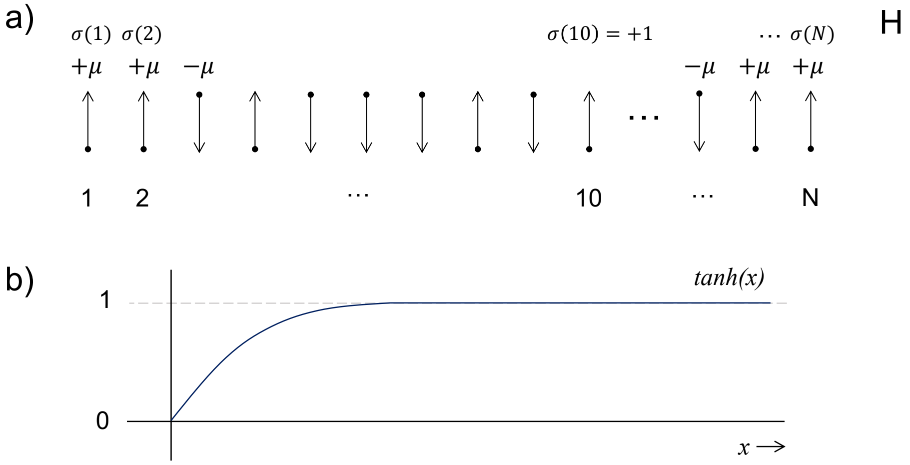

FIG. 8. (a) 1D arrangement of N particles in a background magnetic field H, each with magnetic moment $\pm\mu$. For later convenience we also introduce $\sigma(i)$, which equals ± 1, after which particle i has magnetic moment $\mu\sigma(i)$. (b) Behaviour of $\tanh x$ for x between 0 and ∞.

The energy E_n of the configuration with n upward pointing magnets (in 101) contributes a term of the form $e^{-\beta\mu H(n-m)} = e^{-\beta\mu H[n-(N-n)]}$. However, in the actual partition function this factor has to be added in the sum more than once, since more than one configuration of magnets gives n upward pointing moments. In fact, in this case the same logic by which we determined the number of states in (4) can be used again. The number of ways of arranging n upward-pointing moments out of a total of N is

$$\frac{N!}{n!(N-n)!}, \tag{102}$$

so the factor $e^{-\beta\mu H[n-(N-n)]}$ needs to be added in this many times. The full set of configurations is then the sum of these, from the case of no upward pointing magnets, $n = 0$, up to the case of all upward pointing magnets, $n = N$:

$$Z(\beta) = \sum_{n=0}^{N} \frac{N!}{n!(N-n)!} e^{-\beta\mu H[n-(N-1)]} = \sum_{n=0}^{N} \frac{N!}{n!(N-n)!} X^n Y^{N-n}, \tag{103}$$

where $X = e^{-\beta\mu H}$ and $Y = e^{\beta\mu H}$. At this point, fortune smiles, and the partition function sums produce a closed form. The sum on the right is exactly the binomial formula, so $Z(\beta) = (X + Y)^N$, or

$$Z(\beta) = \left(e^{-\beta\mu H} + e^{\beta\mu H}\right)^N = 2^N \left(\frac{e^{-\beta\mu H} + e^{\beta\mu H}}{2}\right)^N = \Big(2\cosh\left(\beta\mu H\right)\Big)^N, \tag{104}$$

using the definition of the hyperbolic cosine function. With the partition function in hand, we can go about computing the average energy $\langle E\rangle$ and the average magnetization $\langle M\rangle$:

$$\log Z(\beta) = N\log 2 + N\log\cosh(\beta\mu H), \tag{105}$$

so

$$\langle E\rangle = -\frac{\partial \log Z(\beta)}{\partial\beta} = -N\tanh(\beta\mu H)\mu H, \tag{106}$$

whereby from (101)

$$\langle M\rangle = -\tanh(\beta\mu H). \tag{107}$$

The behaviour of $\tanh(x)$, where in our case $x = \beta\mu H$, is illustrated in Figure 8b. Large values of the horizontal axis correspond to small values of $T = 1/\beta$; hence as the temperature decreases the average magnetization converges to -1. This corresponds to a configuration in which all magnets point downward, which is the lowest energy this orientation of H permits. This is a feature of systems described by the Boltmann theory – they are seekers of minimum energy. Conversely, as temperature grows towards infinity our position on Figure 8b moves towards zero, and the magnetization approaches 0 from below. This is another features of systems described by the Boltzmann theory – increasing temperature implies increasing disorder and a tendency for quantities (such as magnetization) which in order to develop require order and cohesion, to go to zero on average.

The 1D Ising model

In the preceding analysis, the model of magnetism, though simple, introduced the counting procedures by which hyperbolic trigonometric functions appear in the form of the partition function, and then predicted a specific trade-off between temperature and magnetization; to wit, the higher the temperature of the crystal, the lower the average magnetization. This reflects the idea that the alignment of moments that gives rise to measureable magnetic fields is destroyed by thermal effects. However, other, more complex behaviour, for instance the existence of a *critical temperature* (in magnetic settings often called the *Curie temperature*), a point at which a discontinuous change in magnetization occurs (and above which a crystal loses its magnetization completely), is not predicted by the previous model.

We next move to a different model, called the *Ising model*, which is, ultimately, capable of producing a critical temperature. We will in fact recognize in this model all of the hallmarks of a phase transition[††]. It too involves a lattice of atoms (e.g., a crystal) with magnetic moments, but this time the moments do not contribute to an energy by being immersed in a background magnetic field. Instead, the moments contribute to the total energy by dint of how they interact with one another.

We return for the moment to the 1D arrangement in Figure 8a, remove the background H, and assert that for any pair of particles, the contribution to the energy is negative if the two

---

[††] The very kind of change that failed to happen with our first attempt at a molecular zipper.

are aligned (e.g., the first two particles, or the 5th and 6th particles) and positive if they oppose (e.g., the second and third). Thus we add to the system (1) a mechanism of communication between neighbouring atoms on the lattice, and (2) a tendency (via reduction in energy) for adjacent atoms to be aligned. Whatever the solution of the problem is, it will involve an attempt to produce fully aligned particles, combated by random variations as the temperature grows.

To accommodate the idea of communication between neighbours, we set the energy contributed by the first two adjacent particles to be

$$-\mathcal{J}\sigma(1)\sigma(2), \tag{108}$$

with the σ defined per Figure 8a. Here $\mathcal{J}$ contains the magnetic features of the system[‡‡]. On the full 1D lattice, the total energy is then

$$E = -\mathcal{J}\sum_{n=1}^{N-1}\sigma(n)\sigma(n+1). \tag{109}$$

This setup has some interesting features from the start. Notice that to produce a low energy, many pairs of aligned moments are needed, but the cases of all atoms aligned up ($\sigma(n)$ and $\sigma(n{+}1)$ both positive) or all aligned down ($\sigma(n)$ and $\sigma(n{+}1)$ both negative) both do equally well. Our expectation is that at high temperature random variations will dominate, but as the temperature drops, an increase in the alignment of moments will be found. But which? This question cannot be resolved without more information — currently it represents a symmetry of the system.

Let us revert once again to an analysis of the statistics of only one of the N particles within the ensemble, as we did in the case of the harmonic oscillator and the law of atmospheres. The one particle alone contributes an energy to the partition function that is the product of two terms that each take on either the value +1 or the value -1, so the contribution itself must be either

$$e^{\beta(+1)\mathcal{J}} \text{ or } e^{\beta(-1)\mathcal{J}}. \tag{110}$$

The partition function is the sum of all (both) of these possible configurations:

$$Z(\beta) = e^{+\beta\mathcal{J}} + e^{-\beta\mathcal{J}} = 2\cosh\beta\mathcal{J}. \tag{111}$$

The average energy is

$$\langle E\rangle = -\frac{1}{Z(\beta)}\frac{\partial \log Z(\beta)}{\partial\beta} = -\mathcal{J}\tanh\beta\mathcal{J}, \tag{112}$$

[‡‡] For completeness, $\mathcal{J} = \mu B$, where as before μ is the magnetic moment of each atom (in this model we assume all atoms are identical), and B is the magnetic field. Here B is often used as the field rather than H, although B is the magnetic flux density rather than the magnetic field proper. The distinction is important but we will not pursue it here.

and so, since energy in the amount $-\mathcal{J}$ is spent per "upward agreement" in orientation, the average magnetic orientation σ must be

$$\langle\sigma\rangle = \tanh\beta\mathcal{J} = \tanh\left(\frac{\mathcal{J}}{T}\right). \tag{113}$$

The function $\tanh x$ (and therefore the average magnetization $\langle\sigma\rangle$) is again illustrated in Figure 9, this time over a wider range of x values. We can observe the symmetry now in action. With $\mathcal{J} = 1$ fixed and positive, with decreasing temperature (i.e., $x \to \infty$) the average $\langle\sigma\rangle$ stabilizes at 1, i.e., all pointing up. With the unit energy fixed and negative it stabilizes at -1.

This sketch of average magnetization versus x bears some resemblance to Figure 7a. And as in the zipper case, it does appear to hint at a phase transition – as the temperature changes, and x varies from the left to the right, a transition seems to be implied, from an overall negative magnetization to an overall positive one. But, again like in the zipper case, on closer inspection we observe that for a system to pass through this transition is impossible, since temperature approaches infinity as we approach $x = 0$ from either side.

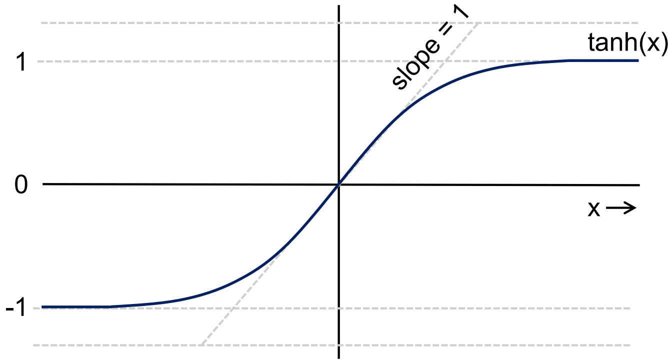

FIG. 9. Sketch of $\tanh(x)$ versus x. Note the slope of the function is 1 at $x = 0$.

That aside, the 1D model is useful in that it allows us to develop an understanding of the effect of communication between lattice points. Let us return to the full lattice, whose energy is

$$E = -\mathcal{J}\sum_{n=1}^{N-1}\sigma(n)\sigma(n+1) = -\mathcal{J}\sum_{n=1}^{N-1}\kappa(n). \tag{114}$$

In the rightmost term we have replaced $\sigma(n)\sigma(n+1)$ with $\kappa(n)$. Doing this tends to shift attention away from the atoms in the lattice, and onto the links between the atoms; for this reason κ is referred to as a *bond variable*. Of course, there are 1 fewer bond variables than there are atom configuration variables, but if we set the first σ, i.e., $\sigma(1)$, equal to +1 (knowing that we will have to come back later and adjust for having done this), then there are the same number of variables κ as there were σ. The partition function will then be a sum of contributions of the form

$$\exp\left[\beta\mathcal{J}\sum_{n=1}^{N-1}\kappa(n)\right], \tag{115}$$

with a contribution included for every instance of the $N-1$ values κ can take on across the lattice. Since it is built of products of the σ values, and since σ can only be $+1$ or -1, so too must κ be either $+1$ or -1. So, the $\kappa(n)$ over $n=(1,...,N-1)$ will be a string of $+1$s and -1s. Suppose there are n instances of $+1$. Then, letting $N'=N-1$,

$$\sum_{k=1}^{N'}\kappa(n)=n-(N'-n), \tag{116}$$

and (115) becomes

$$\exp\left[\beta\mathcal{J}[n-(N'-n)]\right]=\big(e^{\beta\mathcal{J}}\big)^n\big(e^{-\beta\mathcal{J}}\big)^{(N'-n)}=X^nY^{(N'-n)}, \tag{117}$$

where $X=e^{\beta\mathcal{J}}$ and $Y=e^{-\beta\mathcal{J}}$. The partition function is proportional to the sum of all such contributions, with each contribution multiplied by the number of times the situation "n instances of $+1$ and $N'-n$ instances of -1" occurs. This factor is given by the same binomial formula we have now used several times:

$$Z(\beta)\propto\sum_{n=0}^{N'}\frac{N'!}{n!(N'-n)!}X^nY^{(N'-n)}=(X+Y)^{N'}. \tag{118}$$

Substituting X, Y, and $N'=N-1$ back in, and again recognizing the hyperbolic cosine, we have

$$Z(\beta)\propto\big[\cosh\big(\beta\big)\big]^{N-1}. \tag{119}$$

Finally, we remove the $\propto$ sign by multiplying the whole partition function by 2. When we transformed from the σ notation to the κ notation, we fixed $\sigma(1)=+1$. Now that we have completed the analysis, we observe that we could have done the whole thing with the other choice, namely $\sigma(1)=-1$, and gotten exactly the same answer. This means the sum over all possibilities actually needed to happen twice:

$$Z(\beta)=2\big[\cosh\big(\beta\mathcal{J}\big)\big]^{N-1}. \tag{120}$$

From this we can quickly determine the average energy using our standard approach:

$$\langle E\rangle=-\frac{\partial\log Z(\beta)}{\partial\beta}=\frac{\partial}{\partial\beta}\bigg(2+(N-1)\big[\cosh\big(\beta\mathcal{J}\big)\big]\bigg)=-(N-1)\tanh\beta\mathcal{J}.$$

This suggests that $\tanh\beta$ must be the average value of κ. If we multiply the average value of κ by $(N-1)$, we get the average number of $+1$'s contributing to the energy, and if we then multiply that result by the unit energy per link -1, what emerges must be the average energy. Thus

$$\langle\kappa\rangle=\tanh\beta\mathcal{J}=\tanh\left(\frac{\mathcal{J}}{T}\right). \tag{121}$$

The *correlation function* is the average product of $\sigma(i)$ with $\sigma(i{+}n)$. It gives the probability that, if the ith particle is up, the $(i+n)$th particle will also be up. This is

$$\langle \sigma(i)\sigma(i+n)\rangle \,. \tag{122}$$

Since the square of any σ is 1, i.e., $\sigma(j)\sigma(j) = 1$, we can insert pairs into the average above without changing the result, and connect them in alternating pairs to reformulate the construction in terms of κ:

$$\langle \sigma(i)\sigma(i+1)...\sigma(i+n-1)\sigma(i+n)\rangle = \langle \kappa(i)\kappa(i+1)...\kappa(i+n-1)\rangle \,. \tag{123}$$

If we assume that the κ's are all independent, this evaluates to the product of n average κ's. Therefore, by (121),

$$\langle \sigma(i)\sigma(i+n)\rangle = \langle\kappa\rangle^n = \left(\tanh\frac{\mathcal{J}}{T}\right)^n \,. \tag{124}$$

This expresses the fidelity of communication between particles. If the temperature is low, the $\tanh$ function is close to one, and as you step from one atom to another n links away, the communication between them is sustained, since the repeated product of the $\tanh$ with itself n times only reduces slightly. However, as the temperature increases, and $\tanh$ descends towards zero, $\tanh^n$ very rapidly diminishes the correlation. The probability of alignment of two particles separated by n lattice points drops to zero.

Phase transitions

We have thus far twice been led to the brink of discussing phase transitions, but in both situations, although the mathematical wherewithal was in place for a "jump" to take place, from one distinct state to another, physically the jump required that a transit be made across illegal variable values. In both situations it transpires that with slight alterations to the problem the jumps can become realistic. We will cover these alterations in this section.

The multidimensional Ising model

In seeking a phase transition in the Ising model, what we are seeking is a global effect of the communication between adjacent atoms on the lattice to occur. A magnetized crystal has all of its atoms either fully up or fully down. We have seen that either case leads to the same low energy, and so there is no a priori preference for either, and the situation is symmetrical. To select one or the other in the context of the Ising model requires that the atoms communicate any one choice between themselves, via the energy rules we have established. The problem is that if information can only be conveyed via an adjacent atom, and that atom is subject to some degree of randomness, then any trend to the left of a given atom cannot be transmitted to the right, if that atom happens to be pointing in the wrong direction. In fact, this all-or-nothing nature of information transmission in the 1D lattice is the reason why the phase transitions based on communication are impossible here.

Consider instead a 3D lattice, like the one in Figure 10a. Here, even if we stick to the idea of a lattice atom only being able to communicate with its immediate neighbours, now the

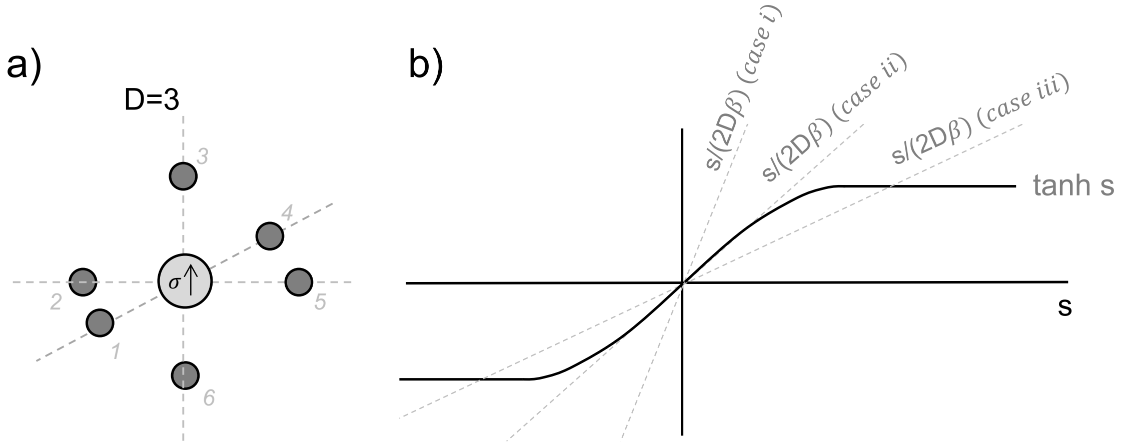

FIG. 10. The multidimensional Ising model. (a) In D dimensions, the number of adjacent particles is $2D$. (b) Mean field theory is an approximate approach in which we equate the average site magnetization with the average surrounding orientations.

number of those neighbours has increased. In fact, every atom in a lattice of dimension D has $2D$ neighbouring atoms. The extension of the Ising model involves including all of the neighbouring atoms in the calculation of the energy contribution of an atom with orientation σ:

$$-\mathcal{J}\sigma\sum_{i=1}^{2D}\sigma(i), \tag{125}$$

with the sum being over all neighbours. From now on we will omit $\mathcal{J}$, as we did previously, assuming units are available within which $\mathcal{J} = 1$. Next we will simplify our analysis by making the *mean field approximation*, in which we replace the sum by $2D$ times the average configuration, which we will call $\bar{\sigma}$. The energy contribution becomes

$$-2D\sigma\bar{\sigma}. \tag{126}$$

Repeating our single-atom analysis in this context, the partition function is the sum of contributions of this energy for all possible configurations, which in this case number 2, one with $\sigma = +1$ and the other with $\sigma = -1$. This produces a similar form as that determined previously:

$$Z(\beta) = e^{\beta(2D\bar{\sigma})} + e^{-\beta(2D\bar{\sigma})} = 2\cosh\big(2D\bar{\sigma}\big). \tag{127}$$

As we did in (114), we can use this Z to compute $\langle E\rangle$ and then the average magnetic orientation $\langle\sigma\rangle$. Skipping a few steps:

$$\langle\sigma\rangle = \tanh\big(\beta 2D\bar{\sigma}\big), \tag{128}$$

emphasizing that $\bar{\sigma}$ is the average *surrounding* orientation, computed via the mean field approximation, and $\langle\sigma\rangle$ is the average *site* orientation. So far in this extension, the mechanism for communication between atoms is not as clear as it was in the 1D model. There, the reduction in energy of a pair of aligned orientations was the way one atom influenced its neighbour. Here, that communication is unclear, since there are multiple neighbours. We include it with an additional step, which is to ask that the average site orientation be equal to the mean field, i.e., that $\langle\sigma\rangle = \bar{\sigma}$. This transforms (128) into an equation for $\bar{\sigma}$:

$$\bar{\sigma} = \tanh\big(\beta 2D\bar{\sigma}\big). \tag{129}$$

To make analysis easier, change variables to $s = 2D\bar{\sigma}$, in which case we have

$$\left(\frac{T}{2D}\right) s = \tanh s. \tag{130}$$

We plot this for several different values of $T = 1/\beta$ in Figure 10b. At high temperature, which corresponds to the steep line, case (i), the solution — which is the point at which the line and the hyperbolic tangent function cross – occurs at the origin, $s = 0$. This corresponds with $\bar{\sigma} = 0$, and signifies that when temperature is high, random fluctuations in magnetic orientation dominate and no average magnetization occurs. As temperature decreases, the slope of the line decreases, as illustrated with cases (ii) and (iii). As we pointed out earlier, the slope of the $\tanh$ function is 1 at the origin, so when $T = 2D$, the slope of the left hand side of the equation and the tangent of the righthand side are equal. This is case (ii). Beyond this critical temperature, e.g., in case (iii), additional solutions of (132) appear, at both positive and negative points along s, where the less-steep line now intersects the curve $\tanh s$. These solutions represent cases of full magnetization. That is, below the critical temperature, a state of complete upward and a state of complete downward magnetization become legitimate solutions. *This* passage of T through a critical temperature represents a phase transition.

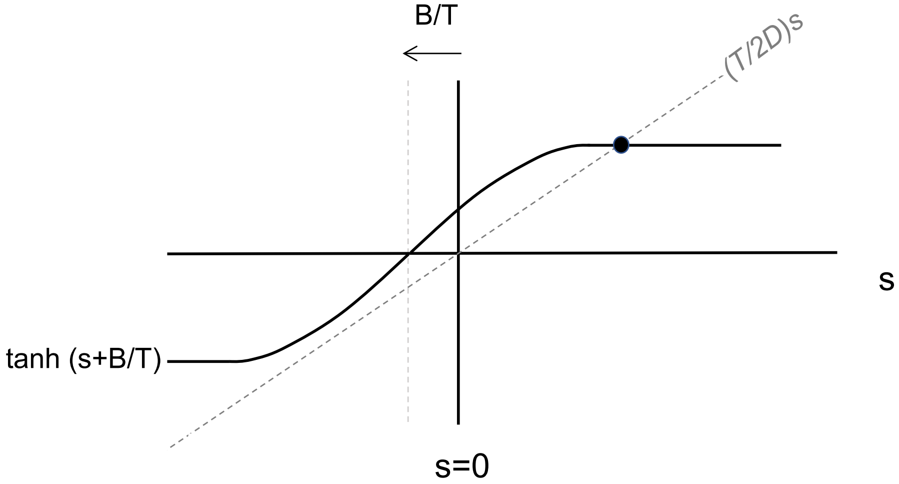

FIG. 11. Framework for analyzing magnetization in the multidimensional Ising model in the presence of a background magnetic field. The intersection of the two curves represents the solution – now, with the $\tanh$ function shifted to the left by B/T, a single solution is available, in which the average magnetization is +1, i.e., the lattice is fully magnetized.

There remains the issue of the symmetry between up and down orientations — still we see no mechanism to choose between the left solution and the right solution; further we see that $s = 0$ remains a solution, which is slightly disconcerting. However, suppose we perturb this system by applying a weak constant magnetic field. To do this is to alter the energy in (125)-(126) by adding a small term proportional to σ:

$$-2D\sigma\bar{\sigma} \rightarrow -\biggl(2D\sigma\bar{\sigma} + B\sigma\biggr). \tag{131}$$

Here B is the magnetic flux density of the field we have added, but for our purposes it is sufficient to think of it as a constant. Propagating this through the same mathematical sequence as before we find that (132) becomes

$$\left(\frac{T}{2D}\right) s = \tanh\left(s + \frac{B}{T}\right). \tag{132}$$

Evidently the problem is similar, but with the $\tanh$ function shifted to the left (assuming $B > 0$) by an amount that depends on the temperature. The situation is illustrated in Figure 11.

Much has happened to clarify the process of magnetization in the Ising model. Now, $s = 0$ is no longer a viable solution, and the presence of the magnetic field has shifted the non-linear function such that a single solution is selected – in this case the "upward" magnetic orientation. Of course, this is the result of a positive B, i.e., it is the upward nature of B that caused the upward magnetization to occur.

The molecular zipper revisited

It turns out that with a slight – and physically warranted – variation to the molecular zipper model, we can also introduce the possibility of a phase transition. We just need to switch to the model in Figure 6b. Here, when the zipper is open, the open links are free to change their orientation, and take on one of some variety of positions. Suppose that that freedom is somewhat restricted, and each open link can take on a discrete, countable number of orientations, a number Kittel called G. If so, it means we counted incorrectly in formulating the zipper partition function, which we recall was built upon contributions of the form

$$e^{-\beta n\varepsilon}. \tag{133}$$

We built up the partition function by counting one of these contributions for each possibility, from $n = 0$, meaning no zipper links were open, through to $n = N - 1$, meaning all but one of the links were open. But, in the new model, if (say) one link is open, we need to count this state G times, since the configuration can come up in G different ways. In fact, there are G^n ways for n links to be open if each one is able to take on G different orientations. So, the partition function for this "degenerate" problem is instead

$$Z(\beta) = \sum_{n=0}^{N-1} G^n e^{-\beta n\varepsilon} = \sum_{n=0}^{N-1} Y^n, \tag{134}$$

where Y differs from the X we used in the earlier derivation only in the presence of the G, i.e., $Y = GX$. The partition function is then as before but with Y in the place of X:

$$Z(\beta) = \frac{1 - Y^N}{1 - Y}, \tag{135}$$

as is the average number of open links $\langle n\rangle$:

$$\langle n\rangle = \frac{Y}{1-Y} - N\frac{Y^N}{1-Y^N}. \tag{136}$$

The plot of *this* result will have the same left panel as before, but a new right panel, since we will now plot Y as a function of temperature instead of X, and since within Y we now have a parameter G. Setting $G = 1.25$, we obtain a counterpart to Figure 12. This time, however, we observe in (b) that Y passes through the value 1 at a well-defined critical temperature, and as Y passes through 1, we observe in (a) that a discontinuous jump in $\langle n\rangle$ occurs. The critical temperature satisfies

$$Y = Ge^{-\varepsilon/T_c} = 1, \tag{137}$$

and so

$$T_c = \frac{\varepsilon}{\log G}, \tag{138}$$

which gives $T_c \approx 4.5$ for $\varepsilon = 1$ and $G = 1.25$, which is consistent with Figure 12b.

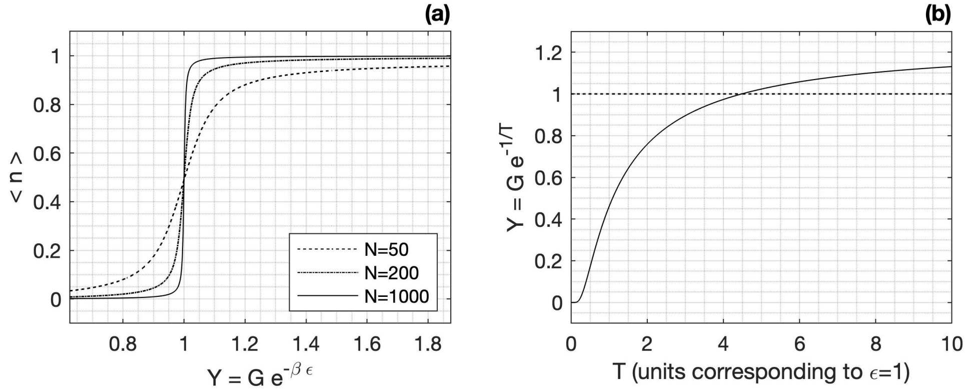

FIG. 12. The zipper model with more than 1 orientation for open links. Here $G = 1.25$ is used; now the phase transition at $Y = 1$ is physically traversed by a system with increasing temperature.

The phase transition is a jump in the probable state of a single zipper from almost completely closed at $T < T_c$ to almost completely open at $T > T_c$. As the number of links in the zipper (N) grows, the transition becomes extremely rapid, in fact discontinuous.

It is the presence of $G > 1$ that produces a finite T_c, in fact $G = 1$ is clearly not permitted in (138), which is why we previously could not bring about a phase transition in the zipper. Why intuitively does it occur though? How does the availability of more than one orientation of an open link lead to the possibility of a phase transition?

STATISTICS OF SYSTEMS THAT EXCHANGE ENERGY AND PARTICLES

Thus far all of our models have grown out of (29), which is the distribution associated with the Canonical Ensemble, i.e., systems which share energy but not particles with their environment. Returning to Figure 1, we recall that initially we permitted two types of exchange between systems, energy, or thermal exchange, and particle, or chemical exchange, but very quickly restricted the ensemble such that chemical exchange would not occur. Let us now relax that restriction, and discuss the Grand Canonical Ensemble.

The distribution and partition function of the GCE

The problem, in words, is much the same as in the case of the Canonical Ensemble; we seek to determine a probability distribution descriptive of the system, and the partition function normalizing it. We do this by considering the occupation numbers implied by Figure 1, and at the appropriate moment re-express these numbers as probabilities. However, the occupation numbers themselves now must change. We still have $\mathcal{N}$ boxes in the ensemble, but now the boxes distinguish themselves by having an energy E_i and a number of particles N_j. We therefore count up not n_i, or how many boxes have energy E_i, but instead n_{ij}, how many boxes have *both* energy E_i *and* particle number N_j. Since the i run over a list of allowable energy values, and the j run over a list of allowable numbers of particles, the n_{ij} comprise an array of numbers:

$$n_{ij} = \begin{bmatrix} n_{11} & n_{12} & \dots & n_{1L} \\ n_{21} & n_{22} & \dots & n_{2L} \\ \vdots & \vdots & \dots & \vdots \\ n_{M1} & n_{M2} & \dots & n_{ML} \end{bmatrix}, \tag{139}$$

where M is, again, some maximum energy index, beyond which the energy is so high that the number of microstates can be assumed to be nil, and where now L is a similar limit on the maximum number of particles in the system. If we cycle through arrangements of boxes, varying particles and energies, each time counting up and recording the array n_{ij}, we observe as before that some n_{ij} arrays are produced by many arrangements of boxes, and some are produced by relatively few. And as before, we assume that the n_{ij} that actually occurs is that which is produced by the largest number of arrangements.

Given a set of occupation numbers n_{ij}, we determine the number of arrangements $W(\{n_{ij}\})$ giving rise to it with the same combinatoric formula we used in the earlier case, but with the now larger set of factorials in the denominator:

$$W(\{n_{ij}\}) = \frac{\mathcal{N}!}{n_{11}!n_{12}!...n_{ML}!} = \frac{\mathcal{N}!}{\prod_{i=1}^{M}\prod_{j=1}^{L} n_{ij}!}. \tag{140}$$

Taking the logarithm, applying Stirling's approximation, and replacing n_{ij} with $P_{ij} = P(E_i, N_j) = n_{ij}/\mathcal{N}$, we obtain

$$\log W(\{n_{ij}\}) = \mathcal{N}S, \tag{141}$$

where

$$S = -\sum_{i=1}^{M}\sum_{j=1}^{L} P_{ij}\log P_{ij} \tag{142}$$

is the entropy. The problem of determining the n_{ij} that are produced most often when cycling through microstates has again become the problem of determining the P_{ij} that minimize $-S$. Again the minimization problem is subject to constraints, in fact one more constraint is now relevant. Proper P_{ij} values must sum to one:

$$\sum_{i=1}^{M}\sum_{j=1}^{L} P_{ij} - 1 = 0. \tag{143}$$

Also, the contributing energy values must add up to a total E_T such that $E = E_T/\mathcal{N}$, recalling that the un-indexed E is a compact way of writing the average $\langle E \rangle$. The total energy must in other words be the sum of the energy of each allowable level multiplied by the number of instances of that level:

$$\sum_{i=1}^{M} \sum_{j=1}^{L} n_{ij} E_i = E_T, \tag{144}$$

so (after dividing through by $\mathcal{N}$), the problem is constrained such that

$$\sum_{i=1}^{M} \sum_{j=1}^{L} P_{ij} E_i - E = 0. \tag{145}$$

Both of these are in concept the same as before. We have to add to these two statements another, concerning the particles. They are permitted to re-distribute themselves amongst the boxes in the ensemble, but they must contribute to a fixed total N_T and produce a well-defined average $N = N_T/\mathcal{N}$ per box as well. That is,

$$\sum_{i=1}^{M} \sum_{j=1}^{L} n_{ij} N_j = N\mathcal{N}. \tag{146}$$

Again dividing through by $\mathcal{N}$, we obtain a new constraint equation:

$$\sum_{i=1}^{M} \sum_{j=1}^{L} P_{ij} N_j - N = 0. \tag{147}$$

This full package is then incorporated in an augmented functional

$$\mathcal{L} = \sum_{i,j}^{M,L} P_{ij} \log P_{ij} + \alpha \Bigg(\sum_{i,j}^{M,L} P_{ij} - 1 \Bigg) + \beta \Bigg(\sum_{i,j}^{M,L} P_{ij} E_i - E \Bigg) + \gamma \Bigg(\sum_{i,j}^{M,L} P_{ij} N_j - N \Bigg),$$

whose derivatives with respect to each P_{ij} set to zero produce

$$\log P_{ij} = -(\alpha + 1) - \beta\big(E_i - \mu N_j\big), \tag{148}$$

where following convention we have changed variables from γ with $\mu = -\gamma/\beta$. Exponentiating and setting $Z = e^{\alpha+1}$, we obtain

$$P(E_i, N_j) = \frac{1}{Z(\beta, \mu)} e^{-\beta\left(E_i - \mu N_j\right)}, \tag{149}$$

where the new partition function is found, by applying the α constraint, to be

$$Z(\beta, \mu) = \sum_{i=1}^{M} \sum_{j=1}^{L} e^{-\beta\left(E_i - \mu N_j\right)}, \tag{150}$$

where the sum is understood to be over all possible states. The formalism has introduced a new parameter μ, which is called the *chemical potential*; as suggested by Figure 1, it is interpreted as the amount of energy a particle in one of the boxes must "pay" to leave the box.

CONCLUSIONS

Equilibrium statistical mechanics provides calculation methods which seem on their face to have applicability in geophysical inversion. However, to develop and test these possibilities it seems prudent to use (or seek) forms in which full, large computational problems are not immediately introduced. (After all, if we take the attitude that massive computation problems are not a worry, we already have good inversion and appraisal methods.) To the author, what is remarkable about the approach is that apparently complex problems do occasionally have tractable, analyzable solutions. It suggests an approach to investigating geophysical application possibilities, which is to stick near to shore, and see to what degree tractable models of statistical mechanics port over. This review is intended to act as a kind of resource as this investigation proceeds.

ACKNOWLEDGEMENTS

The sponsors of CREWES are gratefully thanked for continued support. This work was funded by CREWES industrial sponsors, NSERC (Natural Science and Engineering Research Council of Canada) through the grant CRDPJ 543578-19, and in part by an NSERC-DG.

REFERENCES

Buland, A., and Omre, H., 2003, Bayesian wavelet estimation from seismic and well data: Geophysics, **68**, No. 1, 185–198.

de Figueiredo, L. P., Grana, D., Roisenberg, M., and Rodrigues, B. B., 2019, Gaussian mixture model Markov chain Monte Carlo method for linear seismic inversion: Geophysics, **84**, No. 3, R463–R476.

Fichtner, A., Zunino, A., Gebraad, L., and Boehm, C., 2021, Autotuning Hamiltonian Monte Carlo for efficient generalized nullspace exploration: Geophysical Journal International, **227**, 941–968.

Gleick, J., 1992, Genius: the life and science of Richard Feynman: Pantheon Books, 1st edn.

Keating, S. D., and Innanen, K. A., 2021, Null-space shuttles for targeted uncertainty analysis in full-waveform inversion: Geophysics, **86**, No. 1, R63–R76.

Kestin, J., and Dorfman, J. R., 1971, A course in statistical thermodynamics: Academic Press Inc, 1st edn.

Kittel, C., 1969, Phase transition of a molecular zipper: American Journal of Physics, **37**, 917.

Sears, F. W., 1959, An introduction to thermodynamics, the kinetic theory of gases, and statistical mechanics: Addison-Wesley Publishing, 4th edn.

Susskind, L., 2013, The theoretical minimum: statistical mechanics.

Velis, D. R., and Ulrych, T. J., 1996, Simulated annealing wavelet estimation via fourth-order cumulant matching: Geophysics, **61**, No. 6, 1939–1948.