

Statistical mechanics of classical systems

States and ensembles

- A *microstate* of a statistical system is specified by the complete information about the *states of all microscopic degrees of freedom* of the system. It is impossible in practice to measure or mathematically express the enormous amount of information contained in a microstate. However, this concept is essential for the theoretical formulation of statistical mechanics.
- A *macrostate* of a statistical system represents the observable macroscopic properties of the system, and may be specified by quantities such as temperature T , pressure p , volume V , internal energy E , magnetization \mathbf{M} , etc. Descriptions of non-equilibrium are also of interest, using macroscopic quantities such as position and time dependent temperature $T(\mathbf{r}, t)$, pressure $p(\mathbf{r}, t)$, current density of particle flow $\mathbf{j}(\mathbf{r}, t)$, etc.
- Statistical mechanics studies the dynamics of macroscopic systems with many degrees of freedom starting from the microscopic point of view. The observable *macroscopic* behavior of the system is derived by applying the principles of statistics to the dynamics of all *microscopic* degrees of freedom. In other words, the basic goal is to obtain equations for the evolution of the *macrostate* variables from the known fundamental dynamics and statistics of *microstates*.
- Microstates of classical particles (in a fluid or solid) are typically specified by the instantaneous positions \mathbf{r}_i and momenta \mathbf{p}_i (or velocities $\mathbf{v}_i = \mathbf{p}_i/m_i$) of all particles. One represents a microstate of N particles by a $6N$ -dimensional vector:

$$\boldsymbol{\Omega} = (\mathbf{r}_1, \mathbf{p}_1, \mathbf{r}_2, \mathbf{p}_2, \dots, \mathbf{r}_N, \mathbf{p}_N)$$

- In many cases one needs to generalize this description. For example, particles may have internal degrees of freedom (vibrational and rotational modes of molecules) that need to be included in $\boldsymbol{\Omega}$. If one wants to describe a system of magnets that are free to rotate but not change their location in space, one may want to use $\boldsymbol{\Omega} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \dots, \boldsymbol{\mu}_N)$, where $\boldsymbol{\mu}_i$ is the magnetic moment of the i^{th} magnet. If the number of particles N can change, then it must be included in the definition of $\boldsymbol{\Omega}$ (whose dimensionality is then variable).
- Discrete classical systems have microstates $\boldsymbol{\Omega} = (s_1, s_2, \dots, s_N)$, where s_i is some discrete variable that indicates the state of i^{th} component. For example, two-state variables $s_i \in \{0, 1\}$ would be appropriate for describing a system of many switches in a thermal equilibrium.
- The vector space spanned by all possible microstate vectors $\boldsymbol{\Omega}$ is called *phase space*.
- **Statistical ensemble and microstate probability:** The time evolution of a classical macroscopic system is a trajectory $\boldsymbol{\Omega}(t)$ in phase space. One could in principle determine $\boldsymbol{\Omega}(t)$ from the initial state $\boldsymbol{\Omega}(0)$ by solving the appropriate equations of motion for all of the system's degrees of freedom. This is, however, futile because the system of equations is too large and there is no hope to measure the initial state. Instead, we set up a statistical description of dynamics by defining a statistical ensemble and the probability of microstates.
 - Imagine preparing $\mathcal{N} \rightarrow \infty$ copies of the system through some realistic process in which one controls a small number of the systems' properties. For example, one may control volume, temperature and the number of particles, and make those properties the same in each copy of the system. The obtained collection of systems is called *ensemble*. Since one cannot control the many microscopic degrees of freedom, each member of the ensemble will generally have a different microstate consistent with the controlled parameters. The phase space points $\boldsymbol{\Omega}$ corresponding to the ensemble microstates should span the entire phase space more or less evenly.
 - For continuum systems of particles, consider a small phase space volume:

$$d\Omega = \prod_{i=1}^N d^3r_i d^3p_i$$

centered at the point $\mathbf{\Omega}$ in phase space. Count the number $d\mathcal{N}(\mathbf{\Omega})$ of ensemble microstates within $d\Omega$. The probability density of a microstate $\mathbf{\Omega}$ is:

$$f(\mathbf{\Omega}) = \lim_{\mathcal{N} \rightarrow \infty} \frac{1}{\mathcal{N}} \frac{d\mathcal{N}(\mathbf{\Omega})}{d\Omega}$$

$f(\mathbf{\Omega})$ is a PDF that depends on many random scalar variables. If we realistically prepare a single macroscopic system, then $f(\mathbf{\Omega})d\Omega$ is the probability that the system will be within the phase space volume $d\Omega$ of the microstate $\mathbf{\Omega}$. As any PDF, $f(\mathbf{\Omega})$ is normalized:

$$\int d\Omega f(\mathbf{\Omega}) = 1$$

- For discrete systems whose microstates are $\mathbf{\Omega} = (s_1, s_2, \dots, s_N)$, one can work directly with normalized microstate probabilities $p(\mathbf{\Omega})$:

$$\sum_{\mathbf{\Omega}} p(\mathbf{\Omega}) = 1$$

Equilibrium

- The state of equilibrium is a *macrostate* that does not change in time. This implies that macroscopic state variables do not change in time. The *microstate*, however, may change in time as long as it remains consistent with the given static macrostate.
- **Postulate about the equilibrium probability distribution:** The probability distribution of ensemble microstates must adequately reflect our knowledge about the system's condition. If we knew exactly all details about the system's microstate $\mathbf{\Omega}_0$, then such knowledge would be captured by a perfectly sharp and deterministic distribution:

$$f(\mathbf{\Omega}) = \delta(\mathbf{\Omega} - \mathbf{\Omega}_0) \quad \text{for continuum systems} \quad ; \quad p(\mathbf{\Omega}) = \begin{cases} 1 & , \quad \mathbf{\Omega} = \mathbf{\Omega}_0 \\ 0 & , \quad \mathbf{\Omega} \neq \mathbf{\Omega}_0 \end{cases} \quad \text{for discrete systems}$$

where $\delta(\mathbf{\Omega} - \mathbf{\Omega}_0)$ is the Dirac delta function defined by:

$$\delta(\mathbf{\Omega} - \mathbf{\Omega}_0) = \prod_i \delta(x_i - x_{0i}) \delta(y_i - y_{0i}) \delta(z_i - z_{0i}) \delta(p_{xi} - p_{0xi}) \delta(p_{yi} - p_{0yi}) \delta(p_{zi} - p_{0zi})$$

$$\delta(x) = \begin{cases} 0 & , \quad x \neq 0 \\ \infty & , \quad x = 0 \end{cases} \quad , \quad \int_{-\infty}^{\infty} dx \delta(x) = 1$$

The above distribution is completely biased toward a single microstate $\mathbf{\Omega}_0$. However, in reality we never have so much information about a macroscopic system. Information that we have in an equilibrium is given by state variables such as temperature T , pressure p , chemical potential μ , total energy H , volume V , the number of particles N , etc. Since we have no knowledge about anything else, we should not arbitrarily introduce it into the probability distribution. In other words, the probability distribution should depend only on the quantities we know, and be completely unbiased toward any quantities we do not know - microstate in particular:

$$\begin{aligned} f(\mathbf{\Omega}) &= f\left(T, p, \mu \dots ; H(\mathbf{\Omega}), V(\mathbf{\Omega}), N(\mathbf{\Omega}) \dots\right) && \text{for continuum systems in equilibrium} \\ p(\mathbf{\Omega}) &= p\left(T, p, \mu \dots ; H(\mathbf{\Omega}), V(\mathbf{\Omega}), N(\mathbf{\Omega}) \dots\right) && \text{for discrete systems in equilibrium} \end{aligned}$$

- We anticipated that extensive state variables (like total energy H) depend explicitly on the microstate $\mathbf{\Omega}$, but required that the PDF f or probability p depend on $\mathbf{\Omega}$ only implicitly through macroscopic state variables. Knowing nothing about microstates, all microstates with the same energy, etc. appear equally likely.

- The postulated equilibrium probability distribution must be tested in experiments. In practice, all predictions of this postulate match the experiments because there are so many *microstates* corresponding to the *same macrostate*. The detailed time evolution from one microstate to another looks completely chaotic. One indeed cannot tell anything about which particular *microstate* realizes the observed *macrostate* at any instant of observation (i.e. there is no significant bias).
- The implicit dependence of the equilibrium PDF on total energy, $f(\Omega) = f(H(\Omega))$, can be rigorously derived from the equations of motion. We will show this later.
- We used the notation H for total energy to emphasize its difference from internal energy E . The former is a random state variable $H(\Omega)$ that depends on the microstate, while the latter is the average value $E = \langle H \rangle$. Also, the total energy expressed as a function $H(\Omega)$ of the system's state is called Hamiltonian in mechanics.

- **Microcanonical ensemble** represents perfectly isolated systems in equilibrium. The total energy E of an isolated system is conserved, so that the PDF or probability is:

$$f(\Omega) = C\delta(H(\Omega)-E) \quad \text{continuum} \quad ; \quad p(\Omega) = C'\delta_{H(\Omega),H} \equiv C' \begin{cases} 1 & , H(\Omega) = E \\ 0 & , H(\Omega) \neq E \end{cases} \quad \text{discrete}$$

where C, C' are normalization constants, $\delta(x)$ is the Dirac delta function, and $\delta_{x,y}$ is Kronecker symbol (the discrete analogue of Dirac function). The system's volume and number of particles are also fixed (but we emphasize only the PDF's dependence on energy).

- **Canonical ensemble** represents the systems in equilibrium at given temperature T . The canonical PDF or probability is:

$$f(H(\Omega)) = Ce^{-\beta H(\Omega)} \quad \text{continuum} \quad ; \quad p(\Omega) = C'e^{-\beta H(\Omega)} \quad \text{discrete}$$

where C, C' are normalization constants, and β is a temperature-dependent factor with units of inverse energy (we will show later that $\beta = 1/k_B T$). This PDF is known as Boltzmann distribution. The system exchanges energy with its environment (so it is not conserved), but keeps its volume and particle number fixed.

- One can derive the canonical PDF in the following manner. Consider two systems A and B in contact and thermal equilibrium with each other. Regard A as a member of a canonical ensemble $\mathcal{E}_A(T)$ and B as a member of a separate but equivalent canonical ensemble $\mathcal{E}_B(T)$. There is a unique functional form of the PDF f for all canonical ensembles. It is parametrized by temperature T and cannot explicitly depend on the microstate (since we have equilibrium). It can, however, depend on a few extensive microstate functions such as system's total energy H , volume V , number of particles N , etc. Consequently, the PDFs for finding the systems A and B in particular microstates are $f(T; H_A, V_A, N_A \dots)$ and $f(T; H_B, V_B, N_B \dots)$ respectively. Now consider the joint system $A + B$. Since energy can be exchanged between its parts, $A + B$ is another equilibrium system belonging to the same type of canonical ensemble $\mathcal{E}_{A+B}(T)$. The PDF of $A + B$ is therefore $f(T; H_A + H_B, V_A + V_B, N_A + N_B \dots)$. However, the degrees of freedom that belong to A are completely independent from the degrees of freedom that belong to B (particles are not exchanged between A and B). The properties of probability then ensure:

$$f(T; H_A + H_B, V_A + V_B, N_A + N_B \dots) d\Omega_A d\Omega_B = f(T; H_A, V_A, N_A \dots) d\Omega_A \times f(T; H_B, V_B, N_B \dots) d\Omega_B$$

Cancel the phase space volume factors $d\Omega$ and take the logarithm:

$$\log f(T; H_A + H_B, V_A + V_B, N_A + N_B \dots) = \log f(T; H_A, V_A, N_A \dots) + \log f(T; H_B, V_B, N_B \dots)$$

The most general functional form that satisfies this equation is:

$$\log f(T; H, V, N \dots) = -\beta H - \gamma V - \nu N - \dots$$

where the coefficients β, γ, ν are parameters that characterize the established equilibrium in which energy, forces, particles, etc. can be exchanged with the environment. Specifically, the need to introduce the parameter β can be taken as evidence that temperature exists as a state variable; β is some function of temperature, fixed only by the precise definition of temperature. If we consume $\gamma V + \nu N + \dots$ into the normalization constant C , then:

$$f(T; H, V, N \dots) = C(T; V, N \dots) \times e^{-\beta H}$$

- **Grand canonical ensemble** represents the systems in equilibrium at given temperature T and in contact with a particle reservoir at chemical potential μ (the systems exchange both energy and particles with their environment). The grand canonical PDF or probability is:

$$f(H(\Omega), N) = C e^{-\beta(H(\Omega) - \mu N)} \quad \text{continuum} \quad ; \quad p(\Omega) = C' e^{-\beta(H(\Omega) - \mu N)} \quad \text{discrete}$$

where C, C' are normalization constants, N is the number of particles, and β is a temperature-dependent factor with units of inverse energy.

- The proof is analogous to that for the canonical distribution, but the dependence on N is kept outside of the normalization constant: $f = C \times e^{-\beta H - \nu N}$. Then we may write $\nu = -\beta' \mu$, where β' has the same units as β , in order to obtain the correct units for ω : $f = C e^{-\beta H + \beta' \mu N}$. Lastly, we identify $\beta' = \beta$, so the exponent becomes proportional to $E = H - \mu N$. This is justified by the fact that chemical potential μ is potential energy of a particle that has been taken from the particle reservoir into the system. It is a real potential energy because it can be returned to the reservoir at any time by returning the particle. Therefore, we may regard E as the true total energy of the system, which includes the potential energy with respect to the reservoir. The system with energy E is still in thermal equilibrium, so it should have a canonical PDF ($\beta' = \beta$).

- **Density and number of states:** The equilibrium probability must be normalized over the entire phase space:

$$\int d\Omega f(H(\Omega)) = 1 \quad \text{continuum} \quad ; \quad \sum_{\Omega} p(H(\Omega)) = 1 \quad \text{discrete}$$

Given that equilibrium probability distributions depend only on energy, we can reorganize the normalization conditions as integrals/sums over energy. The weight function of energy in these and similar integrals/sums is the density/number of states.

- In continuum systems, we substitute the identity

$$f(H(\Omega)) = \int dH \delta(H - H(\Omega)) f(H)$$

in the normalization condition:

$$\int d\Omega f(H(\Omega)) = \int dH \int d\Omega \delta(H(\Omega) - H) f(H) = \int dH \omega(H) f(H) = 1$$

$$\omega(H) = \int d\Omega \delta(H(\Omega) - H)$$

The function $\omega(H)$ is the *density of microstates* at energy H ; $\omega(H)dH$ is the *phase space volume* occupied by the microstates whose energy lies in the interval $(H, H + dH)$.

- In discrete systems, we substitute the identity

$$p(H(\Omega)) = \sum_H \delta_{H, H(\Omega)} p(H)$$

in the normalization condition:

$$\sum_{\Omega} p(H(\Omega)) = \sum_H \sum_{\Omega} \delta_{H(\Omega),H} p(H) = \sum_H \Omega(H) p(H) = 1$$

$$\Omega(H) = \sum_{\Omega} \delta_{H(\Omega),H}$$

The function $\Omega(H)$ is the *number of microstates* whose energy is H .

- **Partition function:** The normalization of the canonical ensemble PDF yields:

$$\int dH \omega(H) f(H) = C \int dH \omega(H) e^{-\beta H} = 1 \quad \Rightarrow \quad C = \frac{1}{Z} \quad , \quad Z = \int dH \omega(H) e^{-\beta H}$$

where $\omega(H)$ is the density of states, and Z is the so called partition function. Exactly the same form of the normalization constant C' is obtained for discrete canonical distributions:

$$\sum_H \Omega(H) p(H) = C' \sum_H \Omega(H) e^{-\beta H} = 1 \quad \Rightarrow \quad C' = \frac{1}{Z} \quad , \quad Z = \sum_H \Omega(H) e^{-\beta H}$$

- Partition function depends on various state variables, which are then constrained and related to the internal energy by the normalization condition.

Single and multi-particle probability distributions

- For the sake of generality, let us represent the microstate vector $\Omega = (\omega_1, \omega_2 \dots \omega_N)$ in terms of the state vectors ω_i of individual degrees of freedom. If the system is a fluid or solid of N particles, we would use $\omega_i = (\mathbf{r}_i, \mathbf{p}_i)$, but more generally these could be other continuous or discrete quantities.
- Referring to a single degree of freedom as *particle*, we define the single-particle PDF or probability as:

$$f_1(\omega_i) = \int d\Omega' f(\Omega') \delta(\omega'_i - \omega_i) \quad (\text{continuous}) \quad ; \quad p_1(\omega_i) = \sum_{\Omega'} p(\Omega') \delta_{\omega'_i \omega_i} \quad (\text{discrete})$$

using the microstate PDF $f(\Omega)$ or probability $p(\Omega)$. Note that we used Dirac function $\delta(x - y)$ or Kronecker symbol δ_{xy} (equal to 1 only if $x = y$, otherwise zero) to constrain the integration/summation to only those microstates Ω' whose i^{th} particle's state is fixed to ω_i . The functions f_1 and p_1 capture the statistics of a single particle. If all particles are identical, these functions do not depend on the particle index i , e.g. $f_1(\omega_i) \equiv f_1(\omega)$.

- Similarly, we define the two-particle PDF or probability:

$$f_2(\omega_i, \omega_j) = \int d\Omega' f(\Omega') \delta(\omega'_i - \omega_i) \delta(\omega'_j - \omega_j) \quad (\text{continuous})$$

$$p_2(\omega_i, \omega_j) = \sum_{\Omega'} p(\Omega') \delta_{\omega'_i \omega_i} \delta_{\omega'_j \omega_j} \quad (\text{discrete})$$

and so on. Such multi-particle functions contain information about correlations between particles. Specifically,

$$g_2(\omega_i, \omega_j) = f_2(\omega_i, \omega_j) - f_1(\omega_i) f_1(\omega_j)$$

is a two-body correlation function that equals zero only if the particles are uncorrelated (according to the fundamental property of probability for independent random variables). Assuming that all particles are identical, the multi-particle distribution and correlation functions depend on the coordinates of unspecified particles, as in $f_2(\omega, \omega')$.

- As long as the system has a macroscopically large number N of identical particles, and we are only interested in few-particle $n \ll N$ (rather than many-particle $n \sim N$) PDFs f_n , we can obtain the PDFs by *averaging over the degrees of freedom of a single system*. In other words,

$$f_1(\boldsymbol{\omega}) = \lim_{N \rightarrow \infty} \frac{1}{N} \frac{dN(\boldsymbol{\omega})}{d\boldsymbol{\omega}} \quad , \quad f_2(\boldsymbol{\omega}, \boldsymbol{\omega}') = \lim_{N \rightarrow \infty} \frac{1}{N^2} \frac{dN(\boldsymbol{\omega}, \boldsymbol{\omega}')}{d\boldsymbol{\omega}d\boldsymbol{\omega}'} \quad \dots$$

where $dN(\boldsymbol{\omega})$ is the number of particles (in a single system) whose coordinates are within a small volume $d\boldsymbol{\omega}$ from the state $\boldsymbol{\omega}$, and $dN(\boldsymbol{\omega}, \boldsymbol{\omega}')$ is the number of particle pairs whose coordinates are within a small volume $d\boldsymbol{\omega}d\boldsymbol{\omega}'$ from the joint state $(\boldsymbol{\omega}, \boldsymbol{\omega}')$.

- Proof: Consider the original single-particle PDF and substitute in its definition the defining formula for the objective microstate PDF $f(\boldsymbol{\Omega})$:

$$f_1(\boldsymbol{\omega}) = \int d\boldsymbol{\Omega}' f(\boldsymbol{\Omega}') \delta(\boldsymbol{\omega}'_i - \boldsymbol{\omega}) = \int d\boldsymbol{\Omega}' \delta(\boldsymbol{\omega}'_i - \boldsymbol{\omega}) \times \lim_{N \rightarrow \infty} \frac{1}{N} \frac{dN(\boldsymbol{\Omega}')}{d\boldsymbol{\Omega}'} = \lim_{N \rightarrow \infty} \frac{1}{N} \frac{dN(\boldsymbol{\omega})}{d\boldsymbol{\omega}}$$

When the number $dN(\boldsymbol{\Omega}')$ of ensemble members in the microstate $\boldsymbol{\Omega}'$ is integrated over the constrained phase space volume, one gets the number $dN(\boldsymbol{\omega})$ of ensemble members that contain a particle in the given state $\boldsymbol{\omega}$. Since $dN(\boldsymbol{\omega})$ is obtained by a single counting experiment on the entire ensemble, it can be regarded as a binomially-distributed random variable. Its mean $\langle dN(\boldsymbol{\omega}) \rangle$ is, then, equal to $\mathcal{N}P(\boldsymbol{\omega})$, where $P(\boldsymbol{\omega})$ is the probability that a single ensemble member will have a particle in the state $\boldsymbol{\omega}$. We can alternatively extract $P(\boldsymbol{\omega})$ from a single system by applying the definition of objective probability on the “ensemble” of particles in a given single system of interest:

$$P(\boldsymbol{\omega}) = \lim_{N \rightarrow \infty} \frac{dN(\boldsymbol{\omega})}{N}$$

Therefore:

$$f'_1(\boldsymbol{\omega}) = \lim_{N \rightarrow \infty} \frac{1}{N} \frac{dN(\boldsymbol{\omega})}{d\boldsymbol{\omega}} \rightarrow \left\langle \frac{1}{N} \frac{dN(\boldsymbol{\omega})}{d\boldsymbol{\omega}} \right\rangle = \frac{1}{N} \frac{\mathcal{N}P(\boldsymbol{\omega})}{d\boldsymbol{\omega}} = \lim_{N \rightarrow \infty} \frac{dN(\boldsymbol{\omega})}{Nd\boldsymbol{\omega}} = f_1(\boldsymbol{\omega})$$

The equality indicated by the arrow can be regarded as a consequence of central limit theorem: the random quantity dN/N is an average whose distribution converges to the Gaussian distribution with the same mean and vanishing variance (of the order of $1/N$). Using analogous arguments, one can prove the equivalence between any multi-particle PDFs $f'_n = f_n$, and also generalize to discrete cases.

Fluctuations in equilibrium

- Instantaneous total energy H of a macroscopic system is a random variable (except in the microcanonical ensemble). The standard deviation σ_H of H is negligible in comparison to its mean $E = \langle H \rangle$. In other words, the statistical fluctuations of H are negligible, so all ensembles that describe the same *macrostate* (given by E) are equivalent.

- Proof: A particle energy ϵ_i is a random variable with some finite mean ϵ and standard deviation σ_ϵ (particles are equivalent even if they are correlated by interactions). The total energy H is the sum of N individual particle energies ϵ_i , e.g.:

$$H = \sum_{i=1}^N \epsilon_i \quad , \quad \epsilon_i = \frac{p_i^2}{2m} + U(\mathbf{r}_i) + \frac{1}{2} \sum_{j \neq i} V(\mathbf{r}_j - \mathbf{r}_i)$$

Since all particles are equivalent, the average of H is:

$$E = \langle H \rangle = \left\langle \sum_{i=1}^N \epsilon_i \right\rangle = \sum_{i=1}^N \langle \epsilon_i \rangle = N\epsilon$$

and the variance of H is:

$$\begin{aligned}\sigma_H^2 &= \langle (H - E)^2 \rangle = \left\langle \left(\sum_{i=1}^N (\epsilon_i - \epsilon) \right)^2 \right\rangle = \left\langle \sum_{i=1}^N (\epsilon_i - \epsilon)^2 + \sum_{i \neq j} (\epsilon_i - \epsilon)(\epsilon_j - \epsilon) \right\rangle \\ &= N\sigma_\epsilon^2 + \sum_{i \neq j} \langle (\epsilon_i - \epsilon)(\epsilon_j - \epsilon) \rangle\end{aligned}$$

Any correlations between particles must be local if interactions have short range (Coulomb interactions in plasmas can be included because screening makes them effectively short-ranged). In other words, a particle may be correlated with other particles in its vicinity, but not with the macroscopic majority of particles far away. Therefore, we may express the last term on the right-hand-side in σ_H^2 as:

$$\begin{aligned}\sum_{i \neq j} \langle (\epsilon_i - \epsilon)(\epsilon_j - \epsilon) \rangle &= \sum_i \sum_{\substack{\text{near } i \\ j \neq i}} \langle (\epsilon_i - \epsilon)(\epsilon_j - \epsilon) \rangle + \sum_i \sum_{\substack{\text{far from } i \\ j}} \langle (\epsilon_i - \epsilon)(\epsilon_j - \epsilon) \rangle \\ &= N\phi + \sum_i \sum_{\substack{\text{far from } i \\ j}} \langle (\epsilon_i - \epsilon) \rangle \langle (\epsilon_j - \epsilon) \rangle \\ &= N\phi\end{aligned}$$

We merely separated the sum over the “second” particle j into the contribution from particles that are near the “first” particle i and the contribution from far-away particles. The ones that are near i can interact with it and contribute the amount $N\phi$ proportional the total number N of particles i (the value of ϕ is not important). The far-away particles are uncorrelated with i , so we can use the property of independent random variables that the average of their product equals the product of their averages. Then, by definition of $\epsilon = \langle \epsilon_i \rangle$, the average of $\epsilon_i - \epsilon$ vanishes, and we conclude that the full variance of H is proportional to the number of particles:

$$\sigma_H^2 = N(\sigma_\epsilon^2 + \phi)$$

It can be now clearly seen that the standard deviation of H is much smaller than its mean, $\sigma_H = \text{const} \times \sqrt{N} \ll E = \text{const} \times N$ in the $N \rightarrow \infty$ limit.

- The theorem trivially applies to the microcanonical ensemble, where $\sigma_H \equiv 0$.
- It can be shown using the same argument that all central moments $\langle (H - E)^n \rangle$ in systems with short-range correlations are proportional to N . Therefore, all measures of fluctuations $\sqrt[n]{\langle (H - E)^n \rangle} \propto \sqrt[n]{N}$ are small in comparison to the mean $E \propto N$.
- The same theorem applies also to the fluctuations of other extensive quantities, such as the number of particles N .
- Few body PDF’s f_n (for $N \ll n$) are the same in all types of equilibrium ensembles (microcanonical, canonical, etc.) that correspond to the same macrostate. In this sense (much stricter than previously), all ensembles are equivalent.
 - Proof: Consider N identical particles and focus on $f_1(\omega)$:

$$f_1(\omega) = \int d\Omega' f(\Omega') \delta(\omega'_1 - \omega) = \int d\Omega' f(\Omega') \frac{1}{N} \sum_{i=1}^N \delta(\omega'_i - \omega)$$

The microstate PDF $f(\Omega') \equiv f(H)$ depends only on the energy $H(\Omega')$ in equilibrium, and must produce the mean $E = \langle H(\Omega') \rangle$ fixed by the macrostate specification. The ensemble type determines the system’s boundary condition (interaction with the environment) *in a given macrostate*.

Thus, only the fluctuations of H (the variance and higher-order cumulants) may depend on the type of ensemble. Having this in mind, it is useful to rewrite f_1 as an integral over energy:

$$f_1(\boldsymbol{\omega}) = \frac{1}{N} \int d\Omega' f(\Omega') \sum_{i=1}^N \delta(\omega'_i - \boldsymbol{\omega}) = \frac{1}{N} \int dH f(H) \times \int d\Omega' \delta(H(\Omega') - H) \sum_{i=1}^N \delta(\omega'_i - \boldsymbol{\omega})$$

This looks like the average $\langle X(H, \boldsymbol{\omega}) \rangle$ of the random quantity

$$X(H, \boldsymbol{\omega}) = \int d\Omega' \delta(H(\Omega') - H) \sum_{i=1}^N \delta(\omega'_i - \boldsymbol{\omega})$$

that depends on random total energy H and the given single-particle state $\boldsymbol{\omega}$. By construction, $X(H, \boldsymbol{\omega})$ is proportional to the average number of particles at $\boldsymbol{\omega}$ in the microstates with total energy H . This is a local property of the system's dynamics (mathematically encoded in $H(\Omega)$, the Hamiltonian) and the density of states, but expressed as a function of the *macroscopic total* energy. Small changes of H have very little effect on X ; only small changes of *energy per particle* $\epsilon = H/N$ do affect X :

$$\frac{d^k X}{d\epsilon^k} \sim \mathcal{O}(N^0) \quad \Rightarrow \quad \frac{d^k X}{dH^k} = \frac{1}{N^k} \frac{d^k X}{d\epsilon^k}$$

We can exploit this feature in the expansion of $f_1(\boldsymbol{\omega}) = \langle X(H, \boldsymbol{\omega}) \rangle$ in terms of “central” moments $\langle (H - E)^k \rangle$:

$$\langle X(H, \boldsymbol{\omega}) \rangle = \sum_{k=0}^{\infty} \frac{1}{k!} \frac{d^k X(E, \boldsymbol{\omega})}{dE^k} \langle (H - E)^k \rangle = X(E, \boldsymbol{\omega}) + \frac{1}{2} \frac{d^2 X(E, \boldsymbol{\omega})}{dE^2} \text{Var}(H) + \dots$$

This expansion follows from the Taylor expansion of $\langle X(H, \boldsymbol{\omega}) \rangle$ about $H = E$ under the ensemble average. We proved before that $\text{Var}(H) \propto N$ (all “central” moments are proportional to N). Using $d^k X/dE^k \propto 1/N^k$ we find:

$$f_1(\boldsymbol{\omega}) = \langle X(H, \boldsymbol{\omega}) \rangle = X(E, \boldsymbol{\omega}) + C_2 \frac{1}{N^2} N + C_3 \frac{1}{N^3} N + \dots \xrightarrow{N \rightarrow \infty} X(E, \boldsymbol{\omega})$$

Therefore, $f_1(\boldsymbol{\omega})$ equals the variable X evaluated at the mean energy $H = E$ in the $N \rightarrow \infty$ limit. This way of calculating f_1 is identical to the *microcanonical* ensemble calculation, where total energy is fixed. In other ensembles, energy fluctuations produce corrections (the terms involving constants $C_2, C_3 \dots$), but all of them vanish).

- The proof for other multi-particle PDFs f_n proceeds in the same steps. It breaks down, however, when n becomes comparable to N (because the assumption about the changes of the generalized X with E fails).

Entropy

- The statistical formula for entropy consistent with laws of thermodynamics is:

$$S = -k_B \int d\Omega f(\Omega) \log(f(\Omega)\Delta\Omega) \quad (\text{continuum}) \quad ; \quad S = -k_B \sum_{\Omega} p(\Omega) \log(p(\Omega)) \quad (\text{discrete})$$

In continuum systems, $f(\Omega)d\Omega \rightarrow f(\Omega)\Delta\Omega$ inside the logarithm is microstate probability $p(\Omega)$ in analogy with discrete systems. Then, $\Delta\Omega$ is the elementary phase space volume that cannot contain more than one microstate.

- Proof: Entropy S is defined in thermodynamics as an *independent* state variable related to the transfer of heat in a reversible change from one equilibrium state A to another B :

$$S_B - S_A = \int_A^B \frac{dQ}{T}$$

If the system is unable to do any work, then, by conservation of energy, the amount of transferred heat equals the change of internal energy:

$$S_B - S_A = \int_A^B \frac{dE}{T} \Big|_{dW=0}$$

We will use this relationship and 3rd law of thermodynamics to deduce the statistical definition of entropy. Without loss of generality, we will work with a discrete ensemble and seek an expression of the form:

$$S = \langle \mathcal{S} \rangle = \sum_{\Omega} p(\Omega) \mathcal{S}(\Omega)$$

which properly extracts entropy as a state variable (ensemble average) from the distribution of some physical quantity $\mathcal{S}(\Omega)$. We ought to find $\mathcal{S}(\Omega)$ that is fundamentally independent of microstate energy $H(\Omega)$, because entropy is an independent state variable. Consider a process in which the internal energy of the system is changed at *fixed temperature* and *without any ability for the system to do work* (an example is Joules' experiment with a freely expanding gas that has nothing to push and do work on). Based on this process and the thermodynamic definition of entropy, we find:

$$S_B - S_A = \frac{1}{T}(E_B - E_A) \quad \Rightarrow \quad S = \text{const} + \frac{E}{T}$$

This is a type of an equation of state: S and E are independent at the microscopic level, but related as above *only in equilibrium* under the conditions we fixed. Let us express internal energy E as a desired ensemble average:

$$S = \text{const} + \frac{1}{T} \sum_{\Omega} p(\Omega) H(\Omega)$$

and then try to eliminate the reference to the microstate energy in the sum:

$$\begin{aligned} S &= \text{const} - \frac{1}{\beta T} \sum_{\Omega} p(\Omega) \log(e^{-\beta H(\Omega)}) = \text{const} - \frac{1}{\beta T} \sum_{\Omega} p(\Omega) \log(Z p(\Omega)) \\ &= \text{const}' - \frac{1}{\beta T} \sum_{\Omega} p(\Omega) \log(p(\Omega)) \end{aligned}$$

We temporarily used the canonical probability distribution

$$p(\Omega) = \frac{1}{Z} e^{-\beta H(\Omega)} \quad , \quad Z = \sum_{\Omega} e^{-\beta H(\Omega)}$$

and the fact that partition function Z is a constant whose logarithm we may absorb into const' . However, any probability distribution fixed by the macrostate (i.e. any type of ensemble) is acceptable, because all ensembles are statistically equivalent. We would like to identify:

$$\mathcal{S}(\Omega) \rightarrow -\frac{1}{\beta T} \log(p(\Omega))$$

but this is not adequate because we need a truly microscopic definition of \mathcal{S} that makes no reference to macrostate quantities such as temperature. The problem is eliminated only if βT is constant. In other words,

$$\beta \propto \frac{1}{k_B T}$$

is fundamentally required by the units of β (βE is dimensionless) and the fact that entropy is an independent state variable. We will later show that $\beta = 1/k_B T$ by studying the statistical mechanics of an ideal gas - this is a matter of how exactly temperature is defined. Hence, the meaning of the physical quantity $\mathcal{S}(\Omega)$ that microscopically relates to entropy is derived from the fraction of occurrences of the microstate Ω in the ensemble of microstates prepared from the same macrostate.

- We will now prove that the above constant should be zero using 3rd law of thermodynamics. Canonical distribution (which makes a reference to temperature) shows that the system is certainly in its lowest energy state $H = H_0$ at $T = 0$:

$$\begin{aligned} p(\Omega) &= \frac{1}{\sum e^{-\beta H}} e^{-\beta H} = \frac{1}{\sum e^{-H/k_B T}} e^{-H/k_B T} \xrightarrow{T \rightarrow 0} \frac{1}{e^{-H_0/k_B T}} e^{-H/k_B T} \\ &= e^{-(H-H_0)/k_B T} \xrightarrow{T \rightarrow 0} \delta_{H(\Omega), H_0} \end{aligned}$$

because the partition function in the denominator becomes dominated by the smallest energy H_0 in the spectrum, i.e. $Z \approx e^{-H_0/k_B T}$. All natural systems have a unique lowest energy state, or a few lowest energy states. If the lowest energy state Ω_0 is unique, then $p(\Omega) \rightarrow \delta_{H(\Omega), H_0} = \delta_{\Omega, \Omega_0}$:

$$\begin{aligned} S &= \text{const}' - k_B \sum_{\Omega} p(\Omega) \log(p(\Omega)) \xrightarrow{T \rightarrow 0} \text{const}' - k_B \sum_{\Omega} \delta_{\Omega, \Omega_0} \log(\delta_{\Omega, \Omega_0}) \\ &= \text{const}' - k_B \times 1 \times \log(1) = \text{const}' \rightarrow 0 \end{aligned}$$

Since entropy vanishes at $T = 0$ according to the 3rd law of thermodynamics, we find that $\text{const}' = 0$. If, more generally, there are a few minimum energy states, then entropy defined this way will be finite at $T = 0$, but of the order one (hence negligible), instead of being proportional to the number of particles N as an extensive quantity.

- In discrete microcanonical ensemble at total energy E , entropy equals:

$$S = k_B \log(\Omega(E))$$

where $\Omega(E)$ is the number of microstates at that energy.

- Proof: The microcanonical probability distribution specified by internal energy E is:

$$p(\Omega) = \frac{1}{\Omega(E)} \delta_{H(\Omega), E} \quad , \quad \Omega(E) = \sum_{\Omega} \delta_{H(\Omega), E}$$

where $\Omega(E)$ is the number of microstates at energy E . Hence, entropy in a discrete microcanonical ensemble becomes:

$$\begin{aligned} S &= \text{const}' - k_B \sum_{\Omega} p(\Omega) \log(p(\Omega)) \\ &= \text{const}' - \frac{k_B}{\Omega(E)} \sum_{\Omega} \delta_{H(\Omega), E} \log\left(\frac{\delta_{H(\Omega), E}}{\Omega(E)}\right) \\ &= \text{const}' - \frac{k_B}{\Omega(E)} \sum_{\Omega} \delta_{H(\Omega), E} \log\left(\frac{1}{\Omega(E)}\right) \\ &= \text{const}' - \frac{k_B}{\Omega(E)} \times \Omega(E) \log\left(\frac{1}{\Omega(E)}\right) \\ &= \text{const}' + k_B \log(\Omega(E)) \end{aligned}$$

Note that the Kronecker symbol inside the logarithm (2nd line) has no effect because the only time it matters is when the same Kronecker symbol outside of the logarithm is equal to one. Then, the logarithm is essentially independent of Ω and can be pulled out of the sum. The isolated sum over microstates is just the number of microstates $\Omega(E)$, which is canceled by the normalization of the probability. The final result is that entropy is the logarithm of the number of microstates at given energy E , up to a constant. 3rd law of thermodynamics says that $S = 0$ at zero temperature.

- Having the statistical definition of entropy, it is possible to microscopically derive the 2nd law of thermodynamics. Any system constrained to be in a given macrostate still has many accessible microstates. Out of equilibrium, there could be some bias in the statistical distribution of microstates. For example, one could prepare a gas by putting all of its particles in a small volume at a corner of a large box. Among all existing microstates for the gas in the box, only those with particles confined to the corner are initially likely. The gas will, however, quickly spread through the entire box. In this process, the entropy

$$S = k_B \log(\Omega(E))$$

grows because the number of available microstates of the gas with conserved energy E grows with the volume that the gas occupies. As the system evolves from a less to a more likely (i.e. chaotic) condition, its entropy increases - which is one way to state 2nd law of thermodynamics.

- The laws of nature are invariant under time-reversal. For any possible process, the one unfolding in exactly the opposite order is also possible according to the equation of motion. So, if one drops an egg on the floor and the egg breaks, it is possible in principle to simultaneously impart on every atom exactly the opposite velocity from the one it had at the instant of fall - the pieces of the broken egg would fly upward, reassemble and land back to one's hand. Such a reversed process is never observed, even though it is possible. This is a reflection of the 2nd law of thermodynamics, now interpreted statistically. Every macroscopic system always evolves to more probable and more featureless states. This is the most likely scenario, and when the number of degrees of freedom is huge, the probability of this is "huge". The reversed process is just statistically extremely unlikely. In that sense, there is an *arrow of time*, the causal passage of time from past to future.
- Entropy defined by

$$S \propto - \sum_{\Omega} p(\Omega) \log(p(\Omega))$$

is a mathematical concept, an abstract property of a probability distribution. It expresses the overall uncertainty of measurement outcomes, or the amount of information we lack about the described statistical system. We could have postulated this formula based solely on its mathematical meaning, and then derived all equilibrium distributions from the requirement that entropy (uncertainty) be maximized. The following statements have rigorous mathematical proofs:

- Among all possible ensembles of a statistical system (not just equilibrium ones) with fixed energy, the one with maximum entropy is *microcanonical*.
- Among all possible ensembles of a statistical system with fixed temperature, the one with maximum entropy is *canonical*.
- Among all probability distributions defined on a set of possible measurement outcomes, the one with maximum entropy is *uniform*.
- Among all probability distributions defined on $x \in (0, \infty)$ with fixed mean $\mu = \langle x \rangle$, the one with maximum entropy is *exponential*.
- Among all probability distributions defined on $x \in (-\infty, \infty)$ with fixed mean $\mu = \langle x \rangle$ and variance $\sigma^2 = \text{Var}(x)$, the one with maximum entropy is *Gaussian*.