

Chapter 2

Statistical Mechanics in a Nutshell

Adapted from:

“Understanding Molecular Simulation”
Daan Frenkel and Berend Smit
Academic Press (2001)
pp. 9-22

2.1 Introduction

In this course, we will treat computer simulations methods that are able to describe the properties of thermodynamic ensembles of molecules. Computer simulations allow us to study properties of many-particle systems. However, not all properties can be directly measured in a simulation. Conversely, most of the quantities that can be measured in a simulation do not correspond to properties that are measured in real experiments. To give a specific example: in a molecular dynamics simulation of liquid water, we could measure the instantaneous positions and velocities of all molecules in the liquid. However, this kind of information cannot be compared to experimental data because no real experiment provides us with such detailed information. Rather, a typical experiment measures an average property, averaged over a large number of particles and, usually, also averaged over the time of the measurement. If we wish to use computer simulations as the numerical counterpart of experiments, we must know what kind of averages we should aim to compute in order to explain this, we need to introduce the language of statistical mechanics.

This we shall do here. We will follow a quick (and slightly dirty) derivation of the basic expressions of statistical mechanics. The aim of these derivations is only to show that there is nothing mysterious about concepts such as phase space, temperature and entropy and many of the other statistical mechanical objects that will appear time and again in the remainder of this course.

2.2 Entropy and Temperature

Most of the computer simulations that we discuss are based on the assumption that classical mechanics can be used to describe the motions of atoms and molecules (and a derivation of this classical limit was given in Chapter 1 of the course). This assumption leads to a great simplification in almost all calculations, and it is therefore most fortunate that it is justified in many cases of practical interest. Surprisingly, it turns out to be easier to derive the basic laws of statistical mechanics using the language of quantum mechanics.

We will follow here this route of least resistance. In fact, for our derivation, we need only little quantum mechanics. Specifically, we need the fact that a quantum mechanical system can be found in different states. For the time being, we limit ourselves to quantum states that are eigenvectors of the Hamiltonian \mathcal{H} of the system (i.e., energy eigenstates). For any such state $|i\rangle$, we have that $\mathcal{H}|i\rangle = E_i|i\rangle$, where E_i is the energy of state $|i\rangle$. Most examples discussed in quantum mechanics textbooks concern systems with only a few degrees of freedom (e.g., the one-dimensional harmonic oscillator or a particle in a box). For such systems, the degeneracy of energy levels will be small. However, for the systems that are of interest to statistical mechanics (i.e. systems with $\mathcal{O}(10^{23})$ particles), the degeneracy of energy levels is astronomically large. In what follows, we denote by $\ln \Omega(E, V, N)$ the number of eigenstates with energy E of a system of N particles in a volume V . We now express the basic postulate of statistical mechanics:

Basic Postulate of Statistical Mechanics:

A system with fixed N , V and E is equally likely to be found in any of its $\Omega(E)$ eigenstates.

Much of statistical mechanics follows from this simple (but highly nontrivial) assumption.

To see this, let us first consider a system with total energy E that consists of two weakly interacting subsystems. In this context, weakly interacting means that the subsystems can exchange energy but that we can write the total energy of the system as the sum of the energies E_1 and E_2 of the subsystems.

There are many ways in which we can distribute the total energy over the two subsystems such that $E_1 + E_2 = E$. For a given choice of E_1 , the total number of degenerate states of the system is $\Omega_1(E_1) \times \Omega_2(E_2)$. Note that the total number of states is not the sum but the product of the number of states in the individual systems (since for every possible state $\Omega_1(E_1)$ there are $\Omega_2(E_2)$ possibilities that subsystem 2 can assume). In what follows, it is convenient to have a measure of the degeneracy of the subsystems that is additive. A logical choice is to take the (natural) logarithm of the degeneracy. Hence:

$$\ln \Omega(E_1, E - E_1) = \ln \Omega_1(E_1) + \ln \Omega_2(E - E_1) \quad (2.1)$$

We assume that subsystems 1 and 2 can exchange energy. What is the most likely distribution of the energy? We know that every energy state of the total system is equally likely. But the number of eigenstates that correspond to a given distribution of the energy over the subsystems depends very strongly on the value of E_1 . We wish to know the most likely value of E_1 , that is, the one that maximizes $\ln \Omega(E_1, E - E_1)$. The condition for this maximum is that

$$\left(\frac{\partial \ln \Omega(E_1, E - E_1)}{\partial E_1} \right)_{N, V, E} = 0 \quad (2.2)$$

or, in other words,

$$\left(\frac{\partial \ln \Omega_1(E_1)}{\partial E_1} \right)_{N_1, V_1} = \left(\frac{\partial \ln \Omega_2(E_2)}{\partial E_2} \right)_{N_2, V_2} \quad (2.3)$$

We introduce the shorthand notation

$$\beta(E, V, N) \equiv \left(\frac{\partial \ln \Omega(E, V, N)}{\partial E} \right)_{N, V} \quad (2.4)$$

With this definition, we can write equation (2.3) as

$$\beta(E_1, V_1, N_1) = \beta(E_2, V_2, N_2) \quad (2.5)$$

Clearly, if initially we put all energy in system 1 (say), there will be energy

transfer from system 1 to system 2 until equation (2.3) is satisfied. From that moment on, no net energy flows from one subsystem to the other, and we say that the two subsystems are in (thermal) equilibrium. When this equilibrium is reached, $\ln \Omega$ of the total system is at a maximum. This suggests that $\ln \Omega$ is somehow related to the thermodynamic entropy S of the system. After all, the second law of thermodynamics states that the entropy of a system N , V , and E is at its maximum when the system is in thermal equilibrium. There are many ways in which the relation between $\ln \Omega$ and entropy can be established. Here we take the simplest route; we simply define the entropy to be equal to $\ln \Omega$. In fact, for (unfortunate) historical reasons, entropy is not simply equal to $\ln \Omega$; rather we have

$$S(N, V, E) \equiv k_B \ln \Omega(N, V, E) \quad (2.6)$$

where k_B is Boltzmann's constant, which in S.I. units has the value 1.38066×10^{-23} J/K. With this identification, we see that our assumption that all degenerate eigenstates of a quantum system are equally likely immediately implies that, in thermal equilibrium, the entropy of a composite system is at a maximum. It would be a bit premature to refer to this statement as the second law of thermodynamics, as we have not yet demonstrated that the present definition of entropy is, indeed, equivalent to the thermodynamic definition. We simply take an advance on this result.

The next thing to note is that thermal equilibrium between subsystems 1 and 2 implies that $\beta_1 = \beta_2$. In everyday life, we have another way to express the same thing: we say that two bodies brought into thermal contact are in equilibrium if their temperatures are the same. This suggests that β must be related to the absolute temperature. The thermodynamic definition of temperature is

$$1/T = \left(\frac{\partial S}{\partial E} \right)_{V,N} \quad (2.7)$$

If we use the same definition here, we find that

$$\beta = 1/(k_B T) \quad (2.8)$$

Now that we have defined temperature, we can consider what happens if we have a system (denoted by A) that is in thermal equilibrium with a large heat bath (B). The total system is closed; that is, the total energy $E = E_B + E_A$ is fixed (we assume that the system and the bath are weakly coupled, so that we may ignore their interaction energy). Now suppose that the system A is prepared in one specific quantum state i with energy E_i . The bath then has an energy $E_B = E - E_i$ and the degeneracy of the bath is given by $\Omega_B(E - E_i)$. Clearly, the degeneracy of the bath determines the probability P_i to find system A in state i :

$$P_i = \frac{\Omega_B(E - E_i)}{\sum_j \Omega_B(E - E_j)} \quad (2.9)$$

To compute $\Omega_B(E - E_i)$, we expand $\ln \Omega_B(E - E_i)$ around $E_i = 0$:

$$\ln \Omega_B(E - E_i) = \ln \Omega_B(E) - E_i \frac{\partial \ln \Omega_B(E)}{\partial E} + \mathcal{O}(1/E) \quad (2.10)$$

or, using equations (2.6) and (2.7),

$$\ln \Omega_B(E - E_i) = \ln \Omega_B(E) - E_i/k_B T + \mathcal{O}(1/E) \quad (2.11)$$

If we insert this result in equation (2.9), we get

$$P_i = \frac{\exp(-E_i/k_B T)}{\sum_j \exp(-E_j/k_B T)} \quad (2.12)$$

This is the well-known **Boltzmann distribution** for a system at temperature T . Knowledge of the energy distribution allows us to compute the average energy $\langle E \rangle$ of the system at the given temperature T :

$$\langle E \rangle = \sum_i E_i P_i \quad (2.13)$$

$$\begin{aligned} &= \frac{\sum_i E_i \exp(-E_i/k_B T)}{\sum_j \exp(-E_j/k_B T)} \\ &= -\frac{\partial \ln \sum_i \exp(-E_i/k_B T)}{\partial 1/k_B T} \\ &= -\frac{\partial \ln Q}{\partial 1/k_B T} \end{aligned} \quad (2.14)$$

where, in the last line, we have defined the partition function Q . If we compare equation (2.13) with the thermodynamic relation

$$E = \frac{\partial F/T}{\partial 1/T}$$

where F is the Helmholtz free energy, we see that F is related to the partition function Q :

$$F = -k_B T \ln Q = -k_B T \ln \left(\sum_i \exp(-E_i/k_B T) \right) \quad (2.15)$$

Strictly speaking, F is fixed only up to a constant. Or, what amounts to the same thing, the reference point of the energy can be chosen arbitrarily. In what follows, we can use equation (2.15) without loss of generality. The relation between the Helmholtz free energy and the partition function is often more

convenient to use than the relation between $\ln \Omega$ and the entropy. As consequence, equation (2.15) is the workhorse of equilibrium statistical mechanics.

2.3 Classical Statistical Mechanics

Thus far, we have formulated statistical mechanics in purely quantum mechanical terms. The entropy is related to the density of states of a system with E , volume V , and number of particles N . Similarly, the Helmholtz energy is related to the partition function Q , a sum over all quantum states i of the Boltzmann factor $\exp\left(-\frac{E_i}{k_B T}\right)$. To be specific, let us consider the average value of some observable A . We know the probability that a system at temperature T will be found in an energy eigenstate with energy E_i and we can therefore compute the thermal average of A as

$$\langle A \rangle = \frac{\sum_i \exp(-E_i/k_B T) \langle i|A|i \rangle}{\sum_j \exp(-E_j/k_B T)} \quad (2.16)$$

$\langle i|A|i \rangle$ denotes the expectation value of the operator A in quantum state i . This equation suggests how we should go about computing averages: first we solve the Schrodinger equation for the (many-body system of interest, and next we compute the expectation value of A for all those quantum states that have a nonnegligible statistical weight. Unfortunately, this approach is doomed for all but the simplest systems. First of all, we cannot hope to solve the Schrodinger equation for arbitrary many-body systems. And second, even if we could, the number of quantum states that contribute to the average in equation (2.16) would be so astronomically large $\mathcal{O}(10^{10^{25}})$ that a numerical evaluation of all expectation values would be unfeasible. Fortunately, equation (2.16) can be simplified to a more workable expression in the classical limit. To this end, we first rewrite equation (2.16) in a form that is independent of the specific basis set. We note that $\exp\left(-\frac{E}{k_B T}\right) = \langle i|\exp(-\mathcal{H}/k_B T)|i \rangle$, where \mathcal{H} is Hamiltonian of the system. Using this relation, we can write

$$\begin{aligned} \langle A \rangle &= \frac{\sum_i \langle i|\exp(-\mathcal{H}/k_B T)A|i \rangle}{\sum_j \langle j|\exp(-\mathcal{H}/k_B T)|j \rangle} \\ &= \frac{\text{Tr} \exp(-\mathcal{H}/k_B T)A}{\text{Tr} \exp(-\mathcal{H}/k_B T)} \end{aligned} \quad (2.17)$$

where Tr denotes the trace of the operator. As the value of the trace of an operator does not depend on the choice of the basis set, we can compute thermal averages using any basis set we like. Preferably, we use simple basis sets, such as the set of eigenfunctions of the position or the momentum operator. Next, we use the fact that the Hamiltonian \mathcal{H} is the sum of a kinetic part \mathcal{K} and a potential part \mathcal{U} . The kinetic energy operator is a quadratic function of the momenta of all particles. As a consequence, momentum eigenstates are also eigenfunctions of the kinetic energy operator. Similarly,

the potential energy operator is a function of the particle coordinates. Matrix elements of \mathcal{U} therefore are most conveniently computed in a basis set of position eigenfunctions. However, $\mathcal{H} = \mathcal{K} + \mathcal{U}$ itself is not diagonal in either basis set nor is $\exp[-\beta(\mathcal{K} + \mathcal{U})]$. However, if we could replace $\exp(-\beta\mathcal{H})$ by $\exp(-\beta\mathcal{K}) \exp(-\beta\mathcal{U})$, then we could simplify equation (2.17) considerably. In general, we cannot make this replacement because

$$\exp(-\beta\mathcal{K}) \exp(-\beta\mathcal{U}) = \exp\{-\beta[\mathcal{K} + \mathcal{U} + \mathcal{O}([\mathcal{K}, \mathcal{U}])]\} \quad (2.18)$$

where $[\mathcal{K}, \mathcal{U}]$ is the commutator of the kinetic and potential energy operators while $\mathcal{O}([\mathcal{K}, \mathcal{U}])$ is meant to note all terms containing commutators and higher-order commutators of \mathcal{K} and \mathcal{U} . It is easy to verify that the commutator $[\mathcal{K}, \mathcal{U}]$ is of order $\hbar = h/2\pi$, where h is Planck's constant. Hence, in limit $\hbar \rightarrow 0$, we may ignore the terms of order $\mathcal{O}([\mathcal{K}, \mathcal{U}])$. In that case, we can write

$$\text{Tr} \exp(-\beta\mathcal{H}) \approx \text{Tr} \exp(-\beta\mathcal{U}) \exp(-\beta\mathcal{K}) \quad (2.19)$$

If we use the notation $|\mathbf{r}\rangle$ for eigenvectors of the position operator and $|\mathbf{k}\rangle$ for eigenvectors of the momentum operator, we can express equation (2.19) as

$$\text{Tr} \exp(-\beta\mathcal{H}) = \sum_{\mathbf{r}, \mathbf{k}} \langle \mathbf{r} | e^{-\beta\mathcal{U}} | \mathbf{r} \rangle \langle \mathbf{r} | \mathbf{k} \rangle \langle \mathbf{k} | e^{-\beta\mathcal{K}} | \mathbf{k} \rangle \langle \mathbf{k} | \mathbf{r} \rangle \quad (2.20)$$

All matrix elements can be evaluated directly:

$$\langle \mathbf{r} | \exp(-\beta\mathcal{U}) | \mathbf{r} \rangle = \exp[-\beta\mathcal{U}(\mathbf{r}^N)] \quad (2.21)$$

where $\mathcal{U}(\mathbf{r}^N)$ on the right-hand side is no longer an operator but a function of the coordinates of all N particles. Similarly,

$$\langle \mathbf{k} | \exp(-\beta\mathcal{K}) | \mathbf{k} \rangle = \exp\left[-\beta \sum_{i=1}^N p_i^2 / (2m_i)\right] \quad (2.22)$$

where $p_i = \hbar k_i$ and

$$\langle \mathbf{r} | \mathbf{k} \rangle \langle \mathbf{k} | \mathbf{r} \rangle = 1/V^N \quad (2.23)$$

where V is the volume of the system and N the number of particles. Finally, we can replace the sum over states by an integration over all coordinates and momenta. The final result is

$$\begin{aligned} \text{Tr} \exp(-\beta\mathcal{H}) &\approx \frac{1}{h^{dN} N!} \int d\mathbf{p}^N d\mathbf{r}^N \exp\left\{-\beta \left[\sum_i p_i^2 / (2m_i) + \mathcal{U}(\mathbf{r}^N)\right]\right\} \\ &\equiv Q_{\text{classical}} \end{aligned} \quad (2.24)$$

where d is the dimensionality of the system and the last line defines the classical partition function. The factor $1/N!$ has been inserted afterward to take the indistinguishability of identical particles into account. Every N -particle

quantum state corresponds to a volume h^{dN} in classical phase space, but not all such volumes correspond to distinct quantum states. In particular, all points in phase space that only differ in the labeling of the particles correspond to the same quantum state.

Similarly, we can derive the classical limit for $\text{Tr} \exp(-\beta \mathcal{H}\mathcal{A})$, and finally, we can write the classical expression for the thermal average of the observable A as

$$\langle A \rangle = \frac{\int d\mathbf{p}^N d\mathbf{r}^N \exp\left\{-\beta \left[\sum_i p_i^2/(2m_i) + U(\mathbf{r}^N)\right]\right\} A(\mathbf{p}^N, \mathbf{q}^N)}{\int d\mathbf{p}^N d\mathbf{r}^N \exp\left\{-\beta \left[\sum_j p_j^2/(2m_j) + U(\mathbf{r}^N)\right]\right\}} \quad (2.25)$$

Equations (2.24) and (2.25) are the starting point for virtually all classical simulations of many-body systems.

2.4 Ergodicity

So far, we have discussed the average behavior of many-body systems in a purely static sense: we introduced only the assumption that every quantum state of a many-body system with energy E is equally likely to be occupied. Such an average over all possible quantum states of a system is an *ensemble* average. However, this is not the way we usually think about the average behavior of a system. In most experiments we perform a series of measurements during a certain time interval and then determine the average of these measurements. In fact, the idea behind molecular dynamics simulations is precisely that we can study the average behavior of a many-particle system simply by computing the natural time evolution of that system numerically and averaging the quantity of interest over a sufficiently long time. To take a specific example, let us consider a liquid consisting of atoms. Suppose that we wish to compute the average density of the liquid at a distance r from a given atom i , $\rho_i(r)$. Clearly, the instantaneous density depends on the coordinates \mathbf{r}_j of all particles j in the system. As time progresses, the atomic coordinates will change (according to Newton's equations of motion), and hence the density around atom i will change.

Provided that we have specified the initial coordinates and momenta of all atoms $(\mathbf{r}^N(0), \mathbf{p}^N(0))$, we know, at least in principle, the time evolution of $\rho_i(r, \mathbf{r}^N(0), \mathbf{p}^N(0), t)$. In a molecular dynamics simulation, we measure the time-averaged density $\overline{\rho_i(r)}$ of a system of N atoms, in a volume V , at a constant total energy E :

$$\overline{\rho_i(r)} = \lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t dt' \rho_i(r; t') \quad (2.26)$$

Note that, in writing down this equation, we have implicitly assumed that, for t sufficiently long, the time average does not depend on the initial conditions. This is, in fact, a subtle assumption that is not true in general. However, we shall disregard subtleties and simply assume that, once we have specified N , V , and E , time averages do not depend on the initial coordinates and momenta.

If that is so, then we would not change our result for $\overline{\rho_i(r)}$ if we average over many different initial conditions; that is, we consider the hypothetical situation where we run a large number of molecular dynamics simulations at the same values for N , V , and E , but with different initial coordinates and momenta,

$$\overline{\rho_i(r)} = \frac{\sum_{\text{initial conditions}} \left(\lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t dt' \rho_i(r; \mathbf{r}^N(0), \mathbf{p}^N(0), t') \right)}{\text{number of initial conditions}} \quad (2.27)$$

We now consider the limiting case where we average over all initial conditions compatible with the imposed values of N , V , and E . In that case, we can replace the sum over initial conditions by an integral:

$$\frac{\sum_{\text{initial conditions}} f(\mathbf{r}^N(0), \mathbf{p}^N(0))}{\text{number of initial conditions}} \rightarrow \frac{\int_E d\mathbf{r}^N d\mathbf{p}^N f(\mathbf{r}^N(0), \mathbf{p}^N(0))}{\Omega(N, V, E)} \quad (2.28)$$

where f denotes an arbitrary function of the initial coordinates and momenta $\mathbf{r}^N(0), \mathbf{p}^N(0)$, while $\Omega(N, V, E) = \int_E d\mathbf{r}^N d\mathbf{p}^N$ (we have ignored a constant factor¹).

The subscript E on the integral indicates that the integration is restricted to a shell of constant energy E . Such a "phase space" average corresponds to the classical limit of the *ensemble average* discussed in the previous sections². We denote an ensemble average by $\langle \dots \rangle$ to distinguish it from a time average, denoted by a bar. If we switch the order of the time averaging and the averaging over initial conditions, we find

$$\overline{\rho_i(r)} = \lim_{t \rightarrow \infty} \frac{1}{t} \int dt' \langle \rho_i(r; \mathbf{r}^N(0), \mathbf{p}^N(0), t') \rangle_{NVE} \quad (2.29)$$

However, the ensemble average in this equation does not depend on the time t' . This is so, because there is a one-to-one correspondence between the initial phase space coordinates of a system and those that specify the state of the system at a later time t' . Hence, averaging over all initial phase space coordinates is equivalent to averaging over the time evolved phase space coordinates. For this reason, we can leave out the time averaging in equation

¹ If we consider a quantum mechanical system, then $\Omega(N, V, E)$ is simply the number of quantum states of that system, for given N , V , and E . In the classical limit, the number of quantum states of a d -dimensional system of N distinguishable, structureless particles is given by $\Omega(N, V, E) = (\int d\mathbf{p}^N d\mathbf{r}^N)/h^{dN}$. For N *indistinguishable* particles, we should divide the latter expression by a factor $N!$.

² Here we consider the classical equivalent of the so-called microcanonical ensemble, i.e., the ensemble of systems with fixed N , V , and E . The classical expression for phase space integrals in the microcanonical ensemble can be derived from the quantum mechanical expression involving a sum over quantum states in much the same way that we used to derive the classical constant N, V, T ("canonical") ensemble from the corresponding quantum mechanical expression.

(2.29), and we find

$$\overline{\rho_i(r)} = \langle \rho_i(r) \rangle_{NVE} \quad (2.30)$$

This equation states that, if we wish to compute the average of a function of the coordinates and momenta of a many-particle system, we can *either* compute that quantity by time averaging (the "MD" approach) or by ensemble averaging (the "MC" approach). It should be stressed that the preceding paragraphs are meant only to make equation (2.30) *plausible*, not as a proof.

In fact, that would have been quite impossible because equation (2.30) is not true in general. However, in what follows, we shall simply assume that the "*ergodic hypothesis*", as equation (2.30) is usually referred to, applies to the systems that we study in computer simulations. However, you should be aware that many examples of systems are not ergodic *in practice*, such as glasses and metastable phases, or even *in principle*, such as nearly harmonic solids.