Chapter 2. Principles of Physical Statistics

This chapter is the key part of this course. It is started with a brief discussion of such basic notions of statistical physics as statistical ensembles, probability, and ergodicity. Then the so-called microcanonical distribution postulate is formulated in parallel with the statistical definition of entropy. The next step is the derivation of the Gibbs distribution, which is frequently considered the summit of the statistical physics, and one more, grand canonical distribution, which is more convenient for some tasks - in particular for the derivation of the Boltzmann, Fermi-Dirac, and Bose-Einstein statistics for systems of independent particles.

2.1. Statistical ensembles and probability

As has been already discussed in Sec. 1.1, statistical physics deals with systems in conditions when either the unknown initial conditions, or the system complexity, or the laws of motion (as in the case of quantum mechanics) do not allow a definite prediction of measurement results. The main formalism for the analysis of such systems is the probability theory, so let me start with a very brief review of its basic concepts using informal “physical” language - less rigorous but (hopefully) more transparent than a standard mathematical treatment.\footnote{For the reader interested in reviewing a more rigorous approach, I can recommend, for example, Chapter 18 of the handbook by G. Korn and T. Korn – see MA Sec. 16(ii).}

Consider \( N \gg 1 \) independent similar experiments carried out with \( \textit{apparently} \) similar systems (i.e. systems with identical macroscopic parameters such as volume, pressure, etc.), but still giving, by any of the reasons outlined above, different results of measurements. Such a collection of experiments, together with the fixed method of result processing, is a good example of a statistical ensemble. Let us start from the case when the experiments may have \( M \) different discrete outcomes, and the number of experiments giving the corresponding different results is \( N_1, N_2, \ldots, N_M \), so that

\[
\sum_{m=1}^{M} N_m = N. \tag{2.1}
\]

The probability of each outcome, for the given statistical ensemble, is then defined as

\[
W_m \equiv \lim_{N \to \infty} \frac{N_m}{N}. \tag{2.2}
\]

Though this definition is so close to our everyday experience that it is almost self-evident, a few remarks may still be relevant.

First, probabilities \( W_m \) depend on the exact statistical ensemble they are defined for, notably including the method of result processing. As an example, consider the standard coin tossing. For the ensemble of \( \textit{all} \) tossed coins, the probabilities of both the heads and tails outcomes equal \( \frac{1}{2} \). However, nothing prevents us from defining another statistical ensemble as a set of coin-tossing experiments with the heads-up outcome. Evidently, the probability of finding coins with tails up in this new ensemble is not \( \frac{1}{2} \) but 0. Still, this set of experiments is not only legitimate but also a rather meaningful statistical
ensemble; for example, the exact position and orientation of the tossed coins on the floor, within this restricted ensemble, may be rather random.

Second, a statistical ensemble does not necessarily require $N$ different physical systems, e.g., $N$ different coins. It is intuitively clear that tossing the same coin $N$ times constitutes an ensemble with similar statistical properties. More generally, a set of $N$ experiments with the same system provides a statistical ensemble equivalent to the set of experiments with $N$ different systems, provided that the experiments are kept independent, i.e. that outcomes of past experiments do not affect those of the experiments to follow. Moreover, for most physical systems of interest any special preparation is unnecessary, and $N$ different experiments, separated by sufficiently long time intervals, form a “good” statistical ensemble – the property called ergodicity.$^2$

Third, the reference to infinite $N$ in Eq. (2) does not strip the notion of probability from its practical relevance. Indeed, it is easy to prove (see Chapter 5) that, at very general conditions, at finite but sufficiently large $N$, numbers $N_m$ are approaching their average (or expectation) values$^3$

$$\langle N_m \rangle \equiv W_m N, \quad (2.3)$$

with the relative deviation scale decreasing as $1/\langle N_m \rangle^{1/2}$.

Now let me list those properties of probabilities that we will immediately need. First, dividing Eq. (1) by $N$ and following the limit $N \to \infty$, we get the well-known normalization condition

$$\sum_{m=1}^N W_m = 1; \quad (2.4)$$

just remember that it is true only if each experiment definitely yields one of outcomes $N_1, N_2, \ldots, N_M$. Next, if we have an additive function of results,

$$f = \frac{1}{N} \sum_{m=1}^N N_m f_m, \quad (2.5)$$

where $f_m$ are some definite (deterministic) coefficients, we may define the statistical average (also called the expectation value) of the function as

$$\langle f \rangle \equiv \lim_{N \to \infty} \frac{1}{N} \sum_{m=1}^N \langle N_m \rangle f_m, \quad (2.6)$$

$^2$ The most popular counter-example of a non-ergodic system is an energy-conserving system of particles placed in a potential which is a quadratic form of particle coordinates. Theory of oscillations tells us (see, e.g., CM Sec. 5.2) that this system is equivalent to a set of non-interacting harmonic oscillators. Each of these oscillators conserves its own initial energy $E_j$ forever, so that the statistics of $N$ measurements of one such system may differ from that of $N$ different systems with random distribution of $E_j$, even if the total energy of the system, $E = \sum_j E_j$, is the same. Such non-ergodicity, however, is a rather feeble phenomenon, and is readily destroyed by any of “mixing” mechanisms, such as weak interaction with environment (leading, in particular, to oscillation damping), nonlinear interaction of the components (see, e.g., CM Ch. 4), and chaos (CM Ch. 9), all of them strongly enhanced by increasing the number of particles in the system, i.e. the number of its degrees of freedom. This is why most real-life systems are ergodic; for those interested in non-ergodic exotics, I can recommend the monograph by V. Arnold and A. Avez, Ergodic Problems of Classical Mechanics, Addison-Wesley, 1989.

$^3$ Here, and everywhere in these notes, angle brackets $\langle \rangle$ mean averaging over a statistical ensemble, which is generally different from averaging over time – as it will be the case in quite a few examples considered below.
so that using Eq. (3) we get

\[ \langle f \rangle = \sum_{m=1}^{M} W_m f_m. \]  \hspace{1cm} (2.7)

Notice that Eq. (3) may be considered as the particular form of this general result, for all \( f_m = 1 \).

Next, the spectrum of possible experimental outcomes is frequently continuous. (Think, for example, about the positions of the marks left by bullets fired into a target from a far.) The above formulas may be readily generalized to this case; let us start from the simplest situation when all different outcomes may be described by one continuous variable \( q \), which replaces the discrete index \( m \) in Eqs. (1)-(7). The basic relation for this case is the self-evident fact that the probability \( dW \) of having an outcome within a very small interval \( dq \) near point \( q \) is proportional to the magnitude of that interval:

\[ dW = w(q) dq. \]  \hspace{1cm} (2.8)

Function \( w(q) \), which does not depend on \( dq \), is called the probability density. Now all the above formulas may be recast by replacing probabilities \( W_m \) by products (8), and the summation over \( m \), by integration over \( q \). In particular, instead of Eq. (4) the normalization condition now becomes

\[ \int w(q) dq = 1, \]  \hspace{1cm} (2.9)

where the integration should be extended over the whole range of possible values of \( q \). Similarly, instead by Eq. (5), it is natural to consider a function \( f(q) \). Then instead of Eq. (7), the expectation value of the function may be calculated as

\[ \langle f \rangle = \int w(q) f(q) dq. \]  \hspace{1cm} (2.10)

It is straightforward to generalize these formulas to the case of more variables. For example, results of measurements of a particle with 3 degrees of freedom may be described by the probability density \( w \) defined in the 6D space of its generalized radius-vector \( q \) and momentum \( p \). As a result, the expectation value of a function of these variables may be expressed as a 6D integral

\[ \langle f \rangle = \int w(q,p) f(q,p) d^3q d^3p. \]  \hspace{1cm} (2.11)

Some systems considered in this course consist of components whose quantum properties cannot be ignored, so let us discuss how \( \langle f \rangle \) should be calculated in this case. If by \( f_m \) we mean measurement results, Eq. (7) (and its generalizations) of course remains valid, but since these numbers themselves may be affected by the intrinsic quantum-mechanical uncertainty, it may make sense to have a bit deeper look into this situation. Quantum mechanics tells us\(^4\) that the most general expression for the expectation value of an observable \( f \) in a certain ensemble of macroscopically similar systems is

\[ \langle f \rangle = \sum_{m,m'} W_{mm'} f_m f_{m'} \equiv \operatorname{Tr}(W f). \]  \hspace{1cm} (2.12)

Here \( f_{mm'} \) are the matrix elements of the quantum-mechanical operator \( \hat{f} \) corresponding to the observable \( f \), in a full basis of orthonormal states \( m \),

\[ f_{mm'} = \langle m | \hat{f} | m' \rangle, \]  \hspace{1cm} (2.13)

\(^4\) See, e.g., QM Sec. 6.1.
while coefficients $W_{mm'}$ are elements of the so-called density matrix $W$, which represents, in the same basis, a density operator $\hat{W}$ describing properties of this ensemble. Equation (12) is evidently more general than Eq. (7), and is reduced to it only if the density matrix is diagonal:

$$W_{mm'} = W_m \delta_{mm'},$$  \hspace{1cm} (2.14)

(where $\delta_{mm'}$ is the Kronecker symbol), when the diagonal elements $W_m$ play the role of probabilities of the corresponding states.

Thus the largest difference between the quantum and classical description is the presence, in Eq. (12), of the off-diagonal elements of the density matrix. They have largest values in the pure (also called “coherent”) ensemble, in which the state of the system may be described with state vectors, e.g., the ket-vector

$$|\alpha\rangle = \sum_m \alpha_m |m\rangle,$$  \hspace{1cm} (2.15)

where $\alpha_m$ are some complex coefficients. In this simple case, the density matrix elements are merely

$$W_{mm'} = \alpha_m^* \alpha_{m'},$$  \hspace{1cm} (2.16)

so that the off-diagonal elements are of the same order as the diagonal elements. For example, in the very important particular case of a two-level system, the pure-state density matrix is

$$W = \begin{pmatrix} \alpha_1^* \alpha_1 & \alpha_1^* \alpha_2 \\ \alpha_2^* \alpha_1 & \alpha_2^* \alpha_2 \end{pmatrix},$$  \hspace{1cm} (2.17)

so that the product of its off-diagonal components is as large as that of the diagonal components. In the most important basis of stationary states, i.e. eigenstates of system’s time-independent Hamiltonian, coefficients $\alpha_m$ oscillate in time as

$$\alpha_m(t) = \alpha_m(0) \exp\left\{-i \frac{E_m}{\hbar} t\right\} = |\alpha_m| \exp\left\{-i \frac{E_m}{\hbar} t + i \varphi_m\right\},$$  \hspace{1cm} (2.18)

where $E_m$ are the corresponding eigenenergies, and $\varphi_m$ are constant phase shifts. This means that while the diagonal terms of the density matrix (16) remain constant, its off-diagonal components are oscillating functions of time:

$$W_{mm'} = \alpha_m^* \alpha_{m'} = |\alpha_m| \alpha_{m'} \exp\left\{i \frac{E_m - E_{m'}}{\hbar} t\right\} \exp\{i(\varphi_m - \varphi_{m'})\},$$  \hspace{1cm} (2.19)

Due to the extreme smallness of the Planck constant (on the human scale of things), a miniscule random perturbations of eigenenergies are equivalent to substantial random changes of the phase multiplier, so that the time average of any off-diagonal matrix element tends to zero. Moreover, even if our statistical ensemble consists of systems with exactly the same $E_m$, but different values $\varphi_m$ (which are typically hard to control at the initial preparation of the system), the average values of all $W_{mm'}$ (with $m \neq m'$) vanish again.

---

5 Here I use the Schrödinger picture of quantum mechanics in which the matrix elements $f_{nn'}$ do not evolve in time.
This is why, besides some very special cases, typical statistical ensembles of quantum particles are far from being pure, and in most cases (certainly including the thermodynamic equilibrium), a good approximation for their description is given by the opposite limit of the so-called classical mixture in which all off-diagonal matrix elements of the density matrix equal zero, and its diagonal elements $W_{mm}$ are merely the probabilities $W_m$ of the corresponding eigenstates. In this case, for observables compatible with energy, Eq. (12) is reduced to Eq. (7), with $f_m$ being the eigenvalues of variable $f$.

### 2.2. Microcanonical ensemble and distribution

Let us start with the discussion of physical statistics with the simplest, *microcanonical statistical ensemble* that is defined a set of macroscopically similar closed (isolated) systems with *virtually* the same total energy $E$. Since in quantum mechanics the energy of a closed system is quantized, it is convenient to include into the ensemble all systems with energies $E_m$ within a narrow interval $\Delta E << E$, that is nevertheless much larger than the average distance $\delta E$ between the energy levels, so that the number $M$ of different quantum states within interval $\Delta E$ is large, $M >> 1$. Such choice of $\Delta E$ is only possible if $\delta E << E$; however, the reader should not worry too much about this condition, because the most important applications of the microcanonical ensemble are for very large systems (or very high energies) when the energy spectrum is very dense.\(^7\)

This ensemble serves as the basis for the formulation of a postulate which is most frequently called the *microcanonical distribution* (or sometimes the “main statistical hypothesis”): *in the thermodynamic equilibrium, all possible states of the microcanonical ensemble have equal probability,*

$$W_m = \frac{1}{M} = \text{const.} \tag{2.20}$$

Though in some constructs of statistical mechanics this equality is derived from other axioms, which look more plausible to their authors, I believe that Eq. (20) may be taken as the starting point of the statistical physics, supported “just” by the compliance of all its corollaries with experimental observations.\(^8\)

Note that postulate (20) sheds a light on the nature of the macroscopic irreversibility of microscopically reversible (closed) systems: if such a system was initially in a certain state, its time

\(^6\) The terms “microcanonical”, as well as “canonical” (see Sec. 4 below) are apparently due to J. Gibbs, and I could not find out his motivation for these names. (“Canonical” in the sense of “standard” or “common” is quite appropriate, but why “micro”?)

\(^7\) Formally, the main result of this section, Eq. (20), is valid for any $M$ (including $M = 1$), it is just less informative for small $M$ - and trivial for $M = 1$.

\(^8\) Though I have to move on, let me note that the microcanonical distribution (20) is a very nontrivial postulate, and my advice to the reader to give some thought to this foundation of the whole building of statistical mechanics.
evolution with just miniscule interactions with environment (which is necessary for reaching the thermodynamic equilibrium) would eventually lead to the uniform distribution of its probability among all states with the essentially same energy. Each of these states is not “better” than the initial one; rather, in a macroscopic system, there are just so many of these states that the chance to find the system in the initial state is practically nil - again, think about the ink drop diffusion into a glass of water.

Now let us find a suitable definition of entropy $S$ of a microcanonical ensemble member - for now, in the thermodynamic equilibrium only. Since $S$ is a measure of disorder, it should be related to the amount of information lost when the system goes from the full order to the full disorder, i.e. into the microcanonical distribution (20), or, in other words, the amount of information necessary to find the exact state of your system in a microcanonical ensemble.

In the information theory, the amount of information necessary to make a definite choice between two options with equal probabilities (Fig. 2a) is defined as

$$I(2) = \log_2 2 = 1.$$  \hspace{1cm} (2.21)

This unit of information is called a bit. Now, if we need to make a choice between 4 equally probable opportunities, it can be made in two similar steps (Fig. 2b), each requiring one bit of information, so that the total amount of information necessary for the choice is

$$I(4) = 2I(2) = 2 = \log_2 4.$$ \hspace{1cm} (2.22)

An obvious extension of this process to the choice between $M = 2^m$ states gives

$$I(M) = mI(2) = m = \log_2 M.$$ \hspace{1cm} (2.23)

![Fig. 2.2. “Logarithmic trees” of binary decisions for making a choice between (a) 2 and (b) 4 opportunities with equal probabilities.](image)

This measure, if extended naturally to any integer $M$, is quite suitable for the definition of entropy at equilibrium, with the only difference that, following tradition, the binary logarithm is replaced with the natural one:10

---

9 I will rely on reader’s common sense and intuitive understanding what information is, because in the formal information theory this notion is also essentially postulated - – see, e.g., the wonderfully clear text by J. Pierce, *An Introduction to Information Theory*, Dover, 1980.

10 This is of course just the change of a constant factor: $S(M) = \ln M = \ln 2 \times \log_2 M = \ln 2 \times I(M) \approx 0.693 I(M)$. A review of Chapter 1 shows that nothing in thermodynamics prevents us from choosing such coefficient arbitrarily, with the corresponding change of the temperature scale – see Eq. (1.9). In particular, in the SI units, Eq. (24b) becomes $S = k_B \ln W_m$, so that one bit of information corresponds to the entropy change $\Delta S = k_B \ln 2 \approx 0.693 k_B \approx 0.965 \times 10^{-23}$ J/K. By the way, formula “$S = k \log W$” is engraved on the tombstone of L. Boltzmann (1844-1906) who was the first one to recognize this intimate connection between the entropy and probability.
Using Eq. (20), we may recast this definition in the most frequently used form

\[ S = \ln \frac{1}{W_m} = -\ln W_m. \]  

(2.24b)

(Again, please note that Eq. (24) is valid in the thermodynamic equilibrium only!)

Equation (24) satisfies the major condition for the entropy definition in thermodynamics, i.e. to be a unique characteristic of disorder. Indeed, according to Eq. (20), number \( M \) (and hence any function of \( M \)) are the only possible measures characterizing the microcanonical distribution. We also need this function of \( M \) to satisfy another requirement to the entropy, of being an extensive thermodynamic variable, and Eq. (24) does satisfy this requirement as well. Indeed, mathematics says that for two independent systems the joint probability is just a product of their partial probabilities, and hence, according to Eq. (24b), their entropies just add up.

Now let us see whether Eqs. (20) and (24) are compatible with the 2\textsuperscript{nd} law of thermodynamics. For that, we need to generalize Eq. (24) for \( S \) to an arbitrary state of the system (generally, out of thermodynamic equilibrium), with arbitrary state probabilities \( W_m \). For that, let us first recognize that \( M \) in Eq. (24) is just the number of possible ways to commit a particular system to a certain state \( n \) (\( n = 1, 2, ..., M \)), in a statistical ensemble where each state is equally probable. Now let us consider a more general ensemble, still consisting of a large number \( N >> 1 \) of similar systems, but with a certain number \( N_m = W_m N >> 1 \) of systems in each of \( M \) states, with \( W_m \) not necessarily equal. In this case the evident generalization of Eq. (24) is that the entropy \( S_N \) of the whole ensemble is

\[ S_N = \ln M(N_1, N_2, ...) \]

where \( M(N_1,N_2,...) \) is the number of ways to commit a particular system to a certain state \( n \), while keeping all numbers \( N_n \) fixed. Such number \( M(N_1,N_2,...) \) is clearly equal to the number of ways to distribute \( N \) distinct balls between \( M \) different boxes, with the fixed number \( N_m \) of balls in each box, but in no particular order within it. Comparing this description with the definition of the so-called multinomial coefficients,\(^{11}\) we get

\[ M(N_1,N_2,...) = \frac{N!}{N_1!N_2!...N_M!}, \quad \text{with} \quad N = \sum_{m=1}^{M} N_m. \]

In order to simplify the resulting expression for \( S_N \), we can use the famous Stirling formula in its crudest, de Moivre’s form\(^{12}\) whose accuracy is suitable for most purposes of statistical physics:

\[ \ln(N!)_{N \to \infty} \rightarrow N(\ln N - 1). \]

When applied to our current problem, this gives the following average entropy per system,\(^{13}\)

---

\(^{11}\) See, e.g., MA Eq. (2.3). Despite the intimidating name, Eq. (26) may be very simply derived. Indeed, \( N! \) is just the number of all possible permutations of \( N \) balls, i.e. the ways to place them in certain positions – say, inside \( M \) boxes. Now in order to take into account that the particular order of the balls in each box is not important, that number should be divided by all numbers \( N_n! \) of possible permutations of balls within each box – that’s it.

\(^{12}\) See, e.g., MA Eq. (2.10).
\[ S \equiv \frac{S_N}{N} = \frac{1}{N} \ln \left( \sum_{N_1, N_2, \ldots, N_M} C_{N_1, N_2, \ldots, N_M} \right) = \frac{1}{N} \left[ \ln(N!) - \sum_{m=1}^{M} \ln(N_m!) \right] \]

\[ \rightarrow \frac{1}{N} \left[ N \ln(N-1) - \sum_{m=1}^{M} N_m \ln(N_m - 1) \right] = \ln(N) - \sum_{m=1}^{M} \frac{N_m \ln N_m}{N}. \]

and since this result is only valid in the limit \( N_m \to \infty \) anyway, we may use Eq. (2) to present it as

\[ S = -\sum_{m=1}^{M} W_m \ln W_m = \sum_{m=1}^{M} W_m \ln \frac{1}{W_m}. \]  

(2.29)

This extremely important formula\(^{14}\) may be interpreted as the average of the entropy values given by Eq. (24), weighed with specific probabilities \( W_m \) in accordance with the general formula (7).\(^{15}\)

Now let us find what distribution of probabilities \( W_m \) provides the largest value of entropy (29). The answer is almost evident from a single glance at Eq. (29). For example, if coefficients \( W_m \) are constant (and hence equal to \( 1/M' \)) for a subgroup of \( M' \leq M \) states and equal zero for all others, all \( M' \) nonvanishing terms in the sum (29) are equal to each other, so that

\[ S = M' \frac{1}{M'} \ln M' = \ln M', \]  

(2.30)

so that the closer \( M' \) to its maximum number \( M \) the larger \( S \). Hence, the maximum of \( S \) is reached at the uniform distribution given by Eq. (24).

In order to prove this important fact more strictly, let us find the maximum of function given by Eq. (29). If its arguments \( W_1, W_2, \ldots, W_M \) were completely independent, this could be done by finding the point (in the \( M \)-dimensional space of coefficients \( W_m \)) where all partial derivatives \( \partial S/\partial W_m \) are equal to zero. However, since the probabilities are constrained by condition (4), the differentiation has to be carried out more carefully, taking into account this interdependence:

\[ \left[ \frac{\partial}{\partial W_m} S(W_1, W_2, \ldots) \right]_{\text{cond}} = \frac{\partial S}{\partial W_m} + \sum_{m' \neq m} \frac{\partial S}{\partial W_{m'}} \frac{\partial W_{m'}}{\partial W_m}. \]  

(2.31)

At the maximum of function \( S \), all such expressions should be equal to zero simultaneously. This condition may be presented as \( \partial S/\partial W_m = \lambda \), where the so-called Lagrange multiplier \( \lambda \) is independent of \( m \). Indeed, at such point Eq. (31) becomes

\[ \left[ \frac{\partial}{\partial W_m} S(W_1, W_2, \ldots) \right]_{\text{cond}} = \lambda + \sum_{m' \neq m} \lambda \frac{\partial W_{m'}}{\partial W_m} = \lambda \left( \frac{\partial W_m}{\partial W_m} + \sum_{m' \neq m} \frac{\partial W_{m'}}{\partial W_m} \right) = \lambda \frac{\partial}{\partial W_m} (1) = 0. \]  

(2.32)

\(^{13}\) Strictly speaking, I should use notation \( \langle S \rangle \) here. However, following the style accepted in thermodynamics, I will drop the averaging sign until we will really need them to avoid confusion. Again, this shorthand is not too bad because the relative fluctuations of entropy (as those of any macroscopic variable) are very small at \( N \gg 1 \).

\(^{14}\) With the replacement of \( \ln W_m \) for \( \log_2 W_m \) (i.e. division by \( \ln 2 \)), Eq. (29) is famous as the Shannon (or “Boltzmann-Shannon”) formula for average information \( I \) per symbol in a long communication string using \( M \) different symbols, with probability \( W_m \) each.

\(^{15}\) In some textbooks, this simple argument is even accepted as the derivation of Eq. (29); however, it is evidently less strict than the one outlined above.
For the particular expression (29), condition $\frac{\partial S}{\partial W_m} = \lambda$ yields

$$\frac{\partial S}{\partial W_m} \equiv \frac{d}{dW_m} \left[-W_m \ln W_m\right] = -\ln W_m - 1 = \lambda. \tag{2.33}$$

Equation (33) may hold for all $m$ (and hence the entropy reach its maximum value) only if $W_m$ is independent on $m$. Thus entropy (29) indeed reaches its maximum value (24) at equilibrium.

To summarize, we see that definition (24) of entropy in statistical physics does fit all the requirements imposed on this variable by thermodynamics. In particular, we have been able to prove the 2$^{nd}$ law of thermodynamics, starting from that definition and a more fundamental postulate (20). Now let me discuss one possible point of discomfort with that definition: it depends on the accepted energy interval of the microcanonical ensemble, for whose width $\Delta E$ no exact guidance is offered. However, if the interval $\Delta E$ contains many states, $M \gg 1$, then with a very small relative error (vanishing in the limit $M \to \infty$), $M$ may be presented as

$$M = g(E)\Delta E, \tag{2.34}$$

where $g(E)$ is the density of states of the system:

$$g(E) \equiv \frac{d\Sigma(E)}{dE}, \tag{2.35}$$

$\Sigma(E)$ being the total number of states with energies below $E$. (Note that the average interval $\delta E$ between energy levels, mentioned in the beginning of this section, is just $\delta E = \Delta E/M = 1/g$.) Plugging Eq. (34) into Eq. (24), we get

$$S = \ln M = \ln g(E) + \ln \Delta E, \tag{2.36}$$

so that the only effect of a particular choice of $\Delta E$ is an offset of entropy by a constant, and in Chapter 1 we have seen that such a shift does not affect any measurable quantity. Of course, Eq. (34), and hence Eq. (36) are only precise in the limit when density of states $g(E)$ is so large that the range available for the appropriate choice of $\Delta E$,

$$g^{-1}(E) \ll \Delta E \ll E, \tag{2.37}$$

is sufficiently broad: $M = g(E)E = E/\delta E \gg 1$.

In order to get some feeling of the functions $g(E)$ and $S(E)$ and the feasibility of condition (37), and also to see whether the microcanonical distribution may be directly used for calculations of thermodynamic variables in particular systems, let us apply it to a microcanonical ensemble of many sets of $N \gg 1$ independent, similar harmonic oscillators with eigenfrequency $\omega$. (Please note that the requirement of a virtually fixed energy is applied, in this case, the total energy $E_N$ of the set, rather to a single oscillator - whose energy $E$ may be virtually arbitrary, though certainly less than $E_N \sim NE$.) Basic quantum mechanics teaches us that the eigenenergies of such an oscillator form a discrete, equidistant spectrum:

---

16 This is not to say that these definitions are fully equivalent. Despite all the wealth of quantitative relations given by thermodynamics, it still leaves a substantial uncertainty in the definition of entropy (and hence temperature), while Eq. (24) narrows this uncertainty to an unsubstantial constant.

17 See, e.g., QM Secs. 2.10 and 5.4.
\[ E_m = \hbar \omega \left( m + \frac{1}{2} \right), \quad \text{where } m = 0, 1, 2, \ldots \] (2.38)

If \( \omega \) is kept constant, the zero-point energy \( \frac{\hbar \omega}{2} \) does not contribute to any thermodynamic properties of the system and may be ignored,\(^{18}\) so that for the sake of simplicity we may take that point as the energy origin, and replace Eq. (38) with \( E_m = m \hbar \omega \). Let us carry out an approximate analysis of the system for the case when its average energy per oscillator, 

\[ E = \frac{E_N}{N}, \] (2.39)

is much larger than the energy quantum \( \hbar \omega \). For one oscillator, the number of states with energy \( \varepsilon_1 \) below certain value \( = E_1 >> \hbar \omega \) is evidently \( \Sigma(E_1) \approx E_1/\hbar \omega \) (Fig. 3a). For two oscillators, all possible values of the total energy \( (\varepsilon_1 + \varepsilon_2) \) below some level \( E_2 \) correspond to the points of a 2D square grid within the right triangle shown in Fig. 3b, giving \( \Sigma(E_2) \approx (1/2)(E_2/\hbar \omega)^2 \). For three oscillators, the possible values of the total energy \( (\varepsilon_1 + \varepsilon_2 + \varepsilon_3) \) correspond to those points of the 3D cubic mesh, that fit inside the right pyramid shown in Fig. 3c, giving \( \Sigma(E_3) \approx (1/2)(1/3)(E_3/\hbar \omega)^3 = (1/3!)(E_3/\hbar \omega)^3, \) etc.

Fig. 2.3. Calculating functions \( \Sigma(E_N) \) for the systems of (a) one, (b) two and (c) three quantum oscillators.

An evident generalization of these formulas to arbitrary \( N \) gives the number of states

\[ \Sigma(E_N) \approx \frac{1}{N!} \left( \frac{E_N}{\hbar \omega} \right)^N, \] (2.40)

where coefficient \( 1/N! \) has the geometrical meaning of the (hyper)volume of the \( N \)-dimensional right pyramid with unit sides. Differentiating Eq. (40), we get

\[ g(E_N) = \frac{d\Sigma(E_N)}{dE_N} = \frac{1}{(N-1)!} \left( \frac{E_N}{\hbar \omega} \right)^{N-1}, \] (2.41)

so that

\[^{18}\text{Let me hope that the reader knows that the zero-point energy is experimentally measurable – for example using the famous Casimir effect - see, e.g., QM Sec. 9.1. In Sec. 5.6 below we will discuss another method of experimental observation of that energy.}\]
\[ S_N(E_N) = \ln g(E_N) + \text{const} = -\ln[(N-1)!] + (N-1)\ln E_N - N \ln(h\omega) + \text{const.} \quad (2.42) \]

For \( N \gg 1 \) we can ignore the difference between \( N \) and \( (N-1) \) in both instances, and use the Stirling formula (27) to simplify this result as

\[ S_N(E) - \text{const} \approx N \left( \ln \frac{E_N}{Nh\omega} + 1 \right) \approx N \left( \ln \frac{E}{h\omega} \right) = \ln \left( \frac{E}{h\omega} \right)^N. \quad (2.43) \]

(The second approximation is only valid at very high \( E/h\omega \) ratios, when the logarithm in Eq. (43) is substantially larger than 1, i.e. is rather crude.\(^19\)) Returning for a second to the density of states, we see that in the limit \( N \to \infty \), it is exponentially large:

\[ g(E_N) = e^{S_N} \approx \left( \frac{E}{h\omega} \right)^N, \quad (2.44) \]

so that both conditions (37) may be satisfied within a very broad range of \( \Delta E \).

Now we can use Eq. (43) to find all thermodynamic properties of the system, though only in the limit \( E \gg h\omega \). Indeed, according to thermodynamics (see Sec. 1.2), if the system volume and number of particles are fixed, the derivative \( dS/dE \) is nothing more than the reciprocal temperature – see Eqs. (1.9) or (1.15). In our current case, we imply that the harmonic oscillators are distinct, for example by their spatial positions. Hence, even if we can speak of some volume of the system, it is certainly fixed.\(^20\) Differentiating Eq. (43) over energy \( E \), we get

\[ \frac{1}{T} = \frac{dS_N}{dE_N} = \frac{N}{E_N} = \frac{1}{E}. \quad (2.45) \]

Reading this result backwards, we see that the average energy \( E \) of a harmonic oscillator equals \( T \) (i.e. \( k_B T_K \) is SI units). As we will show in Sec. 5 below, this is the correct asymptotic form of the exact result, valid in our current limit \( E \gg h\omega \).

Result (45) may be readily generalized. Indeed, in quantum mechanics a harmonic oscillator with eigenfrequency \( \omega \) may be described by Hamiltonian

\[ \hat{H} = \frac{\hat{p}^2}{2m} + \frac{\kappa \hat{q}^2}{2}, \quad (2.46) \]

where \( q \) is some generalized coordinate, and \( p \) the corresponding generalized momentum, \( m \) is oscillator’s mass,\(^21\) and \( \kappa \) is the spring constant, so that \( \omega = (\kappa/m)^{1/2} \). Since in thermodynamic equilibrium the density matrix is always diagonal (see Sec. 1 above) in basis of stationary states \( m \), quantum-mechanical averages of the kinetic and potential energies may be found from Eq. (7):

\(^{19}\) Let me offer a very vivid example how slowly does the logarithm function grow at large values of its argument: \( \ln \) of the number of atoms in the visible Universe is less than 200.

\(^{20}\) By the same reason, the notion of pressure \( P \) in such a system is not clearly defined, and neither are any thermodynamic potentials but \( E \) and \( F \).

\(^{21}\) Let me hope that using the same letter for the mass and the state number would not lead to reader’s confusion. I believe that the difference between these uses is very clear from the context.
\[ \langle \frac{p^2}{2m} \rangle = \sum_{m=0}^{\infty} W_m \langle m | \frac{\hat{p}^2}{2m} | m \rangle, \quad \langle \frac{k_\eta^2}{2} \rangle = \sum_{m=0}^{\infty} W_m \langle m | \frac{k_\eta^2}{2} | m \rangle, \]  

\[ (2.47) \]

where \( W_m \) is the probability to occupy \( m \)-th energy level, and bra- and ket-vectors describe the stationary state corresponding to that level.\(^{22}\) However, both classical and quantum mechanics teach us that for any \( m \), the bra-kets under the sums in Eqs. (47), which present the average kinetic and mechanical energies of the oscillator on its \( m \)-th energy level, are equal to each other, and hence each of them is equal to \( E_m/2 \). Hence, even though we do not know the probability distribution \( W_m \) yet (it will be calculated in Sec. 5 below), we may conclude that in the “classical limit” \( T \gg \hbar \omega \),

\[ \langle \frac{p^2}{2m} \rangle = \langle \frac{k_\eta^2}{2} \rangle = \frac{T}{2}. \]

\[ (2.48) \]

Now let us consider a system with an arbitrary number of degrees of freedom, described by a more general Hamiltonian:\(^{23}\)

\[ \hat{H} = \sum_j \hat{H}_j, \quad \hat{H}_j = \frac{\hat{p}_j^2}{2m_j} + \frac{k_j \hat{q}_j^2}{2}, \]

\[ (2.49) \]

with (generally, different) eigenfrequencies \( \omega_j = (\kappa/m_j)^{1/2} \). Since the “modes” (effective harmonic oscillators), contributing into this Hamiltonian, are independent, result (48) is valid for each of the modes. This is the famous \textit{equi-partition theorem}: at thermal equilibrium with \( T \gg \hbar \omega_j \), the average energy of each so-called \textit{half-degree of freedom} (which are defined as variables \( p_j \) or \( q_j \), giving a quadratic term to the system’s Hamiltonian), is equal to \( T/2 \).\(^{24}\) In particular, for each Cartesian coordinate \( q_j \) of a free-moving, non-interacting particle this theorem is valid for \textit{any} temperature, because such coordinates may be considered as 1D harmonic oscillators with vanishing potential energy, i.e. \( \omega_j = 0 \), so that condition \( T \gg \hbar \omega_j \) is fulfilled at any temperature.

At this point, a first-time student of thermodynamics should be very much relieved to see that the counter-intuitive thermodynamic definition (1.9) of temperature does indeed correspond to what we all have known about this notion from our kindergarten physics courses.

I believe that our case study of quantum oscillator systems has been a fair illustration of both the strengths and weaknesses of the microcanonical ensemble approach.\(^{25}\) On one hand, we could calculate virtually everything we wanted in the classical limit \( T \gg \hbar \omega \), but calculations for arbitrary \( T \sim \hbar \omega \), though possible, are difficult, because for that, all vertical steps of function \( \Sigma(E_N) \) have to be carefully

---

\(^{22}\) Note again that though we have committed the energy \( E_N \) of \( N \) oscillators to be fixed (in order to apply Eq. (36), valid only for a microcanonical ensemble at thermodynamic equilibrium), single oscillator’s energy \( E \) in our analysis may be arbitrary - within limits \( \hbar \omega \ll E < E_N \sim NT \).

\(^{23}\) As a reminder, the Hamiltonian of any system whose classical Lagrangian function is an arbitrary quadratic form its generalized coordinates and the corresponding generalized velocities, may be brought to form (49) by an appropriate choice of “normal coordinates” \( q_j \) which are certain linear combinations of the original coordinates – see, e.g., CM Sec. 5.2.

\(^{24}\) This also means that in the classical limit, the heat capacity of a system is equal to the number of its half-degrees of freedom (in SI units, multiplied by \( k_b \)).

\(^{25}\) The reader is strongly urged to solve Exercise 2, whose task is to do a similar calculation for another key (“two-level”) physical system, and to compare the results.
counted. In Sec. 4, we will see that other statistical ensembles are much more convenient for such calculations.

Let me conclude this discussion of entropy with a short notice on deterministic classical systems with a few degrees of freedom (and even simpler mathematical objects called “maps”) that may exhibit essentially disordered behavior, called the deterministic chaos.\(^{26}\) Such chaotic system may be approximately characterized by an entropy defined similarly to Eq. (29), where \(W_m\) are probabilities to find it in different small regions of phase space, at well separated time intervals. On the other hand, one can use an equation slightly more general than Eq. (29) to define the so-called Kolmogorov (or “Kormogorov-Sinai”) entropy \(K\) that characterizes the speed of loss of information about the initial state of the system, and hence what is called the “chaos’ depth”. In the definition of \(K\), the sum over \(m\) is replaced with the summation over all possible permutations \(\{m\} = m_0, m_1, \ldots, m_{N-1}\) of small space regions, and \(W_m\) is replaced with \(W_{\{m\}}\), the probability of finding the system in the corresponding regions \(m\) at time moment \(t_m\), with \(t_m = m \tau\), in the limit \(\tau \to 0\), with \(N \tau = \text{const}\). For chaos in the simplest objects, 1D maps, \(K\) is equal to the Lyapunov exponent \(\lambda > 0\).\(^{27}\) For systems of higher dimensionality, which are characterized by several Lyapunov exponents \(\lambda_i\), the Kolmogorov entropy is equal to the phase-space average of the sum of all positive \(\lambda_i\). These facts provide a much more practicable way of (typically, numerical) calculation of the Kolmogorov entropy than the direct use of its definition.\(^{28}\)

2.3. Maxwell’s Demon, information, and computation

Before proceeding to other statistical distributions, I would like to address one more popular concern about Eq. (24), the direct relation between the entropy and information. Some physicists are still uneasy with entropy being nothing else than the (deficit of) information,\(^{29}\) though to the best of my knowledge, nobody has yet been able to suggest any experimentally verifiable difference between these two notions. Let me give one example of their direct relation, that is essentially a development of the thought experiment suggested by Maxwell as early as in 1867.

Consider a volume containing just one molecule (considered as a point particle), and separated to two equal halves by a movable partition with a door that may be opened and closed at will, at no energy cost (Fig. 4a). If the door is open and the system is in thermodynamic equilibrium, we do not know on which side of the door partition the molecule is. Here the disorder (and hence entropy) are largest, and there is no way to get, from a large ensemble of such systems, any useful mechanical energy.

Now, let us consider that we (as instructed by, in Lord Kelvin’s formulation, an omniscient Maxwell’s Demon) know which side of the partition the molecule is currently located. Then we may

---

\(^{26}\) See, e.g., CM Chapter 9 and literature therein.

\(^{27}\) For the definition of \(\lambda\), see, e.g., CM Eq. (9.9).

\(^{28}\) For more discussion, see, e.g., either Sec. 6.2 of the monograph H. G. Schuster and W. Just, Deterministic Chaos, 4th ed., Wiley-VHS, 2005, or the monograph by Arnold and Avez, cited in Sec. 1.

\(^{29}\) While some of these concerns should be treated with due respect (because the ideas of entropy and disorder are indeed highly nontrivial), I have repeatedly run into rather shallow arguments which stemmed from arrogant contempt to the information theory as an “engineering discipline”, and unwillingness to accept the notion of information on the equal footing with those of space, time, and energy. Fortunately, most leading physicists are much more flexible, and there are even opposite extremes such as J. A. Wheeler’s “it from bit” (i.e. matter from information) philosophy – to which I cannot fully subscribe either.
close the door, so that molecule’s impacts on the partition create, on the average, a pressure force $\not{\mathcal{F}}$ directed toward the empty part of the volume (in Fig. 4b, the right one). Now we can get from the molecule some mechanical work, say by allowing force $\not{\mathcal{F}}$ to move the partition to the right, and picking up the resulting mechanical energy by some deterministic external mechanism. After the partition has been moved to the right end of the volume, we can open the door again (Fig. 4c), equalizing the molecule’s average pressure on both sides of the partition, and then slowly move the partition back to the middle of the volume, without doing any substantial work. With the kind help by Maxwell’s Demon, we can repeat the cycle again and again, and hence make the system to do unlimited mechanical work, fed “only” by information and thermal motion, and thus implementing the perpetual motion machine of the 2nd kind – see Sec. 1.6. The fact that such heat engines do not exist means that the Maxwell’s Demon does not either: getting any new information, at nonvanishing temperature (i.e. at thermal agitation of particles) has a finite energy cost.

![Fig. 2.4. The Maxwell’s Demon paradox: the volume with a single molecule (a) before and (b) after closing the door, and (c) after opening the door in the end of the expansion stage.](a) (b) (c)

In order to evaluate this cost, let us calculate the maximum work per cycle made by the Maxwell’s heat engine (Fig. 4), assuming that it is constantly in thermal equilibrium with a heat bath of temperature $T$. Formula (21) tells us that the information supplied by the demon (what exactly half of the volume contains the molecule) is exactly one bit, $I(2) = 1$. According to Eq. (24), this means that by getting this information we are reducing entropy by $\Delta S_I = -\ln 2$. Now, it would be a mistake to plug this (negative) entropy change into Eq. (1.19). First, that relation is only valid for slow, reversible processes. Moreover (and more importantly), this equation, as well as its irreversible version (1.41), is only valid for a fixed statistical ensemble. The change $\Delta S_I$ does not belong to this category, and may be formally described by the change of the statistical ensemble – from the one consisting of all similar systems (experiments) with an unknown location of the molecule, to the new ensemble consisting of the systems with the molecule in its certain (in Fig. 4, left) half.\[30\]

Now let us consider the slow expansion of the “gas” after the door had been closed. At this stage, we do not need the demon’s help any longer (i.e. the statistical ensemble is fixed), and we can use relation (1.19). At the assumed isothermal conditions ($T = \text{const}$), this relation may be integrated over the whole expansion process, getting $\Delta Q = T \Delta S$. At the finite position, the system’s entropy should be the same as initially, i.e. before the door had been opened, because we again do not know where in the volume the molecule is. This means that the entropy was replenished, during the reversible expansion,

\[30\] This procedure of redefining the statistical ensemble is the central point of the connection between the information theory and physics, and is crucial in particular for any (meaningful :-) discussion of measurements in quantum mechanics - see, e.g., QM Secs. 2.5 and 7.7.
from the heat bath, by $\Delta S = -\Delta S_i = +\ln 2$, so that $\Delta Q = T\Delta S = T\ln 2$. Since by the end of the whole cycle the internal energy $E$ of the system is the same as before, all this heat should have gone into the mechanical energy obtained during the expansion. Thus the obtained work per cycle (i.e. for each obtained information bit) is $T\ln 2 \ (k_B T \ln 2$ in SI units), about $4 \times 10^{-21}$ Joule at room temperature. This is exactly the minimum energy cost of getting one bit of new information about a system at temperature $T$.

The smallness of that amount on the everyday human scale has left the Maxwell’s Demon paradox an academic exercise for almost a century. However, its discussion resumed in the 1960s in the context of energy consumption at numerical calculations, motivated by the exponential (Moore’s-law) progress of the digital integrated circuits, which leads in particular, to a fast reduction of energy $\Delta E$ “spent” (turned into heat) per one binary logic operation. In the current generations of semiconductor digital integrated circuits, $\Delta E$ is of the order of $\sim 10^{-16}$ J,\textsuperscript{31} i.e. still exceeds the room-temperature value of $T\ln 2 = k_B T \ln 2 \approx 3 \times 10^{-21}$ J by more than 4 orders of magnitude. Still, some engineers believe that thermodynamics imposes an important lower limit on $\Delta E$ and hence presents an insurmountable obstacle to the future progress of computation,\textsuperscript{32} so that the issue deserves a discussion.

Let me believe that the reader of these notes understands that, in contrast to naïve popular thinking, computers do not create any new information; all they can do it to reshape (process) it, losing most of input information on the go. Indeed, any digital computation algorithm may be decomposed into simple, binary logical operations, each of them performed by a certain logic circuit called the logic gate. Some of these gates (e.g., logical NOT performed by inverters, as well as memory READ and WRITE operations) do not change the amount of information in the computer. On the other hand, such information-irreversible logic gates as two-input NAND (or NOR, or XOR, etc.) actually erase one bit at each operation, because they turn two input bits into one output bit (Fig. 5a).

In 1961, R. Landauer arrived at the conclusion that each logic operation should turn into heat at least energy

\textsuperscript{31} In the dominating CMOS technology, $\Delta E$ is close to twice the energy $CV^2/2$ of recharging the total capacitance $C$ of the transistor gate electrodes and the wires interconnecting the gates, by the voltage $V$ representing the binary unity. As the technology progresses, $C$ decreases in approximate proportion with the minimum feature size, resulting in the almost proportional decrease of $\Delta E$. (The used voltage $V$ has almost saturated at ~1 V – the value that stems from the bandgap of ~1 eV of the used semiconductor – silicon.)

\textsuperscript{32} Unfortunately, this delusion has resulted in a substantial and unjustified shift of electron device research resources toward using “non-charge degrees of freedom” (such as spin) – as if they do not obey the general laws of statistical physics!
\[ \Delta E_{\text{min}} = T \ln 2 = k_B T \ln 2. \]  
(2.51)

This result may be illustrated with the Maxwell’s Demon machine shown in Fig. 4, operating as a heat pump. At the first stage, with the door closed, it uses external mechanical work \( \Delta E = T \ln 2 \) to reduce the volume in which the molecule is confined from \( V \) to \( V/2 \), pumping heat \( -\Delta Q = \Delta E \) into the heat bath. To model a logically-irreversible logic gate, let us now open the door in the partition, and thus lose 1 bit of information about molecule’s position. Then we will never get work \( T \ln 2 \) back, because moving the partition back to the right, with door open, takes place at zero average pressure. Hence, Eq. (51) gives a fundamental limit for energy loss (per bit) at the logically irreversible computation.

However, in 1973 C. Bennett came up with convincing arguments that it is possible to avoid such energy loss by using only operations that are reversible not only physically, but also logically.\(^{33}\) For that, one has to avoid any loss of information, i.e. any erasure of intermediate results, for example in the way shown in Fig. 5b. (For that, gate F should be physically reversible, with no substantial static power consumption.) In the end of all calculations, after the result has been copied into a memory, the intermediate results may be “rolled back” through reversible gate to be eventually merged into a copy of input data, again without erasing a single bit. The minimal energy dissipation at such reversible calculation tends to zero as the operation speed is decreased, so that the average energy loss per bit may be less than the perceived “fundamental thermodynamic limit” (51).\(^{34}\) The price to pay for this ultralow dissipation is an enormous (exponential) complexity of hardware necessary for storage of all intermediate results. However, using irreversible gates sparingly, it may be possible to reduce the complexity dramatically, so that in future the mostly reversible computation may be able to reduce energy consumption in practical digital electronics.\(^{35}\)

Before we leave Maxwell’s Demon behind, let me use it to discuss, for one more time, the relation between the reversibility of the classical and quantum mechanics of Hamiltonian systems and the irreversibility possible in thermodynamics and statistical physics. In our (or rather Lord Kelvin’s :-) gedanken experiment shown in Fig. 4, the laws of mechanics governing the motion of the molecule are reversible all times. Still, at partition’s motion to the right, driven by molecule’s impacts, the entropy grows, because the molecule picks up heat \( \Delta Q > 0 \), and hence entropy \( \Delta S = \Delta Q/T > 0 \), from the heat bath. The physical mechanism of this irreversible entropy (read: disorder) growth is the interaction of the molecule with uncontrollable components of the heat bath, and the resulting loss of information about the motion of the molecule. Philosophically, the emergence of irreversibility in large systems is a strong argument against the reductionism – a naïve belief that knowing the exact laws of Nature at one level of its complexity, we can readily understand all the phenomena on the higher levels of its organization. In reality, the macroscopic irreversibility of large systems is a wonderful example of a new law (in this case, the 2nd law of thermodynamics) that becomes relevant on the substantially new level of complexity - without defying the lower-level laws. Without such new laws, very little of the higher level organization of Nature may be understood.

\(^{33}\) C. Bennett, IBM J. Res. Devel. 17, 525 (1973); see also a later review C. Bennett, Int. J. Theor. Phys. 21, 905 (1982). To the best of my knowledge, the sub-\( T \ln 2 \) energy loss per logic step is still to be demonstrated experimentally, but at least one research team is closing at this goal.

\(^{34}\) Reversible computation may also overcome the perceived “fundamental quantum limit”, \( \Delta E \Delta t > \hbar \), where \( \Delta t \) is the time scale of the binary logic operation – see K. Likharev, Int. J. Theor. Phys. 21, 311 (1982).

\(^{35}\) The situation is rather different for quantum computation which may be considered as a specific type of reversible but analog computation – see, e.g., QM Sec. 8.5 and references therein.
2.4. Canonical ensemble and the Gibbs distribution

As we have seen in Sec. 2, the microcanonical distribution may be directly used for solving some simple problems, but a further development of this approach (also due to J. Gibbs) turns out to be much more convenient for calculations. Let us consider that a statistical ensemble of similar systems we are studying, each in thermal equilibrium with a much larger heat bath of temperature \( T \) (Fig. 6a). Such an ensemble is called canonical.

![Diagram](image)

Fig. 2.6. (a) System in a heat bath (a canonical ensemble member) and (b) energy spectrum of the composite system (including the heat bath).

Next, it is intuitively evident that if the heat bath is sufficiently large, any thermodynamic variables characterizing the system under study should not depend on the heat bath’s environment. In particular, we may assume that the heat bath is thermally insulated; then the total energy \( E_\Sigma \) of the composite system (consisting of the system of our interest, plus the heat bath) does not change in time. For example, if our system of interest is on its certain (say, \( m^{th} \)) energy level, then

\[
E_\Sigma = E_m + E_{HB}
\]

is conserved. Now let us partition this canonical ensemble into much smaller sub-ensembles, each being a microcanonical ensemble of composite systems whose total energy \( E_\Sigma \) is the same - as discussed in Sec. 2, within a certain small energy interval \( \Delta E_\Sigma \ll E_\Sigma \). According to the microcanonical distribution, probabilities to find the composite system, within this new ensemble, in any state are equal. Still, heat bath energies \( E_{HB} = E_\Sigma - E_m \) (Fig. 6b) of members of this microcanonical sub-ensemble may be different due to the difference in \( E_m \).

The probability \( W(E_m) \) to find the system of our interest (within the selected sub-ensemble) on some energy level \( E_m \) is proportional to the number \( \Delta M \) of such systems in the sub-ensemble. Due to the very large size of the heat bath in comparison with that of the system under study, the heat bath’ density of states \( g_{HB} \) is very high, and \( \Delta E_\Sigma \) may be selected so that

\[
\frac{1}{g_{HB}} \ll \Delta E_\Sigma \ll |E_m - E_m'| \ll E_{HB},
\]

where \( m \) and \( m' \) are any states of the system of our interest. As Fig. 6b shows, in this case we may write \( \Delta M = g_{HB}(E_{HB})\Delta E_\Sigma \). As a result, within the microcanonical ensemble with the total energy \( E_\Sigma \),

---

36 See also exercise problems listed in the end of this chapter.
\[ W_m \propto \Delta M = g_{\text{HB}}(E_{\text{HB}}) \Delta E \approx g_{\text{HB}}(E_{\Sigma} - E_m) \Delta E_{\Sigma}. \] (2.54)

Let us simplify this expression further, using the Taylor expansion with respect to relatively small \( E_m \ll E_{\Sigma} \). However, here we should be careful. As we have seen in Sec. 2, the density of states of a large system is an extremely rapidly growing function of energy, so that if we applied the Taylor expansion directly to Eq. (54), the Taylor series would converge for very small \( E_m \) only. A much broader applicability range may be obtained by taking logarithm of both parts of Eq. (54) first:

\[ \ln W_m = \text{const} + \ln [g_{\text{HB}}(E_{\Sigma} - E_m)] = \ln \Delta E_{\Sigma} = \text{const} + S_{\text{HB}}(E_{\Sigma} - E_m), \] (2.55)

where the second equality results from application of Eq. (56) to the heat bath, and \( \ln \Delta E_{\Sigma} \) has been incorporated into the constant. Now, we can Taylor-expand the (much more smooth) function of energy in the right-hand part, and limit ourselves to the two leading terms of the series:

\[ \ln W_m \approx \text{const} + S_{\text{HB}} \left| \frac{dS_{\text{HB}}}{dE_{\text{HB}}} \right|_{E_m=0} E_m \Delta E_{\Sigma}. \] (2.56)

But according to Eq. (1.9), the derivative participating in this expression is nothing else than the reciprocal heat bath temperature that (due to the large bath size) does not depend on whether \( E_m \) is equal to zero or not. Since our system of interest is in the thermal equilibrium with the bath, this is also the temperature \( T \) of the system – see Eq. (1.8). Hence Eq. (56) is merely

\[ \ln W_m = \text{const} - \frac{E_m}{T}. \] (2.57)

This equation describes a substantial decrease of \( W_m \) as \( E_m \) is increased by several \( T \), and hence our linear approximation (56) is virtually exact as soon as \( E_{\text{HB}} \) is much larger than \( T \) – the condition that is rather easy to satisfy, because as we have seen in Sec. 2, the average energy of each particle is of the order of \( T \).

Now we should be careful again, because so far we have only derived Eq. (57) for a sub-ensemble with fixed \( E_{\Sigma} \). However, since the right-hand part of Eq. (57) includes only \( E_m \) and \( T \) that are independent of \( E_{\Sigma} \), this relation is valid for all sub-ensembles of the canonical ensemble, and hence for the later ensemble as the whole.\(^{37}\) Hence for the total probability to find our system of interest in state with energy \( E_m \), in the canonical ensemble with temperature \( T \), we can write

\[ W_m = \text{const} \times \exp \left( -\frac{E_m}{T} \right) \equiv \frac{1}{Z} \exp \left\{ -\frac{E_m}{T} \right\}. \] (2.58)

This is the famous Gibbs distribution (sometimes called the “canonical distribution”),\(^{38}\) which is frequently arguably the summit of statistical physics,\(^{39}\) because it may be used for a straightforward (or at least conceptually straightforward :-) calculation of all statistical and thermodynamic variables.

---

\(^{37}\) Another way to arrive at the same conclusion is to note that the entropy of the whole canonical ensemble with fixed \( E_m \) has to be a sum of entropies of its microcanonical sub-ensembles (with different \( E_{\Sigma} \)), which participate in Eq. (55). As a result, the logarithm of probability \( W_m \) for our system of interest to have energy \( E_m \) in the whole (canonical) ensemble is just a sum of Eqs. (57) for sub-ensembles with different \( E_{\Sigma} \).

\(^{38}\) The temperature dependence of the type \( \exp \{ -E/T \} \), especially when showing up in rates of certain events, e.g., chemical reactions, is also frequently called the Arrhenius law - after chemist S. Arrhenius who has noticed this
Before I illustrate this, let me first calculate the coefficient $Z$ participating in Eq. (58) for the general case. Requiring, in accordance with Eq. (4), the sum of all $W_m$ to be equal 1, we get

$$Z = \sum_m \exp\left\{-\frac{E_m}{T}\right\},$$

where the summation is formally extended to all quantum states of the system, though in practical calculations, the sum may be truncated to include only the states that are noticeably occupied. This apparently humble normalization coefficient $Z$ turns out to be so important for the relation between the Gibbs distribution (i.e. statistics) and thermodynamics that it has a special name - or actually, two names: either the statistical sum or the partition function. To demonstrate how important $Z$ is, let us use the general Eq. (29) for entropy to calculate its value for the particular case of the canonical ensemble, i.e. the Gibbs distribution of probabilities $W_n$:

$$S = -\sum_m W_m \ln W_m = \ln Z \sum_m \exp\left\{-\frac{E_m}{T}\right\} + \frac{1}{ZT} \sum_m E_m \exp\left\{-\frac{E_m}{T}\right\}.$$

According to the general rule (7), the thermodynamic (i.e. ensemble-average) value $E$ of the internal energy of the system is

$$E = \sum_m W_m E_m = \frac{1}{Z} \sum_m E_m \exp\left\{-\frac{E_m}{T}\right\},$$

so that the second term in the right-hand part of Eq. (60) is just $E/T$, while the first term equals just $\ln Z$, due to the normalization condition (59). (As a parenthetic remark, using the notion of reciprocal temperature $\beta \equiv 1/T$, Eq. (61a), with account of Eq. (59), may be also rewritten as

$$E = -\frac{\partial (\ln Z)}{\partial \beta}.$$

This formula is very convenient for calculations if our prime interest is the average energy $E$ rather than $F$ or $W_n$.) With these substitutions, Eq. (60) yields a very simple relation between the statistical sum and entropy:

$$S = \frac{E}{T} + \ln Z.$$

Using Eq. (1.33), we see that Eq. (62) gives a straightforward way to calculate the free energy $F$ of the system from nothing else than its statistical sum:

$$F = E - TS = T \ln \frac{1}{Z}.$$

law in experimental data. In all cases I am aware of, the Gibbs distribution is the underlying reason of the Arrhenius law.

39 This opinion is shared by several authoritative colleagues, including R. Feynman who climbs on this summit already by page 4 (!) of his brilliant book *Statistical Mechanics*, 2nd ed., Westview, 1998. (Despite its title, this monograph a collection of lectures on a few diverse, mostly advanced topics of statistical physics, rather than its systematic course, so that unfortunately I cannot recommend it as a textbook.)
Now, using the general thermodynamic relations (see especially the circular diagram shown in Fig. 1.7b, and its discussion) we can calculate all thermodynamic potentials of the system, and all other variables of interest. Let me only note that in order to calculate pressure $P$ - e.g., from the second of Eqs. (1.35) - we would need to know the explicit dependence of $F$, and hence of the statistical sum $Z$ on the system volume $V$. This would require the calculation, by appropriate methods of either classical or quantum mechanics, of the volume dependence of eigenenergies $E_m$. I will give numerous examples of such calculations later in the course.\footnote{In many multiparticle systems, the effect of an external field may be presented as a sum of its effects on each particle - frequently described by interaction energy with structure $-\mathcal{F}_j q_j^{(k)}$, where $q_j^{(k)}$ is a generalized coordinate of $k$-th particle. Generally, this energy has to be included directly into energies of particle states $E_m$, used in $Z$, and hence in the free energy $F$ (63). In this case, the thermodynamic equilibrium corresponds to the minimum of $F$ – see Eq. (1.42). On the other hand, for “linear” systems (whose energy is a quadratic-homogeneous form of its generalized coordinates and velocities), equivalent results may be obtained by accounting for the interaction at the thermodynamic level, i.e. by subtracting term $\mathcal{F}_j \langle q_j \rangle = \mathcal{F}_j N \langle q_j^{(k)} \rangle$ from the free energy $F$ calculated in the absence of the field, and then finding the equilibrium as a minimum of the resulting Gibbs energy $G$ – see Eq. (1.43). In this case, any of the approaches is fine, provided only that the same interaction is not counted twice.}

As the final note of this section, Eqs. (59) and (63) may be combined to give a very elegant expression,

$$\exp\left\{ -\frac{F}{T} \right\} = \sum_m \exp\left\{ -\frac{E_m}{T} \right\},$$

which offers a convenient interpretation of free energy as a (rather specific) average of eigenenergies of the system. One more convenient formula may be obtained by using Eq. (64) to rewrite the Gibbs distribution (58) in the form

$$W_m = \exp\left\{ \frac{F - E_m}{T} \right\}.$$  \hspace{1cm} (2.65)

In particular, this expression shows that that since all probabilities $W_m$ are below 1, $F$ is always lower than the lowest energy level. Also, note that probabilities $W_m$ do not depend on the energy reference choice, i.e. on an arbitrary constant added to all $E_m$ (and hence to $E$ and $F$).

### 2.5. Harmonic oscillator statistics

The last property may be immediately used in our first example of the Gibbs distribution application to a particular, but very important system - the harmonic oscillator, for the more general case then was done in Sec. 2, namely for a “quantum oscillator” with an arbitrary relation between $T$ and $\hbar\omega$.\footnote{A simpler task of making a similar calculation for another key quantum-mechanical object, the two-level system, is left for reader’s exercise.} Let us consider a canonical ensemble of similar oscillators, each in a contact with a heat bath of temperature $T$. Selecting the zero-point energy $\hbar\omega/2$ for the origin of $E$, oscillator eigenenergies (38) become $E_m = m\hbar\omega (m = 0, 1, \ldots)$, so that the Gibbs distribution for probabilities of these states is

$$W_m = \frac{1}{Z} \exp\left\{ -\frac{E_m}{T} \right\} = \frac{1}{Z} \exp\left\{ -\frac{m\hbar\omega}{T} \right\},$$  \hspace{1cm} (2.66)
with the statistical sum

\[ Z \equiv \sum_{m=0}^{\infty} \exp\left\{ -\frac{m\hbar \omega}{T} \right\} = \sum_{m=0}^{\infty} \lambda^m, \quad \text{where} \quad \lambda \equiv \exp\left\{ -\frac{\hbar \omega}{T} \right\} \leq 1. \quad (2.67) \]

This series is just an infinite geometric progression (“geometric series”); summing it,\(^{42}\) we get

\[ Z = \frac{1}{1-\lambda} = \frac{1}{1-e^{-\hbar \omega/T}}, \quad \text{(2.68)} \]

so that for the probability \( W_m \) to find the oscillator at each energy level is

\[ W_m = \left(1 - e^{-\hbar \omega/T}\right) e^{-m\hbar \omega/T}. \quad (2.69) \]

As Fig. 7a shows, the probability \( W_m \) to find the oscillator in each particular state (but the ground one, with \( m = 0 \)) vanishes in both low- and high-temperature limits, and reaches its maximum value \( W_m \sim 0.3/m \) at \( T \sim m\hbar \omega \), so that the contribution \( m\hbar \omega W_m \) of each level into the average oscillator energy \( E \) is always smaller than \( \hbar \omega \).

This average energy may be calculated in any of two ways: either using Eq. (7):

\[ E = \sum_{m=0}^{\infty} E_m W_m = \left(1 - e^{-\hbar \omega/T}\right) \sum_{m=0}^{\infty} m\hbar \omega e^{-m\hbar \omega/T}, \quad (2.70) \]

or (simpler) using Eq. (61b), as

\[ \text{Fig. 2.7. Statistical and thermodynamic parameters of a harmonic oscillator, as functions of temperature.} \]

\(^{42}\) See, e.g., MA Eq. (2.8b).
Both methods give (of course) the same famous result,\(^4\)

\[
E = E(\omega, T) = \hbar \omega \frac{1}{e^{\hbar \omega / T} - 1},
\]

which is valid for arbitrary temperature and plays the key role in many fundamental problems of physics. The red line in Fig. 7b shows this \(E\) as a function of normalized temperature. At low temperatures, \(T \ll \hbar \omega\), the oscillator is predominantly in its lowest (ground) state, and its energy (on top of the constant zero-point energy \(\hbar \omega / 2\)) is exponentially small:

\[
E \approx \hbar \omega \exp\{-\hbar \omega / T\} \ll T, \hbar \omega.
\]

On the other hand, in the high-temperature limit the energy tends to \(T\). This is exactly the result (a particular case of the equipartition theorem) that was obtained in Sec. 2 from the microcanonical distribution.

Please note how much simpler is the calculation starting from the Gibbs distribution, even for an arbitrary ratio \(T/\hbar \omega\).

To complete the discussion of thermodynamic properties of the harmonic oscillator, we can calculate its free energy using Eq. (63):

\[
F = T \ln \frac{1}{Z} = T \ln \left(1 - e^{-\hbar \omega / T}\right).
\]

Now entropy may be found from thermodynamics: either from the first of Eqs. (1.35), \(S = -(\partial F / \partial T)_V\), or (even more easily) from Eq. (1.33): \(S = (E - F) / T\). Both relations give, of course, the same result:

\[
S = \frac{\hbar \omega}{T} \frac{1}{e^{\hbar \omega / T} - 1} - \ln \left(1 - e^{-\hbar \omega / T}\right).
\]

Finally, since in the general case the dependence of the oscillator properties (essentially, \(\omega\)) on volume \(V\) in this problem is not specified, such variables as \(P, \mu, G, W,\) and \(\Omega\) are not defined, and we may calculate only the average heat capacity \(C\) per one oscillator:

\[
C = \frac{\partial E}{\partial T} = \left(\frac{\hbar \omega}{T}\right)^2 \frac{e^{\hbar \omega / T}}{\left(e^{\hbar \omega / T} - 1\right)^2} = \left(\frac{\hbar \omega / 2T}{\sinh(\hbar \omega / 2T)}\right)^2.
\]

The calculated thermodynamic variables are shown in Fig. 7b. In the low-temperature limit \(T \ll \hbar \omega\), they all tend to zero. On the other hand, in the high-temperature limit \(T \gg \hbar \omega\), \(F \rightarrow -T \ln(T / \hbar \omega) \rightarrow -\infty\), \(S \rightarrow \ln(T / \hbar \omega) \rightarrow +\infty\), and \(C \rightarrow 1\) (in SI units, \(C \rightarrow k_B\)). Note that the last limit is the direct corollary of the equipartition theorem: each of two “half-degrees of freedom” of the oscillator gives, in the classical limit, a contribution \(C = \frac{1}{2}\) into its heat capacity.

Now let us use Eq. (69) to discuss the statistics of the quantum oscillator described by Hamiltonian (46), in the coordinate representation. Again using density matrix’ diagonality at thermodynamic equilibrium, we may use a relation similar to Eqs. (47) to calculate the probability density to find the oscillator at coordinate \(q\):

\footnote{It was first obtained in 1924 by S. Bose, and is frequently called the Bose distribution – a particular case of the Bose-Einstein distribution - to be discussed in Sec. 8 below.}
\[ w(q) = \sum_{m=0}^{\infty} W_m \psi_m(q) = \sum_{m=0}^{\infty} W_m |\psi_m(q)|^2 = \left(1 - e^{-\hbar \omega / T}\right) \sum_{m=0}^{\infty} e^{-m \hbar \omega / T} |\psi_m(q)|^2, \quad (2.76) \]

where \( \psi_m(q) \) is the eigenfunction of \( m \)-th stationary state of the oscillator. Since each \( \psi_m(q) \) is proportional to the Hermite polynomial\(^{44} \) that requires at least \( m \) elementary functions for its representation, working out the sum in Eq. (76) is a bit tricky,\(^{45} \) but the final result is rather simple: \( w(q) \) is just a normalized \textit{Gaussian distribution} (the “bell curve”),

\[ w(q) = \frac{1}{(2\pi)^{1/2} \alpha_q} \exp \left\{ -\frac{q^2}{2(\alpha_q)^2} \right\}, \quad (2.77) \]

with \( \langle q \rangle = 0 \), and

\[ \langle q^2 \rangle = \frac{\hbar}{2m \omega} \coth \frac{\hbar \omega}{2T}. \quad (2.78) \]

Since \( \coth \xi \) tends to 1 at \( \xi \to \infty \), and diverges as \( 1/\xi \) at \( \xi \to 0 \), Eq. (78) shows that the width of coordinate distribution is constant (and equal to that, \( \hbar/2m \omega \), of the ground-state wavefunction \( \psi_0 \)) at \( T \ll \hbar \omega \), and grows as \( T/m \omega^2 \) at \( T/\hbar \omega \to \infty \).

As a sanity check, we may use Eq. (78) to write the following expression,

\[ \langle U \rangle \equiv \left\langle \frac{kq^2}{2} \right\rangle = \frac{\hbar \omega}{4} \coth \frac{\hbar \omega}{2T} \rightarrow \begin{cases} \frac{\hbar \omega}{4}, & \text{at } T \ll \hbar \omega, \\ \frac{T}{2}, & \text{at } \hbar \omega \ll T, \end{cases} \quad (2.79) \]

for the average potential energy of the oscillator. In order to comprehend this result, let us notice that Eq. (72) for the average full energy \( E \) was obtained by counting it from the ground state energy \( \hbar \omega / 2 \) of the oscillator.\(^{46} \) If we add this energy to the result, we get

\[ E = \frac{\hbar \omega}{e^{\hbar \omega / T} - 1} + \frac{\hbar \omega}{2} = \frac{\hbar \omega}{2} \coth \frac{\hbar \omega}{2T}. \quad (2.80) \]

We see that for arbitrary temperature, \( \langle U \rangle = E/2 \), as we already concluded from Eq. (47). This means that the average kinetic energy, equal to \( E - \langle U \rangle \), is also the same:

\[ \left\langle \frac{p^2}{2m} \right\rangle = \left\langle \frac{kq^2}{2} \right\rangle = \frac{E}{2} = \frac{\hbar \omega}{4} \coth \frac{\hbar \omega}{2T}. \quad (2.81) \]

In the classical limit \( T \gg \hbar \omega \), both energies equal \( T/2 \), reproducing the equipartition theorem result (48).

### 2.6. Two important applications

The results of the previous section, especially Eq. (72), have enumerable applications in physics, but I will have time for a brief discussion of only two of them.

---

\(^{44} \) See, e.g., QM Sec. 2.10.

\(^{45} \) The calculation may be found, e.g., in QM Sec. 7.2.

\(^{46} \) As a quantum mechanics reminder, the ground state energy of the oscillator is not only measurable, but is also responsible for several important phenomena, e.g., the Casimir effect – see, e.g., QM Sec. 9.1.
(i) Blackbody radiation. Let us consider a free-space volume $V$ limited by non-absorbing (i.e. ideally reflecting) walls. Electrodynamics tells us that electromagnetic field in such a cavity may be presented as a sum of “modes” with time evolution similar to that of the usual harmonic oscillator, and quantum mechanics says that the energy of such electromagnetic oscillator is quantized in accordance with Eq. (38), so that at thermal equilibrium the average energy is described by Eq. (72). If volume $V$ is large enough, the number of these modes within a small range $dk$ of the wavevector magnitude $k$ is

$$dN = \frac{gV}{(2\pi)^3}d^3k = \frac{gV}{(2\pi)^3}4\pi k^2 dk,$$

where for electromagnetic waves, the degeneracy factor $g=2$, due to their two different (e.g., linear) polarizations for the same wave vector $k$. With the isotropic dispersion relation for waves in vacuum, $k = \omega/c$, the elementary volume $d^3k$ corresponding to a small interval $d\omega$ is a spherical shell of small thickness $dk = d\omega/c$, and Eq. (82) yields

$$dN = \frac{2V}{(2\pi)^3}4\pi \frac{\omega^2 d\omega}{c^3} = V \frac{\omega^2}{\pi^2 c^3} d\omega.$$

Using Eq. (72), we see that the spectral density of electromagnetic wave energy, per unit volume, is

$$u(\omega) \equiv \frac{E}{V} \frac{dN}{d\omega} = \frac{\hbar \omega^3}{\pi^2 c^3} \frac{1}{e^\hbar \omega /T - 1}.$$

This is the famous Planck’s blackbody radiation law. To understand why its name mentions radiation, let us consider a small planar part, of area $dA$, of a surface that completely absorbs electromagnetic waves incident from any direction. (Such “perfect black body” approximation may be closely approached in special experimental structures, especially in limited frequency intervals.) Figure 8 shows that if the arriving wave was planar, with the incidence angle $\theta$, then power $dP(\omega)$ absorbed by the surface within a small frequency interval $d\omega$ (i.e. energy arriving at the surface within unit time interval), would be equal to the radiation energy within the same frequency interval and inside a cylinder of height $c$, base area $dA \cos \theta$, and hence volume $dV = c dA \cos \theta$:

$$dP(\omega) = u(\omega)d\omega dV = u(\omega)d\omega c dA \cos \theta.$$

Since the thermally-induced field is isotropic, i.e. propagates equally in all directions, this results should be averaged over all solid angles within the polar angle interval $0 \leq \theta \leq \pi/2$:

$$\frac{dP(\omega)}{dA d\omega} = \frac{1}{4\pi} \int dP(\omega) d\Omega = cu(\omega) \frac{1}{4\pi} \int_0^{\pi/2} \sin \theta d\theta \int_0^{2\pi} d\varphi \cos \theta = \frac{c}{4} u(\omega).$$

---

47 See, e.g., EM Sec. 7.9.
48 See, e.g., QM Sec. 9.1.
49 In our current context, the volume should be much larger than $(c\hbar /T)^3$, where $c \approx 3 \times 10^8$ m/s is the speed of light. For room temperature ($T \approx k_B \times 300K \approx 4 \times 10^{21}$ J), that lower bound is of the order of $10^{-16}$ m$^3$.
50 See, e.g., EM Sec. 7.9, or QM Sec. 1.6.
51 Let me hope the reader knows that the law was first suggested in 1900 by M. Planck as an empirical fit for the experimental data on blackbody radiation, and this was the historic point at which the Planck constant $\hbar$ (or rather $\hbar \equiv 2\pi \hbar$) was introduced - see, e.g., QM Sec. 1.1.
Hence the Planck’s formula (84), multiplied by \( c/4 \), gives the power absorbed by such “blackbody” surface. But at thermal equilibrium, this absorption has to be exactly balanced by the surface’s own radiation, due to its finite temperature \( T \).

\[
\text{Fig. 2.8. Calculating the relation between } dP(\omega) \text{ and } u(\omega)d\omega.
\]

I am confident that the reader is familiar with the main features of the Planck law (84), including its general shape (Fig. 9), with the low-frequency asymptote \( u(\omega) \propto \omega^2 \) (due to its historic significance bearing the special name of the Rayleigh-Jeans law), the exponential drop at high frequencies (the Wien law), and the resulting maximum of function \( u(\omega) \), reached at frequency \( \omega_{\text{max}} \),

\[
h \omega_{\text{max}} \approx 2.28T, \quad (2.87)
\]

i.e. at wavelength \( \lambda_{\text{max}} = 2\pi k_{\text{max}} = 2\pi c/\omega_{\text{max}} \approx 2.76 \text{ cm}/T \). Still, I cannot help mentioning two particular values corresponding to visible light (\( \lambda_{\text{max}} \approx 500 \text{ nm} \)) for Sun’s surface temperature \( T_K \approx 6,000 \text{ K} \), and to mid-infrared range (\( \lambda_{\text{max}} \approx 10 \mu\text{m} \)) for the Earth’s surface temperature \( T_K \approx 300 \text{ K} \). The balance of these two radiations, absorbed and emitted by the Earth, determines its surface temperature, and hence has the key importance for all life on our planet. As one more example, the cosmic microwave background (CMB) radiation, closely following the Planck law with \( T_K = 2.725 \text{ K} \) (and hence having maximum density at \( \lambda_{\text{max}} \approx 1.9 \text{ mm} \)), and in particular its weak anisotropy, is a major source of data for all modern cosmology.\(^{52}\)

\[
\text{Fig. 2.9. Frequency dependence of the blackbody radiation density, normalized by } u_0 \equiv T^3/\pi^2h^2c^3, \text{ according to the Planck law (red line) and the Rayleigh-Jeans law (blue line).}
\]

\(^{52}\) For a recent popular book of this topic, see, e.g., S. Singh, \textit{Big Bang: The Origins of the Universe}, HarperCollins, 2005.
Now let us calculate the total energy $E$ of this radiation in some volume $V$. It may be found from Eq. (72) by integration its over all frequencies:\(^{53}\)

$$E = V \int_0^\infty u(\omega) d\omega = V \int_0^\infty \frac{\hbar \omega^3}{\pi^2 c^3} \frac{d\omega}{e^{\hbar \omega/T} - 1} = \frac{VT^4}{\pi^2 c^3} \int_0^\infty \frac{x^3}{e^x - 1} dx = \frac{V}{15\hbar^3 c^3} T^4. \quad (2.88)$$

(The last transition in Eq. (88) uses a table integral equal to $\Gamma(4)\zeta(4) = (3!)^2/(90) = \pi^4/15.\) Using Eq. (86) to recast Eq. (88) into the total power radiated by a blackbody surface, we get the well-known Stefan (or “Stefan-Boltzmann”) law

$$\frac{dP}{dA} = \frac{\pi^2}{60\hbar^3 c^3} T^4 = \sigma T^4, \quad (2.89a)$$

where $\sigma$ is the Stefan-Boltzmann constant

$$\sigma \equiv \frac{\pi^2}{60\hbar^3 c^3} k_B^4 \approx 5.67 \times 10^{-8} \frac{W}{m^2 K^4}. \quad (2.89b)$$

By this time, the thoughtful reader should have an important concern ready: Eq. (84) and hence Eq. (88) are based on Eq. (72) for the average energy of each oscillator, counted from its ground energy $\hbar \omega/2$. However, the radiation power should not depend on the energy origin; why have we not included the ground energy of each oscillator into integration (88), as we have done in Eq. (80)? The answer is that usual radiation detectors only measure the difference between power $P_{\text{in}}$ of the incident radiation (say, that of a blackbody surface with temperature $T$) and their own back-radiation $P_{\text{out}}$ with power corresponding to some effective temperature $T_d$ of the detector (Fig. 10). But however low $T_d$ is, the temperature-independent ground state energy contribution $\hbar \omega/2$ to the back radiation is always there. Hence, the $\hbar \omega/2$ drops out from the difference, and cannot be detected - at least in this simple way. This is the reason why we had the right to ignore this contribution in Eq. (88) – very fortunately, because it would lead to the integral’s divergence at its upper limit. However, let me repeat again that the ground-state energy of the electromagnetic field oscillators is physically real – and important.

---

53 Note that the heat capacity $C_V = (\partial E/\partial T)_V$, following from Eq. (88), is proportional to $T^3$ at any temperature, and hence does not obey the trend $C_V \rightarrow \text{const}$ at $T \rightarrow \infty$. This is the result of the unlimited growth, with temperature, of the number of thermally-excited field oscillators with $\hbar \omega < T$.

54 See, e.g., MA Eqs. (6.8b), (6.6b), and (2.7b).
One more interesting result may be deduced from the free energy $F$ of the electromagnetic radiation, which may be also calculated by integration of Eq. (73) over all the modes, with the appropriate weight:

$$F = \sum_\omega T \ln \left(1 - e^{-h\omega/T}\right) \to \int_0^\infty T \ln \left(1 - e^{-h\omega/T}\right) dN \, d\omega = \int_0^\infty T \ln \left(1 - e^{-h\omega/T}\right) \left(V \frac{\omega^2}{\pi^2 c^3}\right) d\omega. \quad (2.90)$$

Presenting $\omega^2 d\omega$ as $d(\omega^3)/3$, this integral may be readily worked out by parts, and reduced to a table integral similar to that in Eq. (88), yielding a surprisingly simple result:

$$F = -V \frac{\pi^2}{45h^3 c^3} T^4 = -\frac{E}{3}. \quad (2.91)$$

Now we can use the second of general thermodynamic equations (1.35) to calculate pressure:

$$P = \left. \frac{\partial F}{\partial V} \right|_T = \frac{\pi^2}{45h^3 c^3} T^4 = \frac{E}{3V}. \quad (2.92a)$$

This result might be, of course, derived by the integration of the expression for the forces exerted by each mode of the electromagnetic on confining the walls confining it to volume $V$;\(^{55}\) but notice how much simpler the thermodynamic calculation is. Rewritten in the form,

$$PV = \frac{E}{3}, \quad (2.92b)$$

this result may be considered as the equation of state of the electromagnetic field, i.e. from the quantum-mechanical point of view, the photon gas. As we will prove in the next chapter, the equation of state (1.44) of the ideal classical gas may be presented in a similar form, but with a coefficient generally different from Eq. (92). In particular, according to the equipartition theorem, for an ideal gas of nonrelativistic atoms whose internal degrees of freedom are in their ground state, whose whole energy is that of three translational “half-degrees of freedom”, $E = 3N(T/2)$, the factor before $E$ is twice larger than in Eq. (92). On the other hand, a relativistic treatment of the classical gas shows that Eq. (92) is valid for any gas in the ultrarelativistic limit, $T \gg mc^2$, where $m$ is the rest mass of the gas particle. Evidently, photons (i.e. particles with $m = 0$) satisfy this condition.\(^{56}\)

Finally, let me note that Eq. (92) allows the following interesting interpretation. The last of Eqs. (1.60), being applied to Eq. (92), shows that in this particular case the grand potential $\Omega$ equals $(-E/3)$. But according to the definition of $\Omega$, the first of Eqs. (1.60), this means that the chemical potential of the electromagnetic field excitations vanishes:

$$\mu = \frac{F - \Omega}{N} = 0. \quad (2.93)$$

In Sec. 8 below, we will see that the same result follows from Eq. (72) and the Bose-Einstein distribution, and discuss its physical sense.

\(^{55}\) See, e.g., EM Sec. 9.8.

\(^{56}\) Please note that according to Eqs. (1.44), (88), and (92), the difference between the equations of state of the photon gas and an ideal gas of nonrelativistic particles, expressed in the more usual form - as $P = P(V, T)$, is much more dramatic: $P \propto T^{4}V^{0}$ instead of $P \propto T^{3}V^{-1}$. 
(ii) Specific heat of solids. The heat capacity of solids is readily measurable, and in the early 1900s its experimentally observed temperature dependence served as an important test for emerging quantum theories. However, theoretical calculation of $C_V$ is not simple, even for isolators whose specific heat is due to thermally-induced vibrations of their crystal lattice alone. Indeed, a solid may be treated as an elastic continuum only at low relatively frequencies. Such continuum supports three different modes of mechanical waves with the same frequency $\omega$, that obey similar, linear dispersion laws, $\omega = vk$, but velocity $v = v_l$ for one of these modes (the longitudinal sound) is higher than that ($v_t$) of two other modes (the transverse sound). At such frequencies the wave mode density may be described by an evident modification of Eq. (83):

$$dN = V \frac{1}{(2\pi)^3} \left( \frac{1}{v_l^3} + \frac{2}{v_t^3} \right) 4\pi \omega^2 d\omega . \quad (2.94a)$$

For what follows, it is convenient to rewrite this relation in a form similar to Eq. (83):

$$dN = \frac{3V}{(2\pi)^3} 4\pi \frac{\omega^2 d\omega}{v^3}, \quad \text{with } v \equiv \left[ \frac{1}{3} \left( \frac{1}{v_l^3} + \frac{2}{v_t^3} \right) \right]^{-1/3} . \quad (2.94b)$$

However, wave theory shows that as frequency $\omega$ of a sound wave in a periodic structure is increased so that its half-wavelength $\pi/k$ approaches the crystal period $d$, the dispersion law $\omega(k)$ becomes nonlinear before the frequency reaches a maximum at $k = \pi/d$. To make the things even more complex, 3D crystals are generally anisotropic, so that the dispersion law is different in different directions of wave propagation. As a result, the exact statistics of thermally excited sound waves, and hence the heat capacity of crystals, is rather complex and specific for each particular crystal type.

In 1912, P. Debye suggested an approximate theory of the temperature dependence of the specific heat, which is in a surprisingly good agreement with experiment for many insulators, including polycrystalline and amorphous materials. In his model, the linear (acoustic) dispersion law $\omega = vk$, with the effective sound velocity $v$, defined by the latter of Eqs. (94b), is assumed to be exact all the way up to some cutoff frequency $\omega_D$, the same for all three wave modes. This cutoff frequency may be defined by the requirement that the total number of acoustic modes, calculated within this model from Eq. (94b),

$$N = V \frac{1}{(2\pi)^3} \frac{3}{v^3} \int_0^{\omega_D} 4\pi \omega^2 d\omega = \frac{V\omega_D^3}{2\pi^2 v^3} , \quad (2.95)$$

is equal to the universal number $N = 3nV$ of degrees of freedom (and hence of independent oscillation modes) in a system of $nV$ elastically coupled particles, where $n$ is the atomic density of the crystal, i.e. the number of atoms per unit volume. Within this model, Eq. (72) immediately yields the following expression for the average energy and specific heat (in thermal equilibrium at temperature $T$):

---

57 Due to low temperature expansion of solids, the difference between their $C_V$ and $C_P$ is small.
58 In good conductors (e.g., metals), specific heat is contributed (and at low temperatures, dominated) by free electrons – see Sec. 3.3 below.
59 See, e.g., CM Sec. 7.7.
60 See, e.g., CM Sec. 5.3, in particular Fig. 5.5 and its discussion.
\[ E = V \frac{1}{(2\pi)^3} \int_0^{\omega_D} \frac{\hbar \omega}{e^{\hbar \omega / T} - 1} 4\pi \omega^2 d\omega = 3nVT \cdot D(x)_{x=T_D / T}, \quad (2.96) \]

where \( T_D \equiv \hbar \omega_D \) is called the Debye temperature,\(^{61}\) and

\[ c_v \equiv \frac{C_v}{nV} = \frac{1}{nV} \left( \frac{\partial E}{\partial T} \right)_V = 3 \left[ D(x) - x \frac{dD(x)}{dx} \right]_{x=T_D / T}, \quad (2.97) \]

the Debye function. Red lines in Fig. 11 show the temperature dependence of the specific heat \( c_V \) (per atom) within the Debye model. At high temperatures, it approaches a constant value of 3, corresponding to energy \( E = 3nVT \), in accordance with the equipartition theorem for each of 3 degrees of freedom of each atom. (This model-insensitive value of \( c_V \) is known as the Dulong-Petit law.) In the opposite limit of low temperatures, the specific heat is much smaller:

\[ c_V \approx \frac{12\pi^4}{5} \left( \frac{T}{T_D} \right)^3 \ll 1, \quad (2.99) \]

reflecting the reduction of the number of excited waves with \( \hbar \omega < T \) as the temperature is decreased.

As a historic curiosity, P. Debye’s work followed one by A. Einstein, who had suggested (in 1907) a simpler model of crystal vibrations. In this model, all \( 3nV \) independent oscillatory modes of \( nV \) atoms of the crystal have approximately the same frequency, say \( \omega_0 \), and Eq. (72) immediately yields

---

\(^{61}\) In SI units, Debye temperatures \( T_D \) are of the order of a few hundred K for most simple solids (e.g., close to 430 K for aluminum and 340 K for copper), with somewhat lower values for crystals with heavy atoms (~105 K for lead), and reach the highest value ~2200 K for diamond with its relatively light atoms and very stiff lattice.
\[ E = 3nV \frac{\hbar \omega_k}{e^{\hbar \omega_k/kT} - 1}, \quad (2.100) \]

so that the specific heat is functionally similar to Eq. (75):

\[ c_V \equiv \frac{1}{nV} \frac{\partial E}{\partial T} = 3 \left[ \frac{\hbar \omega_k}{2T} \right]^2 \frac{1}{\sinh(\hbar \omega_k / 2T)}. \quad (2.101) \]

This dependence \( c_V(T) \) is shown with blue lines in Fig. 11 (assuming, for the sake of simplicity, \( \hbar \omega_k = T_D \)). At high temperatures, this result does satisfy the universal Dulong-Petit law (\( c_V = 3 \)), but at low temperatures the Einstein’s model predicts a much faster (exponential) drop of the specific heat as the temperature is reduced. (The difference between the Debye and Einstein models is not too spectacular on the linear scale, but in the log-log plot, shown on the right panel of Fig. 11, it is rather dramatic.\(^{62}\)) The Debye model is in a much better agreement with experimental data for simple, monoatomic crystals, thus confirming the conceptual correctness of his wave-based approach.

Note, however, that when a genius such as A. Einstein makes an error, there is probably some deep and important reason behind it. Indeed, crystals with the basic cell consisting of atoms of two or more types (such as NaCl, etc.), feature two or more separate branches of the dispersion law \( \omega(k) \) – see, e.g., Fig. 12.\(^{63}\)

![Fig. 2.12. Dispersion relation for longitudinal waves in a simple 1D model of a solid, with similar interparticle distances \( d \), but alternating particle masses, plotted for a particular mass ratio \( r = 5 \).](image)

While the lower “acoustic” branch is virtually similar to those for monoatomic crystals, and may be approximated by the Debye model, \( \omega = vk \), reasonably well, the upper (“optical”\(^ {64}\)) branch does approach \( \omega = 0 \) at any \( k \). Moreover, for large values of the atom mass ratio \( r \), the optical branches are almost flat, with virtually \( k \)-independent frequencies \( \omega_0 \) that correspond to simple oscillations of each light atom between its heavy counterparts. For thermal excitations of such oscillations, and their

---

\(^{62}\) This is why there is a general “rule of thumb” in science: if you plot your data on a linear rather than log scale, you better have a good excuse ready. (A valid excuse example: the variable you are plotting changes sign within the important range.)

\(^{63}\) This is the exact solution of a particular 1D model of such a crystal – see CM Chapter 5.

\(^{64}\) This term stems from the fact that at \( k \to 0 \), the mechanical waves corresponding to these branches have phase velocities \( v_{ph} \equiv \omega(k)/k \) that are much higher than that of the acoustic waves, and may approach the speed of light. As a result, these waves can strongly interact with electromagnetic (practically, optical) waves of the same frequency, while the acoustic waves cannot.
contribution to the specific heat, the Einstein model (with \( \omega_E = \omega_0 \)) gives a very good approximation, so that the specific heat may be well described by a sum of the Debye and Einstein laws (97) and (101), with appropriate weights.

2.7. Grand canonical ensemble and distribution

As we have seen, the Gibbs distribution is a very convenient way to calculate statistical and thermodynamic properties of systems with a fixed number \( N \) of particles. However, for systems in which \( N \) may vary, another distribution is preferable for some applications. Several examples of such situations (as well as the basic thermodynamics of such systems) have already been discussed in Sec. 1.5. Perhaps even more importantly, statistical distributions for systems with variable \( N \) are also applicable to the ensembles of independent particles on a certain single-particle energy level – see the next section.

With this motivation, let us consider what is called the grand canonical ensemble (Fig. 13). It is similar to the canonical ensemble discussed in the previous section (Fig. 6) in all aspects, besides that now the system under study and the heat bath (in this case typically called the environment) may exchange not only heat but also particles. In all system members of the ensemble, the environments are in both the thermal and chemical equilibrium, and their temperatures \( T \) and chemical potentials \( \mu \) are equal.

![Fig. 2.13. Member of a grand canonical ensemble.](image)

Now let us assume that the system of interest is also in the chemical and thermal equilibrium with its environment. Then using exactly the same arguments as in Sec. 4 (including the specification of a microcanonical sub-ensemble with fixed \( E_\Sigma \) and \( N_\Sigma \)), we may generalize Eq. (55), taking into account that entropy \( S_{\text{env}} \) of the environment is now a function of not only its energy \( E_{\text{env}} = E_\Sigma - E_{m,N} \), \(^{65}\) but also the number of particles \( N_E = N_\Sigma - N \), with \( E_\Sigma \) and \( N_\Sigma \) fixed:

\[
\ln W_{m,N} \propto \ln M = \ln g_{\text{env}} (E_\Sigma - E_{m,N}, N_\Sigma - N) + \ln \Delta E_\Sigma = S_{\text{env}} (E_\Sigma - E_{m,N}, N_\Sigma - N) + \text{const}
\]

\[
\approx S_{\text{env}} \left| \frac{\partial S_{\text{env}}}{\partial E_{\text{env}}} \right|_{E_\Sigma, N_\Sigma} \left| \frac{\partial S_{\text{env}}}{\partial N_{\text{env}}} \right|_{E_\Sigma, N_\Sigma} N + \text{const}.
\]

In order to simplify this relation, we may rewrite Eq. (1.52) in the equivalent form

\[^{65}\] The additional index in the new notation \( E_{m,N} \) for the energy of the system of interest reflects the fact that its eigenvalue spectrum is generally dependent on the number \( N \) of particles in it.
\[ dS = \frac{1}{T} dE + \frac{P}{T} dV - \frac{\mu}{T} dN. \]  
(2.103)

Hence, if entropy \( S \) of a system is expressed as a function of \( E, V, \) and \( N \), then

\[
\left( \frac{\partial S}{\partial E} \right)_{V,N} = \frac{1}{T}, \quad \left( \frac{\partial S}{\partial V} \right)_{E,N} = \frac{P}{T}, \quad \left( \frac{\partial S}{\partial N} \right)_{E,V} = -\frac{\mu}{T}.
\]  
(2.104)

Applying the first one and the last one of these relations to Eq. (102), and using the equality, of the temperatures \( T \) and chemical potentials \( \mu \) of the system under study and its environment, at their equilibrium, discussed in Sec. 1.5, we get

\[
\ln W_{m,N} = S_{\text{env}} (E_\Sigma, N_\Sigma) - \frac{1}{T} E_{m,N} + \frac{\mu}{T} N + \text{const}.
\]  
(2.105)

Again, exactly as at the derivation of the Gibbs distribution in Sec. 4, we may argue that since \( E_{m,N}, T \) and \( \mu \) do not depend on the choice of environment’s size, i.e. on \( E_\Sigma \) and \( N_\Sigma \), the probability \( W_{m,N} \) for a system to have \( N \) particles and be in \( m \)-th quantum state in the whole grand canonical ensemble should also obey a relation similar to Eq. (105). As a result, we get the so-called grand canonical distribution:

\[
W_{m,N} = \frac{1}{Z_G} \exp\left\{ \frac{\mu N - E_{m,N}}{T} \right\}.
\]  
(2.106)

Just as in the case of the Gibbs distribution, constant \( Z_G \) (most often called the grand statistical sum, but sometimes the “grand partition function”) should be determined from the probability normalization condition, now with the summation of probabilities \( W_{m,N} \) over all possible values of both \( m \) and \( N \):

\[
Z_G = \sum_{m,N} \exp\left\{ \frac{\mu N - E_{m,N}}{T} \right\}.
\]  
(2.107)

Now, using the general Eq. (29) to calculate entropy for distribution (106) (exactly like we did it for the canonical ensemble), we get the following expression,

\[
S = -\sum_{m,N} W_{m,N} \ln W_{m,N} = \ln Z_G + \frac{E}{T} - \frac{\mu \langle N \rangle}{T},
\]  
(2.108)

which is evidently a generalization of Eq. (62).\(^{66}\) We see that now the grand thermodynamic potential \( \Omega \) (rather than the free energy \( F \)) may be expressed directly via the normalization coefficient \( Z_G \):

\[
\Omega \equiv F - \mu \langle N \rangle = E - TS - \mu \langle N \rangle = T \ln \frac{1}{Z_G} = -T \ln \sum_{m,N} \exp\left\{ \frac{\mu N - E_{m,N}}{T} \right\}.
\]  
(2.109)

Finally, solving the last equality for \( Z_G \), and plugging the result back into Eq. (106), we can rewrite the grand canonical distribution in the form

\(^{66}\) The average number of particles \( \langle N \rangle \) is of course exactly what was called \( N \) in thermodynamics (see Ch. 1), but I need to keep this explicit notation here to make a clear distinction between this average value of the variable, and its particular values participating in Eqs. (102)-(110).
similar to Eq. (65) for the Gibbs distribution. Indeed, in the particular case when the number $N$ of particles is fixed, $N = \langle N \rangle$, so that $\Omega + \mu N = \Omega + \mu \langle N \rangle \equiv F$, Eq. (110) is reduced right to Eq. (65).

2.8. Systems of independent particles

Now we will use the general statistical distributions discussed above to a simple but very important case when each system we are considering consists of many similar particles whose explicit (physical) interaction is negligible. As a result, each particular energy value $E_{m,N}$ of such a system may be presented as a sum of energies $\varepsilon_k$ of its particles, where index $k$ numbers single-particle energy levels (rather than of the whole system, as index $m$ does).

Let us start with the classical limit. In classical mechanics, the quantization effects are negligible, i.e. there is a virtually infinite number of states $k$ within each finite energy interval. However, it is convenient to keep, for the time being, the discrete-state language, with understanding that the average number $\langle N_k \rangle$ of particles in each of these states, frequently called the state occupancy, is very small. In this case, we may apply the Gibbs distribution to the canonical ensemble of single particles, and hence use it with the substitution $E_{m,N} \rightarrow \varepsilon_k$, so that Eq. (58) becomes

$$
\langle N_k \rangle = c \exp\left(-\frac{\varepsilon_k}{T}\right) \ll 1,
$$

(2.111)

where constant $c$ should be found from the normalization condition:

$$
\sum_k \langle N_k \rangle = 1.
$$

(2.112)

This is the famous Boltzmann distribution.67 Despite its superficial similarity to the Gibbs distribution (58), let me emphasize the conceptual difference between these two results. The Gibbs distribution describes the probability to find the whole system on energy level $E_m$, and it is always valid - more exactly, for a canonical ensemble of systems in thermodynamic equilibrium. On the other hand, the Boltzmann distribution describes occupancy of an energy level of a single particle, and for systems of identical particles is valid only in the classical limit $\langle N_k \rangle \ll 1$, even if the particles do not interact directly.

The last fact may be surprising, because it may seem that as soon as particles of the system are independent, nothing prevents us from using the Gibbs distribution to derive Eq. (111), regardless of the value of $\langle N_k \rangle$. This is indeed true if the particles are distinguishable, i.e. may be distinguished from each other - say by their fixed spatial positions, or by the states of certain internal degrees of freedom (say, spin), or any other “pencil mark”. However, it is an experimental fact that elementary particles of each particular type (say, electrons) are identical to each other, i.e. cannot be “pencil-marked”. For such particles we have to be more careful: even if they do not interact explicitly, there is still some implicit

67 The distribution was first suggested in 1877 by the founding father of statistical physic, L. Boltzmann. For the particular case when $\varepsilon$ is the kinetic energy of a free classical particle (and hence has a continuous spectrum), it is reduced to the Maxwell distribution – see Sec. 3.1 below.
dependence in their behavior, which is especially evident for the so-called fermions (fundamental particles with semi-integer spin) they obey the Pauli exclusion principle that forbids two identical particles to be in the same quantum state, even if they do not interact explicitly.\textsuperscript{68}

Note that here the term “the same quantum state” carries a heavy meaning load here. For example, if two particles are confined to stay in different spatial positions (say, reliably locked in different boxes), they are distinguishable even if they are internally identical. Thus the Pauli principle, as well as other identity effects such as Bose-Einstein condensation, to be discussed in the next chapter, are important only when identical particles may move in the same spatial region. In order to describe this case, instead of “identical”, it is much better to use a more precise (though ugly) term indistinguishable particles.\textsuperscript{69}

In order to take these effects into account, let us examine the effects of nonvanishing occupancy $\langle N_k \rangle \sim 1$ on statistical properties of a system of many non-interacting but indistinguishable particles (at the first stage of calculation, either fermions or bosons) in equilibrium, and apply the grand canonical distribution (109) to a very interesting particular grand canonical ensemble: a subset of particles in the same quantum state $k$ (Fig. 14).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{single-particle-energy-levels.png}
\caption{Grand canonical ensemble of particles in the same quantum state (with eigenenergy $\varepsilon_k$).}
\end{figure}

In this ensemble, the role of the environment is played by the particles in all other states $k' \neq k$, because due to infinitesimal interactions, the particles may change their states. In equilibrium, the chemical potential $\mu$ and temperature $T$ of the system should not depend on the state number $k$, but the grand thermodynamic potential $\Omega_k$ of the chosen particle subset may. Replacing $N$ with $N_k$ - the particular (not average!) number of particles in $k^{th}$ state, and the particular energy value $E_{m,N}$ with $\varepsilon_k N_k$, we may reduce Eq. (109) to

\begin{equation}
\Omega_k = -T \ln \left( \sum_k \exp \left( \frac{\mu N_k - \varepsilon_k N_k}{T} \right) \right) = -T \ln \left[ \sum_k \exp \left( \frac{\mu}{T} \right)^{N_k} \right]
\end{equation}

\textsuperscript{68} See, e.g., QM Sec. 8.1.
\textsuperscript{69} This invites a natural question: what particles are “elementary enough” for the identity? For example, protons and neutrons have an internal structure, in some sense consisting of quarks and gluons; they be considered elementary? Next, if protons and neutrons are elementary, are atoms? molecules? What about really large molecules (such as proteins)? viruses? The general answer to these questions, given by quantum mechanics (or rather experiment :-), is that any particles/systems, no matter how large and complex they are, are identical if they have exactly the same internal structure, and also are exactly in the same internal quantum state – for example, in the ground state of all their internal degrees of freedom.
where the summation should be carried out over all possible values of $N_k$. For the final calculation of this sum, the elementary particle type becomes essential.

In particular, for fermions, obeying the Pauli principle, numbers $N_k$ in Eq. (113) may take only two values, either 0 (state $k$ is unoccupied) or 1 (the state is occupied), and the summation gives

$$\Omega_k = -T \ln \sum_{N_k=0,1} \left( \exp \left( \frac{\mu - \varepsilon_k}{T} \right) \right)^{N_k} = -T \ln \left( 1 + \exp \left( \frac{\mu - \varepsilon_k}{T} \right) \right).$$

(2.114)

Now the average occupancy may be calculated from the last of Eqs. (1.62) – in this case, with $N$ replaced with $\langle N_k \rangle$:

$$\langle N_k \rangle = -\left( \frac{\partial \Omega_k}{\partial \mu} \right)_{T, \varepsilon} = \frac{1}{e^{(\mu - \varepsilon_k)/T} - 1}.$$  

(2.115)

This is the famous Fermi-Dirac distribution, derived in 1926 independently by E. Fermi and P. Dirac.

On the other hand, bosons do not obey the Pauli principle, and for them numbers $N_k$ can take any non-negative integer values. In this case, Eq. (113) turns into the following equality:

$$\Omega_k = -T \ln \sum_{N_k=0}^{\infty} \left( \exp \left( \frac{\mu - \varepsilon_k}{T} \right) \right)^{N_k} = -T \ln \sum_{N_k=0}^{\infty} \lambda^{N_k}, \quad \text{with} \quad \lambda \equiv \exp \left( \frac{\mu - \varepsilon_k}{T} \right).$$

(2.116)

This sum is just the usual geometric progression again, which converges if $\lambda < 1$, giving

$$\Omega_k = -T \ln \frac{1}{1 - \lambda} = -T \ln \frac{1}{1 - e^{(\mu - \varepsilon_k)/T}}, \quad \text{for} \quad \mu < \varepsilon_k.$$  

(2.117)

In this case the average occupancy, again calculated using Eq. (1.62) with $N$ replaced with $\langle N_k \rangle$, obeys the Bose-Einstein distribution,

$$\langle N_k \rangle = -\left( \frac{\partial \Omega_k}{\partial \mu} \right)_{T, \varepsilon} = \frac{1}{e^{(\mu - \varepsilon_k)/T} - 1}, \quad \text{for} \quad \mu < \varepsilon_k,$$

(2.118)

which was derived in 1924 by S. Bose (for the particular case $\mu = 0$) and generalized in 1925 by A. Einstein for an arbitrary chemical potential. In particular, comparing Eq. (118) with Eq. (72), we see that harmonic oscillator excitations, each with energy $\hbar \omega$, may be considered as bosons, with zero chemical potential. We have already obtained this result ($\mu = 0$) in a different way – see Eq. (93). Its physical interpretation is that the oscillator excitations may be created inside the system, so that there is no energy cost $\mu$ of moving them into the system from its environment.

The simple form of Eqs. (115) and (118), and their similarity (besides “only” the difference of the signs before unity in their denominators), is one of most beautiful results of physics. This similarity should not disguise the facts that the energy dependences of $\langle N_k \rangle$, given by these two formulas, are rather different – see Fig. 15. In the Fermi-Dirac statistics, the average level occupancy is finite (and

---

70 As the reader certainly knows, for the electromagnetic field oscillators, such excitations are called photons; for mechanical oscillation modes, phonons. It is important, however, not to confuse these mode excitations with the oscillators as such, and be very careful in prescribing to them certain spatial locations – see, e.g., QM Sec. 9.1.
below 1) at any energy, while in the Bose-Einstein it may be above 1, and even diverges at $\varepsilon_k \to \mu$. However, for any of these distributions, as temperature is increased, it eventually becomes much larger than the difference ($\varepsilon_k - \mu$) for all $k$. In this limit, $\langle N_k \rangle \ll 1$, both distributions coincide with each other, as well as with the Boltzmann distribution (111) with $c = \exp\{\mu/T\}$. The last distribution, therefore, serves as the high-temperature limit for quantum particles of both sorts.

A natural question now is how to find the chemical potential $\mu$ participating in Eqs. (115) and (118). In the grand canonical ensemble as such (Fig. 13), with number of particles variable, the value of $\mu$ is imposed by system’s environment. However, both the Fermi-Dirac and Bose-Einstein distributions are also applicable to equilibrium systems with a fixed but large number $N$ of particles. In these conditions, the role of the environment for some subset of $N' \ll N$ particles is played by the remaining $N - N'$ particles. In this case, $\mu$ may be found by calculation of $\langle N \rangle$ from the corresponding distribution, and then requiring it to be equal to the genuine number of particles in the system. In the next section, we will perform such calculations for several particular systems.

For those and other applications, it will be convenient for us to have ready expressions for entropy $S$ of a general (i.e. not necessarily equilibrium) state of systems of independent Fermi or Bose particles, expressed not as a function of $W_m$ of the whole system – as Eq. (29) does, but as a function of the average occupancy numbers $\langle N_k \rangle$. For that, let us consider a composite system, each consisting of $M >> 1$ similar but distinct component systems, numbered by index $m = 1, 2, \ldots M$, with independent (i.e. not explicitly interacting) particles. We will assume that though in each of $M$ component systems, the number $N_k^{(m)}$ of particles in its $k$-th quantum state may be different (Fig. 16), but their total number $N_k^{(2)}$ in the composite system is fixed:

$$\sum_{m=1}^{M} N_k^{(m)} = N_k^{(2)}. \quad (2.119)$$

![Fig. 2.15. Fermi-Dirac (blue line) and Bose-Einstein (red line) distributions, and the Boltzmann distribution with $c = \exp\{\mu/T\}$ (black line).](image)

![Fig. 2.16. Composite system with a certain distribution of $N_k^{(2)}$ particles in $k$-th state between $M$ component systems.](image)
As a result, the total energy of the composite system is fixed as well,
\[ \sum_{m=1}^{M} N^{(m)}_k \varepsilon_k = N^{(\Sigma)}_k \varepsilon_k = \text{const}, \quad (2.120) \]
so that an ensemble of many such composite systems (with the same \( k \)), in equilibrium, is microcanonical. According to Eq. (24a), the average entropy \( S_k \) per component system may be calculated as
\[ S_k = \lim_{M,N_k \to \infty} \frac{\ln M_k}{M}, \quad (2.121) \]
where \( M_k \) is the number of possible different ways such composite system (with fixed \( N^{(\Sigma)}_k \)) may be implemented.

Let us start the calculation of \( M_k \) with Fermi particles - for which the Pauli principle is valid. Here the level occupancies \( N^{(m)}_k \) may be only equal 0 or 1, so that the distribution problem is solvable only if \( N^{(\Sigma)}_k \leq M \), and evidently equivalent to the choice of \( N^{(\Sigma)}_k \) balls (in arbitrary order) from the total number of \( M \) distinct balls. Comparing this formulation with the binomial coefficient definition,\(^{71}\) we immediately have
\[ M_k = \frac{M!}{\left( M - N^{(\Sigma)}_k \right)!N^{(\Sigma)}_k!}, \quad (2.122) \]
From here, using the Stirling formula (again, in its simplest form (27)), we get
\[ S_k = - \langle N_k \rangle \ln \langle N_k \rangle - \left( 1 - \langle N_k \rangle \right) \ln \left( 1 - \langle N_k \rangle \right), \quad (2.123) \]
where
\[ \langle N_k \rangle \equiv \lim_{M,N_k \to \infty} \frac{N^{(\Sigma)}_k}{M}, \quad (2.124) \]
is exactly the average occupancy of the \( k \)-th single-particle level in each system that was discussed earlier in this section. Since for a Fermi system, \( \langle N_k \rangle \) is always somewhere between 0 and 1, so that entropy (123) is always positive.

In the Bose case, where the Pauli limitation is not valid, the number \( N^{(m)}_k \) of particles on the \( k \)-th level in each of the systems is an arbitrary (positive) integer. Let us consider \( N^{(\Sigma)}_k \) particles and \((M - 1)\) partitions (shown by vertical lines in Fig. 16) between \( M \) systems as \((M - 1 + N^{(\Sigma)}_k)\) similar mathematical objects ordered along one axis. Then \( M_k \) may be calculated as the number of possible ways to distribute the \((M - 1)\) indistinguishable partitions among these \((M - 1 + N^{(\Sigma)}_k)\) distinct objects, i.e. as the following binomial coefficient:\(^{72}\)
\[ M_k = \frac{M + N^{(\Sigma)}_k - 1}{M - 1} C_{M-1} = \frac{(M - 1 + N^{(\Sigma)}_k)!}{(M - 1)!N^{(\Sigma)}_k!}, \quad (2.125) \]

\(^{71}\) See, e.g., MA Eq. (2.2).
\(^{72}\) See also MA Eq. (2.4).
Applying the Stirling formula (27) again, we get the following result,

\[ S_k = -\langle N_k \rangle \ln \langle N_k \rangle + (1 + \langle N_k \rangle) \ln (1 + \langle N_k \rangle), \]  

(2.126)

which again differs from the Fermi case (123) “only” by the signs in the second term, and is valid for any positive \( \langle N_k \rangle \).

Expressions (123) and (126) are valid for an arbitrary (possibly non-equilibrium) case; they may be also used for an alternative derivation of the Fermi-Dirac (115) and Bose-Einstein (118) distributions valid in equilibrium. For that, we may use the method of Lagrange multipliers, requiring (just like it was done in Sec. 2) the total entropy of a system of \( N \) independent, similar particles,

\[ S = \sum_k S_k, \]  

(2.127)

as a function of state occupancies \( \langle N_k \rangle \), to attain its maximum, with the conditions of fixed total number of particles \( N \) and the total energy \( E \):

\[ \sum_k \langle N_k \rangle = N = \text{const}, \quad \sum_k \langle N_k \rangle \epsilon_k = E = \text{const}. \]  

(2.128)

The completion of this calculation is left for reader’s exercise.

In the classical limit, when the average occupancies \( \langle N_k \rangle \) of all states are small, both the Fermi and Bose expressions for \( S_k \) tend to the same limit

\[ S_k = -\langle N_k \rangle \ln \langle N_k \rangle, \quad \text{for } \langle N_k \rangle \ll 1. \]  

(2.129)

This expression, frequently referred to as the Boltzmann (or “classical”) entropy, might be also obtained, for arbitrary \( \langle N_k \rangle \), directly from Eq. (29) by considering an ensemble of systems, each consisting of just one classical particle, so that \( \epsilon_m \to \epsilon_k \) and \( W_m \to \langle N_k \rangle \). Let me emphasize again that for indistinguishable particles, such identification is generally (i.e. at \( \langle N_k \rangle \sim 1 \)) illegitimate even if they do not interact explicitly. As we will see in the next chapter, the indistinguishability affects statistical properties of even classical particles.

2.9. Exercise problems

2.1. A famous example of the macroscopic irreversibility was suggested in 1907 by P. Ehrenfest. Two dogs share \( 2N >> 1 \) fleas. Each flea may jump to another dog, and the rate (i.e. the probability of jumping per unit time) \( \Gamma \) of such an event does not depend on time, and on the location of other fleas. Find the time evolution of the average number of fleas on a dog, and of the flea-related part of dogs’ entropy (at arbitrary initial conditions), and prove that the entropy can only grow.\(^{73}\)

2.2. Use the microcanonical distribution to calculate thermodynamic properties (including entropy, all relevant thermodynamic potentials, and heat capacity), of an ensemble of similar two-level

\(^{73}\) This is essentially a simpler (and funnier :-) version of the particle scattering model used by L. Boltzmann to prove his famed \( H \)-theorem (1872). Besides all the historic significance of that theorem, the model used by Boltzmann (see Sec. 6.2 below) is almost as cartoonish.
systems, in thermodynamic equilibrium at temperature $T$ that is comparable with the energy gap $\Delta$. For each variable, sketch its temperature dependence, and find its asymptotic values (or trends) in the low-temperature and high-temperature limits.

*Hint:* The two-level system is generally defined as any system with just two relevant states whose energies, say $E_0$ and $E_1$, are separated by a finite gap $\Delta \equiv E_1 - E_0$. Its most popular (but not the only!) example is a spin-$\frac{1}{2}$ particle, e.g., an electron, in an external magnetic field.

2.3. Solve the previous problem using the Gibbs distribution. Also, calculate the probabilities of the energy level occupation, and give physical interpretations of your results, in both temperature limits.

2.4. Calculate the low-field magnetic susceptibility $\chi_m$ of a dilute set of non-interacting, spontaneous magnetic dipoles $\mathbf{m}$, in thermal equilibrium at temperature $T$, within two models:

(i) the dipole moment $\mathbf{m}$ is a classical vector of fixed magnitude $m_0$, but arbitrary orientation, and (ii) the dipole moment $\mathbf{m}$ belongs to a quantum spin-$\frac{1}{2}$ particle, and is described by vector operator $\hat{\mathbf{m}} = \gamma \hat{\mathbf{S}}$, where $\gamma$ is the gyromagnetic ratio, and $\hat{\mathbf{S}}$ is the vector operator of particle’s spin.74

*Hint:* The low-field magnetic susceptibility of an isotropic medium is defined75 as

$$\chi_m = \frac{\partial M_z}{\partial H},$$

where $M$ is the (average) magnetization of a unit volume, and axis $z$ is aligned with the direction of the external magnetic field $H$.

2.5. Calculate the low-field magnetic susceptibility of a set of non-interacting, distinguishable particles with an arbitrary spin $s$, neglecting their orbital motion. Compare the result with the solution of the previous problem.

*Hint:* Quantum mechanics76 tells us that the Cartesian component $m_z$ of the magnetic moment of such a particle, in the direction of the applied field, may take $(2s + 1)$ values

$$m_z = \gamma s_m,$$

where $s_m = -s, -s + 1, \ldots, s - 1, s$,

where $\gamma$ is the gyromagnetic ratio of the particle, and $\hbar$ is the Planck’s constant.

2.6.* Derive a general expression for the average interaction potential between two similar magnetic dipoles with fixed magnitude $m$ but arbitrary orientation, at thermal equilibrium. Spell out the result in the low-temperature and high-temperature limits.

2.7.* Analyze the possibility of using a system of non-interacting spin-$\frac{1}{2}$ particles in magnetic field for refrigeration.

*Hint:* See a footnote in Sec. 1.6.

---

74 See, e.g., QM Sec. 4.4. Note that both models assume that the particle’s orbital motion (if any) does not contribute to its magnetic moment.

75 See, e.g., EM Sec. 5.5, in particular Eq. (5.111).

76 See, e.g., QM Sec. 5.7, in particular Eq. (5.197).
2.8. Use the microcanonical distribution to calculate the average entropy, energy, and pressure of a single classical particle of mass $m$, with no internal degrees of freedom, free to move in volume $V$, at temperature $T$.

*Hint:* Try to make a more accurate calculation than has been done in Sec. 2.2 for the system of $N$ harmonic oscillators. For that you will need to know the volume $V_d$ of an $d$-dimensional hypersphere of the unit radius. To avoid being too cruel, I am giving it to you:

$$V_d = \pi^{d/2} / \Gamma\left(\frac{d}{2} + 1\right),$$

where $\Gamma(\xi)$ is the gamma-function.\(^{77}\)

2.9. Solve the previous problem starting from the Gibbs distribution.

2.10. Calculate the average energy, entropy, free energy, and the equation of state of a classical 2D particle (without internal degrees of freedom), free to move within area $A$, at temperature $T$, starting from:

(i) the microcanonical distribution, and
(ii) the Gibbs distribution.

*Hint:* Make the appropriate modification of the notion of pressure.

2.11. A quantum particle of mass $m$ is confined to free motion along a 1D segment of length $a$. Using any approach you like, find the average force the particle exerts on walls of such a “1D quantum well” in thermal equilibrium, and analyze its temperature dependence, focusing on the low-temperature and high-temperature limits.

*Hint:* You may consider series $\Theta(\xi) \equiv \sum_{n=1}^{\infty} e^{-\xi n^2}$ as a known function of $\xi$.\(^{78}\)

2.12. Rotational properties of diatomic molecules (such as $N_2$, CO, etc.) may be reasonably well described using a “dumbbell” model: two point particles, of masses $m_1$ and $m_2$, with a fixed distance $d$ between them. Ignoring the translational motion of the molecule as the whole, use this model to calculate its heat capacity, and spell out the result in the limits of low and high temperatures. (Quantify the conditions.)

2.13. Calculate the heat capacity of a diatomic molecule, using the simple model described in the previous problem, but now assuming that the rotation is confined to one plane.\(^{79}\)

---

\(^{77}\) For its definition and main properties, see, e.g., MA Eqs. (6.6)-(6.9).

\(^{78}\) It may be reduced to the so-called elliptic theta-function $\theta_3(z, \tau)$ for a particular case $z = 0$ - see, e.g., Sec. 16.27 in the Abramowitz-Stegun handbook cited in MA Sec. 16(ii). However, you do not need that (or any other :-) handbook to solve this problem.

\(^{79}\) This is a reasonable model of the constraints imposed on small atomic groups (e.g., ligands) by their environment inside some large molecules.
2.14. Low-temperature specific heat of some solids has a considerable contribution from thermal excitation of spin waves, whose dispersion law scales as $\omega \propto k^2$ at $\omega \to 0$.\(^{80}\) Find the temperature dependence of this contribution to $C_V$ at low temperatures and discuss conditions of its experimental observation.

2.15. A rudimentary “zipper” model of DNA replication is a chain of $N$ links that may be either open or closed - see Fig. on the right. Opening a link increases system’s energy by $\Delta > 0$, and a link may change its state (either open or close) only if all links to the left of it are already open. Calculate the average number of open links at thermal equilibrium, and analyze its temperature dependence in detail, especially for the case $N \gg 1$.

2.16. An ensemble of classical 1D particles of mass $m$, residing in the potential wells

$$U(x) = a|x|^\gamma, \quad \text{with } \gamma > 0,$$

is in thermal equilibrium at temperature $T$. Calculate the average values of its potential energy $U$ and the full energy $E$ using two approaches:

(i) directly from the Gibbs distribution, and

(ii) using the virial theorem of classical mechanics.\(^{81}\)

2.17. For a thermally-equilibrium ensemble of slightly anharmonic classical 1D oscillators, with mass $m$ and potential energy

$$U(q) = \frac{k}{2} x^2 + ax^3,$$

with small coefficient $a$, calculate $\langle x \rangle$ in the first approximation in low temperature $T$.

2.18. * A small conductor (in this context, usually called the single-electron box) is placed between two conducting electrodes, with voltage $V$ applied between them. The gap between one of the electrodes and the island is so narrow that electrons may tunnel quantum-mechanically through this gap (“weak tunnel junction”) – see Fig. on the right. Calculate the average charge of the island as a function of $V$.

* Hint: The quantum-mechanical tunneling of electrons through weak junctions\(^{82}\) between macroscopic conductors, and its subsequent energy relaxation inside the conductor, may be considered as a single inelastic (energy-dissipating) event, so that the only energy relevant for the thermal equilibrium of the system is its electrostatic potential energy.

---

\(^{80}\) Note that by the same dispersion law is typical for elastic bending waves in thin rods – see, e.g., CM Sec. 7.8.

\(^{81}\) See, e.g., CM Problem 1.12.

\(^{82}\) In this context, weak junction means a tunnel junction with transparency so low that the tunneling electron’s wavefunction looses its quantum-mechanical coherence before the electron has time to tunnel back. In a typical junction of a macroscopic area this condition is fulfilled if the effective tunnel resistance of the junction is much higher than the quantum unit of resistance (see, e.g., QM Sec. 3.2) , $R_Q \equiv \hbar/2e^2 \approx 6.5 \text{k}\Omega$. 
2.19. An LC circuit (see Fig. on the right) is at thermodynamic equilibrium with the environment. Find the r.m.s. fluctuation \( \delta V \equiv \langle V^2 \rangle^{1/2} \) of the voltage across it, for an arbitrary ratio \( T/\hbar \omega \), where \( \omega = (LC)^{-1/2} \) is the resonance frequency of this “tank circuit”.

2.20. Derive Eq. (92) from simplistic arguments, representing the blackbody radiation as an ideal gas of photons, treated as ultrarelativistic particles. What do similar arguments give for an ideal gas of classical, nonrelativistic particles?

2.21. Calculate the enthalpy, the entropy, and the Gibbs energy of the blackbody electromagnetic radiation with temperature \( T \), and then use these results to find the law of temperature and pressure drop at an adiabatic expansion of the radiation.

2.22. As was mentioned in Sec. 2.6(i) of the lecture notes, the relation between the visible temperatures \( T_\oplus \) of Sun’s surface and Earth’s surface \( T_o \) follows from the balance of the thermal radiation they emit. Prove that this relation indeed follows, with a good precision, from a simple model in which the surfaces radiate as perfect black bodies with a constant, average temperature.

\textit{Hint}: You may pick up the experimental values you need from any (reliable :-)) source.

2.23. If a surface is not perfectly radiation-absorbing (“black”), the electromagnetic power of its thermal radiation differs from the Stefan law (2.89a) by a frequency-dependent factor \( \varepsilon < 1 \), called emissivity:

\[
\frac{P}{A} = \varepsilon \sigma T^4.
\]

Prove that such surface reflects \((1 - \varepsilon)\) part of incident radiation.

2.24. If two black surfaces, facing each other, have different temperatures (see Fig. on the right), then according to the Stefan radiation law (2.89), there is a net flow of thermal radiation, from a warmer surface to the colder one:

\[
\frac{P_{\text{net}}}{A} = \sigma (T_1^4 - T_2^4).
\]

For many applications (including low temperature experiments) this flow is detrimental. One way to reduce it is to reduce the emissivity \( \varepsilon(\omega) \) of both surfaces – say by covering them with shiny metallic films. An alternative way toward the same goal is to place, between the surfaces, a thin layer (usually called the \textit{thermal shield}), with a low emissivity of both surfaces, and disconnected from any heat bath – see dashed line in Fig. above. Assuming that the emissivity is the same in both cases, find out which way is more efficient.

\textit{Hint}: The definition of emissivity may be found, for example, in the previous problem.

2.25. Two parallel, well conducting plates of area \( A \) are separated by a free-space gap of a constant thickness \( t \ll A^{1/2} \). Calculate the energy of the spontaneous electromagnetic field inside the gap
at thermal equilibrium with temperature $T$. Specify the validity limits of your result. Does the radiation push the plates apart?

2.26. Use the Debye theory to estimate the specific heat of aluminum at room temperature (say, 300 K), and express the result in the following popular units:

(i) eV/K per atom,
(ii) J/K per mole, and
(iii) J/K per gram.

Compare the last number with the experimental value (from a reliable book or online source).

2.27. Use the general Eq. (123) to re-derive the Fermi-Dirac distribution (115) for a system in equilibrium.

2.28. Each of two similar particles, not interacting directly, may take two quantum states, with single-particle energies $\varepsilon$ equal to 0 and $\Delta$. Write down the statistical sum $Z$ of the system, and use it to calculate its average total energy $E$ of the system, for the cases when the particles are:

(i) distinguishable;
(ii) indistinguishable fermions;
(iii) indistinguishable bosons.

Analyze and interpret the temperature dependence of $\langle E \rangle$ for each case, assuming that $\Delta > 0$.

2.29. Calculate the chemical potential of a system of $N >> 1$ independent fermions, kept at fixed temperature $T$, provided that each particle has two non-degenerate energy levels, separated by gap $\Delta$. 
This page is intentionally left blank