

Notes for the course
Theoretical Statistical Physics

Luca Amendola University of Heidelberg

l.amendola@thphys.uni-heidelberg.de

<http://www.thphys.uni-heidelberg.de/~amendola/teaching.html>

v.1.8

January 29, 2024



**UNIVERSITÄT
HEIDELBERG**
ZUKUNFT
SEIT 1386

Contents

1	Thermodynamics	4
1.1	General concepts	4
1.2	Ideal gas	5
1.3	Exact and inexact differentials	6
1.4	Work and heat	7
1.5	First law	8
1.6	Second Law	9
1.7	Thermodynamic potentials	12
1.8	Energy and entropy	13
1.9	Stability conditions	14
2	Introduction to Statistics	17
2.1	Probability and frequency	17
2.2	Probability of combined events	18
2.3	Expected values	19
2.4	Transformation of variables	21
2.5	The main PDFs	21
2.6	Multivariate distributions	24
2.7	Random walks and Fokker-Planck equation	26
3	Statistical mechanics	29
3.1	Statistical definition of entropy	29
3.2	Number of microstates	31
3.3	Microcanonical ensemble	34
3.4	The canonical ensemble	36
3.5	Maximization of entropy	37
3.6	The partition function	38
3.7	Maxwell-Boltzman distribution	40
3.8	Proof that the statistical entropy equals the thermodynamic entropy	42
3.9	Virial and equipartition theorems	43
3.10	Applications of the canonical ensemble: quantum oscillators	45
3.11	Applications of the canonical ensemble: paramagnetism	48
3.12	The macrocanonical ensemble	52
3.13	The chemical potential	55
3.14	Ultra-relativistic ideal gas	55
4	Statistical mechanics of quantum systems	57
4.1	Density operators	57
4.2	Quantum ensembles	60
4.3	Symmetry of many-particle wave functions	62
4.4	Ideal quantum systems	64
4.5	Bose gas	69
4.6	Photon gas	74

4.7	Phonons	75
4.8	Fermi gas	78
4.9	Relativistic fermions	81
5	Phase transitions	84
5.1	General properties of phase transitions	84
5.2	Gas with interacting particles	85
5.3	Critical exponents	89
5.4	Ising model	91
5.5	The Renormalization Group	95
5.6	Mean field theory	98
5.7	Correlation functions	99
6	Mathematical appendix	103
6.1	Partial derivative identities	103
6.2	Liouville's theorem	104
6.3	Moment generating function	106
6.4	Central limit theorem	106
6.5	Estimators of mean and variance.	107
6.6	Parameter estimation: Statistics, sample, bias	108
6.7	Maximum likelihood estimators and Bayes' theorem	109
6.8	Information entropy	111

Acknowledgments

I thank Nils Strodthoff, Nicolò Defenu, and Moritz Drescher for help with the lectures and exercises. Thanks also to R. Erhard, S. Heinze, C. Sünderhauf for pointing out several typos.

These lecture notes are based, among others, on the treatment of the following books, which I recommend for further study:

Greiner, Neise, Stoecker, *Thermodynamics and Statistical Mechanics*, Springer

Pahtria, *Statistical Mechanics*, Elsevier.

The student will benefit a lot from a systematic study of both Greiner's and Pahtria's books, which also contain many more applications than we can cover in the course.

I also employed other texts, for instance:

Schwabl, *Statistical Mechanics*, Springer.

Chandler, *Introduction to Modern Statistical Mechanics*

D. Arovas, *Lecture Notes on Thermodynamics and Statistical Mechanics* (courses.physics.ucsd.edu/2010/Spring/physics)

U. Schwarz, *Lecture Notes in Theoretical Statistical Physics*.

All figures in this text are either created by the Author or used with permission or believed to be in the public domain. If there is any objection to the use of any material, please let me know. The text is released under CC license (<https://creativecommons.org/licenses/by-nc/3.0/>) that is, it is free for non-commercial use, provided appropriate credit is given. I cannot guarantee that also the figures are covered by this license.

Chapter 1

Thermodynamics

1.1 General concepts

1. A thermodynamical system is a collection of a huge number of particles: a gas, a solid etc. The thermodynamic limit is reached when $N, V \rightarrow \infty$ but $\rho = N/V$ remains constant.
2. Large number of particles and fast microchanges (i.e. collisions) require a statistical approach. Every measurement is an average over many states microscopically different (microstates).
3. A system can be characterized by a small number of variables like temperature, volume, pressure, number of particle, charge, dipole, viscosity etc
4. For most of what we say, it is sufficient to consider energy E , volume V , number of particles N . Related quantities are entropy S , temperature T , pressure p , chemical potential μ .
5. A system can be completely isolated (no change in E, V, N), thermally isolated (no transfer of heat), closed (change in E and V but not in N), open (change in all variables).
6. A variable is called a *state variable* if it does not depend on the previous history of a system. Mathematically, this implies that small differences in state variables are *exact differentials*.
7. When two isolated systems are brought into contact (and kept isolated from the external environment), they reach an equilibrium, i.e. their state variables (those that are permitted to be exchanged between the systems) go from a constant but different value before contact to another constant identical value afterwards.
8. Equilibrium is reached “after a long time”, which means after a period of time much longer than some *relaxation time* that depends on the detailed physics of the system. Whenever needed, we will always assume that we are in such state of equilibrium, without trying to answer the question of how much time it takes. In equilibrium, all thermodynamical properties of system stay constant in time.
9. A system in equilibrium is called *homogeneous* if the state variables are the same everywhere (i.e., constant not only in time but also in space). Two or more systems in equilibrium, possibly with different thermodynamic properties, separated by boundaries are called *heterogeneous*.
10. The relations between state variables are called *equations of state*.
11. A state variable can be *intensive* if it does not depend on the size (i.e. on N or V); *extensive* if it depends linearly on the size, i.e. if n identical systems with state variable Q have a value nQ when brought into equilibrium. Intensive variables are p, T, μ . Extensive variables are E, N, V, S .
12. The following obvious rules apply: function of intensive variables only=intensive; intensive \times extensive=extensive; extensive/extensive=intensive; linear combination of extensive variables=extensive.

13. *Zero-th law of thermodynamics*: if systems A and B are in equilibrium with C , then also A and B are in equilibrium.
14. The temperature can be defined by the observable effects it induces on some substances, eg the expansion of a gas or liquid or the radiation emitted by a body or the change in electric resistance. For many gases, a strictly monotonic relation between T and the gas volume is observed. The scale of T can therefore always be defined as (thermodynamic temperature)

$$T = T_0 \frac{V}{V_0} \quad (1.1)$$

assuming P, N constant.

1.2 Ideal gas

1. A ideal gas is defined as a gas of very small classical non-relativistic particles that do not interact with each other except through direct elastic collision (i.e the kinetic energy, included the rotational or vibrational one, is conserved), that move frictionless and do not dissipate energy via radiation or other means, and that is contained into a perfectly elastic-wall container.
2. Every gas at sufficiently high temperature and low pressure behave as a ideal gas.
3. A ideal gas is found to obey the Boyle-Mariotte law:

$$pV = p_0V_0 \quad (1.2)$$

at constant T, N .

4. Combining with (1.1) we obtain

$$\frac{pV}{T} = \text{const} \quad (1.3)$$

5. Since p, T are intensive and V is extensive, the constant must be proportional to N . Therefore we have the equation of a ideal gas

$$pV = NkT \quad (1.4)$$

where experimentally it is found the value of k to be Boltzmann's constant, $k = 1.38 \cdot 10^{-23} JK^{-1}$.

6. This can be written also as

$$p = \rho kT \quad (1.5)$$

$$pV = nRT \quad (1.6)$$

where ρ is the molecule number density, n the number of moles, and $R = Nk/n = N_A k$ the gas constant and $N_A \approx 6 \cdot 10^{23}$ is Avogadro's number.

7. The temperature can be interpreted as a measure of kinetic energy per particle. In fact, denoting with $Nf(\vec{v})d\vec{v}$ the number of particles with velocity in $d\vec{v}$ one obtains that the pressure exerted on the walls of the container of area A along the direction z is

$$p = \frac{1}{A} \int dF_A = \frac{N}{2V} \int dv_x \int dv_y \int dv_z f(\vec{v}) 2mv_z^2$$

(the overall $\frac{1}{2}$ factor arises because only half of the particles move in the positive direction). Assuming an isotropic distribution of velocities in a ideal gas, $f(\vec{v}) = g(v^2) = g_1(v_x^2)g_2(v_y^2)g_3(v_z^2)$, this becomes

$$pV = mN \int d^3v f(|\vec{v}|) v_z^2 = mN \langle v_z^2 \rangle = \frac{mN}{3} \langle v^2 \rangle = \frac{2}{3} N \langle \varepsilon_{KIN} \rangle \quad (1.7)$$

ie pV is proportional to the total kinetic energy of the particles. We are assuming here the particles have only 3 degrees of freedom.

8. Then we see that

$$\langle \varepsilon_{KIN} \rangle = \frac{3}{2}kT \quad (1.8)$$

$$E_{KIN} = \frac{3}{2}NkT \quad (1.9)$$

i.e. $kT/2$ for each degree of freedom.

9. If the molecules have d independent kinetic degrees of freedom (all equally excited), one has

$$E_{KIN} = \frac{d}{2}NkT \quad (1.10)$$

10. Then one also obtains the unit-normalized Maxwell distribution (see Sect. 3.4 for the derivation):

$$f(\vec{v}) = \left(\frac{m}{2\pi kT}\right)^{3/2} \exp\left(-\frac{mv^2}{2kT}\right) \quad (1.11)$$

1.3 Exact and inexact differentials

Let us consider the following differential form

$$\delta f = A(x, y)dx + B(x, y)dy \quad (1.12)$$

This form is called an exact differential if it exists a function $f(x, y)$ such that

$$A = \frac{\partial f}{\partial x}|_y, \quad B = \frac{\partial f}{\partial y}|_x \quad (1.13)$$

In this case, we write df instead of δf . The necessary and sufficient condition for f to exist is

$$\frac{\partial A}{\partial y}|_x = \frac{\partial B}{\partial x}|_y \quad (1.14)$$

because in this case

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x} \quad (1.15)$$

as it should. This can be straightforwardly extended to many dimensions. If $\delta f = A_i dx^i$, (sum over repeated indexes) then the condition for δf to be an exact differential is $A_i = \partial f / \partial x_i$ and therefore

$$\frac{\partial A_i}{\partial x^j} = \frac{\partial A_j}{\partial x^i} \quad (1.16)$$

where all the other variables are held fixed. In this case, $df = \vec{\nabla} f \cdot d\vec{r}$ where the components of the gradient vector are $\nabla_i f = A_i$.

The integral over an exact differential depends only on the initial and final point, not on the path. In fact

$$\int_A^B df = \int_A^B \vec{\nabla} f \cdot d\vec{r} = f_B - f_A \quad (1.17)$$

Consequently, the integral along any closed curve vanishes. If a force can be written as the gradient of a potential, then the work done by the force is independent of the path (conservative force).

1.4 Work and heat

1. If \vec{F} is the force exerted by the system on the environment, the mechanical work done is

$$\delta W = -\vec{F} \cdot d\vec{s} \quad (1.18)$$

where $d\vec{s}$ is the movement along which the force is applied.

2. This can be interpreted in a general way: every work done by the system can be written as a generalized force times a generalized interval. For instance under the action of pressure one has

$$\delta W = -pdV \quad (1.19)$$

3. The work is not an exact differential so it is not a state variable; it is positive if done on the system (the volume decreases so $dV < 0$), negative otherwise.

4. Similarly, the work done on the system when adding a number dN of particles in equilibrium is

$$\delta W = \mu dN \quad (1.20)$$

where μ is the chemical potential (ie, as we will see better soon, the energy gained by a system when a single particle is added).

5. Since energy is an extensive quantity, we have

$$\delta W = \text{intensive} \times d(\text{extensive}) \quad (1.21)$$

6. Beside this mechanical (or electric etc) work, a system can increase its energy also by the transfer from or to the environment of heat, Q . Heat can also be created by doing mechanical work on a system, eg. stirring a liquid: doing a work of 4.183 J on 1g of water increases its temperature from 14.5 to 15.5 Celsius degrees at standard pressure. From now on, however, heat is always counted only if it is transferred from/to the outside. Any other source of internal heat, e.g. friction dissipation, is not included in Q . The sign of δQ is positive if transferred to the system.

7. One can say that δW expresses the change in ordered energy, while δQ the change in disordered energy. It is intuitive that while it is easy to transform ordered motion into disordered motion just by random collisions, the opposite is practically impossible. That is, one can easily (i.e. with a high efficiency) transform mechanical work into heat but not viceversa.

8. When heat is transferred to a system, its energy, as measured by temperature, is changed

$$\delta Q = CdT \quad (1.22)$$

where C is the heat capacity. Its value depends on how the heat is transferred (indeed, Q is not a state variable), so there is a C_V (when volume is kept constant) and a C_p (when pressure is kept constant).

9. The heat capacity is an extensive variable. The specific heat $c = C/(Nm)$, where m is the mass of a molecule, is an intensive variable.
10. A reversible process is a ideal transformation that moves continuously from an equilibrium state to another one. One can define the state variables at every instant. In practice, one can only have quasi-static processes, that move so slowly that every state can be assumed to be in equilibrium. In a reversible transformation, no internal heat (eg due to friction etc) develops. An irreversible process is a process in which the system goes out of equilibrium. When the system goes back to equilibrium, the state variables can be measured and their value will be independent of the process. In irreversible processes, some internal heat always develops.
11. Reversible processes must be engineered with great care; irreversible processes can occur spontaneously. All natural processes are irreversible to a larger or smaller degree.

12. The most important processes are isothermal (constant T , obtained by contact with an infinite heat bath), adiabatic (no heat exchange), isochoric (constant volume), isobaric (constant pressure), cyclic (all the state variables go back to the initial value).
13. For instance, if a isothermal expansion of a ideal gas is reversible, we can apply the ideal gas law at all times and obtain the total work

$$W_{tot} = \int \delta W = - \int p dV = -NkT \int \frac{dV}{V} = -NkT \log \frac{V_2}{V_1} \quad (1.23)$$

Here p is both the internal and the external pressure, since the gas must always be in an equilibrium condition. In an irreversible process the work done will always be smaller than this because some work goes into internal friction, turbulences etc. For instance, if the expansion happens in a previously empty volume the gas particles do not do any work and $W_{tot} = 0$. Similarly, if the volume into which the expansion takes place has a pressure smaller than p_{gas} , again $|W_{tot}(irr)| < |W_{tot}(rev)|$. Due to our sign convention, we have then

$$W_{irr} > W_{rev} \quad (1.24)$$

One can still use Eq. (1.23), but now with p_{ext} (if it has a homogeneous value) instead of p_{gas} (which is ill defined since the gas will not be in equilibrium). So, for two processes that start and end with the same state variables, the total work is always larger for an irreversible process than for a reversible one (as usual counting as positive the work done on the system).

1.5 First law

1. *The first law of thermodynamics* is the conservation of energy

$$dU = \delta W + \delta Q \quad (1.25)$$

Although W, Q are not state variables, U is, so dU is an exact differential.

2. For reversible processes, $\delta W = -pdV$, $\delta Q = C_V dT$ (or $\delta Q = C_p dT$)
3. If in a process the volume is constant, $dU = \delta Q = C_V dT$ so that

$$C_V = \left. \frac{\partial U}{\partial T} \right|_V \quad (1.26)$$

and from (1.10)

$$C_V = \frac{3}{2} Nk \quad (1.27)$$

(for 3 d.o.f.). In this case C_V does not depend on temperature and one has for an ideal gas

$$U = C_V T \quad (1.28)$$

Notice that the relation $dU = C_V dT$ is always valid for reversible processes, although it has been obtained assuming constant volume. The reason is that U is a state function and it does not depend on the path chosen to produce a change dU . For instance, in the next line we apply it to an adiabatic process in which the volume is changing.

4. For an adiabatic process, $dU = \delta W$ and for an ideal gas one has $C_V dT = -NkT dV/V$, i.e.

$$pV^\gamma = const \quad (1.29)$$

where $\gamma = 1 + 2/d = 5/3$ for a monatomic ideal gas, $7/5$ for a diatomic one (where we assume equipartition of energy among the two rotational degrees of freedom of an ideal diatomic molecule, so $C_V = 5Nk/2$).

5. Consider an irreversible isothermal expansion process that brings an ideal gas from V_1 to V_2 (i.e. the system performs work on the environment). Since the process is isothermal and U depends only on T , $dU = 0$. As we have seen, it always will produce less absolute work than a reversible one, (so due to our sign convention the work is larger); because of the first law, it then will also absorb less heat from the environment, so we have

$$Q_{irr} < Q_{rev} \quad (1.30)$$

$$W_{irr} > W_{rev} \quad (1.31)$$

That is, since the system gets heated by the dissipation of internal friction and turbulences, it will need to absorb less heat from the heat bath to keep the same temperature. Similarly, during an isothermal compression, the irreversible process will dissipate more heat to the environment (because it has to dissipate also the internally produced heat) and more work on the system will be needed. Eqs. (1.31) are then valid also during the compression.

6. This property is very general and does not depend on the ideal gas law. In any process in which the total energy does not change (i.e. in isothermal or cyclic processes) an irreversible process will always produce less absolute work on the environment (higher W due to the negative sign) and absorb less heat from it, during an expansion, and requires more work and releases more heat, during a compression.
7. More generally, for any two processes that start with the same state variables and end with the same state variables, the total work (as usual counting as positive the work done on the system) is always larger, and total heat absorbed is always smaller, for an irreversible process than for a reversible one.
8. We can now define a heat capacity at constant pressure, C_P , as the heat δQ_P transferred during a process at $p = const$ in which the temperature varies by dT . For an ideal gas, the work done at constant pressure is $\delta W = -pdV = -kNdT$ and $dU = (3/2)kNdT$. Therefore, by the first law

$$\delta Q_P = dU - \delta W = \frac{5}{2}kNdT \quad (1.32)$$

or $\delta Q = (d+1)kNdT/2$ for a gas with d degrees of freedom. Therefore for an ideal monoatomic gas,

$$C_P = \frac{5}{2}kN \quad (1.33)$$

Since $C_P > C_V$, we can say it takes more heat in a $p = const$ process to achieve the same temperature than in a $V = const$ process, because in the latter case no work is done.

1.6 Second Law

1. In a reversible Carnot cycle (isothermal expansion, adiabatic expansion, isothermal compression, adiabatic compression), one can show that

$$\oint \frac{\delta Q_{rev}}{T} = 0 \quad (1.34)$$

2. This implies that the quantity $\delta Q/T$ is a state variable, called *entropy* (from the Greek, “in transformation”)

$$S = S_A + \int_A^B \frac{\delta Q}{T} \quad (1.35)$$

i.e. $1/T$ is an *integral factor* of δQ , i.e. a factor that transforms a quantity into an exact differential. The entropy is an extensive variable.

3. The total work (on the environment) done in a Carnot cycle is

$$-\Delta W = \oint pdV = \oint TdS \quad (1.36)$$

and the efficiency is

$$\eta = \frac{|\Delta W|}{|\Delta Q_{ABS}|} = 1 - \frac{T_\ell}{T_h} \leq 1$$

where $T_{h,\ell}$ are the high, low heat bath temperatures and ΔQ_{ABS} denotes only the absorbed heat from T_h . The efficiency does not depend on the material, on the pressure, volume etc.

4. Every reversible cycle can be approximated as a number of subsequent Carnot cycles. Therefore, the equation

$$\oint \frac{\delta Q_{rev}}{T} = 0 \quad (1.37)$$

is true for every reversible cycle.

5. The Carnot cycle is the most efficient cycle among all the reversible cycles operating between $T_{h,\ell}$ and, *a fortiori*, among all the irreversible ones. Other cycles (e.g. Stirling's cycle) can be as efficient as Carnot but require several reservoirs at different temperature.
6. In the T, S plane, the Carnot cycle is a rectangle, which has the maximal area between T_h and T_ℓ once the state variables p, V of the initial and final expansion state are chosen.
7. *Second Law of Thermodynamics.* Since $\delta Q_{irr} < \delta Q_{rev}$ and due to the definition of entropy

$$TdS = \delta Q_{rev} \geq \delta Q_{irr} \quad (1.38)$$

If the transformation takes place in a thermally isolated system, such that there is no heat exchange, $\delta Q_{irr} = 0$ and therefore

$$dS \geq 0 \quad (1.39)$$

with the equality only for reversible transformation of isolated systems. As a consequence, the entropy for an isolated system in equilibrium has $dS = 0$, i.e. it is constant and maximal under any possible transformation.

8. In statistical mechanics these statements and the entire thermodynamics will be interpreted in a statistical sense. In particular, the entropy might fluctuate in both directions but for any system the amount of time the system will spend in a configuration of maximal entropy is by far larger than in any other configuration. In this sense, the entropy is always maximal except if one prepares the system in a very special initial condition.
9. From the definition $dS = \delta Q/T$ we can write for reversible transformations

$$dU = TdS - pdV \quad (1.40)$$

or, including the change in particle numbers

$$dU = TdS - pdV + \mu dN \quad (1.41)$$

We say that S, V, N are the *natural variables* for the function U . Since T is intensive and U extensive, the entropy S has to be extensive too. T, p, μ are called conjugate variables of S, V, N .

10. Conversely

$$dS = \frac{1}{T}dU + \frac{p}{T}dV - \frac{\mu}{T}dN \quad (1.42)$$

and U, V, N are the natural variables for S . If one has $U(S, V, N)$ or $S(U, V, N)$ for a system, then all the thermodynamical relation can be obtained. This is not necessarily true for other variables. For instance, from $U(T, N) = 3NkT/2$ one cannot obtain the other equation of state $pV = NkT$, but from $U(S, V, N)$ (see eg Eq. (1.48) below) one can obtain both.

In a system under a transformation that is...	reversible and cyclic	reversible and isolated	reversible and adiabatic	irreversible and isolated	all other cases
the total change in entropy ΔS is	0	0	0	≥ 0	not constrained

Table 1.1: Change of entropy under various circumstances.

11. Since U, S are state variables (dU, dS are exact differentials), a number of relations (equations of state) automatically arise:

$$\left. \frac{\partial U}{\partial S} \right|_{V,N} = T \quad (1.43)$$

$$\left. \frac{\partial U}{\partial V} \right|_{S,N} = -p \quad (1.44)$$

$$\left. \frac{\partial U}{\partial N} \right|_{S,V} = \mu \quad (1.45)$$

and similarly for S , e.g. $\left. \frac{\partial S}{\partial U} \right|_{V,N} = \frac{1}{T}$, etc.

12. By integrating dS , one can obtain the entropy of the ideal gas up to a constant. From eq. (1.42) ($dN = 0$) we have

$$dS = \frac{1}{T} \frac{3}{2} Nk dT + \frac{p}{T} Nk \left(\frac{dT}{T} - T \frac{dp}{p^2} \right) = Nk \left(\frac{5}{2} \frac{dT}{T} - \frac{dp}{p} \right) \quad (1.46)$$

and one obtains

$$S = S_0 + Nk \log \left(\frac{T}{T_0} \right)^{5/2} \left(\frac{p_0}{p} \right) \quad (1.47)$$

The entropy increases with temperature, as it should otherwise $\left. \frac{\partial S}{\partial U} \right|_{V,N} = \frac{1}{T}$ would be negative. Replacing T, p with the ideal gas equations of state one can also obtain

$$S(U, V, N) = S_0 + Nk \log \left(\frac{U}{U_0} \right)^{3/2} \left(\frac{N}{N_0} \right)^{-5/2} \left(\frac{V}{V_0} \right) \quad (1.48)$$

13. The second law applied to two subsystems in contact through a wall that can allow for exchange of heat or volume or particles shows that heat will always flow from hotter to colder bodies, the volumes will change until the pressure on both sides equalizes, and the particles will flow from denser to less dense regions. The second law expresses in fact the tendency to equilibrium of all isolated systems.
14. The same setting will allow to show that S has to be a convex function of U, V and U a concave function of S, V . I.e., the tendency of S to maximize itself is the same as the tendency of U to minimize.
15. By exploiting the fact that $U(\alpha S, \alpha V, \alpha N) = \alpha U(S, V, N)$ one can obtain the *Gibbs-Duhem equation*

$$U = TS - pV + \mu N \quad (1.49)$$

(or $\sum_i \mu_i N_i$ for several species). This is remarkable since in general T, p, μ depend on S, V, N themselves.

16. Combining with dU , one can see that

$$SdT - Vdp + Nd\mu = 0 \quad (1.50)$$

which implies that there must be a relation between the intensive variables T, p, μ , i.e. one of them can be written as a function of the others.

function	form	nat. variables	name
U		S, V, N	energy
F	$U - TS$	T, V, N	free energy (Helmholtz)
H	$U + pV$	S, p, N	enthalpy
G	$U - TS + pV$	T, p, N	Gibbs free energy
Φ	$U - TS - \mu N$	T, V, μ	Grand potential

Table 1.2: Thermodynamic potentials

1.7 Thermodynamic potentials

1. By using Legendre transformations one can introduce several other state functions. They can be useful because in practical experiments one can control, for instance, T or p , but not S, U
2. A Legendre transformation of a function $U(q)$ is defined as a new function $\hat{U}(p)$ such that

$$\hat{U}(p) = U - pq \quad (1.51)$$

where

$$p = \frac{dU}{dq} \quad (1.52)$$

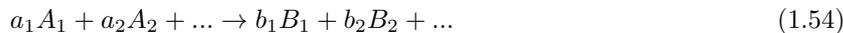
We assume that this relation can be inverted so that we can always write $q = q(p)$. By differentiating U we have $dU = pdq$ and therefore we find

$$d\hat{U} = dU - pdq - qdp = -q(p)dp \quad (1.53)$$

which shows that indeed \hat{U} depends only on p . Therefore, the Legendre transformation creates a new function that depends only on the derivative of the old function. This can be extended to functions of several variables (one can Legendre-transform with respect to a subset of the variables).

3. One introduces in this way several functions, all of them obtained from the energy (the Legendre functions obtained from the entropy are more rarely used)
4. In a system with $T, N = \text{const}$, $dF = dU - TdS = -pdV$, i.e. the change in free energy gives the amount of energy available to do useful work. If the system is isolated, so no work is done or received, the free energy is stationary (and a maximum or minimum depending on which variable; is a minimum when the variables are the standard T, V, N). This is why free energy is useful in engine constructions.
5. We see from Eq. (1.41) that adding a single particle in equilibrium to a system at constant S, V , i.e. for $dN = 1$, the energy changes by $dU = \mu$. However, if only V is kept constant, then μ should be interpreted as the change in free energy, $\mu = dU - TdS$. Adding a particle means increasing U but also S , since new microstates become available. Whether μ is positive or negative depends therefore on the particular system.
6. For processes at constant entropy, minimization of F amounts to minimization of U . So systems at constant entropy minimize the internal energy until they reach equilibrium. For instance, a marble rolling down a bowl with negligible friction so that it does not get heated but still slows down due to air resistance, maintains constant entropy, and its energy (potential plus kinetic) will be minimized when its kinetic energy gets completely dissipated and will stop at the equilibrium point at the bottom of the bowl.
7. For general processes at constant T , neither is entropy maximized nor energy minimized: rather, it's the free energy that is minimized.
8. Similarly, for isobaric processes, the change in enthalpy $H = U + pV$ gives the amount of heat and other non-mechanical work (eg exchange of particles) released or absorbed by the system. Most chemical processes occur at constant (atmospheric) pressure and the enthalpy is therefore the relevant quantity. In an adiabatic, isobaric process, the enthalpy never increases, stays constant for reversible processes and reaches a minimum at equilibrium.

9. Gibbs free energy (also called free enthalpy) $G = U - TS + pV$ gives the amount of non-mechanical work (eg exchange of particles) in isothermal, isobaric processes.
10. If there are several species, $G = \sum_i \mu_i N_i$. Now, let us consider a chemical reaction between particles (or an elementary particle interaction), such that we go from an initial number a_i of particles A_i to a final number b_i of particles B_i :



Then if the process is isothermal and isobaric, G has to remain constant, $dG = 0$ and therefore

$$\sum_i \mu_{A_i} dN_i = 0 \quad (1.55)$$

In a reaction, the change in number of particles of type i is now $dN_{A_i} = -a_i dN$ and $dN_{B_i} = b_i dN$, where dN is a common factor. Then we obtain a condition on the chemical potential that has to be fulfilled in any reaction

$$\sum_i \mu_{A_i} a_i = \sum_i \mu_{B_i} b_i \quad (1.56)$$

11. Finally, one can also define the grand potential $\Phi = U - TS - \mu N = -pV$, which reaches a minimum if no mechanical work is performed.
12. The relations between the thermodynamic potentials are called Maxwell relations and can be obtained by using the identities for partial derivatives.

1.8 Energy and entropy

1. In all branches of physics we learn that systems try to reach their ground state, i.e. the state of minimal energy. A pendulum will stop oscillating, a marble will roll down a bowl and will stop at the bottom, an electron in a hydrogen atom will radiate until it reaches the ground state. The reason is that in any process there is always a fraction of the work done by a system that is irretrievably lost into various forms of friction, so that the energy of the system will decay into its minimal value. Now we show how this result is related to the maximization of entropy.
2. Let us consider an isolated system composed by two subsystems A and B, both initially in equilibrium. System A is thermally isolated, so its entropy is constant, but can exchange work with B. For A, $dU_A = \delta W_A$ since $\delta Q_A = 0$. Now let's assume an amount $\delta W_A < 0$ (since it's work by the system A) is transferred from A to B, for instance through a movement of the separating wall, or radiation, or a long-range force, so that system B acquires an energy $dU_B = -dU_A$ (since $U_A + U_B = \text{const}$). In a spontaneous process, some of this work will go into friction, which generates heat in B, so that the amount of work δW_A transferred to B can be divided into a fraction ϵ as heat, and a fraction $1 - \epsilon$ as work. Then we have

$$dU_B = \delta Q_B + \delta W_B = -dU_A = -\delta W_A > 0 \quad (1.57)$$

and

$$\delta Q_B = -\epsilon \delta W_A > 0, \quad \delta W_B = -(1 - \epsilon) \delta W_A \quad (1.58)$$

3. We can therefore write $\delta Q_B = T dS_B > 0$, i.e. the entropy of system B has increased. This means the process will indeed take place spontaneously. But this can then be repeated over and over again, until a new equilibrium state for the whole system is achieved. Therefore system A will lose energy until the two systems reach an equilibrium state. Total entropy is maximized, while energy of system A is minimized.
4. The conclusion is that a system with constant entropy will tend to a state of minimal energy.

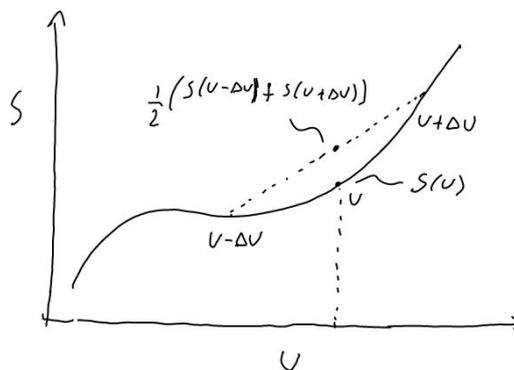


Figure 1.1: Plot of entropy as a function of internal energy.

1.9 Stability conditions

This section follows the discussion in Callen, *Thermodynamics and an Introduction to Thermostatistics*.

1. Suppose we have two systems A and B which exchange energy, volume, and particle number so that the total U, V, N remain constant:

$$U_A + U_B = E, \quad V_A + V_B = V, \quad N_A + N_B = N \quad (1.59)$$

Clearly, for every change, $dE_A = -dE_B, dV_A = -dV_B, dN_A = -dN_B$. Let us consider the change in entropy $S(U, V, N) = S_A(U_A, V_A, N_A) + S_B(U_B, V_B, N_B)$ after the systems are put in contact:

$$dS = dS_A + dS_B = \left[\frac{\partial S}{\partial U} \Big|_A - \frac{\partial S}{\partial U} \Big|_B \right] dU_A + \left[\frac{\partial S}{\partial V} \Big|_A - \frac{\partial S}{\partial V} \Big|_B \right] dV_A + \left[\frac{\partial S}{\partial N} \Big|_A - \frac{\partial S}{\partial N} \Big|_B \right] dN_A \quad (1.60)$$

$$= \left[\frac{1}{T_A} - \frac{1}{T_B} \right] dU_A + \left[\frac{p_A}{T_A} - \frac{p_B}{T_B} \right] dV_A + \left[\frac{\mu_A}{T_A} - \frac{\mu_B}{T_B} \right] dN_A \quad (1.61)$$

When the combined system reaches equilibrium, $dS = 0$. Since the three possible changes, dU_A, dV_A, dN_A , are in general independent, each of the three coefficients in square brackets has to vanish. This shows that two systems brought in contact will reach the same intensive quantities T, p and μ .

2. Now we study the conditions under which this equilibrium state is stable. Suppose the entropy $S(U, V, N)$ has a shape in function of U as in Fig. (1.1). Let us consider now two identical systems separated by a wall and isolated from the external environment. If we transfer an amount ΔU of internal energy from one system to the other, the entropy will change from $2S(U, V, N)$ to

$$S(U - \Delta U, V, N) + S(U + \Delta U, V, N) \quad (1.62)$$

3. As it appears clearly from Fig. (1.1), the entropy will then increase. This implies that if we remove the constraints, this transformation will be favoured and therefore *will* take place. This means that from a state in which the two joined systems were homogeneous, we reach a state of inhomogeneity. Even inside each subsystem, such a spontaneous transfer of energy will take place and the system will spontaneously leave the initial equilibrium state. This particular curve $S(U)$ produces therefore an *instability*.
4. The condition for stability is therefore that the entropy does not increase, i.e.

$$S(U - \Delta U, V, N) + S(U + \Delta U, V, N) \leq 2S(U, V, N) \quad (1.63)$$

By expanding in Taylor series on the lhs we find

$$S - \frac{\partial S}{\partial U} \Delta U + \frac{1}{2} \frac{\partial^2 S}{\partial U^2} \Delta U^2 + S + \frac{\partial S}{\partial U} \Delta U + \frac{1}{2} \frac{\partial^2 S}{\partial U^2} \Delta U^2 \leq 2S \quad (1.64)$$

or

$$\frac{\partial^2 S}{\partial U^2} \Big|_{V,N} \leq 0 \quad (1.65)$$

That is, the curve $S(U)$ has to be concave everywhere for the system to remain stable.

5. An analogous argument also shows that if we fix U, N and change V , we should also impose the condition

$$\frac{\partial^2 S}{\partial V^2} \Big|_{U,N} \leq 0 \quad (1.66)$$

Changing at the same time U, V , we finally obtain the condition

$$\frac{\partial^2 S}{\partial V^2} \frac{\partial^2 S}{\partial U^2} - \left(\frac{\partial^2 S}{\partial V \partial U} \right)^2 \geq 0 \quad (1.67)$$

In other words, the stability conditions requires that $S(U, V, N)$ is a *concave* function of V, U (for fixed N). The surface $S(U, V)$ has to remain below the tangent envelope for any U, V .

6. From Eq. (1.65) and the fact that $(\partial S / \partial U)_{V,N} = 1/T$ one can obtain

$$\frac{\partial^2 S}{\partial U^2} \Big|_{V,N} = -\frac{1}{T^2} \frac{\partial T}{\partial U} \Big|_{V,N} = -\frac{1}{T^2 C_V} \leq 0 \quad (1.68)$$

i.e., the heat capacity C_V should be positive. We will find below several other similar conditions of stability.

7. A parallel argument can be developed for the energy $U(S, V, N)$. If the energy were to decrease after moving some entropy from one subsystem to another one,

$$U(S - \Delta S, V, N) + U(S + \Delta S, V, N) \leq 2U(S, V, N) \quad (1.69)$$

the transformation would be favourable and it will take place, making the system inhomogeneous and therefore unstable. We have to ensure therefore the following conditions

$$\frac{\partial^2 U}{\partial S^2} \Big|_{V,N} = \frac{\partial T}{\partial S} \Big|_{V,N} \geq 0 \quad (1.70)$$

$$\frac{\partial^2 U}{\partial V^2} \Big|_{S,N} = -\frac{\partial p}{\partial V} \Big|_{S,N} \geq 0 \quad (1.71)$$

$$\frac{\partial^2 U}{\partial S^2} \frac{\partial^2 U}{\partial V^2} - \left(\frac{\partial^2 U}{\partial S \partial V} \right)^2 \geq 0 \quad (1.72)$$

That is, the energy surface has to be *convex* on the plane S, V for the system to remain stable.

8. We can extend these conditions to the other thermodynamic potentials, by using the properties of the Legendre transformations. Suppose $U(X)$ is a function and

$$\hat{U} = U - PX \quad (1.73)$$

its Legendre transformation, where $P = \frac{dU}{dX}$. We have then $d\hat{U} = -XdP$, where $X = -d\hat{U}/dP$. Then we have

$$\frac{dX}{dP} = \left(\frac{dP}{dX} \right)^{-1} = \left(\frac{d^2 U}{dX^2} \right)^{-1} \quad (1.74)$$

but at the same time also

$$\frac{dX}{dP} = -\frac{d^2 \hat{U}}{dP^2} \quad (1.75)$$

This implies

$$\left(\frac{d^2 U}{dX^2} \right)^{-1} = -\frac{d^2 \hat{U}}{dP^2} \quad (1.76)$$

That is, if a function $U(X)$ is convex with respect to X , its Legendre transform \hat{U} is concave with respect to the conjugate variable P . The other non-transformed variables have the same concavity of the original function. So, let's consider the free energy $F = U - TS$. Since $U(S)$ is convex, we have that

$$\frac{\partial^2 F}{\partial T^2} \Big|_{V,N} \leq 0 \quad (1.77)$$

because T is the conjugate of S , while

$$\frac{\partial^2 F}{\partial V^2} \Big|_{T,N} \geq 0 \quad (1.78)$$

because V is not transformed in this Legendre transformation. By the same arguments, one can see that

$$\frac{\partial^2 H}{\partial S^2} \Big|_{p,N} \geq 0, \quad \frac{\partial^2 H}{\partial p^2} \Big|_{S,N} \leq 0 \quad (1.79)$$

and

$$\frac{\partial^2 G}{\partial T^2} \Big|_{p,N} \leq 0, \quad \frac{\partial^2 G}{\partial p^2} \Big|_{T,N} \leq 0 \quad (1.80)$$

The general rule is that the *thermodynamic potentials* (i.e the Legendre transforms of the energy) are convex functions of their extensive variables and concave functions of their intensive variables. Exactly the opposite happens with the Legendre transforms of the entropy (sometimes called *Massieu functions*).

9. These conditions can be also understood as providing constraints on physical quantities as heat capacities and compressibilities. The latter are defined as

$$\beta_T = -\frac{1}{V} \frac{\partial V}{\partial p} \Big|_T \quad (1.81)$$

$$\beta_S = -\frac{1}{V} \frac{\partial V}{\partial p} \Big|_S \quad (1.82)$$

i.e. as the relative change in volume when pressure is applied either isothermally (β_T) or adiabatically (β_S). Beside $C_V \geq 0$, as already seen, we have for the isothermal compressibility the stability condition

$$\frac{\partial^2 F}{\partial V^2} \Big|_{T,N} = -\frac{\partial p}{\partial V} \Big|_T = \frac{1}{V\beta_T} \geq 0 \quad (1.83)$$

i.e. $\beta_T > 0$. Since from thermodynamics we know that

$$C_p - C_V = VT \frac{\alpha^2}{\beta_T} > 0 \quad (1.84)$$

$$\frac{C_p}{C_V} = \frac{\beta_T}{\beta_S} \quad (1.85)$$

(where α is the thermal expansion coefficient, $\alpha \equiv \frac{1}{V} \frac{\partial V}{\partial T} \Big|_p$) we have also $C_p \geq 0, \beta_S \geq 0$.

10. These conditions give a mathematical implementation of the principle of Le Chatelier: every inhomogeneity that develops in a system should induce a process that tends to smooth out the inhomogeneity.

Chapter 2

Introduction to Statistics

2.1 Probability and frequency

The standard frequentist definition of the probability of having an event X out of many trials N is

$$P(X) \equiv \lim_{N \rightarrow \infty} \frac{\text{number of occurrences of } X}{N} \quad (2.1)$$

that is, the *frequency* of events. This definition of course assumes that the events X are all equiprobable, which is a kind of circular statement (replacing with “events produced under identical circumstances” does not solve the problem). In fact, this is a definition of *frequency*, rather than of probabilities.

The Bayesian definition, on the other hand, is purely theoretical: the probability is simply assigned to an event basing on our previous experience or on our previous information about it. If we have no information, e.g. we use a die for the first time, we should simply assign equal probability to every possible outcome.

In either case, if X_i with $i = 1, \dots, m$ are all the possible outcomes of the trials (eg, X_1 =”obtaining 1 in a die throw”, X_2 =”obtaining 2” etc)

$$\sum_i^m P(X_i) = 1 \quad (2.2)$$

The sum of all possible probabilities of any given experiments equals 1.

In the limit in which the random variable n_i becomes a continuous variable (eg a temperature, a magnitude etc), we define a probability density or *probability distribution function* (PDF) $f(x)$

$$f(x)dx = P(x) \quad (2.3)$$

and we have, within the domain of x (i.e. all its possible values)

$$\int f(x)dx = 1 \quad (2.4)$$

The two fundamental properties of probability distributions are

$$\int f(x)dx = 1 \quad (2.5)$$

$$f(x) \geq 0 \quad (2.6)$$

We can easily extend the idea to joint events, for instance the probability of obtaining at the same time (non necessarily in a chronological sense) the measurement x in dx (eg a galaxy magnitude) and y in dy (eg the galaxy redshift). Then we have

$$f(x, y)dxdy = P(x, y) \quad (2.7)$$

$$f(x, y) \geq 0 \quad (2.8)$$

$$\int f(x, y)dxdy = 1 \quad (2.9)$$

Immediate consequence of the first law is that if $F(< X) = \int_{-\infty}^X f(x)dx$ is the probability of obtaining a result less than X (F is called *cumulative function*), then the probability of obtaining a result greater than or equal to X is $F(\geq X) = 1 - F(< X)$. So in general if $P(A)$ is the probability of A , the probability of non- A (ie anything but A) is simply $1 - P(A)$, to be denoted as $P(\bar{A})$.

2.2 Probability of combined events

Suppose we throw two dice; the *joint probability* of obtaining 1 in a throw and 2 in the other one is the product of the single-throw probabilities, $P_{12} = P_1P_2$. This is true only because the two throws are independent (do not influence each other). Then we have $P_1 = P_2 = 1/6$ and $P_{12} = 1/36$, as of course one could see by the number of occurrences over the number of experiments. If we have the PDF $f(x, y)$, the event x can be said to be independent of event y if the probability of x , $p(x)dx$, does not depend on y . Now the probability of having x in dx when y is in a sub-range Δ_y is

$$p(x)dx = dx \int_{\Delta_y} dy f(x, y) \quad (2.10)$$

So in order for $p(x)$ not to depend on y it is necessary that f be separable:

$$f(x, y) = q(x)g(y) \quad (2.11)$$

So in this case

$$p(x)dx = q(x)dx \int_{\Delta_y} g(y)dy = Nq(x)dx \quad (2.12)$$

where N is only a normalization constant independent of x (to be fixed by putting $\int p(x)dx = 1$ ie. $N = 1/(\int q(x)dx)$). Then two events x, y are independent if and only if the joint probability $f(x, y)$ is separable. So we define the joint probability as:

If P_A and P_B are the probabilities of the independent events A e B , the probability of having both A and B is P_AP_B . Then

$$P(A \cap B) = P(B \cap A) = P(A)P(B) \quad (2.13)$$

Other two forms of combining probabilities can be defined.

Disjoint P. If P_A and P_B are the prob. of events A and B mutually exclusive (i.e. $A \cap B = 0$), the prob. of A or B is $P_A + P_B$. Therefore

$$P(A \cup B) = P(A) + P(B) \quad (2.14)$$

We have already seen an example of disjoint prob. when we have seen that $P(A) = 1 - P(\bar{A})$. Since A and \bar{A} are mutually exclusive, we can write

$$P(A \cup \bar{A}) = 1 = P(A) + P(\bar{A}) \quad (2.15)$$

So for instance the prob. of having 1 or 2 in a dice roll is $1/6+1/6=1/3$. Considering continuous variables we have

$$p(x \in A \cup B) = \int_A f(x)dx + \int_B f(x)dx \quad (2.16)$$

only if the ranges A and B do not overlap. If they overlap, the events are not mutually exclusive $A \cap B \neq 0$, and we have:

$$p(x \in A \cup B) = \int_A f(x)dx + \int_B f(x)dx - \int_{A \cap B} f(x)dx \quad (2.17)$$

In general therefore

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad (2.18)$$

So for instance if A is the prob. of having one "1" in the first die, whatever the second is, and B the prob. of "1" in the second die, whatever the first is, and we consider the prob. of having at least a "1" in two throws, the event "11" is both A and B . So we have $P(A \cup B) = 1/6 + 1/6 - 1/36 = 11/36$, as we can verify easily since the winning combinations are (11,12,13,14,15,16,21,31,41,51,61) are 11 over 36.

Conditional P. We can also define the conditional probability (prob. of A given B):

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\text{number of cases that are both A and B}}{\text{number of cases that are B}} \quad (2.19)$$

So for instance, the probability of the combination 1-2 *after* obtaining 1 in the first roll equals $(1/36)/(1/6) = 1/6$. Notice that the conditional probability of having A given B is always larger than or equal to the probability of having both A and B .

This extends obviously to continuous variables. The probability of obtaining x in the range $I = (-1, 1)$ given that $x < 0$ is $P(x \in I|x < 0)$. The prob. of having $x < 0$ is

$$P(x < 0) = \int_{<0} f(x)dx \quad (2.20)$$

and the probability of having $x \in I$ and at the same time $x < 0$ is

$$P(x < 0, x \in I) = \int_{-1}^0 f(x)dx$$

Now, the fraction of cases (or area) such that $P(x \in I|x < 0)$ is clearly the fraction $P(x < 0, x \in I)/P(x < 0)$, which agrees with the rule above. In other words, if in 100 measures there are 50 with $x < 0$ and 20 with $-1 < x < 0$ it is clear that the fraction of measures with $x \in I$ among those with $x < 0$ is $20/50=2/5$.

Another example. The prob. of obtaining ≥ 9 in two dice rolls is $10/36$: there are in fact 10 successful events: 36, 45, 46, 55, 56, 66, 63, 54, 64, 65 in 36 possibilities. Which is the prob. of obtaining a score ≥ 9 given that in the first roll the result is 6? We have

$$P(x + y \geq 9|x = 6) = P(x + y \geq 9, x = 6)/P(x = 6) = \frac{4}{36} \frac{6}{1} = \frac{2}{3} \quad (2.21)$$

which indeed is true since if the first die has a 6, then it is sufficient that the second result is 3,4,5,6 to win, i.e. 4 cases out of 6.

The prob. that $B \cup \bar{B}$ occurs is of course always 1, even in the case of conditional prob. We have therefore

$$P(B \cup \bar{B}|A) = 1 = \frac{P(A, B \cup \bar{B})}{P(A)} \quad (2.22)$$

or

$$P(A, B \cup \bar{B}) = P(A) \quad (2.23)$$

In terms of PDF this rule says that integrating a PDF of two variables over the whole domain of one of the two (*marginalization*) we obtain the PDF of the other:

$$\int f(x, y)dy = p(x) \quad (2.24)$$

Clearly if $f(x, y) = p(x)g(y)$, the $p(x)$ is indeed the PDF of x , which of course confirms that a factorizable $f(x, y)$ implies that x and y are independent.

2.3 Expected values

Let's briefly introduce two examples of PDFs.

Uniform distribution.

$f(x) = \text{const.}$ in the range $x \in (a, b)$. We have

$$\int_a^b f(x)dx = \text{const.} \times (b - a) \quad (2.25)$$

and the normalization requires $\text{const} = 1/(b - a)$.

Gauss distribution.

$$f(x) = Ae^{-\frac{(x-x_0)^2}{2\sigma^2}} \quad (2.26)$$

Normalization

$$\int f(x)dx = A \int_{-\infty}^{+\infty} \exp\left(-\frac{(x-x_0)^2}{2\sigma^2}\right)dx = A\sqrt{2\pi\sigma^2} \quad (2.27)$$

from which $A = (2\pi\sigma^2)^{-1/2}$. The parameters x_0 and σ^2 are called mean and variance.

These PDFs, like all other distributions we will introduce later on, can be synthetically characterized by several quantities.

Quantile α : value of x such that

$$\int_{-\infty}^x f(x')dx' = \alpha \quad (2.28)$$

($0 \leq \alpha \leq 1$). If $\alpha = 0.5$ the quantile is called median.

Mode. The value of x such that $P(x)$ is maximal.

Moments or expected values.

The expected value of a quantity $g(x)$ is defined as

$$E[g] = \langle g \rangle \equiv \int g(x)f(x)dx \quad (2.29)$$

The mean is therefore the expectation value of x :

$$E[x] = \int xf(x)dx \quad (2.30)$$

For discrete variables we have

$$E[n] = \sum_1^N n_i P(n_i) \quad (2.31)$$

Since $P(n_i)$ is defined as the number of events n_i divided by the total number of cases, we retrieve the intuitive definition of mean of a variable as the sum of all the values divided by the number of cases.

The variance (or central moment of second order) is defined as

$$E[(x - \langle x \rangle)^2] = \int (x - \langle x \rangle)^2 f(x)dx = \int x^2 f(x)dx - \hat{x}^2 = \langle x^2 \rangle - \langle x \rangle^2 \quad (2.32)$$

(sometimes also $\text{Var}(x)$). For a Gaussian one has

$$E[x] = x_0 \quad (2.33)$$

$$E[(x - \langle x \rangle)^2] = \sigma^2 \quad (2.34)$$

Note that $E[x - \langle x \rangle] = 0$ and $E[y^2] \neq E[y]^2$. For a uniform variable, one has

$$E[x] = \frac{b+a}{2} \quad (2.35)$$

$$E[(x - \langle x \rangle)^2] = \frac{(b-a)^2}{12} \quad (2.36)$$

The variance has great importance in scientific measures. Conventionally in fact the error associated to each measure is given by the square root of the variance, or standard deviation, and is denoted generally with σ also for non-Gaussian distributions.

The n -th order moment and the n -th order central moment are defined respectively as

$$E[x^n] = \int x^n f(x)dx \quad (2.37)$$

$$E[(x - \langle x \rangle)^n] = \int (x - \langle x \rangle)^n f(x)dx \quad (2.38)$$

2.4 Transformation of variables

Given a random variable x and its PDF $f(x)$, we could be interested to derive a PDF of a variable function of x , for instance x^2 or $1/x$ or $y(x)$. If $dy = y'dx$ is the infinitesimal interval of the new variable y as a function of the old one, it is clear that the prob. of having x in $x, x + dx$ must be equal to the one of having y in $y, y + dy$:

$$f(x)dx = g(y)dy \quad (2.39)$$

and therefore the new PDF $g(y)$ is

$$g(y) = f(x) \left| \frac{dx}{dy} \right| \quad (2.40)$$

where the absolute value ensures the positivity of the new PDF. Notice that in general

$$E[g(x)] \neq g(E[x]) \quad (2.41)$$

We can also consider the transformation of variables in the case of many random variables. The transformation from x_1, x_2, \dots to y_1, y_2, \dots can be performed introducing the Jacobian of the transformation

$$f(x_i)d^n x = g(y_i)d^n y \quad (2.42)$$

from which

$$g(y_i) = f(x_i)|J| \quad (2.43)$$

where $J_{ij} = \partial x_i / \partial y_j$ and $|\dots|$ denotes the determinant.

2.5 The main PDFs

Binomial PDF. Let us consider N independent events, eg the scores 1 – 3 – 2 – 6 etc in a series of dice rolls, or the sequence *TTHHTH* of heads/tails in coin tosses. We want to evaluate the probability that a joint event, eg 8 heads out of 10 tosses, or three times a 1 out of 4 dice rolls, regardless of the order in the sequence, i.e. considering the events as indistinguishables. This is exactly the same kind of statistics we need in eg the statistics of a gas, which depends on the probability for indistinguishable particles to be in a given region of phase space.

We need first of all to evaluate the number of possible sequences. If we have N different elements, eg a, b, c , we can permute the N elements $N!$ times. For instance, $N = 3$ elements can be combined $3! = 6$ times: *abc, acb, cab, cba, bac, bca*. Then $N!$ is the number of permutations of distinguishable elements.

Suppose now we have only two elements, eg head or tail, or event A and any other event \bar{A} . Then many permutations are identical, for instance *HHTTT* remains the same by exchanging the two H s and the three T s. Suppose we have n times one of the two elements and, therefore, $N - n$ the number of the other. Then, among the total $N!$ permutations, a fraction $n!$ is identical because we permute the same identical n element, and a fraction $(N - n)!$ will also be identical for the same reason. How many distinguishable combinations will we obtain? Clearly

$$\frac{\text{total permutations}}{(\text{permutations among } n)(\text{permutations among } N - n)} = \frac{N!}{n!(N - n)!} \equiv \binom{N}{n} \quad (2.44)$$

For instance, if $N = 4$ and $n = 2$ (as in *TTHH*) we will have $4!/2!/2! = 6$ equivalent combinations (*HHTT, HTHT, TTHH, THTH, THHT, HTTH*). Notice that for $n = 0$ we define $n! = 1$.

The binomial PDF generalizes this calculation to the case in which we have a series of n independent events A each with the same probability p (eg for “head” the prob. is $1/2$, for a 2 in a dice roll is $1/6$ etc). In this case, the occurrence of n events A or prob. p out of N implies the occurrence of $N - n$ events \bar{A} with prob. $1 - p$. All this implies a joint prob. of

$$p^n(1 - p)^{N - n} \quad (2.45)$$

But clearly we have $\binom{N}{n}$ of such combinations and therefore the binomial prob. will be

$$P(n; N, p) = \frac{N!}{n!(N - n)!} p^n(1 - p)^{N - n} \quad (2.46)$$

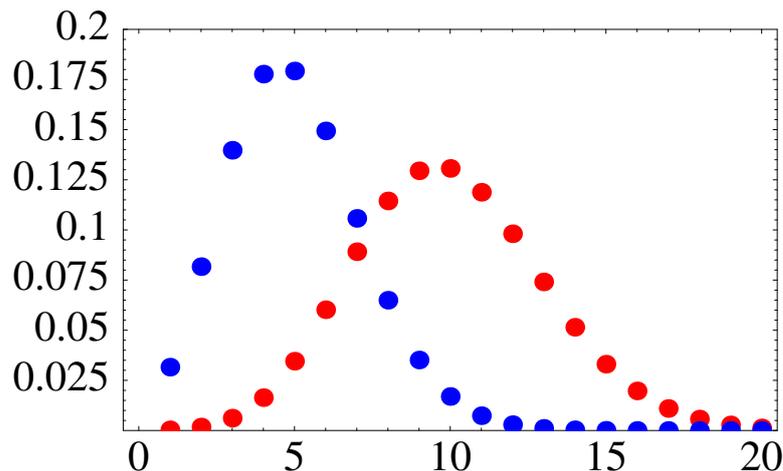


Figure 2.1: Binomial for $N = 120$ and $p = 1/12$ (red dots) e $p = 1/24$ (blu dots).

where n is the discrete random variable $0 \leq n \leq N$ (number of events A) while N, p are the distribution parameters. Notice that by employing the rules of the binomial we have, as indeed we should have expected:

$$\sum_n P(n; N, p) = (p + (1 - p))^N = 1 \quad (2.47)$$

It is also intuitive that the mean of events A of prob. (frequency) p out of N events should be the fraction p of N and indeed

$$E[n] = Np \quad (2.48)$$

$$\sigma^2 = E[(n - Np)^2] = Np(1 - p) \quad (2.49)$$

Let's demonstrate the first one:

$$E[n] = \sum_n nP(n; N, p) = \sum_{n=0}^N \frac{nN!}{n!(N-n)!} p^n (1-p)^{N-n} \quad (2.50)$$

$$= \sum_{n=1}^N \frac{N(N-1)!}{(n-1)!(N-n)!} p p^{n-1} (1-p)^{N-n} \quad (2.51)$$

$$= Np \sum_{n'=0}^{N'} \frac{(N')!}{(n')!(N'-n')!} p^{n'} (1-p)^{N'-n'} = Np \quad (2.52)$$

The binomial distribution for $N = 1$ is called Bernoulli distribution:

$$P(n; 1) = p^n (1-p)^{1-n} \quad (2.53)$$

for $n = 0, 1$. It applies for instance to a single toss of a coin and gives the probability that an event with probability p , e.g. tail, happens ($n = 1$) or does not happen ($n = 0$).

Poissonian PDF. Let us consider now the limit of the Binomial for $N \rightarrow \infty$ and $p \rightarrow 0$ (rare events), but keeping $Np = \nu$ finite. We obtain then the Poissonian PDF

$$P(n; \nu) = \frac{\nu^n}{n!} e^{-\nu} \quad (2.54)$$

The moments are

$$E[n] = e^{-\nu} \sum_n n \frac{\nu^n}{n!} = \nu \quad (2.55)$$

$$E[(n - \nu)^2] = \nu \quad (2.56)$$

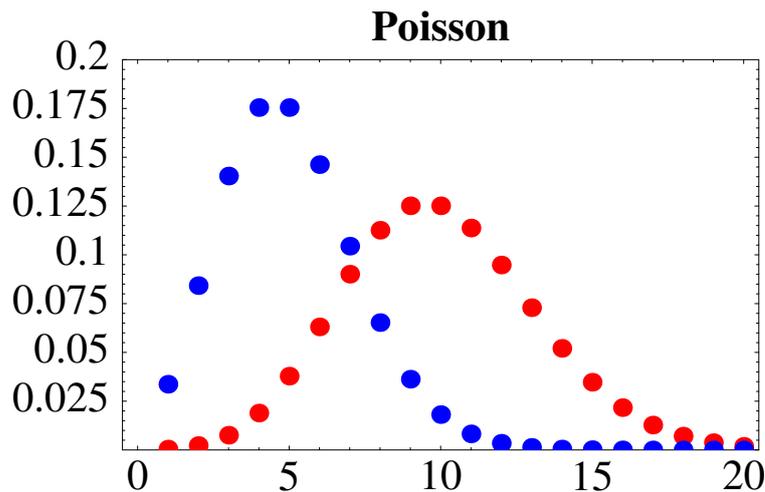


Figure 2.2: Poissonian for $\nu = 10$ (red) and $\nu = 5$ (blue). Note the similarity to the Binomial.

For large n , we can assume that n is a continuous variable. In this case we generalize to

$$P(x; \nu) = \frac{\nu^x}{\Gamma(x+1)} e^{-\nu} \quad (2.57)$$

where $\Gamma(x)$ (equal to $(x-1)!$ for x integer) is the gamma function

$$\Gamma(x) = \int_0^{\infty} e^{-t} t^{x-1} dt \quad (2.58)$$

Gaussian PDF. For large ν , the Poissonian is well approximated by the Gaussian with mean and variance ν . The Gaussian is defined as:

$$G(\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (2.59)$$

and has mean μ and variance σ^2 . Defining the new variable $z = (x - \mu)/\sigma$ the Gaussian becomes the Normal distribution:

$$N(x) = G(0, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (2.60)$$

We can define the *error function*

$$\text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \quad (2.61)$$

so that the cumulative function $F(x) = \int_{-\infty}^x G(x; 0, 1) dx$ becomes

$$F(x) = \frac{1}{2} \left[1 + \text{erf}\left(\frac{x}{\sqrt{2}}\right) \right] \quad (2.62)$$

The prob. that the gaussian variable x distributed as $G(x; \mu, \sigma)$ in in the range $(\mu - a, \mu + a)$ is

$$P(x \in (-a, a)) = \text{erf}\left(\frac{a}{\sigma\sqrt{2}}\right) \quad (2.63)$$

The Gaussian PDF is of such great importance not only because is the large- ν limit of the Poissonian but also because of the Central Limit Theorem (see Sect. 6.4):

Every random variable X sum (or linear combination) of many independent variables x_i (i.e. $X = \sum_i x_i$) is distributed approximately as a Gaussian of mean $\sum_i \mu_i$ and variance $\sigma_X^2 = \sum_i \sigma_i^2$ in the limit $n \rightarrow \infty$ independently of the individual PDFs.

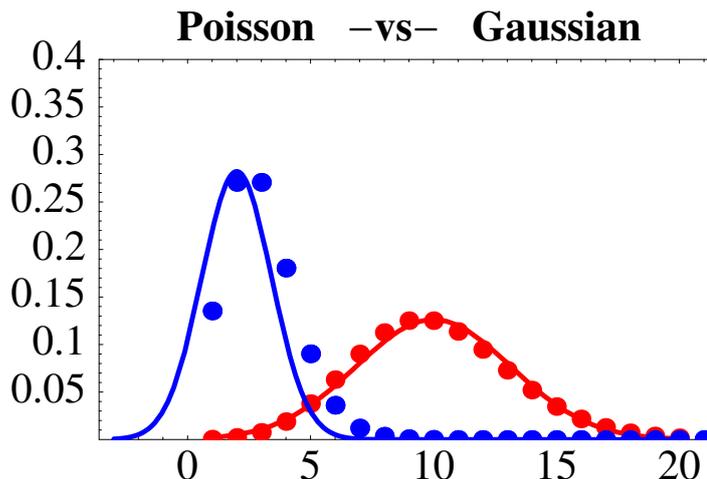


Figure 2.3: Comparing Poissonian and Gaussian PDFs for $\nu = 2$ (blue) and $\nu = 10$ (red).

In practice, the CLT can be applied in many experimental situations in which the error is the sum of many independent causes: reading errors, instrumental noise, contaminations etc. In these cases, the measure can be assumed to be gaussian distributed.

Three important values of the cumulative function are

$$F(\mu + j\sigma) - F(\mu - j\sigma) = \text{erf}\left(\frac{j}{\sqrt{2}}\right) = 0.68, 0.95, 0.997 \quad (2.64)$$

for $j = 1, 2, 3$: these give the prob. of finding x at $j = 1, 2, 3\sigma$ from the mean μ . Conventionally, errors are quoted at 1σ even for non-Gaussian distributions.

Exponential distribution. The PDF is given by

$$f(x) = \lambda e^{-\lambda x} \quad (2.65)$$

for $x \geq 0$ and 0 otherwise. The mean and the variance are

$$E[x] = \frac{1}{\lambda} \quad (2.66)$$

$$V[x] = E[(x - E[x])^2] = \frac{1}{\lambda^2} \quad (2.67)$$

2.6 Multivariate distributions

So far we have seen mostly distributions of single variables. We have however already defined the joint probability

$$f(x, y) dx dy \quad (2.68)$$

of having x, y in the area $dx dy$. The definition of probability requires now that

$$\int f(x, y) dx dy = 1 \quad (2.69)$$

It is clear that we can extend this definition to many variables x_i in the volume $d^n x$. For independent variables $f(x, y) = f_1(x)f_2(y)$.

Analogously to the 1D case, we define the means

$$E[x] = \int x f(x, y) dx dy = \mu_x \quad (2.70)$$

$$E[y] = \int y f(x, y) dx dy = \mu_y \quad (2.71)$$

and the covariance matrix

$$C_{ij} = E[(x_i - \mu_i)(x_j - \mu_j)] \quad (2.72)$$

$$= \int (x_i - \mu_i)(x_j - \mu_j)f(x_1, x_2)dx_1dx_2 \quad (2.73)$$

where x_i is the vector of random variables and μ_i the mean vector. The elements along the diagonal are the variances σ_i^2 of the individual random variables. If x_1, x_2 are independent then

$$C_{12} = \int (x_1 - \mu_x)(x_2 - \mu_y)f_1(x_1)f_2(x_2)dx_1dx_2 \quad (2.74)$$

separates out and by definition of mean $C_{12} = 0$: then the covariance matrix of independent variables is diagonal (however in general $C_{12} = 0$ does not imply independent variables, but only uncorrelated variables).

The degree of correlation is indicated by the weight of the off-diagonal elements. For any two variables we define

$$\rho_{xy} \equiv \frac{C_{xy}}{\sigma_x\sigma_y} \quad (2.75)$$

If we have the multivariate PDF $f(x, y)$ we can obtain the PDF of the individual variables by integrating out the other one:

$$g(x) = \int f(x, y)dy \quad (2.76)$$

This new PDF (*marginalized* over y) gives the probability of having x in dx whatever is y . We realize immediately that

$$E[x] = \int xg(x)dx \quad (2.77)$$

and similarly for all the other moments of x . All these definitions extend immediately to n dimensions, e.g.

$$C_{ij} = E[(x_i - \mu_i)(x_j - \mu_j)] \quad (2.78)$$

$$= \int (x_i - \mu_i)(x_j - \mu_j)f(x_1, x_2, x_3, \dots)d^n x \quad (2.79)$$

Multinomial distribution. The binomial distribution can be generalized to the case in which there are not just two possible outcomes with probability p and $1 - p$ but k possible outcomes each with probability p_i , $i = 1, \dots, k$, with the constraint that the outcomes exhaust all the possibilities, so $\sum_i p_i = 1$. Now the probability of having a particular sequence of independent outcomes formed by x_1 outcomes of type 1, x_2 of type 2, etc will be

$$p_1^{x_1} p_2^{x_2} p_3^{x_3} \dots p_k^{x_k} \quad (2.80)$$

Just as for the binomial distribution, accounting for all the possible internal permutations leads to the multinomial distribution, i.e. the probability that in a sequence of N trial one finds x_1 items of type 1, x_2 of type 2 etc. This is given by

$$P(x_1, x_2, \dots, x_k) = \frac{N!}{x_1!x_2!\dots x_k!} p_1^{x_1} p_2^{x_2} p_3^{x_3} \dots p_k^{x_k} \quad (2.81)$$

(provided $\sum_i p_i = 1$). The expected values and variances are

$$E[x_i] = Np_i \quad (2.82)$$

$$Var[x_i] = Np_i(1 - p_i) \quad (2.83)$$

Here we have however also a non-zero covariance

$$Cov[x_i x_j] = -np_i p_j \quad (2.84)$$

The negative value reflects the fact that if x_i is large (i.e. several items of type i are extracted), then is more likely to have fewer items j , since the total number of outcomes is fixed to N .

Multivariate gaussian. The most interesting case of multivariate PDF is the multivariate Gaussian. Let us consider the most general Gaussian of two variables x_1, x_2 (with zero mean for simplicity)

$$G(x_1, x_2) = N \exp\left[-\frac{1}{2(1-\rho^2)}\left(\frac{x_1^2}{\sigma_1^2} + \frac{x_2^2}{\sigma_2^2} - 2\frac{\rho x_1 x_2}{\sigma_1 \sigma_2}\right)\right] \quad (2.85)$$

where $N = 1/(2\pi\sigma_1\sigma_2\sqrt{1-\rho^2})$. The covariance matrix is

$$C_{ij} = \int x_i x_j f(x_1, x_2) dx_1 dx_2 = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \quad (2.86)$$

and so $\rho = C_{xy}/\sigma_x\sigma_y$ is indeed the correlation coefficient. For $\rho = 1$ the distribution is degenerate, i.e. $\det C = 0$.

This PDF can be written as

$$G(x_1, x_2) = N \exp\left[-\frac{1}{2}(X_i C_{ij}^{-1} X_j)\right] \quad (2.87)$$

where we defined the vector $X_i \equiv (x_i - \mu_i)$ and $N = 1/2\pi\sqrt{\det C}$. This can be immediately generalized to n variables:

$$G(x_i, i = 1\dots n) = \frac{1}{(2\pi)^{n/2}\sqrt{\det C}} \exp\left[-\frac{1}{2}(X_i C_{ij}^{-1} X_j)\right] \quad (2.88)$$

The contours of equiprobability $P = \text{const}$ are ellipsoids with principal axes oriented along the eigenvectors and semiaxes proportional to the square root of the eigenvalues of \mathbf{C} .

2.7 Random walks and Fokker-Planck equation

This section follows the treatment in U. Schwarz, *Lecture Notes in Theoretical Statistical Physics*.

We can apply now some of these tools to a physical system characterized by random movements of particles. Let's start in one dimension. Imagine a particle (or a person) doing a random walk, i.e. a number N of steps of size Δx in one dimension, of which $n_r = i$ with probability p to the right and $n_l = N - i$ with probability $q = 1 - p$ to the left. If each step takes a time Δt , one has

$$T = N\Delta t, \quad \text{total time} \quad (2.89)$$

$$X = m\Delta x, \quad \text{total distance} \quad (2.90)$$

where

$$m = n_r - n_l = 2i - N \quad (2.91)$$

We see that the only random variable is i and we expect it to follow a binomial distribution with parameters N, p . A fraction p of the N steps is therefore expected to be on the right, so:

$$\langle i \rangle = pN \quad (2.92)$$

with a variance

$$\langle (i - \langle i \rangle)^2 \rangle = \langle i^2 \rangle - \langle i \rangle^2 = Np(1-p) \quad (2.93)$$

so that

$$\langle i^2 \rangle = Np(1-p) + p^2 N^2 \quad (2.94)$$

Then the average distance after time T is

$$\langle X \rangle = \langle m\Delta x \rangle = \langle (2i - N)\Delta x \rangle = (2\langle i \rangle - N)\Delta x \quad (2.95)$$

$$= (2p - 1)N\Delta x = (2p - 1)\frac{\Delta x}{\Delta t} T \quad (2.96)$$

and we can define the average velocity

$$v \equiv \frac{\langle X \rangle}{T} = (2p - 1)\frac{\Delta x}{\Delta t} \quad (2.97)$$

We assume now the right or left steps are equiprobable, so $p = q = 1/2$ and $\langle X \rangle = 0$: as expected, in this symmetric random walk, the average position remains at the origin. The variance is then

$$\sigma^2 \equiv \langle X^2 \rangle = \langle (2i - N)^2 \rangle \Delta x^2 \quad (2.98)$$

$$= (4\langle i^2 \rangle - 4\langle i \rangle N + N^2) \Delta x^2 \quad (2.99)$$

$$= (4Np(1-p) + 4p^2N^2 - 4N^2p + N^2) \Delta x^2 \quad (2.100)$$

$$= 2(4pq \frac{\Delta x^2}{2\Delta t} T + ((2p-1) \frac{\Delta x}{\Delta t})^2 T^2) \quad (2.101)$$

For $p = q = 1/2$ we find then

$$\sigma^2 = 2DT \quad (2.102)$$

where we defined the diffusion coefficient

$$D = \frac{\Delta x^2}{2\Delta t} \quad (2.103)$$

It is important to remark that the dispersion σ grows with \sqrt{T} . If the motion were “ballistic”, i.e. a particle moving either to the right or to the left with constant speed, it would have obviously covered a distance that grows linearly with time. In a random walk, also called Brownian or diffusive motion, the particle would instead “explore” on average a distance $\sigma \propto T^{1/2}$, even maintaining its average position at the origin.

This can be generalized to d equivalent dimensions by performing the same operation on the position vector $\vec{R} = \sum_{i=1}^{i=d} \vec{r}_i$, where \vec{r}_i is the vector along direction i . As above, in the symmetric case, we have

$$\langle r_i \rangle = 0 \quad (2.104)$$

and therefore also $\langle \vec{R} \rangle = 0$. Assuming steps in a direction i are uncorrelated with steps in any other direction j , i.e.

$$\langle \vec{r}_i \vec{r}_j \rangle_{i \neq j} = 0 \quad (2.105)$$

one finds (using $\langle r^2 \rangle = \langle X^2 \rangle$ from the 1-dimensional case,

$$\sigma^2 = \langle \vec{R}^2 \rangle = \langle \sum r_i \cdot \sum r_j \rangle = \sum \langle r^2 \rangle \quad (2.106)$$

$$= 2dDT \quad (2.107)$$

(here the central dot is a scalar product). If we apply this to macroscopic (but still small) particles suspended freely in a fluid randomly hit by the fluid molecules, we see that the macroscopic particles will execute a random walk with Δx equal to the mean intermolecular distance and Δt the mean interval between two consecutive hits. One can expect that the diffusion coefficients, being proportional to a velocity squared times Δt , be also proportional to the temperature, and in fact it can be shown that $D = \mu k_B T$, with μ being a coefficient that depends on the viscosity of the fluid.

This approach can be expanded by deriving an equation for the time and space evolution of a probability distribution $P(x, t)$ that represent the probability of finding a particle at position x and time t in a system of many particles. We can think of $P dx dt$ as the fraction of particles that are in the interval $dx dt$. Then the fraction of particles at position x at the time $t + \Delta t$ must be the sum of the probability of being at the left position $x - \Delta x$ at the earlier epoch t weighted by the probability p of jumping to the right, plus the probability to be at the right position $x + \Delta x$ weighted by the probability $q = 1 - p$ of jumping to the left. We are clearly disregarding the probability of longer jumps, but they could be taken into account by summing over a probability distribution of jumps. This gives the following conservation equation

$$P(x, t + \Delta t) = pP(x - \Delta x, t) + qP(x + \Delta x, t) \quad (2.108)$$

(for now we assume p, q are constants). One could add the number of particle at x, t minus the particles that jump right or left from there, but that would exactly zero, because we are assuming at each step all particles jump for a given location to either left or right. Then we can expand around x, t for small time and space intervals as follows

$$P(x, t + \Delta t) = P + \dot{P} \Delta t + \dots \quad (2.109)$$

$$P(x \pm \Delta x, t) = P \pm P' \Delta x + \frac{1}{2} P'' \Delta x^2 + \dots \quad (2.110)$$

where a prime denotes $\partial/\partial x$ (whenever the arguments are not explicit, we mean they are x, t). Then, inserting into the conservation equation, we find

$$P + \dot{P}\Delta t = (p + q)P + (q - p)P'\Delta x + \frac{1}{2}P''(p + q)\Delta x^2 \quad (2.111)$$

Since $p + q = 1$, taking the continuum limit this becomes the Fokker-Planck equation

$$\dot{P} = -vP' + DP'' \quad (2.112)$$

where as before

$$v = (p - q)\frac{\Delta x}{\Delta t}, \quad \text{drift velocity} \quad (2.113)$$

$$D = \frac{\Delta x^2}{2\Delta t}, \quad \text{diffusion coefficient} \quad (2.114)$$

The Fokker-Planck equation is a partial derivative equation that governs the evolution of the distribution $P(x, t)$ in space and time, given an initial distribution. If we now assume that v, D also depend on space and time, we obtain the full form of the 1-dimensional Fokker-Planck equation:

$$\dot{P} = -(vP)' + (DP)'' \quad (2.115)$$

where now $v = v(x, t)$ and $D = D(x, t)$ (if they are constant, we fall back to (2.112)).

For a Dirac initial condition $P(x, t = 0) = \delta(x)$, i.e. all particles of the system are localized at the origin, the solution of Eq. (2.112) is a Gaussian

$$P(x, t) = \frac{1}{\sqrt{4\pi Dt}} \exp -\frac{(x - vt)^2}{4Dt} \quad (2.116)$$

from which we read the same 1-dimensional variance $\sigma^2 = 2Dt$ as derived above. In presence of both drift and diffusion, the distribution moves its peak and increases its variance, until the initial localized distribution spreads everywhere. The Fokker-Planck equation can be generalized to include the full phase-space distribution.

If we had taken also the second order Δt^2 term in (2.109), then we would have got a term $\ddot{P}\Delta t/2$ in Eq. (2.111), and we show now under which conditions this is negligible with respect to the $P''D$ term. For this to happen, the condition

$$\left| \frac{\ddot{P}}{P''} \right| \ll \left(\frac{\Delta x}{\Delta t} \right)^2 = \frac{2D}{\Delta t} \quad (2.117)$$

should be satisfied. From Eq. (2.111) we can calculate (we take now a symmetric problem, so $v = 0$)

$$\left| \frac{\ddot{P}}{P''} \right|_{x=0} = \frac{3}{2} \frac{D}{t} \quad (2.118)$$

so we see that the $\ddot{P}\Delta t^2$ term is negligible when $t \gg \Delta t$, i.e. after a long enough time (at least near the peak). Same condition is obtained comparing $\ddot{P}\Delta t$ to \dot{P} .

Chapter 3

Statistical mechanics

3.1 Statistical definition of entropy

1. Consider a thermodynamical system with $E, V, N = \text{const}$ (these state variables are denoted collectively as a *macrostate*).
2. The microstate of such a system is represented by the collection of all positions $\mathbf{q} = \vec{q}$ and momenta $\mathbf{p} = \vec{p}$ of every particle at a given time. There are therefore $6N$ such numbers. A microstate is then represented by a point moving in a $6N$ -dimensional phase space.
3. If we know the Hamiltonian of the system, $H = H(\mathbf{q}, \mathbf{p})$ (we assume H not to depend explicitly on time since we are discussing equilibrium states), then we can evolve the system through the Hamiltonian equations

$$\dot{\mathbf{q}} = \frac{\partial H}{\partial \mathbf{p}}, \quad \dot{\mathbf{p}} = -\frac{\partial H}{\partial \mathbf{q}} \quad (3.1)$$

and every function $A(t, \mathbf{q}, \mathbf{p})$ of \mathbf{q}, \mathbf{p} can be obtained by the standard evolution equation

$$\frac{dA}{dt} = \frac{\partial A}{\partial t} + \{A, H\} \quad (3.2)$$

where the Poisson brackets stand for

$$\{A, H\} = \sum_{\nu}^{3N} \left(\frac{\partial A}{\partial q_{\nu}} \frac{\partial H}{\partial p_{\nu}} - \frac{\partial A}{\partial p_{\nu}} \frac{\partial H}{\partial q_{\nu}} \right) \quad (3.3)$$

4. The volume of microstates of E in the shell $E, E + \Delta E$ can be evaluated in the continuum limit as

$$\Delta\omega = \int_{E \leq H \leq E + \Delta E} d^{3N}p d^{3N}q = \int_{E \leq H \leq E + \Delta E} d\omega \quad (3.4)$$

where $d\omega = d^{3N}p d^{3N}q$ is the infinitesimal phase-space cell.

5. The volume of microstates at exactly the value E corresponds to the (hyper)area

$$\sigma(E) = \int_{E=H} d\sigma \quad (3.5)$$

where $d\sigma$ is an infinitesimal area element on the surface $E = H(p, q)$.

6. Clearly in the classical continuum case, the number of microstates is infinite, since there can be a microstate for every value of \mathbf{q}, \mathbf{p} . A finite number can be obtained only if we define a small, but finite, cell, and we assume that values of \mathbf{p}, \mathbf{q} within that cell belong to the same microstate.

7. So we define the number Ω of microstates belonging to the same macrostate

$$\Omega(E, V, N) = \frac{\sigma(E, V, N)}{\sigma_0} = \frac{1}{\sigma_0} \int_{E=H, V=const, N=const} d\sigma \quad (3.6)$$

8. In order to evaluate Ω we first evaluate the volume $\omega(E)$ of microstates with energy less than E , and then we obtain

$$d\omega = \sigma(E)dE \quad (3.7)$$

so that

$$\sigma(E) = \left. \frac{\partial \omega}{\partial E} \right|_{V, N} \quad (3.8)$$

and finally

$$\Omega(E, V, N) = \frac{1}{\sigma_0} \frac{\partial \omega}{\partial E} \quad (3.9)$$

9. The fundamental postulates of statistical mechanics state that a) all microstates belonging to the same macrostate (*possible microstates*) are in turn reached by the system (ergodicity postulate); b) all microstates of equal energy are equiprobable (equiprobability postulate) and c) the equilibrium state is the one with the largest number of possible microstates belonging to the same macrostate (equilibrium postulate).
10. A more exact definition of ergodicity is: over long periods of time, the time spent by a system in some region of the phase space of microstates with the same energy is proportional to the volume of this region, i.e., all accessible microstates are equiprobable over a long period of time.
11. The collection of all possible microstates belonging to a given macrostate is called an *ensemble*.
12. Because of the ergodic hypothesis, the time a system spends in a particular microstate is proportional to the probability of that microstate. As a consequence, a measurement, i.e. an average over time, is equal to an average over all the possible microstates. The latter is called an *ensemble average*.
13. The ergodic postulate has not been demonstrated in full generality; it remains therefore a working hypothesis that may fail in some cases.
14. For non thermodynamical system, the ergodicity is easily broken: eg a pendulum only visits states along the initial plane (neglecting Earth's rotation...), although all other planes are equally energetic.
15. If we consider two non-interacting systems, each with a number Ω_i of possible microstates, the number of possible microstates of the combined system is clearly

$$\Omega(E, V, N) = \Omega_1(E_1, V_1, N_1)\Omega_2(E_2, V_2, N_2) \quad (3.10)$$

where $E = E_1 + E_2$ etc. Therefore, $\log \Omega$ is an extensive quantity.

16. We now define a function, called statistical entropy, that is extensive and depends on Ω alone. Therefore we require

$$S(\Omega) = S_1(\Omega_1) + S_2(\Omega_2) \quad (3.11)$$

$$\Omega = \Omega_1\Omega_2 \quad (3.12)$$

which implies immediately that $S \sim \log \Omega$, i.e.

$$S = k \log \Omega \quad (3.13)$$

We will show soon that k has to be identified with Boltzmann's constant.

17. The system reaches equilibrium when Ω (or equivalently $\log \Omega$) is maximized, i.e. when entropy is maximized. Since this is true also for the thermodynamic entropy already introduced in Sect. 1, we hypothesize that the two quantities are the same, up to a constant. Later on, we will show indeed that the statistical entropy so defined coincides with the thermodynamical entropy.

18. This is not really a full demonstration that $S = k \log \Omega$ (and other slightly different definitions are possible) since we just assumed that S depends only on Ω . In fact, there is no such a full proof and one might as well take this definition as a fundamental postulate of statistical mechanics. As we will see soon, it is however consistent with the thermodynamical definition.

3.2 Number of microstates

1. We can now evaluate the new definition of S for an ideal gas, where the Hamiltonian is

$$H = \sum \frac{p_\nu^2}{2m} \quad (3.14)$$

2. The volume of microstates is

$$\omega = \int_{H < E} d^{3N}q d^{3N}p = V^N \int_{H < E} d^{3N}p = V^N \int_{\sum p^2 < 2mE} d^{3N}p \quad (3.15)$$

ie the volume of the $3N$ -sphere of radius $R = \sqrt{2mE}$.

3. A standard calculation gives for the volume of a N -sphere of radius R

$$V_N(R) = \frac{\pi^{N/2}}{\frac{N}{2}\Gamma(\frac{N}{2})} R^N \quad (3.16)$$

where $\Gamma(n) = (n-1)!$ for integer n , and is the Gamma function $\Gamma(z) \equiv \int_0^\infty x^{z-1} e^{-x} dx$ for positive real arguments. A few useful values are $\Gamma(1) = 1, \Gamma(1/2) = \sqrt{\pi}, \Gamma(3/2) = \sqrt{\pi}/2$. Obviously, we recover the standard definitions for $N = 2, 3$.

4. Therefore

$$\omega = \frac{\pi^{3N/2}}{\frac{3N}{2}\Gamma(\frac{3N}{2})} (2mE)^{3N/2} V^N \quad (3.17)$$

and the number of microstates on the energy surface is

$$\Omega(E, V, N) = \frac{1}{\sigma_0} \frac{\partial \omega}{\partial E} = \frac{2m\pi^{3N/2}}{\sigma_0 \Gamma(\frac{3N}{2})} (2mE)^{3N/2-1} V^N \quad (3.18)$$

5. The entropy is then

$$S(E, V, N) = k \log \left(\frac{2m\pi^{3N/2}}{\sigma_0 \Gamma(\frac{3N}{2})} (2mE)^{3N/2-1} V^N \right) \quad (3.19)$$

Note that the argument of the log is indeed dimensionless, since $mE \sim p^2, V \sim q^3, \sigma_0 \sim p^{3N} q^{3N} / E \sim p^{3N-2} q^{3N} m$ (here \sim means “has the same dimension as”).

6. For $N \gg 1$ we can use the Stirling approximation

$$\log \Gamma(n) \approx n \log n - n \quad (3.20)$$

(notice that $\log \Gamma(n) \approx \log \Gamma(n+1) = n!$ in the limit $n \gg 1$) so (redefining $\sigma_0 \rightarrow (\sigma_0 E)^{1/N}$)

$$S(E, V, N) = Nk \left(\frac{3}{2} + \log \left(\frac{V}{\sigma_0} \right) \left(\frac{4\pi mE}{3N} \right)^{3/2} \right) \quad (3.21)$$

Just to have an idea, the Stirling formula approximates the exact result to better than 0.1% for $n > 400$.

7. From this expression we can immediately obtain the standard thermodynamical relations (the residual dependence of σ_0 on E is negligible for large N)

$$\frac{1}{T} = \left. \frac{\partial S}{\partial E} \right|_{V,N} = \frac{3}{2} \frac{Nk}{E} \quad (3.22)$$

$$\frac{p}{T} = \left. \frac{\partial S}{\partial V} \right|_{E,N} = \frac{Nk}{V} \quad (3.23)$$

which give the equations of state of ideal gases, $E = 3NkT/2$ and $pV = NkT$. This confirms that k is indeed Boltzmann constant.

8. The entropy so defined however is not an extensive quantity, since it depends on $\log V$ rather than on $\log V/N$ (Gibbs' paradox). This problem is corrected by noting that, from a quantum mechanical point of view, elementary particles are indistinguishable and therefore the counting of microstates Ω has to be divided by the number of permutations among N particles, i.e. by $N!$
9. Then we have

$$S(E, V, N) = Nk \left(\frac{3}{2} + \log \left(\frac{V}{\sigma_0} \right) \left(\frac{4\pi m E}{3N} \right)^{3/2} \right) - k \log N! \quad (3.24)$$

$$= Nk \left(\frac{5}{2} + \log \left(\frac{V}{N\sigma_0} \right) \left(\frac{4\pi m E}{3N} \right)^{3/2} \right) \quad (3.25)$$

(using Stirling's formula) which is now extensive. The dependence on E, V, N is the same as we have already seen in Eq. (1.48).

10. Let us now count the number of quantum mechanical microstates. Assuming a single particle in a box of size L one has the energy levels

$$\varepsilon_n = \frac{h^2}{8mL^2} (n_x^2 + n_y^2 + n_z^2) \quad (3.26)$$

with $n_{x,y,z} \geq 1$ are integer values, so the vector \vec{n} characterizes completely the quantum state.

11. If the total energy per particle is ε , the microstates are limited to a shell of radius $R = \sqrt{n_x^2 + n_y^2 + n_z^2} = (L/h)\sqrt{8m\varepsilon}$ in \vec{n} space.
12. For N particles we have an energy sphere of radius

$$E = \frac{h^2}{8mL^2} \sum_i^{3N} n_i^2 \quad (3.27)$$

The number of possible microstates are then the number of points on the surface of this sphere or, more exactly, on the energy shell of radius E and thickness δE , to be quantified below.

13. Notice that there is only a single ground microstate, ie $n_i = 1$ for every i . Since the ground state is reached when $T = 0$, this shows that the entropy vanishes when $T \rightarrow 0$. This is a general property of quantum systems and is called *Third Law of Thermodynamics*.
14. If we define the dimensionless energy

$$E^* = \frac{8mL^2}{h^2} E = \sum_i^{3N} n_i^2 \quad (3.28)$$

then we see that the volume of the $3N$ -sphere of radius E^* is equal to the number of grid points in the $n_{x,y,z}$ space

$$V_n = \frac{\pi^{3N/2}}{2 \Gamma(\frac{3N}{2})} (E^*)^{3N/2} \quad (3.29)$$

We need however to consider only positive values of $n_{x,y,z}$, i.e. take only the positive octant of this sphere. This means dividing by 2^{3N} . Then we obtain that the number of microstates inside E is

$$\Sigma = V_{n>0} = \frac{\pi^{3N/2}}{2^{3N} \frac{3N}{2} \Gamma(\frac{3N}{2})} (E^*)^{3N/2} \quad (3.30)$$

$$= \left(\frac{V}{h^3}\right)^N \frac{(2\pi m E)^{3N/2}}{\frac{3N}{2} \Gamma(\frac{3N}{2})} \quad (3.31)$$

($V = L^3$) which is identical to Eq. (3.17) if each phase-space cell is taken to be h^3 in size. One can say that h is the quantum of phase-space.

15. Now the number of microstates on the energy shell of radius $R = \sqrt{2mE}$ and thickness δR is

$$\Omega = \Sigma(R) - \Sigma(R - \delta R) = \left(\frac{V}{h^3}\right)^N \frac{(2\pi m)^{3N/2}}{\frac{3N}{2} \Gamma(\frac{3N}{2}) (2m)^{3N/2}} (R^{3N} - (R - \delta R)^{3N}) \quad (3.32)$$

$$= \left(\frac{V}{h^3}\right)^N \frac{(2\pi m)^{3N/2}}{\frac{3N}{2} \Gamma(\frac{3N}{2}) (2m)^{3N/2}} R^{3N} \left(1 - \left(1 - \frac{\delta R}{R}\right)^{3N}\right) \quad (3.33)$$

$$= \left(\frac{V}{h^3}\right)^N \frac{(2\pi m E)^{3N/2}}{\frac{3N}{2} \Gamma(\frac{3N}{2})} \left(1 - \left(1 - \frac{\delta R}{R}\right)^{3N}\right) \quad (3.34)$$

Then, since the scatter Δn_i of the quantum number n_i at the surface is around unity, we see that

$$\frac{\delta R}{R} = \frac{1}{2} \frac{\delta E^*}{E^*} = \sum_{\text{surface}} \frac{\Delta n_i}{n_i} \approx \frac{1}{3N} \quad (3.35)$$

and therefore for $N \rightarrow \infty$,

$$\left(1 - \frac{\delta R}{R}\right)^{3N} \approx \left(1 - \frac{1}{3N}\right)^{3N} \rightarrow e^{-1} \quad (3.36)$$

In fact, this correction is in general much smaller, because we can take a shell of states such that Δn_i is equal to a few units, rather than 1. Discarding this little correction, finally we obtain

$$\Omega = \left(\frac{V}{h^3}\right)^N \frac{(2\pi m E)^{3N/2}}{\frac{3N}{2} \Gamma(\frac{3N}{2})} \quad (3.37)$$

One should notice that δR disappears, and that the number of microstates at the energy surface is approximately equal to the number of microstates in the entire volume! That is, for large N , the microstates are essentially all on the energy surface.

16. We still need to include Gibbs' correction factor $N!$. So finally

$$\Omega = \left(\frac{V}{h^3}\right)^N \frac{(2\pi m E)^{3N/2}}{N! \left(\frac{3N}{2}\right)!} \quad (3.38)$$

(we used $\frac{3N}{2} \Gamma(\frac{3N}{2}) = \left(\frac{3N}{2}\right)!$) and we obtain the *Sackur-Tetrode formula*

$$S = Nk \left(\frac{5}{2} + \log \left[\frac{V}{N h^3} \left(\frac{4\pi m E}{3N} \right)^{3/2} \right] \right) \quad (3.39)$$

$$= Nk \left(\frac{5}{2} + \log \frac{v}{\lambda^3} \right) \quad (3.40)$$

(replacing $E = 3NkT/2$) where the constant σ_0 is now no longer arbitrary and where we have defined the volume per particle $v = V/N$ and the thermal wavelength

$$\lambda = \frac{h}{\sqrt{2\pi m k T}} \quad (3.41)$$

Notice that the energy in the Sackur-Tetrode formula is just the kinetic energy of the system. One can see that the dependence on E, V, N is the same as in the thermodynamic treatment (1.48), thereby confirming that the statistical definition of the entropy is the same as the thermodynamic one (a more general proof is given below). As we will see later on, this derivation is only valid for $v/\lambda^3 \gg 1$, i.e. large temperatures, which represents the classical limit.

17. From S one can derive all the thermodynamic equations of state. For instance, the chemical potential is

$$\mu = -T \left. \frac{\partial S}{\partial N} \right|_{E,V} = -kT \log \frac{v}{\lambda^3} \quad (3.42)$$

Pay attention to keeping E constant: one needs to use Eq. (3.39) rather than (3.40), because this second expression, where E has been replaced, is not in the natural variables U, V, N and some information is lost.

18. Since we approximated the number of microstates with a continuous function $\Omega(E)$, the limit $T \rightarrow 0$ we obtain, namely $S \rightarrow -\infty$, is not correct. Since as we noticed there is just a single microstate of lowest energy, the correct limit is $S(T \rightarrow 0) \rightarrow 0$, as we will see when a proper quantum-mechanical calculation is realized.

3.3 Microcanonical ensemble

1. We can now introduce an important quantity, the probability distribution of microstates, $\rho(\mathbf{q}, \mathbf{p})$, where \mathbf{q}, \mathbf{p} as usual define a representative point in the $6N$ phase-space. The fraction of microstates in the volume $d^{3N}p d^{3N}q$ is then defined to be

$$\rho(\mathbf{q}, \mathbf{p}) d^{3N}p d^{3N}q \quad (3.43)$$

and of course the distribution ρ is normalized to unity in the entire allowed (i.e., compatible with the macrostate) phase-space

$$\int \rho(\mathbf{q}, \mathbf{p}) d^{3N}p d^{3N}q = 1 \quad (3.44)$$

2. Any quantity $f(\mathbf{q}, \mathbf{p})$ has then an ensemble average

$$\langle f \rangle = \int f(\mathbf{p}, \mathbf{q}) \rho(\mathbf{q}, \mathbf{p}) d^{3N}p d^{3N}q \quad (3.45)$$

which by the ergodic hypothesis is equal to the time average, i.e. to the measurement of f over a macrostate. To every observable quantity one must then associate a function $f(\mathbf{q}, \mathbf{p})$.

3. If the macrostate has constant energy (and constant V, N), the fundamental postulate of statistical mechanics is that the microstates are equiprobable. Since there are Ω of them, we have

$$\rho = \rho_{mc} = \frac{1}{\Omega} \quad (3.46)$$

This is called the *microcanonical ensemble*.

4. This can be written as

$$\rho_{mc} = \frac{\delta(E - H(\mathbf{p}, \mathbf{q}))}{\omega(E)} \quad (3.47)$$

(δ is Dirac's delta and $\omega(E)$ the volume of available microstates, in this case the surface at $E = \text{const}$) so that the normalization condition is automatically fulfilled. In practice we assume a small shell ΔE of equiprobable states and put

$$\rho_{mc} = \frac{1}{\Omega(E)} \quad \text{for } E \leq H \leq E + \Delta E \quad (3.48)$$

and zero otherwise.

5. Since we should always consider phase-space cells of size h^3 for the correct counting of microstates, we can include this factor into the normalization and write always

$$\frac{d^{3N}p d^{3N}q}{h^{3N}} \quad (3.49)$$

in the averages. Gibbs' factor $N!$ should also be included for indistinguishable particles.

6. The distribution ρ , assumed independent of time because we are considering equilibrium states, is conserved along the trajectories of the system in phase space. In fact from Liouville's theorem (see 6.2) we have

$$\frac{d\rho}{dt} = \sum_i \{\rho, H\} = 0 \quad (3.50)$$

This guarantees that the fundamental assumption of equiprobability of microstates, if true at one moment, remains true at all times. It follows also that ρ commutes with H and therefore has to be a function of integrals of motion, e.g. H itself.

7. Imagine a closed region dV of phase-space delimited by a number of representative points (let's think of them as particles for a moment). This region will move in the phase space along with the representative points. Particles in this region cannot move out because otherwise they will cross other particle trajectories, but crossing is forbidden since a point in phase space defines the initial conditions, and to each initial condition there is (classically) only one trajectory. So since the density ρ is constant because of Liouville's theorem, and the number ρdV of representative particles in dV is also constant, it means the volume dV can be distorted along the flow but must remain of the same size.
8. Since every observable is an average of a function, also the entropy should be defined as the average of some quantity. We have then (omitting the brackets on the l.h.s.)

$$S(E, V, N) = \frac{k}{h^{3N}} \int d^{3N}p d^{3N}q \rho(\mathbf{q}, \mathbf{p}) \log \Omega = \frac{k}{h^{3N}} \int d^{3N}p d^{3N}q \rho_{mc} [-\log \rho_{mc}] \quad (3.51)$$

$$= -k \langle \log \rho_{mc} \rangle \quad (3.52)$$

where the last expression, since ρ_{mc} is constant, equals $k \log \Omega$, as it should. The definition of entropy, called Gibbs' entropy¹, as

$$S = -k \langle \log \rho \rangle \quad (3.53)$$

is indeed the most general one, as it applies to all possible ensembles.

9. In this form, the entropy coincides with the so called uncertainty function, or information entropy. Suppose we have a discrete probability distribution p_i . The information entropy is defined as

$$H(p_i) = - \sum_i p_i \log p_i \quad (3.54)$$

This function quantifies the amount of ignorance with respect to a given system. A state that is certain (ie, no ignorance, $p_i = 0$ or 1) does not contribute to it; it is maximized by equal probabilities if one does not have any further information and it always increases if more information is added to the system.

10. It is easy to show that the distribution p_i that maximizes the uncertainty $H(p_i)$ subject to the constraint $\sum p_i = 1$ is the microcanonical distribution. The same principle of maximum entropy allows to derive other ensemble distributions when other constraints are imposed.
11. As already mentioned, in quantum mechanics, the ground state of a system is reached for $T \rightarrow 0$. In all cases of physical interest, there is only one ground state, or a small number. The number of microstates corresponding to $T \rightarrow 0$ is therefore one or a very small number. Then $S(T \rightarrow 0) = 0$, or a number much smaller than S at any given finite temperature. We will assume then from now on that the reference state of entropy is the one at zero temperature for which $S = 0$ (*Third Law of Thermodynamics*).

¹Unfortunately, there is at least another quantities called Gibbs' entropy, defined as $S = k \log \Omega_{\leq E}$ where $\Omega_{\leq E}$ is the number of microstates of energy less than or equal to E (instead of our definition of microstates on the surface of constant E).

3.4 The canonical ensemble

1. Consider now a closed system S characterized by energy E_S embedded in a heat bath or reservoir R with energy E_R , both in equilibrium at temperature T . The heat bath is, as usual, supposed to be very large so its temperature does not change even if some heat is exchanged.
2. Since $E_R \gg E_S$, we have

$$\frac{E_S}{E} = 1 - \frac{E_R}{E} \ll 1 \quad (3.55)$$

where $E = E_S + E_R$. We ask for the probability of finding the system S in a state with energy E_i . The probability of S being in E_i must be the same as the probability of R being in $E - E_i$. This probability is then equal to the fraction of microstates of the heat bath for the energy $E - E_i$, or proportional to the number of microstates, since all microstates with *total* fixed energy E are equiprobable:

$$p_i \propto \Omega_R(E - E_i) \quad (3.56)$$

3. We can now expand $\log \Omega_R$ (or the entropy S) for $E_i \ll E$:

$$S = k \log \Omega_R(E - E_i) \approx k \log \Omega_R(E) - \frac{\partial}{\partial E} [k \log \Omega_R(E)] E_i + \dots \quad (3.57)$$

Now, since

$$\frac{\partial S_R}{\partial E} \Big|_{V,N} = \frac{k \partial \log \Omega_R}{\partial E} \Big|_{V,N} = \frac{1}{T} \quad (3.58)$$

(we are by construction at constant volume and number of particles) we find

$$\Omega_R(E - E_i) \approx \Omega_R(E) \exp\left(-\frac{E_i}{kT}\right) \quad (3.59)$$

4. The term we neglected in (3.57) is

$$\frac{1}{2} \frac{\partial^2 S}{\partial E^2} \Big|_R E_i^2 = \frac{E_i^2}{2} \left(\frac{\partial}{\partial E} \frac{1}{T} \right)_R \quad (3.60)$$

$$= -\frac{E_i^2}{2T^2} \left(\frac{\partial E}{\partial T} \right)_R^{-1} = -\frac{E_i^2}{2T^2 C_{V,R}} \quad (3.61)$$

where $C_{V,R}$ is the heat capacity of the reservoir. This term is negligible because C_V is an extensive quantity and becomes infinite when the heat bath become infinite.

5. Since $E = \text{const}$, this means that the probability that R is in $E - E_i$, and therefore that S is in E_i , is

$$p_i \propto \exp\left(-\frac{E_i}{kT}\right) \quad (3.62)$$

Properly normalized over all possible microstates, this means

$$p_i = \frac{\exp\left(-\frac{E_i}{kT}\right)}{\sum_i \exp\left(-\frac{E_i}{kT}\right)} \quad (3.63)$$

or, in continuous notation

$$p_c = \frac{\exp[-\beta H(\mathbf{p}, \mathbf{q})]}{h^{-3N} \int d^{3N} p d^{3N} q \exp[-\beta H(\mathbf{p}, \mathbf{q})]} \quad (3.64)$$

where $\beta = 1/kT$ and H is the Hamiltonian. The integral extends over all possible values of \mathbf{p}, \mathbf{q} . This is the *canonical phase-space density*, or distribution, or ensemble.

6. Any quantity can therefore be obtained as an average of some function. For instance the average energy (or average hamiltonian) is

$$U = \langle H \rangle = \frac{\sum_i E_i \exp(-\beta E_i)}{\sum_i \exp(-\beta E_i)} \quad (3.65)$$

3.5 Maximization of entropy

1. The canonical distribution can be obtained also as the distribution that maximizes the entropy given the constraints

$$\langle E \rangle = \int E p(\Gamma) d\Gamma, \quad 1 = \int p(\Gamma) d\Gamma \quad (3.66)$$

where for simplicity we define the integration over the phase space simply as $d\Gamma = h^{-3N} d^{3N} p d^{3N} q$. Then we have

$$S = -k \langle \log p \rangle = -k \int p \log p d\Gamma \quad (3.67)$$

To extremize S given the constraints, we use the Lagrange multipliers

$$S/k = \int p \log p d\Gamma - \beta \langle E \rangle - \int E p d\Gamma + A(1 - \int p d\Gamma) \quad (3.68)$$

and obtain by differentiating p

$$\delta S/k = \int [dp \log p + dp + \beta E dp - A dp] d\Gamma = 0 \quad (3.69)$$

which requires

$$dp(\log p + \beta E + 1 - A) = 0 \quad (3.70)$$

Since this must be true for every non-zero interval dp , we obtain the canonical distribution

$$p_c \propto e^{-\beta E} \quad (3.71)$$

2. In order to identify β we can proceed as follows. We define the *partition function*

$$Z = \sum_i \exp(-\beta E_i) \quad (3.72)$$

or

$$Z = \int d\Gamma \exp[-\beta H(\mathbf{p}, \mathbf{q})] \quad (3.73)$$

3. Then we get for the entropy

$$S = -k \langle \log p \rangle = k\beta \langle H \rangle + k \log Z = k\beta U + k \log Z \quad (3.74)$$

and therefore

$$\frac{1}{T} = \frac{\partial S}{\partial U} = kU \frac{\partial \beta}{\partial U} + k\beta + \frac{\partial}{\partial U} (k \log Z) \quad (3.75)$$

It follows

$$\beta = \frac{1}{kT} \quad (3.76)$$

as expected. This shows in a very clear way that the equilibrium distribution maximizes the entropy subject to some constraints.

4. Notice moreover that Eq. (3.74) can be written as

$$TS = U + kT \log Z \quad (3.77)$$

and therefore the free energy is

$$F = U - TS = -kT \log Z \quad (3.78)$$

The canonical ensemble can then be written also as

$$p_c = \frac{\exp(-\frac{E}{kT})}{Z} = \exp\left(\frac{F - E}{kT}\right) \quad (3.79)$$

5. The use of the Hamiltonian in Eq. (3.64) makes it clear that if $H(\mathbf{q}, \mathbf{p})$ is invariant with respect to the change of the individual coordinates p_i, q_i , then the number of indistinguishable microstates such that one (any) particle is at q_1, p_1 , another (any other) at q_2, p_2 etc. is $N!$. This means that the partition function, i.e. the sum over all possible distinguishable microstates, should be divided by $N!$.

3.6 The partition function

1. Since the canonical partition function gives $F(T, V, N)$, all thermodynamical quantities can be obtained from it. The relation between partition function and free energy is one of the most important results of statistical physics.
2. Directly from the partition function, we can obtain all the moments of the energy

$$\langle E \rangle = -\frac{\partial Z}{Z \partial \beta} \quad (3.80)$$

$$\langle E^2 \rangle = \frac{\partial^2 Z}{Z \partial \beta^2} \quad (3.81)$$

$$\langle E^n \rangle = (-1)^n \frac{\partial^n Z}{Z \partial \beta^n} \quad (3.82)$$

so we can say that Z acts as a moment generating function.

3. For an ideal gas,

$$Z = \frac{1}{N! h^{3N}} \int d^{3N} \mathbf{p} d^{3N} \mathbf{q} \exp[-\beta H(\mathbf{p}, \mathbf{q})] \quad (3.83)$$

$$= \frac{1}{N! h^{3N}} V^N \int d^{3N} \mathbf{p} \exp\left[-\beta \sum_{\nu} \frac{p_{\nu}^2}{2m}\right] \quad (3.84)$$

$$= \frac{1}{N! h^{3N}} V^N \prod_{\nu} \int d^3 p_{\nu} \exp\left[-\beta \frac{p_{\nu}^2}{2m}\right] \quad (3.85)$$

$$= \frac{V^N}{N!} \left(\frac{2\pi m k T}{h^2}\right)^{3N/2} = \frac{V^N}{N! \lambda^{3N}} \quad (3.86)$$

(\prod_i is the productory symbol) from which $F = -kT \log Z$ and we can obtain, for instance,

$$p = -\frac{\partial F}{\partial V}|_{T, N}, \quad S = -\frac{\partial F}{\partial T}|_{V, N} \quad (3.87)$$

from which we recover the Sackur-Tetrode entropy formula (approximating $\log N! \approx N \log N - N$). Notice that in the canonical ensemble, the phase-space integrals extend to infinity, since there is no longer any restriction on the energy.

4. The thermal wavelength

$$\lambda \equiv \frac{h}{(2\pi m k T)^{1/2}} \quad (3.88)$$

is up to order of unity constants equal to $h/(mE_{kin})^{1/2} \sim h/p$, i.e. it represents the average de Broglie wavelength for particles moving with momentum p at temperature T . As long as $\lambda \ll (V/N)^{1/3}$, i.e. a thermal wavelength much smaller than the typical interparticle distance, the quantum effects can be neglected. That is, the particle's wavefunctions do not overlap. Typical values of λ at room temperature are around 10^{-11} m. Typical mean interparticle distances in ordinary gases are of the order of $10^{-8} - 10^{-9}$ m, so much larger than λ .

5. For non-interacting identical particles, one obviously has $H(\mathbf{q}, \mathbf{p}) = \sum_i h_i(p_i, q_i)$ and

$$Z(T, V, N) = \frac{1}{N! h^{3N}} \int d^{3N} \mathbf{p} d^{3N} \mathbf{q} \exp[-\beta H(\mathbf{p}, \mathbf{q})] = \frac{Z(T, V, 1)^N}{N!} \quad (3.89)$$

and analogously ($Z_1 = Z(T, V, 1)$)

$$\rho_N = N! \frac{e^{-\beta h_1} e^{-\beta h_2} \dots}{Z_1^N} \quad (3.90)$$

while the probability of finding one particle at q, p is just

$$p_1 = \frac{e^{-\beta h(p,q)}}{Z_1} \quad (3.91)$$

as if it were a system of a single particle.

6. The observable quantities, as already mentioned, are the averages of some properly defined functions. For instance, the density of the i -th particle is

$$\rho_i(\vec{r}) = \langle \delta(\vec{r}_i - r) \rangle \quad (3.92)$$

the total density is

$$\rho(\vec{r}) = \left\langle \sum_i \delta(\vec{r}_i - r) \right\rangle \quad (3.93)$$

7. Since in the canonical distribution the energy is not fixed, we can evaluate its fluctuations, defined as the variance of the distribution. First we define the *density of energy states* $g(E)$ by calculating the number of microstates in the energy shell $E, E + dE$ as

$$d\Gamma = \frac{1}{h^{3N}} \int_{E \leq H \leq E+dE} d^{3N} \mathbf{p} d^{3N} \mathbf{q} = g(E) dE \quad (3.94)$$

so that $g(E)dE$ gives the number of microstates in the energy interval dE , and the associated probability of finding a system in the range $E, E + \Delta E$ is

$$p_c dE = Z^{-1} g(E) dE \exp(-\beta E) \quad (3.95)$$

The density of states increases with the energy, while the exponential decreases: this means that there should be a maximum somewhere. The maximum of the distribution ρ_c occurs at that E that solves the equation

$$\frac{\partial p_c}{\partial E} = \frac{1}{Z} \left(\frac{\partial g}{\partial E} - g\beta \right) e^{-\beta E} = 0 \quad (3.96)$$

and one can find that the maximum coincides with the constant value of the microcanonical ensemble. In fact, since $S = k \log \Omega = k \log[g(E)\Delta E]$, the condition becomes (assuming the shell ΔE is fixed independent of E)

$$\frac{k \partial \log g}{\partial E} \Big|_{E_{max}} = \frac{\partial S}{\partial E} \Big|_{E_{max}} = \frac{1}{T} \quad (3.97)$$

which is exactly the relation we should employ in the microcanonical ensemble to find the constant energy given the temperature. Moreover, this also coincides with the average

$$U \equiv \langle E \rangle = -\frac{\partial}{\partial \beta} \log Z(\beta) = F + TS \quad (3.98)$$

The variance is then

$$\begin{aligned} \sigma_E^2 &= \langle E^2 \rangle - \langle E \rangle^2 = \frac{\partial^2 Z}{Z \partial \beta^2} - \left(\frac{\partial Z}{Z \partial \beta} \right)^2 \\ &= \frac{\partial^2 \log Z}{\partial \beta^2} = -\frac{\partial U}{\partial \beta} = kT^2 \frac{\partial U}{\partial T} \Big|_{V,N} = kT^2 C_V \end{aligned} \quad (3.99)$$

This is an example of a fluctuation-dissipation relation, i.e. the relation among a fluctuation quantity, σ_E^2 , and a “dissipation”, i.e. the response of a system variable, U , to a change in an external control parameter, here T .

8. Since C_V and E are both extensive, the relative variance $\sigma/\langle E \rangle$ decreases as $1/\sqrt{N}$ for large N , as expected. Since the fluctuations around the mean energy E are extremely small, the canonical ensemble gives for all practical purposes the same thermodynamical relations as the microcanonical ensemble. On the other hand, the derivation of thermodynamics, i.e. of F or S , is much simpler for the canonical than for the microcanonical ensemble.

9. The heat capacity might become infinity during a phase transition: when water boils, the heat goes into the formation of bubbles while the temperature stays constant, so that $\partial T/\partial U = C_V^{-1} \rightarrow 0$. In this case, the fluctuations are no longer negligible and the system goes well out of equilibrium.
10. Since we now quantified the thickness of the energy shell, we can assume that $\Delta E/E \approx N^{-1/2}$ is the typical uncertainty of the energy of a classical system. Now, as we have seen, the number Ω of microstates in a shell of thickness ΔE and the number Σ in the entire volume from 0 to E are related by

$$\Omega = \frac{d\Sigma}{dE} \Delta E = \frac{Ed\Sigma}{dE} \frac{\Delta E}{E} = \frac{Ed\Sigma}{dE} \frac{1}{N^{1/2}} \quad (3.100)$$

If, as we have estimated, Σ increases like $E^{3N/2}$, we see that, writing $\Sigma = aE^{3N/2}$, one has

$$\Omega = \frac{3N}{2} aE^{3N/2} \frac{1}{N^{1/2}} = \frac{3N}{2} \Sigma \frac{1}{N^{1/2}} \quad (3.101)$$

and therefore

$$\log \Omega = \log \Sigma + \log \frac{3N^{1/2}}{2} \quad (3.102)$$

Clearly, since $\log \Sigma$ increases like N , the last term is negligible for N sufficiently large. Therefore, the formulation of the entropy of the microcanonical system in terms of the full volume of microstates ($S = k \log \Sigma$) or in terms of a thin energy shell ($S = k \log \Omega$), are practically equivalent. There are so many more states near the outer surface of the energy sphere that including also those inside the volume does not make much of a difference.

3.7 Maxwell-Boltzmann distribution

1. For an ideal gas

$$Z(T, V, 1) = V \lambda^{-3} \quad (3.103)$$

and

$$p_1 = \frac{e^{-\frac{p^2}{2mKT}}}{V h^{-3} (2\pi m k T)^{3/2}} \quad (3.104)$$

This is probability that a single particle is in the region $h^{-3} V dp^3$ of phase space. The probability of it being in $d^3 p$ is just $P = p_1 V h^{-3}$, and the probability of it being in $d^3 v$ can be obtained by the transformation relation $p(v) d^3 v = P(p) d^3 p$. Since $p = mv$ for a non-relativistic particle, we obtain

$$p(v) = m^3 P(p) = \left(\frac{m}{2\pi k T} \right)^{3/2} e^{-\frac{mv^2}{2kT}} \quad (3.105)$$

2. We have therefore also obtained the Maxwell-Boltzmann distribution of velocities for a single particle, correctly normalized to unity, as

$$p(\vec{v}) d^3 v = \left(\frac{m}{2\pi k T} \right)^{3/2} \exp\left(-\frac{mv^2}{2kT}\right) d^3 v \quad (3.106)$$

If one is only interested in the speed $v = |\vec{v}|$, one can write $d^3 v = v^2 \sin \theta d\theta d\phi dv$ and integrate out the angles, obtaining

$$p(v) dv = \left(\frac{m}{2\pi k T} \right)^{3/2} 4\pi v^2 \exp\left(-\frac{mv^2}{2kT}\right) dv \quad (3.107)$$

which has a maximum at $v_p = \sqrt{2kT/m}$ and a mean speed $\langle v \rangle = 2v_p/\sqrt{\pi}$ (notice that instead of course $\langle \vec{v} \rangle = 0$).

3. We can now find the density of energy states $g(E)$ by applying Eq. (3.94). We first recall the integral in Eq. (3.31), which gives the number of microstates in the an energy volume

$$\Sigma = \frac{1}{h^{3N}} \int_{E \leq H} d^{3N} \mathbf{p} d^{3N} \mathbf{q} = \left(\frac{V}{h^3} \right)^N \frac{(2\pi mE)^{3N/2}}{\frac{3N}{2} \Gamma(\frac{3N}{2})} \quad (3.108)$$

Then we obtain $g(E)$ as

$$g(E)dE = \frac{d\Sigma}{dE} dE \quad (3.109)$$

This gives

$$g(E) = \left(\frac{V}{h^3} \right)^N \frac{(2\pi mE)^{3N/2}}{\Gamma(\frac{3N}{2})E} \quad (3.110)$$

(the Gibbs' factor $1/N!$ should be included "by hand" for indistinguishable particles). Then if Γ represents the phase-space coordinates, we can write $p(\Gamma)d\Gamma = p(E)dE$, and since $d\Gamma = g(E)dE$ we have $p(E) = p(\Gamma)g(E)$.

4. Then we see that the canonical distribution can be written as

$$p(E) = \frac{g(E)e^{-\beta E}}{Z} = AE^{-1+3N/2}e^{-\beta E} \quad (3.111)$$

where A is an overall factor. This distribution peaks at $dp(E)/dE = 0$ i.e. for

$$E = \frac{3}{2}NkT \quad (3.112)$$

(for $N \gg 1$) as of course we should have expected. So although the canonical distribution in phase space has a maximum for $E = 0$, in the energy space the maximum is at the ideal gas energy. The average energy is

$$\langle E \rangle = \frac{\int_0^\infty E g(E) e^{-\beta E} dE}{\int_0^\infty g(E) e^{-\beta E} dE} = \frac{\Gamma(2 + \frac{3N}{2})}{\beta \Gamma(1 + \frac{3N}{2})} = \frac{(1 + \frac{3N}{2})!}{\beta (\frac{3N}{2})!} = (1 + \frac{3N}{2})kT \quad (3.113)$$

which again in the large N limit coincides with the expected result.

5. It is a simple exercise to show that

$$\langle E^m \rangle = \beta^{-m} \frac{\Gamma(1 + m + \frac{3N}{2})}{\Gamma(1 + \frac{3N}{2})} = \beta^{-m} \frac{(m + \frac{3N}{2})!}{\frac{3N}{2}!} \quad (3.114)$$

6. Then, we find the variance (one has to be careful in taking the $N \rightarrow \infty$ limit only at the end)

$$\sigma_E^2 = \langle E^2 \rangle - \langle E \rangle^2 = \frac{\frac{3N}{2}!(2 + \frac{3N}{2})! - [(1 + \frac{3N}{2})!]^2}{\beta^2 [(\frac{3N}{2})!]^2} \quad (3.115)$$

$$= \frac{\frac{3N}{2}!(2 + \frac{3N}{2})! - (1 + \frac{3N}{2})!(1 + \frac{3N}{2})(\frac{3N}{2})!}{\beta^2 [(\frac{3N}{2})!]^2} \quad (3.116)$$

$$= \frac{(2 + \frac{3N}{2})(1 + \frac{3N}{2}) - (1 + \frac{3N}{2})(1 + \frac{3N}{2})}{\beta^2} \quad (3.117)$$

$$= \frac{1}{\beta^2} (1 + \frac{3N}{2}) \rightarrow \frac{3N}{2} (kT)^2 \quad (3.118)$$

which coincides with Eq. (3.99) for $C_V = 3Nk/2$. So the use of the density of energy states allows one to perform several calculations by integrating over a single variable, E , rather than over the full $6N$ phase space coordinates.

3.8 Proof that the statistical entropy equals the thermodynamic entropy

We have now two definitions of entropy, a thermodynamic one Eq. (1.38) and a statistical one, Eq. (3.53). From each expression one can derive the same thermodynamic relations and therefore we expect they refer to the same physical quantity. However, one can have other definitions of statistical entropy that give by differentiation the same thermodynamic quantities. In this section we establish their identity in a general way.

Let us consider the Hamiltonian of a gas of interacting particles with potential $U(\vec{q})$ which depends only on the interparticle distance and is symmetric with respect to the exchange of any two particles. The partition function is then

$$Z = \frac{1}{N!h^{3N}} \int d^{3N}\mathbf{p} d^{3N}\mathbf{q} \exp \left[-\beta \sum_{\nu} \frac{p_{\nu}^2}{2m} - \beta U(\vec{q}) \right] \quad (3.119)$$

$$= \frac{1}{N!h^{3N}} \prod_{\nu} \int d^3 p_{\nu} \exp \left[-\beta \frac{p_{\nu}^2}{2m} \right] \int d^{3N}\mathbf{q} \exp [-\beta U(\vec{q})] \quad (3.120)$$

$$= \frac{1}{N!\lambda^{3N}} V \int d^3 q_2 \dots d^3 q_N \exp [-\beta U(\vec{q})] = \frac{1}{N!\lambda^{3N}} W \quad (3.121)$$

where $W = V \int d^3 q_2 \dots d^3 q_N \exp [-\beta U(\vec{q})]$ and in the last line we used the fact that one can rewrite U to be independent of one of the particle's coordinate, e.g. q_1 . The distribution function in the canonical ensemble is then

$$p = \frac{\exp \left[-\beta \sum_{\nu} \frac{p_{\nu}^2}{2m} - \beta U(\vec{q}) \right]}{Z} \quad (3.122)$$

and Gibbs' entropy is defined as (Gibbs' factor $N!$ plays no role here, so it can also be neglected everywhere)

$$S(V, T) = -k \int p \log p \frac{d^{3N}\mathbf{p} d^{3N}\mathbf{q}}{N!h^{3N}} \quad (3.123)$$

$$= k \int \frac{\exp \left[-\beta \sum_{\nu} \frac{p_{\nu}^2}{2m} - \beta U(\vec{q}) \right]}{Z} \left[\beta \sum_{\nu} \frac{p_{\nu}^2}{2m} + \beta U(\vec{q}) + \log W - \log(N!\lambda^{3N}) \right] \frac{d^{3N}\mathbf{p} d^{3N}\mathbf{q}}{N!h^{3N}} \quad (3.124)$$

$$= k\beta(\langle K \rangle + \langle U(\vec{q}) \rangle) + k \log W - k \log N! - 3kN \log \lambda \quad (3.125)$$

where we used the fact that the average of constants, here $\log W, \log N!, \log \lambda^{3N}$, are equal to the constants themselves. Now we consider a reversible transformation in T and V . We need now two relations. The first one is

$$\frac{\partial \log W}{\partial \beta} = -\langle U \rangle \quad (3.126)$$

This can be seen from the following equation:

$$\begin{aligned} \langle U \rangle &= \frac{1}{Z} \int \frac{d^{3N}\mathbf{p} d^{3N}\mathbf{q}}{N!h^{3N}} U \exp \left[-\beta \sum_{\nu} \frac{p_{\nu}^2}{2m} - \beta U(\vec{q}) \right] \\ &= \frac{1}{ZN!h^{3N}} \int d^{3N}\mathbf{p} \exp \left[-\beta \sum_{\nu} \frac{p_{\nu}^2}{2m} \right] \int d^{3N}\mathbf{q} U \exp [-\beta U(\vec{q})] \\ &= \frac{\int d^{3N}\mathbf{p} \exp \left[-\beta \sum_{\nu} \frac{p_{\nu}^2}{2m} \right] \int d^{3N}\mathbf{q} U \exp [-\beta U(\vec{q})]}{\frac{N!h^{3N}}{N!h^{3N}} \int d^{3N}\mathbf{p} \exp \left[-\beta \sum_{\nu} \frac{p_{\nu}^2}{2m} \right] \int d^{3N}\mathbf{q} \exp [-\beta U(\vec{q})]} \\ &= \frac{\int d^{3N}\mathbf{q} U \exp [-\beta U(\vec{q})]}{\int d^{3N}\mathbf{q} \exp [-\beta U(\vec{q})]} = -\frac{\partial W}{W \partial \beta} \end{aligned}$$

The second relation is

$$\frac{\partial \log W}{\partial V} = \beta \langle p \rangle \quad (3.127)$$

This can be obtained from the first of the Eqs. (3.87). In fact $F = -kT \log Z = -kT \log W + kT \log(N! \lambda^{3N})$, and since the volume V is only in W , we get Eq. (3.127).

Therefore, the change in S is

$$dS = \frac{\partial S}{\partial \beta} d\beta + \frac{\partial S}{\partial V} dV \quad (3.128)$$

$$= k\beta(d\langle K \rangle + d\langle U \rangle) + k(\langle K \rangle + \langle U \rangle)d\beta + k \frac{\partial \log W}{\partial \beta} d\beta + k \frac{\partial \log W}{\partial V} dV - 3kN \frac{d\lambda}{\lambda} \quad (3.129)$$

$$= \frac{d\langle K \rangle + d\langle U \rangle}{T} + k(\langle K \rangle + \langle U \rangle)d\beta - k\langle U \rangle d\beta + \frac{\langle p \rangle dV}{T} - \frac{3}{2} kN \frac{d\beta}{\beta} \quad (3.130)$$

$$= \frac{d\langle K \rangle + d\langle U \rangle}{T} + k\langle K \rangle d\beta + \frac{\langle p \rangle dV}{T} - \frac{\langle K \rangle}{T} \frac{d\beta}{\beta} \quad (3.131)$$

$$= \frac{d\langle K \rangle + d\langle U \rangle}{T} + k\langle K \rangle d\beta + \frac{\langle p \rangle dV}{T} - k\langle K \rangle d\beta \quad (3.132)$$

$$= \frac{d\langle K \rangle + d\langle U \rangle + \langle p \rangle dV}{T} = \frac{\delta Q}{T} \quad (3.133)$$

where we employed $\langle K \rangle = N \langle \frac{p_i^2}{2m} \rangle = \frac{3}{2} NkT$. We see finally that on the rhs we have the general expression for $\delta Q/T$, i.e. the thermodynamic definition of entropy. This completes our proof (for more detail, see Jaynes, *Gibbs vs Boltzmann Entropies*, American Journal of Physics, 391, 1965).

Historically, the entropy was introduced by Boltzmann as

$$S_B = -kN \int p_1 \log p_1 d^3 \mathbf{p} d^3 \mathbf{q} \quad (3.134)$$

where p_1 is the probability distribution for a single particle

$$p_1 = \frac{\lambda^3}{V} \exp \left[-\beta \sum_{\nu} \frac{p_{\nu}^2}{2m} \right] \quad (3.135)$$

It is not difficult to show that $S = S_B$ only for non-interacting particles. Gibbs' entropy is therefore more general.

3.9 Virial and equipartition theorems

1. Given a system characterized by a Hamiltonian $H(\mathbf{q}, \mathbf{p})$, we want to calculate now the mean value of $x_i \frac{\partial H}{\partial x_k}$ where x_i with $i = 1, \dots, 6N$ represents any one of the coordinates \mathbf{q}, \mathbf{p} . We have then

$$\langle x_i \frac{\partial H}{\partial x_k} \rangle = \frac{1}{h^{3N}} \int d^{6N} x \rho(\mathbf{x}) x_i \frac{\partial H}{\partial x_k} \quad (3.136)$$

where ρ can be either the micro- or the canonical distribution.

2. Let us begin with the canonical calculation. As it will be clear, it is much simpler than the microcanonical one. One has in fact

$$\langle x_i \frac{\partial H}{\partial x_k} \rangle = \frac{1}{Z h^{3N}} \int d^{6N} x e^{-\beta H} x_i \frac{\partial H}{\partial x_k} = \frac{1}{Z h^{3N}} \int d^{6N-1} x dx_k \frac{\partial H}{\partial x_k} e^{-\beta H} x_i \quad (3.137)$$

$$= \frac{1}{Z h^{3N}} \int d^{6N-1} x dx_k \frac{\partial e^{-\beta H}}{(-\beta) \partial x_k} x_i = \frac{1}{\beta Z h^{3N}} \int d^{6N-1} x dx_k e^{-\beta H} \frac{\partial x_i}{\partial x_k} \quad (3.138)$$

$$= \frac{\delta_{ik}}{\beta Z h^{3N}} \int d^{6N-1} x dx_k e^{-\beta H} = \frac{\delta_{ik}}{\beta Z h^{3N}} \int d^{6N} x e^{-\beta H} = \frac{\delta_{ik}}{\beta} = \delta_{ik} kT \quad (3.139)$$

where δ_{ij} is the Kronecker symbol (equal 0 for any $i \neq j$ and 1 otherwise), we employed integration by parts in the second line, and discarded the boundary terms assuming that H is infinity (so $e^{-\beta H} \rightarrow 0$) whenever a coordinate q reaches to the container wall (where the potential energy is infinite) or the momentum p goes to infinity.

3. We now do the same calculation in the microcanonical ensemble. Then we have $\rho = 1/\Omega$ in the energy shell and zero outside, so

$$\langle x_i \frac{\partial H}{\partial x_k} \rangle = \frac{1}{\Omega h^{3N}} \int_{E \leq H \leq E + \Delta E} d^{6N} x x_i \frac{\partial(H - E)}{\partial x_k} \quad (3.140)$$

(we used the fact that $\partial E / \partial x_i = 0$).

4. This can be evaluated as

$$\langle x_i \frac{\partial H}{\partial x_k} \rangle = \frac{1}{\Omega h^{3N}} \Delta E \frac{\partial}{\partial E} \int_{0 \leq H \leq E} d^{6N} x x_i \frac{\partial(H - E)}{\partial x_k} \quad (3.141)$$

$$= -\frac{1}{\Omega h^{3N}} \Delta E \frac{\partial}{\partial E} \int_{0 \leq H \leq E} d^{6N} x (H - E) \frac{\partial x_i}{\partial x_k} \quad (3.142)$$

where we integrate by parts with respect to x_k , and as before we neglected the boundary term because the variable x_k on the boundary evolves according to the equation $E = H(\mathbf{x})$ and therefore lies always on the energy surface, where $H - E = 0$.

5. Finally, since $\frac{\partial x_i}{\partial x_k} = \delta_{ik}$, we have

$$\langle x_i \frac{\partial H}{\partial x_k} \rangle = -\frac{\delta_{ik}}{\Omega h^{3N}} \Delta E \frac{\partial}{\partial E} \int_{0 \leq H \leq E} d^{6N} x (H - E) \quad (3.143)$$

$$= \frac{\delta_{ik}}{\Omega h^{3N}} \Delta E \frac{\partial}{\partial E} \int_{0 \leq H \leq E} d^{6N} x = \frac{\delta_{ik} \Sigma}{\Omega} \Delta E \quad (3.144)$$

(here one should include the differentiation also of the integral limit, but it turns out to be zero), where $\Sigma = h^{-3N} \int_{0 \leq H \leq E} d^{6N} x$ counts the number of microstates in the energy sphere $0 \leq H \leq E$.

6. Now Ω is the number of microstates in the energy shell, and therefore

$$\Omega = \frac{\partial \Sigma}{\partial E} \Delta E \quad (3.145)$$

so finally

$$\langle x_i \frac{\partial H}{\partial x_k} \rangle = \frac{\delta_{ik}}{\frac{\partial \log \Sigma}{\partial E}} \quad (3.146)$$

7. Now we employ again the approximation we used before, namely that

$$\log \Sigma \approx \log \Omega \quad (3.147)$$

ie that the number of microstates in the energy shell of thickness $\frac{\Delta E^*}{E^*} \approx 1$ is almost equal to the entire number of microstate in the energy sphere; that is, the “skin” of the energy sphere contains the bulk of the microstates. (Alternatively, one could have defined $S = k \log \Sigma$ from the start, as actually proposed by Gibbs.)

8. This gives the *equipartition theorem*

$$\langle x_i \frac{\partial H}{\partial x_k} \rangle = k \frac{\delta_{ik}}{\frac{\partial S}{\partial E}|_{N,V}} = \delta_{ik} kT \quad (3.148)$$

That is, every term $x_i \partial H / \partial x_k$ in the Hamiltonian accounts for a kT in a thermodynamic system.

9. If x_i is a coordinate or a momentum, this means respectively

$$\langle x_i \frac{\partial H}{\partial x_i} \rangle = -\langle q_i \dot{p}_i \rangle = -\langle q_i F_i \rangle = kT \quad (3.149)$$

$$\langle x_i \frac{\partial H}{\partial x_i} \rangle = \langle p_i \dot{q}_i \rangle = kT \quad (3.150)$$

10. In the second case, we can also recognize $\langle p_i \dot{q}_i \rangle = 2\langle K_{q_i} \rangle$, i.e. twice the kinetic energy in the direction q_i , so we have the mean kinetic energy for a particle i (three degrees of freedom)

$$\langle K_i \rangle = \frac{3}{2}kT \quad (3.151)$$

11. Similarly from (3.149), and summing over the three directions, we can write for the i -th particle

$$-\langle \vec{q} \cdot \vec{F} \rangle_i = \langle \vec{q} \cdot \vec{\nabla} V \rangle_i = 3kT \quad (3.152)$$

and assuming a power law potential $V \propto r^\alpha$ one has

$$\langle \vec{r} \cdot \vec{\nabla} V \rangle_i = \langle r \frac{\partial V}{\partial r} \rangle_i = \alpha \langle V \rangle_i \quad (3.153)$$

(where $\vec{\nabla} V(r) = \frac{dV}{dr} \frac{\vec{r}}{r}$ and $r = |\vec{r}|$) so that, for three degrees of freedom of a single particle,

$$\alpha \langle V \rangle + i = -\langle \vec{q} \cdot \vec{F} \rangle_i = 3kT \quad (3.154)$$

and for N particles we obtain the *virial theorem*

$$\langle K \rangle = \frac{\alpha}{2} \langle P \rangle = \frac{3}{2}NkT \quad (3.155)$$

where $P = NV_i$. For quadratic potential, $\alpha = 2$, the kinetic energy, $K = NK_i$, and potential energy, $P = NV_i$, are equal on average.

12. For any quadratic Hamiltonian

$$H = \sum_{\nu}^{3N} (A_{\nu} p_{\nu}^2 + B_{\nu} q_{\nu}^2) \quad (3.156)$$

one has then

$$\langle H \rangle = \frac{1}{2}fkT \quad (3.157)$$

where $f = 6N$ is the number of degrees of freedom, regardless of A_{ν}, B_{ν} . So the equipartition theorem applied to quadratic Hamiltonians says that the energy associated with each degree of freedom at temperature T is $kT/2$. Therefore, we also find that the heat capacity in the classical limit is $C_V \rightarrow \frac{1}{2}k \times$ d.o.f..

13. The equipartition theorem applied to an ideal gas gives as expected the equation of state $pV = NkT$.
14. The use of integrals instead of sums over discrete states is crucial for this derivation of the equipartition theorem, as we will see in the next section.

3.10 Applications of the canonical ensemble: quantum oscillators

1. The canonical ensemble is also called Boltzmann statistics when applied to classical systems.
2. Even quantum systems can be discussed with the Boltzmann statistics under some limitations.
3. Let us consider a system of fixed quantized harmonic oscillators. The energy spectrum for every oscillator of frequency ω is

$$\varepsilon_n = \hbar\omega(n + \frac{1}{2}) \quad (3.158)$$

with $n = 0, 1, 2, \dots$. The quantum number n takes now the place of the phase-space variables p, q .

4. Assuming the particles are non-interacting, we have for N distinguishable oscillators (eg atoms fixed in a lattice)

$$Z(T, V, N) = [Z(T, V, 1)]^N \quad (3.159)$$

where the one-particle partition sum is

$$\begin{aligned} Z(T, V, 1) &= \sum_n \exp -\beta \varepsilon_n = e^{-\frac{\beta \hbar \omega}{2}} \sum_n (e^{-\beta \hbar \omega})^n \\ &= \frac{e^{-\frac{\beta \hbar \omega}{2}}}{1 - e^{-\beta \hbar \omega}} = [2 \sinh(\frac{\beta \hbar \omega}{2})]^{-1} \end{aligned}$$

where we used the sum rule of geometric series

$$\sum_{n=0}^{\infty} r^n = \frac{1}{1-r} \quad (3.160)$$

if $r < 1$. Notice also the useful rule obtained by differentiation

$$\sum_{n=0}^{\infty} n r^n = \frac{r}{(1-r)^2} \quad (3.161)$$

and the generalization for $N > 0$ (binomial expansion)

$$\frac{1}{(1-x)^N} = \sum_{\ell=0}^{\infty} \binom{N+\ell-1}{\ell} x^\ell \quad (3.162)$$

5. Finally

$$Z(T, V, N) = [2 \sinh(\frac{\beta \hbar \omega}{2})]^{-N} \quad (3.163)$$

and the free energy is

$$F = -kT \log Z = NkT \log[2 \sinh(\frac{\beta \hbar \omega}{2})] = \frac{N}{2} \hbar \omega + NkT \log(1 - e^{-\beta \hbar \omega}) \quad (3.164)$$

where the first term represents the zero-point energy.

6. From F , all other thermodynamic relations follow. For instance

$$\mu = \frac{\partial F}{\partial N}|_{T,V} = \frac{F}{N} \quad (3.165)$$

$$p = -\frac{\partial F}{\partial V}|_{N,V} = 0 \quad (3.166)$$

(no pressure because the oscillators are fixed at their location).

7. The entropy and average energy are

$$S = -\frac{\partial F}{\partial T}|_{V,N} = Nk \left[\frac{\beta \hbar \omega}{e^{\beta \hbar \omega} - 1} - \log(1 - e^{-\beta \hbar \omega}) \right] \quad (3.167)$$

$$U = F + TS = N\hbar\omega \left[\frac{1}{2} + \frac{1}{\exp(\beta \hbar \omega) - 1} \right] \quad (3.168)$$

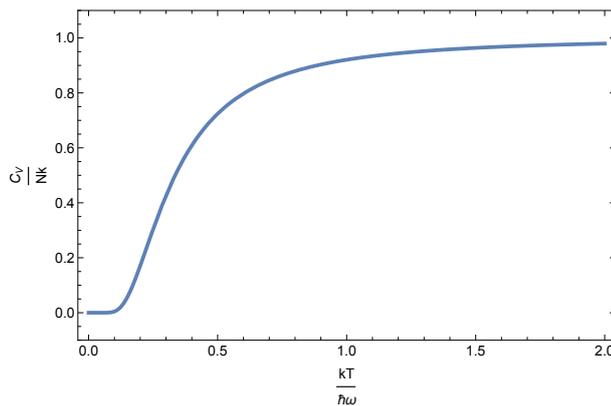
Notice that the entropy vanishes for $\beta \hbar \omega \rightarrow \infty$ i.e. at zero temperature, as required by the Third Law. The energy can be written also as

$$U = N \langle \varepsilon_n \rangle \quad (3.169)$$

where $\langle \varepsilon_n \rangle$ is the mean energy of an oscillator and is

$$\langle \varepsilon_n \rangle = \hbar \omega \left[\frac{1}{2} + \frac{1}{\exp(\beta \hbar \omega) - 1} \right] \quad (3.170)$$

which can of course also be calculated directly as $\langle \varepsilon_n \rangle = -\frac{\partial \log Z_1}{\partial \beta}$.

Figure 3.1: Plot of C_V/Nk as a function of temperature for the quantum oscillators.

8. Notice that the classical equipartition theorem $U = NkT$ is recovered only in the limit of $T \rightarrow \infty$, i.e. when the discrete sum can be approximated as an integral. In the opposite limit $T \rightarrow 0$ one has instead $U = N\hbar\omega/2$, ie the sum of all the zero-point energies.

9. Similarly, one can also evaluate the mean quantum number $\langle n \rangle$

$$\langle n \rangle = \frac{\langle \varepsilon_n \rangle - \frac{\hbar\omega}{2}}{\hbar\omega} = \frac{1}{\exp(\beta\hbar\omega) - 1} \quad (3.171)$$

10. The heat capacity is now

$$C_V = \frac{\partial U}{\partial T}|_{V,N} = Nk(\beta\hbar\omega)^2 \frac{\exp(\beta\hbar\omega)}{(\exp(\beta\hbar\omega) - 1)^2} \quad (3.172)$$

and we see that for $T \rightarrow 0$, the capacity approaches zero. That is, for very small temperatures, the system cannot absorb any heat, ie no oscillator gets excited. In the opposite limit of large temperatures, $C_V = Nk$, as we expected from the virial theorem for N particles with two degrees of freedom each (kinetic and potential).

11. We use now again the concept of density of energy states. Consider the probability of being in microstates of a given energy E , each one occurring with the same probability p_c . If there are $g(E)$ of them, the probability will be the sum of all the p 's:

$$p(E) = p_c(E)g(E) \quad (3.173)$$

$$p(E)dE = p_c(E)g(E)dE \quad (3.174)$$

where the second line applies in the continuous limit, so $g(E)dE$ is the number of microstates in dE . So instead of considering separately the probability of each microstate, we can consider the weighted probability $p_c(E)g(E)$. The function $g(E)$ is called *density of energy states*. Accordingly, the partition function becomes $\sum g(E)p(E)$, where *the sum now is over the energy levels and not over the microstates*.

12. We can now obtain the density of states $g(E)$ for our model. With the use of the binomial expansion (3.162) we can rewrite the partition function as

$$Z(T, V, N) = [2 \sinh(\frac{\beta\hbar\omega}{2})]^{-N} = \left[\frac{\exp(-\frac{1}{2}\beta\hbar\omega)}{1 - \exp(-\beta\hbar\omega)} \right]^N \quad (3.175)$$

$$= \sum_{\ell=0}^{\infty} \binom{\ell + N - 1}{\ell} \exp[-\beta\hbar\omega(\frac{1}{2}N + \ell)] \quad (3.176)$$

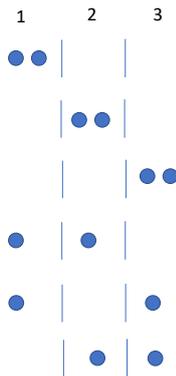


Figure 3.2: There are six ways to distribute $\ell = 2$ quanta of energy in $N = 3$ particles (oscillators). This corresponds to $\binom{\ell + N - 1}{\ell}$, since there are $N - 1$ bars and ℓ balls.

This expression can be compared to the general expression that defines the density of states $g(E)$ for discrete states

$$Z(T, V, N) = \sum_i g(E_i) e^{-\beta E_i} \quad (3.177)$$

to obtain

$$g_\ell \equiv g(E_\ell) = \binom{\ell + N - 1}{\ell} \quad (3.178)$$

$$E_\ell \equiv \hbar\omega\left(\ell + \frac{N}{2}\right) \quad (3.179)$$

13. The energy states E_ℓ can be interpreted as the energy of ℓ quanta of energy $\hbar\omega$ each, plus the zero-point energy of N oscillators. The density of states for discrete states is called degeneracy factor or multiplicity, g_ℓ . It corresponds to the number of ways one can distribute ℓ indistinguishable quanta of energy among N different (distinguishable) possible states (i.e., how many states with energy ℓ among N oscillators). See Fig. 3.2 for an example.
14. It is possible now to show that the entropy $S = k \log \Omega$ coincides with $S = k \log g_\ell$, since g_ℓ gives the number of microstates at any given energy.

3.11 Applications of the canonical ensemble: paramagnetism

1. If a system of infinitesimal magnetic dipoles (i.e., particles with spin) with magnetic moment $\vec{\mu}_i$ is under the action of an external homogeneous magnetic field \vec{H} , the energy of the system is

$$E = - \sum_i^N \vec{\mu}_i \cdot \vec{H} \quad (3.180)$$

The minus sign implies that the energy is minimized when the spins align with the external field. That is, it takes some work to disalign the spins.

2. Neglecting the translational degrees of freedom (dipole fixed in a lattice), the dipole can only change their orientation with respect to an axis z that is chosen to be along \vec{H} . The degrees of freedom are then the

two angle θ, ϕ for each dipole. The Hamiltonian (with a hat to distinguish from the magnetic field) is then

$$\hat{H} = - \sum_i^N \vec{\mu}_i \cdot \vec{H} = - \sum_i^N \mu_i H \cos \theta_i \quad (3.181)$$

3. We are also neglecting the spin-spin interactions. These will be studied in Sect. (5.4).

4. If the dipoles are all identical, then the partition function of this system is

$$Z(T, H, N) = \int d\Omega_1 \int d\Omega_2 \dots \int d\Omega_N \exp \left(\beta \mu H \sum_i^N \cos \theta_i \right) \quad (3.182)$$

where $d\Omega_i = \sin \theta_i d\theta_i d\phi_i$ and where one see that H takes the role of the volume here.

5. Since we are neglecting any interaction among the dipoles, we have

$$Z(T, H, N) = Z(T, H, 1)^N \quad (3.183)$$

$$Z(T, H, 1) = \int \sin \theta d\theta d\phi \exp(\beta \mu H \cos \theta) \quad (3.184)$$

$$= 2\pi \int_{-1}^{+1} dx \exp(\beta \mu H x) \quad (3.185)$$

$$= 4\pi \frac{\sinh(\beta \mu H)}{\beta \mu H} \quad (3.186)$$

6. Then the probability for a dipole to be oriented in $\theta + d\theta, \phi + d\phi$ is

$$\rho(\theta, \phi) d\Omega = \frac{\exp(\beta \mu H \cos \theta)}{Z(T, H, 1)} \sin \theta d\theta d\phi \quad (3.187)$$

7. The mean magnetic dipole can then be evaluated as the mean of the vector $\vec{\mu} = \mu \{ \sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta \}$. One finds $\langle \mu_x \rangle = \langle \mu_y \rangle = 0$ and

$$\langle \mu_z \rangle = \frac{\mu}{Z_1} \int \cos \theta \exp[\beta \mu H \cos \theta] \sin \theta d\theta d\phi \quad (3.188)$$

$$= \mu \left[\coth(\beta \mu H) - \frac{1}{\beta \mu H} \right] \quad (3.189)$$

One can also write

$$\langle \mu_z \rangle = \frac{1}{\beta} \frac{\partial}{\partial H} \log Z(T, H, 1) = - \frac{\partial}{\partial H} F(T, H, 1) \quad (3.190)$$

where $F(T, H, 1) = -kT \log Z(T, H, 1) = F(T, H, N)/N$.

8. The total dipole $\langle D_z \rangle = N \langle \mu_z \rangle$ in the direction z is then

$$\langle D_z \rangle = - \frac{\partial}{\partial H} F(T, H, N) \quad (3.191)$$

This shows that here the averages of μ_z and H (or better NH , so to obtain an extensive quantity as V) take the role of p, V in the context of gases. Notice that

$$\langle D_z \rangle = N \langle \mu_z \rangle = - \frac{\langle E \rangle}{H} \quad (3.192)$$

9. Explicitly, we find

$$F(T, H, N) = -NkT \log \frac{4\pi \sinh(\beta \mu H)}{\beta \mu H} \quad (3.193)$$

$$\langle D_z \rangle = N\mu \left[\coth(\beta \mu H) - \frac{1}{\beta \mu H} \right] \quad (3.194)$$

The behavior of $\langle D_z \rangle / N\mu$ as a function of $x = \beta\mu H$ is a curve that rises from 0 for zero magnetic field H or large temperatures (no net alignment) to a saturation value 1 (full alignment) for large fields or small temperatures. This behavior (no net magnetization for $H = 0$) is called *paramagnetic*.

10. The entropy can be derived as $S = -\partial F / \partial T|_{H,N}$, and then the magnetization energy is

$$U = F + TS = N(kT - \mu H \coth(\frac{\mu H}{kT})) \quad (3.195)$$

For $T \rightarrow \infty$, $\coth(\mu H/kT) \rightarrow kT/\mu H$ and therefore $U \rightarrow 0$. That is, if the temperature is very large, the dipoles will move randomly in all directions, and the average projection on the \vec{H} axis tends to zero. For small temperatures, instead, one gets $U \rightarrow -N\mu H$, since the dipoles become all aligned with \vec{H} .

11. Finally, one can get the heat capacity

$$C_H = \frac{\partial U}{\partial T}|_{H,N} = Nk \left(1 - \frac{x^2}{\sinh^2 x} \right) \quad (3.196)$$

where $x = \beta\mu H$. One sees then that $C_H \rightarrow 0$ for large temperatures (see Fig. 3.4). This is typical for systems with an energy upper bound. That is, once the system is completely randomized at high temperature, it cannot absorb any more energy. Moreover, $C_H \rightarrow Nk$ for small temperatures. Of course, this takes into account only the magnetization energy: the system can still absorb heat in the other degrees of freedom (rotation, vibrations, etc) that we neglect here.

12. For quantum systems, the spin states are quantized. The simplest case is when each dipole can only assume two states, up or down, with energies $E = -g\mu_B H\sigma$ and $\sigma = \pm 1/2$ (here g is the gyromagnetic ratio, equal to 2 for the intrinsic spin of electrons and $\mu_B = e\hbar/2m_e c$ is Bohr's magneton). Then the one-particle partition function is

$$Z(T, H, 1) = \sum_{\sigma=-1/2}^{\sigma=+1/2} \exp(-\beta E) = 2 \cosh(\beta\varepsilon) \quad (3.197)$$

with $\varepsilon = \mu_B H$ for electrons.

13. The free energy is then

$$F(T, H, N) = -NkT \log[2 \cosh(\beta\varepsilon)] \quad (3.198)$$

from which again all thermodynamic quantities can be obtained. In particular, the average magnetization is

$$\langle D_z \rangle = -\frac{\partial F}{\partial H}|_{T,N} = N\mu_B \tanh \beta\varepsilon \quad (3.199)$$

(see Fig. 3.3) and the average spin per particle is

$$\sigma \equiv \langle D_z \rangle / N\mu_B = \tanh \beta\varepsilon \quad (3.200)$$

14. As usual, from the entropy $S = -\partial F / \partial T|_{H,N}$ we obtain the energy

$$U = F + TS = -N\mu_B H \tanh(\beta\varepsilon) \quad (3.201)$$

and the heat capacity

$$C_H = \frac{\partial U}{\partial T} = Nk(\beta\varepsilon)^2 \cosh^{-2}(\beta\varepsilon) \quad (3.202)$$

For small and large $x = \beta\varepsilon$ the heat capacity can be approximated as $C_H \sim x^2 e^{-2x}$. So now one sees again that $C_H \rightarrow 0$ for high temperature, but also for small temperatures, since the dipoles cannot absorb arbitrarily small amount of energy and if they are totally aligned, they remain so. The heat capacity has therefore a maximum at approximately $\varepsilon \approx kT$. This behavior (see Fig. 3.5) is called *Schottky heat capacity*.

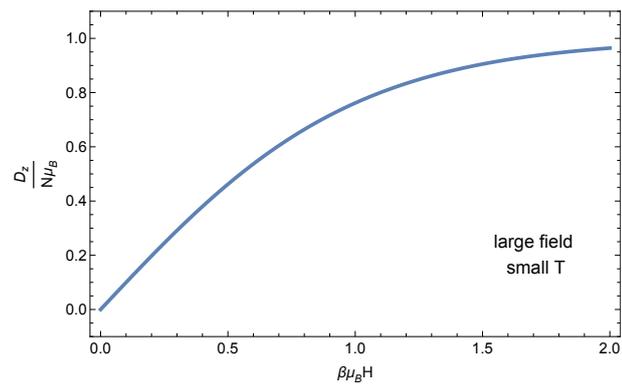


Figure 3.3: Plot of $D_z/N\mu$ as a function of temperature for a paramagnet.

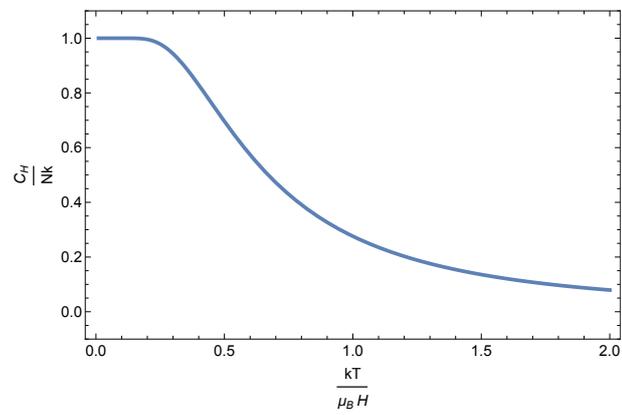


Figure 3.4: Plot of C_H/Nk as a function of temperature for a classical paramagnet.

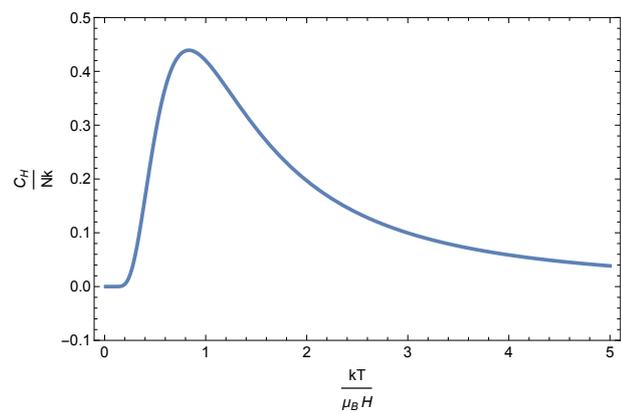


Figure 3.5: Plot of C_H/Nk as a function of temperature for a quantum paramagnet.

3.12 The macrocanonical ensemble

1. If we repeat the steps of Sect. 3.4 to a subsystem S that can exchange both energy and particles with the environment R , we find the probability for S to be in a state with energy $E_i \ll E$ and particles $N_S \ll N$ to be

$$p_{i,N} \propto \Omega_R(E - E_i, N - N_S) \approx \Omega_R(E, N) \left[1 - \frac{\partial \log \Omega_R}{\partial E} E_i - \frac{\partial \log \Omega_R}{\partial N} N_S + \dots \right] \quad (3.203)$$

and

$$\log p_s \propto \log \Omega_R(E - E_i, N - N_S) \approx \log \Omega_R(E, N) - \frac{\partial \log \Omega_R}{\partial E} E_i - \frac{\partial \log \Omega_R}{\partial N} N_S + \dots$$

2. Therefore since $S_R = k \log \Omega_R$ and

$$\frac{\partial S_R}{\partial E} \Big|_{V,N} = \frac{1}{T}, \quad \frac{\partial S_R}{\partial N} \Big|_{V,E} = -\frac{\mu}{T}, \quad (3.204)$$

we obtain

$$p_s \propto \Omega_R(E, N) \exp \left[-\frac{\partial \log \Omega_R}{\partial E} E_i - \frac{\partial \log \Omega_R}{\partial N} N_S \right] \quad (3.205)$$

$$= \Omega_R(E, N) \exp \left(-\frac{E_i}{kT} + \frac{\mu N_S}{kT} \right) \quad (3.206)$$

The prefactor can of course be absorbed into the normalization.

3. The macrocanonical (or grand-canonical) distribution is then

$$p_{i,N} = \frac{\exp(-\beta(E_i - \mu N))}{\sum_N \sum_i \exp(-\beta(E_i - \mu N))} \quad (3.207)$$

4. As before, this distribution can also be obtained as the distribution that maximizes the uncertainty $\sum p \log p$ when the constraints on the average energy and the average number of particles are enforced (see later).

5. A relation between the macrocanonical partition function

$$\mathcal{Z} = \sum_N \sum_i \exp(-\beta(E_i - \mu N)) \quad (3.208)$$

and the grand potential Φ also exists, analogously to the relation between free energy and the canonical ensemble. The grand potential is

$$\phi = U - TS - \mu N \quad (3.209)$$

from which

$$d\Phi = -SdT - pdV - Nd\mu \quad (3.210)$$

which shows that $\Phi = \Phi(T, V, \mu)$. It follows that from Φ we can obtain all thermodynamical quantities:

$$\frac{\partial \Phi}{\partial T} = -S, \quad \frac{\partial \Phi}{\partial V} = -p, \quad \frac{\partial \Phi}{\partial \mu} = -N \quad (3.211)$$

Now we have

$$S = -k \langle \log p \rangle = -\frac{1}{\beta T} \langle -\beta(E - \mu N) \rangle + \frac{\log \mathcal{Z}}{\beta T} \quad (3.212)$$

$$= \frac{\langle E \rangle - \mu \langle N \rangle}{T} + \frac{\log \mathcal{Z}}{\beta T} \quad (3.213)$$

and therefore

$$\Phi = -TS + U - \mu N = -kT \log \mathcal{Z} \quad (3.214)$$

Given \mathcal{Z} , therefore, one can get all thermodynamical quantities.

6. One can also write the macrocanonical partition sum as a weighted sum of the canonical partition function

$$\mathcal{Z}(T, V, \mu) = \sum_N \left(e^{\frac{\mu}{kT}} \right)^N Z(T, V, N) \quad (3.215)$$

The factor $z = e^{\frac{\mu}{kT}}$ is called *fugacity*.

7. For noninteracting, indistinguishable systems one has

$$\mathcal{Z}(T, V, \mu) = \sum_N \frac{1}{N!} \left(e^{\frac{\mu}{kT}} \right)^N Z(T, V, 1)^N \quad (3.216)$$

$$= \sum_N \frac{1}{N!} \left[e^{\frac{\mu}{kT}} Z(T, V, 1) \right]^N \quad (3.217)$$

$$= \exp \left[Z(T, V, 1) \exp\left(\frac{\mu}{kT}\right) \right] \quad (3.218)$$

where we used the sum rule

$$\sum_i \frac{x^n}{n!} = e^x \quad (3.219)$$

8. In many cases then the macrocanonical partition function can be simply written in terms of the canonical one. In fact, all partition functions are mathematically connected by Laplace transformations (see e.g. Greiner's textbook, p. 247).

9. We can now show that the relative fluctuations in the macrocanonical ensemble go as $N^{-1/2}$, and are therefore negligible. We need to calculate

$$\sigma_N^2 = \langle (N - \langle N \rangle)^2 \rangle = \langle N^2 \rangle - \langle N \rangle^2 \quad (3.220)$$

We have

$$\langle N \rangle = \frac{\sum_{i,N} N e^{-\beta(E_i - \mu N)}}{\sum_{i,N} e^{-\beta(E_i - \mu N)}} = \frac{1}{\beta \mathcal{Z}} \frac{d\mathcal{Z}}{d\mu} \quad (3.221)$$

$$\langle N^2 \rangle = \frac{\sum_{i,N} N^2 e^{-\beta(E_i - \mu N)}}{\sum_{i,N} e^{-\beta(E_i - \mu N)}} = \frac{1}{\beta^2 \mathcal{Z}} \frac{d^2 \mathcal{Z}}{d\mu^2} \quad (3.222)$$

$$= \frac{1}{\beta^2 \mathcal{Z}} \frac{d}{d\mu} \frac{d\mathcal{Z}}{d\mu} = \frac{1}{\beta \mathcal{Z}} \frac{d}{d\mu} (\langle N \rangle \mathcal{Z}) \quad (3.223)$$

$$= \frac{1}{\beta} \frac{d}{d\mu} \langle N \rangle + \frac{\langle N \rangle}{\beta \mathcal{Z}} \frac{d\mathcal{Z}}{d\mu} = \frac{1}{\beta} \frac{d}{d\mu} \langle N \rangle + \langle N \rangle^2 \quad (3.224)$$

Therefore,

$$\sigma_N^2 = \langle N^2 \rangle - \langle N \rangle^2 = \frac{1}{\beta} \frac{d}{d\mu} \langle N \rangle \quad (3.225)$$

so we see that $\sigma_N^2 \sim N$ is extensive, and therefore the relative fluctuation vanish for large N :

$$\frac{\sigma_N}{\langle N \rangle} \sim \langle N \rangle^{-1/2} \quad (3.226)$$

10. We can also use the Gibbs-Duhem relation (1.50) to write this result in another way. From

$$N d\mu = -S dT + V dp \quad (3.227)$$

and for $T = \text{const}$, we get

$$N \frac{\partial \mu}{\partial N} \Big|_{T,V} = V \frac{\partial p}{\partial N} \Big|_{T,V} \quad (3.228)$$

Moreover, we can use the general relation (which is immediate to verify for an ideal gas)

$$\frac{\partial p}{\partial N}|_{T,V} = -\frac{V}{N} \frac{\partial p}{\partial V}|_{T,N} \quad (3.229)$$

to write

$$\frac{\partial \mu}{\partial N}|_{T,V} = -\frac{V^2}{N^2} \frac{\partial p}{\partial V}|_{T,N} \quad (3.230)$$

Inserting this in Eq. (3.225), we find (as usual in thermodynamic notation, we disregard the average brackets, since all quantities are supposed to be averages)

$$\frac{\sigma_N^2}{N^2} = \frac{1}{\beta N^2} \frac{d}{d\mu} N = -\frac{1}{V^2} \frac{\partial V}{\partial p}|_{T,N} = \frac{kT}{V} \beta_V \quad (3.231)$$

where we defined the compressibility (see Eq. 1.81)

$$\beta_V = -\frac{1}{V} \frac{\partial V}{\partial p}|_{T,N} \quad (3.232)$$

Here again, since β_V is an intensive quantity, the relative number fluctuations vanish in the thermodynamic limit $V \rightarrow \infty$.

11. A similar argument holds for the energy fluctuations which now can be written as

$$\frac{\sigma_E^2}{E^2} = \frac{\sigma_{can}^2}{U^2} + \frac{\sigma_N^2}{U^2} \left(\frac{\partial U}{\partial N}|_{T,V} \right)^2 \quad (3.233)$$

Since the fluctuations around the equilibrium state are very small, the macrocanonical ensemble gives the same thermodynamic relations as the other ensembles.

12. During the isothermal compression of a gas, below a certain critical temperature, there is a phase transition during which the pressure does not change, in spite of compression, because the gas condenses in the liquid phase. Then one has $(\partial p / \partial V)_T = 0$ and consequently $\beta_V \rightarrow \infty$. This shows that during a phase transition the fluctuations are no longer negligible.
13. In analogy to what we did in Sec. (3.5), we can derive the macrocanonical distribution by maximizing the entropy with two Lagrange multipliers that enforce energy and number constraints (here $d\Gamma$ includes also a sum over N)

$$-S = \int p \log p d\Gamma - \beta(\langle E \rangle - \int E p d\Gamma) + \alpha(\langle N \rangle - \int N p d\Gamma) \quad (3.234)$$

We have then

$$-\delta S = \int \log p \delta p d\Gamma + \int \delta p d\Gamma + \beta \int E \delta p d\Gamma - \alpha \int N \delta p d\Gamma \quad (3.235)$$

$$= \int (\log p + 1 + \beta E - \alpha N) \delta p d\Gamma \quad (3.236)$$

and therefore

$$p \sim \exp(-\beta E + \alpha N) \quad (3.237)$$

We can now identify

$$\beta = \frac{1}{k_B T}, \quad \alpha = \frac{\mu}{k_B T} \quad (3.238)$$

to obtain the macrocanonical distribution.

3.13 The chemical potential

1. As we have seen, the chemical potential is the change in energy when a particle is added to it, at constant entropy and volume. Keeping a constant entropy is not trivial, and the resulting change can be positive or negative.
2. Let us consider a simple example: a system with two particles, e.g. two harmonic oscillators. Two quanta of energy 2ε distributed among the two particles can be assigned both to one of the particles (so two possible states of the system, let's call them 11, 22), or one for each particles (12), for a total of three possible combinations, i.e. three possible energy states. The entropy is $k \log 3$.
3. If we add one particle, the total number of combinations when still having 2ε of energy is six: 11, 22, 33, 12, 13, 23. The entropy is $k \log 6$. So we added a particle but the entropy changed, so we can't say what the chemical potential μ is.
4. In order to add one particle while keeping the number of states constant to three, there is only one solution: reducing the quanta to 1ε . In this case, in fact, we have again three states: one ε to particle 1, or 2, or 3. Then we see that the energy to add a particle at constant entropy is actually negative and equal to $-\varepsilon$.
5. This can be generalized. The Sackur-Tetrode entropy for N particles is $S(U, V, N)$. Adding one particle while changing U to $U + \mu$ is $S'(U + \mu, V, N + 1)$. We can get μ by requiring that $S = S'$, i.e. solving for μ the equation

$$S(U, V, N) = S'(U + \mu, V, N + 1) \quad (3.239)$$

in the limit of $N \gg 1$ and $U \gg \mu$. The result is Eq. (3.42).

3.14 Ultra-relativistic ideal gas

1. The relativistic relation between energy and momentum is

$$E = \sqrt{m^2c^4 + p^2c^2} \quad (3.240)$$

For a massive particle at rest, $E = mc^2$. For a massive non-relativistic particle, $p \ll mc$, $E \approx mc^2 + \frac{p^2}{2m}$. Finally, for ultra-relativistic particles (eg, photons)

$$E = pc \quad (3.241)$$

So the energy depends linearly with p , rather than quadratically.

2. For photons, we also know that $E = \hbar\omega = \hbar ck$, so the momentum vector can be written as $\vec{p} = \hbar\vec{k}$, where \vec{k} is the wavevector.
3. A gas of photons in equilibrium at constant T follows Planck's law (see later). In this case, the energy density is not a free variable, but depends only on T according to Stefan-Boltzmann's law. Therefore if we bring a particle in or out from the photon gas system at constant T , the energy does not change (as also the pressure and the number density). This implies $\mu = 0$. Essentially, the reason is that photon number is not conserved so adding a photon at equilibrium does not change the number of microstates. Therefore, the macrocanonical ensemble reduces to the canonical one.
4. Then, also assuming indistinguishability, we have

$$Z = \frac{1}{N!} \prod_i^N \left(\int \frac{d^3p_i d^3q_i}{h^3} e^{-\beta cp_i} \right) \quad (3.242)$$

$$= \frac{V^N}{N! h^{3N}} \left(\int d^3p_i e^{-\beta cp_i} \right)^N \quad (3.243)$$

$$= \frac{V^N}{N! h^{3N}} (4\pi \int_0^\infty p^2 dp e^{-\beta cp})^N \quad (3.244)$$

$$= \frac{V^N}{N! (\beta ch)^{3N}} (4\pi \int_0^\infty x^2 dx e^{-x})^N = \frac{1}{N!} \left(\frac{V}{\lambda^3} \right)^N \quad (3.245)$$

where we used the formula

$$\Gamma(3) = \int_0^{\infty} x^2 dx e^{-x} = 2 \quad (3.246)$$

and defined a new relativistic thermal wavelength

$$\lambda_{rel} \equiv \frac{hc}{2\pi^{1/3}kT} \quad (3.247)$$

that has the same role as the previously defined non-relativistic wavelength λ_{nr} . We can notice that, up to some order of unity constants

$$\left(\frac{\lambda_{rel}}{\lambda_{nr}}\right)^2 = \frac{mc^2}{kT} \quad (3.248)$$

and that $mc^2 \gg kT$ for a classical non-relativistic gas.

5. The free energy is then

$$F = -kT \log Z = kT \log N! - NkT \log(V/\lambda^3) \quad (3.249)$$

$$= kT(N \log N - N) - NkT \log(V/\lambda^3) \quad (3.250)$$

$$= -NkT[(1 - \log N + \log(V/\lambda^3))] \quad (3.251)$$

$$= -NkT(1 + 3 \log \frac{d}{\lambda}) \quad (3.252)$$

where

$$d \equiv (V/N)^{1/3} \quad (3.253)$$

is the mean interparticle distance. Quantum effects become important for $\lambda_{rel} \gg d$, i.e. when the photon wavefunctions overlap because they are larger than the typical interparticle separation. The overlap of the wavefunctions can be treated only in a fully quantum-mechanical calculation, as we will see in the next chapter.

6. We can now obtain the thermodynamic relations:

$$S = -\frac{\partial F}{\partial T} = kN \left[4 + 3 \log \left(\frac{d}{\lambda} \right) \right] \quad (3.254)$$

and

$$p = -\frac{\partial F}{\partial V} = \frac{NkT}{V} \quad (3.255)$$

just like for non-relativistic particles, since the dependence of F on the volume is exactly the same.

7. Moreover,

$$E = F + TS = F - T \frac{\partial F}{\partial T} = 3NkT \quad (3.256)$$

and $C_V = 3Nk$. We see that this is larger than for the non-relativistic gas, $C_V = 3Nk/2$. In fact, if we rederive the virial theorem for a kinetic term that is linear in the momentum p , rather than quadratic, we find that $E = kT \times \text{d.o.f.}$

Chapter 4

Statistical mechanics of quantum systems

4.1 Density operators

This subsection follows in part the notes by S. J. van Enk at http://pages.uoregon.edu/svanenk/solutions/Mixed_states.pdf, to which we refer for further details. We also benefited from Griffith, *Quantum Mechanics*.

1. In quantum mechanics we denote the solution of the Schrodinger equation a *pure state* $|\psi\rangle$. Any linear combination (superposition) of pure states is another pure state. We will always consider normalized vectors in the following, i.e. $\langle\psi|\psi\rangle = 1$.
2. The dual bra vector $\langle\phi|$ is a linear operator in Hilbert space that transforms ket vectors in C -numbers, $\langle\phi|\psi\rangle$. This is completely analogous to the scalar product of vectors, but applies also to vectors with infinite degrees of freedom, i.e. functions.
3. An operator \hat{O} is a linear operation that transforms vectors into other vectors; any operator and its hermitian conjugate can be expressed as a linear combination of

$$\hat{O} = |\phi\rangle\langle\psi| \quad (4.1)$$

$$\hat{O}^\dagger = |\psi\rangle\langle\phi| \quad (4.2)$$

4. The expected value of \hat{O} on the state $|\psi\rangle$ is

$$\langle\hat{O}\rangle = \langle\psi|\hat{O}|\psi\rangle \quad (4.3)$$

Only hermitian operators can represent physical observables, because they produce real values of $\langle\hat{O}\rangle$.

5. A special case is the hermitian operator

$$\hat{\rho} = |\psi\rangle\langle\psi| \quad (4.4)$$

which is obviously a projection operator, i.e.

$$(\hat{\rho})^2 = \hat{\rho} \quad (4.5)$$

6. Denoting with $|n\rangle$ an orthonormal basis on Hilbert space, any vector can be uniquely decomposed in this basis as $|\psi\rangle = \sum_n |n\rangle\langle n|\psi\rangle$. We can then define the orthonormality relation as

$$I = \sum_n |n\rangle\langle n| \quad (4.6)$$

and obviously $I|\psi\rangle = |\psi\rangle$.

7. The elements of an operator \hat{O} in the basis $|n\rangle$ are

$$O_{mn} = \langle n|\hat{O}|m\rangle \quad (4.7)$$

and its Trace is then

$$\text{Tr}\hat{O} = \sum_n O_{nn} = \sum_n \langle n|\hat{O}|n\rangle \quad (4.8)$$

Notice that the Trace is a number, not an operator. A crucial property of the trace is that it is invariant with respect to a basis transformation, provided the transformation is unitary, i.e. (using matrix notation) if $O' = U^*OU$ with $U^*U = 1$. In fact, since $\text{Tr}(AB) = \text{Tr}(BA)$ we have

$$\text{Tr}O' = \text{Tr}(U^*(OU)) = \text{Tr}(OUU^*) = \text{Tr}(O) \quad (4.9)$$

8. By inserting the identity $I = \sum_n |n\rangle\langle n|$ we find that for any operator \hat{O} we can write the expected value on the state $|\psi\rangle$ as

$$\langle \hat{O} \rangle = \langle \psi|\hat{O}|\psi\rangle = \sum_n \langle \psi|\hat{O}|n\rangle \langle n|\psi\rangle \quad (4.10)$$

$$= \sum_n \langle n|\psi\rangle \langle \psi|\hat{O}|n\rangle = \sum_n \langle n|\hat{\rho}_\psi \hat{O}|n\rangle \quad (4.11)$$

$$= \text{Tr}(\hat{\rho}_\psi \hat{O}) = \text{Tr}(\hat{O} \hat{\rho}_\psi) \quad (4.12)$$

9. Analogously, we can express probabilities and overlaps in terms of $\hat{\rho}$

$$|\langle \phi|\psi\rangle|^2 = \text{Tr}(\hat{\rho}_\psi |\phi\rangle\langle \phi|) = \text{Tr}(|\phi\rangle\langle \phi| \hat{\rho}_\psi) \quad (4.13)$$

10. Since the expectation values of arbitrary operators and probabilities and overlaps can be expressed by the operator $\hat{\rho}$, the state $|\psi\rangle$ is completely characterized by $\hat{\rho}$, which is called *density operator*.

11. Since $|\psi\rangle$ is normalized, we have

$$\text{Tr}\hat{\rho} = 1 \quad (4.14)$$

12. For an example, take a particle moving in a 1D space x . The wavefunction is

$$\langle x|\psi\rangle = \psi(x) \quad (4.15)$$

Inserting twice the identity $\int dx |x\rangle\langle x|$, the density matrix can be written then as

$$\hat{\rho} = |\psi\rangle\langle \psi| = \int dx \int dx' |x'\rangle\langle x'|\psi\rangle\langle \psi|x\rangle\langle x| \quad (4.16)$$

The matrix elements of $\hat{\rho}$ are then

$$\langle x'|\hat{\rho}|x\rangle = \rho(x', x) = \rho_{x'x} = \psi(x')\psi^*(x) \quad (4.17)$$

One can see then that

$$\text{Tr}(\hat{\rho}) = \int dx |\psi(x)|^2 = 1 \quad (4.18)$$

13. Consider now a mixture of \mathcal{N} pure states, or *mixed state*

$$\hat{\rho} = \sum_{k=1}^{\mathcal{N}} p_k |\psi_k\rangle\langle \psi_k| \quad (4.19)$$

where the weights p_k obey

$$0 < p_k \leq 1, \quad \sum p_k = 1 \quad (4.20)$$

Now $\hat{\rho}$ is no longer a projection operator. So the rule $(\hat{\rho})^2 = \hat{\rho}$ is only valid for pure states and indeed can be used to distinguish between pure and mixed states in a basis invariant way.

14. In general, we can think of mixed state as a collection of pure states, each with associated probability p_k . One reason we consider such mixed states is because the quantum states are hard to isolate, and hence often entangled (“mixed”) to the environment. A mixed state can be for instance a beam of photons each with its own polarization or any collection of pure states prepared with probability p_k , i.e. an ensemble of pure states.
15. Expectation values, i.e. experimentally measured values, are obtained as

$$\langle \hat{O} \rangle = \text{Tr}(\hat{\rho} \hat{O}) = \sum_k p_k \text{Tr}(|\psi_k\rangle \langle \psi_k| \hat{O}) \quad (4.21)$$

i.e. a sum of $O_k = \text{Tr}(|\psi_k\rangle \langle \psi_k| \hat{O})$ weighted by p_k .

16. Since $\hat{\rho}$ is hermitian, we can diagonalize it:

$$\hat{\rho} = \sum_{k=1}^M \lambda_k |\phi_k\rangle \langle \phi_k| \quad (4.22)$$

where the M states $|\phi_k\rangle$ are orthogonal and span the Hilbert space. Since $\hat{\rho}|\phi_i\rangle = \lambda_i|\phi_i\rangle$, the λ_i are the eigenvalues of $\hat{\rho}$.

17. In statistical physics, the probability for the k -th system of an ensemble formed by \mathcal{N} elements to be in a state $|n\rangle$ (we assume from now on that we adopt the orthonormal *energy* wavefunctions as basis) is different from system to system. If the system is in $|\psi^k\rangle = \sum_n a_n^k |n\rangle$, the probability of obtaining as a result of a measurement the energy value corresponding to n is $|a_n^k|^2$.
18. The coefficients a_n^k evolve according to

$$i\hbar \dot{a}_n^k(t) = \sum_m H_{nm} a_m^k(t) \quad (4.23)$$

where H_{nm} are the matrix elements of \hat{H} and of course for all k the normalization

$$\sum_n |a_n^k|^2 = 1 \quad (4.24)$$

is maintained.

19. Now the density operator is defined as the operator with matrix elements

$$\rho_{nm} = \sum_{k=1}^{\mathcal{N}} \lambda_k a_n^k(t) a_m^{k*}(t) \quad (4.25)$$

The diagonal elements ρ_{nn} are then the ensemble average over the probabilities $|a_n^k|^2$; the latter themselves are averages over possible measurements on a single system. Then ρ_{nn} is the probability that a system chosen at random in the ensemble is in the state n . The normalization condition

$$\sum_n \rho_{nn} = 1 \quad (4.26)$$

applies.

20. By differentiating Eq. (4.25) with respect to time and inserting (4.23) and its conjugate, we can see that the operator $\hat{\rho}$ obeys the quantum generalization of Liouville equation (*von Neumann equation*)

$$i\hbar \dot{\rho}_{mn} = (\hat{H}\hat{\rho} - \hat{\rho}\hat{H})_{mn} \quad (4.27)$$

or

$$i\hbar \dot{\hat{\rho}} = [\hat{H}, \hat{\rho}] \quad (4.28)$$

Here again, as in the classical case, for a system to be in equilibrium $\hat{\rho}$ must not depend on time and therefore it commutes with \hat{H} ; therefore it has to depend on \hat{H} , itself independent of time and can be diagonalized with \hat{H} . Therefore $\hat{\rho}$, just as \hat{H} , is represented by a diagonal matrix in the basis formed by the energy eigenstates:

$$\rho_{nm} = \rho_n \delta_{nm} \quad (4.29)$$

In other representations $\hat{\rho}$ is not diagonal, but is always symmetric (otherwise the probability to make a transition from a state to another one would be different from the inverse and equilibrium could not be maintained).

21. Quantum statistics therefore consists in expressing the density operator $\hat{\rho}$ in terms of \hat{H} . Once we have $\hat{\rho}$, the expectation value of any observable is $\text{Tr}(\hat{\rho}\hat{O})$, regardless of the basis.

4.2 Quantum ensembles

1. We can consider now the same ensembles we studied so far. For instance, the canonical ensemble ρ_c becomes now a density operator

$$\hat{\rho}_c = \frac{\exp(-\beta\hat{H})}{\text{Tr} \exp(-\beta\hat{H})} \quad (4.30)$$

(let us be reminded that $\text{Tr} \exp(-\beta\hat{H})$ is a number, not an operator). The expression $e^{-\beta\hat{H}}$ should be interpreted as

$$e^{-\beta\hat{H}} = \sum_n \frac{(-\beta\hat{H})^n}{n!} \quad (4.31)$$

Since $\hat{H}^m|n\rangle = (E_n)^m|n\rangle$, one has

$$e^{-\beta\hat{H}}|m\rangle = \sum_n \frac{(-\beta\hat{H})^n}{n!}|m\rangle = \sum_n \frac{(-\beta E_m)^n}{n!}|m\rangle = e^{-\beta E_m}|m\rangle \quad (4.32)$$

Therefore, the matrix elements of $\hat{\rho}_c$ in the energy basis are

$$\rho_{nm} = \frac{\langle n | \exp(-\beta\hat{H}) | m \rangle}{\text{Tr} \exp(-\beta\hat{H})} = \frac{\exp(-\beta E_n)}{\text{Tr} \exp(-\beta E_n)} \delta_{nm} \quad (4.33)$$

which obviously satisfies the normalization condition $\sum_n \rho_{nn} = 1$ and is indeed in the form (4.29). The expression Tr in the last expression just means “sum over n ”.

2. The average of any observable \hat{O} is then

$$\langle \hat{O} \rangle = \text{Tr}(\hat{\rho}\hat{O}) = \frac{\text{Tr} [\exp(-\beta\hat{H}) \hat{O}]}{\text{Tr} \exp(-\beta\hat{H})} \quad (4.34)$$

$$= \sum_{n,m} \rho_{nm} O_{mn} \quad (4.35)$$

if O_{nm} are the elements of \hat{O} in the energy basis.

3. For instance, the average energy is

$$U = \langle \hat{H} \rangle = \text{Tr}(\hat{\rho}\hat{H}) = \frac{\text{Tr} \left[\exp(-\beta\hat{H}) \hat{H} \right]}{\text{Tr} \exp(-\beta\hat{H})} \quad (4.36)$$

$$= \frac{\sum_n [\exp(-\beta E_n) E_n]}{\sum_n \exp(-\beta E_n)} \quad (4.37)$$

$$= -\frac{\partial}{\partial \beta} \log \text{Tr} \exp(-\beta\hat{H}) \quad (4.38)$$

$$= -\frac{\partial}{\partial \beta} \log Z(T, V, N) \quad (4.39)$$

just as in the classical case.

4. For the grand canonical ensemble one has, similarly,

$$\hat{\rho}_c = \frac{\exp(-\beta(\hat{H} - \mu\hat{N}))}{\text{Tr} \exp(-\beta(\hat{H} - \mu\hat{N}))} \quad (4.40)$$

where now also \hat{N} is an operator (equal to N only for systems with constant number of particles).

5. For instance, let us evaluate the average energy of a single free particle,

$$U = \langle \hat{H} \rangle = \text{Tr}(\hat{\rho}\hat{H}) \quad (4.41)$$

We can choose any basis, since the Trace is invariant with respect to transformations. Since for a free particle $\hat{H} = \hat{p}^2/2m$, we can choose the momentum basis $\langle \vec{k} |$, for which the matrix that represents \hat{H} is diagonal. We have

$$\hat{H}|\vec{k}\rangle = E|\vec{k}\rangle \quad (4.42)$$

where

$$E = \frac{\hbar^2 \vec{k}^2}{2m} \quad (4.43)$$

where in a box of size L one has $\vec{k} = \frac{\pi}{L}(n_x, n_y, n_z)$. This basis is assumed orthonormalized, so $\langle \vec{k} | \vec{k}' \rangle = \delta(\vec{k} - \vec{k}')$. We need first to find the elements of the density operator. We have

$$\langle \vec{k} | \exp(-\beta\hat{H}) | \vec{k}' \rangle = \exp(-\beta \frac{\hbar^2 \vec{k}^2}{2m}) \delta_{\vec{k}\vec{k}'} \quad (4.44)$$

while the canonical partition function is

$$Z(T, V, 1) = \sum_{\vec{k}} \langle \vec{k} | \exp(-\beta\hat{H}) | \vec{k}' \rangle = \sum_{\vec{k}} \exp(-\beta \frac{\hbar^2 \vec{k}^2}{2m}) \quad (4.45)$$

6. When the eigenvalues \vec{k} lie very close together, we can approximate the sum as an integral. To do this, we replace

$$\sum_{\vec{k}} \rightarrow \frac{V}{(2\pi)^3} \int d^3k \quad (4.46)$$

since the elementary cell for \vec{k} has volume in k -space equal to $(2\pi/L)^3$. Then we obtain

$$Z(T, V, 1) = \frac{V}{(2\pi)^3} \int d^3k \exp(-\beta \frac{\hbar^2 \vec{k}^2}{2m}) = \frac{V}{(2\pi)^3} \left(\frac{2m\pi}{\beta\hbar^2} \right)^{3/2} = \frac{V}{\lambda^3} \quad (4.47)$$

and finally the elements of the density operator

$$\rho_{\vec{k}\vec{k}'} = \frac{\lambda^3}{V} \exp(-\beta \frac{\hbar^2 \vec{k}^2}{2m}) \delta_{\vec{k}\vec{k}'} \quad (4.48)$$

Since the elements of \hat{H} in the momentum basis are $\hbar^2 \vec{k}^2 / 2m$, the average energy of a single particle is now

$$\langle \hat{H} \rangle = \text{Tr}(\hat{\rho} \hat{H}) = \sum_{\vec{k}\vec{k}'} \rho_{\vec{k}\vec{k}'} H_{\vec{k}\vec{k}'} \quad (4.49)$$

$$= \sum_{\vec{k}\vec{k}'} \frac{\lambda^3}{V} \exp(-\beta \frac{\hbar^2 \vec{k}^2}{2m}) \frac{\hbar^2 \vec{k}^2}{2m} \delta_{\vec{k}\vec{k}'} \quad (4.50)$$

$$= \frac{V}{(2\pi)^3} \frac{\lambda^3}{V} \frac{\hbar^2}{2m} \int d^3k \vec{k}^2 \exp(-\beta \frac{\hbar^2 \vec{k}^2}{2m}) \quad (4.51)$$

$$= \frac{\lambda^3}{(2\pi)^3} \frac{\hbar^2}{2m} 4\pi \int_0^\infty dk k^4 \exp(-\beta \frac{\hbar^2 k^2}{2m}) \quad (4.52)$$

$$= \frac{\lambda^3}{2\pi^2} \frac{\hbar^2}{2m} \left(\frac{2m}{\beta \hbar^2} \right)^{5/2} \int_0^\infty dx x^{3/2} \exp(-x) \quad (4.53)$$

$$= \frac{3}{2} kT \quad (4.54)$$

(the last integral is $\Gamma(5/2) = 3\sqrt{\pi}/4$) in exact correspondance with the classical result.

7. The same result could have been obtained much faster by the usual relation between partition function and average energy (applied in this case to a single particle)

$$\langle \hat{H} \rangle = - \frac{\partial \log Z(T, V, 1)}{\partial \beta} \quad (4.55)$$

8. For N free particles, the partition function is simply

$$Z(T, V, N) = (Z(T, V, 1))^N = \frac{V^N}{\lambda^{3N}} \quad (4.56)$$

Notice that the Gibbs factor is missing, since we have not yet accounted for indistinguishable particles.

4.3 Symmetry of many-particle wave functions

1. Let us consider now a system of two particles, and a Hamiltonian that is the sum of two independent Hamiltonians. The Hamiltonian is then invariant under a permutation of the two particles. Writing the Hamiltonian as a differential operator, we know that in such case, the solution can be written as a separable function: $\psi(r_1, r_2) = \phi_a(r_1)\phi_b(r_2)$, which should be read as ‘‘particle 1 is in state a , particle 2 in state b ’’. However, this wavefunction is not invariant under permutation, since in general $\phi_a(r_1)\phi_b(r_2) \neq \phi_b(r_1)\phi_a(r_2)$. We can however always symmetrize ψ in two ways, obtaining a symmetric and an antisymmetric wavefunction,

$$\psi^S(r_1, r_2) = A[\phi_a(r_1)\phi_b(r_2) + \phi_a(r_2)\phi_b(r_1)] \quad (4.57)$$

$$\psi^A(r_1, r_2) = A[\phi_a(r_1)\phi_b(r_2) - \phi_a(r_2)\phi_b(r_1)] \quad (4.58)$$

In the second case, the wavefunction gets an overall minus sign under permutation, but this is not a problem since wavefunctions are always defined up to a phase. This scheme can be applied to any number of particles.

2. More in general, if the Hamiltonian of a system is invariant under a certain operation (symmetry), there must be an operator which commutes with \hat{H} and can therefore be diagonalized along with \hat{H} . In our case, first applying a Hamiltonian operator to get the energy states and then a permutation is the same as first permuting the particles and then applying the Hamiltonian.
3. For a Hamiltonian that is invariant under the exchange between two particles, eg under exchange of particles i and j , we associate the permutation (or transposition) operator P_{ij} such that

$$\hat{P}_{ij}\psi(\dots r_i \dots r_j \dots) = \psi(\dots r_j \dots r_i \dots) \quad (4.59)$$

Permuting twice among i, j one returns to the original wavefunction, so $\hat{P}_{ij}^2 = \hat{I}$. Its eigenvalues are then obtained by noting that

$$\hat{P}_{ij}\psi = \lambda\psi \quad (4.60)$$

$$\hat{P}_{ij}^2\psi = \psi = \lambda^2\psi \quad (4.61)$$

which therefore gives the two eigenvalues $\lambda = \pm 1$. Under an exchange of particles i, j , therefore, the eigenfunctions of \hat{P}_{ij} either remain the same ($\lambda = 1$; symmetric states) or change sign ($\lambda = -1$; antisymmetric states). This implies that there exist a class of particles such that no two of them can occupy the same state (*Pauli principle*). These particles are the *fermions*. All the other particles are called *bosons*. A general permutation can be decomposed into a number P of 2-particle transpositions \hat{P}_{ij} .

4. If the Hamiltonian is invariant with respect to arbitrary permutations, there is a generalized permutation operator \hat{P} that operates such permutations. Since \hat{P} can be diagonalized simultaneously with \hat{H} , one can express any wavefunction as a sum of eigenstates of the permutation operator. This means that one can take any wavefunction and symmetrize or antisymmetrize it by a linear combination of permuted wavefunctions

$$\psi^S = \sum_P \hat{P}\psi \quad (4.62)$$

$$\psi^A = \sum_P (-1)^P \hat{P}\psi \quad (4.63)$$

where the sum extends over all distinguishable permutations.

5. Take now a system of N non-interacting particles, such that $\hat{H} = \sum_i \hat{h}(r_i, p_i)$. This Hamiltonian is clearly invariant under any permutation. Its energy eigenstates are simply products of the 1-particle eigenstates

$$\psi = \prod_i^N \phi_{k_i} = \phi_{k_1}(r_1)\phi_{k_2}(r_2)\dots \quad (4.64)$$

and consequently the most general normalized A, S wavefunctions can be formed by a A, S linear combination of eigenfunctions

$$\psi^S = A \sum_P \hat{P} \left(\prod_i^N \phi_{k_i} \right) \quad (4.65)$$

$$\psi^A = \frac{1}{\sqrt{N!}} \sum_P (-1)^P \hat{P} \left(\prod_i^N \phi_{k_i} \right) \quad (4.66)$$

where A is a normalization constant still to be determined (one has to take into account the fact that if several particles have the same \hat{k} , then the permutation should not be counted when forming Ψ^S since they corresponds to the same state, so are not distinguishable). The second normalization constant, $1/\sqrt{N!}$, simply takes into account the fact that there are $N!$ permutation among N numbers.

6. Notice that

$$\sum_P (-1)^P \hat{P} \left(\prod_i \phi_{k_i} \right) = \det \Phi \quad (4.67)$$

where Φ is the matrix formed by a first row equal to $\phi_{k_1}(r_1), \phi_{k_1}(r_2), \phi_{k_1}(r_3), \dots$, followed by a second row equal to $\phi_{k_2}(r_1), \phi_{k_2}(r_2), \phi_{k_2}(r_3), \dots$ etc. This is called *Slater determinant*. If two of the quantum numbers k_i are equal, i.e. if two particles occupy the same quantum state, the determinant vanishes and there is no antisymmetric wavefunction.

7. A powerful theorem (the *spin-statistics theorem*) that can be demonstrated in quantum field theory says that fermions have half-integer spin, while bosons have integer spin, and they keep this feature until they exist.
8. The constant A can be determined only if we know how many permutations are possible. If there are n_i bosons in the state k_i , the normalization is $A = (N!n_1!n_2!\dots)^{-1/2}$.

4.4 Ideal quantum systems

1. A quantum system of identical particles can therefore be characterized by either a symmetric or antisymmetric wavefunction

$$|k_1, k_2, \dots, k_N\rangle^{A,S} \quad (4.68)$$

and correspondingly the canonical partition function is

$$Z(T, V, N) = \text{Tr}[\exp(-\beta\hat{H})] \quad (4.69)$$

$$= \frac{1}{N!} \sum \langle k_1, k_2, \dots, k_N | \exp(-\beta\hat{H}) | k_1, k_2, \dots, k_N \rangle^{A,S} \quad (4.70)$$

2. Both A, S states can be approximated for non-interacting particles at high temperatures by the Maxwell-Boltzmann partition function, which can be simply written for indistinguishable quantum states as above with the factorized wavefunction

$$|k_1, k_2, \dots, k_N\rangle^{MB} = |k_1\rangle|k_2\rangle\dots|k_N\rangle \quad (4.71)$$

3. In the energy representation

$$\hat{H}|k_1, k_2, \dots, k_N\rangle^{A,S} = E|k_1, k_2, \dots, k_N\rangle^{A,S} \quad (4.72)$$

where $E = \sum \varepsilon_{k_i}$, so that $\exp(-\beta\hat{H})$ diagonalizes into $\exp(-\beta E)$. The energies ε_k are the 1-particle energy levels.

4. As we have seen already, for the MB statistics, the partition function factorizes:

$$Z^{MB}(T, V, N) = \frac{1}{N!} [Z(T, V, 1)]^N \quad (4.73)$$

5. In order to find a similar simplification for fermions and bosons, we introduce the concept of occupation numbers. Instead of expressing a state by enumerating the quantum numbers for each particle, as in $|k_1, k_2, \dots, k_N\rangle^{A,S}$, we can express it by enumerating how many particles sit in state k_1 , how many in k_2 etc., as in

$$|n_1, n_2, \dots\rangle^{A,S} \quad (4.74)$$

with the convention that n_1 refers to the *occupation number* of the lowest-energy state, the number n_2 to the second level and so on. For instance, if there are six particles, two in state k_1 , three in k_2 , and one in k_3 , instead of writing $|k_1, k_1, k_2, k_2, k_2, k_3\rangle$ we write $|2, 3, 1\rangle$. Clearly the constraint

$$N = \sum_k n_k \quad (4.75)$$

must be fulfilled, and the total energy eigenvalues can be written in function of the occupation numbers as

$$E = \sum_k^{\infty} \varepsilon_k n_k \quad (4.76)$$

(notice that the latter expression is not a constraint, since in the canonical and grand canonical ensembles E is not constant).

6. The ordered sequence n_k , along with the specification of the symmetry of the state, completely determine the state of the quantum system.
7. For bosons, the occupation numbers can be any value from zero to N (*Bose-Einstein statistics*, BE). For fermions, only $n = 0, 1$ is allowed, due to Pauli principle (*Fermi-Dirac statistics*, FD).
8. We define now the number operator through the relation

$$\hat{N}|n_1, n_2, \dots\rangle^{A,S} = N|n_1, n_2, \dots\rangle^{A,S} \quad (4.77)$$

and the occupation number operator through

$$\hat{n}_k|n_1, n_2, \dots\rangle^{A,S} = n_k|n_1, n_2, \dots\rangle^{A,S} \quad (4.78)$$

where $n_k = 0, 1, 2, \dots$ for bosons and $n_k = 0, 1$ for fermions.

9. The orthonormality relation now becomes

$$\langle n'_1, n'_2, \dots | n_1, n_2, \dots \rangle^{A,S} = \delta_{n_1 n'_1} \delta_{n_2 n'_2} \dots \quad (4.79)$$

10. It is then clear that the density operator acting on a state defined by the occupation numbers is

$$\langle n'_1, n'_2, \dots | \hat{\rho} | n_1, n_2, \dots \rangle^{A,S} = Z^{-1} \exp(-\beta \sum_{k=1}^{\infty} n_k \varepsilon_k) \delta_{n'_1 n_1} \delta_{n'_2 n_2} \dots \quad (4.80)$$

where

$$Z(T, V, N) = \sum_{\{n_k\}} \exp(-\beta \sum_{k=1}^{\infty} n_k \varepsilon_k) \quad (4.81)$$

and of course the sum $\sum_{\{n_k\}}$ must run over all the configurations compatible with the constraint $N = \sum n_k$ and with the Pauli principle.

11. The diagonal entries of the matrix representing $\hat{\rho}$

$$P\{n_k\} = \langle n_1, n_2, \dots | \hat{\rho} | n_1, n_2, \dots \rangle^{A,S} = Z^{-1} \exp\left(-\beta \sum_{k=1}^{\infty} n_k \varepsilon_k\right) \quad (4.82)$$

are then to be interpreted as the probability of finding the particular configuration $\{n_k\}$.

12. Similarly, for the grand canonical density operator one finds

$$\langle n'_1, n'_2, \dots | \hat{\rho} | n_1, n_2, \dots \rangle^{A,S} = \mathcal{Z}(T, V, \mu)^{-1} \exp(-\beta \sum_{k=1}^{\infty} n_k (\varepsilon_k - \mu)) \delta_{n'_1 n_1} \delta_{n'_2 n_2} \dots \quad (4.83)$$

Now however there is no longer the constraint $N = \sum n_k$, since N is not fixed in the grand canonical case. Finally, the probability of a particular configuration is

$$P\{n_k\} = \mathcal{Z}^{-1} \exp\left(-\beta \sum_{k=1}^{\infty} n_k (\varepsilon_k - \mu)\right) \quad (4.84)$$

13. Now comparing Eq. (4.73) and Eq. (4.81), one realizes that it is possible to write the partition function in a very general way

$$Z(T, V, N) = \sum_{\{n_k\}} g\{n_k\} \exp\left(-\beta \sum_{k=1}^{\infty} n_k \varepsilon_k\right) \quad (4.85)$$

where the *statistical weight* g is defined as

$$\begin{aligned} g\{n_k\} &= \frac{1}{n_1! n_2! \dots} \quad \text{MB} \\ g\{n_k\} &= 1, \quad \text{BE} \\ g\{n_k\} &= 1, \text{ if all } n_k = 0, 1, \quad \text{FD} \\ &= 0, \text{ otherwise} \end{aligned}$$

Notice that the $N!$ factor of Eq. (4.70) now disappears, since we are summing over *distinct* configurations $\{n_k\}$.

14. For the MB statistics, this statement can be proved by making use of the *multinomial theorem*, according to which

$$(x_1 + x_2 + \dots + x_m)^N = \sum_{n_1, n_2, \dots, n_m}^N \frac{N!}{n_1! n_2! \dots n_m!} \prod_{t=1}^m x_t^{n_t} \quad (4.86)$$

(with $\sum_k n_k = N$). Then we see that

$$\frac{1}{N!} \sum_{n_1, n_2, \dots}^N \frac{N!}{n_1! n_2! \dots} \exp(-\beta \varepsilon_1)^{n_1} \exp(-\beta \varepsilon_2)^{n_2} \dots = \frac{1}{N!} \left(\sum_{k=1}^{\infty} \exp(-\beta \varepsilon_k) \right)^N \quad (4.87)$$

$$= \frac{1}{N!} [Z(T, V, 1)]^N \quad (4.88)$$

15. The same goes for the grand canonical partition function

$$\mathcal{Z}(T, V, \mu) = \sum_{\{n_k\}} g\{n_k\} \exp\left(-\beta \sum_{k=1}^{\infty} n_k (\varepsilon_k - \mu)\right) \quad (4.89)$$

Here the sum is not constrained by $N = \sum_k n_k$ and many calculations become easier

16. Now, for BE, the latter expression can be simplified as follows

$$\mathcal{Z}(T, V, \mu) = \sum_{n_1} \sum_{n_2} \dots \prod_{k=1}^{\infty} \exp(-\beta n_k (\varepsilon_k - \mu)) \quad (4.90)$$

$$= \sum_{n_1} \exp(-\beta n_1 (\varepsilon_1 - \mu)) \sum_{n_2} \exp(-\beta n_2 (\varepsilon_2 - \mu)) \dots \quad (4.91)$$

that is, as a product of sums:

$$\mathcal{Z}^{BE}(T, V, \mu) = \prod_{k=1}^{\infty} \sum_{n_k=0}^{\infty} [\exp(-\beta (\varepsilon_k - \mu))]^{n_k} \quad (4.92)$$

$$= \prod_{k=1}^{\infty} \frac{1}{1 - z \exp(-\beta \varepsilon_k)} \quad (4.93)$$

(z is the fugacity), where we have employed the geometric sum rule (3.160).

17. For the Fermi-Dirac statistics we have

$$\mathcal{Z}^{FD}(T, V, \mu) = \prod_{k=1}^{\infty} \sum_{n_k=0}^1 (\exp(-\beta(\varepsilon_k - \mu)))^{n_k} \quad (4.94)$$

$$= \prod_{k=1}^{\infty} (1 + z \exp(-\beta\varepsilon_k)) \quad (4.95)$$

18. The same calculation also leads to the MB grand canonical partition function

$$\mathcal{Z}^{MB}(T, V, \mu) = \prod_{k=1}^{\infty} \exp[z \exp(-\beta\varepsilon_k)] \quad (4.96)$$

which coincides with Eq. (3.218).

19. The three grand canonical partition functions can be summarized as

$$\log \mathcal{Z} = \frac{1}{a} \sum_{k=1}^{\infty} \log(1 + az \exp(-\beta\varepsilon_k)) \quad (4.97)$$

with $a = 1$ (FD), $a = -1$ (BE) and $a \rightarrow 0$ (MB). We see that the classical MB case is recovered for $z \rightarrow 0$: we will often employ this limit in the following to obtain the classical results. Notice that μ depends on T , so $T \rightarrow \infty$ does not imply $z = e^{\mu/KT} \rightarrow 1$.

20. From the grand canonical partition function one obtains the grand potential

$$\Phi(T, V, \mu) = -kT \log \mathcal{Z}(T, V, \mu) = U - TS - \mu N = -pV \quad (4.98)$$

and all the thermodynamical relations.

21. From the general recipe for the average on an observer,

$$\langle \hat{O} \rangle = \text{Tr}(\hat{\rho} \hat{O}) = \frac{\text{Tr} \left[\exp \left(-\beta(\hat{H} - \mu \hat{N}) \right) \hat{O} \right]}{\text{Tr} \exp \left(-\beta(\hat{H} - \mu \hat{N}) \right)} \quad (4.99)$$

one obtain several important relations, as for instance the expected total particle number

$$N = \sum_{k=1}^{\infty} \langle \hat{n}_k \rangle = kT \frac{\partial}{\partial \mu} \log \mathcal{Z} = \sum_{k=1}^{\infty} \frac{1}{z^{-1} \exp(\beta\varepsilon_k) + a} \quad (4.100)$$

where the fugacity is $z = \exp(\beta\mu)$, from which it follows the *average occupation number*

$$\langle \hat{n}_k \rangle = \frac{1}{\exp(\beta(\varepsilon_k - \mu)) + a} \quad (4.101)$$

The one-particle energy levels ε_k are to be determined by the one-particle Hamiltonian \hat{h} . Notice that $\mu < 0$ for bosons, so there is no singularity for $\varepsilon_k = \mu$.

22. The BE average occupation number always is larger than MB, while the FD is smaller. For small occupation numbers, the three distributions tend to the MB one (Figs. 4.1,4.2).

23. Finally, the average energy is

$$U = \langle \hat{H} \rangle = - \frac{\partial \log \mathcal{Z}}{\partial \beta} \Big|_{z, V} = \sum_{k=1}^{\infty} \frac{\varepsilon_k}{z^{-1} \exp(\beta\varepsilon_k) + a} \quad (4.102)$$

where one should notice that the derivative is performed by keeping z constant, i.e. $\beta\mu$ constant (not just μ).

24. If the particles (bosons or fermions) have $g_s = 2s + 1$ spin degrees of freedom (for instance, $g_s = 2$ for spin 1/2 fermions) or g_s polarizations states (e.g. $g_s = 2$ for photons), then the occupation numbers must be multiplied by g_s , also called *spin degeneracy*.

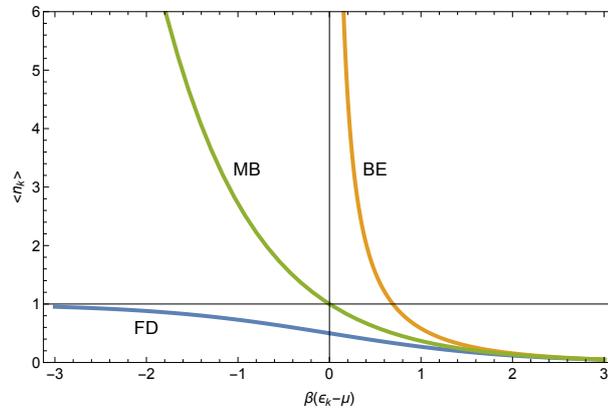


Figure 4.1: Plot of $\langle n_k \rangle$ for BE, FD and MB statistics.

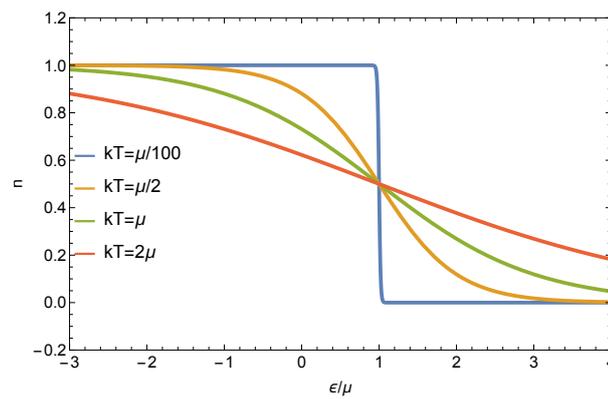
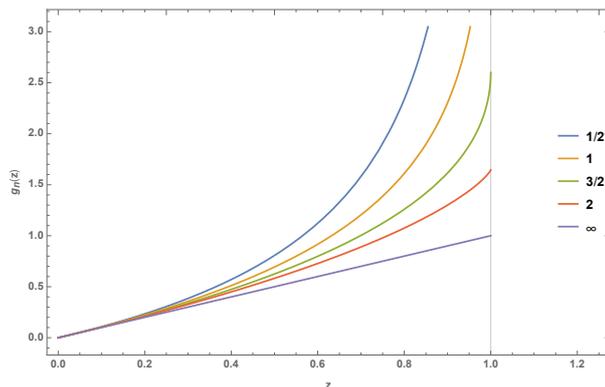


Figure 4.2: Plot of $\langle n_k \rangle$ for a fermion gas.


 Figure 4.3: Plot of $g_n(z)$.

4.5 Bose gas

1. From Eq. (4.100) written for a boson gas, ie $a = -1$, one can obtain the fugacity of μ from N . Since $N \geq \langle n_k \rangle$ one sees that $\varepsilon_k > \mu$ for any quantum number k . If the smallest energy level is zero (as in an ideal gas in an infinite volume) then $\varepsilon_{k=0} = 0$ and $\mu \leq 0$. This in turn implies $0 \leq z \leq 1$.
2. For a large volume, the sum in k can be approximated by an integral

$$\sum_{k=1}^{\infty} \rightarrow \frac{V}{(2\pi)^3} \int d^3k = \frac{2\pi V}{h^3} (2m)^{3/2} \int_0^{\infty} \varepsilon^{1/2} d\varepsilon = \int_0^{\infty} g(\varepsilon) d\varepsilon \quad (4.103)$$

where we used $\varepsilon = \hbar^2 \vec{k}^2 / 2m$ (valid for non relativistic particles) and where

$$g(\varepsilon) = \frac{2\pi V}{h^3} (2m)^{3/2} \varepsilon^{1/2} = \frac{V}{\lambda^3} \frac{2}{\sqrt{\pi}} \beta^{3/2} \varepsilon^{1/2} \quad (4.104)$$

is the one-particle density of states.

3. However the integral is not a good approximation near $\varepsilon = 0$, where the discreteness of the energy levels is important. In order to take this into account, the grand canonical partition must therefore be written as

$$\log \mathcal{Z} = - \sum_k \log(1 - z \exp(-\beta \varepsilon_k)) \quad (4.105)$$

$$= - \int g(\varepsilon) \log(1 - z \exp(-\beta \varepsilon)) d\varepsilon - \log(1 - z) \quad (4.106)$$

Also, we have

$$N = \int \frac{g(\varepsilon)}{z^{-1} \exp(\beta \varepsilon) - 1} d\varepsilon + \frac{z}{1 - z} \quad (4.107)$$

The last term is indeed the expected number of particles in the ground state.

1. These integrals can be performed by defining the function

$$g_n(z) = \frac{1}{\Gamma(n)} \int_0^{\infty} \frac{x^{n-1} dx}{z^{-1} \exp(x) - 1} \quad (4.108)$$

(see Fig. 4.3) for $0 \leq z \leq 1$ and $n > 1$. Two useful values for $\Gamma(n)$ are $\Gamma(3/2) = \sqrt{\pi}/2$ and $\Gamma(5/2) = 3\sqrt{\pi}/4$. We have then

$$\log \mathcal{Z} = \frac{V}{\lambda^3} g_{5/2}(z) - \log(1 - z) \quad (4.109)$$

$$N = \frac{V}{\lambda^3} g_{3/2}(z) + \frac{z}{1 - z} = N_\varepsilon + N_0 \quad (4.110)$$

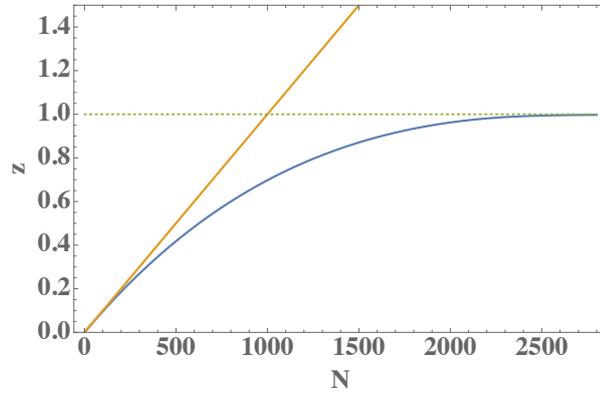


Figure 4.4: Plot of z as a function of N . The straight orange line is the approximation $z \sim N$ of Eq. (4.114).

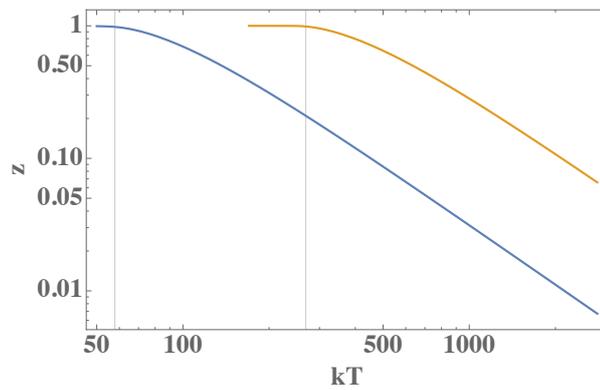


Figure 4.5: Plot of z as a function of temperature for $N = 10^3$ (bottom curve) and $N = 10^4$ (top curve), and $V/\lambda^3 = 1$. For large T , the behavior is $z \sim T^{-3/2}$. The vertical lines mark the critical T_c below which a BEC forms.

where the last term denotes the number of particles in excited states and the number of particles in the ground state, respectively. The equation for N can be used to obtain, numerically, z as a function of N (see Fig. 4.4). For very small z the ground state is empty; for $z \rightarrow 1$, however, it gets progressively filled up.

2. The Bose-Einstein function $g_n(z)$ has the useful Taylor expansion for $0 \leq z \leq 1$

$$g_n(z) = \sum_{k=1}^{\infty} \frac{z^k}{k^n} \quad (4.111)$$

which shows that for small z , $g_n(z) \sim z$. The value at $z = 1$ equals the Riemann function $\zeta(n)$ defined as

$$g_n(1) = \zeta(n) = \sum_{k=1}^{\infty} \frac{1}{k^n} \quad (4.112)$$

Some important values are $\zeta(3/2) = 2.612$, $\zeta(4) = \pi^4/90$. Another useful property is the recursion

$$\frac{\partial}{\partial z} g_n(z) = \frac{1}{z} g_{n-1}(z) \quad (4.113)$$

3. For very small z we have then $N \approx N_\varepsilon = Vz/\lambda^3$, i.e.

$$z = \frac{\lambda^3 N}{V} \quad (4.114)$$

or $z \sim T^{-3/2}$. Then we recover the result for a classical ideal gas obtained from the Sackur-Tetrode formula Eq. (3.42)

$$\mu = -kT \log \frac{V}{\lambda^3 N} \quad (4.115)$$

For very large temperatures therefore $z \rightarrow 0$ (notice that this could not have been guessed from the definition $z = e^{\mu/kT}$, since μ depends on T as well). For very small T , instead, $N_\varepsilon \rightarrow 0$ and $z \rightarrow N_0/(1 + N_0) \rightarrow 1$ (taking as usual $N \sim N_0 \rightarrow \infty$). This behavior is represented in Fig. (4.5).

4. The maximum number of particles that can be contained in excited states is

$$N_{\varepsilon(max)} = \frac{V}{\lambda^3} \zeta(3/2) \quad (4.116)$$

where ζ is the Riemann function and $\zeta(3/2) = g_{3/2}(1) \approx 2.612$. If $N > N_{\varepsilon(max)}$ then a fraction of particles will have to stay in the ground state: this is called *Bose-Einstein condensation* (BEC). Predicted by Bose and Einstein in the 20s, the BEC has been observed only in 1995.

5. This happens then when

$$\frac{N\lambda^3}{V} = \rho\lambda^3 > \zeta(3/2) \quad (4.117)$$

($\lambda = h/(2\pi mkT)^{1/2}$ and $\rho \equiv N/V$) ie for large densities and/or small temperatures. In this situation, the de Broglie thermal wavelength λ is larger than the mean particle separation and the wavefunctions overlap. When this condition is satisfied, the system will start to condense more and more into the ground state. Any additional particle added to the system will also condense. The entire system can be described as a single coherent quantum state. Since high densities often imply a departure from an ideal gas (or even liquefaction or solidification), BECs are typically produced at extremely low temperatures, of the order of nK .

6. In order for the ground state to be significantly populated, N_0/N has to be close to unity. For very large N this is possible only if $z \approx 1$. So in a Bose-Einstein condensation we can approximate $z = 1$.

7. The critical temperature at which a BEC begins is

$$kT_c = \left(\frac{N}{V}\right)^{2/3} \frac{h^2}{2\pi m \zeta^{2/3}(3/2)} \quad (4.118)$$

In practice, the temperature for all interesting physical systems is extremely low, of the order of nanoKelvins, except for Helium-4 (2.17K), which however cannot be described by a simple non-interacting gas. At these low temperatures, a typical gas liquifies or solidifies. In order to maintain the gas phase, the systems have to be very rarefied, i.e. N/V very small.

8. The pressure can be easily obtained by Eq. (3.214), $pV = kT \log \mathcal{Z}$. This gives

$$p = \frac{kT}{\lambda^3} g_{5/2}(z) - \frac{kT}{V} \log(1-z) \quad (4.119)$$

The second term becomes negligible in the thermodynamic limit $V \rightarrow \infty$, even when $z \rightarrow 1$ (because then $V^{-1} \log(1-z) \rightarrow -V^{-1} \log(1+N_0) \rightarrow 0$ for large V, N_0 such that $N_0/V = \text{const}$). This is due to the fact that the ground state does not contribute to pressure since it has no kinetic energy. In a BEC we can put $z = 1$ and therefore we can approximate

$$p \approx \frac{kT}{\lambda^3} \zeta(5/2) \quad (4.120)$$

We see then than in a BEC the pressure depends only on temperature, not on volume. Every new particle added to the system will have to go to the ground state, without contributing to pressure. The same applies for a volume compression, which decreases N_ε and therefore increases N_0 for a given N .

9. On the other hand, for $z \ll 1$, $p = kTz/\lambda^3$. Since at the same time $N \approx N_\varepsilon = Vz/\lambda^3$, we obtain $pV = NkT$, i.e. the classical ideal gas law.
10. Eq. (4.120) shows that along an isothermal compression the curve $p(V)$ is horizontal, just as during a gas-liquid phase transition. Here the phases are the condensed state and the excited particles. *The Bose-Einstein condensation takes place in momentum space rather than in physical space.*
11. For the energy we have

$$U = \int \frac{g(\varepsilon)\varepsilon}{z^{-1} \exp(\beta\varepsilon) - 1} d\varepsilon = \frac{3}{2} kT \frac{V}{\lambda^3} g_{5/2}(z) \quad (4.121)$$

(there is no ground state term now since $\varepsilon_0 = 0$). For small z this reproduces the classical ideal gas case (using Eq. 4.114)

$$U = \frac{3}{2} kT \frac{V}{\lambda^3} g_{5/2}(z) \approx \frac{3}{2} kT \frac{V}{\lambda^3} z = \frac{3}{2} NkT \quad (4.122)$$

while for $z \rightarrow 1$

$$U = \frac{3}{2} kT \frac{V}{\lambda^3} g_{5/2}(1) = \frac{3}{2} kT \frac{V}{\lambda^3} \zeta(5/2) \quad (4.123)$$

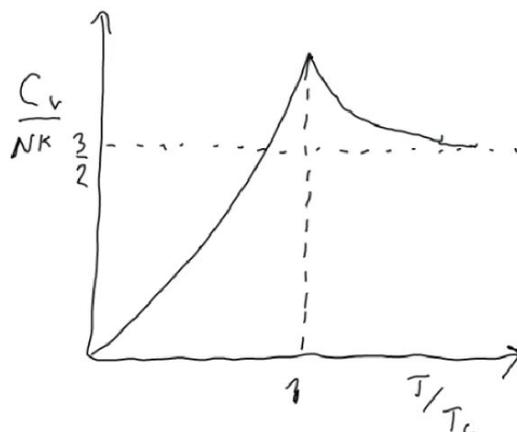
12. The heat capacity can be evaluated using the thermodynamical relations

$$U = -\frac{\partial \log \mathcal{Z}}{\partial \beta} \Big|_{z,V} \quad (4.124)$$

$$C_V = \frac{\partial U}{\partial T} \Big|_{V,N} \quad (4.125)$$

(notice that to evaluate U we fix z , but to evaluate C_V we need $\partial z / \partial T|_{V,N}$). A good approximation valid for $T > T_c$ is as follows (see similar calculation in Sec. 4.8)

$$\frac{C_V}{Nk} \approx \frac{15}{4} \frac{g_{5/2}}{g_{3/2}} - \frac{9}{4} \frac{g_{3/2}}{g_{1/2}} \quad (4.126)$$

Figure 4.6: Plot of C_V for a boson gas.

while, for $T < T_c$, differentiation of Eq. (4.123) gives immediately

$$\frac{C_V}{Nk} \approx \frac{15}{4} \zeta(5/2) \frac{V}{N\lambda^3} \quad (4.127)$$

For very small z , $g_n(z) \approx z$ for any n , and we see that for $T > T_c$ the heat capacity $C_V \rightarrow 3NkT/2$, i.e. the classical result. For $T < T_c$, instead, we see that the heat capacity goes to zero as $C_V \sim T^{3/2}$.

13. The plot of C_V/Nk shows a cusp at $T = T_c$ in the limit $N \rightarrow \infty$ (Fig. 4.6). This is a typical property of second-order phase transitions: the heat capacity is continuous but its derivative is discontinuous.
14. Combining p and U , one can immediately show that the equation of state for a boson for any temperature is the standard one for a monoatomic ideal gas ($3N$ degrees of freedom, so $(3/2)NkT = U$)

$$p = \frac{2}{3} \frac{U}{V} \quad (4.128)$$

(to avoid confusion, notice that here U is just the kinetic energy; since the rest mass energy $U_0 = Nmc^2$ is much larger than the kinetic energy in the non-relativistic regime we are considering, one has also $p \ll \rho \equiv U_0/V$, so that in General Relativity or Cosmology one says that matter has equation of state $p \approx 0$, meaning that $p \ll \rho$).

15. One finds also, as expected from the Third Law, that $S \rightarrow 0$ for $T \rightarrow 0$. In fact, since for $T \rightarrow 0$ we can consider in $\Phi = -kT \log \mathcal{Z}$ (Eq. 4.106) only the ground state, we have

$$S = -\frac{\partial \Phi}{\partial T} |_{V, \mu} = k \frac{\partial}{\partial kT} (kT \log(1 - e^{\mu/kT})) |_{V, \mu} \quad (4.129)$$

$$= k \log(1 - e^{\mu/kT}) + \frac{k^2 T \mu e^{\mu/kT}}{(kT)^2 (1 - e^{\mu/kT})} \quad (4.130)$$

$$\rightarrow 0 \quad (4.131)$$

since $e^{\mu/kT} \rightarrow 0$ for $T \rightarrow 0$ because μ is negative and we are keeping μ constant, and $T^{-1} e^{\mu/kT} \rightarrow 0$ in the same limit. Notice that in some cases, like here, we perform partial derivative keeping μ constant, but in other cases, like in Eq. (4.124), we need to keep z , not μ , constant, and in yet other cases, like Eq. (4.125), neither are kept constant and one has to consider that μ depends on T as well.

4.6 Photon gas

1. For a gas of photons we can proceed exactly in the same manner, except now

$$\varepsilon_k = \hbar\omega = \hbar c|\vec{k}| \quad (4.132)$$

Then we obtain for the number of states in $d\varepsilon$

$$g(\varepsilon)d\varepsilon = \frac{2V}{(2\pi)^3}d^3k = \frac{8\pi V k^2}{(2\pi)^3}dk = \frac{V}{(\hbar c)^3} \frac{\varepsilon^2}{\pi^2}d\varepsilon \quad (4.133)$$

where the initial factor of 2 comes from considering both polarization states ($g_s = 2$). Then we obtain

$$g(\varepsilon) = \frac{V}{(\hbar c)^3} \frac{\varepsilon^2}{\pi^2} \quad (4.134)$$

Moreover, we need to take a vanishing chemical potential, $\mu = 0$, since massless bosons in the state $\varepsilon \rightarrow 0$ can be created without expenditure of energy (photon number is not conserved due to the process of absorption and reemission of the container's walls).

2. The average occupation number and energy are then

$$\langle n_k \rangle = \frac{1}{e^{\beta\varepsilon_k} - 1} \quad (4.135)$$

$$\langle \varepsilon_k \rangle = \frac{\varepsilon_k}{e^{\beta\varepsilon_k} - 1} \quad (4.136)$$

The total energy is

$$U = \int \frac{g(\varepsilon)\varepsilon d\varepsilon}{e^{\beta\varepsilon} - 1} = \int \frac{g_\omega(\omega)\hbar\omega d\omega}{e^{\beta\varepsilon} - 1} \quad (4.137)$$

where $g_\omega(\omega) = g(\varepsilon)(d\varepsilon/d\omega) = g(\varepsilon)\hbar$. Denoting with $u(\omega)d\omega$ the amount of energy density in the frequency interval $d\omega$, we can define the *spectral distribution* $u(\omega)$ of the energy density as

$$u(\omega) = \frac{1}{V}g(\varepsilon)\frac{\hbar^2\omega}{e^{\beta\varepsilon} - 1} = \frac{\hbar\omega^3}{c^3\pi^2(e^{\beta\hbar\omega} - 1)} \quad (4.138)$$

This is indeed the famous Planck's black body radiation formula. For small ω (long wavelengths), the behavior can be approximated as $u(\omega) \sim \omega^2$, which is the Rayleigh-Jeans approximation, while for large ω one recovers Wien's approximation $u(\omega) \sim \omega^3 e^{-\beta\hbar\omega}$. Planck's law has a peak at $\hbar\omega \approx 2.82kT$ (see Fig. 4.7).

3. The total energy density is then

$$\frac{U}{V} = \int_0^\infty u(\omega)d\omega = \frac{(kT)^4}{\pi^2\hbar^3c^3} \int \frac{x^3 dx}{e^x - 1} = \frac{\pi^2}{15} \frac{(kT)^4}{\hbar^3c^3} \quad (4.139)$$

(the last integral equals $\zeta(4)\Gamma(4) = \pi^4/15$). The characteristic T^4 behavior is called Stefan's Law. Stefan's constant σ is obtained when writing the expression for the radiation that leaks out of a cavity (black body box). In this case the radiation flow per unit area is $Uc/4V = \sigma T^4$ where $\sigma = \pi^2 k^4/60\hbar^3c^2$.

4. The grand canonical partition function for a photon gas is

$$\log \mathcal{Z} = \frac{pV}{kT} = - \sum_k \log(1 - e^{-\beta\varepsilon}) \rightarrow - \int g(\varepsilon) \log(1 - e^{-\beta\varepsilon})d\varepsilon \quad (4.140)$$

The additional ground state term is now always negligible in the limit of large V with respect to the integral. With an integration per parts we obtain

$$\log \mathcal{Z} = \frac{\beta}{3} \int \frac{\varepsilon g(\varepsilon)d\varepsilon}{\exp(\beta\varepsilon) - 1} = \frac{\beta}{3} \frac{V}{(\hbar c)^3} \frac{1}{\pi^2} \int \frac{(\beta\varepsilon)^3 d(\beta\varepsilon)}{\exp(\beta\varepsilon) - 1} \quad (4.141)$$

$$= \frac{1}{3} \frac{V}{(\hbar c\beta)^3} \frac{1}{\pi^2} \int \frac{x^3 dx}{\exp x - 1} = \frac{1}{3} \frac{V}{(\hbar c\beta)^3} \frac{1}{\pi^2} \zeta(4)\Gamma(4) \quad (4.142)$$

$$= \frac{V}{(\hbar c\beta)^3} \frac{\pi^2}{45} \quad (4.143)$$

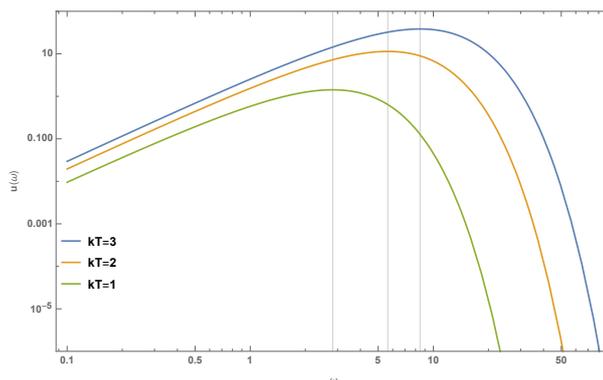


Figure 4.7: Planck spectrum for various temperatures (kT is in units of \hbar/sec ; the units of u are arbitrary). The vertical lines mark the peak at $2.82kT$.

from which we obtain all the thermodynamical relations. For instance we have the relativistic equation of state

$$p = \frac{\pi^2 (kT)^4}{45 (\hbar c)^3} = \frac{1}{3} \frac{U}{V} = \frac{1}{3} \rho \quad (4.144)$$

where ρ is the energy density. Here, contrary to Eq. (4.128), we are in a relativistic regime so this equation of state holds also in GR and cosmology. The two general-relativistic expressions for massive and massless particles are then $p = 0$ and $p = \rho/3$, respectively.

4.7 Phonons

1. Phonons are collective excitations of atoms in a lattice. We only consider phonons produced thermally by random excitations of a monoatomic lattice, not acoustic phonons produced by external perturbations or optical phonons produced by diatomic ions. Because of their properties that are similar to the wave representation of particles, phonons are also called quasi-particles.
2. A solid can be schematically approximated as a collection of atoms of equal mass m at positions x_1, x_2, \dots, x_N (each of the x 's represent the three coordinates of the atoms) oscillating around their average positions $\bar{x}_1, \bar{x}_2, \dots$, where the potential is supposed to possess a minimum. Around this minimum, the potential can be approximated as

$$V(x_1, x_2, \dots, x_N) \approx V_0 + \frac{1}{2} \sum_{i,k} \frac{\partial^2 V}{\partial x_i \partial x_k} \Big|_{\min} (x_i - \bar{x}_i)(x_k - \bar{x}_k) = \sum_{i,k} \alpha_{ik} \xi_i \xi_k \quad (4.145)$$

where $\xi_n = x_n - \bar{x}_n$. We now put the irrelevant constant V_0 to zero. The corresponding Hamiltonian is

$$H = \sum_i \frac{1}{2} m \dot{\xi}_i^2 + \sum_{i,k} \alpha_{ik} \xi_i \xi_k \quad (4.146)$$

3. The real, symmetric matrix α_{ij} can now be diagonalized (simultaneously with the kinetic term $m \delta_{ij} \dot{\xi}_i \dot{\xi}_j$ since the identity matrix commutes with any matrix), by introducing new coordinates q_i

$$H = \sum_i \frac{1}{2} m (\dot{q}_i^2 + \omega_i^2 q_i^2) \quad (4.147)$$

where now the sum extends to $3N$ for N atoms each with three spatial degrees of freedom. The coordinates q_i , also called normal modes, are linear combinations of the atoms' coordinates x_i . Their oscillations represent therefore collective motions, i.e. waves propagating through the solid: the *phonons*. The frequencies ω_i are the phonon eigenfrequencies.

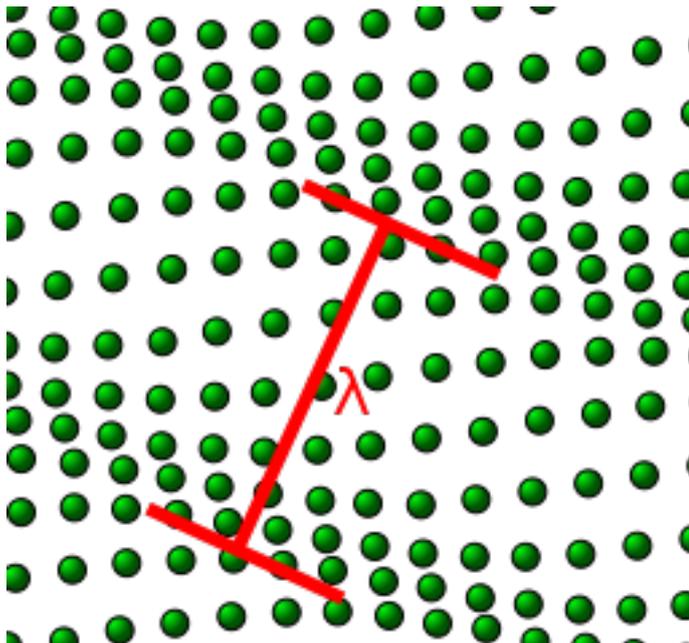


Figure 4.8: Illustration of phonons in a square lattice. Credit: Wikipedia, FlorianMarquardt

4. To simplify further the problem we adopt now Einstein's assumption that all eigenfrequencies are the same. The problem is then identical to $3N$ independent, distinguishable, quantum oscillators with frequency ω . However the frequency ω has to be interpreted not as the oscillation frequency of a single atom but as the frequency of the phonon waves. Although the oscillators are distinguishable, the phonons are not, and we can apply the BE statistics.
5. The average excitation level for each oscillator is

$$\langle n \rangle = \frac{1}{e^{\beta\hbar\omega} - 1} \quad (4.148)$$

and the average energy is $\langle \varepsilon \rangle = \hbar\omega(\frac{1}{2} + \langle n \rangle)$. In view of the phonon interpretation, the average excitation level should be interpreted as the number of indistinguishable excitation quanta per oscillator, each of them of energy $\hbar\omega$. So instead of having many oscillators at different energy levels, we have many waves each carrying a $\hbar\omega$ quantum of energy. The two pictures are completely interchangeable.

6. In this way, we see the phonons as very similar to photons. They are the energy quanta of the oscillations in a solid and each quantum carries an energy $\Delta\varepsilon = \hbar\omega$. Their velocity of propagation is however not the speed of light but some value c_s (speed of sound) which in general also depends on whether the wave oscillates transversally or longitudinally with respect to the direction of propagation \vec{k} .
7. The problem can then be immediately solved using the Bose-Einstein distribution. Just as for the photon case, the phonons are not conserved and the chemical potential vanishes. The energy is found to be

$$U = 3N\langle \varepsilon \rangle = 3N\hbar\omega \left(\frac{1}{2} + \frac{1}{e^{\beta\hbar\omega} - 1} \right) \quad (4.149)$$

and the heat capacity becomes

$$C_V = \frac{\partial U}{\partial T} \Big|_{V,N} = 3Nk(\beta\hbar\omega)^2 \frac{e^{\beta\hbar\omega}}{(e^{\beta\hbar\omega} - 1)^2} \quad (4.150)$$

and is formally identical to the one for quantum oscillators, see Eq. (3.172) and Fig. (3.1).

8. The heat capacity goes to zero as expected for $T \rightarrow 0$ while it takes the classical value $C_V = 3Nk$ for large temperatures, in agreement with the virial theorem for $6N$ degrees of freedom ($3N$ kinetic terms, $3N$ potential terms) .
9. The assumption of a single frequency is of course too crude. Debye generalized the Einstein model by assuming a continuous spectrum of frequencies up to the largest allowed by the discrete nature of the atom lattice, the cut-off frequency ω_c , corresponding to a wavelength $\lambda_c = 2\pi v/\omega_c$ of the order of the mean interparticle distance, if v is the wave propagation speed. The density of states $g(\varepsilon)$ for a system that obeys a linear relation $\varepsilon = \hbar c_s k$ is Eq. (4.133), which was however for two polarizations; so for one polarization we have

$$g(\varepsilon) = \frac{V}{2(\hbar c_s)^3} \frac{\varepsilon^2}{\pi^2} \quad (4.151)$$

For phonons, we have in fact three polarization, two transversal with velocity c_t and one longitudinal with velocity c_l , so

$$g(\varepsilon) = \frac{V}{(\hbar c_t)^3} \frac{\varepsilon^2}{\pi^2} + \frac{V}{2(\hbar c_l)^3} \frac{\varepsilon^2}{\pi^2} \quad (4.152)$$

10. The total number of normal modes is $3N$, so we must impose the condition

$$3N = \int_0^{\varepsilon_c} g(\varepsilon) d\varepsilon \quad (4.153)$$

($\varepsilon_c = \hbar\omega_c$) to obtain the cut-off frequency

$$\omega_c^3 = 18\pi^2 \frac{N}{V} \left(\frac{1}{c_l^3} + \frac{2}{c_t^3} \right)^{-1} \quad (4.154)$$

which indeed shows that $\lambda_c = 2\pi c_{l,t}/\omega_c$ is approximately the interparticle mean separation $(V/N)^{1/3}$. Then we see that

$$g(\varepsilon) = 9N \frac{\varepsilon^2}{(\hbar\omega_c)^2} \quad (4.155)$$

11. Then, since $g(\omega) = g(\varepsilon)\hbar$, one obtains

$$U = \int_0^{\omega_c} g(\omega) \langle \varepsilon \rangle d\omega = 9 \frac{N\hbar}{\omega_c^3} \int_0^{\omega_c} \omega^3 d\omega \left(\frac{1}{e^{\beta\hbar\omega} - 1} + \frac{1}{2} \right) \quad (4.156)$$

and

$$C_V = 9Nk \left(\frac{kT}{\hbar\omega_c} \right)^3 \int_0^{x_c} dx \frac{x^4 e^x}{(e^x - 1)^2} \quad (4.157)$$

with $x_c = \hbar\omega_c/kT$, and one can see that $C_V \sim T^3$ for small T (the integral evaluates to the constant $4\pi^4/15$ for $x_c \rightarrow \infty$), in better agreement with experiments than the exponential behavior of the Einstein model. The cut-off frequency ω_c depends on the transverse and longitudinal speed of sound and in practice is estimated from measurements. Even the Debye model is however just a rough approximation to the real cases.

12. The heat capacity does not take into account the degrees of freedom associated with the free electrons in a metal. This however is a very good approximation at room temperature. The electrons are fermions and obey the Fermi distribution; as we will see in the next section, their heat capacity goes to zero for temperatures below their Fermi temperature, which in metals happens to be of the order of $10^4 \div 10^5$ K (10eV)

4.8 Fermi gas

1. For a gas of non-interacting fermions, the grand canonical partition function is

$$\log \mathcal{Z} = \sum_k \log(1 + z \exp(-\beta \varepsilon_k)) \quad (4.158)$$

where again the fugacity is to be evaluated as a function of N using the relation

$$N = \sum_{k=1}^{\infty} \langle \hat{n}_k \rangle = \sum_{k=1}^{\infty} \frac{1}{z^{-1} \exp(\beta \varepsilon_k) + 1} \quad (4.159)$$

Now the fugacity z can take any positive value.

2. Converting sums to integrals, we have

$$\log \mathcal{Z} = \int_0^{\infty} g(\varepsilon) \log(1 + z \exp(-\beta \varepsilon)) d\varepsilon \quad (4.160)$$

$$N = \int_0^{\infty} \frac{g(\varepsilon)}{z^{-1} \exp(\beta \varepsilon) + 1} d\varepsilon \quad (4.161)$$

where for non relativistic particles with spin s we have

$$g(\varepsilon) = g_s \frac{2\pi V}{h^3} (2m)^{3/2} \varepsilon^{1/2} \quad (4.162)$$

where $g_s = 2s + 1$ is the degeneracy of states due to the spin.

3. The additional $k = 0$ term that was necessary in the boson case, see Eqs. (4.110), is now negligible in the thermodynamic limit: in fact, $z/(1+z)$ is always smaller than unity. This implies immediately that there is no Bose-Einstein condensations for fermions, due to the Pauli principle.
4. As in the boson case, we define the integral function

$$f_n(z) = \frac{1}{\Gamma(n)} \int_0^{\infty} \frac{x^{n-1} dx}{z^{-1} \exp(x) + 1} = \sum_{k=1}^{\infty} (-1)^{k-1} \frac{z^k}{k^n} \quad (4.163)$$

(see Fig. 4.9) and express the integrals as (integrating by parts $\varepsilon^{1/2}$ in 4.160)

$$\log \mathcal{Z} = \frac{g_s V}{\lambda^3} f_{5/2}(z) \quad (4.164)$$

$$N = \frac{g_s V}{\lambda^3} f_{3/2}(z) \quad (4.165)$$

5. The function $f_n(z)$ is a monotonous function that grows as z for $z \ll 1$ and as $(\log z)^n/n!$ for $z \gg 1$, and is $(1 - 2^{1-n})\zeta(n)$ for $z = 1$. They obey the recursion relation

$$\frac{\partial}{\partial z} f_n(z) = \frac{1}{z} f_{n-1}(z) \quad (4.166)$$

6. Since we have \mathcal{Z} we can evaluate all the thermodynamical equations of state. So we have for instance the internal energy

$$U = -\frac{\partial}{\partial \beta} \log \mathcal{Z}|_{z,V} = \frac{3}{2} N k T \frac{f_{5/2}(z)}{f_{3/2}(z)} \quad (4.167)$$

At small densities, high temperatures one has $z \ll 1$ and one obtains the usual EOS of the ideal gas, $U = 3NkT/2$ and $p = 2U/3V = (2/3)\rho$.

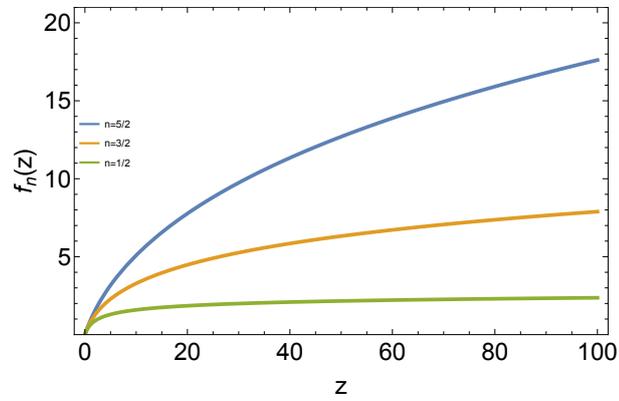
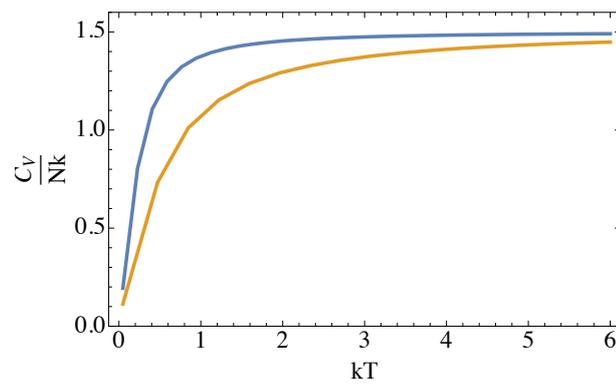
Figure 4.9: Plot of $f_n(z)$.

Figure 4.10: Heat capacity for fermions for two values of the gas density (higher curve, lower density, i.e. “more classical”).

7. The specific heat can be evaluated by the usual expression $C_V = (\partial U/\partial T)|_{V,N}$. One need however also $(\partial z/\partial T)|_{V,N}$. This can be evaluated by making use of the recursion relation. One has in fact

$$\frac{\partial z}{\partial T} \frac{1}{z} f_{1/2} = \frac{\partial z}{\partial T} \frac{\partial}{\partial z} f_{3/2} = \frac{\partial f_{3/2}}{\partial T} \quad (4.168)$$

Now from Eq. (4.165) one has that $f_{3/2} = N\lambda^3/Vg_s$ depends on T only through λ^3 , so that

$$\frac{\partial f_{3/2}}{\partial T} = -\frac{3}{2T} \frac{N\lambda^3}{Vg_s} = -\frac{3}{2T} f_{3/2} \quad (4.169)$$

and finally

$$\frac{\partial z}{\partial T}|_{V,N} = -\frac{3}{2} \frac{z}{T} \frac{f_{3/2}}{f_{1/2}} \quad (4.170)$$

8. Then we have (we put $f' = \partial f/\partial z$)

$$C_V = \frac{\partial U}{\partial T}|_{V,N} = \frac{3}{2} Nk \frac{f_{5/2}}{f_{3/2}} + \frac{3}{2} NkT \left(\frac{f'_{5/2}}{f_{3/2}} - \frac{f_{5/2} f'_{3/2}}{f_{3/2}^2} \right) \frac{\partial z}{\partial T} \quad (4.171)$$

$$= \frac{3}{2} Nk \frac{f_{5/2}}{f_{3/2}} - \frac{9}{4} NkT \left(\frac{f_{3/2}}{z f_{3/2}} - \frac{f_{5/2} f_{1/2}}{z f_{3/2}^2} \right) \frac{z}{T} \frac{f_{3/2}}{f_{1/2}} \quad (4.172)$$

$$= \frac{15}{4} Nk \frac{f_{5/2}}{f_{3/2}} - \frac{9}{4} Nk \frac{f_{3/2}}{f_{1/2}} \quad (4.173)$$

See Fig. (4.10).

9. For large temperatures, $f_n(z) \rightarrow z$ and one recovers the classical result $C_V = 3Nk/2$.
10. From the expressions for U, N and from $\mu \equiv kT \log z$ and $pV/kT = \log \mathcal{Z}$ one obtains also the free energy and entropy

$$F = N\mu - pV = NkT \left(\log z - \frac{f_{5/2}}{f_{3/2}} \right) \quad (4.174)$$

$$S = \frac{U - F}{T} = Nk \left(\frac{5}{2} \frac{f_{5/2}}{f_{3/2}} - \log z \right) \quad (4.175)$$

11. For $T \rightarrow 0$, $z \gg 1$ and we can use the approximation $f_n(z) = (\log z)^n/n!$. Moreover, the mean occupation number can be approximated by a step function

$$\langle n_k \rangle \approx \Theta(\mu - \varepsilon) \quad (4.176)$$

where $\Theta(x) = 0$ for $x < 0$ and 1 otherwise, i.e. all the states are occupied by a single (or g_s) fermion for $\varepsilon < \mu$ and are empty above this threshold.

12. Then one gets the equations of state for a degenerate ($T \rightarrow 0$) fermion ideal gas

$$p \approx g_s \left(\frac{2\pi m}{h^2} \right)^{3/2} \mu^{5/2} \frac{8}{15\sqrt{\pi}} \quad (4.177)$$

$$n \equiv \frac{N}{V} \approx g_s \left(\frac{2\pi m}{h^2} \right)^{3/2} \mu^{3/2} \frac{4}{3\sqrt{\pi}} \quad (4.178)$$

Contrary to the boson case, now the pressure does not vanish for $T \rightarrow 0$ and is actually independent of T . Combining these two equations, we see that the $p - n$ equation of state for the degenerate non-relativistic Fermion gas can be written as

$$p = K_{nr} n^{5/3} \quad (4.179)$$

where K_{nr} is independent of T .

13. Also, the expected total energy is

$$U = \int_0^\infty \varepsilon g(\varepsilon) \Theta(\mu - \varepsilon) = g_s V \left(\frac{2\pi m}{\hbar^2} \right)^{3/2} \mu^{5/2} \frac{4}{5\sqrt{\pi}} \quad (4.180)$$

From these relations we obtain the average energy per particle

$$\frac{U}{N} = \frac{3}{5} \mu_F \quad (4.181)$$

where

$$\varepsilon_F \equiv \mu_F = \frac{\hbar^2}{2m} \left(\frac{6\pi^2 N}{g_s V} \right)^{2/3} \quad (4.182)$$

This value of the chemical potential is called *Fermi energy*, ε_F : it gives the highest energy (per particle) that the fermions reach when they are all in the ground state, i.e. for $T \rightarrow 0$. The average energy per particle in this state is $3\varepsilon_F/5$. The Fermi momentum is

$$p_F = \sqrt{2m\varepsilon_F} = \hbar \left(\frac{6\pi^2 N}{g_s V} \right)^{1/3} \quad (4.183)$$

14. One finds that

$$\frac{\mu_F}{kT} = \frac{\hbar^2}{2mkT} \left(\frac{6\pi^2 N}{g_s V} \right)^{2/3} = \frac{\lambda^2}{\ell^2} \left(\frac{6\pi^2}{g_s (4\pi)^{3/2}} \right)^{2/3} = \mathcal{O}(1) \left(\frac{\lambda}{\ell} \right)^2 \quad (4.184)$$

where $\ell = (V/N)^{1/3}$ is the mean interparticle separation. This shows that for $kT \gg \mu_F$ the gas behaves classically, while for $kT \ll \mu_F$ the quantum properties are important and the gas becomes degenerate.

15. The free electrons in a metal have a density $n \approx 10^{23} \text{ cm}^{-3}$ and therefore $\mu_F \approx 8\text{eV}$ (actually a bit less because the mass gets an effective correction due to interactions), while room temperature equals $1/40\text{eV}$, so $\mu_F/kT \gg 1$. On the other hand $mc^2/kT \gg 1$, so the electrons are non-relativistic (see next section). Therefore free electrons in a metal at room temperature is an example of a non-relativistic almost ideal Fermi gas at the $T \rightarrow 0$ limit.

16. The specific heat for $T \rightarrow 0$ is linear in T (rather than $T^{3/2}$ as for bosons),

$$\frac{C_V}{Nk} \approx \frac{\pi^2 kT}{2 \varepsilon_F} \quad (4.185)$$

and goes to zero for small temperatures (notice that if we replace the large z approximation $f_n \rightarrow (\log z)^n/n!$ in Eq. (4.173) we get zero, so we need to use a more accurate approximation; see Greiner's book pp. 348-349 for the full calculation). This means that, for instance, the electrons in a metal, or the nuclear energy states, for which $kT \ll \varepsilon_F$, do not contribute at all to the heat capacity, instead of contributing $kT/2$ per degree of freedom as expected classically.

17. One finds also, as expected, that $S \rightarrow 0$ for $T \rightarrow 0$, just as for the Bose condensate. In fact for $T \rightarrow 0$, i.e. $z \rightarrow \infty$, one has $f_n \approx (\log z)^n/n!$ and

$$S \rightarrow Nk \left[\frac{5}{2} \frac{\Gamma(5/2)(\log z)^{5/2}}{\Gamma(7/2)(\log z)^{3/2}} - \log z \right] \rightarrow 0 \quad (4.186)$$

as expected. The degenerate Fermi gas is the most ordered state of fermion systems.

4.9 Relativistic fermions

1. In Special Relativity, for particles of mass m , one has

$$\varepsilon^2 = m^2 c^4 + \hbar^2 k^2 c^2 \quad (4.187)$$

Therefore, one converts the integral $V(2\pi)^{-3}d^3k = V(2\pi)^{-3}4\pi k^2 dk$ over momenta into an integral over energy as

$$\frac{V4\pi k^2}{(2\pi)^3\hbar^3}dk = \frac{V4\pi}{(2\pi)^3\hbar^3c^3}(\varepsilon^2 - m^2c^4)^{1/2}\varepsilon d\varepsilon \quad (4.188)$$

and the density of states, including spin degeneracy, is

$$g(\varepsilon)d\varepsilon = \frac{Vg_s}{(2\pi)^3\hbar^3c^3}4\pi k^2 dk = \frac{Vg_s}{2\pi^2\hbar^3c^3}(\varepsilon^2 - m^2c^4)^{1/2}\varepsilon d\varepsilon \quad (4.189)$$

2. One obtains therefore the following general equations for the number density, the energy density, and the pressure, valid in both the relativistic and non-relativistic regimes

$$n = \frac{N}{V} = \int \frac{g(\varepsilon)}{z^{-1}\exp(\beta\varepsilon) + 1}d\varepsilon = \frac{g_s}{2\pi^2\hbar^3c^3} \int_{mc^2}^{\infty} \frac{(\varepsilon^2 - m^2c^4)^{1/2}}{e^{\beta(\varepsilon-\mu)} + 1} \varepsilon d\varepsilon \quad (4.190)$$

$$\rho = \frac{E}{V} = \int \frac{g(\varepsilon)\varepsilon}{z^{-1}\exp(\beta\varepsilon) + 1}d\varepsilon = \frac{g_s}{2\pi^2\hbar^3c^3} \int_{mc^2}^{\infty} \frac{(\varepsilon^2 - m^2c^4)^{1/2}}{e^{\beta(\varepsilon-\mu)} + 1} \varepsilon^2 d\varepsilon \quad (4.191)$$

$$\begin{aligned} p &= \frac{kT \log \mathcal{Z}}{V} = \frac{kT}{V} \int g(\varepsilon) \log(1 + z \exp(-\beta\varepsilon)) d\varepsilon = kT \frac{g_s}{2\pi^2\hbar^3c^3} \int_{mc^2}^{\infty} (\varepsilon^2 - m^2c^4)^{1/2} \varepsilon \log(1 + z \exp(-\beta\varepsilon)) d\varepsilon \\ &= \frac{1}{3} \frac{g_s}{2\pi^2\hbar^3c^3} \int_{mc^2}^{\infty} \frac{(\varepsilon^2 - m^2c^4)^{3/2}}{e^{\beta(\varepsilon-\mu)} + 1} d\varepsilon \end{aligned} \quad (4.192)$$

where as usual in the last integral we performed an integration by parts ($d(\varepsilon^2 - m^2c^4)^{3/2}/3 = (\varepsilon^2 - m^2c^4)^{1/2}\varepsilon d\varepsilon$). The same expressions, but with a -1 in the denominator instead of $+1$, apply to bosons. The integrals can only be solved numerically. These equations play an important role in many fields. In cosmology, for instance, they express the abundance of particles as a function of temperature and therefore of cosmic expansion.

3. For large temperatures, or equivalently for $z \rightarrow 0$, the occupation number can be replaced by the MB distribution $e^{-\beta(\varepsilon-\mu)}$. One sees then that $p = nkT$ in both the non-relativistic and the relativistic regime, as we already noticed in Sec. 3.14.
4. The Fermi energy now is given by

$$n = \frac{g_s}{2\pi^2\hbar^3c^3} \int_{mc^2}^{\mu_F} (\varepsilon^2 - m^2c^4)^{1/2} \varepsilon d\varepsilon = \frac{g_s}{6\pi^2\hbar^3c^3} (\mu_F^2 - m^2c^4)^{3/2} \quad (4.193)$$

which for $\mu_F = mc^2 + \varepsilon_F$ with $\varepsilon_F \ll mc^2$ reduces to the non-relativistic value Eq. (4.182). In general, then, the Fermi energy (only the kinetic part $\varepsilon_F = \mu_F - mc^2$ is usually considered) becomes

$$\frac{\varepsilon_F}{mc^2} = \left[\frac{\hbar^2}{m^2c^2} \left(\frac{6\pi^2n}{g_s} \right)^{2/3} + 1 \right]^{1/2} - 1 \quad (4.194)$$

5. For a degenerate Fermion gas one uses Eq. (4.176) and the p, ρ -integrals become

$$\rho = \frac{g_s}{2\pi^2\hbar^3c^3} \int_{mc^2}^{\mu_F} (\varepsilon^2 - m^2c^4)^{1/2} \varepsilon^2 d\varepsilon \quad (4.195)$$

$$= \frac{1}{8} \frac{g_s}{2\pi^2\hbar^3c^3} \left(\mu_F \sqrt{\mu_F^2 - m^2c^4} (2\mu_F^2 - m^2c^4) + m^4c^8 \log \left(\frac{mc^2}{\sqrt{\mu_F^2 - m^2c^4} + \mu_F} \right) \right) \quad (4.196)$$

$$\begin{aligned} p &= \frac{1}{3} \frac{g_s}{2\pi^2\hbar^3c^3} \int_{mc^2}^{\mu_F} (\varepsilon^2 - m^2c^4)^{3/2} d\varepsilon \\ &= \frac{1}{8} \frac{g_s}{6\pi^2\hbar^3c^3} \left(\mu_F \sqrt{\mu_F^2 - m^2c^4} (2\mu_F^2 - 5m^2c^4) - 3m^4c^8 \log \left(\frac{mc^2}{\sqrt{\mu_F^2 - m^2c^4} + \mu_F} \right) \right) \end{aligned} \quad (4.197)$$

The expressions can be algebraically simplified by defining the dimensionless momentum y as $y^2 = (\mu_F^2 - m^2 c^4)/m^2 c^4 \equiv p_F^2/m^2 c^2$, so that

$$p = \frac{1}{8} \frac{g_s m^4 c^8}{6\pi^2 \hbar^3 c^3} \left(y \sqrt{y^2 + 1} (2y^2 - 3) + 3 \log \left(y + \sqrt{y^2 + 1} \right) \right) \quad (4.198)$$

$$\rho = \frac{1}{8} \frac{g_s m^4 c^8}{2\pi^2 \hbar^3 c^3} \left(y \sqrt{y^2 + 1} (2y^2 + 1) - \log \left(y + \sqrt{y^2 + 1} \right) \right) \quad (4.199)$$

$$n = \frac{g_s m^4 c^8}{6\pi^2 \hbar^3 c^3} y^3 \quad (4.200)$$

In the ultra-relativistic limit we can take $y \gg 1$ and we find

$$n \sim y^3 \quad (4.201)$$

$$p = \frac{1}{3} \rho \sim y^4 \quad (4.202)$$

and we obtain the $p - n$ equation of state for a degenerate, ultra-relativistic Fermionic gas

$$p = K_{ur} n^{4/3} \quad (4.203)$$

where again K_{ur} is independent of T .

6. We see then that Fermion gases can exist in four qualitatively different regimes, each with their own $p - n$ equation of state: classical and non-relativistic ($p = nkT$), classical and relativistic ($p = nkT$), degenerate and non-relativistic ($p = K_{nr} n^{5/3}$), degenerate and relativistic ($p = K_{ur} n^{4/3}$). The gas become relativistic for $mc^2/kT \ll 1$ and degenerate for $\mu_F/kT \gg 1$.
7. Neutrinos are extremely light fermions, with mass smaller than 1eV. They are therefore relativistic, because essentially any process that create them, e.g. in stars, occurs at much higher energy. The cosmological neutrino background has been relativistic (i.e. $mc^2 \ll kT$) for the entire cosmic history except very recently. For neutrinos, therefore, in most cases one can put $m = 0$ in the above equations.
8. An important application of Fermi gas is to the electrons and helium atoms in a white dwarf. At a typical temperature of 10^7 K the helium atoms are completely ionized and the electrons behave as a relativistic (since $p_F \approx m_e c$) but cold (since $kT \approx 1\text{keV} \ll \varepsilon_F$) gas, while the atoms are non-relativistic. The pressure of the degenerate electron gas supports the star from gravitational collapse if the mass is smaller than the Chandrasekhar mass

$$M_0 = \frac{9}{64} \left(\frac{3\pi}{\alpha^3} \right) \left(\frac{\hbar c}{G m_n^2} \right)^{3/2} m_n \quad (4.204)$$

where m_n is the mass of the nucleons and α is a inhomogeneity correction factor of order unity. This is roughly 1.44 solar masses. See Greiner's textbook, example 14.5 p. 359 for the full calculation.

9. The formulae in this chapter make large use of the constants k , \hbar and c . A system of units, called *natural units*, can be defined in which temperature are expressed in energy units, velocities are expressed in units of c and actions (= energy \times time) in units of \hbar . This would then be very convenient, because in this system one has by definition $k = c = \hbar = 1$. Of the three standard units (mass, length, time) only one (or a combination) would then be needed to express all physical quantities, and very often the energy unit (in electronvolt) is chosen. For instance, a length is converted in units of 1/energy by multiplying it by $1/\hbar c$; an unit of time also acquires dimensions of 1/energy when multiplied by $1/\hbar$, etc.. Finally, if one uses also units in which $G = 1$, called Planck units, then all physical quantities become dimensionless.

Chapter 5

Phase transitions

5.1 General properties of phase transitions

1. Real systems display behavior much more complicated than those that have been studied so far. The most dramatic example is the so-called phase transition, i.e. a rapid departure from one state of equilibrium in the transition to another one. Phase transitions assume many different forms: condensation of a gas into a liquid, solidification, boiling, spontaneous magnetization, superfluidity, etc.
2. Although very different phenomena, the phase transitions have also some common “universal” aspects: in particular, they all display a discontinuity in some quantity and they all show the establishment of long-range order in a previously disordered system. For instance, in the process of condensation of a gas through volume compression, the density changes abruptly at the boundary between the gas and its liquid phase and the compressibility at constant temperature becomes infinite (the volume of the system gas+liquid is reduced but the pressure does not change). In a ferromagnetic substance, the residual magnetization that remains when the external field is switched off changes sign depending on whether the external field is reduced to zero from the positive or the negative side, etc.
3. At the same time, the systems develop some form of long-range order or correlation: larger and larger regions of liquid phase form, or the individual spins of the ferromagnetic substance align spontaneously. It is clear that a necessary condition to achieve long-range correlations is that the particles interact in some way. In fact, the stronger the interaction is, the more likely a system undergoes a phase transition for a given temperature.
4. Phase transitions develop normally only below some critical temperature T_c : the long-range correlation is destroyed at high temperatures. For instance, a substance may develop ferromagnetic properties below some temperature (in this case called Curie temperature T_{Curie}) and paramagnetic (i.e. no residual or spontaneous magnetization) above it.
5. Phase transitions are sometimes called *first order or discontinuous* if an intensive quantity called order parameter (e.g. density for gas-liquid condensation, specific magnetization for a ferromagnet) changes discontinuously from a phase to another one. They are called *second order or continuous* if the quantity changes continuously but its derivative is discontinuous (i.e., the order parameter has a spike).
6. A more modern classification is based on the existence of *latent heat*, i.e. the amount of heat that is absorbed or released during the transition while the temperature remains constant (and therefore the heat capacity is infinite). The latent heat is non-zero if the heat goes into the new phase (e.g. from the liquid to the gas) rather than into the kinetic energy of the old phase. If there is latent heat, the transition is called first order or *discontinuous*, otherwise second order or *continuous*. So the most typical example of first order transition is the liquid-gas transition below critical temperature (where phases coexist), while ferromagnetism is an example of second order transition, as is also liquid-gas at or above the critical temperature.

7. The study of phase transitions is one of the most active area of research in statistical mechanics. The general goals are to find universal relations that apply to many different systems and to approach the description of real systems. In most cases, the problem is tractable only numerically through simulations. There are only a few analytical result available, two of which we will review next, namely the condensation in real gases and the Ising model.

5.2 Gas with interacting particles

1. If the molecules of a gas interact, the Hamiltonian will include a potential

$$H = \sum_i \frac{p_i^2}{2m} + \sum_{i < k} U_{ik}(|\vec{r}_i - \vec{r}_k|) \quad (5.1)$$

The classical canonical partition function can be integrated over the momenta as before

$$Z(T, V, N) = \frac{1}{N! h^{3N}} \left(\frac{2\pi m}{\beta} \right)^{3N/2} Q(V, T) \quad (5.2)$$

where we still have to evaluate

$$Q(V, T) = \int d^{3N} r \prod_{i < k} \exp(-\beta U_{ik}) \quad (5.3)$$

which contains the space part.

2. For an ideal gas, $\exp(-\beta U_{ik}) \rightarrow 1$. A first order deviation from ideal systems can be therefore obtained by expanding in

$$f_{ik} = \exp(-\beta U_{ik}) - 1 \quad (5.4)$$

Then we obtain

$$\prod_{i < k} (1 + f_{ik}) \approx 1 + \sum_{i < k} f_{ik} + \sum'_{i, k, m, \ell} f_{ik} f_{m\ell} + \dots \quad (5.5)$$

(whenever we put a prime on a sum it means the sum has to obey some additional constraint). In this section we consider only the first two terms, i.e. we put $f_{ik} \ll 1$. A systematic way to perform the expansion term by term is to use Mayer's cluster method, see for instance Pathria, *Statistical Mechanics*.

3. The integral Q is now

$$Q(V, T) \approx V^N + V^{N-2} \sum_{i < k} \int d^3 r_i \int d^3 r_k [\exp(-\beta U_{ik}) - 1] \quad (5.6)$$

If we assume that the potential U_{ik} has the same form U for all pairs, we can simplify the integration using relative coordinates $\vec{r} = \vec{r}_i - \vec{r}_j$ and center-of-mass coordinates $\vec{R} = (\vec{r}_i + \vec{r}_j)/2$. These latter can then be integrated over and we obtain

$$Q(V, T) = V^N + V^{N-1} \frac{N(N-1)}{2} \int d^3 r [\exp(-\beta U(r)) - 1] \quad (5.7)$$

(the factor $N(N-1)/2 \approx N^2/2$ counts the number of pairs with $i < k$). Finally, we put

$$a(\beta) = \int d^3 r [\exp(-\beta U(r)) - 1] \quad (5.8)$$

$$= 4\pi \int r^2 dr [\exp(-\beta U(r)) - 1] \quad (5.9)$$

so that finally

$$Z(T, V, N) \approx \frac{1}{N!} \left(\frac{2\pi mkT}{h^2} \right)^{3N/2} \left(V^N + V^{N-1} \frac{N^2}{2} a(T) \right) \quad (5.10)$$

$$= \frac{1}{N!} \left(\frac{V^N}{\lambda^{3N}} \right) \left(1 + \frac{N^2}{2V} a(T) \right) \quad (5.11)$$

4. This gives a first correction over the ideal gas EOS in the limit of small a :

$$p = -\frac{\partial F}{\partial V}|_{T,N} = \frac{\partial}{\partial V}(kT \log Z) = \frac{NkT}{V} - kT \frac{\frac{a}{2} \frac{N^2}{V^2}}{1 + \frac{a}{2} \frac{N^2}{V^2}} \quad (5.12)$$

$$\approx \frac{NkT}{V} \left(1 - \frac{a}{2} \frac{N}{V}\right) \quad (5.13)$$

5. Various potentials have been proposed to model the strong repulsion for small distances and the attraction for larger distances that occur for real gas molecules. A simple analytical result can be obtained with the Sutherland potential (Fig. 5.1), which is attractive as $U = -U_0(r_0/r)^6$ up to the small distance r_0 , below which is infinitely repulsive (solid spheres). Then we have (we assume $\beta U_0 \ll 1$)

$$a(\beta) = \int d^3r [\exp(-\beta U(r)) - 1] \quad (5.14)$$

$$= 4\pi \int_0^{r_0} r^2 dr [-1] + 4\pi \int_{r_0}^{\infty} r^2 dr [\exp(\beta U_0(r_0/r)^6) - 1] \quad (5.15)$$

$$\approx -\frac{4\pi}{3} r_0^3 + 4\pi \beta U_0 \int_{r_0}^{\infty} r^2 dr (r_0/r)^6 \quad (5.16)$$

$$= -\frac{4\pi}{3} r_0^3 (1 - \beta U_0) \quad (5.17)$$

from which the EOS becomes

$$p = \frac{kT}{v} \left(1 + \frac{2\pi r_0^3}{3v} \left(1 - \frac{U_0}{kT}\right)\right) \quad (5.18)$$

where $v = V/N$, i.e. the inverse of the number density. Since the correction terms are supposed to be small, this expression can be transformed as

$$p + \frac{2\pi r_0^3}{3v^2} U_0 = \frac{kT}{v} \left(1 + \frac{2\pi r_0^3}{3v}\right) \approx \frac{kT}{v} \left(1 - \frac{2\pi r_0^3}{3v}\right)^{-1} \quad (5.19)$$

and finally

$$\left(p + \frac{2\pi r_0^3}{3v^2} U_0\right) \left(v - \frac{2\pi r_0^3}{3}\right) = kT \quad (5.20)$$

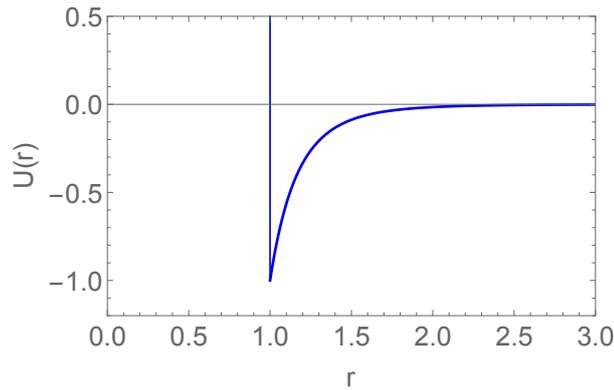
i.e. just like the Van der Waals EOS $(p + a/v^2)(v - b) = kT$, with

$$a = \frac{2\pi r_0^3}{3} U_0 \quad (5.21)$$

$$b = \frac{2\pi r_0^3}{3} \quad (5.22)$$

where we see that, up to constants, a measures the strength of the interaction and b measures the volume of the particles.

6. Due to the various limiting assumptions, in particular the assumption of constant density, the isothermals of the Van der Waals potential display an unphysical behavior, since they contain regions where the compressibility is negative, i.e. the slope of $p(V)$ is positive (e.g. the interval $d-e$ in Fig. 5.2). This would cause the system to go irreversibly out of equilibrium, instead of performing a phase transition. Although a physical behavior can be obtained from first principles through an exact or numerical evaluation, the so-called Maxwell construction (horizontal lines in the pV plane such that the area below is the same as for the Van der Waals isotherm, e.g. the segment $a-c$ in Fig. 5.2) gives a reasonable approximation in a much easier way.
7. The Van der Waals EOS, corrected through the Maxwell construction, gives a qualitatively accurate description of phase transitions, including the existence of a critical temperature, of co-existent phases and of discontinuities.

Figure 5.1: Sutherland potential $U(r)$ with $r_0 = 1$ (arbitrary units)

8. The critical temperature occurs at the inflection point of $p(v)$, i.e. for $\partial p/\partial v = \partial^2 p/\partial v^2 = 0$, i.e. for

$$p_c = \frac{a}{27b^2}, \quad v_c = 3b, \quad T_c = \frac{8a}{27b} \quad (5.23)$$

The ratio

$$K = \frac{kT_c}{P_c v_c} = 8/3 \quad (5.24)$$

is then a universal quantity, regardless of the particular gas. For an ideal gas, $K = 1$, so K can be seen as a deviation from ideal gas condition. The Van der Waals equation can then be written in a universal or reduced form

$$\left(p_r + \frac{3}{v_r^2}\right)(3v_r - 1) = 8T_r \quad (5.25)$$

where $p_r = p/p_c$, $v_r = v/v_c$, $T_r = T/T_c$ are all dimensionless quantities written in units of the critical values. This expression is called “*law of corresponding states*”, because different systems have corresponding (i.e. universal) forms. It is of course only an approximation to real behavior.

9. The equation can be written as a third order polynomial in v_r :

$$3p_r v_r^3 - (p_r + 8T_r)v_r^2 + 9v_r - 3 = 0 \quad (5.26)$$

On any isotherm ($T_r = \text{const}$) one has then three roots for v_r with constant p_r lines. In general, they can all three be real, or one real and two complex. More in detail, one can easily confirm that for some temperatures (lower than T_c , i.e. $T_r < 1$) there are three real roots, while above T_c only one.

10. Above the critical temperature, the fluid is called supercritical and shares with continuity (i.e., without a sharp phase transition) properties of liquids (at high pressures) and of gases (at low pressures). For water, the critical values are $T_c = 647\text{K}$, $p_c = 218 \text{ atm}$, $v_r = 0.0032 \text{ m}^3/\text{kg}$. As it can be seen, the pressure is very high, as with most other substances (Helium is an exception).
11. Near criticality, we can put $p_r = 1 + \pi$, $v_r = 1 + \psi$, $T_r = 1 + t$ and study the behavior for small deviations from 1 when moving along various directions. Then we obtain a polynomial of third order in ψ :

$$3\psi^3 + 8(\pi - t)\psi^2 + (7\pi - 16t)\psi + 2(\pi - 4t) = 0 \quad (5.27)$$

where we just neglected a higher order term $3\psi^3\pi$. On the critical isotherm $t = 0$ and at the lowest order in π, ψ we obtain

$$\pi \approx -\frac{3}{2}\psi^3 \quad (5.28)$$

which shows how flat is the critical isotherm around criticality.

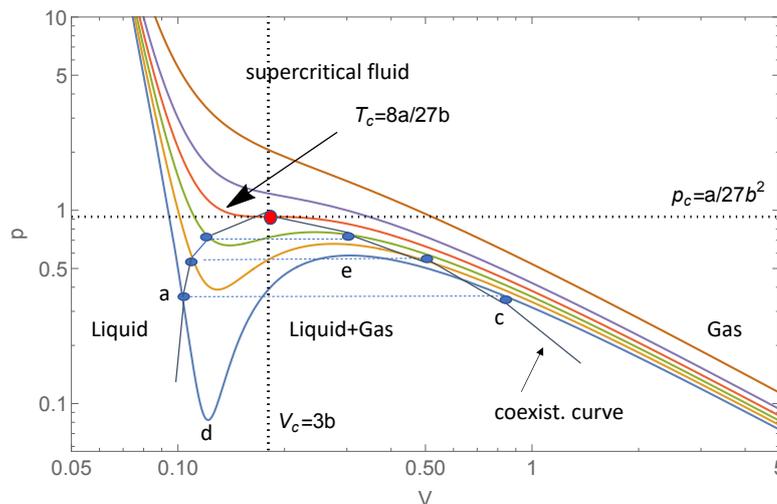


Figure 5.2: Van der Waals isotherms. The liquid+gas phase is bounded by the coexistence curve. The horizontal lines like $a - c$ represent the Maxwell constructions, while the interval $d - e$ is the unphysical region. The big red dot is the critical point.

12. We then consider moving near criticality along $\pi \approx 4t$. Inserting this equivalence in (5.27) we get three roots for ψ , one at $\psi = 0$ and (neglecting the higher order term $8(\pi - t)\psi = -24t\psi$)

$$\psi_{1,2} = \pm 2|t|^{1/2} \quad (5.29)$$

where the upper sign is for the gas phase (higher ψ , i.e. higher $v = 1/n$, and therefore smaller density) while the lower sign is for the liquid phase. The curve $\psi^2 = -4t$ is the inverted parabola that delimits (near criticality) the coexistence region, and is called *coexistence curve*.

13. Along the coexistence curve (and therefore for $T < T_c$) we see that the difference of number densities of the liquid and the gas is given by

$$n_\ell - n_g = \frac{1}{v_\ell} - \frac{1}{v_g} = \frac{1}{v_c(1 + \psi_\ell)} - \frac{1}{v_c(1 + \psi_g)} \approx \frac{1}{v_c}(\psi_g - \psi_\ell) = \frac{2}{v_c\sqrt{T_c}}|T - T_c|^{1/2} \quad (5.30)$$

while in the $T > T_c$ regime one has obviously $n_\ell = n_g$. The parameter $\delta n = n_\ell - n_g$ can be called a *order parameter* of the liquid-gas phase transition: the transition occurs when varying the order parameter, in particular when it goes from zero (gas = disordered phase) to a finite value (liquid = more ordered phase). One has then

$$\delta n \sim |T - T_c|^{1/2} \quad T < T_c \quad (5.31)$$

$$\delta n = 0, \quad T \geq T_c \quad (5.32)$$

The exponent $\beta = 1/2$ is called *critical exponent* (see next section) and characterizes this kind of transition.

14. We can also estimate the dimensionless isothermal compressibility $K_T \equiv p_c\beta_T = -p_cV^{-1}\partial V/\partial p|_{T,N} \sim -\partial\psi/\partial\pi|_t$. From Eq. (5.27) we have

$$\pi = \frac{8t + 16t\psi + 8t\psi^2 - 3\psi^3}{(1 + \psi)^2(2 + 3\psi)} \quad (5.33)$$

Then we obtain

$$K_T = -\frac{\partial\psi}{\partial\pi}\Big|_t = -\left(\frac{\partial\pi}{\partial\psi}\right)_t^{-1} \approx \frac{(1+\psi)^3(2+3\psi)^2}{6(4t(1+\psi)^3 + \psi^2(3+4\psi))} \quad (5.34)$$

Now moving along $\psi = 0$ for $T \geq T_c$ i.e. $t > 0$ we obtain

$$K_T \approx \frac{1}{6t} \quad (5.35)$$

but if we approach the critical point along the coexistence curve for which (5.29) is valid, then for $t < 0$ we get

$$K_T \approx \frac{1}{12|t|} \quad (5.36)$$

The different behavior shows a discontinuity as $t \rightarrow 0$, again a fundamental property of phase transitions. So, analogously to δn , we have now

$$K_T \sim \alpha_1 |T - T_c|^{-1} \quad T < T_c \quad (5.37)$$

$$K_T \sim \alpha_2 |T - T_c|^{-1}, \quad T \geq T_c \quad (5.38)$$

with different positive coefficients. The exponents $\gamma_{\pm} = -1$ of the two laws are again called critical exponents. The various laws just found also show then another typical property of phase transitions, namely the existence of power laws in the vicinity of the critical point.

15. More details about the Van der Waals phase transitions in Pathria, *Statistical Mechanics*, Chap. 10 and 12.

5.3 Critical exponents

1. As already mentioned, one of the goal of the study of phase transition is to find universal properties near the transition, independent of the details of the systems. Since the transitions are always characterized by singularities or discontinuities, one expect to be able to describe them with inverse or fractional power laws. The exponents of these power laws are called critical exponents and are found to be rather general.
2. In any phase transition one can identify one (or more) quantity that undergoes one of these singularities/discontinuities when a conjugate field is varied. In a liquid-gas transition, for instance, the pressure and volume are such a pair: the pressure is a smooth function of the volume before or after the transition, but the compressibility becomes infinite during the coexistence.
3. We define *order parameter* m a quantity that undergoes a discontinuity (in the function itself or some of its derivatives) when its conjugate field (*ordering field*) h vanishes. For a liquid-gas condensation, the order parameter is the density difference, as we have already seen, and the ordering field is the pressure. (Note that as long as the number of particles stays constant, volume and number density are interchangeable quantities.) We take the number density difference $m = \delta n = n_\ell - n_g$, and identify the ordering field as $h = p - p_c$. Then we see indeed that for $h \rightarrow 0$, m is such that it is zero for $T \geq T_c$ (homogeneous gas phase) while it is positive for $T < T_c$. In each phase transitions one can identify such ordering parameters and fields. For ferromagnetism, for instance, m is the average magnetization and h the external field.
4. The derivative

$$\chi_0 = \left(\frac{\partial m}{\partial h}\right)_{T; h \rightarrow 0} \quad (5.39)$$

is called in general *susceptibility* (for the liquid-gas systems, it would be in fact the isothermal compressibility 5.35). It represents the response of the system to an external action. During the phase transition the susceptibility becomes singular for $T \rightarrow T_c$.

Critical exponents	Magnetic systems	Gas-liquid systems	Binary fluid mixtures	Binary alloys	Ferroelectric systems	Superfluid He ^{4(f)}	Mean field results
α, α'	0.0-0.2	0.1-0.2	0.05-0.15	–	–	-0.026	0
β	0.30-0.36	0.32-0.35	0.30-0.34	0.305±0.005	0.33-0.34	–	1/2
γ	1.2-1.4	1.2-1.3	1.2-1.4	1.24±0.015	1.0±0.2	inaccessible	1
γ'	1.0-1.2	1.1-1.2	–	1.23±0.025	1.23±0.02	inaccessible	1
δ	4.2-4.8	4.6-5.0	4.0-5.0	–	–	inaccessible	3
ν	0.62-0.68	–	–	0.65±0.02	0.5-0.8	0.675	1/2
η	0.03-0.15	–	–	0.03-0.06	–	–	0

Table 5.1: Table of experimental critical exponents (entries from Pathria, Statistical Mechanics)

5. The behavior of the ordering parameters/fields and of the susceptibility near a transition is characterized by power laws in function of temperature or of the field h , as in Eq. (5.29) or (5.35). This lets us introduce the following critical exponents:

$$m \sim (T_c - T)^\beta, (h \rightarrow 0, T \leq T_c) \quad (5.40)$$

$$\chi_0 \sim \begin{cases} (T - T_c)^{-\gamma}, (h \rightarrow 0, T \geq T_c) \\ (T_c - T)^{-\gamma'}, (h \rightarrow 0, T \leq T_c) \end{cases} \quad (5.41)$$

$$m|_{T=T_c} \sim h^{1/\delta}. (h \rightarrow 0, T = T_c) \quad (5.42)$$

The last expression corresponds to Eq. (5.28) for the liquid-gas case.

6. Finally, we define two other critical exponents based on the heat capacity C_V :

$$C_V \sim \begin{cases} (T - T_c)^{-\alpha}, T \geq T_c \\ (T_c - T)^{-\alpha'}, T \leq T_c \end{cases} \quad (5.43)$$

In case the exponent is very small, one can have instead a logarithmic behavior, e.g. $C_V \sim \log(1/|T - T_c|)$ or a behavior that is better fitted by a function $f \sim (|t|^{-\lambda} - 1)/\lambda$ (where $t = T - T_c$). The derivative df/dt would be in any case a power law.

7. For a gas-liquid transition we have

$$\beta = \frac{1}{2}, \gamma = \gamma' = 1, \delta = 3, \alpha = \alpha' = 0 \quad (5.44)$$

(we have not demonstrated the C_V behavior for the liquid-gas transition; see Pathria's textbook for more information). Remarkably, the same values are found for magnetic systems in the so-called mean-field approximation. Although these specific values are not a very good approximation to real cases, the fact that different systems possess similar critical exponents is indeed verified experimentally (see Tab. 5.1).

8. The critical exponents are not completely independent. Some general inequality among them have been shown to exist, like the Ruskbrooke inequality

$$\alpha' + 2\beta + \gamma' \geq 2 \quad (5.45)$$

9. The estimation of the critical exponents from first principles is an important goal of research in statistical physics.

5.4 Ising model

1. In this section we investigate a relatively simple model of phase transitions, the Ising model. This model is a schematical approximation of a system of interacting magnetic dipoles but with minor modifications can also be employed to model binary alloys, spin glasses and other many-particle systems.
2. The Ising model, investigated for the first time by Lenz and Ising in the 20's, consists in N magnetic dipoles fixed on a regular lattice at locations indexed by i which can take the two spin values $\sigma_i = \pm 1$ and interact through the Hamiltonian

$$H_J = -J \sum_{nn} \sigma_i \sigma_k \quad (5.46)$$

where the sum extends only to the nearest neighbors. The idea is that the interaction over larger distances can be neglected. Assuming a constant $J > 0$, the neighboring spin will tend to have the same sign (i.e. to align themselves), since then the energy is minimized.

3. The number of nearest neighbors depend of course on the space dimensionality. They are 2 in 1 dimension, 4 in 2 dimensions, and $2d$ in general (for a cubic lattice).
4. On top of the interaction Hamiltonian there is also the magnetic energy in presence of an external magnetic field H that we already studied in Sect. (3.11)

$$H_0 = - \sum_i H \mu \sigma_i \quad (5.47)$$

where μ is the magnetic dipole.

5. The magnetization

$$\langle M \rangle = \sum_i \mu \langle \sigma_i \rangle \quad (5.48)$$

that remains when the external field is switched off is called spontaneous or residual or permanent magnetization and the material is called *ferromagnetic*. A paramagnetic behavior, as we have seen, occurs when no spontaneous magnetization develops (either because there is no interaction or because it is not sufficient to create spontaneous magnetization). The onset of spontaneous magnetization below a certain temperature T_{Curie} is an example of phase transitions.

6. The Ising model allows to find a relation between $\langle M \rangle$ and temperature. Such a relation is analytical and exact in 1 (Ising) or 2 (Onsager) dimensions, and approximated or numerical beyond this.
7. In one dimension, the partition function of a Ising model (that we denote Q in this section) is relatively simple since there are only two neighbors for each spin

$$Q(\beta, N, H) = \sum_{spins} e^{-\beta(H_0 + H_I)} \quad (5.49)$$

$$= \sum_{\sigma_1 = \pm 1} \sum_{\sigma_2 = \pm 1} \dots \exp(\beta \mu H \sum_i \sigma_i + \beta J \sum_i \sigma_i \sigma_{i+1}) \quad (5.50)$$

This can be written in a symmetric form if we identify σ_{N+1} with σ_1 (periodic boundary conditions). Then we have in fact

$$Q(\beta, N, H) = \sum_{\sigma_1 = \pm 1} \sum_{\sigma_2 = \pm 1} \dots \exp(\beta \mu H \sum_i \sigma_i + \beta J \sum_i \sigma_i \sigma_{i+1}) \quad (5.51)$$

$$= \sum_{\sigma_1 = \pm 1} \sum_{\sigma_2 = \pm 1} \dots \exp \left[\beta \sum_i \left(\frac{1}{2} \mu H (\sigma_i + \sigma_{i+1}) + J \sigma_i \sigma_{i+1} \right) \right] \quad (5.52)$$

8. This sum can be calculated in various ways. One that makes use of *transfer matrices* is the following. Let us represent the two states spin up and spin down by the two orthonormal vectors $|\sigma_a\rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$ and $|\sigma_b\rangle = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$. Then we have by completeness

$$\sum_{k=a,b} |\sigma_k\rangle\langle\sigma_k| = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = I \quad (5.53)$$

(identity matrix) for every site i .

9. Now let us define an operator \hat{P} acting on the $|\sigma\rangle$ space such that

$$\langle\sigma_i|\hat{P}|\sigma_{i+1}\rangle = \exp\beta \left[J\sigma_i\sigma_{i+1} + \frac{1}{2}\mu H(\sigma_i + \sigma_{i+1}) \right] \quad (5.54)$$

The explicit form of the operator is

$$\hat{P} = \begin{pmatrix} \exp(\beta(J + \mu H)) & \exp(-\beta J) \\ \exp(-\beta J) & \exp(\beta(J - \mu H)) \end{pmatrix} \quad (5.55)$$

as it can be seen by explicit evaluation. For instance, if both spin at sites i and $i + 1$ are up, one has

$$\langle\sigma_{i,a}|\hat{P}|\sigma_{i+1,a}\rangle = (1, 0)\hat{P} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = e^{\beta(J+\mu H)} \quad (5.56)$$

which indeed is (5.54) for $\sigma_i = \sigma_{i+1} = 1$. Then we have

$$Q(\beta, N, H) = \sum_{\sigma_1=\pm 1} \sum_{\sigma_2=\pm 1} \dots \exp \left[\beta \sum_i^N \left(\frac{1}{2}\mu H(\sigma_i + \sigma_{i+1}) + J\sigma_i\sigma_{i+1} \right) \right] \quad (5.57)$$

$$= \sum_{\sigma_1=\pm 1} \sum_{\sigma_2=\pm 1} \dots \langle\sigma_1|\hat{P}|\sigma_2\rangle\langle\sigma_2|\hat{P}|\sigma_3\rangle\dots\langle\sigma_N|\hat{P}|\sigma_1\rangle \quad (5.58)$$

$$= \sum_{\sigma_1=\pm 1} \langle\sigma_1|\hat{P}^N|\sigma_1\rangle = \text{Tr}\hat{P}^N \quad (5.59)$$

(employing (5.53)).

10. The Trace of \hat{P} can be calculated by summing its eigenvalues. The eigenvalues $\lambda_{1,2}$ are obtained by the equation $\det(\hat{P} - \lambda I) = 0$, i.e.

$$\lambda^2 - 2\lambda \exp(\beta J) \cosh(\beta\mu H) + 2 \sinh(2\beta J) = 0 \quad (5.60)$$

from which

$$\lambda_{1,2} = e^y \left[\cosh x \pm \sqrt{\sinh^2 x + e^{-4y}} \right] \quad (5.61)$$

where $x = \beta\mu H$ and $y = \beta J$. The Trace of \hat{P}^N is then simply $Q = \lambda_1^N + \lambda_2^N$ and we obtain the free energy

$$F(N, H, T) = -kT \log (\lambda_1^N + \lambda_2^N) \quad (5.62)$$

11. In the case without interaction, $J = 0$, we should obtain the paramagnetic behavior (3.198). We have in fact in this case $\lambda_1 = 2 \cosh \beta\mu H$ and $\lambda_2 = 0$ and therefore

$$F(N, H, T)_{J=0} = -NkT \log (2 \cosh \beta\mu H) \quad (5.63)$$

When instead we switch off the external field H we obtain $\lambda_1 = 2 \cosh \beta J$, $\lambda_2 = 2 \sinh \beta J$. Since $\lambda_1 > \lambda_2$, for large N the term λ_2^N is negligible and we obtain

$$F(N, 0, T) = -NkT \log (2 \cosh \beta J) \quad (5.64)$$

12. The total magnetic moment is given by

$$\langle M \rangle = -\frac{\partial F}{\partial H}|_{N,T} = -\beta\mu \frac{\partial F}{\partial x}|_{N,T} \quad (5.65)$$

$$= -N\mu \frac{\lambda_1^{N-1}\lambda'_1 + \lambda_2^{N-1}\lambda'_2}{\lambda_1^N + \lambda_2^N} \quad (5.66)$$

$$= \frac{N\mu \sinh x}{(e^{-4y} + \sinh^2 x)^{1/2}} \frac{\lambda_1^N - \lambda_2^N}{\lambda_1^N + \lambda_2^N} \quad (5.67)$$

where we used Eq. (5.61) to obtain

$$\lambda'_{1,2} \equiv \frac{d\lambda_{1,2}}{dx} = \frac{\sinh x}{\sqrt{\sinh^2 x + e^{-4y}}} (\pm \lambda_{1,2}) \quad (5.68)$$

(+ for λ_1 and $-$ for λ_2).

13. Again neglecting λ_2^N we obtain simply

$$\langle M \rangle = \frac{N\mu \sinh x}{(e^{-4y} + \sinh^2 x)^{1/2}} \quad (5.69)$$

For $J \rightarrow 0$, i.e. $y \rightarrow 0$, we obtain $M = N\mu \tanh x$. It appears that $M \rightarrow 0$ for $x \rightarrow 0$, which implies that there is no spontaneous magnetization at finite temperature when the external field is switched off. The one-dim Ising model is indeed not ferromagnetic.

14. From these expressions also other quantities like the average spins up or down N_{\pm} can be evaluated. One can in fact write

$$\langle M \rangle = \mu \sum_i \sigma_i = \mu(N_+ - N_-) \quad (5.70)$$

where N_{\pm} are the average number of spins up or down. Then, since $N = N_+ + N_-$, we have

$$N_{\pm} = \frac{1}{2} \left(N \pm \frac{\langle M \rangle}{\mu} \right) \quad (5.71)$$

and therefore

$$N_{\pm} = \frac{1}{2} N \left(1 \pm \frac{\sinh x}{[e^{-4y} + \sinh^2 x]^{1/2}} \frac{\lambda_1^N - \lambda_2^N}{\lambda_1^N + \lambda_2^N} \right) \quad (5.72)$$

As expected, in the limit of high temperature $x \rightarrow 0$ there is an equal number of spins up and down. In the limit of high external field, $x \rightarrow \infty$, all spins are aligned.

15. One can also estimate the spin correlation, i.e. $\langle \sigma_i \sigma_j \rangle$ and from this quantities like N_{++} or N_{--} , i.e. the average number of neighboring pairs with positive or negative spin. It is interesting to note that even in absence of external field one finds that there is a non-zero correlation among spins due to their self-interaction.
16. The Ising model is not ferromagnetic only in one dimension. The reason is that in one dimension each spin has only two neighbors and this interaction is too weak to provide the strong correlation needed to produce a phase transition. The energy difference between a completely spin-up system with maximal magnetization ($-NJ$) and a system with the first half spin up and the second half down, for a net vanishing total magnetization ($(-N+4)J$), is very small, and there is no energetic advantage in a phase transition.
17. In any dimension larger than one, analytical or numerical results show that there is always a residual magnetization, i.e. a discontinuity in M for $H \rightarrow 0$ below some temperature (Curie temperature).
18. For 2 dimensions the general solution has been found by Onsager in the 40s:

$$Q(\beta, N, H = 0) = (2 \cosh(\beta J) e^J)^N \quad (5.73)$$

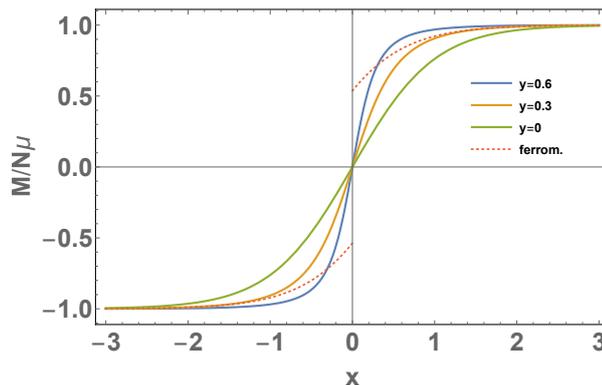


Figure 5.3: Magnetization of the 1D Ising model, for increasing interaction strength y . The dotted line represents a ferromagnetic transition.

where

$$I = (2\pi)^{-1} \int_0^\pi d\phi \log\left[\frac{1}{2}(1 + (1 - \kappa^2 \sin^2(\phi))^{1/2})\right] \quad (5.74)$$

with $\kappa = 2 \sinh(2\beta J) / \cosh^2(2\beta J)$. It can be shown that a phase transition occurs here for any $T < T_C = 2.269J/k$, where T_C is the root of $\sinh 2\beta J = 1$.

19. The energy of the system is as usual (we consider all quantities constant except T)

$$U = -kT \frac{d \log Q}{d\beta} = kT^2 \frac{d \log Q}{dT} \quad (5.75)$$

20. The heat capacity is singular near the phase transition:

$$C = \frac{\partial U}{\partial T} \Big|_{H=0} = \frac{2}{T} U + kT^2 \frac{d^2 \log Q}{dT^2} \approx kT^2 \frac{d^2 \log Q}{dT^2} \sim -\log |T - T_C| \quad (5.76)$$

(near T_c , the regular first term in C is negligible with respect to the singular second term) and the magnetization for $T < T_C$ is

$$M/N \sim (T_C - T)^\beta \quad (5.77)$$

with $\beta \approx 1/8$, while of course $M = 0$ for $T > T_c$ (see Fig. 5.4). Notice how this behavior is qualitatively similar to the liquid-gas order parameter Eq. (5.32).

21. The exponents of these power laws are the critical exponents introduced in the previous section. In three dimensions, the specific heat and the specific magnetization behave like $C/N \sim |T - T_C|^\alpha$, $M/N \sim (T_C - T)^\beta$, with $\alpha \approx 0.125$ and $\beta \approx 0.313$.
22. The Ising model can help investigating various kinds of phase transitions. For instance, one can think of a system of particles in a container. If we divide up the volume into cells so small that at most they contain either zero or one particle, then we can associate a state up (cell is occupied) or down (cell is empty) with each cell. When two particles are in nearest-neighbor cells, the system acquires an energy $-\varepsilon$, which means the system goes into a preferred, lower energy state if the particles are close together. This reproduces an interaction among particles. Now the system is completely equivalent to the Ising model and can be studied with the same means, reproducing a phase transitions for certain values of temperature $1/\beta$ and of interaction energy ε .

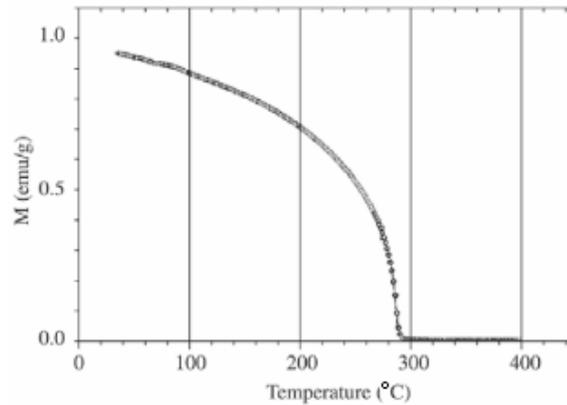


Figure 8. Magnetization-temperature curve of the powder calcined at 1450 °C when subsequently subjected to a 240 Oe magnetic field.

Figure 5.4: Magnetization versus temperature of a garnet ferrite, showing the typical transition from paramagnetic to ferromagnetic behavior (from Caffarena and Ogasawara, *Mat. Res.* [online]. 2003, vol.6, n.4, pp. 569-576).

5.5 The Renormalization Group

This section follows very closely the treatment in Chandler, *Introduction to Modern Statistical Mechanics*, to which we refer for further details.

1. One important technique for solving the Ising model is the Renormalization Group (RG) method. We can illustrate the main points of this method already in one dimension.
2. Let us start with Eq. (5.52) for $H = 0$

$$Q(K, N) = \sum_{\sigma_1, \sigma_2, \sigma_3, \dots} \exp K(\sigma_1 \sigma_2 + \sigma_2 \sigma_3 + \sigma_3 \sigma_4 + \dots) \quad (5.78)$$

where $K = J/kT$ (previously called y) and write it as

$$Q(K, N) = \sum_{\sigma_1, \sigma_2, \dots} \exp K(\sigma_1 \sigma_2 + \sigma_2 \sigma_3) \exp K(\sigma_3 \sigma_4 + \sigma_4 \sigma_5) \dots \quad (5.79)$$

3. Now we can integrate over all the even numbered spins $\sigma_2, \sigma_4, \dots$:

$$Q(K, N) = \sum_{\sigma_1, \sigma_3, \dots} [\exp K(\sigma_1 + \sigma_3) + \exp -K(\sigma_1 + \sigma_3)][\exp K(\sigma_3 + \sigma_5) + \exp -K(\sigma_3 + \sigma_5)] \dots \quad (5.80)$$

In this way we have removed half of the spins. If we could recast this expression in a form similar to (5.78), but with possibly different K, N , then we would have obtained a recursive relation that could perhaps be solved. That is, we need to be able to write *for every* σ, σ' an expression as the following one

$$\exp K(\sigma + \sigma') + \exp -K(\sigma + \sigma') = f(K) e^{K' \sigma \sigma'} \quad (5.81)$$

If this is possible, then we obtain

$$Q(K, N) = f(K)^{N/2} Q(K', N/2) \quad (5.82)$$

since now only $N/2$ spins are left. A transformation that achieves this is called a *Kadanoff transformation*.

4. If σ, σ' are both either $+1$ or -1 , we have

$$e^{2K} + e^{-2K} = f(K)e^{K'} \quad (5.83)$$

and if they have opposite signs then

$$2 = f(K)e^{-K'} \quad (5.84)$$

These are two conditions in two unknowns, $f(K)$ and K' . The solution is

$$K' = \frac{1}{2} \log \cosh 2K \quad (5.85)$$

$$f(K) = 2 \cosh^{1/2} 2K \quad (5.86)$$

Then we can see that we can form the recursion relation

$$\log Q = \frac{N}{2} \log f(K) + \log Q(K', N/2) \quad (5.87)$$

or by putting $\log Q = Ng(K)$

$$g(K) = \frac{1}{2} \log f(K) + \frac{1}{2}g(K') \quad (5.88)$$

or

$$g(K') = 2g(K) - \log[2\sqrt{\cosh(2K)}] \quad (5.89)$$

5. Eq. (5.89) along with the Kadanoff transformation provide a recursive way to renormalize the partition function. Since the recursive property can be seen as forming a group, the scheme is called *renormalization group*.

6. Now $Q(K, N)$ can be evaluated by iterating the recursion. We can start by inverting (5.85)

$$K = \frac{1}{2} \cosh^{-1} e^{2K'} \quad (5.90)$$

and, inverting (5.84) and inserting it in (5.88),

$$g(K) = \frac{1}{2}[g(K') + \log 2 + K'] \quad (5.91)$$

We begin the iteration with a very small K' eg $K' = 0.01$. Since K' is very small, we can approximate $Q(K', N) \approx Q(0, N) = \sum_{spins} 1 = 2^N$ (i.e. no interaction). Now by definition $g(K') = (1/N) \log Q(K', N)$ and therefore $g(K') = \log 2$. Inserting K' and $g(K')$ in Eqs. (5.90) and (5.91) we obtain $K = 0.100334$ and $g(K) = 0.698$. Next, we can use these as new primed values of $K', g(K')$ and iterate. After just a few iterations, one obtains results in agreement with the exact values of the previous section (for $H = 0$), i.e.

$$g(K) \rightarrow \frac{\log Q}{N} \Big|_{H=0} = \frac{\log \lambda_1^N + \lambda_2^N}{N} \Big|_{H=0} \approx \log \lambda_1 = \log (2 \cosh \beta J) \quad (5.92)$$

7. The iteration proceeds until it reaches one of the fixed points for K , where any new iteration does not change the result. For the 1D Ising model, the fixed points are at 0 and ∞ ; at these fixed points the temperature is either zero or infinity: these are called trivial fixed point. The iteration process can be seen as a continuous change in scale, i.e. viewing the system averaged over a larger and larger scale. When we reach a fixed point, it means that a change in scale does not change the properties of the system. Since this “scale-invariance” is a property of the phase transitions associated with the existence of long-range correlations, we can conclude that the existence of a fixed point at finite temperature signals the existence of a phase transition. The fact that in the Ising 1D model there is no fixed point at finite values is a consequence of the absence of phase transitions.

8. The key ideas of the RG approach are then: the integration over a subset of degrees of freedom, the rescaling Kadanoff procedure that generates a recursion relation, and the search for non trivial fixed points.

9. In 2D, things become quite more complicated; here we give only a schematic derivation (see Chandler, sec. 5.7 for details). The procedure is the same: integrating out some degrees of freedom in order to construct a recursive relation. Now there are 4 neighbors for every spin. Consider spins 5 and 6, whose nearest neighbors are spins 1,2,3,4,7,8 (see Fig. 5.5). Now integrating over 5 and 6 we obtain terms like

$$Q(K, N) = \sum_{\sigma_1, \sigma_2, \dots} [\exp K(\sigma_1 + \sigma_2 + \sigma_3 + \sigma_4) + \exp -K(\sigma_1 + \sigma_2 + \sigma_3 + \sigma_4)] \quad (5.93)$$

$$\times [\exp K(\sigma_2 + \sigma_3 + \sigma_7 + \sigma_8) + \exp -K(\sigma_2 + \sigma_3 + \sigma_7 + \sigma_8)] \dots \quad (5.94)$$

Again, such a form should be valid for all possible spin configurations. However now there will be several distinct choices of spin values, of which only four give distinct terms, for instance $\sigma_1 = \sigma_2 = \sigma_3 = \sigma_4 = \pm 1$, but we still have only two possible Kadanoff transformations (5.85) and a complete solution is not possible. We need more parameters, K_1, K_2, K_3 . So we can try to obtain 4 equations like the following one

$$e^{K(\sigma_1 + \sigma_2 + \sigma_3 + \sigma_4)} + e^{-K(\sigma_1 + \sigma_2 + \sigma_3 + \sigma_4)} = f(K) e^{\frac{1}{2}K_1(\sigma_1\sigma_2 + \sigma_2\sigma_3 + \sigma_3\sigma_4 + \sigma_4\sigma_1) + K_2(\sigma_1\sigma_3 + \sigma_2\sigma_4) + K_3(\sigma_1\sigma_2\sigma_3\sigma_4)} \quad (5.95)$$

which can now be solved for $f(K), K_1, K_2, K_3$ as a function of K , obtaining in particular

$$f(K) = 2 \cosh^{1/2}(2K) \cosh^{1/8}(4K) \quad (5.96)$$

$$K_1 = \frac{1}{4} \log \cosh(4K) \quad (5.97)$$

$$K_2 = \frac{1}{8} \log \cosh(4K) \quad (5.98)$$

$$K_3 = \frac{1}{8} \log \cosh(4K) - \frac{1}{2} \log \cosh(2K) \quad (5.99)$$

If we insert this back into $Q(K, N)$ we obtain a complicated expression in which there are sums over nearest neighbors, over next-to-nearest neighbors and over 4 spins at once.

10. This is still too complicated. We therefore now neglect K_3 and write

$$Q(K, N) = f(K)^{N/2} \sum_{spins} e^{\frac{1}{2}K_1(\sigma_1\sigma_2 + \sigma_2\sigma_3 + \sigma_3\sigma_4 + \sigma_4\sigma_1) + K_2(\sigma_1\sigma_3 + \sigma_2\sigma_4)} e^{\frac{1}{2}K_1(\sigma_2\sigma_3 + \sigma_3\sigma_8 + \sigma_7\sigma_8 + \sigma_7\sigma_2) + K_2(\sigma_2\sigma_8 + \sigma_7\sigma_3)}$$

Then we see that every nearest neighbor (nn) pair, e.g. $\sigma_2\sigma_3$, is repeated twice (and the same would be for $\sigma_1\sigma_4$ if we added more terms), while the next nearest-neighbors (nnn) like $\sigma_2\sigma_4$, only once. Then we can write

$$Q(K, N) = f(K)^{N/2} \sum_{N/2 spins} \exp[K_1 \sum_{nn} \sigma_i \sigma_j + K_2 \sum_{nnn} \sigma_i \sigma_j] \quad (5.100)$$

and we simplify again by assuming that the sum can be written as sums over only nn but with a different K' :

$$K_1 \sum_{nn} \sigma_i \sigma_j + K_2 \sum_{nnn} \sigma_i \sigma_j \approx K' \sum_{nn} \sigma_i \sigma_j \quad (5.101)$$

and therefore finally we obtain a recursion relation

$$Q(K, N) = f(K)^{N/2} \sum_{N/2 spins} \exp[K' \sum_{nn} \sigma_i \sigma_j] = f(K)^{N/2} Q(K', N/2) \quad (5.102)$$

11. It turns out that neglecting the four spin combination entirely, an approximated recursive relation can be found

$$g(K) = \frac{1}{2} [\log f(K) + g(K')] \quad (5.103)$$

$$g(K') = 2g(K) - \log[2 \cosh^{1/2}(2K) \cosh^{1/8}(4K)] \quad (5.104)$$

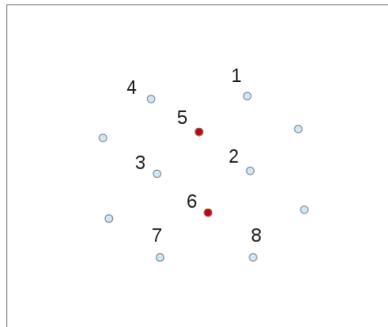


Figure 5.5: Summing over spins 5 and 6.

Approximating $K' = K_1 + K_2$ we obtain from Eq. (5.98), the recursive relation

$$K' = \frac{3}{8} \log \cosh(4K) \quad (5.105)$$

Now this equation has a fixed point: the value $K_c \approx 0.50698$ in fact solves the equation $K = \frac{3}{8} \log \cosh(4K)$, which means that once we reach that value we don't move away from it. However, it is an unstable fixed point, meaning that if we start from anywhere near K_c , the next iterations bring us away from it, i.e. $K' < K$ for $K < K_c$ and $K' > K$ for $K > K_c$. The fixed points at 0 and ∞ (the trivial fixed points) are instead stable. $K_c \approx 0.5$ means $T_c \approx 2J/k$, not far from the exact value $\approx 2.3J/k$ obtained in the previous section.

12. The heat capacity near T_c is

$$C \sim \frac{d^2 g(K)}{dK^2} \approx |T - T_c|^{-\alpha} \quad (5.106)$$

where we replaced the T -derivative with the K -derivative, since $T = J/kK$. The heat capacity diverges around the critical temperature, with $\alpha \approx 0.131$. (The exact result from Onsager's solution is however a logarithm divergence).

5.6 Mean field theory

1. The Ising model and many other systems out of equilibrium can also be approximated by the mean-field theory method. The idea here is to approximate the interactions on a given spin with a smooth average field rather than with the discrete lattice.
2. So instead of the exact form of the energy

$$E = -\mu H \sum_i \sigma_i - \sum_{i,j} J_{ij} \sigma_i \sigma_j \quad (5.107)$$

where $J_{ij} = J$ for nearest neighbors and 0 otherwise we seek an approximated form

$$E_{mf} = -\mu(H + \Delta H) \sum_i \sigma_i \quad (5.108)$$

which is then equivalent to a system with no interactions and a modified field $H + \Delta H$.

3. Let us replace $\sigma_i = m + \delta\sigma_i$, where $m = \langle \sigma_i \rangle$ is independent of i , and $\delta\sigma_i$ is the deviation from the mean value. Then we can write

$$E = -\mu H \sum_i \sigma_i - \sum_{i,j} J_{ij} (m^2 + 2m\delta\sigma_i + \delta\sigma_i\delta\sigma_j) \approx -\mu H \sum_i \sigma_i - \sum_{i,j} J_{ij} (m^2 + 2m\delta\sigma_i) \quad (5.109)$$

$$= -\mu H \sum_i \sigma_i - \sum_{i,j} J_{ij} (m^2 + 2m(\sigma_i - m)) = -\mu H \sum_i \sigma_i - \sum_{i,j} J_{ij} (-m^2 + 2m\sigma_i) \quad (5.110)$$

where we crucially neglect the quadratic fluctuations, assuming they are small. We now put $J_{ij} = J$ and assume there are z nearest neighbors, so that

$$\sum_{i,j} \dots = \sum_{nn} \dots = \frac{1}{2} \sum_i \sum_{j \in nn(i)} \dots \quad (5.111)$$

(the one-half factor is needed otherwise each nn bond is counted twice). Then we obtain

$$E \approx \frac{1}{2} J m^2 \sum_i \sum_{j \in nn(i)} 1 - \mu H \sum_i \sigma_i - J z m \sum_i \sigma_i = \left(\frac{1}{2} J N m^2 z \right) - \mu H_{mf} \sum_i \sigma_i \quad (5.112)$$

where

$$H_{mf} \equiv H + \frac{1}{\mu} J z m \quad (5.113)$$

where $z = 2D$ is the number of nearest neighbors in D dimensions. In this way we obtain a mean field acting on the spins instead of the true instantaneous value. In this sense, the spins act as if they were uncoupled but in a modified mean external field H_{mf} independent of i . The constant term in bracket can be neglected since it just gives an overall additive constant to the free energy $\sim \log Q$.

4. Such a non-interacting field has been already solved in Sect. 3.11. The partition function is

$$Q_{mf} = (2 \cosh[\beta\mu(H + \Delta H)])^N \quad (5.114)$$

where $\Delta H = \frac{1}{\mu} J z m$ and the average spin $m \equiv \langle \sigma \rangle$ is given by the equation (see Eq. 3.200)

$$m = \tanh[\beta\mu(H + \Delta H)] = \tanh[\beta\mu(H + \frac{1}{\mu} J z m)] \quad (5.115)$$

This is a transcendental equation for m which can easily be solved numerically. For $H = 0$, it gives non-zero solutions for m only if $\beta J z > 1$ i.e. for $T < T_C$ where (for $z = 2D$)

$$T_C = \frac{2DJ}{k} \quad (5.116)$$

This is therefore the critical temperature.

5. The mean field approximation is completely wrong in one dimension. In this case, in fact, the predicted $T_C = 2J/k$ is well different from the exact result, $T_C = 0$ (no phase transition). In 2 dimensions the mean field result $T_C = 4J/k$ is not so far from the exact result $T_C = 2.3J/k$ and in higher dimensions it gets closer to the numerical estimates. The predicted behavior near the critical point is however not really good. The critical exponent β is in fact predicted to be $1/2$ instead of the exact value 0.313 in 3D.

5.7 Correlation functions

This section follows closely the discussion in Pathria, *Statistical Mechanics*, Sect. 11.11, to which we refer for further details.

1. Phase transitions have another fundamental property: a strong correlation between distant points. In fact, this can be taken as the property that distinguishes phase transitions from ordinary fluctuations around an equilibrium state, in which the correlation among particles rapidly (typically, exponentially) decays away with distance. The residual magnetization of a ferromagnet can then be seen as the establishment of a long-range correlation among the spins, even if the microscopical interaction acts only among nearest neighbors.

2. We define the correlation function for pair of spins as

$$g(i, j) = \langle \sigma_i \sigma_j \rangle - \langle \sigma_i \rangle \langle \sigma_j \rangle \quad (5.117)$$

If the spins are uncorrelated obviously $g(i, j) = 0$. Notice that

$$\sum_{i,j} g(i, j) = \langle (\sum_i \sigma_i)^2 \rangle - \langle \sum_i \sigma_i \rangle^2 = \frac{1}{\mu^2} (\langle M^2 \rangle - \langle M \rangle^2) \quad (5.118)$$

where $M = \mu \sum_i \sigma_i$ is the total magnetization and $\langle M \rangle$ its average.

3. Let us consider again the Ising partition function

$$Q(\beta, H, J) = \sum_{\{\sigma_i\}} \exp(\beta\mu H \sum_i \sigma_i + \beta J \sum_i \sigma_i \sigma_{i+1}) \quad (5.119)$$

It follows that

$$\frac{\partial}{\partial H} \log Q = \frac{1}{Q} \sum_{\{\sigma_i\}} (\beta\mu \sum_i \sigma_i) \exp(\beta\mu H \sum_i \sigma_i + \beta J \sum_i \sigma_i \sigma_{i+1}) \quad (5.120)$$

$$= \beta\mu \langle \sum_i \sigma_i \rangle = \beta \langle M \rangle \quad (5.121)$$

and similarly

$$\frac{1}{Q} \frac{\partial^2}{\partial H^2} Q = \frac{1}{Q} \sum_{\{\sigma_i\}} (\beta\mu \sum_i \sigma_i)^2 \exp(\beta\mu H \sum_i \sigma_i + \beta J \sum_i \sigma_i \sigma_{i+1}) = \beta^2 \langle M^2 \rangle \quad (5.122)$$

Therefore, the magnetic susceptibility is

$$\chi = \frac{\partial \langle M \rangle}{\partial H} = \frac{1}{\beta} \frac{\partial^2}{\partial H^2} \log Q \quad (5.123)$$

4. Moreover we have

$$\frac{\partial^2}{\partial H^2} \log Q = \frac{1}{Q} \frac{\partial^2 Q}{\partial H^2} - \frac{1}{Q^2} \left(\frac{\partial Q}{\partial H} \right)^2 = \beta^2 (\langle M^2 \rangle - \langle M \rangle^2) \quad (5.124)$$

and therefore, because of Eq. (5.118), the magnetic susceptibility is

$$\chi = \frac{\partial \langle M \rangle}{\partial H} = \beta (\langle M^2 \rangle - \langle M \rangle^2) \quad (5.125)$$

$$= \beta \mu^2 \sum_{i,j} g(i, j) \quad (5.126)$$

This is called a *fluctuation-dissipation* relation. We have seen another example for ideal gases in Eq. (3.99). This relation shows that a large (or infinite) susceptibility, typical of a phase transition, implies also a strong correlation.

5. If the system is statistically homogeneous, its properties do not depend on the particular location. In this case we expect χ to depend only on the inter-particle separation \vec{r} and not on the individual position i, j . We may then convert the sum into an integral as $\sum_{ij}[\dots] \rightarrow \sum_i \sum_j[\dots] \rightarrow \sum_i a^{-d} \int[\dots] d\vec{r} = N a^{-d} \int[\dots] d\vec{r}$ and therefore

$$\chi = N\beta\mu^2 \sum_{\vec{r}} g(\vec{r}) \approx \frac{N\beta\mu^2}{a^d} \int g(\vec{r}) d\vec{r} \quad (5.127)$$

where a^d is the volume in d dimensions of an elementary lattice cell, such that $Na^d = V$.

6. There are no exact calculations of $g(\vec{r})$. An approximate form can be obtained in mean field theory. In this case we have seen in fact that

$$\langle\sigma_i\rangle = \tanh(\beta\mu H_{mf}) \quad (5.128)$$

where

$$H_{mf} = H_i + \frac{1}{\mu} J \sum_{nn} \langle\sigma_i\rangle \quad (5.129)$$

Contrary to what we have done so far, we consider now a non-homogeneous external field H_i . We then think of $\langle\sigma_i\rangle$ as a field over space and write

$$\sum_{nn} \langle\sigma_i\rangle \approx z\langle\sigma_i\rangle + \sum_j \nabla_j \langle\sigma_i\rangle dx_j + \frac{1}{2} \sum_{\ell m} \nabla_\ell \nabla_m \langle\sigma_i\rangle dx_\ell dx_m + \dots \quad (5.130)$$

where $\nabla_i = \partial/\partial x_i$. The term in $\nabla\langle\sigma_i\rangle$ cancels out because of the symmetric configuration of nearest neighbor spins. Expanding up to second order, and keeping only the isotropic term $\nabla^2 = \sum_i \nabla_i \nabla_i$, we obtain then

$$H_{mf} \approx H_i + \frac{z}{\mu} J \langle\sigma_i\rangle + \frac{cJa^2}{\mu} \nabla^2 \langle\sigma_i\rangle \quad (5.131)$$

Here, z is the number of nearest neighbors, c is a order-of-unity factor that depends on the particular lattice. Finally, a is an effective lattice length constant, equal to the typical interparticle distance.

7. Approximating for small x

$$\langle\sigma_i\rangle = \tanh(\beta\mu H_{mf}) \approx \beta\mu H_i + z\beta J \langle\sigma_i\rangle + c\beta J a^2 \nabla^2 \langle\sigma_i\rangle \quad (5.132)$$

we obtain near the critical point $H_i = 0$ the approximate expression

$$\beta\mu H_i \approx (1 - z\beta J) \langle\sigma_i\rangle - c\beta J a^2 \nabla^2 \langle\sigma_i\rangle \quad (5.133)$$

Near the critical point, $zJ = kT_c$, as we have seen in the mean field approximation. For $T \approx T_c$ we have then

$$(t - c'a^2 \nabla^2) \langle\sigma_i\rangle = h_i \quad (5.134)$$

where $c' = c/z$ and where we introduced the ordering field and the order parameter, respectively:

$$h_i = \beta\mu H_i \quad t = \frac{T - T_c}{T_c} \approx \frac{\beta_c - \beta}{\beta_c} \quad (5.135)$$

Differentiating, we obtain

$$(t - c'a^2 \nabla^2) \frac{\partial \langle\sigma_i\rangle}{\partial h_j} = \delta_{ij} \quad (5.136)$$

where we considered every local value h_i as independent of the field h_j at different locations. Now, one can show that

$$\frac{\partial \langle\sigma_i\rangle}{\partial h_j} = g(i, j) \quad (5.137)$$

In fact, since $\langle M \rangle = \mu \sum_i \langle\sigma_i\rangle$, we can expand and write

$$\delta \langle M \rangle = \mu \sum_i \sum_j \left\langle \frac{\partial \sigma_i}{\partial H_j} \delta H_j \right\rangle \quad (5.138)$$

Putting for simplicity $\delta H_j = \delta H$, we see that

$$\frac{\delta \langle M \rangle}{\delta H} = \beta \mu^2 \sum_i \sum_j \langle \frac{\partial \sigma_i}{\partial h_j} \rangle \quad (5.139)$$

and comparing with (5.126) we recover (5.137). Then we have an equation for the correlation function. In the continuous limit, $g(i, j) \rightarrow g(r)$ and $\delta_{ij} \rightarrow \delta(r)$, so

$$(t - c'a^2 \nabla^2)g(r) \approx \delta(r) \quad (5.140)$$

As is well known, the Laplacian ∇^2 in spherical coordinates becomes $r^{-2}(d/dr)r^2(d/dr)$ and one can show by direct substitution that $g(r) = (a/r)e^{-r/\xi}$ is a solution. Alternatively, one can go to Fourier space, in which $g(r) \rightarrow \hat{g}(k)e^{i\mathbf{k}\cdot\mathbf{r}}$; this means that

$$\hat{g}(k) = \frac{1}{t + c'a^2 k^2} \quad (5.141)$$

In real space, the Fourier anti-transform gives a simple expression in three dimensions:

$$g(\vec{r}) = \int \frac{e^{-i\mathbf{k}\cdot\mathbf{r}} d^3k}{t + c'a^2 k^2} = 2\pi \int \frac{e^{-ikr \cos \theta} k^2 dk d(\cos \theta)}{t + c'a^2 k^2} \sim \frac{a}{r} e^{-r/\xi} \quad (5.142)$$

where $\xi = a(c'/t)^{1/2}$ is the *correlation length*, which diverges as expected for $T \rightarrow T_c$. This law, first derived by Ornstein and Zernike in 1914-16, well exemplifies the transition from exponentially decaying correlations for $r \gg \xi$ (and therefore for $T \gg T_c$) to a power law for $r \ll \xi$.

8. For any dimension d and at small scales $r \ll \xi$ we also obtain a power law behavior

$$g \sim \frac{a^{d-2}}{r^{d-2}} \quad (5.143)$$

9. The scale ξ measures the scale below which the system is correlated. As the critical temperature is approached, the correlation length diverges to infinity. This is exactly the sort of long-range order that is created during a phase transition. It is this macroscopic behavior that allows to describe phase transitions in a way that does not depend on the microscopic features and leads to some kind of universal laws.

Chapter 6

Mathematical appendix

6.1 Partial derivative identities

Here we collect some useful identities concerning partial derivatives.

1. Chain rule. Consider a function $f(x, y)$ and suppose x, y depend on another variable s , i.e. $x(s), y(s)$. Then

$$\frac{df}{ds} = \frac{\partial f}{\partial x} \frac{dx}{ds} + \frac{\partial f}{\partial y} \frac{dy}{ds} \quad (6.1)$$

2. Chain rule. Consider a function $f(x, y)$ and suppose x, y depend on two other variables, i.e. $x(s, r), y(s, r)$. Then

$$\frac{\partial f}{\partial s} = \frac{\partial f}{\partial x} \frac{\partial x}{\partial s} + \frac{\partial f}{\partial y} \frac{\partial y}{\partial s} \quad (6.2)$$

$$\frac{\partial f}{\partial r} = \frac{\partial f}{\partial x} \frac{\partial x}{\partial r} + \frac{\partial f}{\partial y} \frac{\partial y}{\partial r} \quad (6.3)$$

(where the variable kept fixed is always the other one belonging to the same pair x, y or s, r). As a particular case, if we have $f[x(s, y), y]$, i.e. if the transformation is from x, y to s, y (in other words, we transform only x) then

$$\frac{\partial f}{\partial s} \Big|_y = \frac{\partial f}{\partial x} \Big|_y \frac{\partial x}{\partial s} \Big|_y + \frac{\partial f}{\partial y} \Big|_x \frac{\partial y}{\partial s} \Big|_y = \frac{\partial f}{\partial x} \Big|_y \frac{\partial x}{\partial s} \Big|_y \quad (6.4)$$

$$\frac{\partial f}{\partial y} \Big|_s = \frac{\partial f}{\partial x} \Big|_y \frac{\partial x}{\partial y} \Big|_s + \frac{\partial f}{\partial y} \Big|_x \frac{\partial y}{\partial y} \Big|_s = \frac{\partial f}{\partial x} \Big|_y \frac{\partial x}{\partial y} \Big|_s + \frac{\partial f}{\partial y} \Big|_x \quad (6.5)$$

3. Product rule

$$\frac{\partial}{\partial x}(uv) = u \frac{\partial v}{\partial x} + v \frac{\partial u}{\partial x} \quad (6.6)$$

4. Commutation rule

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x} \quad (6.7)$$

5. Reciprocals

$$\left(\frac{\partial f}{\partial x} \right)_y = \left(\frac{\partial x}{\partial f} \right)_y^{-1} \quad (6.8)$$

but note the variable held fixed has to be the same, so if, e.g., $f(x, y, z)$

$$\left(\frac{\partial f}{\partial x} \right)_y \neq \left(\frac{\partial x}{\partial f} \right)_z^{-1} \quad (6.9)$$

6. Triple product rule. If $f(x, y) = \text{const}$, then

$$\left(\frac{\partial f}{\partial x}\right)_y \left(\frac{\partial y}{\partial f}\right)_x \left(\frac{\partial x}{\partial y}\right)_f = -1 \quad (6.10)$$

Analogously, if x, y, z are not independent variables, i.e. if a relation $f(x, y, z) = 0$ exists (which is identical to say that $z(x, y) = \text{const}$ as in the previous case) then

$$\left(\frac{\partial x}{\partial y}\right)_z \left(\frac{\partial y}{\partial z}\right)_x \left(\frac{\partial z}{\partial x}\right)_y = -1 \quad (6.11)$$

7. Transformation of variables. If we have n functions y_i of n variables x_j , one can build the Jacobi matrix of the transformation from \mathbf{y} to \mathbf{x}

$$J_{ij} \equiv \frac{\partial y_i}{\partial x_j} \quad (6.12)$$

(this means implicitly that all the other $x_{m \neq j}$ are kept fixed). Its inverse, obtained with the rules for the inversion of a matrix, is

$$(J^{-1})_{ij} \equiv \frac{\partial x_i}{\partial y_j} \quad (6.13)$$

(this means first take the inverse, then take the i, j entry) where again all the other $y_{m \neq j}$ are kept fixed. It follows

$$\sum_m (J^{-1})_{im} J_{mj} = \sum_m \frac{\partial x_i}{\partial y_m} \frac{\partial y_m}{\partial x_j} = \delta_{ij} \quad (6.14)$$

Notice that

$$\frac{\partial x_i}{\partial y_j} \neq \left(\frac{\partial y_i}{\partial x_j}\right)^{-1} \quad (6.15)$$

since the variables kept fixed are different on the two sides.

6.2 Liouville's theorem

A flow of particles that are neither destroyed nor created (“conserved”) and do not collide against each other is governed by the collisionless Boltzmann equation, also called Liouville's equation when generalized to the thermodynamical phase-space. Let the number dN of particles in a space volume dV and with moments within volume $dp_x dp_y dp_z$ be

$$dN = f(t, \mathbf{x}, \mathbf{p}) dx dy dz dp_x dp_y dp_z \quad (6.16)$$

where f is the *distribution function* and the space-momentum volume element is called the phase-space volume. Points in the full thermodynamic $6N$ -phase-space are also conserved and do not “collide” (two different microstates will always remain different, i.e. they will never occupy the same $6N$ -point, at least classically, because otherwise the same initial conditions in terms of positions and momenta would give rise to two different trajectories, breaking classical determinism). So although to fix the ideas we begin with f as a particle number density, we will at the end generalize the result to the number of microstates in a $6N$ -dimensional cell $d^{3N} q d^{3N} p$,

$$dN_M = f(t, \mathbf{q}, \mathbf{p}) d^{3N} q d^{3N} p \quad (6.17)$$

Let's first consider a 1-dimensional example. Since the particle number is conserved, the number of particles with momentum p_x entering in a space volume $dx dy dz$ in direction x in the time interval dt is $\rho dx dy dz = \rho v_x dt dy dz$ (v_x is the x -component of the velocity and ρ is the number density of particles), minus the number exiting from the other side; if the particles move only along x , this has to be equal to the change in the number of particles inside the volume itself, $d\rho dV$. Then we have the continuity equation (the incoming velocity is taken negative)

$$d\rho dx dy dz = -d(\rho v_x) dt dy dz \quad (6.18)$$

or

$$\frac{\partial \rho}{\partial t} = -\frac{\partial(\rho v_x)}{\partial x} \quad (6.19)$$

This simple derivation can be extended to all six phase-space coordinates. The demonstration makes use of the following relations, valid in any number of dimensions and any choice of coordinates \mathbf{w} :

$$\frac{\partial}{\partial t} \int f dV = - \int_{\partial V} (f \mathbf{v}) \cdot \mathbf{n} dA = - \int_V \nabla \cdot (f \mathbf{v}) dV \quad (6.20)$$

where $\mathbf{v} = \dot{\mathbf{w}}$ and \mathbf{n} is the unit normal to the surface ∂A enclosing V . The first equality applies when f is a conserved quantity (mass, or number of particles, or number of microstates), the second is the divergence theorem and is always valid. Since the volume V is arbitrary, the integrands of the first and third member must be equal. For the phase-space coordinates this gives the generalized version of Eq. (6.19), namely

$$\frac{\partial f}{\partial t} + \sum_{i=1}^6 \frac{\partial (f \dot{w}_i)}{\partial w_i} = 0 \quad (6.21)$$

where $w = \{x, y, z, p_x, p_y, p_z\} = \{\mathbf{q}, \mathbf{p}\}$ is the phase-space vector of coordinates. Notice that if we allow for collisions, i.e. instantaneous interactions, f would not be differentiable, and an extra term taking into account this should be included. Now since $\dot{\mathbf{q}} = \partial H / \partial \mathbf{p}$ and $\dot{\mathbf{p}} = -\partial H / \partial \mathbf{q}$, where H is the Hamiltonian, we have

$$\sum_{i=1}^6 \frac{\partial (f \dot{w}_i)}{\partial w_i} = \sum_{i=1}^6 \left(f \frac{\partial \dot{w}_i}{\partial w_i} + \dot{w}_i \frac{\partial f}{\partial w_i} \right) \quad (6.22)$$

$$= f \left(\frac{\partial^2 H}{\partial \mathbf{q} \partial \mathbf{p}} - \frac{\partial^2 H}{\partial \mathbf{p} \partial \mathbf{q}} \right) + \sum_{i=1}^6 \dot{w}_i \frac{\partial f}{\partial w_i} \quad (6.23)$$

$$= \sum_{i=1}^6 \dot{w}_i \frac{\partial f}{\partial w_i} \quad (6.24)$$

Then we are left finally with Liouville's theorem, i.e. that the total differential of f vanishes:

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \sum_{i=1}^6 \dot{w}_i \frac{\partial f}{\partial w_i} = 0 \quad (6.25)$$

Intuitively, this means that f is constant along the flow of the particles (although does not have to be constant for a fixed position, i.e. $\partial f / \partial t$ is not necessarily zero).

Finally, since $\dot{w}_i = \{\dot{q}_i, \dot{p}_i\}$, this can be written also as

$$\frac{df}{dt} = \frac{\partial f}{\partial t} + \sum_i \left(\frac{\partial f}{\partial q_i} \frac{\partial H}{\partial p_i} - \frac{\partial f}{\partial p_i} \frac{\partial H}{\partial q_i} \right) \quad (6.26)$$

$$= \frac{\partial f}{\partial t} + \sum_i \{f, H\} = 0 \quad (6.27)$$

where the Poisson brackets are defined as $\{A, B\} \equiv \frac{\partial A}{\partial \mathbf{q}} \frac{\partial B}{\partial \mathbf{p}} - \frac{\partial B}{\partial \mathbf{q}} \frac{\partial A}{\partial \mathbf{p}}$. Eq. (6.27) is completely general, as long as f is a conserved quantity, and can be applied also to non-equilibrium states.

Since Eq. (6.20) is valid in any number of dimensions, this result extends directly to the full $6N$ -dimensional phase-space of thermodynamic systems, where f represents the density of microstates in the $6N$ volume element $d^{3N} q d^{3N} p$. Clearly, if f is independent of time, i.e. it represents an equilibrium distribution, $\partial f / \partial t = 0$ and the Poisson bracket vanishes. This means that f must depend on \mathbf{p}, \mathbf{q} only through constants of motion of the system. Typically, for a thermodynamic system, for which we are not interested in translations or rotations, the energy is the only constant of motion, and therefore we might quite generally assume that $f = f(H)$, as we do when deriving the canonical ensemble. In the grand-canonical ensemble, also N is a relevant constant of motion for every microstate, so $f = f(H, N)$ (obviously, in general, both H and N can vary from microstate to microstate, unless they are constrained).

6.3 Moment generating function

The moments are the most important descriptors of a probability distribution function (PDF). It is therefore useful to be able to calculate them easily. To this scope, one introduces the moment generating function (MGF), defined for a single variable as

$$m_x(t) \equiv \langle e^{tx} \rangle = \int e^{tx} f(x) dx \quad (6.28)$$

It is easy to show that

$$\left. \frac{d^r m_x(t)}{dt^r} \right|_{t=0} = \langle x^r \rangle \quad (6.29)$$

One has in fact

$$\left. \frac{d^r m_x(t)}{dt^r} \right|_{t=0} = \left. \frac{d^r \langle e^{tx} \rangle}{dt^r} \right|_{t=0} = \left\langle \left. \frac{d^r e^{tx}}{dt^r} \right|_{t=0} \right\rangle \quad (6.30)$$

$$= \langle x^r e^{tx} \rangle_{t=0} = \langle x^r \rangle \quad (6.31)$$

Analogously, one can define the MGF for central moments:

$$m_{x-\mu}(t) \equiv \langle e^{t(x-\mu)} \rangle = \int e^{t(x-\mu)} f(x) dx \quad (6.32)$$

Suppose now we have two independent variables x, y distributed as $f(x), g(y)$. Let us find the MGF of the sum $s = x + y$. We have $p(x, s) ds dx = f(x)g(y) dx dy$ and therefore

$$p(s, x) = f(x)g(y)|J| \quad (6.33)$$

where the Jacobian of the transformation from (x, y) to (s, x) is

$$J = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \quad (6.34)$$

and therefore $|J| = 1$. Then the distribution of s is given by marginalizing p over x ,

$$q(s) = \int p(s, x) dx = \int f(x)g(y) dx \quad (6.35)$$

It follows

$$m_s(t) = \langle e^{ts} \rangle = \int e^{ts} q(s) ds = \int e^{ts} p(s, x) dx ds \quad (6.36)$$

$$= \int e^{t(x+y)} f(x)g(y) dx dy = \langle e^{tx} \rangle \langle e^{ty} \rangle = m_x(t) m_y(t) \quad (6.37)$$

i.e. the MGF of the sum of two independent variables is the product of the two MGFs. This extends obviously to the sum of any number of independent variables.

If the MGF exists, then two PDFs with the same MGF are identical; in other words, the MGF characterizes completely the PDF (under some general smoothness conditions).

6.4 Central limit theorem

The MGF helps us to demonstrate the Central Limit Theorem, according to which the Gaussian is the asymptotic distribution of the sum of n independent identically distributed (IID) random variables in the limit of $n \rightarrow \infty$.

Let x_i with $i = 1, \dots, n$ be n IID random variables with mean μ and variance σ^2 , with an unknown PDF. The CLT states that the variable

$$Y \equiv \frac{\hat{x} - \mu}{\sigma/\sqrt{n}} \quad (6.38)$$

where $\hat{x} = \sum_i x_i/n$ tends to a Gaussian variable for $n \rightarrow \infty$. Let us define the normal variables

$$z_i = \frac{x_i - \mu}{\sigma} \quad (6.39)$$

with $\langle z_i \rangle = 0$ and $\langle z_i^2 \rangle = 1$. Clearly

$$Y = \frac{1}{\sqrt{n}} \sum_i z_i \quad (6.40)$$

Let us find the MGF of Y . By the property of additive variables we have

$$m_Y(t) = \langle e^{Yt} \rangle = \langle e^{z_i t / \sqrt{n}} \rangle^n \quad (6.41)$$

Now

$$\langle e^{z_i t / \sqrt{n}} \rangle^n = \left\langle 1 + \frac{z_i t}{\sqrt{n}} + \frac{z_i^2 t^2}{2!n} + \frac{z_i^3 t^3}{3!n^{3/2}} + \dots \right\rangle^n \quad (6.42)$$

$$= \left(1 + \frac{\langle z_i \rangle t}{\sqrt{n}} + \frac{\langle z_i^2 \rangle t^2}{2!n} + \frac{\langle z_i^3 \rangle t^3}{3!n^{3/2}} + \dots \right)^n \quad (6.43)$$

Since $\langle z_i \rangle = 0$ and $\langle z_i^2 \rangle = 1$ we obtain for $n \gg 1$

$$\langle e^{z_i t / \sqrt{n}} \rangle^n = \left(1 + \frac{t^2}{2!n} + \frac{\langle z_i^3 \rangle t^3}{3!n^{3/2}} + \dots \right)^n \quad (6.44)$$

$$\approx \left(1 + \frac{t^2}{2n} \right)^n \approx 1 + \frac{t^2}{2} \approx e^{\frac{1}{2}t^2} \quad (6.45)$$

which is the MGF of a Normal Gaussian variable, QED.

The importance of this theorem is that it guarantees that if the errors in a measure are the results of many independent errors due to various parts of the experiment, then they are expected to be distributed in a Gaussian way. It can be extended to the case of independent variables with different mean and variances but in this case the condition $n \gg 1$ is not a sufficient condition for normality.

6.5 Estimators of mean and variance.

In this section we assume always independent and identically distributed variables. The sample mean

$$\hat{x} \equiv \frac{\sum_i x_i}{n} \quad (6.46)$$

is an unbiased estimator of $\mu = E[x]$. In fact

$$E[\hat{x}] = \frac{1}{n} \sum_i \int x_i f(x) dx = \frac{\sum_i \mu}{n} = \mu \quad (6.47)$$

Notice also that even a single measure, eg x_1 , is an unbiased estimator of the mean: $E[x_1] = \mu$. The choice should then be the estimator of minimal variance. As we see below, the variance of the mean goes like $1/n$; the mean is therefore a better estimator than a single measure, or any mean of a subset of measures.

The sample variance

$$s^2 = \frac{1}{n-1} \sum_i (x_i - \hat{x})^2 \quad (6.48)$$

is an unbiased estimator of $\sigma^2 = E[(x - \mu)^2]$. Notice that both x_i and \hat{x} are random variables; if the mean is known in advance, the denominator of the unbiased estimator would be n instead of $n-1$.

Analogously, the unbiased estimator of the covariance is

$$\hat{V}_{ij} = \frac{1}{n-1} \sum_k (x_k - \hat{x}_i)(x_k - \hat{x}_j) \quad (6.49)$$

For two variables this is

$$\hat{V}_{xy} = \frac{n}{n-1}(\hat{xy} - \hat{x}\hat{y}) \quad (6.50)$$

where in this specific instance we use the $\hat{}$ to denote the sample average, for instance

$$\hat{xy} = \frac{1}{n} \sum_i x_i y_i \quad (6.51)$$

Finally, an estimator of the correlation coefficient is

$$r = \frac{\hat{V}_{xy}}{s_x s_y} = \frac{\hat{xy} - \hat{x}\hat{y}}{\sqrt{(\hat{x}^2 - \hat{x}^2)(\hat{y}^2 - \hat{y}^2)}} \quad (6.52)$$

(note $\hat{x}^2 \neq \hat{x}^2$). This is only asymptotically unbiased, even if \hat{V}_{xy}, s_x, s_y are unbiased quantities; however is often used because of its simplicity.

We can now estimate the variance of the mean:

$$\text{Var}[\hat{x}] = E[\hat{x}^2] - (E[\hat{x}])^2 = E\left[\left(\frac{1}{n} \sum_i x_i\right)\left(\frac{1}{n} \sum_j x_j\right)\right] - \mu^2 \quad (6.53)$$

$$= \frac{1}{n^2} \sum_{i,j} E[x_i x_j] - \mu^2 \quad (6.54)$$

$$= \frac{1}{n^2} [(n^2 - n)\mu^2 + n(\mu^2 + \sigma^2)] - \mu^2 = \frac{\sigma^2}{n} \quad (6.55)$$

where in the last step we have employed $E[x_i x_j] = \mu^2$ for $i \neq j$ and $E[x_i^2] = \mu^2 + \sigma^2$ (and that there are $n^2 - n$ combinations $i \neq j$ and n of $i = j$). The same result can be readily obtained by the law of error propagation, which in this case is exact since the mean is a linear combination of random variables. This is a very important result: the standard deviation of the mean is a factor $1/\sqrt{n}$ smaller wrt the standard deviation of a single measure. So if we perform another measurement, we expect it to deviate by $\sim \sigma$ wrt the mean; but if we take another mean of a set of n similar measurements, then we expect the new mean to deviate from the old one only by $\sim \sigma/\sqrt{n}$.

6.6 Parameter estimation: Statistics, sample, bias

So far we have analysed the theoretical properties of the distributions. However, what we really normally have is a number of measurements $x_1, x_2, x_3 \dots x_n$. If the measures are independent, we can assume that the joint PDF of the full set x_i is

$$f_{\text{sample}}(x_i) = f(x_1)f(x_2)f(x_3)\dots f(x_n) \quad (6.56)$$

Our problem now is to derive from the n measures the estimates of the population parameters, that is, the parameters that characterize $f(x)$, for instance the mean μ and the variance σ^2 . We need to find then functions of the data x_i , generally called *statistics*, such that they approximate the parameters of the $f(x)$. We have already seen an example of estimator: the mean

$$\hat{x} = \frac{\sum_i x_i}{n} \quad (6.57)$$

(now we use a hat to denote the estimator as a random variable, rather than any specific estimate) is in fact an estimator of $\mu = E[x]$. We can certainly have several estimators for any given parameter; here we see now which are the main properties that “good” estimator should possess.

Let θ be the parameter of $f(x)$ to be estimated and $\hat{\theta}$ the estimator, function of the n measures x_i . If $\hat{\theta}$ approximates θ in the limit of large n , the estimator is said to be *consistent* :

$$\lim_{n \rightarrow \infty} P(|\hat{\theta} - \theta| > \epsilon) = 0 \quad (6.58)$$

for every $\epsilon > 0$.

The expected value of $\hat{\theta} = \hat{\theta}(x_1, x_2, \dots)$ is by definition

$$E[\hat{\theta}] = \int \hat{\theta} f(x_1; \theta) f(x_2; \theta) \dots dx_1 \dots dx_n \quad (6.59)$$

If the *bias*

$$b = E[\hat{\theta}] - \theta \quad (6.60)$$

is zero for every n , the estimator $\hat{\theta}$ is *unbiased*. If $b \rightarrow 0$ only for large n , the estimator is said to be asymptotically *unbiased*. The bias is a systematic error because it does not depend on how good the measures are but on our choice of the estimator. At least in principle, one can always choose a better estimator or a unbiased one.

We define also consider the variance of the estimator:

$$V[\hat{\theta}] = E[(\hat{\theta} - E[\hat{\theta}])^2] = \int (\hat{\theta} - E[\hat{\theta}])^2 f(x_1; \theta) f(x_2; \theta) \dots dx_1 \dots dx_n \quad (6.61)$$

The variance of $\hat{\theta}$ is a statistical error because is unavoidable (although it can be minimized), since it depends ultimately on the fact that $\hat{\theta}$ is a random variable.

We define also the mean square error

$$MSE = E[(\hat{\theta} - \theta)^2] = E[(\hat{\theta} - E(\theta) + E(\theta) - \theta)^2] = V[\hat{\theta}] + b^2 \quad (6.62)$$

which can be indeed interpreted as the sum of the statistical and systematic errors.

6.7 Maximum likelihood estimators and Bayes' theorem

Let us suppose we know, or have good reasons to suspect, that a random variable x , e.g., the apparent magnitude of a supernova, has a probability distribution function (PDF) $f(x; \theta)$ that depends on an *unknown* parameter θ , e.g., the absolute magnitude. The “;” (previously we used the vertical bar |) is meant to distinguish the random variables x from the parameter θ . As we have already seen, such a probability is called a *conditional probability* of having the data x given the theoretical parameter θ . We may for instance suppose that the apparent magnitude m is distributed as a Gaussian variable with a given variance σ^2 (the observational error on m), but we do not know one of the cosmological parameters that enter the expected value $m_{\text{th}} = 5 \log_{10} d_L(z; \Omega_m^{(0)}, \Omega_\Lambda^{(0)}) + \text{constant}$, where d_L is the luminosity distance.

If we repeat the measure and we obtain x_1, x_2, x_3, \dots , then the law of joint probability tells us that the probability of obtaining x_1 in the interval dx_1 around x_1 , x_2 in the interval dx_2 around x_2 and so forth is

$$f(x_i; \theta) dx_i \equiv \prod_i f_i(x_i; \theta) dx_i = f_1(x_1; \theta) f_2(x_2; \theta) f_3(x_3; \theta) \dots dx_1 dx_2 dx_3 \dots, \quad (6.63)$$

if the measures are independent of each other. Clearly, for every θ this multivariate PDF will assume a different value. It is logical to *define* the best θ as the value for which $\prod_i f(x_i; \theta)$ is maximal. Indeed, if we generate random variables distributed as $f(x; \theta)$, the most likely outcome for x is that value maximizing $f(x; \theta)$. Conversely, if we have a particular outcome x , then our best bet is to assume that θ is such as to maximize the occurrence of that x . We used as an example independent data and a single parameter but this is by no means necessary. We define the best θ_α as those parameters that maximizes the joint function $f(x_1, x_2, \dots, x_n; \theta_1, \theta_2, \dots, \theta_m)$. Since in general we have many parameters to estimate, we write the function simply $f(x_i; \theta_\alpha)$, meaning all the x_i 's and all the θ_α 's.

The maximum likelihood method of parameter estimation consists therefore in finding the parameters that maximize the *likelihood function* $f(x_i; \theta_\alpha)$ by solving the system

$$\frac{\partial f(x_i; \theta_\alpha)}{\partial \theta_\alpha} = 0, \quad \alpha = 1, \dots, m. \quad (6.64)$$

Let us denote the solutions of these equations as $\hat{\theta}_\alpha$. They are functions of the data x_i and therefore are random variables, just as the data are. So the classical *frequentist* approach would try to determine the distribution of the $\hat{\theta}_\alpha$ s knowing the distribution of the x_i s; if this is possible, one can assign probabilities to $\hat{\theta}_\alpha$'s ranges, for

instance determine the interval of $\hat{\theta}_\alpha$ that contains 95% probability that a particular set of data has been drawn from the theoretical distribution (we will see this in later chapters). One problem with this approach is that it is often too difficult to derive the $\hat{\theta}_j$'s distribution analytically and very demanding to derive them numerically through simulated datasets. But the main problem is that this approach does not take into account what we already know concerning the theoretical parameters, for instance the result of previous experiments. To handle this information properly we need to switch to the *Bayesian* approach. Instead of looking for the probability $f(x_i; \theta_\alpha)$ of having the data given the model, we estimate the probability $L(\theta_\alpha; x_i)$ of having the model given the data.

This problem is solved by the fundamental theorem of conditional probabilities, the so-called Bayes' theorem¹:

$$P(T; D) = \frac{P(D; T)P(T)}{P(D)}, \quad (6.65)$$

where we denote the known data x_i with D and the unknown theory (that is, the theoretical parameters θ_α) with T . On the r.h.s., $P(D; T)$ is the conditional probability of having the data given the theory; $P(T)$ and $P(D)$ are the probability of having the theory and the data, respectively; finally, on the l.h.s., $P(T; D)$ is the conditional probability of having the theory given the data. Bayes' theorem is a consequence of the definition of conditional probability given in Eq. (2.19) $P(A; B) \equiv P(A, B)/P(B)$ and of the symmetry of the joint probability $P(A, B)$ (the probability of having both A and B) under the exchange of A, B .

It follows that

$$P(\theta_\alpha; x_i) = \frac{f(x_i; \theta_\alpha)p(\theta_\alpha)}{g(x_i)}, \quad (6.66)$$

where $p(\theta_\alpha)$ is called the *prior* probability for the parameters θ_α , while $g(x_i)$ is the PDF of the data x_i . The final function $P(\theta_\alpha; x_i)$ (or simply $P(\theta_\alpha)$ for shortness) is called *posterior*. The posterior contains the information we are looking for: the probability distribution of the parameters given that we observed the data x_i and that we have some prior knowledge about the parameters themselves. In fact the whole method in the Bayesian context should be called "the posterior method" rather than the "likelihood" method. Of course, the frequentist and the Bayesian estimator coincide if the prior is uniform; however, the PDF of the parameter would still be different.

Since $P(\theta_\alpha; x_i)$ is a probability distribution function for θ_α , it has to be normalized to unity:

$$\int P(\theta_\alpha; x_i) d^n \theta_\alpha = 1 = \frac{\int f(x_i; \theta_\alpha) p(\theta_\alpha) d^n \theta_\alpha}{g(x_i)}, \quad (6.67)$$

and therefore

$$\int f(x_i; \theta_\alpha) p(\theta_\alpha) d^n \theta_\alpha = g(x_i). \quad (6.68)$$

The integral on the l.h.s. is called *evidence* and the same name is sometimes given also to $g(x_i)$. The function $g(x_i)$ does not depend on the parameters θ_α and therefore it is of no help in estimating the parameters. From the point of view of $P(\theta_\alpha)$ it is just a normalization factor. The prior $p(\theta_\alpha)$ is also often unknown. Normally we do not know the probability distribution of theories, that is, whether for instance the Λ CDM model is more probable, from an absolute point of view, than a modified gravity model or whether the cosmological constant density fraction $\Omega_\Lambda^{(0)} = 0$ is more probable than $\Omega_\Lambda^{(0)} = 0.7$. However, we often *do know* something which, while not quite absolute in any sense, still represents some form of information independent of the data at hand. Namely, we know the results of previous experiments. If an experiment convincingly excluded, say, $\Omega_\Lambda^{(0)} < 0.5$, then we could use this information, putting $p(\Omega_\Lambda^{(0)} < 0.5) = 0$. If instead we believe that $\Omega_\Lambda^{(0)} = 0.7 \pm 0.1$, then we could use as $p(\Omega_\Lambda^{(0)})$ a Gaussian with mean 0.7 and standard deviation 0.1. These are typical *prior distributions*. An improper prior, i.e. one which is not normalized to unity, can also be employed. For instance, one can assume a uniform prior in the entire range from $-\infty$ to $+\infty$.

Once we have $P(\theta_\alpha)$ we need to search the maximum to obtain the maximum likelihood estimators $\hat{\theta}_i$. Because of the priors, this will differ in general from the maximum of $f(x_i; \theta_\alpha)$. Equation (6.64) is then replaced by

$$\frac{\partial P(\theta_\alpha)}{\partial \theta_\alpha} = 0, \quad \alpha = 1, \dots, n. \quad (6.69)$$

¹Reverend Thomas Bayes (1702–1761) studied what in modern terminology is the binomial distribution and introduced the concept of conditional probability. His work was published posthumously in 1763.

If, as usually the case, we do not know the denominator $g(x_i)$ in Eq. (6.66), the posterior P is not normalized and its normalization has to be recalculated. The overall normalization N is the integral over the parameter space:

$$N = \int P(\theta_\alpha) d^n \theta_\alpha, \quad (6.70)$$

where the integral extends to the whole parameter domain. From the normalized likelihood [i.e. $P(\theta_\alpha)/N$ which we keep calling $P(\theta_\alpha)$], we can derive the regions of confidence (or *belief*) for the parameters. These are defined as regions $R(\alpha)$ of constant $P(\theta_\alpha)$ for which

$$\int_{R(\alpha)} P(\theta_\alpha) d^n \theta = \alpha. \quad (6.71)$$

The region $R(\alpha)$ is the region for which the integral above evaluates to $0 < \alpha < 1$ (remember that now P is normalized and therefore its integral over the whole domain is 1). To find R one evaluates

$$\int \hat{P}(p_i) d^n \theta = \alpha_i, \quad (6.72)$$

where $\hat{P}(p_i) = P$ if $P > p_i$ and 0 elsewhere (i.e. the volume lying within the curve of “height” p_i , smaller than the peak of P). By trial and error (or by interpolating over a grid of p_i) one finds the preferred α_i . The typical choices are $\alpha = 0.683, 0.954, 0.997$ (also denoted as $1, 2, 3\sigma$, respectively), but sometimes other reference values are also employed. The value p_i that corresponds to α_i is the level at which we have to cut P to find the region $R(\alpha_i)$.

6.8 Information entropy

The concept of information entropy was introduced by Shannon in the 50s to estimate how much information is on average transmitted through cables, in order to optimize the efficiency of communications. In other words, it’s a measure of order: the more ordered a signal is (e.g., a text containing only “a”s), the less information it conveys, the smaller the entropy.

Imagine a reporter located in A needs to communicate every day to the journal in B some information, eg. the sport results. The transmitted text, encoded through a binary system of “0” and “1”, contains the sport, the score, the scorers, the weather condition, the audience, and other information. Let’s focus just on the sport discipline. Suppose in order to specify the kind of sport, the reporter just sends a code from a table ordered according to sport popularity that contains all disciplines, eg “1” is for football, “2” for tennis, etc. This can be encoded by approximately $b_i = \log_2 N_i/2 = \log N_i - 1$ bits, where N_i is the entry position in the table for sport i . E.g., if football is the first entry, it just needs a “0”; tennis will need “1” so one bit is sufficient; basket needs “10”, two bits, and so forth, so $2^{(N_i-1)}$ bits for position N_i or $N_i + 1$. Clearly, the less popular a sport is, the more bits it needs to be encoded, so N_i has to be an inverse function of popularity p_i , which is just the frequency with which that discipline is mentioned. The simplest choice is therefore $N_i = 1/p_i$, but any inverse power would not change the final result.

What is the average number of bits for the first field in the report? Sport i is mentioned with frequency p_i and every time requires $\log N_i$ bits (for large N_i we can neglect the -1). Then the average length is

$$S = \sum_i p_i \log_2 N_i = - \sum_i p_i \log p_i \quad (6.73)$$

We see that if $N_i = p_i^{-k}$, one would just rescale S ; the same if we use another log basis. In all generality, we can then say that

$$S = -k \sum_i p_i \log p_i = -k \langle \log p_i \rangle \quad (6.74)$$

is the average length of the report for the field “sport discipline”. Every other field will have the same form, so can be added to S . That is, one can send any message by compiling a table containing all possible “words” ranked in order of usage frequency, and then just send the table entry. So instead of saying “hello world” we will say, e.g., “word number 251 and word number 676” in the frequency table. S will measure the average

number of bits we need for such messages. A word that is always used ($p_i = 1$), or never used, and therefore never transmitted, ($p_i = 0$), has zero entropy, so it uses no “cable resources”. So here “amount of information” strictly means “communication of data weighted by their frequency”.