CHEM 630P

Statistical Mechanics

Lecture Notes
Part 1: Fundamentals

by Guillaume Lamoureux

Department of Chemistry and Biochemistry
Concordia University
Montréal, Québec, Canada
Email: guillaume.lamoureux@concordia.ca

Compiled on September 23, 2009
Latest version available at http://faculty.concordia.ca/glamoure/teaching.html

1 Probability theory

1.1 Probability distributions and moments

Exercise:
Deduce the probabilities p_i ($i = 2$ to 12) of a random variable X corresponding to the outcome of rolling two ideal dice.

Exercise:
Write a program that generates the probabilities p_i ($i = n$ to $6n$) for the outcome of rolling n dice.

A *random variable* X is a quantity for which we cannot predict with certainty which value it will take every time we measure it, but for which we can tell the *probability* that it takes a given value. If X is a discrete variable that can take any one of n values $x_1, x_2, \ldots, x_n$, the probabilities of each value occuring are defined as $p_1, p_2, \ldots, p_n$. These values can be empirically obtained from a very large number of observations, or deduced from basic assumptions on the nature of the elementary process giving rise to X. If X is a continuous variable, the probability of the outcome of a measurement is described by a continuous probability density $w(x)$, defined such that the probability that the measurement of X falls within the interval $[x, x + dx]$ is equal to $w(x)dx$. Over the full range of values that X can take, those infinitesimal probabilities sum up to one:

$$\text{Normalization:} \qquad \int dx\ w(x) \overset{\text{def}}{=} 1. \tag{1}$$

The mean value (or *expectation value*) of a variable X distributed according to $w(x)$ is

$$\text{Mean value:} \qquad \langle X \rangle = \int dx\ w(x) x. \tag{2}$$

More generally, we define the mth moment of the distribution as the mean value of the mth power of X:

$$m\text{th moment:} \qquad \langle X^m \rangle = \int dx\ w(x) x^m. \tag{3}$$

Wikipedia:
Taylor series

We can write the mean value of any function $f(X)$ in terms of the moments, using a Maclaurin series of the function:

$$\begin{aligned} \langle f(X) \rangle &= \int dx\ w(x) f(x) = \int dx\ w(x) \sum_{m=0}^{\infty} \frac{1}{m!} \left. \frac{\partial^m f}{\partial x^m} \right|_0 x^m \\ &= \sum_{m=0}^{\infty} \frac{1}{m!} \left. \frac{\partial^m f}{\partial x^m} \right|_0 \langle X^m \rangle. \end{aligned} \tag{4}$$

Wikipedia:
Fourier transform

The *characteristic function*, ϕ, is defined as the Fourier transform of the distribution:

$$\phi(k) \overset{\text{def}}{=} \int dx\ w(x) e^{ikx} = \langle e^{ikX} \rangle. \tag{5}$$

$w(x)$, $\phi(k)$, and the moments $\langle X^m \rangle$ contain the same information, and can be obtained from any one of the other two. To summarize, we have

$$\langle X^m \rangle = \int dx\ w(x) x^m = \left. \frac{\partial^m \phi}{\partial (ik)^m} \right|_0, \tag{6}$$

$$\phi(k) = \int dx\, w(x) e^{ikx} = \sum_{m=0}^{\infty} \frac{(ik)^m}{m!} \langle X^m \rangle, \tag{7}$$

$$w(x) = \frac{1}{2\pi} \int dk\, \phi(k) e^{-ikx} = \frac{1}{2\pi} \int dk\, e^{-ikx} \sum_{m=0}^{\infty} \frac{(ik)^m}{m!} \langle X^m \rangle. \tag{8}$$

Exercise:
Verify Eqs. (6) to (8)

The moment $\langle X^m \rangle$ corresponds to the mth coefficient of the MacLaurin series of $\phi(k)$, and for that reason $\phi(k)$ is also called the *moment-generating function*.

Knowing $w_X(x)$, the probability distribution of random variable X, one can calculate $w_Y(y)$, the distribution of any variable $Y = f(X)$:

$$w_Y(y) = \int dx\, w_X(x) \delta(f(x) - y) = \langle \delta(f(X) - y) \rangle, \tag{9}$$

Wikipedia:
Dirac delta function

where δ is Dirac δ-function, defined such that

$$\int dz\, \delta(z - a) g(z) = g(a) \tag{10}$$

for any function g.

Exercise:
Prove that the transformed distribution $w_Y(y)$ is normalized as well.

1.1.1 Binomial distribution

Wikipedia:
Binomial distribution

The *binomial distribution* is the function $B(k; n, p)$ that describes the probability of observing k "successes" in n attempts, if each attempt has a probability p of being successful (and, conversely, a probability $1 - p$ of being unsuccessful).

$$B(k; n, p) \overset{\text{def}}{=} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}. \tag{11}$$

It is a *discrete* distribution, defined only for integer values of k going from 0 to n. By definition, the number of successes from n attempts is any number from 0 to n, and B is normalized such that

$$\sum_{k=0}^{n} B(k; n, p) = 1. \tag{12}$$

The binomial distribution is often encountered in problems related to chance games, but more generally in any system where a global outcome is formed by combining multiple random variables taking discrete values independently from one another. It the context of statistical mechanics, this number of variables is usually very large and the binomial distribution can be approximated by more convenient distributions.

1.1.2 Poisson distribution

Wikipedia:
Poisson distribution

The *Poisson distribution*, $P(k; \lambda)$, is a limiting case of the binomial distribution when n becomes infinitely large, but np remains constant, and equal to λ.

$$P(k; \lambda) \overset{\text{def}}{=} \frac{\lambda^k e^{-\lambda}}{k!}. \tag{13}$$

The Poisson distribution is often used in situations where the "successes" and "failures" happen on a continuous basis—at any point in time, for instance—but the number of "successes" remains *countable*. Since the number of "successful" events can be arbitrarily high, the distribution is defined from $k = 0$ to ∞ but becomes vanishingly small for large values of k. In the context of the statistical mechanics of molecules, these events could be molecular collisions, electronic transitions, or specific conformational changes.

1.1.3 Normal distribution

Wikipedia:
Normal distribution

The *normal distribution*, $N(x; \mu, \sigma^2)$, is another limiting case of the binomial distribution, when n becomes infinitely large.

$$N(x; \mu, \sigma^2) \overset{\text{def}}{=} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2} \tag{14}$$

The normal distribution is used in situations where variable X combines an "uncountable" number of microscopic events—uncountable in the sense that it is very large and that we are not interested in how many "successes" were actually observed, but only in what their overall effect is. The parameters μ and σ are called the *mean* and the *standard deviation* and, in connection to the binomial distribution, correspond to the limits

$$\lim_{n\to\infty} np = \mu \qquad \text{and} \qquad \lim_{n\to\infty} np(1-p) = \sigma^2. \tag{15}$$

The characteristic function of $N(x; \mu, \sigma^2)$ is

$$\phi(k; \mu, \sigma^2) = \langle e^{ikX} \rangle = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{+\infty} dx\ e^{-(x-\mu)^2/2\sigma^2} e^{ikx}. \tag{16}$$

Using the definite integral formula

$$\int_{-\infty}^{+\infty} dx\ e^{-(ax^2+bx+c)} = \sqrt{\frac{\pi}{a}} e^{(b^2-4ac)/4a}, \tag{17}$$

we find

$$\phi(k; \mu, \sigma^2) = e^{ik\mu - k^2\sigma^2/2} = e^{ik\mu + (ik)^2\sigma^2}. \tag{18}$$

The moments of the distribution are therefore

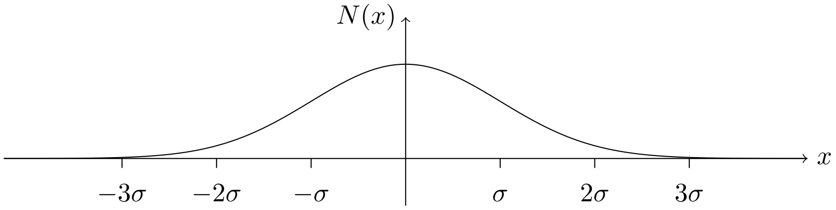
$$\langle X \rangle = \mu, \tag{19}$$
$$\langle X^2 \rangle = \mu^2 + \sigma^2, \tag{20}$$
$$\langle X^3 \rangle = \mu^3 + 3\mu\sigma^2, \tag{21}$$
$$\langle X^4 \rangle = \mu^4 + 6\mu^2\sigma^2 + 3\sigma^4. \tag{22}$$

Exercise:
Derive Equations (19) to (22)

For $\mu = 0$, we find

$$\langle X^m \rangle = \frac{\partial^m}{\partial(ik)^m} e^{(ik)^2\sigma^2}\bigg|_0 = \begin{cases} \dfrac{m!}{2^{m/2}(m/2)!}\sigma^m & \text{for } m \text{ even,} \\ 0 & \text{for } m \text{ odd.} \end{cases} \tag{23}$$

For a normally distributed variable, we expect 68.2% of our measurements to fall within $\pm\sigma$ of the mean, 95.4% fall within $\pm 2\sigma$, and 99.7% fall within $\pm 3\sigma$.



1.2 Cumulants

The previous discussion can be generalized to distributions of multiple variables, where x is replaced by a vector $\mathbf{x} = (x_1, x_2, \ldots)$ and $w(\mathbf{x})d\mathbf{x}$ corresponds to the probability that $\mathbf{X} = (X_1, X_2, \ldots)$ lies in the volume element $d\mathbf{x}$ centered around $\mathbf{x}$. We define the generalized moments of n variables as following:

$$\begin{aligned} \langle X_1^{m_1} \cdots X_n^{m_n} \rangle &= \int dx_1 \cdots dx_n \; w(x_1, \ldots, x_n) x_1^{m_1} \cdots x_n^{m_n} \\ &= \int d\mathbf{x} \; w(\mathbf{x}) \prod_{i=1}^{n} x_i^{m_i}. \end{aligned} \tag{24}$$

The generalized characteristic function is defined as the n-dimensional Fourier transform of w:

$$\phi(\mathbf{k}) \overset{\text{def}}{=} \int d\mathbf{x} \; w(\mathbf{x}) e^{i\mathbf{k}\cdot\mathbf{x}} = \langle e^{i\mathbf{k}\cdot\mathbf{X}} \rangle. \tag{25}$$

It is also the generating function of the moments:

$$\langle X_1^{m_1} \cdots X_n^{m_n} \rangle = \frac{\partial^{m_1}}{\partial(ik_1)^{m_1}} \cdots \frac{\partial^{m_n}}{\partial(ik_n)^{m_n}} \phi(\mathbf{k}) \bigg|_{\mathbf{0}}. \tag{26}$$

We define the *cumulants*, κ (or $\langle \cdots \rangle_{\mathrm{c}}$), as the quantities generated from the logarithm of the moment-generating function $\phi(\mathbf{k})$:

$$\begin{aligned} \kappa_{1,\ldots,n}^{m_1,\ldots,m_n} &\overset{\text{def}}{=} \langle X_1^{m_1} \cdots X_n^{m_n} \rangle_{\mathrm{c}} \\ &\overset{\text{def}}{=} \frac{\partial^{m_1}}{\partial(ik_1)^{m_1}} \cdots \frac{\partial^{m_n}}{\partial(ik_n)^{m_n}} \ln \phi(\mathbf{k}) \bigg|_{\mathbf{0}}. \end{aligned} \tag{27}$$

The cumulants can be used to write an expansion of the generalized moments:

$$\langle X_1 \rangle = \kappa_1, \tag{28}$$

$$\langle X_1 X_2 \rangle = \kappa_1 \kappa_2 + \kappa_{12}, \tag{29}$$

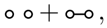
$$\begin{aligned} \langle X_1 X_2 X_3 \rangle &= \kappa_1 \kappa_2 \kappa_3 \\ &+ \kappa_{12}\kappa_3 + \kappa_{13}\kappa_2 + \kappa_{23}\kappa_1 \\ &+ \kappa_{123}, \end{aligned} \tag{30}$$

$$\begin{aligned}\langle X_1X_2X_3X_4\rangle &= \kappa_1\kappa_2\kappa_3\kappa_4\\ &+ \kappa_{12}\kappa_3\kappa_4+\kappa_{13}\kappa_2\kappa_4+\kappa_{14}\kappa_2\kappa_3\\ &\qquad +\kappa_{23}\kappa_1\kappa_4+\kappa_{24}\kappa_1\kappa_3+\kappa_{34}\kappa_1\kappa_2\\ &+ \kappa_{12}\kappa_{34}+\kappa_{13}\kappa_{24}+\kappa_{14}\kappa_{23}\\ &+ \kappa_{123}\kappa_4+\kappa_{124}\kappa_3+\kappa_{134}\kappa_2+\kappa_{234}\kappa_1\\ &+ \kappa_{1234}. \end{aligned} \tag{31}$$

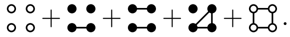
Here, we have simplified the notation by dropping the upper indices ($\kappa_1 = \kappa_1^1$, $\kappa_{12} = \kappa_{12}^{11}$, etc.). There is a method to the madness: orders of the cumulants are formed in combinations that sum up to the order of the moment (e.g., for the 4th-order moment, we have combinations $\kappa_i\kappa_j\kappa_k\kappa_l$, $\kappa_{ij}\kappa_{kl}$, $\kappa_{ijk}\kappa_l$, and κ_{ijkl}) and for each combination of orders, all distinct permutations of the variables are represented. Using a diagrammatic notation, we can write

Exercise:
Derive Eqs. (28) to (30) from the definitions of the moments and cumulants.

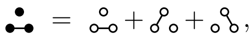
$\langle X_1\rangle =$  , (32)

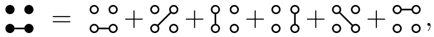
$\langle X_1X_2\rangle =$  , (33)

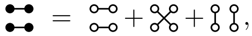
$\langle X_1X_2X_3\rangle =$  , (34)

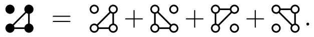
$\langle X_1X_2X_3X_4\rangle =$  . (35)

The "full" diagrams represent the sum of all variable permutations:

 , (36)

 , (37)

 , (38)

 . (39)

For a univariate distribution, we have

$$\begin{aligned}\langle X\rangle &= \langle X\rangle_{\mathrm{c}}, &(40)\\ \langle X^2\rangle &= \langle X\rangle_{\mathrm{c}}^2+\langle X^2\rangle_{\mathrm{c}}, &(41)\\ \langle X^3\rangle &= \langle X\rangle_{\mathrm{c}}^3+3\langle X^2\rangle_{\mathrm{c}}\langle X\rangle_{\mathrm{c}}+\langle X^3\rangle_{\mathrm{c}}, &(42)\\ \langle X^4\rangle &= \langle X\rangle_{\mathrm{c}}^4+6\langle X^2\rangle_{\mathrm{c}}\langle X\rangle_{\mathrm{c}}^2+3\langle X^2\rangle_{\mathrm{c}}^2+4\langle X^3\rangle_{\mathrm{c}}\langle X\rangle_{\mathrm{c}}+\langle X^4\rangle_{\mathrm{c}}. &(43)\end{aligned}$$

The second cumulant represents the average of the squared fluctuations around the mean:

$$\langle X_iX_j\rangle_{\mathrm{c}} = \langle X_iX_j\rangle - \langle X_i\rangle\langle X_j\rangle = \langle(X_i-\langle X_i\rangle)(X_j-\langle X_j\rangle)\rangle. \tag{44}$$

1.2.1 Independent variables

Cumulants have a few properties that make them very important for the theory of statistical mechanics. If the moments of two independent random variables X_1 and X_2 can be generated from the characteristic functions $\phi_{X_1}(k)$ and $\phi_{X_2}(k)$, the moments of the sum $Y = X_1 + X_2$ can be generated with the product of the functions:

$$\phi_Y(k) = \phi_{X_1}(k)\phi_{X_2}(k). \tag{45}$$

Exercise:
Prove Equation (45)

If their cumulants are generated from functions $K_{X_1}(k) = \ln \phi_{X_1}(k)$ and $K_{X_2}(k) = \ln \phi_{X_2}(k)$, the cumulant of the sum can be generated from the *sum* of the generating functions:

$$K_Y(k) = K_{X_1}(k) + K_{X_2}(k). \tag{46}$$

A consequence of this is that the mth cumulant of a sum is the sum of the mth cumulants:

$$\text{Additivity:} \qquad \langle X_1 + X_2|^m\rangle_\mathrm{c} = \langle X_1^m\rangle_\mathrm{c} + \langle X_2^m\rangle_\mathrm{c}. \tag{47}$$

Wikipedia:
Extensive quantity

No such simple relationship exists for the moments. Equation (47) means that the cumulants of an extensive quantity are extensive quantities as well. The cumulants also have the following properties (for any constants a and b):

$$\begin{aligned} \text{Equivariance:} \quad & \langle X + b\rangle_\mathrm{c} = \langle X\rangle_\mathrm{c} + b, & (48)\\ \text{Invariance:} \quad & \langle X + b|^m\rangle_\mathrm{c} = \langle X^m\rangle_\mathrm{c} \quad \text{for } m \geq 2, & (49)\\ \text{Homogeneity:} \quad & \langle aX|^m\rangle_\mathrm{c} = a^m \langle X^m\rangle_\mathrm{c}. & (50) \end{aligned}$$

More generally, the multidimentional distribution of n independent variables, $w(\mathbf{x})$, can be written as the product of n individual distributions:

$$w(\mathbf{x}) = \prod_{i=1}^{n} w_i(x_i), \tag{51}$$

and the moment- and cumulant-generating functions are simple combinations of the individual generating functions:

$$\phi(\mathbf{k}) = \prod_{i=1}^{n} \phi_i(k_i) \qquad \text{and} \qquad K(\mathbf{k}) = \sum_{i=1}^{n} K_i(k_i) = \sum_{i=1}^{n} \ln \phi_i(k_i). \tag{52}$$

The generalized moments become simple products of the individual moments:

$$\langle X_1^{m_1} \cdots X_n^{m_n}\rangle = \langle X_1^{m_1}\rangle \cdots \langle X_n^{m_n}\rangle \tag{53}$$

and the generalized cumulants $\langle X_1^{m_1} \cdots X_n^{m_n}\rangle_\mathrm{c}$ are nonzero only if no more that one of the m's is greater than zero. (In other words, all mixed cumulants are zero, which is why cumulants are additive in the first place.) More generally, the average of a product of any two functions of independent variables is the product of the individual averages:

$$\langle f(x_i) g(x_j)\rangle = \langle f(x_i)\rangle \langle g(x_j)\rangle \quad \text{if } i \neq j. \tag{54}$$

1.3 Central limit theorem

Wikipedia:
Central limit theorem

The central limit theorem states that the sum of a large number of identically distributed random variables has a normal distribution, regardless of the shape of the distribution of those individual variables. We can prove the theorem by imagining a random variable Y corresponding to the sum of n random variables X drawn from the same probability distribution $w(x)$. But first, let us discuss some more properties of the normal distribution.

1.3.1 Normal distribution

From Eq. (18), we find that the cumulant-generating function of the univariate normal distribution is

$$K(k) = ik\mu + (ik)^2\sigma^2. \tag{55}$$

Because $K(k)$ is a second-degree polynomial in ik, the only two nonzero cumulants are $\kappa_1 = \mu$ and $\kappa_2 = \sigma^2$. For an n-dimensional system, the normal distribution has the form

$$N(\mathbf{x}; \boldsymbol{\mu}, \mathbf{S}) = \frac{1}{(2\pi)^{n/2}} \frac{1}{\sqrt{\det \mathbf{S}}} e^{-(\mathbf{x}-\boldsymbol{\mu})\cdot\mathbf{S}^{-1}\cdot(\mathbf{x}-\boldsymbol{\mu})/2} \tag{56}$$

where $\boldsymbol{\mu}$ is an n-dimensional vector corresponding to $\langle \mathbf{X} \rangle$ and $\mathbf{S}$ is the $n \times n$ *covariance* matrix ($S_{ij} = \langle X_i X_j \rangle_\text{c}$). If all variables are independent, the covariance matrix is diagonal, its determinant is the product of the diagonal elements (σ_i^2), and

$$N(\mathbf{x}; \boldsymbol{\mu}, \mathbf{S}) = \prod_{i=1}^{n} \frac{1}{\sigma_i\sqrt{2\pi}} e^{-(x-\mu_i)^2/2\sigma_i^2} = \prod_{i=1}^{n} N(x_i; \mu_i, \sigma_i^2). \tag{57}$$

Note that it is usually possible to find a linear transformation $\mathbf{x}' = \mathbf{U}\mathbf{x}$ that diagonalizes the covariance matrix and therefore allows the distribution to be written as a product of n normal distributions of the independent variables $\mathbf{x}'$.

1.3.2 Proof

The cumulants of a variable Y corresponding to the average of n independent variables $X_1, \ldots, X_n$ having the same probability distribution are

$$\langle Y^m \rangle_\text{c} = \frac{1}{n^m}\left(\langle X_1^m \rangle_\text{c} + \cdots + \langle X_n^m \rangle_\text{c}\right) = \frac{1}{n^{m-1}} \langle X^m \rangle_\text{c}. \tag{58}$$

As n gets arbitrarily large, all cumulants except κ_1 vanish, which means—as one would expect—that the distribution of Y becomes an arbitrarily narrow peak centered at $\langle Y \rangle$. This is commonly referred to as the *law of large numbers*: the average of any quantity can be calculated to an arbitrarily high precision using an arbitrarily large number of measurements.

Wikipedia:
Law of large numbers

If Y corresponds to the *standardized* sum of n independent variables:

$$Y = \frac{1}{\sqrt{n}}(X_1 + X_2 + \cdots + X_n), \tag{59}$$

its cumulants are

$$\langle Y^m \rangle_\text{c} = \frac{1}{n^{m/2}}\left(\langle X_1^m \rangle_\text{c} + \cdots + \langle X_n^m \rangle_\text{c}\right) = \frac{1}{n^{m/2-1}} \langle X^m \rangle_\text{c}. \tag{60}$$

As n gets arbitrarily large, only the two first cumulants survive ($\langle Y \rangle_\text{c} = \sqrt{n}\langle X \rangle_\text{c}$ and $\langle Y^2 \rangle_\text{c} = \langle X^2 \rangle_\text{c}$), which means that Y is normally distributed with a mean $\mu = \sqrt{n}\langle X \rangle_\text{c} = \sqrt{n}\langle X \rangle$ and a variance $\sigma^2 = \langle X^2 \rangle_\text{c} = \langle X^2 \rangle - \langle X \rangle^2$. This proves the central limit theorem—provided we accept the assumption that the only distribution that has $\kappa_i = 0$ for $i > 2$ is the normal distribution.

There is a more general statement of the theorem, which holds for "locally correlated" variables as well. This is an important generalization, since most variables encountered in statistical mechanics are not strictly independent, but simply dependent on a small number of "neighboring" variables. The sum of a large number of such variables will also have a normal distribution.

2 Classical mechanics

2.1 Newtonian, Lagrangian, and Hamiltonian formulations

Reading:
Section 1-2 of McQuarrie [1976]

Wikipedia:
Newton's laws of motion

The most common description of the evolution of a system of N classical particles over time is Newton's second law, which relates the instantaneous accelation of each particle (labelled $i = 1, 2, \ldots, N$) to the force applied on it:

$$\boxed{m_i \mathbf{a}_i = \mathbf{F}_i.} \tag{61}$$

For a large number of interacting particles, these equations are used as following:

1. Knowing the positions $\mathbf{x}(t)$ and velocities $\mathbf{v}(t)$ of all particles at a given time t, and provided with the *force law* that these particles obey, we calculate the atomic forces $\mathbf{F}_i(t)$;
2. The atomic accelerations $\mathbf{a}_i(t)$ are obtained from $\mathbf{F}_i(t)/m_i$;
3. Those accelerations are used to calculate the positions and velocities at a later time $t + \delta t$;
4. The procedure is repeated from $t + \delta t$ to $t + 2\delta t$, and so on.

The "$F = ma$" form of Newton's equation requires the accelerations, velocities, and positions—as well as the force law—to be described in Cartesian coordinates:

$$m_i \ddot{x}_i = F_i^x, \tag{62}$$
$$m_i \ddot{y}_i = F_i^y, \tag{63}$$
$$m_i \ddot{z}_i = F_i^z. \tag{64}$$

Any other coordinate system requires these $3N$ equations to be re-written in a completely different form—which is not always very convenient. There exist two main alternative formulations based not on the forces and accelerations—which are vectors, and depend on the coordinate system—but on scalar functions of *generalized coordinates*.

Wikipedia:
Generalized coordinates

The *Lagrangian formulation* of the equations of motion is defined in terms of $3N$ generalized coordinates q:

$$\boxed{\frac{d}{dt}\left(\frac{\partial \mathscr{L}}{\partial \dot{q}_j}\right) = \frac{\partial \mathscr{L}}{\partial q_j}.} \tag{65}$$

(Here, the index j runs from 1 to $3N$.) The function $\mathscr{L}$ is the *Lagrangian* of the system, defined as the total kinetic energy of the system (K), minus its total potential energy (U):

$$\mathscr{L}(q, \dot{q}) \overset{\text{def}}{=} K - U. \tag{66}$$

K is typically a function of the generalized velocities ($\dot{q}$) only, and U is typically a function of the generalized coordinates (q) only—but not necessarily. The coordinates q can be the usual $3N$ Cartesian coordinates or any other coordinates that uniquely define the configuration of the system.

The *Hamiltonian formulation* of the equations of motion is defined in terms of $3N$ generalized coordinates, q, and $3N$ *generalized momenta*, p, defined as following:

$$p_j \overset{\text{def}}{=} \frac{\partial \mathscr{L}}{\partial \dot{q}_j}. \tag{67}$$

Hamilton's equations are

$$\boxed{\dot{q}_j = \frac{\partial \mathscr{H}}{\partial p_j} \qquad \text{and} \qquad \dot{p}_j = -\frac{\partial \mathscr{H}}{\partial q_j}.} \tag{68}$$

The function $\mathscr{H}$ is the *Hamiltonian* of the system, defined as the following Legendre transform of the Lagrangian:

Wikipedia:
Legendre transform

$$\mathscr{H}(q, p) \overset{\text{def}}{=} \sum_j p_j \dot{q}_j - \mathscr{L}(q, \dot{q}). \tag{69}$$

(Notice that the Legendre transform makes $\mathscr{H}$ a function of the generalized momenta p instead of the generalized velocities $\dot{q}$.) If K is quadratic in terms of the generalized velocities and if U depends only on the generalized coordinates, the Hamiltonian corresponds to the total energy of the system (E):

Exercise:
Formulate the classic "projectile motion" and "harmonic oscillator" problems using the Newtonian, Lagrangian, and Hamiltonian approaches. Show that they lead to the same equations of motion.

$$\mathscr{H} = K + U = E. \tag{70}$$

The time derivative of any property A function of q, p, and t can be expressed as following (using the chain rule):

$$\begin{aligned} \frac{dA}{dt} &= \sum_j \left(\frac{\partial A}{\partial q_j}\dot{q}_j + \frac{\partial A}{\partial p_j}\dot{p}_j\right) + \frac{\partial A}{\partial t} \\ &= \sum_j \left(\frac{\partial A}{\partial q_j}\frac{\partial \mathscr{H}}{\partial p_j} - \frac{\partial A}{\partial p_j}\frac{\partial \mathscr{H}}{\partial q_j}\right) + \frac{\partial A}{\partial t} \\ &= \{\mathscr{H}, A\} + \frac{\partial A}{\partial t}, \end{aligned} \tag{71}$$

where we have used the Poisson bracket notation of two functions $A(q, p, t)$ and

Wikipedia:
Poisson bracket

$B(q,p,t)$:

$$\{A,B\} \overset{\text{def}}{=} \sum_j \left(\frac{\partial A}{\partial q_j}\frac{\partial B}{\partial p_j} - \frac{\partial A}{\partial p_j}\frac{\partial B}{\partial q_j} \right) = -\{B,A\}. \tag{72}$$

2.2 Liouville's theorem

We introduce a function $\rho(q,p,t)$ called the *phase space distribution*, defined such that $\rho(q,p,t)\,dq\,dp$ is proportional to the probability of finding a system at time t within the volume element $dq\,dp$ around point (q,p), knowing *only* that this system evolves according to a given Hamiltonian $\mathscr{H}$—not knowing what its initial conditions were, for instance. The time evolution of ρ can be calculated from Eq. (71):

$$\frac{d\rho}{dt} = \{\mathscr{H},\rho\} + \frac{\partial \rho}{\partial t}. \tag{73}$$

There is a subtle but essential difference in the meaning of quantities $d\rho/dt$ and $\partial\rho/\partial t$. On one hand, the partial derivative $\partial\rho/\partial t$ represents the rate of flow of the trajectories through the volume element around a fixed point (q,p). This flow can be expected to be faster in some regions of the phase space that in other. On the other hand, the total derivative $d\rho/dt$ represents the variation of ρ around a point $(q(t),p(t))$ moving with time.

Wikipedia:
Liouville's theorem (Hamiltonian)

Liouville's theorem states that the phase space distribution function is constant over time:

$$\frac{d\rho}{dt} = 0. \tag{74}$$

In other words, if instead of propagating a single point (q_0,p_0) over time we propagate all trajectories in a differential element $dq\,dp$ around this point, we observe that, although the shape of this infinitesimal element may change considerably over time, its actual *volume* does not. The density of possible trajectories contained in that element is constant over time. If trajectories "fan out" across one direction, they will get closer in other directions. At any fixed point in phase space, the trajectory flow is therefore given by

$$\frac{\partial\rho}{\partial t} = \{\rho,\mathscr{H}\}. \tag{75}$$

Any system that is not under the influence of an external force has a phase space distribution that does not explicitly depend on time—it depends only on its internal variables q and p—and therefore $\partial\rho/\partial t = 0$ and

$$\{\rho,\mathscr{H}\} = 0, \tag{76}$$

Wikipedia:
Constant of motion

which means that ρ is a constant of motion and is a function of the Hamiltonian only: $\rho = \rho(\mathscr{H}(q,p))$. Such a system is said to be in *thermodynamic equilibrium*.

We define the *partition function*, Z, as the integral of ρ over the entire phase space:

$$Z \overset{\text{def}}{=} \int dq\, dp\ \rho(\mathscr{H}(q,p)). \tag{77}$$

It represents the number of distinct possible trajectories obeying the Hamiltonian. The average of any quantity $A = A(q,p)$ over all possible trajectories obeying $\mathscr{H}$ corresponds to

$$\langle A \rangle = \frac{1}{Z} \int dq\, dp\ \rho(\mathscr{H}(q,p)) A(q,p). \tag{78}$$

At first glance, the partition function seems to serve strictly as a normalization factor, that could be absorbed by defining a probability distribution in phase space $w \overset{\text{def}}{=} \rho/Z$. We will see later on that Z is also a powerful moment-generating function—and its logarithm, a cumulant-generating function.

3 Review of thermodynamics

3.1 Internal energy

The total energetic content of a system composed of a large number of particles is represented by its *internal energy*, E. If we exert work on such a system, and bring it from thermodynamic conditions X_i to thermodynamic conditions X_f, its internal energy will increase. If those constraints X_f are relaxed, and the system is left to return to the original conditions X_i, a fraction of the internal energy accumulated during the i → f process is reconverted into work but, unless the system is otherwise isolated from the rest of the world, a fraction is dissipated as heat.

One recovers the maximum amount of work by letting the conditions X_f go back to their initial values X_i slowly enough that the system is at all time in a state of quasi-equilibrium—such a process is called *quasi-static*. The full amount of work initially done on the system can be recovered upon only if both the i→f and f→i processes are quasi-static, and *only* if the system interacts with its environment through the constraint X—and not through any other thermodynamic condition X'.

For example, the work exerted on a piston to compress a confined gas from a volume V_i to a volume $V_\mathrm{f} < V_\mathrm{i}$ can be fully recovered only if (1) the piston and the cylinder are made of a perfectly insulating material (no heat exchange with the environment) and if (2) the piston is perfectly airtight (no gas particles going in or out of the volume), and *only* if (3) both the compression and the decompression are performed infinitely slowly. (Here, we also presume that the piston moves inside the cylinder without any friction.) Only under these ideal conditions will an amount of work ΔW be totally transferred to the internal energy of the gas ($E_\mathrm{f} = E_\mathrm{i} + \Delta W$), and totally restored upon decompression.

From a strictly thermodynamic point of view, the state of a system is governed by the fact that the internal energy is an *extensive function* of the other

Wikipedia:
Extensive quantity

extensive variables describing the system, such as its entropy S_B, its volume V, and the number of particles N it is composed of:

$$E = E(S_\mathrm{B}, V, N). \tag{79}$$

A system λ times bigger will have λN particles, a volume λV, and an entropy λS_B, and its internal energy will be λ times larger. In other words,

$$E(\lambda S_\mathrm{B}, \lambda V, \lambda N) = \lambda E(S_\mathrm{B}, V, N). \tag{80}$$

Wikipedia:
Euler's homogeneous function theorem

Because of Euler's theorem, such an extensive function obeys the following equation:

$$E(S_\mathrm{B}, V, N) = \left(\frac{\partial E}{\partial S_\mathrm{B}}\right)_{V,N} S_\mathrm{B} + \left(\frac{\partial E}{\partial V}\right)_{N,S_\mathrm{B}} V + \left(\frac{\partial E}{\partial N}\right)_{S_\mathrm{B},V} N. \tag{81}$$

Wikipedia:
Intensive quantity

Since E is extensive, each of the three terms should be extensive as well, and the coefficients of S_B, V, and N are necessarily *intensive*. We define those conjugate intensive variables as follows:

$$T \overset{\text{def}}{=} \left(\frac{\partial E}{\partial S_\mathrm{B}}\right)_{V,N}, \quad -p \overset{\text{def}}{=} \left(\frac{\partial E}{\partial V}\right)_{N,S_\mathrm{B}}, \quad \mu \overset{\text{def}}{=} \left(\frac{\partial E}{\partial N}\right)_{S_\mathrm{B},V}, \tag{82}$$

corresponding to the temperature, the pressure (*minus* the pressure, actually), and the chemical potential of the system. We therefore have the following fundamental equation:

$$E(S_\mathrm{B}, V, N) = T S_\mathrm{B} - pV + \mu N. \tag{83}$$

It was said earlier that the function E depends only on the extensive variables, therefore its exact differential can be written as

$$dE = T\,dS_\mathrm{B} - p\,dV + \mu\,dN. \tag{84}$$

In other words, the internal energy of a system can be increased either by heating it ($dS_\mathrm{B} > 0$), by compressing it ($dV < 0$), or by inserting particles ($dN > 0$)—or by doing a combination of the three. Those three variables define the thermodynamic conditions of the system—what we previously called "X".

A single number of particles N is sufficient to describe the equilibrium state of the system only of all particles are identical. In general, we need as many "composition" variables as we have chemical species in the system: N_1, N_2, etc. Equation (83) then becomes

$$E(S_\mathrm{B}, V, N_1, N_2, \ldots) = T S_\mathrm{B} - pV + \sum_i \mu_i N_i. \tag{85}$$

This accounts for the fact that each individual molecule of chemical species i contributes to the total internal energy by an amount μ_i, depending on the species.

For a microscopic point of view, the internal energy corresponds to the sum of the potential energy (U) and the kinetic energy (K) of all particles composing the system:

$$E = U + K. \tag{86}$$

There may seem to be a profound disconnection between this microscopic definition of E and the thermodynamic formula of Eq. (85). One of the purposes of statistical mechanics is actually to provide a connection between those two equations.

3.2 Thermodynamic functions and relations

3.2.1 Helmholtz free energy

Wikipedia: Legendre transform

We define *Helmholtz potential*, F, as the following Legendre transform of internal energy E:

$$\begin{aligned} F(T, V, N) &\overset{\text{def}}{=} E(S_{\mathrm{B}}, V, N) - T S_{\mathrm{B}} \qquad (87) \\ &= -pV + \mu N. \end{aligned}$$

In general, the Legendre transform of an extensive function E of an extensive variable X_i forms a new extensive function E_i that is independent of X_i but that depends on its conjugate *intensive* variable $\xi_i \overset{\text{def}}{=} \partial E / \partial X_i$:

$$E_i(\xi_i) \overset{\text{def}}{=} E(X_i) - \xi_i X_i. \tag{88}$$

Helmholtz potential depends on the temperature instead of the entropy, and is well suited to describe the state of a system kept at constant temperature and volume. For any isothermal (and non-diffusive) process, its variation corresponds to the mechanical work provided by a compression:

$$dF = -\underbrace{S_{\mathrm{B}}\, dT}_{0} - p\, dV + \underbrace{\mu\, dN}_{0} = -p\, dV. \tag{89}$$

It also correspond to the work that gets ideally restored when the system spontaneously dilates. A system at constant temperature—and constant V and N—is in thermodynamic equilibrium if $dF = 0$.

3.2.2 Gibbs free energy

We define *Gibbs potential*, G, as the following double Legendre transform of the internal energy:

$$\begin{aligned} G(T, -p, N) &\overset{\text{def}}{=} F(T, V, N) + pV \qquad (90) \\ &= \mu N. \qquad (91) \end{aligned}$$

Gibbs potential is well suited to describe the state of a system kept at constant temperature and pressure. For any isothermal and isobaric (and non-diffusive) process, its variation is strictly zero:

$$dG = -\underbrace{S_{\mathrm{B}}\,dT}_{0} + \underbrace{V\,dp}_{0} - \underbrace{\mu\,dN}_{0} = 0. \tag{92}$$

In other words, it is the conserved quantity in any process happening at constant pressure and temperature—which is the most common situation in chemistry and biochemistry. For *diffusive* processes (for which $dN \neq 0$), it corresponds to the energy required to insert of remove particles from the system. A system at constant temperature and pressure—and constant N—is in thermodynamic equilibrium if $dG = 0$.

Thermodynamic functions F and G are called *free energies* because they represent the maximum amount of energy theoretically available as work according to a given thermodynamical constraint. In itself, a spontaneous reaction $A \to B$ liberates an internal energy ΔE corresponding to the difference in kinetic and potential energy between the reactants and the products. However, if the system in which the reaction happens is kept at constant temperature, only $\Delta F = \Delta E - T\Delta S_{\mathrm{B}}$ (Helmholtz free energy) can be converted into work: a contribution $T\Delta S_{\mathrm{B}}$ goes into maintaining the system at constant temperature. At constant temperature and pressure, as it is usually the case for *in vivo* biochemical processes, only $\Delta G = \Delta E - T\Delta S_{\mathrm{B}} + p\Delta V$ (Gibbs free energy) is available: an additional contribution $-p\Delta V$ goes into maintaining the system at constant pressure. In living organisms, the energy stored in a chemical bond or in a molecular affinity is measured in "units" of G—which makes Gibbs free energy the fundamental energetic "currency" of life.

3.2.3 Reformulation based on entropy

To establish a connection with statistical mechanics, it is useful to reformulate the definitions of all previous thermodynamic functions using *entropy* as the fundamental extensive function—instead of the internal energy. By writing entropy as a dimensionless quantity $S \overset{\mathrm{def}}{=} S_{\mathrm{B}}/k_{\mathrm{B}}$ and by defining $\beta \overset{\mathrm{def}}{=} 1/k_{\mathrm{B}}T$ we have, according to Eq. (83),

$$S(E, V, N) = \beta E + \beta pV - \beta\mu N. \tag{93}$$

In this new equation, the conjugate intensive variables are β, βp, and $-\beta\mu$ (equivalent to variables T, $-p$, and μ used previously). The equivalent to Helmholtz free energy is $-\beta F$, defined as

$$\begin{aligned} -\beta F(\beta, V, N) &\overset{\mathrm{def}}{=} S(E, V, N) - \beta E \\ &= \beta pV - \beta\mu N, \end{aligned} \tag{94}$$

and the equivalent to Gibbs free energy is $-\beta G$, defined as

$$-\beta G(\beta, \beta p, N) \overset{\mathrm{def}}{=} -\beta F(\beta, V, N) - \beta pV \tag{95}$$

$$= -\beta\mu N, \tag{96}$$

Depending on the thermodynamic constraints exerted on the system, its equilibrium state corresponds to extrema in either S, $-\beta F$, or $-\beta G$.

- In the *microcanonical* ensemble (E, V, and N constant), equilibrium is attained when entropy S is maximum with respect to any virtual variation of the conditions E, V, or N.
- In the *canonical* ensemble (β, V, and N constant), equilibrium is attained when $-\beta F$ is maximum with respect to β, V, and N.
- In the *isothermal-isobaric* ensemble (β, βp, and N constant), equilibrium is attained when $-\beta G$ is maximum with respect to N but minimum with respect to β and βp.

In the more common parlance of the Legendre transforms of the internal energy, thermodynamic equilibrium is attained when, at constant temperature, the Helmholtz free energy F decreases for any change in T and increases for any change in V or N and when, at constant temperature and pressure, the Gibbs free energy G decreases for any change in T or p and increases for any change in N.

4 Equilibrium ensembles and fluctuations

Reading:
Chapter 3 of Chandler [1987]

4.1 Microcanonical ensemble

We now consider the thermodynamic system discussed in Section 3 as a fluid of N identical particles described by $3N$ generalized coordinates q and their $3N$ conjugate momenta p. For a given configuration of the system, the instantaneous internal energy is equal to its Hamiltonian $\mathscr{H}(q,p)$. If those N particles are contained in a fixed volume V in thermal insulation, the internal energy of the system is equal to a constant E and the probability density of finding the system at point (q,p) in phase space is

$$w(q,p;E) = \frac{\delta(\mathscr{H}(q,p)-E)}{\displaystyle\int dq\,dp\;\delta(\mathscr{H}(q,p)-E)}. \tag{97}$$

Obviously, the probability is nonzero only for states that have an internal energy consistent with the thermodynamic constraint. What is less obvious, however, is that *all* points consistent with the constraint $\mathscr{H}(q,p) = E$ are equiprobable. This assumption, called the *fundamental postulate* of statistical mechanics, amounts to saying that an arbitrary large collection of systems prepared at random [started from various initial conditions such that $\mathscr{H}(q_0,p_0) = E$] will collectively spend the same amount of time in all microscopic states consistent with the thermodynamic constraints, and therefore that all states will contribute equally to the average properties of the system—whether they are forming a single trajectory or not.

A stronger version of that postulate, called the *ergodic hypothesis*, assumes

Wikipedia:
Ergodic hypothesis

that a *single* system will eventually visit all microscopic states consistent with the constraints. This would imply that, instead of accumulating statistics over a large number of randomly prepared systems, one could simply measure a single system for a long enough time. While this is a reasonable assumption for systems that are large and chaotic, it is not true in general.

The phase space distribution of the system is given by

$$\rho(q,p;E) = C_N \delta(\mathscr{H}(q,p) - E). \tag{98}$$

Wikipedia:
Action (physics)

Wikipedia:
Plank constant

The constant C_N ensures that $\delta(\mathscr{H} - E)\, dq\, dp$ is unitless and size-consistent. Function $\delta(\mathscr{H} - E)$ has units of inverse energy and should be multiplied by an arbitrary reference energy E_0. Each surface element $dq_i dp_i$ has units of action (because q_i and p_i are conjugate variables) and, to make contact with the more general *quantum* version of the same distribution, each one is divided by Plank constant, h—which corresponds to the quantum of action.

$$C_N \overset{\text{def}}{=} \frac{1}{N!}\frac{E_0}{h^{3N}}. \tag{99}$$

The factor $1/N!$ accounts for the fact that any of the $N!$ permutations of the N identical particles corresponds to the same state of the system, which should be counted only once. This factor would have a different form if the particles were not all identical or if the system was described by quantum mechanics instead of classical mechanics. The value of E_0 has no incidence on any of the properties of the system.

The collection of random identical systems distributed according to Eq. (98) is called the *microcanonical ensemble* of the Hamiltonian $\mathscr{H}$. The partition function for this ensemble is

$$\begin{aligned}\Omega(E,V,N) &= \int dq\, dp\ \rho(q,p;E) \\ &= C_N \int dq\, dp\ \delta(\mathscr{H}(q,p) - E).\end{aligned} \tag{100}$$

In principle, the $3N$ "coordinate" integrals are performed over the entire space, but it is assumed that those integrals have no contribution outside volume V, presumably because a single particle going through the walls of the container would require the system to have an energy higher that E. Conventionally, the volume dependence of Ω is represented by the domain of integration for the $3N$ "coordinate" integrals. For a system with discrete energy levels, the microcanonical partition function is simply the total number of micro-states that have an energy E:

$$\Omega(E,V,N) = n(E). \tag{101}$$

The entropy S is related to the microcanonical partition function Ω by the simple equation

$$\boxed{e^S = \Omega(E,V,N).} \tag{102}$$

This equation is the fundamental connection between thermodynamics and statistical mechanics. It is commonly written as $S_\mathrm{B} = k_\mathrm{B} \ln \Omega$ and serves as a definition of the thermodynamic entropy.

Wikipedia:
Boltzmann constant

According to conventional thermodynamics, the temperature of the system is defined in terms of the change of entropy under a (virtual) variation of the internal energy. Using the connection between the entropy and the microcanonical partition function, we have

$$\frac{1}{T} \overset{\text{def}}{=} \left(\frac{\partial S_\mathrm{B}}{\partial E}\right)_{V,N} = \left(\frac{\partial (k_\mathrm{B} \ln \Omega)}{\partial E}\right)_{V,N} . \tag{103}$$

Equivalently, we can write this definition in terms of the inverse temperature $\beta = 1/k_\mathrm{B} T$:

$$\beta \overset{\text{def}}{=} \left(\frac{\partial S}{\partial E}\right)_{V,N} = \left(\frac{\partial (\ln \Omega)}{\partial E}\right)_{V,N} . \tag{104}$$

4.2 Canonical ensemble

Let us now consider a system of N particles contained in a fixed volume V, but kept at constant temperature T through heat exchange with a second system that acts as a *heat bath*. This heat bath is considered large enough that any energy transfer with the system leaves its temperature unaffected—and equal to T. The probability density of finding such a system at point (q, p) in phase space is

$$w(q, p; \beta) = \frac{e^{-\beta \mathscr{H}(q,p)}}{\displaystyle\int dq\, dp\; e^{-\beta \mathscr{H}(q,p)}}, \tag{105}$$

were β is the inverse temperature of the heat bath. The combination "$e^{-\beta E}$" (where E is the energy of the system) is often called the "Boltzmann factor", and Eq. (105) simply states that the probability of finding a system in state (q, p) is proportional to the Boltzmann factor of that state. (The denominator is simply providing the correct normalization.) For a system existing only in *discrete* states labelled $\nu = 0, 1, 2$, etc. and having energies E_ν, the probability of each state is

$$p_\nu(\beta) = \frac{e^{-\beta E_\nu}}{\displaystyle\sum_\nu e^{-\beta E_\nu}}. \tag{106}$$

The expression for the probability in the microcanonical ensemble, proportional to $\delta(\mathscr{H} - E)$, was a direct consequence of the fundamental postulate, but this new "$e^{-\beta \mathscr{H}}$" expression needs to be proven.

We consider that the heat bath itself is formed of N_b particles of a type different from the first N particles. The Hamiltonian of the overall system is the sum of the Hamiltonians of the two subsystems, plus a term $\mathscr{W}$ describing their thermal interaction:

$$\mathscr{H}'(\omega, \omega_\mathrm{b}) = \mathscr{H}(\omega) + \mathscr{H}_\mathrm{b}(\omega_\mathrm{b}) + \mathscr{W}(\omega, \omega_\mathrm{b}). \tag{107}$$

To shorten the notation, we have represented (q, p) as ω and $(q_\mathrm{b}, p_\mathrm{b})$ as ω_b. If we assume that the subsystems are very weakly coupled, we can neglect $\mathscr{W}$ and write $\mathscr{H}' = \mathscr{H} + \mathscr{H}_\mathrm{b}$. The internal energy of the overall system is E'. The energies of the subsystems may fluctuate, but their sum is always $E + E_\mathrm{b} = E'$. The microcanonical distribution function of the overall system is

$$\rho'(\omega, \omega_\mathrm{b}; E') = C_N C_{N_\mathrm{b}} \delta(\mathscr{H}(\omega) + \mathscr{H}_\mathrm{b}(\omega_\mathrm{b}) - E') \tag{108}$$

and the corresponding microcanonical partition function is

$$\begin{aligned} \Omega'(E', V', N') &= \int d\omega \int d\omega_\mathrm{b} \; \rho'(\omega, \omega_\mathrm{b}; E') \\ &= C_N C_{N_\mathrm{b}} \int d\omega \int d\omega_\mathrm{b} \; \delta(\mathscr{H}(\omega) + \mathscr{H}_\mathrm{b}(\omega_\mathrm{b}) - E'). \end{aligned} \tag{109}$$

We can write ρ, the distribution function for the system of N particles by integrating the full distribution ρ' over the "bath" variables, and normalizing:

$$\begin{aligned} \rho(\omega; E') &= \frac{\displaystyle\int d\omega_\mathrm{b} \; \rho'(\omega, \omega_\mathrm{b}; E')}{\displaystyle\int d\omega \int d\omega_\mathrm{b} \; \rho'(\omega, \omega_\mathrm{b}; E')} \\ &= \frac{C_N C_{N_\mathrm{b}} \displaystyle\int d\omega_\mathrm{b} \; \delta(\mathscr{H}(\omega) + \mathscr{H}_\mathrm{b}(\omega_\mathrm{b}) - E')}{C_N C_{N_\mathrm{b}} \displaystyle\int d\omega \int d\omega_\mathrm{b} \; \delta(\mathscr{H}(\omega) + \mathscr{H}_\mathrm{b}(\omega_\mathrm{b}) - E')} \\ &= \frac{C_N}{\Omega'(E')} \Omega_\mathrm{b}(E' - \mathscr{H}(\omega)). \end{aligned} \tag{110}$$

We write the formula for $\ln \rho$ as a Taylor expansion around E':

$$\begin{aligned} \ln \rho(\omega; E') &= \ln C_N - \ln \Omega'(E') + \ln \Omega_\mathrm{b}(E' - \mathscr{H}(\omega)) \\ &= \ln C_N - \ln \Omega'(E') + \ln \Omega_\mathrm{b}(E') - \mathscr{H}(\omega) \left. \frac{\partial(\ln \Omega_\mathrm{b})}{\partial E_\mathrm{b}} \right|_{E'} \\ &\quad + \tfrac{1}{2} \mathscr{H}^2(\omega) \left. \frac{\partial^2(\ln \Omega_\mathrm{b})}{\partial E_\mathrm{b}^2} \right|_{E'} - \ldots \end{aligned} \tag{111}$$

The first three terms are independent of the configuration ω of the system, and can be assimilated to a single constant, $\ln C$. Because the heat bath is a very large system, $\partial(\ln \Omega_\mathrm{b})/\partial E_\mathrm{b}$ is practically equal to $\partial(\ln \Omega')/\partial E'$, which corresponds to β. Coefficient $\partial^2(\ln \Omega_\mathrm{b})/\partial E_\mathrm{b}^2$ is negligible. We will see later that it corresponds to the energy fluctuations of the heat bath, $\langle E_\mathrm{b}^2 \rangle_\mathrm{c}$, which are necessarily equal to $\langle E^2 \rangle_\mathrm{c}$, the energy fluctuations of the system itself (because $E + E_\mathrm{b}$

is constant), and which become negligible compared to E in the thermodynamic limit $N \to \infty$. In the thermodynamic limit, we have

$$\ln \rho(\omega; \beta) = \ln C - \beta \mathscr{H}(\omega). \tag{112}$$

Consequently, $\rho(\omega; \beta) = Ce^{-\beta \mathscr{H}(\omega)}$, which translates into Eq. (105) once constant C is fixed by the normalization condition.

We define Z, the *canonical* partition function of the system, as

$$Z(\beta, V, N) = C_N \int dq\, dp\; e^{-\beta \mathscr{H}(q,p)}. \tag{113}$$

Although we are not interested in using it, there is a similar partition function for the heat bath: $Z_{\mathrm{b}}(\beta, V_{\mathrm{b}}, N_{\mathrm{b}})$. The two subsystems have distinct volumes and numbers of particles, but because they are in thermal contact their thermodynamic equilibria are defined by a common variable, β.

The canonical partition function formally corresponds to the Laplace transform of $\bar{\Omega} \stackrel{\mathrm{def}}{=} d\Omega/dE$, the *density of states*:

Wikipedia:
Laplace transform

$$Z(\beta, V, N) = \int dE\; \bar{\Omega}(E, V, N) e^{-\beta E}. \tag{114}$$

For a system with discrete energy levels E_l,

$$Z(\beta, V, N) = \sum_l n(E_l) e^{-\beta E_l}, \tag{115}$$

where $n(E_l)$ is the number of micro-states having an energy E_l. If the sum is performed on *states* ν rather than energy levels l,

$$Z(\beta, V, N) = \sum_\nu e^{-\beta E_\nu}. \tag{116}$$

The canonical partition function Z is related to Helmholtz free energy, $-\beta F$, in the same way the microcanonical partition function Ω is related to the entropy:

$$e^{-\beta F} = Z(\beta, V, N). \tag{117}$$

Because $-\beta F = \ln Z$ (which itself has the form of a characteristic function), the cumulants of the internal energy can be generated from the derivatives of $-\beta F$ with respect to $-\beta$. In particular,

$$\langle E \rangle_{\mathrm{c}} = \langle E \rangle = \left(\frac{\partial(-\beta F)}{\partial(-\beta)} \right)_{V,N} = -\left(\frac{\partial \ln Z}{\partial \beta} \right)_{V,N}, \tag{118}$$

$$\langle E^2 \rangle_{\mathrm{c}} = \langle E^2 \rangle - \langle E \rangle^2 = \left(\frac{\partial^2(-\beta F)}{\partial(-\beta)^2} \right)_{V,N} = \left(\frac{\partial^2 \ln Z}{\partial \beta^2} \right)_{V,N}. \tag{119}$$

4.3 Generalized ensembles

As we have shown, the partition function of the *canonical* ensemble is

$$Z(\beta, V, N) = \int dE\ \bar{\Omega}(E, V, N) e^{-\beta E} \tag{120}$$

and corresponds to the number of states of a system in thermal equilibrium with a heat bath. The procedure used to derive this expression can be generalized to any type of thermodynamic interactions. In general, the variation of entropy of a system in interaction with its environment—which we will call a “reservoir”—may have any of the following contributions [see Eq. (93)]:

$$dS = \beta\, dE + \beta p\, dV - \beta\mu\, dN, \tag{121}$$

or more generally

$$dS = \beta\, dE + \sum_i \xi_i\, dX_i, \tag{122}$$

where X_i are extensive variables (V, N, etc.) and ξ_i are their conjugate intensive variables (βp, $-\beta\mu$, etc.). While all extensive variables involved in the interaction may fluctuate within the subsystem, they will remain constant for the overall system, and the microcanonical partition function of this overall system can be used to generate the partition functions of each of the subsystems.

4.3.1 Isothermal-isobaric ensemble

The *isothermal-isobaric* ensemble defines a system in thermal and mechanical interaction with a reservoir ($dE \neq 0$ and $dV \neq 0$). Its partition function is

$$\Delta(\beta, \beta p, N) \overset{\text{def}}{=} \int dV\ \bar{Z}(\beta, V, N) e^{-\beta p V}, \tag{123}$$

where $\bar{Z}$ is the canonical partition function obtained using a density of states $\bar{\Omega} \overset{\text{def}}{=} d^2\Omega/dE\, dV$. In terms of the density of states,

$$\begin{aligned} \Delta(\beta, \beta p, N) &= \int dE\, dV\ \bar{\Omega}(E, V, N) e^{-\beta E} e^{-\beta p V} \\ &= \int dE\, dV\ \bar{\Omega}(E, V, N) e^{-\beta H}, \end{aligned} \tag{124}$$

where H is the instantaneous enthalpy of the system ($H = E + pV$).

The isothermal-isobaric partition function Δ is related to Gibbs free energy as following:

$$e^{-\beta G} = \Delta(\beta, \beta p, N). \tag{125}$$

The combination $-\beta G$ can be used as a generating function of the cumulants of the volume:

$$\langle V\rangle_{\mathrm{c}} \;=\; \langle V\rangle = \left(\frac{\partial(-\beta G)}{\partial(-\beta p)}\right)_{\beta,N} = -\left(\frac{\partial \ln\Delta}{\partial\beta p}\right)_{\beta,N}, \tag{126}$$

$$\langle V^2\rangle_{\mathrm{c}} \;=\; \langle V^2\rangle - \langle V\rangle^2 = \left(\frac{\partial^2(-\beta G)}{\partial(-\beta p)^2}\right)_{\beta,N} = \left(\frac{\partial^2 \ln\Delta}{\partial(\beta p)^2}\right)_{\beta,N}. \tag{127}$$

It can also be used to generate the cumulants of the enthalpy $(E + pV)$:

$$\langle H\rangle_{\mathrm{c}} \;=\; \langle H\rangle = \left(\frac{\partial(-\beta G)}{\partial(-\beta)}\right)_{\beta p,N} = -\left(\frac{\partial \ln\Delta}{\partial\beta}\right)_{\beta p,N}, \tag{128}$$

$$\langle H^2\rangle_{\mathrm{c}} \;=\; \langle H^2\rangle - \langle H\rangle^2 = \left(\frac{\partial^2(-\beta G)}{\partial(-\beta)^2}\right)_{\beta p,N} = \left(\frac{\partial^2 \ln\Delta}{\partial\beta^2}\right)_{\beta p,N}. \tag{129}$$

Therefore, the average internal energy of the system is

$$\langle E\rangle = \langle H\rangle - p\langle V\rangle = -\left(\frac{\partial \ln\Delta}{\partial\beta}\right)_{\beta p,N} + p\left(\frac{\partial \ln\Delta}{\partial\beta p}\right)_{\beta,N}. \tag{130}$$

4.3.2 Grand-canonical ensemble

The *grand-canonical* ensemble defines a system in thermal and diffusive interaction with a reservoir ($dE \neq 0$ and $dN \neq 0$). Its partition function is

$$\Xi(\beta, V, -\beta\mu) = \int dN\;\bar{Z}(\beta, V, N)e^{\beta\mu N}. \tag{131}$$

where $\bar{Z}$ is the canonical partition function obtained using a density of states $\bar{\Omega} \overset{\mathrm{def}}{=} d^2\Omega/dE\,dN$. We can also write

$$\Xi(\beta, V, -\beta\mu) = \int dE\,dN\;\bar{\Omega}(E, V, N)e^{-\beta E}e^{\beta\mu N} \tag{132}$$

For a system with discrete states ν,

$$\Xi(\beta, V, -\beta\mu) = \sum_\nu e^{-\beta E_\nu}e^{\beta\mu N_\nu}. \tag{133}$$

The thermodynamic potential of the grand-canonical ensemble, sometimes called the “grand potential”, is defined as

$$e^{-\beta\Phi} \overset{\mathrm{def}}{=} \Xi(\beta, V, -\beta\mu). \tag{134}$$

The potential $-\beta\Phi$ is a function of β, V, and $\beta\mu$. By analogy with Eq. (95), it is equal to $-\beta F + \beta\mu N$, and therefore equal to $\beta p V$. It can be used as a generating function of the cumulants of the number of particles:

$$\langle N\rangle_{\mathrm{c}} \;=\; \langle N\rangle = \left(\frac{\partial(-\beta\Phi)}{\partial\beta\mu}\right)_{\beta,V} = \left(\frac{\partial \ln\Xi}{\partial\beta\mu}\right)_{\beta,V}, \tag{135}$$

$$\langle N^2\rangle_{\mathrm{c}} \;=\; \langle N^2\rangle - \langle N\rangle^2 = \left(\frac{\partial^2(-\beta\Phi)}{\partial(\beta\mu)^2}\right)_{\beta,V} = \left(\frac{\partial^2 \ln\Xi}{\partial(\beta\mu)^2}\right)_{\beta,V}. \tag{136}$$

It can also be used to generate the cumulants of $E - \mu N$:

$$\begin{aligned}
\langle E - \mu N \rangle_{\mathrm{c}} &= \left(\frac{\partial(-\beta\Phi)}{\partial(-\beta)} \right)_{V,\beta\mu} = -\left(\frac{\partial \ln \Xi}{\partial \beta} \right)_{V,\beta\mu}, && (137) \\
\langle E - \mu N |^2 \rangle_{\mathrm{c}} &= \left(\frac{\partial^2(-\beta\Phi)}{\partial(-\beta)^2} \right)_{V,\beta\mu} = \left(\frac{\partial^2 \ln \Xi}{\partial \beta^2} \right)_{V,\beta\mu}. && (138)
\end{aligned}$$

Therefore, the average internal energy of the system is

$$\langle E \rangle = \langle E - \mu N \rangle + \mu \langle N \rangle = -\left(\frac{\partial \ln \Xi}{\partial \beta} \right)_{V,\beta\mu} + \mu \left(\frac{\partial \ln \Xi}{\partial \beta\mu} \right)_{\beta,V}. \tag{139}$$

4.4 Potentials as generating functions

The connection between the thermodynamic potentials S, $-\beta F$, $-\beta G$, and $-\beta\Phi$ and the partition functions from statistical mechanics allows us to use a given potential $\phi(\xi_i)$ as a generating function of the cumulants of the corresponding extensive variable X_i:

$$\langle X_i^m \rangle_{\mathrm{c}} = \left(\frac{\partial^m \phi}{\partial(-\xi_i)^m} \right)_{\neg\xi_i}. \tag{140}$$

The "$\neg\xi_i$" ("not ξ_i") notation means that all variables other than ξ_i are held constant. (We use the notation "X" for an extensive variable and "ξ" for its conjugate intensive variable.)

In practical terms, this means that many properties of the system can be calculated—or measured—from derivatives of the free energies relative to temperature, pressure, or chemical potential. We have seen that, for a system at constant temperature, the internal energy can be obtained from Helmholtz free energy:

$$\langle E \rangle_{\mathrm{c}} = \langle E \rangle = \left(\frac{\partial(-\beta F)}{\partial(-\beta)} \right)_{V,N}. \tag{141}$$

The heat capacity at constant volume, C_V, defined as the derivative of $\langle E \rangle$ with respect to temperature, can therefore be expressed in terms of the variance of the energy:

$$\begin{aligned}
C_V &\overset{\mathrm{def}}{=} \left(\frac{\partial \langle E \rangle}{\partial T} \right)_{V,N} = \frac{1}{k_{\mathrm{B}} T^2} \left(\frac{\partial \langle E \rangle}{\partial(-\beta)} \right)_{V,N} \\
&= \frac{1}{k_{\mathrm{B}} T^2} \left(\frac{\partial^2(-\beta F)}{\partial(-\beta)^2} \right)_{V,N} = \frac{\langle E^2 \rangle_{\mathrm{c}}}{k_{\mathrm{B}} T^2}. && (142)
\end{aligned}$$

Similarly, for a system at constant temperature and pressure, the enthalpy can be obtained from Gibbs free energy:

$$\langle H \rangle_{\mathrm{c}} = \langle H \rangle = \left(\frac{\partial(-\beta G)}{\partial(-\beta)} \right)_{\beta p,N}, \tag{143}$$

and the heat capacity at constant pressure, C_p, can be expressed in terms of the variance of the enthalpy:

$$\begin{aligned} C_p &\overset{\text{def}}{=} \left(\frac{\partial\langle H\rangle}{\partial T}\right)_{\beta p,N} = \frac{1}{k_{\mathrm{B}}T^2}\left(\frac{\partial\langle H\rangle}{\partial(-\beta)}\right)_{\beta p,N} \\ &= \frac{1}{k_{\mathrm{B}}T^2}\left(\frac{\partial^2(-\beta G)}{\partial(-\beta)^2}\right)_{\beta p,N} = \frac{\langle H^2\rangle_{\mathrm{c}}}{k_{\mathrm{B}}T^2}. \end{aligned} \tag{144}$$

In other words, the heat capacity is related to enthalpy fluctuations at constant pressure and to *energy* fluctuations at constant temperature.

The isothermal compressibility, κ, is defined as

$$\kappa \overset{\text{def}}{=} -\frac{1}{V}\left(\frac{\partial V}{\partial p}\right)_T. \tag{145}$$

It can be expressed in terms of the volume fluctuations in the isothermal-isobaric ensemble:

$$\begin{aligned} \kappa &\overset{\text{def}}{=} -\frac{1}{\langle V\rangle}\left(\frac{\partial\langle V\rangle}{\partial p}\right)_{\beta,N} = \frac{\beta}{\langle V\rangle}\left(\frac{\partial\langle V\rangle}{\partial(-\beta p)}\right)_{\beta,N} \\ &= \frac{\beta}{\langle V\rangle}\left(\frac{\partial^2(-\beta G)}{\partial(-\beta p)^2}\right)_{\beta,N} = \frac{\beta}{\langle V\rangle}\langle V^2\rangle_{\mathrm{c}}. \end{aligned} \tag{146}$$

References

David Chandler. *Introduction to Modern Statistical Mechanics.* Oxford University Press, New York, 1987.

John Gunn. *Notes de cours de mécanique statistique.* Département de chimie, Université de Montréal, 1998.

Shang-Keng Ma. *Statistical Mechanics.* World Scientific, Singapore, 1985.

Donald A. McQuarrie. *Statistical Mechanics.* Harper & Row, New York, 1976.

Franz Schwabl. *Statistical Mechanics.* Springer-Verlag, Berlin, 2002.

Mark E. Tuckerman. *Lecture notes on Statistical Mechanics.* Department of Chemistry, New York University, 2003.