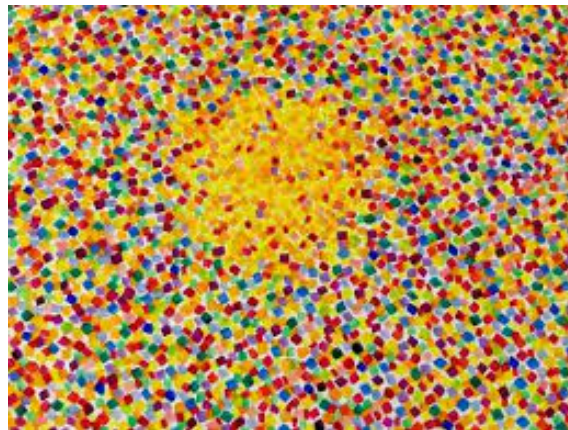# Thermostatistique 2023-24 (LU3PY403)

Fabio Pietrucci, Sorbonne Université

June 21, 2024

XXX add in CM4 esempio in cui estraggo valori A da distrib. non gaussiana (con una spalla, ad es), mostrare che std.dev. è stabile facendo sempre più misure, mentre std.dev. della media scende come sqrt(N). Mostrare anche che media è distribuita come una gaussiana!

Organisation

6 ECTS: 12x2h cours, 12x2h TD, 12h projet bibliographique.

Modalité d'évaluation:

CC mi-parcours, sur table en présentiel (30/100);

présentation orale avec diapositives du projet bibliographique (30/100) (non rattrapable);

CC final, sur table en présentiel (40/100).

Seconde session: épreuve orale unique qui permet de rattraper les deux CC (70/100).

Main books:

H.B. Callen, "Thermodynamics and an introduction to thermostatistics" [1];

C. Texier & G. Roux, "Physique statistique" [2].

Secondary:

F. Reif, "Fundamentals of Statistical and thermal physics" [3];

E. Brunet, T. Hocquet, X. Leyronas, polycopié 3P011 (http://www.lps.ens.fr/ ebrunet/) [4] .

Example of exercise book: S.B. Cahn et al, "A guide to physics problems - part 2".

Contents

1   CM1

Outline: in the first weeks you'll get basic material necessary to understand thermodynamics and statistical physics, like equations of motion (importance of energy), some differential calculus and some probability theory.

1.1   What is thermodynamics (TD)?

Empirical macroscopic behavior of systems is often simple, a few variables that appear well defined are sufficient to describe it: equilibrium conditions. However, if we go deeper, definition of equilibrium is not trivial.

Energy conservation (1st law) + entropy maximization (2nd law) allow to build very successful and elegant theory, but we need to set many parameters and we lack microscopic understanding. This because TD is a science of relationships between physical quantities, not of the quantities themselves. See the very nice discussion in the last chapter of ref. [1].

TD is a key tool, very largely employed in modern science due to its generality. Many basic questions remain open: how to interpret (and compute) entropy in tricky situations? Can we apply the second principle to the universe as a whole? How can we apply it to living organisms? What about the arrow of time? etc...

1.2   What is statistical physics (SP)?

Historically, it fills gap between microscopic laws of motion (Newton/Schrödinger) and macroscopic behavior, including TD. Much more generally, both at the microscopic and macroscopic scale, it makes sense of experiments and simulations. Its probabilistic methods have large applications beyond physics.

Contrary to what is written in the beginning of almost any book on SP, we should not think to SP as "unfortunately we have too many particles making it impossible to predict/observe detailed motion, so let's resort to a second-choice theory...", instead it gives accurate and insightful predictions of new phenomena whether we know or not the detailed particle motion! In other words, even if we know perfectly the microscopic details we still need SP. Very simple basis (prob. theory concepts) and hypotheses (equal a priori probabilities, ergodicity) lead to very deep and varied consequences. It is not a mere transfer of knowledge from micro to macro: plenty of new phenomena and physical laws emerge at different scales, as introduced well by Anderson's article "more is different". Not necessarily $N \to \infty$: very often we can apply to $N \sim 100$ or smaller, let's think to $N \gg 1$, rather.

Today, it is unclear to me how much the relationship between information theory (very popular today) and statistical physics is just an analogy, and how much it is deeper. I think it can be misleading to attribute too much physical relevance to information theory.

A part of researchers in SP focus on studying simple models (Ising, lattice gas, Hubbard...) in great mathematical detail. In general, well-built empirical models are interesting, albeit simplified, since they play a central role in science and have a deep connection with our ability to understand the world as well as with our creativity and problem solving skills. However they are not sufficient to understand the real world.

Another part of researchers in SP focus on realistic systems formed by thousands to millions of particles, like solids/liquids/gases/plasmas with simple or complex compositions, advanced materials, biological systems (from proteins, DNA, RNA, lipids to whole cells and spectacular nanomachines like ATP synthase), chemical reactions. Here the challenge is finding suitable theoretical techniques, as well as numerical methods, to extract/predict relevant physical observables (populations, free-energy barriers, kinetic rates...) from simulations or experiments. A lot of research is still needed both in equilibrium and non-equilibrium domains.

1.3 Basic jargon: systems, variables, equilibrium, processes...

By now you should be used to the jargon: isolated/open/closed/rigid/etc systems, energy, heat and work, entropy, NVTE etc conditions, equilibrium / out of equilibrium, transformation.

However, being used to something does not guarantee to understand it very well (think to mathematics, or to car driving...): the above concepts are seldom explained in detail. We'll do our best to dig into them in this course.

For instance, what is equilibrium? A good definition prescribes lack of macroscopic changes (i.e., changes in macroscopic variables) and lack of currents. So, an isolated pure system (like a liquid or gas) with inhomogeneous density $\rho(\mathbf{r}, t)$ is not in equilibrium: currents change density, irreversible diffusion takes place tending to equilibrium:

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot \mathbf{j} \neq 0 \ , \quad \frac{\partial \rho}{\partial t} = D\nabla^2 \rho$$

(another example: a solid between a hot and a cold reservoir, now energy currents are present even though there is a stationary energy density profile).

Let's look carefully at above equations: they apply to rather macroscopic, coarse scale (let's say above micrometer and above nanosecond – recall atoms scales are Angstrom and femtosecond), so they are rather in the domain of TD. However there appears the letter "t", so it is not TD (in a strict definition, too strict perhaps...), when we are not yet at equilibrium. After equilibrium is reached we can use TD. Equilibrium corresponds to zero time derivative, i.e. "t" disappearing from the game, regarding macroscopic quantities.

However, even in equilibrium atoms move a lot: how to reconcile with above? What is equilibrium from microscopic viewpoint? The key is to introduce averages: stationary behavior of average properties (and lack of currents) is equilibrium. Usually we need to average over both space and time, something that is automatically done in most experiments (we can look at explicit examples, like the nice illustration of Texier 3.1).

How and why a given system evolves towards equilibrium? What is the mechanism? A difficult and central question, we'll see examples throughout the course. The microscopic behavior of atoms (forces etc.) is similar both in equilibrium and out-of-equilibrium conditions; something is different in the collective behavior of many particles (note that "many" can be smaller than 100). We need probability theory.

1.4 Microscopic illustration

Many things can be understood looking at 100 gas or liquid particles in a cubic box (see also Texier 3.1). One can consider Lennard-Jones interaction, but also water and molten silicon, to become familiar with atomic scale of length, time, energy (important!). We can use molecular dynamics, numerical solution of equations of motion to find atomic trajectories, or Monte Carlo simulations (where, however, we lack realistic trajectories).

One can plot density on a finer or coarser grid, time evolution, inspect subvolumes, show different initial conditions, evolution to equilibrium. Interesting to see averages and fluctuations: the latter reduce if the coarse graining is more aggressive, both in space and time, as we will see.

Example of 5 gas atoms in a box: if we count how often all atioms are found in left half, we get with a good approximation $1/2^5 = 0.03125$. It is a simple example of how SP allows predicting result of experiment, irrespective of our knowledge or ignorance about the detailed motion (in MD we know everything), based on a correct application of probability theory. Now one can repeat with more atoms, with water etc. and get a glimpse of the second principle.

If we wait "long enough", the system "typically" relaxes to equilibrium, irrespective of details of interaction between atoms! A crucial property that we have troubles to demonstrate theoretically. How long? Difficult to say. Connection with second principle and entropy increase, it will be clarified later

on. What does it mean "typically"? Tricky case: metastability, when a system can get trapped in a state for a long time before reaching a more stable state, it is the rule rather than the exception around us.

We can also look at energy evolution, temperature and pressure, and so on. Food for lots of future thought.

Another crucial concept we will explore: all interesting, realistic systems are cahotic, chaos plays an important role in the applicability or probability theory to physics and in the irreversibile aspect of many phenomena.

1.5   Unspoken assumptions

When we consider examples like the one above, we expect a certain behavior from matter, because we are used to it: e.g., density remains finite, no collapse, atoms repel each other at very short distance; atoms are attracted to each other, they are always sticky irrespective of the element (hence the name "condensed" matter). In short, there are many unspoken assumptions, that are necessary to make TD key concepts work, e.g., extensivity of properties like energy or entropy, the idea that surface effects are negligible compared to bulk ones, etc.

Explanation of this behavior comes from quantum mechanics, and often is very complex to work out. E.g. short range repulsion is a complex combination of Pauli principle (important but insufficient: also hydrogen atoms repel each other!) and of the behavior of the solutions of Schrödinger equation when atoms get very close together, not fully understood still today (calculations are very difficult). Roger Balian has a nice discussion in the second volume of his book "From microphysics to macrophysics". Note the constrast with the classical Coulomb attraction between two opposite charges, which diverges at zero distance. As stated by Feynman, one can also invoke the uncertainty principle: "The electrostatic forces pull the electron as close to the nucleus as possible, but the electron is compelled to stay spread out in space over a distance given by the Uncertainty Principle. If it were confined in too small a space, it would have a great uncertainty in momentum. But that means it would have a high expected energy– which it would use to escape from the electrical attraction. The net result is an electrical equilibrium not too different from the idea of Thompson–only is it the negative charge that is spread out, because the mass of the electron is so much smaller than the mass of the proton."

Similar situation with attraction between atoms, tendency to form chemical bond: a subtle and complex QM phenomenon, not fully understood (it occupies a big scientific community: quantum chemists and beyond), very difficult to compute accurately – in fact, it is considered a task for quantum computers (McArdle et al, Rev. Mod. Phys. 92, 015003, 2020). To learn more, you can read Cohen-Tannoudji et al. vol. II ch. XI(G), or Feynman Lectures on Physics vol. III ch. 10 (www.feynmanlectures.caltech.edu/).

Simple models like Lennard-Jones potential try to capture such effects with a simple functional form of the internuclear distance, a typical approach in molecular dynamics simulations, where you "average out" the effect of the electrons, obtaining a "force field". If you compute carefully (from QM) the energy profile for different pairs of atoms, you will always find something qualitatively similar to Lennard-Jones.

Another unspoken assumption: we expect matter to loose quickly the memory of the initial conditions. We expect also that different parts of a system, not too close together (say > 1 nm), are not correlated. Same thing for two times separated by a sizable interval. Such decorrelations are quite ubiquitous, but, once again, non-trivial to understand from a theoretical viewpoint. Without decorrelations we would have a hard time in applying simple statistical tools to physics!

1.6   The role of energy

The central quantity that connects microscopic laws of motion (classic or quantum), SP and TD is the energy. It has many important roles. One could argue that it cannot be easily observed and

quantified, however it is a very useful device to rationalize the motion of the particles. Note that motion, trajectories, is what we can directly observe: an interesting question is which physical quantities are "real" in physics and which ones are artificial, only useful calculation tricks?

In fact, both Hamilton eq and Schrödinger eq are based on H, so time evolution is computed thanks to the energy expression (Texier 3.A has short and advanced summary): for N particles we have, using Cartesian coordinates and corresponding momenta

$$\sum_{j=1}^{N} \frac{\mathbf{p}_j^2}{2m_j} + U(\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N)$$

In general, the potential energy is a complex function of the $3N$ coordinates. Very often, in practice, it comes from the Born-Oppenheimer approx.: U contains the quantum ground state energy of the electrons at fixed nuclei, a complicated many-body function of nuclear coordinates, plus the Coulomb repulsion between nuclei. In simple situations, we try to approximate U with two-body interactions:

$$U(\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N) \approx \sum_{j=1}^{N-1} \sum_{k=j+1}^{N} u(\mathbf{r}_j, \mathbf{r}_k)$$

an example being the Lennard-Jones potential, modeling short-range repulsion and long-range van der Waals attraction:

$$u(\mathbf{r}_j, \mathbf{r}_k) = 4\epsilon \left[\left(\frac{\sigma}{r_{jk}} \right)^{12} - \left(\frac{\sigma}{r_{jk}} \right)^{6} \right] \quad , \quad r_{jk} = |\mathbf{r}_j - \mathbf{r}_k|$$

Note that, as a rule, interactions between particles complicate a lot the problem from the SP viewpoint (because the particles are not statistically independent!), but of course they also introduce very interesting phenomena like metastability, polymorphism, phase transitions, etc.

In general, we can adopt generalized coordinates – instead of Cartesian ones – based on the symmetry of U (e.g., spherical coordinates): we end up again with $n = 3N$ positions and the same number of momenta, $H(q_1, q_2, ..., q_n, p_1, p_2, ..., p_n)$. The aim is simplifying the writing of U, the price to pay is a more complicated kinetic energy, a quadratic form with a non-diagonal mass matrix, but often it is worth the effort for an easier overall study. The classical equations of motion, in Hamilton form so independent of the choice of the generalized coordinates, are

$$\begin{cases} \dot{q}_j & = \frac{\partial H}{\partial p_j} \\ \dot{p}_j & = -\frac{\partial H}{\partial q_j} \end{cases}$$

It is important, from now on, to be able to write cleanly and with a rigorous notation the classical Hamiltonian and the equations of motion for many particles.

In the quantum case we use Schrödinger eq:

$$i\hbar \frac{\partial \psi}{\partial t} = H\psi$$

where, in the position representation, ψ depends on the coordinates of all particles (e.g., nuclei and electrons): note the difference with respect to classical mechanics, we don't have positions and momenta but only positions (or only momenta), we "loose" half the degrees of freedom.

Classical and quantum time evolution of generic observables is also computed from the energy:

$$\frac{dA}{dt} = \{A, H\} + \frac{\partial A}{\partial t} \equiv \sum_j \left(\frac{\partial A}{\partial q_j} \frac{\partial H}{\partial p_j} - \frac{\partial H}{\partial q_j} \frac{\partial A}{\partial p_j} \right) + \frac{\partial A}{\partial t}$$

$$\frac{d\langle A \rangle}{dt} = \frac{1}{i\hbar} \langle [A, H] \rangle + \langle \frac{\partial A}{\partial t} \rangle \equiv \frac{1}{i\hbar} \langle (AH - HA) \rangle + \langle \frac{\partial A}{\partial t} \rangle$$

Moreover, energy conservation is related to the symmetry of physical laws with respect to time translations, and such symmetry, like others, are at the foundation of thermodynamics: see the excellent final chapter of Callen, with also a description of Noether's theorem, broken symmetry, the central limit theorem as a basis for simple statistics in thermodynamics, etc. Excellent citations therein!.

You can try replacing A with H in the expressions above, to demonstrate energy conservation. Note however that if you introduce dissipation of energy, clearly energy is no more conserved (we leave the playground of Hamiltonian conservative systems) and time-reversal symmetry is broken: irreversible phenomena, the second principle is again in action. Hence, people like Eddington tried to connect the "arrow of time" (apparent or true?) to the second principle. Subtle to grasp, since Boltzmann who scratched a lot his head. To give an idea of the difficulties, Landau begins his SP book by convincing the reader that if we observe entropy to grow from time t_0 towards the future, we should expect it to grow also from t_0 towards the past! We'll see examples with also suggested articles to go deeper.

We'll see later that energy is also key in SP to predict the equilibrium probability of states, e.g. with the canonical law $p_\alpha \propto e^{-E_\alpha/kT}$, where α indicates a microstate (see later...), and through the expression of probabilities the energy allows also to define the entropy, e.g. in the canonical ensemble $S = -k \sum_\alpha p_\alpha \log p_\alpha$. Conclusion: energy is truly a key tool!

2 CM2

2.1 Types of transformations

To discuss transformations, we need plenty of concepts that you already saw: heat, work, entropy, temperature... We recall key points, and the two principles.

We need to define quasi-static, reversible and irreversible transformations, a little bit tricky. We change some macroscopic variable (V, E, T...): if done very slowly, we remain rather close to equilibrium states. In the unrealistic limit where the transformation is infinitely slow, we traverse a continuum of equilibrium states: quasi-static transformation. An idealization, where time disappears from the game.

In practice, a real transformation passes through non-equilibrium states, i.e., states of the system that require, to be specified, all microscopic coordinates, not only macroscopic ones. We start from an equil. state (the system has been left at fixed constraints for a very long time), we change constraints to new fixed values so system becomes out of equil., after long-enough time system will be again in equil. state. When out-of-equil., quantities like P, T can be computed microscopically (e.g., in a simulation) but they are not in equil., not stationary (show numerical illustration).

Key concept: after changing constraints, system needs a certain relaxation time τ to reach equilibrium again. If we move constraints (e.g. we change the volume from V_1 to V_2) very slow w.r.t. τ, during this movement it looks like system is in equil. even though it is changing all the time.

2.2 First law of TD

As we saw in CM1+TD1, energy is conserved in an isolated system, which is also a statement of the first law of TD. It the system is not isolated, its possible energy changes (i.e., exchanges with the environment) are classified identifying two contributions, work and heat, that are quite different in nature: $\Delta E = W + Q$ (a more detailed statement of the first law). The work concerns macroscopic mechanical coordinates (e.g., the volume), the heat concerns "hidden" microscopic coordinates. We adopt the convention of positive heat and work when they raise the energy of the system.

An important point is made, e.g., in Atkins's book Ref. [5] (excellent source): "The distinction between work and heat is made in the surroundings. The fact that a falling weight may stimulate thermal motion in the system is irrelevant to the distinction between heat and work: work is identified as energy transfer making use of the organized motion of atoms in the surroundings, and heat is identified as energy transfer making use of thermal motion in the surroundings. In the adiabatic compression of a gas, for instance, work is done on the system as the atoms of the compressing weight descend in an orderly way, but the effect of the incoming piston is to accelerate the gas molecules to higher average speeds. Because collisions between molecules quickly randomize their directions, the orderly motion of the atoms of the weight is in effect stimulating thermal motion in the gas. We observe the falling weight, the orderly descent of its atoms, and report that work is being done even though it is stimulating thermal motion."

As clearly discussed by Callen, the key to the simplicity of the macroscopic description of TD lies in the coarse nature of macroscopic measurement w.r.t. the atomic scales of time and distance. Onlys such combinations of coordinates that are time-indep. result observable in TD, and they are connected to conserved quantities. In the spatial case, only long-wavelength modes survive to averaging, and they correspond to the volume, so we care about length variation of the whole system. Mechanics, elasticity, electricity are concerned with surviving coordinates. TD studies consequences of microsc. behavior that appear at macroscopic scale.

The internal energy is just the total energy (kinetic plus potential), excluding the kinetic energy for the motion of the system as a whole. Note that E is constant in the isolated case, so we do not need to do any average. However if we are in the canonical case, i.e., contact with a heat bath, clearly E fluctuates and from the TD point of view we should consider the average. In the TD limit of very large system size the relative fluctuations go to zero, however the absolute fluctuations diverge! $\sigma_E = \sqrt{\langle E^2 \rangle - \langle E \rangle^2} \propto \sqrt{N}$, $\langle E \rangle \propto N$, $\sigma_E / \langle E \rangle \propto 1/\sqrt{N}$, we will demonstrate later.

The (internal) energy E, often indicated with U, is a state function, so dE is exact differential, while δW and δQ are not, they are just small quantities: $dE = \delta W + \delta Q$. This means that W and Q depend on the path of the transformation, not just on the endpoints, and it is absurd to speak of "the work of a state of the system" or the "heat of a state of the system", contrary to energy, since W and Q are always associated with a transformation. However, when $\delta W = 0$, $dE = \delta Q$ becomes an exact differential, and likewise for δW when $\delta Q = 0$.

2.3 Examples

(Brunet notes 2.1, 2.2 show examples, and ways to transfer work and heat).

Consider a process where $W = 0$, $\Delta E = Q$: even though heat is connected with thermal agitation, i.e., kinetic energy, it would be wrong to identify ΔE with a pure increase (or decrease) of kinetic energy. When we supply heat to the system, e.g. via the contact with a hotter thermal bath, atoms in the system are accelerated to higher velocities hence higher kinetic energy, however such kinetic energy is quite immediately partially transformed into potential energy (think to an oscillator). There is a fair redistribution (the same happens if $\Delta E = W$), equipartition in the case of classical harmonic systems (see later), there could even be a phase transition. This is an example of how TD tells us little information (here, ΔE only), knowing what happens inside the system requires statistical physics and detailed microscopic information.

So picture of heat transfer is more complicated than just kinetic energy, even though the latter defines temperature (through equipartition):

$$K = \sum_{i=1}^{N} \frac{1}{2} m_i \mathbf{v}_i^2 = 3N \frac{k_B T}{2} \ , \quad k_B T = \frac{2}{3} \frac{K}{N} = \frac{2}{3} \bar{k}$$

In this def. for a finite system, of course T fluctuates, even if there is a thermostat (as can be showed with a simulation).

2.4 Entropy definition in TD

A central problem in TD is the following: start from system in equil., now remove (change) some constraints, can we predict the new equil. state it will reach? The solution is found by entropy maximization (Callen 1.10).

We often focus on quasi-static (i.e., super slow), reversible (i.e., no change in the entropy of the universe) transformations since they can be written explicitly through equilibrium quantities, i.e.

$$dE = \delta Q + \delta W = T\,dS - p\,dV + \sum \mu_i dN_i$$

This expression, or the equivalent one for dS, is the fundamental relation/equation, or "identité thermodynamique". Note that here we do not have anymore an isolated system (that would have $E =$const): here dS can be different from zero even though the transformation is reversible because in the formula above E, S, V, N are quantities of the system under observation, which is a sub-system immersed in a larger system able to exchange heat, work and particles.

Note that out-of-equil. we could not even have constant values of $p, T, \mu...$ Once again, quasi-static transf. can be reversible (for the universe $\Delta S_{\text{tot}} = 0$) or irreversible ($\Delta S_{\text{tot}} > 0$). See CM3 for troubles if we try to use "quasi-static TdS" above in slow free expansion of a gas into vacuum: it can't be used since $dS > 0$, $T =$const, but $dE = 0$, as free expansion is an intrinsically irreversible process, no matter the speed. It is important to distinguish quasi-static and reversible processes.

In free expansion of a gas into vacuum, or when you put in contact hot and cold body, new equil. state is reached irreversibly, by maximizing the state function S_{tot}, which strictly speaking is defined only for equil. states (but see Derrida notes for definitions in local equil.). Extremum principle, as usual in many branches on physics.

Following Callen, we define entropy through postulates, to be justified a posteriori (another typical approach in physics):

postulate 1: there exist particular states, called equil. states (of "simple systems"), that macroscopically are completely characterized by the extensive parameters E, V, N (where N denotes collectively the set of all N_i for the different species...). If necessary, we could add dipole moment, elastic strain, etc. Note that in practice we are satisfied with long-lived metastable states, that appear to be at equil. on long times but they are not on even longer times. For instance on geological times isotopes would decay and so on (see also the out-of-equil. character of the whole universe!).

postulate 2: there exists an entropy function $S(E, V, N)$, defined for equil. states, such that the set of extensive parameters of subsystems composing the overall system are determined by maximizing S w.r.t. the manifold of constrained equil. states that are compatible with the system.

Difficult statement, clearer with an example: consider an overall isolated system made by 2 isolated parts, each one in equil., separated by adiabatic wall; now we make the wall diathermal: how does energy redistributes at equil.? To find E_1, E_2 of the two parts we start considering a virtual situation where the wall is still adiabatic, but we consider all possible pairs E_1, E_2 compatible with the constraint of fixed total $E = E_1 + E_2$. This trick of imagining an adiabatic wall (where actually it is diathermal) allows defining S for different sistuations (we conceptually avoid the transient when the system relaxes to equil., where S would difficult todefine), then we pick the maximal S among all virtual possibilities and this will be the entropy for the system with diathermal wall.

In principle, this solves TD problems, if we know $S(E, V, N)$, i.e., "the fundamental relation": it contains all TD information, beautiful. In practice it seems difficult to proceed very far without SP: the latter allows to compute explicitly the entropy.

postulate 3: the entropy of a composite system (i.e., made by large parts, even in contact) is additive (extensive), $S = S_1 + S_2$ (here we neglect the role of interfaces, if any); S is continuous and differentiable, and it increases monotonically with E (see later the case of phase transitions).

It follows that S is homogeneous 1st order function: $S(\lambda E, \lambda V, \lambda N) = \lambda S(E, V, N)$. $(\partial S / \partial E)_{V,N} > 0$ will be identified with $1/T$. This postulate also implies that we can invert $E = E(S, V, N)$, a single-valued, continuous and differentiable function. So, as fundamental relation we ca use $S(E, V, N)$ or $E(S, V, N)$, they are equivalent and contain all TD information.

postulate 4: $S = 0$ in the state for which $\partial E / \partial S = 0$ (i.e., $T = 0$).

This is Planck's extension of Nernst's 3rd law of TD ("The entropy of a thermodynamic system goes to a constant as the temperature goes to zero"). It requires QM, and it is not necessary for the bulk of TD.

Note that $S(E, V, N)$ contains the same information as $E(S, V, N)$; at first sight it is strange that E depends on S, however it is more intuitive to think that it depends on T, and the two are conjugated. Think to ideal gas: it is impossible to define E only through N, V, we need to know how much kinetic energy we have, and it is related to temperature and to entropy. We will see explicitly Sackur-Tetrode expression in CM3/TD3, these matters will become more clear.

2.5   Second law of TD

We recall main points, since it was studied before. Clear explanations can also be found on wikipedia.

The first law does not put any restriction on transforming energy under different forms: based on the first law, one might imagine that work could be freely transformed into heat (true, think to electric heaters, that have no "efficiency label"), and heat into work (false, there are well-defined limits to the efficiency of such conversion, see Carnot's cycle). The second law instead regulates possible and impossible transformations.

In reversible processes $dS = \delta Q_{rev}/T$, where we consider a closed system in equil. with bath at common temperature T, that exchanges heat. In real life, processes are irreversible and feature $dS > \delta Q_{irr}/T$, where T now is the temperature of the bath. In general we can still define $dS = \delta Q_{rev}/T$,

if we mean that, even if the process going from 1 to 2 is irreversible, one can imagine an idealized reversible process with the same endpoints, and use the reversible heat of this fictitious process to compute the entropy variation in the real process.

Spontaneous processes are those that can be observed: they feature $\Delta S > 0$, where here we mean the entropy of an isolated system, or the total entropy of system+bath (again, the whole is an isolated system). Entropy increases until the system reaches equilibrium, but TD cannot not say anything about the characteristic speed of the evolution, it might be very slow (rare events, see later).

The second law will become more interesting when introducing the statistical viewpoint.

This is how Arthur Eddington characterized the Second Law (citation from ... ?): "The Law that entropy increases-the Second Law of Thermodynamics-holds, I think, the supreme position among the laws of Nature. If someone points out to you that your pet theory of the universe is in disagreement with Maxwell's equations—then so much for Maxwell's equations. If it is found to be contradicted by observation—well, these experiments do bungle things sometimes. But if your theory is found to be against the Second Law of Thermodynamics I can give you no hope; there is nothing for it but to collapse in deepest humiliation." The nature of the physical world, Eddington: "Now the primary laws of physics taken one by one all declare that they are entirely indifferent as to which way you consider time to be progressing, just as they are indifferent as to whether you view the world from the right or the left. This is true of the classical laws, the relativity laws, and even of the quantum laws. It is not an accidental property; the reversibility is inherent in the whole conceptual scheme in which these laws find a place. Thus the question whether the world does or does not 'make sense' is outside the range of these laws. We have to appeal to the one outstanding law - the second law of thermodynamics - to put some sense into the world. It opens up a new province of knowledge, namely, the study of organization; and it is in connection with organization that a direction of time flow and a distinction between doing and undoing appears for the first time."

A safe approach to the second principle: we prepare an isolated system with some initial conditions (i.e., positions and momenta of all atoms, or spin configuration, etc.), we let it evolve naturally in time with its own dynamics (say Hamiltonian), it reaches equilibrium, we measure/compute S. Then we make a change of constraints (e.g., we increase the volume), we fix again all constraints, let it evolve naturally, after equil. is reached we measure/compute S': we find $S' \geq S$. This procedure allows looking at entropy only in equil. situations, where it is well-defined. This is why textbooks emphasize constraints in the discussion of thermodynamics, having constraints means fixing the macroscopic features (like N, V, E) of a system, so that we can consider it isolated and we can safely reach equilibrium (at least we hope: if the system components can undergo chemical or subatomic reactions, or phase transitions, or other rare events, we can be misled in detecting equilibrium! But we can still hope for local-in-time equil. if the metastable state is long-lived).

Less safe but still accepted (see Ref. [6]): the system is not in strict equil., but only in "local equil." , each local region of the system of a sufficient size appears stationary and without currents on a timescale τ long compared to atomic motion but short compared to the whole experiment, but adjacent regions have different macroscopic parameters; now we can try computing/measuring S_k in each region, and on a timescale $\gg \tau$ we will see the total $S = \sum S_k$ to increase, in agreement with the second law (the procedure allows also treating metastable systems with sufficiently-long lifetime, see previous paragraph). Probably this is what Landau refers to in figure 1 of his book. If instead we want to define/estimate $S(t)$ in a strongly out-of-equil. experiment, we could have troubles.

Note that still today there is a lot of confusion and heated discussion about the meaning and even the applicability of the second law. We'll try to clarify as much as we can with examples...

2.6   Conditions for equilibrium

We start from the fundamental relation, either $dE = ...$ or $dS = ...$, consider an isolated system divided in two subsystems, so $0 = dE = dE_1 + dE_2$ (energy conservation), or $0 = dS = dS_1 + dS_2$ (entropy maximization). In the latter case, if we consider heat transfer, i.e., $dE_i \neq 0$ but $dV_i = 0$, $dN_i = 0$ (no

mechanical- or chemical work exchange between 1 and 2) then

$$0 = dS = \frac{\partial S}{\partial E_1} dE_1 + \frac{\partial S}{\partial E_2} dE_2 = \frac{1}{T1} dE_1 + \frac{1}{T2} dE_2 = \left(\frac{1}{T1} - \frac{1}{T2} \right) dE_1$$

hence $T_1 = T_2$ (we could have done the same steps in the energy representation). Note that the above relation, once combined with the 2nd law, tells us also that energy (heat) flows from the hot to the cold body: if we start from equilibrium state with separated 1 and 2 and $T_1 > T_2$, after allowing heat exchange we reach new equilibrium maximizing S, so that $\Delta S > 0$ and $\Delta S = (T_1^{-1} - T_2^{-1})\Delta E_1$ imply $\Delta E_1 < 0$.

Similar approach yields $p_1 = p_2$ for volume exchanges and $\mu_1 = \mu_2$ for particle exchanges.

Note: we imposed $dS = 0$ but, to maximize S, we should also have $d^2 S < 0$; such stability conditions will be studied when approaching phase transitions (Callen ch. 8,9).

3 CM3

3.1 Some more details about TD?

In principle, it remains some basic TD material to explain, like types of work, types of heat transfer, stability, positivity of specific heat (Brunet 3.3.2), thermostats and barostats (Brunet 3.3.4) etc. (see Brunet ch 2 and 3, very clear). However, this is not a full course on thermodynamics, a selection of topics is analyzed.

3.2 Measuring entropy

An important practical point is how to measure entropy: when $W = 0$, $\Delta E = Q$ and we can do calorimetry, i.e., we measure ΔE for a process by measuring the heat flow, using the change in temperature as a proxy. For details see the clear discussion in Atkins [5] chapters 3A, 3B.

In practice, $S_2 = S_1 + \int_1^2 \delta q_{rev}/T$ and we can replace $\delta q_{rev} = C_V dT$ or $C_p dT$, $C_V = (\partial E/\partial T)_V$ (for ideal gas it is $3Nk/2$). But if we pass through a phase transition, we need to add the latent heat contribution to the entropy: $\Delta H_{trans}/T_{trans}$.

In other experiments, one can measure the equilibrium population of two different macrostates A, B of the system as a function of T: as we will se later, the ratio of probabilities depends on the free-energy difference $p_A/p_B = \exp[-\beta(F_A - F_B)]$, so that the T dependence provides $\Delta F = \Delta E - T\Delta S$. Once again, entropy is directly observable in experiments, irrespective of our degree of knowledge of the internal parts of the system.

Entropy is difficult to grasp theoretically, but at the same time it is something we can measure with experiments, it is not an opinion (even if scientists information-theory addicted seem to believe it is...). Let's keep the feet on earth as much as possible. As always, theoretical physics needs to be compatible with experiments, otherwise it becomes a mathematical game.

3.3 Entropy definition in SP

The fundamental bridge between TD ans SP, as well as between microscopic physics and much of macroscopic physics/chemistry/biology etc., is Boltzmann formulation $S = k_B \log \Omega(N, V, E)$, where Ω denotes the number of accessible microstates of the isolated system at the given macroscopic conditions (constraints) N, V, E, as you already saw in L2. This deceptively simple formula is supported by many plausible arguments (that of course cannot be demonstrated to be true: accord between the resulting theory and exp. supports the theory), and it has tons of implications. Much of the following lectures will deal with this. A funny example: hard spheres can crystallize at fixed T even though they do not reduce energy, but by gaining entropy, counter-intuitive (see D. Frenkel article "order through disorder: entropy-driven phase transitions" for biblio project).

Actually the formula, carved on Boltzmann's grave in Vienna, is due to Planck, and it should include an additive constant, unfortunately lost in translating from German, that is important in so-called Gibbs paradox, concerning entropy of mixing of identical gases, an intriguing issue. Note how Gibbs addressed the problem in his book, before the discovery of QM, and QM is not necessarily the only issue here: see D. Frenkel, Molecular Physics 112, 2325, 2014.

Important point: in SP we always start from isolated system, that's where we understand better the microscopic physics! Just Hamilton eq. or Schrödinger eq. (see CM1), with conservative forces, no strange additional ingredients. Then we can partition into a small and a big part etc. Let's get a flavor for the origin of Boltzmann entropy formula, and let's check if it is consistent with TD definition: it is not obvious they are the same object!

3.4 Some views of the entropy

A good starting point is the ideal gas in a box, that you saw in the thermodynamics L2 course with N. Menguy. If we have N non-interacting particles in a box of volume V, and we remove a wall (constraint) so that they can suddenly diffuse into a twice larger volume, ΔS between the two states results from doubling the available spatial configurations for each particle, so that the increase in the number of N-particles microstates is 2^N:

$$\frac{\Omega_2}{\Omega_1} = 2^N \ , \quad \Delta S = k \log \frac{\Omega_2}{\Omega_1} = Nk \log 2$$

One can think as a sequence of systems with fixed wall (constraint), closely spaced in volume, as sequence of equilibrium states virtually depicting a quasi-static transformation, albeit irreversible. Total ΔS is the same for this virtual experiment or from abrupt removal of wall $V \to 2V$, since S is state function. However, even making the expansion infinitely slow does not allow us to use the quasi-static expression $\delta Q = TdS$, since at each infinitesimal step $dS > 0$ while $\delta Q = 0$. Callen calls slow free expansion "essentially irreversible, non-quasi-static", whereas for expansion against a slow devrease of external pressure $dS = 0$, $dE = -pdV$ and you get quasi-static, reversible (isentropic), adiabatic expansion (see Callen problem 4.2-2).

We'll see in TD3 that a rapid increase of Ω results not only from ΔV but also from ΔE: remember that $S = S(E, V, N)$.

As we will see, the logarithm in the entropy is very handy, giving $S \propto N$ (extensive entropy, 1st order homogeneous function), while Ω is a cumbersome quantity to use in practical calculations, since it typically grows exponentially with N (the precise law depends on the specific system). See the TD for an example.

3.5 Illustration: classical phase space and lattice gas

Calculations with continuous $\mathbf{r}, \mathbf{p}$ are cumbersome (integrals...), and it is not obvious how to pass from phase space volumes to a discrete number Ω. The customary solution consists in plugging in a bit of QM: due to Heisenberg uncertainty principle, the most precise localization of a particle in terms of $\mathbf{r}, \mathbf{p}$ corresponds to a small "cell" of extension $dx\,dy\,dz\,dp_x dp_y dp_z \approx h^3$, so for N particles it makes sense to divide phase space in a grid of cells of extension h^{3N}, as an operative hypothesis. As usual, the machine is justified a posteriori since final results will be in agreement with experiments.

Hence, classically, we compute Ω as "number of phase-space cells of volume h^{3N} compatible with the constraint $H(\mathbf{r}_1, ..., \mathbf{r}_N, \mathbf{p}_1, ..., \mathbf{p}_N) = E$". Note that $\Delta S = k \log(\Omega_2/\Omega_1)$ will not have any dependence on the constant h.

A nice simple illustration is provided by the lattice gas, a model without momenta, only configurational entropy. Follow nice explanation of Brunet ch. 5.

3.6 Introduction to the quantum description

Following Texier ch 3.2 and 3.B: in QM the microstate of a system is specified from the wavefunction $\psi(\mathbf{r})$, or ket $|\psi\rangle$, belonging to Hilbert space. In practice we can specify quantum numbers to identify the quantum (micro)state of the system. For instance, electrons, protons and neutrons have $s = 1/2$ and we can identify the spin state of a single particle using $m = \pm 1/2$, i.e. $|\uparrow\rangle$ and $|\downarrow\rangle$, a 2D space. The state of a free particle in a box can be identified with the momentum $\mathbf{p} = \hbar\mathbf{k}$, conserved (eigenstate of $H =$ kinetic energy), with $\psi_\mathbf{k}(\mathbf{r}) = e^{i\mathbf{kr}}$; for a cubic box of side L, assuming $V = \infty$ out of the box, boundary conditions (zero wavefunction out of the box) gives quantization: $\psi_\mathbf{k}(\mathbf{r}) = \sin(k_x x)\sin(k_y y)\sin(k_z z)$, with $k_\alpha = n_\alpha \pi/L$, $n_\alpha = 0, 1, 2, ...$ so the state is characterized as $|\mathbf{k}\rangle$ or $|\mathbf{n}\rangle$, with energy $E_\mathbf{k} = \hbar^2 k^2/2m$. Clearly there can be degeneracy: for instance state $|\mathbf{n}\rangle = |551\rangle$ has the same energy as $|515\rangle$ and $|155\rangle$, due to evident symmetry (cubic box), but also as $|711\rangle$, $|171\rangle$, $|117\rangle$ ("accidental" degeneracy), so total degeneracy is $g = 6$. In general, we can indicate with $|\mathbf{l}\rangle$ the

microstate, where $\mathbf{l}$ is the array of all quantum numbers: for free electrons in a box we have $4N$ such numbers, due to spin and momentum. Since eigenstates of H are stationary, they play an important role. As usual, H is important!

Notable differences between classical and quantum SP: Pauli principle puts some important restrictions on the occupation of one-particle states, not present in classical physics, so it complicates things. On the other hand, in QM for a finite volume the possible values of the quantum numbers (e.g., the momentum) are discretized, which simplifies the statistical treatment since we have sums over a countable set of microstates, while in classical phase space we have integrals over a continuum, more cumbersome. We'll see in CM4 that in probability theory we need care when changing continuum variables, since we can get non-trivial effects (e.g., microstates that are equiprobable for a given variable s are no more equiprobable after the simple transformation leading to the variable s^3).

3.7   What is a macrostate?

Let's go back to the definition of macrostate: usually we do not know in which precise microstate the system is found at time t, and even if we do, immediately after typically it will change microstate (unless, in QM, it is a stationary state and there is absolutely no perturbation: we'll have later a deeper discussion). So from the viewpoint of TD, or, more broadly, of the kind of average properties that are usually measured in exp., we have to do with a multitude of possible microstates, each one visited with a certain probability.

Therefore a key quantity in SP, both classical and quantum, is the probability density in the space of microstates (phase space or Hilbert space). It turns out that at equilibrium such density ρ is particularly simple, and it depends on few variables like N, V, E, consistently with the TD notion of macrostate. So a given equilibrium macrostate is specified in SP by its probability density ρ in microstates-space. For the classical and quantum cases we have $\rho(\{\mathbf{q}_j, \mathbf{p}_j\})$ and $\rho = \sum_{\mathbf{l}} p_{\mathbf{l}} |\mathbf{l}\rangle\langle\mathbf{l}|$, respectively. More details in the following lectures.

4 CM4

4.1 Basic probability theory

Aim: sketch main results about distributions, averages, independence, law of large numbers, central limit theorem. Basic info is in Texier's book. Some demonstrations are done in TD.

Basic steps:

1. define idealized experiment

2. define all possible outcomes (sample space)

3. assign a probability p_k to each outcome ($0 < p_k \leq 1$, $\sum_k p_k = 1$)

each of these steps include arbitrary **hypotheses**: their validity can be tested by comparing experimental results with predictions from our calculations (e.g., Maxwell-Boltzmann vs Bose-Einstein vs Fermi-Dirac statistics...)

widespread hypothesis: **equal a priori probabilities**. Example: an atom can occupy any sub-region of volume v of a box of volume V with probability v/V. More general: all microstates of the same energy have the same probability.

- **sampling without replacement**
  e.g., drawing k cards from a deck of n cards: $n(n-1)(n-2)...(n-k+1) = \frac{n!}{(n-k)!}$ possible arrangements in specific order, or $\frac{n!}{k!(n-k)!}$ irrespective of the order

- **sampling with replacement**
  e.g., drawing k cards from a deck of n cards, putting back the card and shuffling after each draw: n^k possible arrangements

(for $n \gg k$ they are similar) e.g., probability that sampling k cards with replacement we get k different cards is therefore $\frac{n!}{(n-k)!} \frac{1}{n^k}$

4.2 Binomial distribution

if we repeat n independent experiments, each with 2 outcomes ("success" and "failure" with probabilities p and $q = 1 - p$), the probability to observe k successes irrespective of the order is

$p_B(k; n, p) = \binom{n}{k} p^k q^{n-k} = \frac{n!}{k!(n-k)!} p^k q^{n-k}$

$\sum_{k=0}^{n} p_B(k; n, p) = (p + q)^n = 1$

$\langle k \rangle = \sum_{k=0}^{n} k \, p_B(k; n, p) = np$

$\langle k^2 \rangle - \langle k \rangle^2 = npq$

$\frac{\sqrt{\langle k^2 \rangle - \langle k \rangle^2}}{\langle k \rangle} = \sqrt{\frac{q}{p}} \frac{1}{\sqrt{n}}$

important result: relative fluctuations decay like $1/\sqrt{n}$

4.3 Law of large numbers

for large n the observed fraction of successes k/n tends to the theoretical average p (justifying the interpretation of probability as frequency)

$P(|k/n - p| < \epsilon) \to 1$

i.e., it is unlikely to observe a deviation from p larger than any fixed ϵ

4.4 Moment generating function

a useful tool to compute moments of a generic distribution:

$\langle e^{sk} \rangle = \sum_k e^{sk} p_X(k) = 1 + s\langle k \rangle + \frac{1}{2}s^2\langle k^2 \rangle + ...$

$\langle k^r \rangle = \frac{d^r}{ds^r}\langle e^{sk} \rangle|_{s=0}$

4.5 Poisson distribution

approx. of binomial distrib. for large n, when $p \ll 1$, $\langle k \rangle = np \equiv \lambda \sim 1$:

$p_B(k;n,p) \approx p_P(k;\lambda) = \frac{\lambda^k}{k!}e^{-\lambda}$

$\langle k \rangle = \langle k^2 \rangle - \langle k \rangle^2 = \lambda \quad$ (for $p \ll 1$ $np \approx npq$)

e.g., for an average density of ρ points per unit surface, with uniform random distribution, a surface S (of any shape) contains on average $\lambda = \rho S$ points, and the probability to observe k points in S is $p_P(k;\lambda)$.

Similarly, for an average rate of ν events per unit time (e.g. radioactive decays, chemical reactions, protein unfolding...), with uniform random distribution, a time interval of duration t contains on average $\lambda = \nu t$ events, and the probability to observe k events in a time t is $p_P(k;\lambda)$.

Note that Poisson law is the law for a random, uniform distribution of points in a region, whenever we ask the question: "what is the prob. to find k points in a sub-region whose size is such that $\langle k \rangle$ is of the order of one?".

4.6 Normal distribution

approx. of binomial distrib. for large n, when $1 \ll \langle k \rangle = np \ll n$:

$$p_B(k;n,p) \approx p_G(k;\langle k \rangle,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}e^{-(k-\langle k \rangle)^2/2\sigma^2}$$

where $\sigma^2 \equiv \langle k^2 \rangle - \langle k \rangle^2 = npq$, while k now becomes a continuous variable, and p_G a continuous distribution:

$\langle f(x) \rangle = \int_{-\infty}^{\infty} f(x)\, p_G(x)\, dx$

$p_G(x)\, dx = $ prob. to observe system between x and $x + dx$

4.7 Random variables (also called stochastic variables)

definition: a set of numbers x, each one with a prob. $p(x)$

where x can be discrete ($\sum_x p(x) = 1$) or continuous ($\int p(x)dx = 1$)

multivariate case (e.g., 3-dim. position or velocity of atom) ("loi jointe"):

$p(\mathbf{x}) \equiv p(x_1, x_2, ..., x_n)$

$\int d\mathbf{x}\, p(\mathbf{x}) \equiv \int dx_1 \int dx_2 ... \int dx_n\, p(x_1, x_2, ..., x_n) = 1$

marginal distribution ("loi marginale") = keep information only about subset of variables:

$f(x) = \int dy\, p(x,y)\, , \quad g(y) = \int dx\, p(x,y)$

conditional prob. of y for a given value of x (Bayes' rule):

$p(y|x) = \frac{p(x,y)}{\int dy\, p(x,y)} = \frac{p(x,y)}{f(x)}\, , \quad \int dy\, p(y|x) = 1$
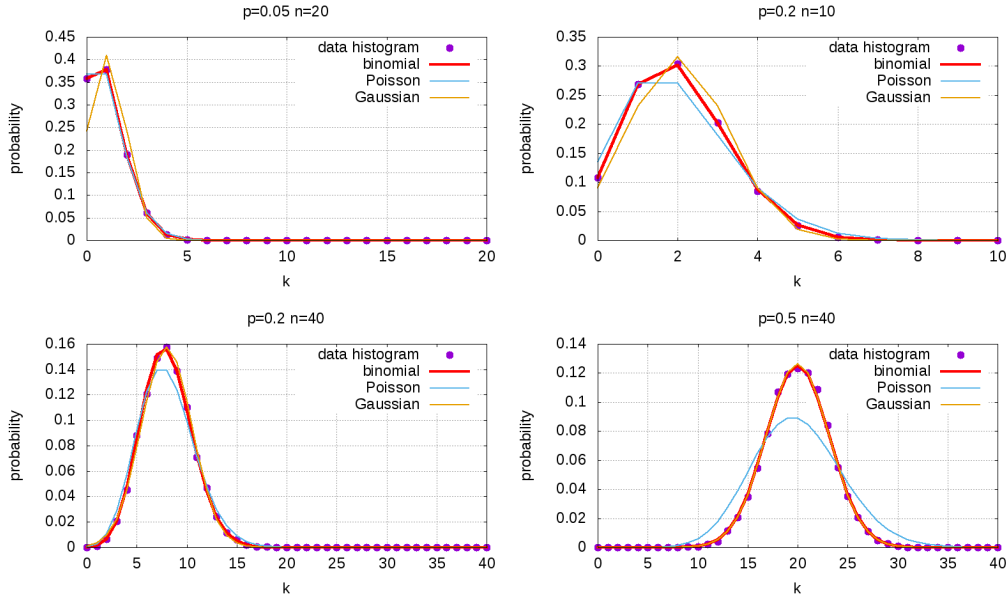
Figure 1: Data histograms (10^4 experiments) versus theoretical distributions.

if x and y are **independent** $p(x,y) = f(x)g(y)$

general rules:

$\langle u(x) \rangle = \int dx \ u(x)p(x)$

$\mathrm{Var}(x) \equiv \sigma_x^2 = \langle x^2 \rangle - \langle x \rangle^2 \ , \quad \mathrm{Var}(ax+b) = a^2 \mathrm{Var}(x)$

Chebyshev's inequality: $P(|x - \langle x \rangle| \geq t) \leq \frac{\mathrm{Var}(x)}{t^2}$ (it is unlikely to observe deviations from the mean much larger than σ_x)

Law of large numbers: for $n \to \infty$ and for any $\epsilon > 0$

$P(|\frac{1}{n} \sum_1^n x_j - \langle x \rangle| > \epsilon) \to 0$ (frequency of observation can be interpreted as probability)

in general: $\langle xy \rangle = \int dx \int dy \ xy \ p(x,y) \neq \langle x \rangle \langle y \rangle$

$\mathrm{Cov}(x,y) = \langle (x - \langle x \rangle)(y - \langle y \rangle) \rangle = \langle xy \rangle - \langle x \rangle \langle y \rangle$

$\mathrm{Var}(x+y) = \mathrm{Var}(x) + \mathrm{Var}(y) + 2\mathrm{Cov}(x,y)$

4.8 Fluctuations

Sums and products of multiple random variables $x_1, x_2, ... x_n$:

$\langle s \rangle = \langle \sum_j x_j \rangle = \sum_j \int dx_1 \int dx_2 ... \int dx_n \ x_j \ p(x_1, x_2, ..., x_n) = \sum_j \langle x_j \rangle$

$\langle s^2 \rangle = \langle \sum_j x_j \sum_k x_k \rangle = \sum_j \langle x_j^2 \rangle + \sum_j \sum_{k \neq j} \langle x_j x_k \rangle$

n independent random variables with the same distribution:

$\langle s \rangle = n\langle x \rangle \ , \quad \langle s^2 \rangle = n\langle x^2 \rangle + n(n-1)\langle x \rangle^2$

$\sigma_s^2 = n(\langle x^2 \rangle - \langle x \rangle^2) = n\sigma_x^2$

$\frac{\sigma_s}{\langle s \rangle} = \frac{\sigma_x}{\langle x \rangle} \frac{1}{\sqrt{n}}$

This is an important result: sum of many independent contributions has relative fluctuations decaying like $1/\sqrt{n}$ (e.g., weakly interacting parts of a big system, or successive processes in time like random walk).

4.9 Simple but important: the logic of estimating statistical uncertainties (error bars)

It cannot be underestimated how important it is to understand that error over the average of independent observables has behavior $\sigma/\sqrt{n}$. The formula comes simply from observing that the variance of a sum is the sum of the variances, $\mathrm{Var}(\sum A_k) = \sum \mathrm{Var}(A_k) = n\mathrm{Var}(A)$, so the variance of the arithmetic average is $\mathrm{Var}(A)/n$.

In short, it means that if we measure one value only, say A_1, it will be dispersed around the theoretical (ideal) $\langle A\rangle$ with a spread $\sigma_A = \sqrt{\mathrm{Var}(A)}$ (of course, we cannot know the variance if we do a single measurement... unless we have a theoretical prediction); if instead we measure n times $A_1, A_2, ..., A_n$ then the average of the measured values $\sum A_k/n$ will be dispersed around $\langle A\rangle$ with a spread $\sigma_A/\sqrt{n} = \sqrt{\mathrm{Var}(A)/n}$. So, the statistical uncertainty decreases with n, it is worth repeating many times an experiment (or a simulation), if we can afford it, to increase the precision of the measured average.

This applies to a huge number of relevant physical observables A, since very often A is the sum of many independent contributions (due to space and time decorrelations). Think to the temperature or the density measured at many different points of a sample, each point being sufficiently far away to be uncorrelated. Or you prepare many similar samples and you do a measure on each of them. Or you measure on a single system (at equilibrium) at different times, separated so that each measurement is uncorrelated to the others.

One can check if measurements are, indeed, uncorrelated, by inspecting the correlation function: $C(t) = \langle A(t_0)A(t_0 + t)\rangle$. This is either obtained from theory or it is estimated from the observed data (note that, at equil., $C(t)$ does not depend on t_0. For an in-depth discussion see the excellent book of Papoulis, "Probability, Random Variables and Stochastic Processes"). In any case we expect it to decay to zero in a certain time τ, in many systems and for many observables. If it is the case, we take measurements spaced by an interval τ (or longer), so that they are independent. Similarly considering space instead of time, there is typically a correlation length. Adding non-independent (correlated) samples does not decrease the statistical error, as it is easy to demonstrate.

4.10 Central limit theorem

Given n independent random variables x_j, all with a same distribution $p(x_1) = p(x_2) = ... = p(x_n)$ of generic shape and finite variance, for $n \gg 1$ the sum $s = \sum x_j$ has a Gaussian distribution:

$$P(s) \to \tfrac{1}{\sqrt{2\pi}\sigma_s} e^{-(s-\langle s\rangle)^2/2\sigma_s^2}, \quad \langle s\rangle = n\langle x\rangle, \quad \sigma_s^2 = n\sigma_x^2$$

(actually the theorem holds also for $p(x_1) \neq p(x_2) \neq ... \neq p(x_n)$, as long as all variances exist).

This is an explanation of the fact that Gaussian distributions are ubiquitous: many many observables (not only in physics) are the sum of random variables. Note also that any arithmetic average is a sum, just multiplied by a normalization.

You can find a demonstration in many places, basically you write the law of the sum from the joint law as a product of individual laws:

$$P(s) = \int dx_1 \int dx_2 ... \int dx_n p(x_1)p(x_2)...p(x_n)\, \delta\left(s - \sum x_k\right)$$

and you express the Dirac delta as

$$\delta\left(s - \sum x_k\right) = \frac{1}{2\pi}\int dk\, e^{-ik(s-\sum x_k)}$$

giving

$$P(s) = \frac{1}{2\pi}\int dk\, e^{-iks}\left(\int dx\, e^{ikx}p(x)\right)^n$$

the integral in parenthesis is the characteristic function of $p(x)$ with imaginary argument. Assuming for simplicity $\langle x \rangle = 0$ and approximating the characteristic function to second order in x gives

$$\left(\int dx \, e^{ikx} p(x) \right)^n \approx \left(1 - \frac{\sigma_x^2 k^2}{2} \right)^n = \left(1 - \frac{\sigma_s^2 k^2}{2n} \right)^n \approx e^{-\sigma_s^2 k^2 / 2}$$

finally, Fourier back-transforming

$$P(s) = \frac{1}{\sqrt{2\pi}\sigma_s} e^{-s^2/2\sigma_s^2}$$

and of course replace s^2 in the exp with $(s - \langle s \rangle)^2$ if $\langle x \rangle \neq 0$.

Note that for n smaller than 100 you get already a nice Gaussian shape, until you look far from the average: the queues can be quite different from the Gaussian ones if n is not very large, be careful. Large deviation theory deals with this.

5 CM5

We go deeper intro entropy, including some discussion of information theory, and we see how stochastic models (like the random walk) are at once faithful approximations of the real dynamics, and good illustrations of difficult concepts like the second principle, irreversible behavior, approach to equilibrium, kinetics.

5.1 A preliminary warning

Basic idea: in isolated system we have many available equiprobable states with a mechanism generating transitions between them (e.g., some kind of dynamics) $\implies$ S grows whenever constraints are removed, i.e., when more microstates become available. In principle, to define entropy in the orthodox way we should have constraints fixed for a long time (how long? difficult question), so that the system can explore each microstate over and over again. This condition would ensure that the occupation time of each microstate divided by the total time corresponds well to $1/\Omega$ (the theoretical probability), so that it makes sense to write $S = k \log \Omega$, since Ω is really the number of accessible (better, accessed) microstates. This is called ergodicity.

However, if we take this strict condition literally, it never happens! We never have enough time to guarantee this, even in relatively small systems of thousands of particles we would need astronomic time scales, while our experiments take seconds, hours or days (not to speak of simulations: nanoseconds or microseconds). If the system is not isolated but in contact with a thermostat, microstates are no more equiprobable but some are exponentially less probable than others, does it improve or worsen the situation? The system+thermostat is anyway an isolated system.

Despite of these problems our formulas for entropy generally give correct results compared with experiments... This apparent paradox is not trivial to solve, and seldom discussed. Probably, in many systems, phase space is a huge boring repetition of similar motifs, like a huge land where each squared kilometer is rather flat and contains very similar things, like the same amount of rocks, trees, ponds etc. In, this case, if you sample one squared kilometer or 50 or 1000 the statistical properties (like average tree density and so on) would be very similar. It is difficult at this stage to get much clearer ideas.

5.2 Some arguments motivating $S = k \log \Omega$: prob. theory & information

I follow Texier ch. 4. To build statistical physics we pass from the mechanical description, i.e. a trajectory following the laws of motion, to a probabilistic description, i.e. we look at the occupation of the different possible states. In a plot of observable vs time, we take as basis for the analysis not the time axis but the state axis.

The central question passes then from "how to compute the time evolution?" to "how to compute the probability distribution?". Empirical viewpoint: let's first take a very long and fully-detailed trajectory, visiting many times each state, then let's count probabilities as occupation frequencies (like tossing coins), $p_\alpha = t_\alpha / t_{tot}$, i.e., the time spent in a state (accumulated over many visits) over the total time of the trajectory. Another viewpoint, from information theory: if we have Ω microstates and we know nothing about how much they are occupied, we adopt (tentatively) the principle of minimal information (or minimum bias), i.e. we assign equal prob. $1/\Omega$ to each microstate. It looks like the equal a priori prob. postulate ("postulat fondamental" in French): it can be shown experimentally that it works for isolated system with fixed energy. However, it does not work for a system in contact with a thermostat: no matter if we know nothing about the probabilities, in this case we cannot just adopt the information theory viewpoint and say that each microstate has equal prob. because we are ignorant...

Maximal ignorance: we only know the number of accessible microstates Ω, uniform $p_\alpha = 1/\Omega$. We could introduce entropy as measure of missing information: it can only depend on Ω, $S(\Omega)$. If Ω grows, then S should grow because we know less (dispersion over more microstates). If we know perfectly, then $p_\alpha = \delta_{\alpha\gamma}$ and the missing information is zero, $S = 0$. If we have two independent systems with uniform distributions, $p_\alpha = 1/\Omega$ and $p'_\gamma = 1/\Omega'$, then $P_{\alpha\gamma} = 1/\Omega\Omega'$, as total number of

joint microstates is $\Omega\Omega'$ (consider, e.g., 3×3 table of possible microstates...). By correspondence with entropy in thermodynamics, we want missing information S to be extensive, so $S_{tot} = S(\Omega) + S(\Omega')$. Putting all together these properties we find that $S \propto \log(\Omega)$, and taking $S = k_B \log(\Omega)$ allows to match statistical and thermodynamic entropies.

So far, our use of information theory and the introduction of e.a.p.p. postulate appear a plausible tentative, to be discarded if it does not match with the laws of mechanics, i.e., with Hamiltonian dynamics for isolated systems! But it works (with further decisions: e.g., use momenta instead of velocities, something difficult to see in information theory, by the way): trajectories in phase space are compatible with this probabilistic approach. Simple example: gas with collisions fills uniformly the configuration space in the box, if given enough time, so it is compatible with uniform probabilities in the configuration space. So e.a.p.p. is behind calculation of ideal gas entropy using integrals in phase space with uniform weight in the allowed region (i.e., for fixed energy, on the surface of the momenta hypersphere).

As a counterexample we could consider a crystal, as system without uniform space probabilities: however this is spontaneous symmetry breaking (see Anderson "More is different" beautiful article), infinitely many translations are allowed, in principle, so that in principle prob. is uniform. However, in practice on a given timescale we do not observe equal probabilities for all locations: the problem of ergodicity...

I somehow disagree with Balian, "Information in statistical physics" [1] where he (like many physicists) claims we need stat. phys. because i) of our ignorance about details of the system in classical physics (I contest that even if we know everything stat. phys. keeps working equally well and gives useful results, we need it irrespective of microscopic information), ii) because of intrinsic impossibility to know all information in quantum physics, thus we need to get used to probabilistic language. However I prefer to keep distinct the different problems: in classical physics the frequentist interpretation of probability gives empirical results (e.g., in simulations) that give great match between stat. phys. and microscopic physics, without any fundamental need of information theory concepts based on ignorance of the observer. The situation in quantum physics is peculiar due to the influence of the observer on the state of the system, I will not discuss this here.

Note that, even when I know everything about the system, for example in a computer simulation where the trajectory of the particles (or the evolution of the spins in the Ising model) is perfectly known, I can still estimate probabilities from frequencies of occupation (on a fine or coarse resolution) and compute entropy or any other statistical property finding that $S = k \log \Omega$ works well in isolated system, while Gibbs formula $S = -k \sum p_\alpha \log p_\alpha$ (we'll see in CM6) works also for the fixed T case. The point I try to make is that our ignorance or knowledge of the microscopic trajectories does not change the results, we still get the same thermodynamic quantities in agreement with experiments if we use the right formulas. The right formulas are those coming from theories that could pass the test of experiments, I don't think such formulas can be "demonstrated a priori" to be true based on information theory. The latter theory seems to give analogies rather than foundations. Note also that, historically, Boltzmann and Gibbs introduced the correct ways of computing entropy decades before Shannon works, with Shannon entropy being a copy of Gibbs entropy.

Sometimes we read that, if we know perfectly the situation of the system, i.e. which microstate it occupies, then the information is maximal, the state has prob. 1 and the others zero, and we get zero entropy. This statement requires careful analysis: it is true that $\Omega = 1$ and $S = 0$ when the system is stuck forever in one microstate (e.g., the ground state at zero Kelvin), however if we know that at a given t the system is in a given microstate, while it will move to other microstates immediately after, we cannot conclude that $S = 0$ just because we have perfect information: entropy must depend on the total number of accessible microstates at equilibrium. If this number is Ω, then $S = k \log \Omega$ irrespective of our knowledge or ignorance of the sequence of microstates populated in time, for example $1 \to 3 \to 10 \to 4 \to 2 \to 15...$

As mentioned in CM3, entropy can be directly measured experimentally, irrespective on our information about the inner variables of the system, e.g., from the T-dependence of the occupation of macrostates $p_A/p_B = \exp[-\beta(F_A - F_B)]$, with $\Delta F = \Delta E - T\Delta S$.

---

[1] https://hal.archives-ouvertes.fr/hal-00012785

5.3 Microscopic reversibility versus macroscopic irreversibility

It is instructive to look at simulation where N particles wander in a box and, from time to time, localize themselves again in the left side of a box: this is not in contradiction with second principle, for two reasons: First, if the particles localize for a short time on one side and then wander again in the whole box, we should not say that entropy reduced at that time, since entropy is not an instantaneous property but requires the system to explore the accessible states. Second, in a large system we should wait an enormous time to see all the particles going to the left of the box, so the dynamics of a macroscopic system appears irreversible (imagine starting with all particles on the left: they go everywhere and never come back within the age of the universe).

Nevertheless, the microscopic equations of motion are time-reversible, while macroscopic evolution (e.g., the diffusion eq.) appears irreversible, and this caused and still causes a lot of controversies. For example, Landau in figure 1 of his book Stat Phys 1 or Ambegaokar and Clerk in their article "Entropy and time" for the bibliographic project, are tempted to define entropy as a function of time $S(t)$, leading to apparent decrease of entropy during the evolution of an isolated system.

Instead of introducing $S(t)$, a tricky and slippery notion, I prefer to consider large deviation from equilibrium of an observable. Let's just take number of gas particles in a small region, where gas can be real but at low density. $n(t)$ displays all the necessary phenomenology, one can estimate likelihood of fluctuations away from average (and compare with binomial law). If initially we are very far from average, n will "appear to relax" towards the average, albeit always fluctuating. It can come back to initial situation, just it takes very long if $n(0)$ is very different from average. If it takes too long to be practically observable, we say two things: 1) the evolution is (practically) irreversible, 2) the initial condition was very special thus prepared "by someone", because we do not expect it to occur spontaneously in the available time. All of this could apply to the whole universe? Just we do not seem to be close to reaching equilibrium yet, maybe because there are all sorts of subatomic reactions going on at different energy and time scales since the big bang? Landau argues in the beginning of his book that we should not consider the universe as a whole as an isolated system. By the way the big bang is just an extrapolation of an observed trend of smaller and smaller universe in the past, we cannot be sure there was really a zero-size singularity, right? Disclaimer: cosmology is far from my field of expertise; when I tried to pose questions about the applicability to the whole universe of the second principle to colleagues experts of high-energy physics and string theory, I could never receive an answer... I will try again.

Poincaré recurrence theorem (classical and quantum) adds some interesting information: if we have an isolated conservative system with some initial conditions, the trajectory will bring it back arbitrarily close to such initial conditions after the so-called recurrence time. The latter can be huge in a system with many degrees of freedom. The basic idea is that an isolated system has a finite and fixed accessible number of states (microstates in phase space or in quantum mechanics), so, given enough time, it will visit any one of them, including the initial conditions. Note that, according to the theorem, a broken egg will certainly form again spontaneosuly some day in the future, if we wait long enough! What saves the (apparent) irreversibility of macroscopic evolution would then just be the enormous recurrence times.

Interesting readings to go deeper: Ter Haar Rev. Mod. Phys. 1955 "Foundations of statistical mechanics" is an analysis of reversibility vs irreversibility, ergodic theorems, H-theorem, and so on. Key point seems to be Markovianization, introducing the idea of decorrelation and relaxation time, i.e., the system loosing memory of initial conditions. If relaxation time is much shorter than Poincaré recurrence time, the system appears as irreversible and the second principle is safe. Becker "Theory of heat" section C on microcanonical ensemble clearly refers also to these points.

5.4 Random walk and diffusion

(Follow the slides on moodle)

One can stress usefulness of Markov processes in statistical physics, the simplicity of their formulation, and how they naturally bring equilibrium distribution and irreversible approach to it. The price to pay is a non-trivial connection to reversible Hamilton eq., their non-trivial justification, that comes

from coarse graining (neglect of degrees of freedom + coarse time scale).

The figures in the slides justify, empirically, how random walk is a good model of a system's trajectory when looking at a single particle with a bad time resolution. Note that the resolution is the key: even without large difference in mass between heavy solute (e.g., colloid) and light solvent, still we observe Markovian random walk if the time resolution is coarse-enough (e.g., water diffusion in water, one atom in a crystal, etc.). So the crucial point is understanding that when we coarse grain the Hamiltonian deterministic evolution of an isolated system we obtain an evolution of the coarse-grained variables that is apparently random and irreversible (only apparently, of course), and it also appears Markovian only if the time resolution is coarse enough. These concepts are not easy to appreciate in depth because they require familiarity with the theory of stochastic processes.

6 CM6

(The beginning of the lecture is dedicated to correct together the CC1 exam)

Here we introduce a fundamental tool: the canonical ensemble, with the connected expressions for probability of microstates, entropy, and free energy. These will be the workhorses for much of future work.

6.1 From NVE to NVT: from energy to free energy

Recall the fundamental relation: $dE = TdS - pdV + \mu dN$ it tells us several things: S, V, N are the control variables (fixed for isolated system at equilibrium), there are three pairs of conjugate variables (T, S), $(-p, V)$, (μ, N), related between them such that $T = \partial E/\partial S$, $-p = \partial E/\partial V$, $\mu = \partial E/\partial N$. Recall also that it is equivalent to use the entropy representation, $dS = ...$, obtained trivially from $dE = ...$, that changes variables to E, V, N. We can do more subtle changes of variables, to fit experimental constraints: often T instead of E or S is controlled, or p instead of V. We apply Legendre transformation: to switch the control variable inside a conjugated pair, e.g., from S to T, we subtract from E the product of the pair:

$$F = E - TS \; , \quad dF = dE - S\,dT - T\,dS = -S\,dT - p\,dV + \mu\,dN$$

we introduce Helmoltz free energy $F(T, V, N)$, very useful, and we identify $-S = \partial F/\partial T$, $-p = \partial F/\partial V$, $\mu = \partial F/\partial N$. We'll see further transformations in the future, but the latter one is arguably the most widespread and useful. So far we worked with macroscopic TD, let's consider the SP viewpoint next.

6.2 The canonical probabilities

In the isolated system case, NVE (microcanonical), we linked macro to micro by postulating $S(N, V, E) = k \log \Omega(N, V, E)$. In the NVT case (canonical) there is a thermostat ensuring constant T, E is no more conserved, how do we proceed? Well, again as before: we consider a large isolated system (since we know how to deal with it) and identify a small subsystem (our target) in thermal contact with the rest (the thermostat, or bath), without exchange of volume or particles. We make a single, crucial hypothesis: the two regions interact relatively weakly (bulk energies much larger than interface energies), so that $E \approx E_1 + E_2$, $S \approx S_1 + S_2$, $\Omega \approx \Omega_1 \Omega_2$. This latter expression, stating statistical independence, is the key to microscopic derivations (here we demonstrate $e^{-\beta E_\alpha}$ as in Callen and, partially, Reif, similar to Texier; Brunet passes through properties of thermostats).

In particular, we can ask the question: how probable (in the frequency sense) is it to observe the small subsystem 1 in microstate α, irrespective of the microstate γ of subsystem 2? All microstates of the total system, i.e., pairs (α, γ), are equally probable, but a given α might correspond to more γ than another, yielding more microstates of the whole system, hence higher probability. In particular, since $E = E_1 + E_2$, reducing E_1 means increasing E_2 hence larger Ω_2 (since, generally, it grows exponentially fast with energy: think to many particles, many ways to satisfy a given total energy in subsystem 2), hence high probability, and the opposite for increasing E_1 (but we always have $E_1 \ll E_2$, since subsystem 1 is much smaller than 2).

In practice, if we consider a single α, $\Omega_1 = 1$ and $p_\alpha = \Omega_2/\Omega$ is the fraction of microstates compatible with α. The only functional dependence on α is in Ω_2: $p_\alpha = \Omega_2(E_2)/\Omega(E) = \Omega_2(E - E_\alpha)/\Omega(E)$ so we can consider, from the viewpoint of the index α, p_α equal to a constant times a function of E_α.

Since $E - E_\alpha$ is close to E (subsystem 1 is much smaller than subsystem 2), we try a Taylor expansion, using $\log \Omega_2$, instead of Ω_2 since the latter grows too fast, exponentially fast, the expansion would not converge. E.g., expanding $f = (1 + x)^N$ with $x \ll 1$ gives troubles: $1 + Nx + N(N-1)x^2/2 + ...$ and for very large N we can have $Nx > 1$ and the second order term larger than the first order term; expanding instead $\log f = N \log(1 + x)$ poses no troubles, each term is proportional to

N and higher order terms are smaller than lower order terms. The entropy is well-behaved, extensive, easier to manipulate mathematically. Another viewpoint is just: let's manipulate the entropy, a physical quantity that we can measure.

$$p_\alpha = \exp\left\{k^{-1}[S_2(E - E_\alpha) - S(E)]\right\} = \exp\left\{k^{-1}[S_2(E - E_\alpha) - S_1(E_1) - S_2(E_2)]\right\}$$

It is important to note, here, that $E_1 \equiv \langle E_1 \rangle \equiv \sum_\alpha E_\alpha p_\alpha$, since subsystem 1 is not isolated but in contact (albeit "weak") with subsystem 2, and vice versa $E_2 \equiv \langle E_2 \rangle = E - \langle E_1 \rangle$. You need some elasticity to follow the steps without getting lost.

Focusing on the argument of the exponential, with a small trick we make $S_2(E_2)$ disappear:

$$S_2(E - E_\alpha) - S_1(E_1) - S_2(E_2) = S_2(E - E_1 + E_1 - E_\alpha) - S_1(E_1) - S_2(E - E_1)$$

$$\approx S_2(E - E_1) + \left(\frac{\partial S_2}{\partial E_2}\right)_{E - E_1}(E_1 - E_\alpha) - S_1(E_1) - S_2(E - E_1)$$

$$= \frac{1}{T_2}(E_1 - E_\alpha) - S_1(E_1)$$

where we can call $T_2 \equiv T$ ($\beta = 1/kT$), since subsystem 2 is a very large bath, it imposes a constant temperature to the small subsystem 1. We reach an important result:

$$p_\alpha = \exp\left\{\beta[(E_1 - TS_1) - E_\alpha]\right\} = e^{\beta F_1} e^{-\beta E_\alpha}$$

Now we notice two last things. First, we have an expression containing – formally – only quantities relative to subsystem 1, i.e., F_1 and E_α, even though T is there to remind us of the existence of a thermal bath; hence, from now on, we just write F, E, S instead of F_1, E_1, S_1, since we focus on subsystem 1. Second, probabilities must be normalized: we identify the partition function as the exponential of the free energy:

$$\sum_\alpha p_\alpha = 1 \implies e^{-\beta F} = \sum_\alpha e^{-\beta E_\alpha} \equiv Z$$

$$p_\alpha = \frac{e^{-\beta E_\alpha}}{Z}$$

This is a fundamental expression, the canonical probability, the most used formula in statistical physics.

Take-home-message: if we increase the energy of the small subsystem and reduce by the same amount the energy of the large bath, we increase a bit Ω_1 but we reduce a lot Ω_2, it is not symmetric due to the difference in size, so the net effect is reducing strongly the total number of microstates $\Omega = \Omega_1 \cdot \Omega_2$ of the subsystem+bath, hence the exponential decrease of probability.

Be careful, however, since this reasoning concerns a single microstate α: sometimes there is a group (let's call it A) formed by a vast number of microstates all with the same energy, so the total probability $p_A = Z^{-1} \sum_{\alpha \in A} e^{-\beta E_\alpha}$ to find the system in one of them can be large even though their energy E_α is quite high; in such a case we say that entropy wins over energy. We'll see more about this in the following lectures. As a simple example, one can discuss two ions in water, or two proteins, that attract each other (water being the bath): when they form a complex, energy is lower but the complex is rigid, few microstates, while when they are far away the energy is higher but there is a number of same-energy microstates proportional to the size of the box, so for a large enough box the dissociated state is more probable, even for a strongly attractive interaction!

Z will play an important role in many theoretical calculations, even though it is not something we measure; rather, we can measure the free-energy difference between two similar systems prepared in different ways, for example liquid and solid, reactant and product of a reaction, etc: the ratio of the corresponding partition functions is the exponential of the free-energy difference, a fruitful point of view for complex systems, as we will see with phase transitions.

7 CM7

We go deeper in the canonical ensemble, trying to understand the exponential in the probability formula and giving a new, more general definition of entropy.

7.1 Illustration of the Boltzmann weight

To illustrate better why we get an exponential (the so-called Boltzmann weight) in the formula $p_\alpha = e^{-\beta E_\alpha}/Z$, let's consider explicitly the ideal gas: according to the reasoning in the previous section $p_\alpha \propto \Omega_2(E - E_\alpha)$, with $E_\alpha \ll E$, and in the ideal gas we know that $\Omega \propto E^{3N/2}$, so for large N_2 (i.e., large bath) we find

$$p_\alpha \propto (E - E_\alpha)^{3N_2/2} = E^{3N_2/2}\left(1 - \frac{E_\alpha}{3N_2/2}\frac{3N_2/2}{E}\right)^{3N_2/2} \approx E^{3N_2/2}e^{-E_\alpha/\epsilon}$$

where we used the fact that $3N_2/2E$ is constant for growing N_2 (since $E \approx E_2$ and the energy is extensive), and the fact that $(1 + x/n)^n$ tends to e^x (see the analogy when the Poisson law is obtained as approximation of the binomial law). The exponential is an excellent approximation already for $N_2 \sim 100$, as can be checked by plotting the curves. Note that $\epsilon = E/(3N_2/2)$ corresponds here to kT, explaining why we commonly say that kT is a sort of "thermal energy" available to the microscopic degrees of freedom. In the case here the equipartition theorem predicts that each of the $3N_2$ momentum degrees of freedom (the only ones contributing to the energy of the ideal gas, in a quadratic fashion) have, on average, $kT/2$ energy.

The behavior $\Omega \propto E^{cN}$, with c a constant close to unity, is typical of many different systems, not only the ideal gas, and it is associated to the fact that the system is made of many statistically independent parts. For example, a collection of harmonic oscillators has $\Omega \propto E^{3N}$, leading again to the exponential Boltzmann weight as above.

7.2 Probabilities of macrostates depend on energy and entropy

At first sight, $p_\alpha = e^{-\beta E_\alpha}/Z$ seems to say that the only thing that matters to determine the probabilities is the energy: microstates of lower energy are more probable to be observed than those of higher energy. The reality is more subtle: in almost any experiment or simulation on a realistic, complex system, we do not study single microstates but, rather, collections of many microstates, called macrostates. For example, the states analyzed are the liquid versus solid phase of a sample, the folded versus unfolded conformation of a protein in water, etc.

In all these cases, we are interested into the probability of a macrostate A, for example the liquid phase (many microstates correspond to the liquid, but also to the crystal considering thermal motion around the ideal positions): $p_A = \sum_{\alpha \in A} p_\alpha = Z_A/Z = e^{-\beta(F_A - F)}$ where we introduced two new notions: the partition function restricted to a macrostate, Z_A, and the free energy of a macrostate, F_A. We will see that this viewpoint is very useful in many studies of phase transitions, chemical reactions etc.

A consequence of $p_A = \sum_{\alpha \in A} p_\alpha$ is that two given macrostates, one composed by few low-energy microstates and the other by many high-energy microstates, can be equally probable! The first one, we say, is stabilized by its low average energy, the second by its high entropy.

A real example is protein folding: in the folded state, compact and quite rigid, the energy is low but the entropy is also low, while in the unfolded state, formed by a huge number of disordered open-chain configurations, energy is quite high but entropy is also high: depending on T, one dominates over the other (as you can imagine, higher T stabilizes the unfolded state and destabilizes the folded one). In general, both energy and entropy are important to determine the probabilities of macrostates.

Finally, note that $p_A/p_B = \exp[\beta(F_B - F_A)] \equiv \exp(\beta\Delta F)$, and, since $\Delta F = \Delta E - T\Delta S$, if we measure the change of p_A/p_B in a small T-interval, where we can assume that ΔE and ΔS are

almost constant, we can fit the results to extract the latter two quantities. Experiments of this kind are routinely performed in different fields.

7.3   The phase-space viewpoint

One could study the prob. density of the position, $\rho(x)$, for 1D harmonic oscillator comparing NVE and NVT: we pass from "catenaria" (analytical $\rho(x)$ computed in PPM homework 2022) to Gaussian. When we add thermostat, we allow different harmonic oscillator orbits in phase space, and those smaller (lower E) are more probable (see previous exercise), so we get Gaussian cloud... A numerical illustration could also discuss the passage from linearized to chaotic pendulum.

7.4   General (Gibbs) definition of entropy

Thanks to the expression for p_α combining together TD (F) and SP (E_α), we are now ready to find a new explicit microscopic expression for S, easier to apply than $k \log \Omega$, that works – we will see – for all ensembles NVE, NVT, NpT, ...

$$F = -kT \log Z = \langle E \rangle - TS \ \Rightarrow \ TS = \sum E_\alpha \frac{e^{-\beta E_\alpha}}{Z} + \beta^{-1} \log Z$$

$$= -\beta^{-1} \sum \log \left(e^{-\beta E_\alpha} \right) p_\alpha + \beta^{-1} \log Z$$

$$= -\beta^{-1} \sum \left[\log \left(\frac{e^{-\beta E_\alpha}}{Z} \right) + \log Z \right] p_\alpha + \beta^{-1} \log Z = -\beta^{-1} \sum p_\alpha \log p_\alpha$$

This is a very important result, Gibbs expression for the entropy, $S = -k \sum p_\alpha \log p_\alpha$, that can be checked to hold also in isolated system ($p_\alpha = 1/\Omega \Rightarrow S = k \log \Omega$) and in other ensembles if we use the corresponding expressions for p_α. We will explore several properties of of Gibbs entropy.

The formula is found also from other reasoning, e.g., using Lagrange multipliers and the multinomial generalization of the binomial coefficient, in case of many identical subsystems, or from information theory ideas (Shannon entropy, see e.g. Texier 4.2.3). Here it comes purely from physical approach, using our understanding of isolated system plus the hypothesis of almost-independence $\Omega_{tot} \approx \Omega_1 \Omega_2$.

Remark: even when the two subsystems are not independent, we can still make a gedankenexperiment (i.e., consider a virtual situation) where an adiabatic wall renders 1 and 2 isolated systems, really independent from each other, with fixed E_1 and E_2, so that we compute total entropy S_{tot} from $\Omega_{tot} = \Omega_1 \Omega_2$, we try varying E_1 and E_2 at fixed E_{tot}, all of this to be able to compute probabilities of the case where the wall is diathermal. It is the same approach used to introduce the idea of entropy maximisation in equilibrium.

For the brave, the mathematically-oriented article of Goldstein et al, "Gibbs and Boltzmann Entropy in Classical and Quantum Mechanics" (https://arxiv.org/pdf/1903.11870) gives an interesting viewpoint, according to which i) Boltzmann entropy is the pillar we can build upon, ii) we have to be careful when defining Gibbs entropy in phase space since some common views are misleading, and iii) information theory does not really help defining or understanding physical entropy. I tend to agree with the authors of this article, and I add that molecular dynamics simulations very strongly support Boltzmann viewpoint. Amazing how in 2024 scientists keep debating in a very heated way about these fundamental points!

7.5   Extensivity and maximisation of canonical entropy

If we have two subsystems, A and B separated by an adiabatic wall (so they are independent from each other), the joint probability of a microstate of the whole system is the product of the marginal probabilities, as usual, and this leads to the extensivity of Gibbs entropy:

$$P_{\alpha\gamma}^{AB} = p_\alpha^A \cdot p_\gamma^B$$

$$S_{AB}/k = -\sum_\alpha \sum_\gamma p_\alpha^A p_\gamma^B \log(p_\alpha^A p_\gamma^B) = -\sum_\alpha \sum_\gamma p_\alpha^A p_\gamma^B (\log p_\alpha^A + \log p_\gamma^B)$$

$$= -\sum_\alpha \left(\sum_\gamma p_\gamma^B\right) p_\alpha^A \log p_\alpha^A - \sum_\gamma \left(\sum_\alpha p_\alpha^A\right) p_\gamma^B \log p_\gamma^B = S_A + S_B$$

where we used the normalization of the marginal probabilities $\sum_\alpha p_\alpha^A = \sum_\gamma p_\gamma^B = 1$. Note that the situation above corresponds to neglecting the interaction between the subsystems A and B: in the canonical ensemble $E_{AB} \approx E_A + E_B \implies P_{\alpha\gamma}^{AB} = e^{-\beta(E_\alpha^A + E_\gamma^B)}/Z = (e^{-\beta E^A}/Z_A) \cdot (e^{-\beta E_\gamma^B}/Z_B) = p_\alpha^A \cdot p_\gamma^B$. However the result is valid also in the microcanonical ensemble (remember that Gibbs entropy formula remains valid).

Now we demonstrate that the canonical probabilities maximise the entropy. Consider a system where in two different situations: in the first case we have a canonical probability distribution $p_\alpha = e^{-\beta E_\alpha}/Z$, in the second case we imagine (a gedanken experiment, again) the system to have a different probability distribution $p_\alpha' \neq p_\alpha$ (of course the microstates are the same and E_α are the same, since the system is the same, they have just different probabilities). To make a fair comparison we also consider the two situations to have the same average energy:

$$\sum p_\alpha E_\alpha = \sum p_\alpha' E_\alpha = \langle E \rangle$$

let's now compare the entropies:

$$\frac{S'-S}{k} = \sum_\alpha \left(-p_\alpha' \log p_\alpha' + p_\alpha' \log p_\alpha - p_\alpha' \log p_\alpha + p_\alpha \log p_\alpha\right) = \sum_\alpha p_\alpha' \log \frac{p_\alpha}{p_\alpha'} + \sum_\alpha (p_\alpha - p_\alpha') \log p_\alpha$$

The second term is

$$\sum_\alpha (p_\alpha - p_\alpha')(-\beta E_\alpha - \log Z) = -\beta(\langle E \rangle - \langle E \rangle) - \log Z(1-1) = 0$$

so that

$$\frac{S'-S}{k} = \sum_\alpha p_\alpha' \log \frac{p_\alpha}{p_\alpha'} \leq \sum_\alpha p_\alpha' \left(\frac{p_\alpha}{p_\alpha'} - 1\right) = 0$$

where we used $\log x \leq x - 1$ (easy to prove), so we find that $S \geq S'$, canonical probabilities maximise entropy with respect to other distributions, at fixed $\langle E \rangle$. Note that this is valid also in the microcanonical ensemble, because we can still write the probabilities with the canonical formula for the special case of fixed $E_\alpha \equiv E \; \forall \alpha$:

$$p_\alpha = \frac{e^{-\beta E}}{Z} \;, \quad Z = \sum_{\alpha=1}^{\Omega} e^{-\beta E} = e^{-\beta E} \sum_{\alpha=1}^{\Omega} 1 = e^{-\beta E} \cdot \Omega \implies p_\alpha = \frac{1}{\Omega}$$

8 CM8

## 8.1 Negative temperature and population inversion

When temperature T is interpreted as proportional to the average kinetic energy, $3NkT/2 = \langle K \rangle$ (equipartition), as it happens in a system where particles can move and have a velocity, then T cannot be negative. Likewise, in a system whose energy spectrum has no upper limit, the thermodynamic definition $T = \partial E / \partial S$ gives $T > 0$ since increasing E typically leads to an increase of S.

However, one can consider (as Onsager did) cases where the energy of the system is bounded from above, i.e., with E having a maximum. For instance, a simple case is a two-level system where each of the N particles can have energy either 0 or $\epsilon > 0$: in this case $E \in [0, N\epsilon]$. Note that the ratio of the occupation probability for the single-particle states, $p_\epsilon / p_0 = e^{-\beta\epsilon}$, when a thermostat is present, tends to zero at low-T and to 1 at high-T. If one starts with an initial state where all particles are in the zero-energy state, increasing E increases the number of available configurations and thus $S = k \log \Omega$ increases, giving $T > 0$.

Consider what happens if we start from all particles in the excited state: first of all, this situation should not normally happen, because, as written above, when in equilibrium with a thermostat $p_\epsilon / p_0 \leq 1$, so that the situation $n > N/2$, called population inversion, is usually realised with some artificial means, like in lasers (very interesting topic!). When we start with $n > N/2$, decreasing E leads to an increase of S, thus, formally, $T < 0$. Note that here we are considering the temperature of the two-level system, not the (positive) temperature of a thermostat. If n particles are excited, the number of microstates in the isolated system is $\Omega = N!/(n!(N-n)!)$ and S has a maximum for $n = N/2$, so T will be positive for $0 \leq E < N\epsilon/2$ and negative for $N\epsilon/2 < N \leq N\epsilon$. Using Stirling approximation

$$S/k = \log \Omega \approx N \log N - n \log n - (N - n) \log(N - n)$$

so that $S(n = 0) = S(n = N) = 0$, while the top of the curve $S(n = N/2)$ corresponds to the most likely macrostate (related to the equilibrium state when in contact with a thermostat), as expected.
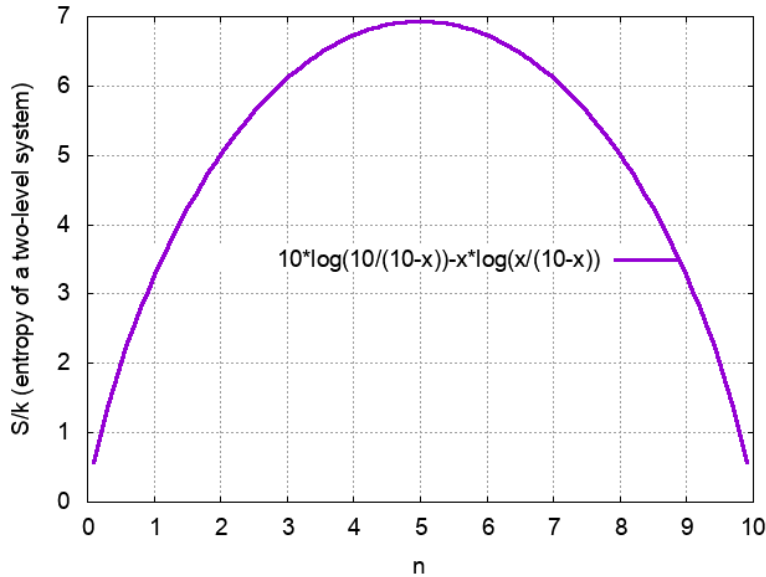
Using $1/T = \partial S / \partial E$ one finds

$$T = \frac{\epsilon}{k \log[(N - n)/n]}$$

that behaves as expected (for $n = N/2$ the temperature is not defined, it suddenly jumps from very positive to very negative values across this point).

In this and other examples (like spin systems, that lack kinetic energy) when the system with $T < 0$ is put into contact with a thermostat at $T' > 0$, it is curious to observe that heat (energy) flows towards the reservoir, so that "negative temperatures are hotter than positive ones".

This treatment applies also to the end of the article of Ambegaokar & Clerk, Entropy and time, given for the bibliographic project.

9 CM9

9.1 The ergodic approach and the arrow of time

The present lecture is based on slides (available on moodle) that are shown on the projector. An interesting discussion can also be found in the book of Texier.

The aim is to compare the frameworks of Gibbs ensembles versus the one of the ergodic trajectory, in simple terms.

To illustrate theory, showing that it works, in particular for entropy estimation, one can consider several examples: random walk, molecular dynamics, Ising Monte Carlo. The message is that details of dynamics do not matter, what matters is that we have a working mechanism that induces transitions between microstates.

Consider phase-space trajectories for a collection of N hos interacting among them: even for the NVE case, looking at one of the ho shows Maxwell-Boltzmann law, typical of canonical NVT situation. In other words, a subset of d.o.f. of NVE system behaves like a canonical subsystem: we do not need a large subsystem in contact with a larger bath, just one atom immersed in 100 others can display canonical statistics. Similar for LJ gas and liquid. The rest of the system acts, in practice, as a thermal bath: this is why canonical averages are used very widely.

Note: sometimes ensembles with their prob. densities are introduced referring to our lack of information about the precise microstate occupied, however if we have a trajectory passing through a multitude of microstates in ergodic fashion, we still use proficiently the ensemble framework, even though we know microstate at every time, just because we care about macroscopic average over long time. In other words, $p_i < 1$ does not necessarily mean ignorance, it can mean that trajectory (perfectly known, for instance) spends a p_i fraction of the total time in state i So motivation for ensembles and densities is not only lack of information, it is, more generally, the passage to the study of average or coarse-grained properties.

An interesting talk on the link between clocks and the second principle is given G. Milburn on youtube (`https://youtu.be/4tVbTF8DTwQ?si=WiaZEPTqiuMNr0N8`) and in the article "The thermodynamics of clocks" for the bibliographic project.

10   CM10

10.1   The Euler equation

To find useful relations about free energies and phase diagrams, we start from a simple derivation based on the extensivity of $E(S, V, N)$ and $S(E, V, N)$: if we call X_j the variables of E beyond S

$$E(\lambda S, \lambda X_1, \lambda X_2, ...) = \lambda E(S, X_1, X_2, ...)$$

now we will demonstrate Euler theorem on 1st order differential forms: we differentiate both members with respect to λ

$$\frac{\partial E(\lambda S, ...)}{\partial(\lambda S)} \frac{\partial(\lambda S)}{\partial \lambda} + \sum \frac{\partial E(\lambda S, ...)}{\partial(\lambda X_j)} \frac{\partial(\lambda X_j)}{\partial \lambda} =$$

$$\frac{\partial E(\lambda S, ...)}{\partial(\lambda S)} S + \sum \frac{\partial E(\lambda S, ...)}{\partial(\lambda X_j)} X_j = E(S, X_1, X_2, ...)$$

now taking $\lambda = 1$ we find an interesting and simple relation (Euler equation):

$$\frac{\partial E}{\partial S} S + \sum \frac{\partial E}{\partial X_j} X_j = E$$

therefore for a simple system $E = TS - pV + \mu_1 N_1 + ...$, and in the same way one finds for the entropy

$$S = \frac{1}{T} E + \frac{p}{T} V - \frac{\mu_1}{T} N_1 + ...$$

Interesting: starting from $E(S, V, N) = TS - pV + \mu N$ and Legendre transforming, we find Helmoltz free energy $F(T, V, N) = E - TS = -pV + \mu N$, enthalpy $H(S, p, N) = E + pV = TS + \mu N$, and Gibbs free energy (or free enthalpy) $G = E - TS + PV = \mu N$. These simple relations are attractive (in particular, it shows that molar $g = G/N$ is just equal to the chemical potential), but we must not forget the 3 (or more) natural variables of each function.

10.2   Gibbs-Duhem relation and number of degrees of freedom

We apply Euler equation to obtain a useful relation about phase coexistence:

$$E = TS - pV + \mu N \quad \Rightarrow \quad dE = TdS + SdT - pdV - Vdp + \mu dN + Nd\mu$$

and if we subtract the fundamental relation $dE = TdS - pdV + \mu dN$ we find Gibbs-Duhem relation

$$SdT - Vdp + Nd\mu = 0 \quad \Rightarrow \quad d\mu = -sdT + vdp$$

where we introduce the intensive s (entropy per atom, or per molecule) and v (volume per atom, or per molecule). Note also that the same applies to the molar free-enthalpy $g = G(N, p, T)/N$ since it is equal to μ thanks to Euler equation. We will use Gibbs-Duhem relation when analyzing the slope of the coexistence line between two phases.

This indicates that we cannot vary independently the chemical potential, temperature and pressure, but one is constrained by the other two, therefore the number of intensive parameters able to change independently, called the thermodynamic degrees of freedom, is equal to the number of thermodynamic variables minus 1. For instance, we have 2 (intensive) degrees of freedom for variables S, V, N, i.e., in a one-component system. We have 3 (intensive) degrees of freedom for variables S, V, N_1, N_2, i.e., in a two-component system, and so on.

10.3   Relation between free energies and work

$F = E - TS$, recall it is defined at equilibrium, and now we consider the variation at fixed T:

$$(\Delta F)_T = \Delta E - T\Delta S = Q + W - T\Delta S$$

where Q is heat received by the system (when positive it increases E of the system) and W work done on the system (same thing), while ΔS is the entropy change in the system, equal to universe minus thermostat:

$$\Delta S_{\text{univ}} = \Delta S + \Delta S_{\text{therm}} \geq 0$$

Since the thermostat is huge, any finite Q transferred to the much smaller system is negligible, and can be considered as reversible transfer from the thermostat viewpoint (not from the system viewpoint, in general), so $Q = -T\Delta S_{\text{therm}}$ and we find

$$(\Delta F)_T = W - T\Delta S_{\text{univ}} \leq W$$

We conclude that for a reversible process $(\Delta F)_T = W_{\text{rev}}$, while for an irreversible one the free energy change is the same (since F is a state function) but the work is different because its difference with the entropy change must remain the same: $W_{\text{irr}} = (\Delta F)_T + T\Delta S_{\text{univ}}^{\text{irr}} > (\Delta F)_T$. Now there are two cases: if some work is done on the system, so $W > 0$, it will be larger or equal than $(\Delta F)_T > 0$. If the system does some work on the surroundings, instead, $W < 0$, then $(\Delta F)_T < 0$ will be more negative than W, so in this case $|W| \leq |(\Delta F)_T|$, hence the interpretation "the free energy change at constant T is, in absolute value, the maximum work that the system can do on the surroundings, i.e., the maximum work that can be extracted from the system".

Similarly for G, one finds that at constant p, T (i.e., in contact with barostat and thermostat) $(\Delta G)_{pT} = Q + W - T\Delta S + p\Delta V = W' - T\Delta S_{\text{univ}}$ where $W' = W + p\Delta V$ is the non-mechanical work. Again minus the (negative) free enthalpy change is the maximum non-mechanical work that the system can do on the surroundings.

Finally, let's consider the enthalpy change at constant p: $(\Delta H)_p = \Delta E + p\Delta V = Q$, so the heat absorbed by the system during an isobaric transformation is the enthalpy variation.

10.4   Minimization of thermodynamic potentials

(Taken from Brunet lectures section 4.1, simplified) The central principle states that at equilibrium the entropy of an isolated system, here the universe (subscript "univ"), is maximized. Consider the universe as large, with a big part that we call thermostat (subscript "therm"), at fixed T (we neglect fluctuations) and a smaller part called "the system" or subsystem (no subscript):

$$S_{\text{univ}} = S + S_{\text{therm}}$$

We assume that the thermostat is at equilibrium and cannot be significantly perturbed by processes in the small system, while the system maybe out of equilibrium: as in deriving the canonical probabilities, we assume the system can exchange with the thermostat only heat, with no change in N or V. Approach to equilibrium consist in maximizing S_{univ}, i.e., in minimizing the potential

$$\Phi = -TS_{\text{univ}}$$

This formulation is general, but we specialize the way we write Φ for different experimental situations.

now, if N, V, T are fixed in the small system, a change in the latter starting out of equilibrium results in

$$\Delta \Phi = -T\Delta S_{\text{univ}} = -T(\Delta S + \Delta S_{\text{therm}})$$

and since V does not change $\Delta E = Q + W = Q = -T\Delta S_{\text{therm}}$, so in this case $\Delta \Phi = \Delta E - T\Delta S$ it means that, at equilibrium, the free energy of the system $F = E - TS$ takes the minimum value (at fixed N, V, T, with respect to varying internal degrees of freedom...). Since F is only defined at

equilibrium, we work with a decreasing Φ before reaching equilibrium, and once it is reached we call such potential $\Phi_{\min} \equiv F(N, V, T)$.

A slightly different situation: we consider as fixed T and p of the thermostat and barostat, now the small system can exchange heat but also change volume. We keep as basic principle the minimization of $\Phi = -TS_{\text{univ}}$, as before, but now from the viewpoint of the small system $\Delta E = Q + W = -T\Delta S_{\text{therm}} - p\Delta V$, thus $\Delta \Phi = -T\Delta S_{\text{univ}} = -T\Delta S + \Delta E + p\Delta V$ is negative, Φ decreases until reaching a minimum at equilibrium, and such minimum corresponds to the value $G = E - TS + PV$. Once again, we call $\Phi_{\min} \equiv G(N, p, T)$.
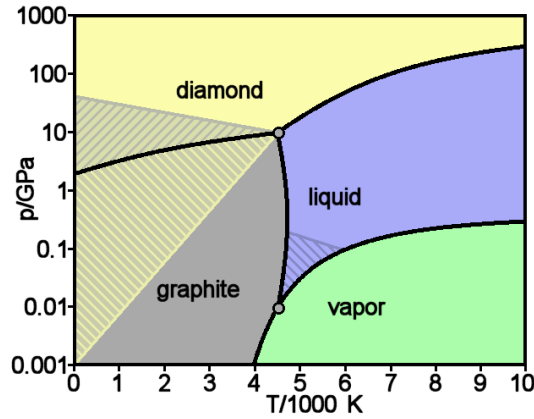
10.5 Phase diagram

Some of this material can be found in Reif 8.5, Brunet ch 7.

Consider a typical experiment on pure substance (e.g., water) at controlled p, T, so the relevant potential is $G(N, p, T)$, with

$$G = E - TS + pV \ , \quad dG = -SdT + Vdp + \mu dN$$

But now we admit (empirically) the existence of different phases: solid, liquid, gas; therefore $dG = -SdT + Vdp + \sum \mu_i dN_i$.

We can plot the phases on p, T diagram: phase diagram. At each point, the equilibrium state of the system is characterized by G being the minimum value of the potential Φ. So if we consider two phases A, B and we do not allow p, T to vary (we sit on a point in the diagram) but we allow N_A and N_B to vary, with $dN_B = -dN_B$, we have

$$dG = \mu_A dN_A + \mu_B dN_B = (\mu_A - \mu_B)dN_A = 0$$

So for $dG = 0$ to be true there are 3 cases: if $\mu_A = \mu_B$ the condition is true for any dN_A, so we have coexistence, A and B are in equilibrium, we can have any fraction of the two. If $\mu_A > \mu_B$ we can reduce G by $dN_A < 0$, so A tends to disappear and only B (the lowest-chemical potential phase) exists at equilibrium. The opposite is true if $\mu_A < \mu_B$. So we have naturally the regions of stability od A and B and the regions of coexistence, which are lines: $\mu_A \equiv \mu_A(p, T)$, $\mu_B \equiv \mu_B(p, T)$ (N is fixed) and the equation $\mu_A(p, T) = \mu_B(p, T)$ leaves one parameter free, so if we fix p we obtain T or vice-versa. Saying that A and B have same G (they are equiprobable) does not mean that interconversion is free: since phases are structurally very different in first order transition, not connected by continuous geometric transformation and geometrically separated by interface, we have barrier ΔG^* and rare jumps. We wait for random fluctuation to bring us on the other side (see discussion in CM11). Actually we can also consider $\mu_A(p, T) = \mu_B(p, T) = \mu_C(p, T)$, which leaves no free parameter, it is the triple point.

Note that, in principle, the phase diagram only gives information about the most stable phase at each condition p, T: this information is of limited interest, since in reality, around us, often we find a

phase that survives for a long time (metastability, due to kinetic barriers) even though it is not the most stable one. An example is diamond, observed at ambient conditions; sometimes the mestastability is also depicted on a phase diagram, as in the figure here (see shaded regions) schematically representing the phase diagram of carbon.

10.6 Clapeyron equation

When we consider phases like solid, liquid and gas, they correspond to very different arrangements of the atoms, with different molar volumes v and entropies s (related to specific heats), as it is intuitive. But looking at dG, such quantities are first derivatives of G with respect to p and T, so they change discontinuously in such transitions, that are called 1st order transitions.

As explained in Callen 9.3: consider two nearby points on a coexistence line, (p, T) and $(p', T') \equiv (p + dp, T + dT)$:

$$\mu_A(p, T) = \mu_B(p, T) \ , \ \mu_A(p', T') = \mu_B(p', T')$$

$$\mu_A(p', T') - \mu_A(p, T) = \mu_B(p', T') - \mu_B(p, T)$$

$$= -s_A dT + v_A dp = -s_B dT + v_B dp$$

$$(s_B - s_A)dT = (v_B - v_A)dp \ , \ \ \frac{dp}{dT} = \frac{\Delta s}{\Delta v} = \frac{l}{T\Delta v}$$

where $l = T\Delta s$ is the latent heat of the transition. This is the Clausius-Clapeyron equation. Therefore, the slope of the $p(T)$ curve has implications on the signs of the quantities. Usually, when you pass from solid to liquid and from liquid to gas, both entropy and volume increase, so positive slope. But in water close to $T = 0°C$, for instance, melting increases entropy but decreases volume, so slope is negative.

Le Chatelier principle is nicely discussed by Callen.

11 CM11

A short account of the theory of first order phase transitions, illustrated using the interesting phenomenon of nucleation. Based on Callen ch. 9, and on articles for the bibliographic project.

11.1 Summary of rare events, microstates vs macrostates, order parameters

Recall from the previous lectures that when we look at a complex system (i.e., with a lot of degrees of freedom) considering only a few generalized coordinates, e.g., one order parameter s, $\rho(s)$ represents the marginal probability density after integrating $e^{-\beta H}/Z$ over all the other degrees of freedom. So, fixing s corresponds to fix a macrostate formed by many microstates, and if we restrict the partition function to a fixed s (or a small interval) we get
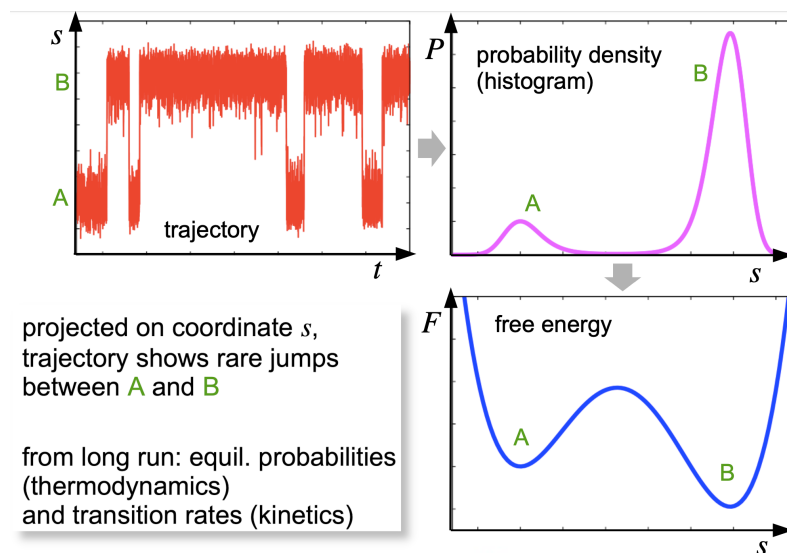
$$\frac{Z_s}{Z_{s'}} = \frac{\rho(s)}{\rho(s')} = e^{-\beta[F(s)-F(s')]}$$

so that the free-energy landscape as a function of s is usually written

$$F(s) = -kT \log \rho(s)$$

(even though, strictly speaking, the formula above has wrong units because of ρ: usually one considers that a reference is used to fix the zero value of the free-energy, in practice it is like dividing $\rho(s)$ by an arbitrary ρ_0 in the log, but this is immaterial since the zero is arbitrary, only differences matter).

If we observe a simulation or experiment of a phase transition, chemical reaction, protein folding etc. projected on an order parameter s we obtain something like the graph below. Remember: the experiment/simulation should be long enough so that we reversibly sample the different states until their average population is stationary (with respect to a longer experiment, see the discussion of error bars in CM4) and we can claim that we have the equilibrium picture (see CM9), only at this point we can identify the log of the histogram with the free-energy landscape! Unfortunately, often we cannot reach this situation of excellent statistical quality.

projected on coordinate s, trajectory shows rare jumps between A and B

from long run: equil. probabilities (thermodynamics) and transition rates (kinetics)

Coming back to the case of the phase diagram (replacing Helmoltz with Gibbs free energy), on the coexistence line there is an arbitrary amount of phases A or B due to equal probability, because the free energies of A and B are equal. However the barrier between them slows down the typical time that you have to wait to observe the transition between A and B: only in an extremely long experiment (typically unfeasible) you really observe equal average amount of A and B. As usual, kinetics prevents equilibrium from being easily observed, at least in many cases.

11.2 Latent heat

During a 1st order phase transition at constant p the system receives (or expels) energy, called latent heat (or enthalpy) of fusion/vaporization etc. Recall that in isobaric transformations $(\Delta H)_p = \Delta E + p\Delta V = Q$. Melting 1 kg of ice at 0 °C and ambient p requires to inject more than 300 kJ of energy, crystallizing the liquid releases the same amount.

It is important to obtain also some microscopic insight. During the melting transition of a crystal, we pass from a state of low potential energy among atoms (favorable arrangement of atoms) and low entropy (ordered arrangement) to a state of high potential energy and high entropy. The system loses potential energy, kinetic energy does not change if T is constant, so the thermostat must provide heat to the system: it provides kinetic energy, converted into potential during the transition. The system absorbs heat, $(\Delta H)_p \equiv Q > 0$, it is called the latent heat of fusion. The latent heat of solidification is equal and opposite.

Recall, from CM10, that the latent heat (together with the molar volume change) is simply connected to the slope of the coexistence line in the phase diagram. Compare also with Callen fig. 9.5 (here below): $dG/dT = -S$, increasing T we pass from solid to liquid to gas and the slope $G(T)$ is always negative (entropy grows with T, and it jumps discontinuously during phase transitions, jumps being the absorbed latent heats. We can make the transition in different ways, always supplying/expelling the same latent heat: e.g., by changing T at constant p, or by changing p at constant T.

Note that, during the transition at coexistence conditions (e.g., on the melting line in the phase diagram), potential energy among atoms, entropy and volume can change between the two phases, but in such a way that $\Delta\mu \equiv \Delta g = 0$. If, instead, the transition occurs out of the coexistence line (e.g., if liquid water crystallizes below the melting T, from supercooled liquid, a common situation in the clouds), then the system passes from a (metastable) phase at higher g to a phase at lower g, thus $\Delta g < 0$. This is characteristic of a spontaneous irreversible transition, where the entropy of the universe is maximized, i.e., $\Delta\Phi < 0$ (see CM10).

11.3 Metastable phases and the role of time (kinetics)

As mentioned before, at coexistence conditions or even out of them, observing a phase transition requires some time, possibly a very long one. This is due to the presence of a "free-energy barrier" separating the two phases, since 1st order transitions are discontinuous in terms of atomic geometry (think to the big difference between a crystal and a liquid), and growing a liquid region into a crystal (or the inverse), requires growing an interface where atoms are neither crystalline neither liquid. Such frontier atoms experience unfavorable, high-energy geometries: bonds are broken and not yet rearranged, angles are distorted, available volumes are not optimal... The result is an unfavorable contribution to the free energy of the system, growing with the size of the interface between the phases, until we convert completely the system.

Note that, at coexistence conditions, there is no bulk free-energy gain in passing from one phase to another one ($\mu_A = \mu_B$): if the geometry of the system is such that we can increase A or B without increasing the size of the interface (think to cylindrical vessel with an interface parallel to the ends), we can change freely the composition, there is no barrier. But if the growing region has, e.g., a spherical shape inside the system, then the growth is unfavorable.

If now we move away from the coexistence line in the phase diagram, the barrier decreases gradually until disappearing at long-enough distance in p, T-space (see Callen figures in ch.9). Before disappearing, we can have the very interesting phenomenon of nucleation.

Note: in a finite system it is possible also to have a transition from a lower-g phase to a higher-g one, however it will be rare and it will revert to the most stable phase (more rapidly, since the barrier will be lower). In general, we can have "random" fluctuations allowing the system to jump between any pair of adjacent states, but the rate decreases exponentially with the barrier height, and the residence time decreases exponentially with the increase of free-energy in the new phase.
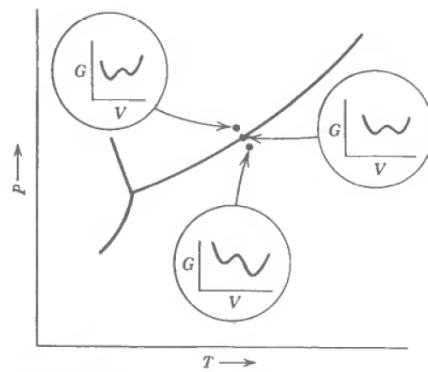
FIGURE 10.1
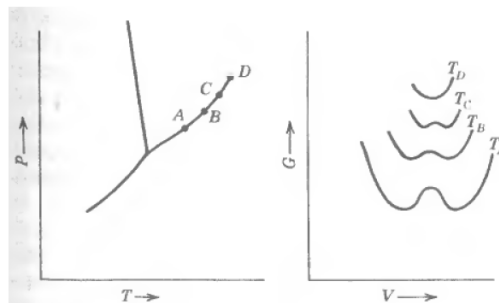Competition of two minima of the Gibbs potential near the coexistence curve.

FIGURE 9.6
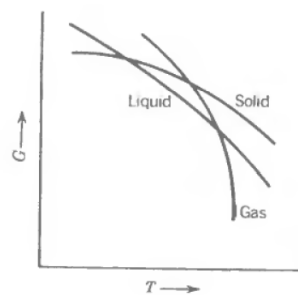The two minima of G corresponding to four points on the coexistence curve. The minima coalesce at the critical point D.

FIGURE 9.5
Minima of the Gibbs potential as a function of T.

11.4   Nucleation and growth of a crystal from a liquid

Good references (also for bibliographic project): G.C. Sosso et al, Chem. Rev. 116, 7078 (2016) (general theory and simulations, see figure below), J.J. De Yoreo et al, Reviews in Mineralogy and Geochemistry 54, 57 (2003) (biomineralization, general theory, experiments).

A short summary of classical nucleation theory: we start from supercooled liquid (or supersaturated solution), i.e., below melting point; it is a metastable state since the crystal would be more stable, $\mu_L - \mu_C = \Delta\mu > 0$. We assume small nuclei of the crystal can spontaneously form due to "random" fluctuations in the local structure of the liquid. We assume such nuclei to have a perfectly crystalline structure, like the bulk solid, with an interface of vanishing thickness between liquid and solid (capillarity approximation). Clearly such hypothesis is not so realistic if the nucleus includes a small number of atoms. The interface is an unfavorable region, from the viewpoints of atomic geometry and energetics, being between two well-defined phases, so it contributes a positive term in the free energy, $\gamma > 0$.
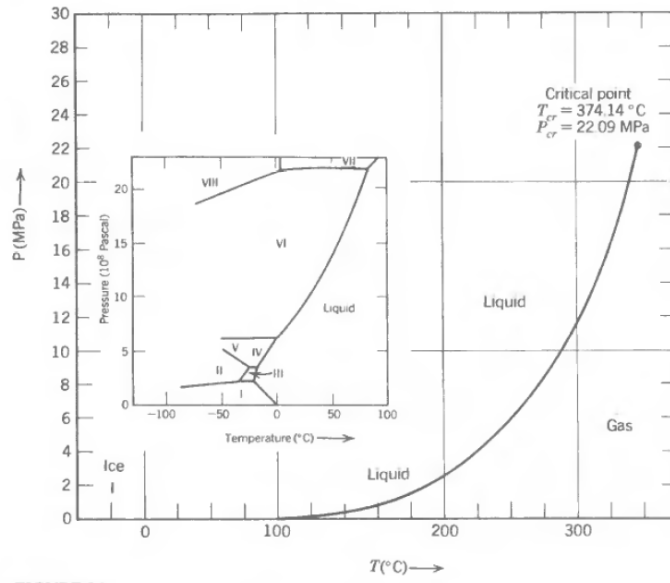
FIGURE 9.1
Phase diagram of water. The region of gas-phase stability is represented by an indiscernibly narrow horizontal strip above the positive temperature axis in the phase diagram (small figure). The background graph is a magnification of the vertical scale to show the gas phase and the gas–liquid coexistence curve.
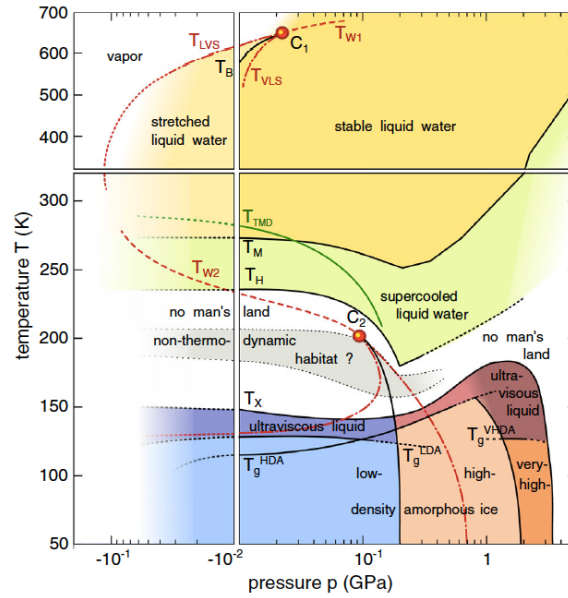
We write the free energy variation due to the formation of a spherical nucleus of radius r into the liquid as

$$\Delta G(r) = 4\pi r^2 \gamma - \frac{4\pi}{3} r^3 \Delta\mu$$

where we consider the interfacial free energy per unit surface and the chemical potential difference per unit volume. The function $\Delta G(r)$ is a free energy landscape: it describes the free-energy for a fixed value of the order parameter r, that monitors the evolution of the 1st order phase transition (this is a general framework, the nature of the order parameter will vary with the transition considered).

The landscape has a maximum for a nucleus formed by the following critical number of atoms (easy

43

to compute):

$$n^* = \frac{32\pi}{3} \rho_C \frac{\gamma^3}{\Delta\mu^3}$$

where ρ_C is the number density in the crystal. The corresponding value of the free-energy (the barrier height) is

$$\Delta G^* = \frac{16\pi}{3} \frac{\gamma^3}{\Delta\mu^2}$$

At n^* there is an unstable equilibrium: growing or shrinking of the nucleus decreases the free energy, there is 50% probability to evolve in either direction. If at this point the crystal grows, the transition will proceed until transforming the whole liquid, since $\mu_C < \mu_L$ at the p, T under consideration.
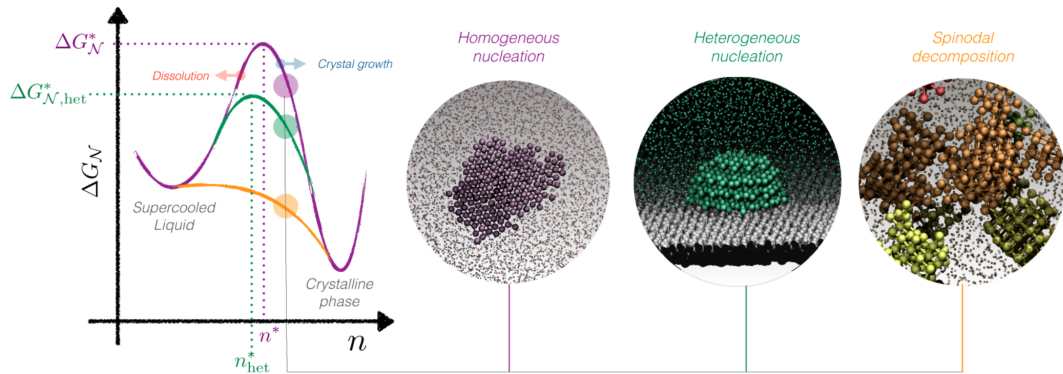
As always when free-energy barriers are present, successful nucleation events will occur with a rate

$$\nu \equiv \frac{1}{\tau} = Ae^{-\Delta G^*/kT}$$

where τ is the average time we need to wait before a transition occurs (also called the mean first passage time). The rationale of this expression (studied by Arrhenius, Eyring, Kramers etc.) is that we wait for a random jump, with a number of attempts per unit time A that multiplies the ratio between the probability to be at the top of the barrier and the probability to be at the bottom of the starting well $\exp(-\Delta G^*/kT)$. In the case of nucleation, we note that the rate will grow linearly with the amount of liquid, since the event can occur anywhere inside it with equal probability.

The quantities γ, $\Delta\mu$, n^*, ΔG^*, ν depend on the p, T conditions. For instance, the lower the temperature below the melting point (the degree of supercooling) the lower the barrier (until it disappears) and the smaller the critical size, but also the slower the characteristic dynamics of the liquid atoms, so the global effect of cooling on the nucleation can be non-monotonous. If we cool the system strongly enough and fast enough, we could also trap it into an amorphous or glassy metastable state! In this case, we could need to heat up the glass to allow crystal nucleation.

We notice how $\Delta G^*/kT \gg 1$ implies a very rare transition: it was estimated that for pure liquid water at 0°C in absence of impurities or interfaces, successful nucleation inside an amount of liquid of the size of all the oceans united would require longer than the age of the universe! It is remarkable, given that we are speaking of a single localized nucleation event in a huge volume and in a huge time. In practice, ice nucleation on the earth and in the atmosphere occurs below 0°C and with the help of dust particles, bacteria (!), etc., while solid water is usually amorphous in the interstellar space.

Classical nucleation theory is limited by its own simplistic assumptions: the nucleus is actually not spherical, when small the crystal is not perfect (it could even be amorphous), different crystal polymorphs could appear and mix and interconvert during nucleation and growth, the interfacial free energy is a complex quantity depending on the type of interface etc. Nevertheless, it is a basis to start studying nucleation of materials, precipitation from solution of salts or methane clathrates or drugs or nanostructures, formation of amyloid (protein) fibrils in neurodegenerative diseases, and so on.

44

We speak of kinetic aspects when discussing rates and barriers, we speak of thermodynamic aspects when discussing the phase diagram, the most stable state at any condition etc. In nature, we are surrounded by a competition between kinetics (time) and thermodynamic (stability). Often, systems appear as trapped into metastable states (not the most stable ones) due to their initial conditions and detailed history. Examples are diamond (that should evolve to graphite), as well as many materials and molecules of practical use, but also ourselves. On ordinary timescales a large number of systems are trapped out of equilibrium, and ingenious synthesis techniques (as well as kitchen recipes at home) consist in driving a system along a complex path to achieve a useful metastable state, that remains trapped by barriers for a time long-enough for its practical use (including eating it!).

11.5   How can we model a phase transition?

We can use computer simulations at different levels of detail. Ccompared to experiments, simulations can be useful to test our theories, to understand the nucleation mechanism, or to extract useful quantities like barriers and rates.

- Molecular dynamics (MD): very realistic, all the details of the atomic trajectories are simulated in time by numerically solving Hamilton's equations (or Schroedinger equation, in small systems); very expensive to simulate, but the most insightful and close to experimental reality.

- Monte Carlo: less detailed, instead of real trajectories there are "unphysical jumps" between microstates, still there is a realistic Hamiltonian for atoms; we can explore realistic geometries with realistic energies and probabilities (many applications, e.g., in Ising models of magnetic materials).

- Random-walk model (Langevin, Fokker-Planck, master equations...): coarse-grained description, we do not see anymore atoms but only the random fluctuations of an order parameter (indicating the progress of the transition from liquid to crystal) on a free-energy landscape; still, the statistical properties of the order parameter, both static and time-dependent, can be correct, similarly to the diffusion problem.

Some examples of MD simulations can be showed with the projector, including melting of a crystal followed by cooling, where due to insufficient time the system does not goes back to crystal but to an amorphous state (glass). Related examples are chemical reactions or protein folding: same framework based on free-energy landscape, order parameter framework, stochastic effective dynamics (random walk on a landscape). This is a very general approach, widespread in modern research.

12 CM12

In this last lecture we give a look to the modifications introduced by the quantum nature of particles. We cannot go very deep, but it is possible to understand the main features of quantum statistical physics, and the connection with the classical theory in terms of a particular limit (low density, high T).

12.1 Introduction to quantum statistical physics

We follow Callen [1] ch. 18 and Reif [3] ch. 9. A wavefunction for N identical non-interacting particles (ideal gas)

$$\Psi_{\mathbf{s}_1,...,\mathbf{s}_N}(\mathbf{q}_1,...,\mathbf{q}_N)$$

is specified using N complete sets of one-particle quantum numbers $\mathbf{s}_j$ (for example, n, l, m, m_s for an electron in the hydrogen atom, or $\mathbf{k}, m_s$ for a free electron in a box) and it is a function of N sets of "coordinates" $\mathbf{q}_j$ (for example, x, y, z, σ when projected on the representation of Cartesian coordinates and z projection of the spin).

Particles are either fermions (semi-integer spin) or bosons (integer spin). Fermions have anti-symmetric wavefunctions under exchange of particle labels: $\Psi(..., \mathbf{q}_i, ..., \mathbf{q}_j, ...) = -\Psi(..., \mathbf{q}_j, ..., \mathbf{q}_i, ...)$, while bosons have symmetric ones (no change of sign). More details are studied in relativistic quantum mechanics.

Fermions are somehow closer to classical impenetrable particles: they cannot occupy the same place when having the same quantum numbers (however electrons form pairs with spin pointing in opposite directions in atomic and molecular orbitals). There is a sort of effective repulsion, due to symmetry, Pauli exclusion principle. This can be seen since Ψ must change sign under exchange, but when the exchanged particles have the same quantum numbers $\mathbf{s}_i = \mathbf{s_j}$ and the same coordinates $\mathbf{q}_i = \mathbf{q}_j$ then the form mathematically is the same: it must change and not change sign at the same time, which is possible only if it vanishes. It can be seen with two particles occupying one-electron states ϕ_α and ϕ_β (we obtain simple forms of Ψ because the particles are non-interacting):

$$\Psi(\mathbf{q}_1, \mathbf{q}_2) = \frac{1}{\sqrt{2}}[\phi_\alpha(\mathbf{q}_1)\phi_\beta(\mathbf{q}_2) - \phi_\beta(\mathbf{q}_1)\phi_\alpha(\mathbf{q}_2)]$$

it can be checked that exchanging $\mathbf{q}_1$ and $\mathbf{q}_2$ Ψ changes sign. But if $\alpha = \beta$, there is no way to do the trick, Ψ is identically zero, it means that we cannot put the two fermions in one-particles states with identical sets of quantum numbers. If they have the same m_s, in particular, they cannot be in the same spatial orbital, they must occupy different regions in space.

Bosons are somehow closer to classical waves: they can accumulate in large numbers in the same quantum state (like a collection of harmonic oscillators), and when they are mass-less they can be freely created and annihilated in pairs (photons). Similarly to the case above, we can form a symmetric wavefunction for two particles as

$$\Psi(\mathbf{q}_1, \mathbf{q}_2) = \frac{1}{\sqrt{2}}[\phi_\alpha(\mathbf{q}_1)\phi_\beta(\mathbf{q}_2) + \phi_\beta(\mathbf{q}_1)\phi_\alpha(\mathbf{q}_2)]$$

Note that, in this case, if $\alpha = \beta$ we must use a simpler expression: $\Psi(\mathbf{q}_1, \mathbf{q}_2) = \phi_\alpha(\mathbf{q}_1)\phi_\alpha(\mathbf{q}_2)$

12.2 Comparison of classical and quantum statistics

The different behavior in terms of occupation of one-particle states gives very different statistics and probabilities, so implications for statistical mechanics are very important. The intrinsic indistinguishabe character both of fermions and bosons (exchanging labels does not give a different microstate, the sign change in Ψ does not affect observables) gives a very different statistics compared to classical distinguishable particles.

The best is considering a simple example: two particles (say A and B) in 3 possible states. As can be seen in the tables reported from Reif's book, Maxwell-Boltzmann gives 9 possible states of the whole

Maxwell-Boltzmann statistics: The particles are to be considered distinguishable. Any number of particles can be in any one state.

1	2	3
AB	. . .	. . .
. . .	AB	. . .
. . .	. . .	AB
A	B	. . .
B	A	. . .
A	. . .	B
B	. . .	A
. . .	A	B
. . .	B	A

system (note that this is a wrong statistics, no particles obey MB, but we can have a classical limit under suitable situations, see later), Bose-Einstein gives 6 states, Fermi-Dirac gives 3 states. Furthermore, BE favors particles being in the same state compared to MB, the opposite for FD. More calculations of this kind will be done in the TD.

Bose-Einstein statistics: The particles are to be considered *indistinguishable.* Any number of particles can be in any one state. The indistinguishability implies that $B = A$, so that the three states in the MB case which differed only in interchange of A and B are now no longer to be counted as distinct. The enumeration is then as follows:

1	2	3
AA	. . .	. . .
. . .	AA	. . .
. . .	. . .	AA
A	A	. . .
A	. . .	A
. . .	A	A

There are now three distinct ways of placing the particles in the same state. There are three distinct ways of placing the particles in different states. Hence there exist a total of $3 + 3 = 6$ possible states for the whole gas.

Fermi-Dirac statistics: The particles are to be considered as *indistinguishable.* No more than one particle can be in any one state. The three states in the BE case where two particles were in the same state must be eliminated in this case. One is thus left with the following enumeration:

1	2	3
A	A	. . .
A	. . .	A
. . .	A	A

There exist now only a total of 3 possible states for the whole gas.

This example shows one further qualitative feature of interest. Let

$$\xi = \frac{\text{probability that the two particles are found in the same state}}{\text{probability that tha two particles are found in different states}}$$

Then we have for the three cases

$$\xi_{MB} = \tfrac{3}{6} = \tfrac{1}{2}$$
$$\xi_{BE} = \tfrac{3}{3} = 1$$
$$\xi_{FD} = \tfrac{0}{3} = 0$$

Thus in the BE case there is a greater relative tendency for particles to bunch together in the same state than in classical statistics. On the other hand, in the FD case there is a greater relative tendency for particles to remain apart in different states than there is in classical statistics.

12.3 Occupation numbers: Bose-Einstein and Fermi-Dirac laws

In quantum statistics it is natural to work in the occupation number formalism, let's see how it works (see also Texier ch. 11). Considering the simple case of independent (uncorrelated) particles, i.e., having negligible interactions, where the joint probability is the product of the marginal ones, we can identify single-particle states. Examples are the eigenstates of the hydrogen atom, of a particle in a box, of a free particle (a plane wave), etc. Instead of specifying the eigenstate of each particle, since anyway we cannot distinguish one particle from another it is more natural to specify how many particles occupy each single-particle eigenstate. The only thing that matters are the occupation numbers.

As in the classical case, one starts by writing the partition function Z, for example in the canonical NVT case: care must be taken in summing only over the allowed N-particle microstates, noting that what is allowed is different in MB/FD/BE cases. For the simple case of two one-particle energy eigenstates ϵ_a, ϵ_b, the partition function for one particle is $z = e^{-\beta\epsilon_a} + e^{-\beta\epsilon_b}$; for 2 particles we find

$$Z_{MB} = e^{-\beta 2\epsilon_a} + 2e^{-\beta(\epsilon_a+\epsilon_b)} + e^{-\beta 2\epsilon_b} = z^2$$

$$Z_{BE} = e^{-\beta 2\epsilon_a} + e^{-\beta(\epsilon_a+\epsilon_b)} + e^{-\beta 2\epsilon_b} \neq z^2/2!$$

$$Z_{FD} = e^{-\beta(\epsilon_a+\epsilon_b)} \neq z^2/2!$$

so factorization disappears in the quantum case: the problem, we will see here below, is that we fix the total particle number N. The formula $Z_{MB}/N!$ is not a good approximation to the quantum case.

We pass to occupation numbers $n_{\mathbf{s}}$, i.e., the number of particles that occupy the one-particle state with quantum numbers $\mathbf{s}$:

$$N = \sum_{\mathbf{s}} n_{\mathbf{s}} \ , \quad E = \sum_{\mathbf{s}} n_{\mathbf{s}} \epsilon_{\mathbf{s}}$$

It turns out that using occupation numbers and working in the grand-canonical ensemble instead of the canonical one, the partition function is factorized into single-particle parts, simplifying a lot the calculations.

The grand-canonical ensemble, a very useful one, corresponds to contact with a thermostat and a particle reservoir (think to a permeable membrane), and is obtained replacing $N \to \mu$ with a Legendre transform of the canonical ensemble

$$J = F - \mu N = E - TS - \mu N \ , \quad dJ = -SdT - pdV - Nd\mu$$

where $J(\mu, V, T)$ is the grand potential. By the same technique employed for the canonical ensemble one finds

$$p_\alpha = \frac{1}{\Xi} e^{-\beta(E_\alpha - \mu N_\alpha)} \ , \quad \Xi = \sum_\alpha e^{-\beta(E_\alpha - \mu N_\alpha)} \ , \quad S = -k_B \sum_\alpha p_\alpha \log p_\alpha$$

Here we cannot go into details of derivations, anyway the method is always the same: maximize the entropy to find the equilibrium distribution.

In the case of independent particles, the grand partition function is always factorized, but in terms of single-particle states (not in terms of particles!):

$$\Xi = \sum_\alpha e^{-\beta(E_\alpha - \mu N_\alpha)} = \sum_{\alpha \equiv \{n_{\mathbf{s}}\}} e^{-\beta \sum_{\mathbf{s}} n_{\mathbf{s}}(\epsilon_{\mathbf{s}} - \mu)} = \prod_{\mathbf{s}} \sum_{n_{\mathbf{s}}} e^{-\beta n_{\mathbf{s}}(\epsilon_{\mathbf{s}} - \mu)} = \prod_{\mathbf{s}} \xi_{\mathbf{s}}$$

(note that each microstate α coincides with a given set $\{n_{\mathbf{s}}\}$). Therefore we use the grand-canonical ensemble with occupation numbers, anyway in the limit of large systems the averages computed with different ensembles coincide. In BE statistics, any occupation number from zero to infinite is allowed, whereas in FD statistics $n_{\mathbf{s}} = 0$ or 1 for each one-particle state.

Large differences show at low T: for bosons, all particles fall into the same ground state (if non-degenerate). For fermions, we cannot put more than 1 in each state so even at low T we are forced to fill up to a relatively high energy (e.g., electrons in a metal).

It is useful to compute the average occupation numbers: the probability to occupy microstate $\alpha \equiv \{n_{\mathbf{s}}\}$ is formally factorized into a product of probabilities to observe $n_{\mathbf{s}}$ particles in state $\mathbf{s}$

$$p_{\{n_{\mathbf{s}}\}} = \prod_{\mathbf{s}} \frac{1}{\xi_{\mathbf{s}}} e^{-\beta n_{\mathbf{s}}(\epsilon_{\mathbf{s}} - \mu)}$$

thus, according to this viewpoint,

$$\langle n_{\mathbf{s}} \rangle = \frac{1}{\xi_{\mathbf{s}}} \sum_{n_{\mathbf{s}}} n_{\mathbf{s}} e^{-\beta n_{\mathbf{s}}(\epsilon_{\mathbf{s}} - \mu)} \equiv \frac{1}{\beta} \frac{\partial}{\partial \mu} \log \xi_{\mathbf{s}}$$

Let's apply to fermions:

$$\xi_{\mathbf{s}}^{FD} = 1 + e^{-\beta(\epsilon_{\mathbf{s}} - \mu)}$$

$$\langle n_{\mathbf{s}}^{FD} \rangle = \frac{1}{e^{\beta(\epsilon_{\mathbf{s}} - \mu)} + 1}$$

where, interestingly, the chemical potential μ appears here by imposing a normalization condition $\sum \langle n_{\mathbf{s}}^{FD} \rangle = N$.

For bosons one finds:

$$\xi_{\mathbf{s}}^{BE} = \sum_{n_{\mathbf{s}}=0}^{\infty} e^{-\beta n_{\mathbf{s}}(\epsilon_{\mathbf{s}} - \mu)} = \frac{1}{1 - e^{-\beta(\epsilon_{\mathbf{s}} - \mu)}}$$

with the condition $e^{-\beta(\epsilon_{\mathbf{s}} - \mu)} < 0$, i.e., $\mu < \epsilon_{\mathbf{s}}$ for each state, so μ must be lower than the ground state.

$$\langle n_{\mathbf{s}}^{BE} \rangle = \frac{1}{e^{\beta(\epsilon_{\mathbf{s}} - \mu)} - 1}$$

(note that in the case of photons we do not constrain N so that we have $\mu = 0$) (Reif par. 9.3 shows how, in general, shifting all energies by a constant does not affect average occupation numbers). Finally, it is interesting to find the expected classical result:

$$\langle n_{\mathbf{s}}^{MB} \rangle = N \frac{e^{-\beta \epsilon_{\mathbf{s}}}}{\sum e^{-\beta \epsilon_{\mathbf{s}'}}}$$

MB statistics is a good approximation in the classical limit: it can be shown (Reif par. 9.8) that for sufficiently low density or sufficiently high T (the so-called non-degenerate quantum gas) the condition $\sum \langle n_{\mathbf{s}} \rangle = N$ results in a chemical potential that becomes so low that, for each $\mathbf{s}$, $e^{\beta(\epsilon_{\mathbf{s}} - \mu)} \gg 1$, $\langle n_{\mathbf{s}} \rangle \ll 1$. This means that μ needs to become much lower in energy than the ground state level, and under these conditions we recover the classical MB expression for $\langle n_{\mathbf{s}} \rangle$, with a bonus: we directly obtain also the $1/N!$ factor in front of the partition function, that avoids Gibbs paradox. Thus, starting from the quantum gas, the factor does not need to be artificially added a posteriori.

12.4 Ergodicity in the quantum world

In classical physics, we naturally observe ergodicity (albeit difficult to demonstrate formally) even in the case of isolated system with Hamiltonian evolution (no external perturbations), with perfectly-defined initial conditions, thanks to chaos and mixing. Dynamics is deterministic and reversible, still we can use probability theory and we observe e.a.p.p. if we define probability of a microstate (in phase space) as fraction of time spent into it, over a long-enough time.

In quantum physics the situation appears even more complex to understand in detail, at a fundamental level, even though in practice we also observe e.a.p.p. and we apply the same probability theory. Now if initial condition of isolated system is exactly defined, the wavefunction (we call it a pure state) evolves in a deterministic way according to

$$|\psi(t)\rangle = \sum_{\alpha=1}^{\Omega} c_{\alpha} e^{-iE_{\alpha}t} |\alpha\rangle$$

where $H|\alpha\rangle = E_\alpha|\alpha\rangle$; however, in this case the probability to observe at any time the system into state $|\alpha\rangle$ is not uniform, but fixed from the initial decomposition of the state vector:

$$p_\alpha = \langle\psi|\alpha\rangle\langle\alpha|\psi\rangle = |c_\alpha|^2 \neq \frac{1}{\Omega}$$

so the e.a.p.p. is violated for a pure state in an isolated quantum system, we lack the equivalent of classical chaos and mixing properties. The dynamics remains time-reversible like in classical physics (see Tolman [7] par. 95), and the e.a.p.p. is restored using two approaches: on one side, unavoidable external perturbations, on the other side, the fact that the initial state of the system cannot be measured with perfect precision in a finite time (see Tolman par. 98).

Due to Heisenberg's principle $\Delta E \Delta t \gtrsim h$ a measure taking a reasonable time, not too long, will result in an uncertainty ΔE large enough to contain many energy eigenstates (we expect the spectrum to be quite dense when N is large). To take this into account, we introduce (with an ad hoc assumption) the e.a.p.p. by postulating equal probabilities p_α for all microstates within ΔE: the wavefunctions corresponding to the different possible states are mixed with random phases in a "density matrix":

$$\rho = \sum p_\alpha|\alpha\rangle\langle\alpha| \ , \quad \sum p_\alpha = 1 \ , \quad \mathrm{Tr}(\rho) = 1 \ , \quad \langle A\rangle = \mathrm{Tr}(\rho A)$$

This is a probability mixture, different from a pure state displaying interference. The justification for introducing probabilities so abruptly can be based on time-dependent perturbation theory (considering that $\{|\alpha\rangle\}$ are not exactly eigenstate, they display transitions) and detailed balance (see Tolman), but it is quite unsatisfactory from my viewpoint.

Argument from https://farside.ph.utexas.edu/teaching/sm1/lectures/node26.html is that, if we assume master equation (=Markovianity) and we assume symmetry $K_{ij} = K_{ji}$ based on QM transition prob (time-dep. perturbation theory), then we find $dS/dt > 0$ and $p_i =$ the same for each state in equilibrium, i.e., equal a priori prob. However, as the author says, we assumed Markovianity so the proof is not really general. In my StatPhys slides I use the same equation to evaluate dS/dt (coarse-grained case), saying that since all p_i are equal, then $K_{ij} = K_{ji}$, a little bit inversing the argument for the general case. If it is truly equivalent to assume one or the other (meaning, there is a double implication), then the symmetry of the matrix from QM could be a better starting point.

To go deeper: "eigenstate thermalization hypothesis".

If, instead of being isolated, the system is in contact with a thermostat, building upon our understanding of the classical case we directly define

$$\rho = \frac{1}{Z}\sum e^{-\beta E_\alpha}|\alpha\rangle\langle\alpha| \ , \quad Z = \sum e^{-\beta E_\alpha}$$

Note that for a pure state the treatment greatly simplifies, there is no more a mixture of probabilistic nature, but a single quantum state: the density matrix becomes simply the projector on the state $\rho = |\psi\rangle\langle\psi|$.

With density matrices we can define entropy $S = -k\,\mathrm{Tr}(\rho\log\rho)$, we can write an equivalent of Liouville equation (von Neumann equation, or Moyal equation using Wigner function), we can study the measurement process, we can deal with entanglement, qubits etc. These are advanced topics not covered here, insightful treatments can be found in the books of Pathria and of Balian.

As usual, it can be seen in more advance studies that going beyond the independent-particles approximation introduces mathematical complications, but also very interesting phenomena.

References

[1] Herbert B Callen. *Thermodynamics and an Introduction to Thermostatistics, second edition*. John Wiley and Sons, 1985.

[2] Christophe Texier and Guillaume Roux. *Physique statistique: des processus élémentaires aux phénomènes collectifs*. Dunod, 2020.

[3] Frederick Reif. *Fundamentals of statistical and thermal physics*. Waveland Press, 2009.

[4] Eric Brunet, Thierry Hocquet, and Xavier Leyronas. Cours de thermodynamique (l3 3p011, also in english), 2019.

[5] Peter Atkins and Julio de Paula. *Physical chemistry: thermodynamics, structure, and change*. WH Freeman and Company New York, 2014.

[6] Sheldon Goldstein and Joel L Lebowitz. On the (boltzmann) entropy of non-equilibrium systems. *Physica D: Nonlinear Phenomena*, 193(1-4):53–66, 2004.

[7] Richard Chace Tolman. *The principles of statistical mechanics*. Courier Corporation, 1979.