Part IIA Quantum Physics, Lent 2004

Dr. R.R. Horgan

December 22, 2011

Contents

1 Introduction

This course of sixteen lectures is concerned with explaining how the principles of quantum mechanics and quantum statistical physics underlie some important and applicable phenomena in modern physics. In particular we will study the laser and the bi-polar transistor. Over many years these devices have been developed in sophisticated and diverse ways to serve the many applications with which we are very familiar. However, our object is to avoid (or try to avoid) getting bogged down in the complicated details of the systems whose workings we are trying to understand. The underlying theory is what we are aiming at, and it will be enough to study the simplest and cleanest systems which exhibit the phenomena in which we are interested. The complete theory of such devices does need advanced techniques of theoretical physics but it is important to realise that the phenomena themselves can be explained conceptually and successfully using the basic ideas of quantum mechanics and statistical physics. Nevertheless, to then convert this understanding into the design of efficient working systems requires a lot of calculation and analysis of the details of real experimental systems.

This course will assume that students have taken the Part IB "Quantum Mechanics" course and knowledge of the Part II "Foundations of Quantum Mechanics" course will be useful at times, but all other necessary theory will be developed in the lectures. In particular, the ideas of quantum statistical physics will be presented for the first time. The new ideas presented in course are very general in nature and are of great interest in their own right: the concepts underpinning them are universal and of wide-ranging applicability. Nevertheless, the focus is on practical application which acts as the vehicle to help us understand the important mathematical and physical techniques we are using.

2 Review of Quantum Mechanics

Quantum mechanics is constructed in terms of **operators** acting as transformations on the **linear vector space** of **wavefunctions**

(i) Important **operators** are the linear momentum $\boldsymbol{p}$, the position vector $\boldsymbol{x}$ and the Hamiltonian H, where

$$H \;=\; \frac{1}{2m}\boldsymbol{p}^2 \;+\; V(\boldsymbol{x}) \;. \tag{2.1}$$

$V(\boldsymbol{x})$ is the **potential energy** function and H is the "energy" operator. In the Schrödinger representation we have:

$\boldsymbol{x}$ represented by multiplication by $\boldsymbol{x}$;
$\boldsymbol{p}$ represented by $-i\hbar\nabla$,

which imply the commutation relations:

$$[x_i, p_j] = i\hbar\delta_{ij} \;, \qquad [x_i, x_j] \;=\; [p_i, p_j] \;=\; 0 \;. \tag{2.2}$$

Then from eqn. (2.1) the Hamiltonian H is represented by

$$H \;=\; -\frac{\hbar^2}{2m}\nabla^2 \;+\; V(\boldsymbol{x}) \;. \tag{2.3}$$

H and $\boldsymbol{p}$ are differential operators in the Schrödinger representation and so act on sufficiently differentiable functions of $\boldsymbol{x}$. These functions are called **wavefunctions** and embody all of the physics of the system under study. These wavefunctions are functions of position, $\boldsymbol{x}$, and time, t. If we know the wavefunction, $\psi(\boldsymbol{x}, t)$, for all $\boldsymbol{x}$ and t we can predict the outcome of all experimental measurements on the system.

(ii) The dynamics of the system which allow us to calculate $\psi(\boldsymbol{x}, t)$ are given by Schrödinger's equation for the time evolution of $\psi(\boldsymbol{x}, t)$. This is the "equation of motion" of quantum mechanics. Schrödinger's equation is:

$$i\hbar \frac{\partial}{\partial t} \psi(\boldsymbol{x}, t) \;=\; H\psi(\boldsymbol{x}, t) \; . \tag{2.4}$$

(iii) **Energy eigenfunctions** $\psi_n(\boldsymbol{x}, t)$ satisfy

$$H\psi_n(\boldsymbol{x}, t) \;=\; E_n \psi_n(\boldsymbol{x}, t) \; , \tag{2.5}$$

where n labels and distinquishes the different eigenfunctions. Together with eqn. (2.4), this equation implies that

$$\psi_n(\boldsymbol{x}, t) \;=\; \phi_n(\boldsymbol{x})\, e^{-iE_n t/\hbar} \; , \tag{2.6}$$

with

$$H\phi_n(\boldsymbol{x}) \;=\; E_n \phi_n(\boldsymbol{x}) \; . \tag{2.7}$$

This last equation is Schrödinger's time-independent equation for the **energy eigenvalues**, E_n. The wavefunctions $\psi_n(\boldsymbol{x}, t)$ are particularly important and all called "stationary wavefunctions (or states)". Equation (2.7) is a differential equation which can be solved once the boundary conditions are given. Generally, the boundary conditions are such that only particular values for the E_n are allowed. The sequence $E_0, E_1 \ldots$ give all the allowed **observable** values for the energy of the system.

The $\psi_n(\boldsymbol{x}, t)$ form a **complete basis** for the linear space of wavefunctions. Hence, the most general solution to Schrödinger's time-dependent equation, eqn. (2.4) can be written using this basis as:

$$\psi(\boldsymbol{x}, t) \;=\; \sum_n a_n \, \psi_n(\boldsymbol{x}, t) \; , \tag{2.8}$$

which, using eqn. (2.6), becomes:

$$\psi(\boldsymbol{x}, t) \;=\; \sum_n a_n \, \phi_n(\boldsymbol{x})\, e^{-iE_n t/\hbar} \; . \tag{2.9}$$

This way of writing the general solution to eqn. (2.4) is a standard Maths Methods technique of solving all such equations subject to given boundary conditions: it is not peculiar to quantum mechanics.

(iv) The **inner product** (like a "dot product") of two wavefunctions ψ and χ is denoted by (χ, ψ) and is given by

$$(\chi, \psi) \;=\; \int d^3x \, \chi^*(\boldsymbol{x}, t)\, \psi(\boldsymbol{x}, t) \; . \tag{2.10}$$

In particular for two energy eigenfunctions ψ_m and ψ_n

$$(\psi_m, \psi_n) \;=\; \delta_{mn} \qquad \Longrightarrow \qquad (\phi_m, \phi_n) \;=\; \delta_{mn} \; . \tag{2.11}$$

I.e., the energy eigenfunctions are **orthonormal**.

The **matrix element** of an operator A is denoted by $(\chi, A\psi)$ and is given by

$$(\chi, A\psi) \;=\; \int d^3x\, \chi^*(\boldsymbol{x},t)\,(A\psi(\boldsymbol{x},t)) \;. \tag{2.12}$$

The hermitian conjugate $A^\dagger$ of A is defined by

$$(A^\dagger\chi, \psi) \;=\; (\chi, A\psi) \;. \tag{2.13}$$

(v) The solution to eqn. (2.4) is then constructed as follows:

  (a) Solve Schrödinger's time-independent equation (2.7) using the physical boundary conditions to find the $\{E_n\}$, and the corresponding $\phi_n(\boldsymbol{x})$.

  (b) Construct the $\psi_n(\boldsymbol{x},t)$ using eqn. (2.6) and express the general solution, $\psi(\boldsymbol{x},t)$, using the ψ_n as a basis as given in eqn. (2.8).

  (c) Since Schrödinger's time-independent equation is first-order in time it is an initial value problem and we only need to know the solution at $t=0$ to solve for all t. Suppose we are given $\psi(\boldsymbol{x},0) = f(\boldsymbol{x})$. Then from eqn (2.6) we see that

$$\psi(\boldsymbol{x},0) \;=\; f(\boldsymbol{x}) \;=\; \sum_n a_n\,\phi_n(\boldsymbol{x}) \;. \tag{2.14}$$

  Hence, using orthonormality, eqn. (2.11), we determine the coefficients $\{a_n\}$ by

$$(\phi_n, f) \;=\; a_n \;. \tag{2.15}$$

  The general solution is given by substituting the $\{a_n\}$, which are usually complex, back into eqn. (2.8).

(vi) Other operators have associated eigenvalues and eigenfunctions too. In general, we have

$$Ag_n(\boldsymbol{x}) \;=\; \lambda_n g_n(\boldsymbol{x}) \;. \tag{2.16}$$

If $A^\dagger = A$ then A is said to be hermitian and the eigenvalues λ_n are real. All observables have real eigenvalues and are thus represented by hermitian operators. Like the ϕ_n the g_n are orthonormal and form a complete basis for the space of wavefunctions:

$$\psi(\boldsymbol{x},t) \;=\; \sum_n c_n(t)\,g_n(\boldsymbol{x}) \;. \tag{2.17}$$

with

$$c_n(t) \;=\; (g_n, \psi) \;. \tag{2.18}$$

Note: in the special case $A = H$ we found

$$\psi(\boldsymbol{x},t) \;=\; \sum_n a_n\, e^{-iE_nt/\hbar}\,\phi_n(\boldsymbol{x}) \;,$$

which identifies in this case

$$c_n(t) \;=\; a_n\, e^{-iE_nt/\hbar} \qquad \text{and} \qquad g_n(\boldsymbol{x}) \;=\; \phi_n(\boldsymbol{x}) \;. \tag{2.19}$$

In general, however, the time-dependence of the $\{c_n(t)\}$ is not so simple: the energy operator is very special.

(vii) The interpretation of this expansion is explained by the question

"A system is described by the wavefunction $\psi(\boldsymbol{x},t)$. What is the probability $P_n^A(t)$ of measuring the value of the observable represented by the (hermitian) operator A to have value λ_n? "

The answer is

$$P_n^A(t) \;=\; |c_n(t)|^2 \;=\; |(g_n,\psi)|^2 \;. \tag{2.20}$$

In the special case $A = H$ we see from eqn. (2.19) that

$$P_n^H(t) \;=\; |a_n\, e^{-iE_nt/\hbar}|^2 \;=\; |a_n|^2 \;,$$

which establishes that P_n^H is independent of time. Clearly, we must have

$$\sum_n |c_n(t)|^2 \;=\; 1 \;, \tag{2.21}$$

which is ensured by imposing

$$\int d^3x\, |\psi(\boldsymbol{x},t)|^2 \;=\; 1 \;. \tag{2.22}$$

2.1 Example of two-state system

To illustrate the review above we consider a two-state system. Such a system has only two energy eigenfunctions ψ_1 and ψ_2 with energies E_1 and E_2 respectively: $H\phi_n = E_n\phi_n,\;\; n = 1, 2$. The general solution for the wavefunction is thus

$$\begin{aligned} \psi(\boldsymbol{x},t) \;&=\; a_1\, e^{-iE_1t/\hbar}\, \phi_1(\boldsymbol{x}) \;+\; a_2\, e^{-iE_2t/\hbar}\, \phi_2(\boldsymbol{x}) \;, \\ |a_1|^2 \;&+\; |a_2|^2 \;=\; 1 \end{aligned} \tag{2.1.1}$$

Let an observable, A, operate on the eigenfunctions $\phi_1,\; \phi_2$ as

$$A\,\phi_1 \;=\; \phi_2 \;, \qquad\qquad A\,\phi_2 \;=\; \phi_1 \;. \tag{2.1.2}$$

What are the eigenfunctions of A ? Let a generic eigenfunction be $g(\boldsymbol{x})$ with

$$A\,g(\boldsymbol{x}) \;=\; \lambda\, g(\boldsymbol{x}) \;. \tag{2.1.3}$$

Using that $\phi_1,\; \phi_2$ form a complete basis we must have that

$$\begin{aligned} g \;&=\; b_1\,\phi_1 \;+\; b_2\,\phi_2 \;, \\ |b_1|^2 \;&+\; |b_2|^2 \;=\; 1 \;, \end{aligned} \tag{2.1.4}$$

where b_1 and b_2 are determined as follows.

Multiply both sides of eqn. (2.1.4) by ϕ_1^* and integrate over all space. Repeat with ϕ_2^*. Using orthonormality of the ϕ_n we find

$$(\phi_1, g) \;=\; b_1 \;, \quad (\phi_2, g) \;=\; b_2 \;. \tag{2.1.5}$$

This is just taking the inner product of eqn. (2.1.4) with ϕ_1 and ϕ_2.

Now operate with A on both sides of eqn. (2.1.4) and use eqns. (2.1.2) and (2.1.3) to give

$$\begin{aligned} A\,g \;&=\; \lambda\, g \\ \Longrightarrow \qquad b_1A\phi_1 \;+\; b_2A\phi_2 \;&=\; \lambda b_1\,\phi_2 \;+\; \lambda b_2\,\phi_1 \;, \end{aligned} \tag{2.1.6}$$

Taking the inner product of this equation with the ϕ_1 and ϕ_2 we find, using eqn. (2.1.5),

$$\begin{array}{rcl} \lambda\, b_1 & = & b_2 \\ \lambda\, b_2 & = & b_1 \end{array} \qquad \Longrightarrow \qquad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \lambda \begin{pmatrix} b_1 \\ b_2 \end{pmatrix} \tag{2.1.7}$$

Clearly, the problem is now a 2x2 matrix eigenvalue problem. There are two eigenvectors corresponding to two eigenvalues. These are

$$\begin{array}{cc} \sqrt{\tfrac{1}{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} & \qquad \sqrt{\tfrac{1}{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} \\ \\ \lambda = \lambda_1 \;=\; 1 & \qquad \lambda = \lambda_2 \;=\; -1 \end{array} \tag{2.1.8}$$

Then we have

$$\begin{aligned} g_1 &= \sqrt{\frac{1}{2}}\,\phi_1 + \sqrt{\frac{1}{2}}\,\phi_2 \\ g_2 &= \sqrt{\frac{1}{2}}\,\phi_1 - \sqrt{\frac{1}{2}}\,\phi_2 \ , \end{aligned} \tag{2.1.9}$$

with corresponding eigenvalues $\lambda_1 = 1$ and $\lambda_2 = -1$.

Now suppose that the system has been prepared at $t = 0$ to be in the eigenstate with eigenfunction $g_1(\boldsymbol{x})$ of A. This is the initial condition for the evolution of the system. This means we set $\psi(\boldsymbol{x}, 0) = g_1(\boldsymbol{x})$. This initial condition can be ensured, for instance, by taking an ensemble (i.e., collection) of many similar systems and measuring the observable A for each at $t = 0$. Then consider **only** those systems in the ensemble for which the measurement yielded the value λ_1.

Thus using familiar notation

$$\Psi(\boldsymbol{x}, 0) \;=\; g_1(\boldsymbol{x}) \;=\; a_1\,\phi_1(\boldsymbol{x}) \;+\; a_2\,\phi_2(\boldsymbol{x}) \ , \tag{2.1.10}$$

and from eqn. (2.1.9) we can identify

$$a_1 \;=\; \sqrt{\frac{1}{2}} \quad \text{and} \quad a_2 \;=\; \sqrt{\frac{1}{2}} \ . \tag{2.1.11}$$

Thus the wavefunction $\psi(\boldsymbol{x}, t)$ is given by (eqn. (2.6))

$$\psi(\boldsymbol{x}, t) \;=\; \sqrt{\frac{1}{2}}\left(e^{-iE_1 t/\hbar}\phi_1(\boldsymbol{x}) \;+\; e^{-iE_2 t/\hbar}\phi_2(\boldsymbol{x})\right) \ . \tag{2.1.12}$$

Using eqn. (2.1.9) we can rewrite this equation as

$$\psi(\boldsymbol{x}, t) \;=\; c_1(t)\, g_1(\boldsymbol{x}) \;+\; c_2(t)\, g_2(\boldsymbol{x}) \ , \tag{2.1.13}$$

where

$$c_1(t) = \frac{1}{2}\left(e^{-iE_1 t/\hbar} \;+\; e^{-iE_2 t/\hbar}\right) \qquad c_2(t) = \frac{1}{2}\left(e^{-iE_1 t/\hbar} \;-\; e^{-iE_2 t/\hbar}\right) \ . \tag{2.1.14}$$

Now let's answer the question:

"Given that at $t = 0$ the system is in the eigenstate with eigenfunction g_1 of operator A, what is the probability, $P_1^A(t)$ that the observable corresponding to A will be measured to be λ_1 at time t? "

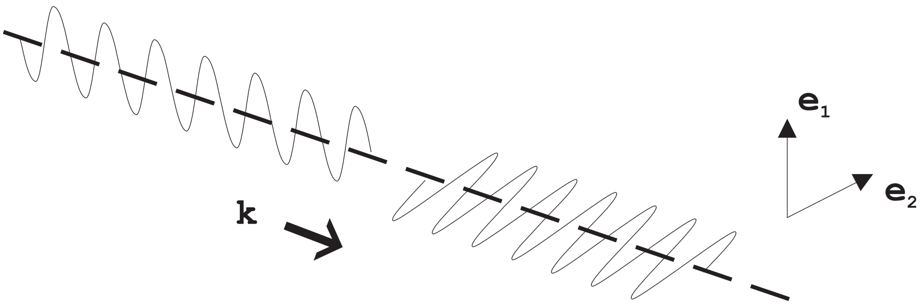
From all that has gone before we know that

$$P_1^A(t) \;=\; |c_1(t)|^2 \;=\; \cos^2\left(\frac{(E_2 - E_1)t}{2\hbar}\right) \; . \tag{2.1.15}$$

This means that from our ensemble of prepared systems a proportion $P_1^A(t)$ will yield the value λ_1 on measurement of the observable corresponding to the operator A.

3 Photon polarization states

A photon of light consists of an electromagnetic wave oscillating transversely to direction of propagation of the beam. The energy, E of the photon is given in terms of its momentum, $\boldsymbol{k}$, by $E \;=\; \hbar\omega \;=\; c|\boldsymbol{k}|$. The unit vector lying in the plane of oscillation of the photon's electric field and perpendicular to the propagation direction, $\hat{\boldsymbol{k}}$, is called the **polarization vector**, denoted $\boldsymbol{e}$. Clearly, there are two independent polarization vectors which can be denoted $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$: $\boldsymbol{e}_1 \cdot \hat{\boldsymbol{k}} \;=\; \boldsymbol{e}_2 \cdot \hat{\boldsymbol{k}} \;=\; 0$.



In general we have

$$\boldsymbol{e} \;=\; \cos\theta\,\boldsymbol{e}_1 \;+\; \sin\theta\,\boldsymbol{e}_2 \; . \tag{3.1}$$

Thus the photon has two possible polarization states which we represent by

$$\psi_1 \;=\; \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \psi_2 \;=\; \begin{pmatrix} 0 \\ 1 \end{pmatrix} , \tag{3.2}$$

with associated polarization vectors $\boldsymbol{e}_1$ and $\boldsymbol{e}_2$, respectively.

Using eqn. (3.1) the general state of a **plane-polarized** photon with polarization vector $\boldsymbol{e}$ is thus

$$\psi \;=\; \cos\theta\,\psi_1 \;+\; \sin\theta\,\psi_2 \;=\; \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} \; . \tag{3.3}$$

The inner product is

$$(\chi, \psi) \;=\; (\chi^*)^T \cdot \psi \; , \tag{3.4}$$

i.e., the dot product in two complex dimensions: χ and ψ are vectors in a 2D complex space.

Note, for example:

(i) The polarization angle θ is given by

$$\cos\theta \;=\; (\psi_1, \psi) \qquad\qquad \sin\theta \;=\; (\psi_2, \psi)\ . \tag{3.5}$$

(ii) The probability of measuring the photon to have polarization in the 1-direction (i.e., $\boldsymbol{e}_1$ in the diagram) is (c.f. eqn. (2.20))

$$P_1 \;=\; |(\psi_1, \psi)|^2 \;=\; \cos^2\theta\ . \tag{3.6}$$

When a beam of light passes through a sheet of polaroid all the emergent photons have the **same** polarization which is determined by the orientation of the polaroid. In this way we can prepare **plane-polarized** photons in a given polarization state.

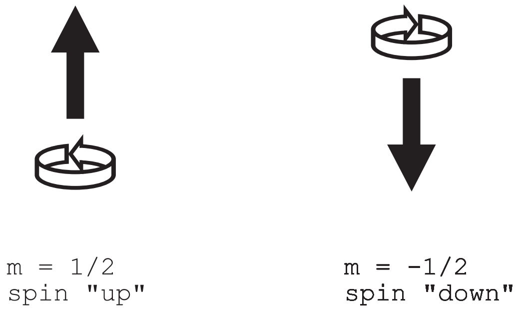
Circularly polarized light has wavefunctions

$$\psi_\pm \;=\; \sqrt{\frac{1}{2}}\,(\psi_1 \;\pm\; i\psi_2) \;=\; \sqrt{\frac{1}{2}}\begin{pmatrix} 1 \\ \pm i \end{pmatrix}\ . \tag{3.7}$$

4 Electron spin

We associate angular momentum with orbital motion of particles, say electrons in an atom. In quantum mechanics the orbital angular momentum operator is $\boldsymbol{L} \;=\; \boldsymbol{x} \wedge \boldsymbol{p}$. We can find simultaneous eigenfunctions for the operators $\boldsymbol{L}^2$ and L_z with eigenvalues $\hbar^2 l(l+1)$ and $\hbar m$ respectively, with the restriction that l and m are **integers** with $l \geq 0$ and $-l \leq m \leq l$.

However, there is another kind of angular momentum which has nothing to do with motion of a particle but is an *intrinsic* property of the particle itself. This is called **spin**. The electron has spin angular momentum $l \;=\; 1/2$ and so the allowed values of m are $\pm 1/2$. (Note: half-integer values of l and m were **not** allowed in the case of orbital angular momentum.) Thus the electron has an internal degree of freedom which is a kind of discrete internal coordinate and can only take one of the two values $\pm 1/2$. Classically the spin might be pictured as the rotation of the particle on its own axis but it is, in fact, a purely quantum phenomenon since its magnitude is proportional to $\hbar$.



The presence of spin is detected because it gives rise to a **magnetic moment**, **M**, just as a circulating current does: the electron is a minute bar-magnet like a compass needle. Thus an electron in a magnetic field has a potential energy which is given by

$$E \;=\; -\,\mathbf{B}\cdot\mathbf{M}\ . \tag{4.1}$$

where **B** is the field. The existence of electron spin can be detected in atomic spectra (fine-structure) because of atomic energy-level shifts due to the electron magnetic moment interacting with internal magnetic fields in the atom. Also by applying external magnetic fields to electron beams or to atoms, this energy can again be detected as shifts in the energy levels. Nuclear Magnetic Resonance (NMR) relies on this effect for proton spin in the nucleus.

The important message is that the electron has **two** internal states which we shall from now on label by $s \;=\; \pm 1/2$.

5 Many-particle wavefunctions

Consider a system of two particles which is described by the wavefunction $\psi(\boldsymbol{x}_1, \alpha_1, \boldsymbol{x}_2, \alpha_2, t)$, where α_i stands for other degrees of freedom: e.g., for an electron it is the spin which takes values in $[-1/2, 1/2]$. The state is normalized:

$$\sum_{\alpha_1\alpha_2} \int |\psi(\boldsymbol{x}_1, \alpha_1, \boldsymbol{x}_2, \alpha_2, t)|^2 \; d^3x_1 d^3x_2 \;=\; 1 \;, \tag{5.1}$$

and the function

$$P(\boldsymbol{x}_1, \alpha_1, \boldsymbol{x}_2, \alpha_2, t) \;=\; |\psi(\boldsymbol{x}_1, \alpha_1, \boldsymbol{x}_2, \alpha_2, t)|^2 \tag{5.2}$$

gives the probability density at time t for finding particle 1 at $\boldsymbol{x}_1$ with spin α_1 and particle 2 at $\boldsymbol{x}_2$ with α_2. The single particle probability density for particle 1 is given by

$$P(\boldsymbol{x}_1, \alpha_1, t) \;=\; \sum_{\alpha_2} \int |\psi(\boldsymbol{x}_1, \alpha_1, \boldsymbol{x}_2, \alpha_2, t)\,|^2 \; d^3x_2 \;, \tag{5.3}$$

and likewise for particle 2.

The Hamiltonian will in general take the form:

$$H \;=\; -\frac{\hbar^2}{2m_1}\nabla_1^2 \;-\; \frac{\hbar^2}{2m_2}\nabla_2^2 \;+\; V(\boldsymbol{x}_1, \boldsymbol{x}_2) \;, \tag{5.4}$$

where the subscripts "1" and "2" label the two particles. (Note: In principle H can depend on the "coordinates" α_i, e.g., for electrons it would depend on spin if magnetic fields acted on the system. However, this is an unnecessary complication for our purposes.)

In the special case that $V(\boldsymbol{x}_1, \boldsymbol{x}_2) \;=\; V_1(\boldsymbol{x}_1) + V_2(\boldsymbol{x}_2)$ the Hamiltonian is **separable** which means that:

$$H \;=\; H_1(\boldsymbol{x}_1) \;+\; H_2(\boldsymbol{x}_2) \;. \tag{5.5}$$

We can write

$$H_1\phi_n^{(1)}(\boldsymbol{x}_1, \alpha_1) = E_n^{(1)}\phi_n^{(1)}(\boldsymbol{x}_1, \alpha_1) \qquad H_2\phi_n^{(2)}(\boldsymbol{x}_2, \alpha_2) = E_n^{(2)}\phi_n^{(2)}(\boldsymbol{x}_2, \alpha_2) \;,$$

$$\Longrightarrow$$

$$H\,\phi_{n_1}^{(1)}(\boldsymbol{x}_1, \alpha_1)\phi_{n_2}^{(2)}(\boldsymbol{x}_2, \alpha_2) \;=\; (E_{n_1}^{(1)} + E_{n_2}^{(2)})\,\phi_{n_1}^{(1)}(\boldsymbol{x}_1, \alpha_1)\phi_{n_2}^{(2)}(\boldsymbol{x}_2, \alpha_2) \;. \tag{5.6}$$

Thus in the separable case the two-particle eigenfunctions take the form

$$\psi_{n_1n_2}(\boldsymbol{x}_1, \alpha_1, \boldsymbol{x}_2, \alpha_2, t) \;=\; \psi_{n_1}^{(1)}(\boldsymbol{x}_1, \alpha_1, t)\psi_{n_2}^{(2)}(\boldsymbol{x}_2, \alpha_2, t) \;, \tag{5.7}$$

with energy $E^{(1)}_{n_1} + E^{(2)}_{n_2}$. These eigenfunctions form a complete basis for the space of all two-particle wavefunctions and thus

$$\psi(\boldsymbol{x}_1, \alpha_1, \boldsymbol{x}_2, \alpha_2, t) \;=\; \sum_{n_1 n_2} c_{n_1 n_2} \psi_{n_1 n_2}(\boldsymbol{x}_1, \alpha_1, \boldsymbol{x}_2, \alpha_2, t) \;. \tag{5.8}$$

This analysis can easily be extended to systems of N particles.

5.1 Identical particles

Suppose the system consists of N identical particles. Then labelling them $1, 2, \ldots$etc. wrongly distinquishes them. For example, there is no way of distinquishing

$$\psi(\boldsymbol{x}_1, \alpha_1, \boldsymbol{x}_2, \alpha_2, t) \qquad \text{from} \qquad \psi(\boldsymbol{x}_2, \alpha_2, \boldsymbol{x}_1, \alpha_1, t) \;, \tag{5.1.1}$$

experimentally. They must be the **same** state with the **same** energy. They may differ by a phase since this does not affect probabilities or energies. Somehow we must construct the states to take into account the indistinquishability of the particles. This is done in one of the two following ways which have far reaching consequences for most physical systems:

(i) **Fermi-Dirac statistics** states that the wavefunction must be **antisymmetric** under interchange of any two particles. In the example above this means $1 \;\leftrightarrow\; 2$ with a phase factor of -1. Particles which obey Fermi-Dirac statistics are called **fermions** and include all spin-1/2 particles and in particular the electron.

(ii) **Bose-Einstein statistics** states that the wavefunction for identical particles is **symmetric** under interchange of any pair of particles. Particles which obey Bose-Einstein statistics are called **bosons** which includes the photon.

Thus we have:

$$\begin{aligned} \text{Fermi-Dirac:} \qquad & \psi(\boldsymbol{x}_1, \alpha_1, \boldsymbol{x}_2, \alpha_2, t) \;=\; -\,\psi(\boldsymbol{x}_2, \alpha_2, \boldsymbol{x}_1, \alpha_1, t) \;, \\ \text{Bose-Einstein:} \qquad & \psi(\boldsymbol{x}_1, \alpha_1, \boldsymbol{x}_2, \alpha_2, t) \;=\; +\,\psi(\boldsymbol{x}_2, \alpha_2, \boldsymbol{x}_1, \alpha_1, t) \;. \end{aligned} \tag{5.1.2}$$

In the separable case for two particles

$$H \;=\; -\,\frac{\hbar^2}{2m}\nabla_1^2 \;-\; \frac{\hbar^2}{2m}\nabla_2^2 \;+\; V(\boldsymbol{x}_1) \;+\; V(\boldsymbol{x}_2) \;, \tag{5.1.3}$$

and the wavefunction becomes

$$\Psi \;=\; \sqrt{\tfrac{1}{2}}\,(\psi_1(\boldsymbol{x}_1, \alpha_1, t)\psi_2(\boldsymbol{x}_2, \alpha_2, t) \;\mp\; \psi_1(\boldsymbol{x}_2, \alpha_2, t)\psi_2(\boldsymbol{x}_1, \alpha_1, t)) \;, \tag{5.1.4}$$

with $(-)$ for Fermi-Dirac and $(+)$ for Bose-Einstein.

5.2 The Exclusion Principle

Consider two electrons and neglect the interaction between them. The Hamiltonian is then separable. Put one in single particle state $\psi_1(\boldsymbol{x}, s, t)$ and the other in $\psi_2(\boldsymbol{x}, s, t)$. The total energy is $E \;=\; E_1 + E_2$ and the correct wavefunction is

$$\Psi(\boldsymbol{x}_1, s_1, \boldsymbol{x}_2, s_2, t) \;=\; \sqrt{\tfrac{1}{2}}\,(\psi_1(\boldsymbol{x}_1, s_1, t)\psi_2(\boldsymbol{x}_2, s_2, t) \;-\; \psi_1(\boldsymbol{x}_2, s_2, t)\psi_2(\boldsymbol{x}_1, s_1, t)) \;. \tag{5.2.1}$$

If $\psi_1 \;=\; \psi_2$ then $\Psi \;=\; 0$. This is the **Exclusion Principle** which states that

> “No two electrons may occupy identical single particle states.”

For a system of N electrons with separable Hamiltonian the energy eigenfunctions take the form

$$\Psi(\boldsymbol{x}_1, s_1, \ldots, \boldsymbol{x}_N, s_N, t) \;=\; \Phi(\boldsymbol{x}_1, s_1, \ldots, \boldsymbol{x}_N, s_N, t)\, e^{-iEt/\hbar}\;, \qquad (5.2.2)$$

where Φ can be written most conveniently as

$$\Phi \;=\; \sqrt{\frac{1}{N!}} \begin{vmatrix} \phi_1(\boldsymbol{x}_1, s_1) & \phi_1(\boldsymbol{x}_2, s_2) & \ldots\ldots & \phi_1(\boldsymbol{x}_N, s_N) \\ \phi_2(\boldsymbol{x}_1, s_1) & \phi_2(\boldsymbol{x}_2, s_2) & \ldots\ldots & \phi_2(\boldsymbol{x}_N, s_N) \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \phi_N(\boldsymbol{x}_1, s_1) & \phi_N(\boldsymbol{x}_2, s_2) & \ldots\ldots & \phi_N(\boldsymbol{x}_N, s_N) \end{vmatrix}\;, \qquad (5.2.3)$$

with $E \;=\; \sum_{i=1}^{N} E_i$. Fermi-Dirac statistics are automatically satisfied.

This is the **Slater determinant** and eigenfunctions of this kind form the complete basis set of functions for all atomic physics calculations.

E.g., for $N = 3$ we have $E = E_1 + E_2 + E_3$ and

$$\Phi(\boldsymbol{x}_1, s_1, \boldsymbol{x}_2, s_2, \boldsymbol{x}_3, s_3) \;=\; \sqrt{\frac{1}{6}} \begin{vmatrix} \phi_1(\boldsymbol{x}_1, s_1) & \phi_1(\boldsymbol{x}_2, s_2) & \phi_1(\boldsymbol{x}_3, s_3) \\ \phi_2(\boldsymbol{x}_1, s_1) & \phi_2(\boldsymbol{x}_2, s_2) & \phi_2(\boldsymbol{x}_3, s_3) \\ \phi_3(\boldsymbol{x}_1, s_1) & \phi_3(\boldsymbol{x}_2, s_2) & \phi_3(\boldsymbol{x}_3, s_3) \end{vmatrix}\;, \qquad (5.2.4)$$

In the case that the Hamiltonian does not depend on the spin of the electron the energy eigenstates which differ just by spin orientation have the **same** energy and **same** spatial wavefunction. In this circumstance we may ignore the presence of spin in many of our calculations except to remember that any **given** state may be occupied by at most two electrons: one with spin “up” $(+1/2)$ and one with spin “down” $(-1/2)$. For example, in the lithium atom, which has three electrons, the (non-degenerate) ground state is occupied by two electrons with opposing spins and the Exclusion Principle then requires that the third electron occupy the first excited state.

The Exclusion Principle is the key to atomic structure and the periodic table – if it were not true then all atomic electrons would be in the lowest (i.e., ground) state and the chemical richness of the elements would not occur.

(For detailed theory of spin see the Easter term course on “Symmetries and Groups in Quantum Physics”)

6 Quantum Statistical Physics

To deal with large systems of many particles statistical methods are essential. We can write down the classical equations of motion or the Schrödinger equation for the system but, with typically 10^{23} particles in the system, we have no hope of solving for the full microscopic motion or wavefunction of the system. There are even conceptual difficulties of which chaos is one manifestation. However, we shall assume that the system is in **equilibrium** and so then we can describe its properties by a few **macroscopic** variables such as pressure, temperature, volume, etc.

(1) The whole system can occupy any one of a sequence of macroscopic energy eigenstates each of which is uniquely specified by a complete (and very large) set of quantum numbers. For a **given** energy eigenvalue there will be a very large number of states with that energy. That is, the **degeneracy** of the energy level is very large.

(2) The important principle underlying statistical physics is that

> **The system will occupy all macroscopic eigenstates of a given energy with equal probability.**

An important example that has wide applicability is of a system which consists of N non-interacting particles.

(i) A given particle can occupy any one of a sequence of single-particle energy eigenstates of energy E_i, $i = 1, 2, \ldots, \infty$. In general, there is more than one state with a given energy and the the number, g_i, of states with a given energy E_i is called the degeneracy of that energy level.

(ii) A given macroscopic energy eigenstate of the whole system will be described by a wavefunction which is the product of the individual single-particle wavefunctions in a form which generalizes equation 5.7 or 5.1.4. The whole system will have **very** many such eigenstates with the **same** macroscopic energy and so a given energy level has a very large degeneracy.

(iii) Consider all macroscopic eigenstates of a **given** energy E and total particle number N. Let the number of particles with energy E_i be N_i: the N_i are the **occupation numbers** of the single particle levels. All macroscopic observables will be functions only of the N_i and so we shall average over all other quantum numbers which distinguish the states.

(iv) There will be many macroscopic energy eigenstates associated with the same set of occupation numbers $\{N_i\}$ and equilibrium is characterized by the set $\{\overline{N_i}\}$ which is **most probable**. Using (2) above this will be that set of occupation numbers which is associated with the largest number of macroscopic states of the system.

(v) Because N is large the most probable set is **overwhelmingly dominant** and there are uniques values for the macroscopic quantities which then characterize the equilibrium.

6.1 The Boltzmann Distribution

The number of particles with energy E_i is N_i, and we have

$$N \;=\; \sum N_i \;, \qquad\qquad E \;=\; \sum N_i E_i \;. \tag{6.1.1}$$

Now,

(i) there are

$$\frac{N!}{N_1!N_2!\ldots}$$

of partitioning N into the set $\{N_i\}$, and

(ii) for a **given** energy E_i there are $g_i^{N_i}$ ways of assigning the N_i particles since each particle has a choice of g_i levels.

Thus the number of ways of realising this distribution is

$$\frac{N!}{N_1!N_2!\ldots}\, g_1^{N_1} g_2^{N_2} \ldots \ . \tag{6.1.2}$$

Hence, the probability of realising the partition of N into the set $\{N_i\}$ is

$$P(\{N_i\}) \ \propto \ \prod_i \frac{g_i^{N_i}}{N_i!} \ . \tag{6.1.3}$$

The $N!$ is an irrelevant factor since P must be normalized to unity.

Now maximize P with N and E fixed. Use Lagrange multipliers and maximize

$$\begin{aligned} f(\{N_i\}) &= \log P - \alpha N - \beta E \\ &= \sum_i \left[(N_i \log g_i - \log N_i!) \ - \ N_i\,(\alpha + \beta E_i) \right] \ , \end{aligned} \tag{6.1.4}$$

with respect to each of the N_i. Stirling's formula for $N_i \to \infty$ gives

$$\log N_i! \ \approx \ N_i(\log N_i - 1) \ + \ \mathrm{O}(\log N_i) \ ,$$

and so the stationary point is given by $N_i = \overline{N}_i$, for each i, where $\overline{N}_i$ satisfies

$$\begin{aligned} \log g_i \quad &- \quad (\log \overline{N}_i - 1) \ - \ 1 \ - (\alpha + \beta E_i) \ = \ 0 \ , \\ \Longrightarrow \\ \overline{N}_i \quad &= \quad g_i\, e^{-\alpha - \beta E_i} \end{aligned} \tag{6.1.5}$$

Using the constraints from eqn. (6.1.1), we can determine α by

$$N \ = \ e^{-\alpha}\, Z \qquad \Longrightarrow \qquad \alpha \ = \ \log \frac{Z}{N} \ ,$$

where

$$Z \ = \ \sum_i g_i\, e^{-\beta E_i} \qquad \textbf{the partition function} \tag{6.1.6}$$

$$\overline{N}_i \ = \ \frac{N}{Z}\, g_i\, e^{-\beta E_i} \qquad \textbf{the mean occupation number} \tag{6.1.7}$$

Note that the mean occupation number for **one given** level of energy E is thus $\frac{N}{Z} e^{-\beta E}$. We also have

$$E \ = \sum_i \overline{N}_i E_i \ = \ \frac{N}{Z} \sum_i g_i E_i e^{-\beta E_i} \ = \ - N \frac{d}{d\beta} \log\, Z \ . \tag{6.1.8}$$

The partition function, Z, is fundamental to statistical physics.

The equilibrium state is characterized by the value of β, and β^{-1} is identified with the **temperature**:

$$\beta \ = \ \frac{1}{kT} \ . \tag{6.1.9}$$

T is the temperature in degrees Kelvin (^0K) and k is Boltzmann's constant:

$$k \ = \ 1.38 \cdot 10^{-23}\,\text{Joules(J)}/\ ^0K \ = \ 0.862 \cdot 10^{-4}\,\text{electron-volts(eV)}\ ^0K \ .$$

[Remember, a **Joule(J)** is a unit of energy and one Joule is the work done when a force of one Newton acts over a distance of one metre: one Joule = one Newton-metre, and 1eV = $1.60 \cdot 10^{-19} J$ (see later for more on eV).]

Why is this interpretation sensible? Bring two systems together each of which are **separately in equilibrium** characterized by N_1, E_1, β_1 and N_2, E_2, β_2 respectively. Together they make a system with $N = N_1 + N_2$ and $E = E_1 + E_2$ but which is initially **not** in equilibrium since it is not characterized by a single value of β. The whole system will move to equilibrium and will then be characterized by a single value of β, β_e. The two systems are still distinquishable but they are **now in equilibrium with each other** since they have a common value of β, namely β_e. This is exactly what we expect of temperature also: an important condition for two systems to be in equilibrium is that they have the **same** temperature.

The occupation number distribution as a function of energy given in eqn. (6.1.7) is called the **Boltzmann distribution**.

6.1.1 The example of a perfect gas

A perfect gas consists of a system of particles which do not interact with each other. The gas is in a cubical box with walls at temperature, T and energy will be exchanged between the particles and the walls when the particle collide (inelastically) with the walls. The gas will come into equilibrium with the box and will have temperature, T, also.

Let the number of particles be N, the box be of side L with volume $V = L^3$, and the mass of each particle be m. The wavefunction for one particle in the box is

$$\phi(\boldsymbol{x}) \ = \ \sqrt{\frac{8}{V}}\sin\left(\frac{\pi l_1 x_1}{L}\right)\sin\left(\frac{\pi l_2 x_2}{L}\right)\sin\left(\frac{\pi l_3 x_3}{L}\right) \ , \qquad (6.1.1.1)$$

with energy

$$E \ = \ \frac{\hbar^2\pi^2}{2mL^2}(l_1^2 + l_2^2 + l_3^2) \quad \text{and} \quad p^2 \ = \ 2mE \ , \qquad (6.1.1.2)$$

with l_i integers and $l_i > 0$. In the limit that $L \to \infty$ we can treat the energy as taking values in a continuous spectrum and then the Boltzmann distribution gives the **occupation number density**, $n(E)$, as a function of energy:

$$\frac{dN(E)}{dE} \ \equiv \ n(E) \ = \ Cg(E)e^{-E/kT} \ , \qquad (6.1.1.3)$$

where $g(E)dE$ is the number of energy states in the interval $(E, E + dE)$, C is a normalization constant, and $N(E)$ is the number of particles with energy less than E . Let $\boldsymbol{l} = (l_1, l_2, l_3)$ be a vector in the positive octant of a 3D space. Then fixed E means fixed length radius $|\boldsymbol{l}|$:

$$R(E)^2 \ = \ |\boldsymbol{l}|^2 \ = \ \frac{2mL^2E}{\pi^2\hbar^2} \ , \qquad (6.1.1.4)$$

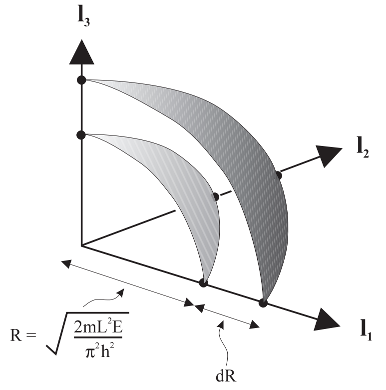
and the number of states (in the limit that $|\boldsymbol{l}| \gg 1$) in $(E, E + dE)$ is the volume of the shell in the positive quadrant between spheres of radii $R(E)$ and $R(E) + (dR/dE)dE$

(see the figure). Using eqn (6.1.1.4) this volume is

$$g(E)dE = \frac{4\pi}{8}R^2\frac{dR}{dE}dE = \frac{(2m)^{3/2}}{4\pi^2\hbar^3}VE^{1/2}dE ,$$

$$\Longrightarrow$$

$$g(E) = \frac{(2m)^{3/2}}{4\pi^2\hbar^3}VE^{1/2} . \qquad (6.1.1.5)$$



We can now calculate a number of macroscopic properties of the gas. Of course, we must always remember that the outcome must be compatible with the $L \to \infty$, $|\boldsymbol{l}| \to \infty$ limits. Note that $g(E) \propto$ volume, V.

(i) First we must normalize the number distribution by calculating C in eqn. (6.1.1.3). We can absorb all other constants into C and write:

$$\begin{aligned} N &= C\int_0^\infty g(E)e^{-E/kT}\,dE \\ &= C'\int_0^\infty e^{-E/kT}E^{1/2}dE . \end{aligned} \qquad (6.1.1.6)$$

At this point we note the results

$$\begin{aligned} \int_0^\infty u^2e^{-\lambda u^2}\,du &= \frac{1}{4}\sqrt{\frac{\pi}{\lambda}} , \\ \int_0^\infty u^4e^{-\lambda u^2}\,du &= \frac{3}{8}\sqrt{\frac{\pi}{\lambda}} . \end{aligned} \qquad (6.1.1.7)$$

Using the change of variable $u^2 = E$, we find

$$\int_0^\infty e^{-E/kT}E^{1/2}dE = 2\int_0^\infty u^2e^{-u^2/kT}\,du = \frac{\pi^{1/2}}{2}(kT)^{3/2} . \qquad (6.1.1.8)$$

Thus we get

$$C' = \frac{2N}{\pi^{1/2}}\frac{1}{(kT)^{3/2}} . \qquad (6.1.1.9)$$

Then we have

$$n(E) = C'\,e^{-E/kT}E^{1/2} . \qquad (6.1.1.10)$$

(ii) The mean energy is given by

$$\begin{aligned}\overline{E} &= \int_0^\infty E\,n(E)\,dE\\ &= C'\int_0^\infty E^{3/2}e^{-E/kT}dE\ .\end{aligned}$$

Substituting $u^2 = E$ we get

$$\begin{aligned}\overline{E} &= 2C'\int_0^\infty u^4\,e^{-u^2/kT}\,du\\ &= C'\frac{3}{4}\pi^{1/2}(kT)^{5/2}\ .\end{aligned}$$

Substituting for C' gives

$$\boxed{\overline{E} \;=\; \tfrac{3}{2}NkT \qquad \textbf{equipartition of energy}} \qquad (6.1.1.11)$$

This is a fundamental result that relates the mean energy per particle to the temperature. **Equipartition of energy** says that

“ The mean energy **per degree of freedom** is $\frac{1}{2}\,kT$ ”.

In three dimensions we get the result in the box above and the generalization to D dimensions is obvious. Our result corresponds to a value of (eqn. (6.1.1.4))

$$|\boldsymbol{l}|^2 \;=\; \frac{3kTmL^2}{\hbar^2\pi^2}\ . \qquad (6.1.1.12)$$

For **fixed** temperature, T, we see that $|\boldsymbol{l}| \to \infty$ as $L| \to \infty$ which justifies the assumptions of continuity made earlier. With $\hbar = 1.05\cdot 10^{-34} Jsecs$ and $K = 1.38\cdot 10^{-23} J/\,^0K$ we find

$$|\boldsymbol{l}|^2 \;=\; 1.52\cdot 10^{19}\,L^2T\ , \qquad (6.1.1.13)$$

where L is in metres and T is ^0K. Clearly, for all reasonable physical situations our assumptions are valid! (Note: $T = 300\,^0K$ is about room temperature.)

6.2 Distribution for Indistinquishable Particles

In the derivation of the most probable distribution above we used quantum methods but treated the particles as **distinquishable**, i.e., we counted the number of ways the particles could be distributed amongst the states treating **each** rearrangement as distinct even if the individual energy-level occupation numbers were **unchanged**. This procedure must be modified for indistinquishable particles.

For example, consider a system with just two states ϕ_1 and ϕ_2 and two particles in the system. Before we would count 4 ways of distributing the particles between the states:

	(1)	(2)	(3)	(4)
ϕ_1	1 2	1	2	−−
ϕ_2	−−	2	1	1 2

If the particles are **indistinquishable** then (2) and (3) are the **same** distribution since out of the two possible quantum states $\phi_1(\boldsymbol{x}_1)\phi_2(\boldsymbol{x}_2)$ and $\phi_1(\boldsymbol{x}_2)\phi_2(\boldsymbol{x}_1)$ only:

$$\frac{1}{\sqrt{2}}\left(\phi_1(\boldsymbol{x}_1)\phi_2(\boldsymbol{x}_2)\ +\ \phi_1(\boldsymbol{x}_2)\phi_2(\boldsymbol{x}_1)\right) \qquad (6.2.1)$$

survives for Bose particles, and only

$$\frac{1}{\sqrt{2}}\left(\phi_1(\boldsymbol{x}_1)\phi_2(\boldsymbol{x}_2)\ -\ \phi_1(\boldsymbol{x}_2)\phi_2(\boldsymbol{x}_1)\right) \qquad (6.2.2)$$

survives for Fermi particles. In addition configurations (1) and (4) are forbidden by the Exclusion Principle in the case of fermions. So for bosons **three** distributions or configurations are distinct and for fermions only **one** configuration survives.

In the general case of many particles and levels we must treat bosons and fermions separately to get the right occupation number distributions. We first consider bosons and derive the Bose-Einstein distribution.

6.2.1 Bose-Einstein statistics

As before consider levels of energy E_i and degeneracy g_i. Assign N_i particles to the g_i levels of energy E_i. The number of different arrangements of the N_i particles among those levels is given by the number of ways of writing N_i as a sum of non-negative integers,

$$N_i\ =\ n_1 + n_2 + \ldots + n_{g_i}\ . \qquad (6.1.2.1)$$

Then there are n_1 in level 1, n_2 in level 2, and so on. This is the number of ways of dividing N_i dots by $g_i - 1$ strokes:

$$\underbrace{\underset{n_1}{\bullet\bullet}\ \mid\ \underset{n_2}{\bullet\bullet}\ \mid\ \bullet\bullet\ \ldots\ldots\ \underset{n_j}{\bullet}\ \mid\mid\ \underset{n_{j+1}}{\bullet\bullet}\ \mid\ \bullet\bullet\ \ldots}_{N_i\ \text{dots},\ g_i-1\ \text{strokes}}$$

The number of ways of arranging the lines and dots is calculated by considering $N_i + g_i - 1$ locations each occupied by a dot or a line. Assign the dots first. The first dot can be put into one of $(N_i + g_i - 1)$ locations, the next into one of $(N_i + g_i - 2)$ etc. When the N_i dots are assigned the lines just fill up the rest. Hence, so far there are

$$(N_i + g_i - 1)(N_i + g_i - 2)\ldots(g_i)\ =\ \frac{(N_i + g_i - 1)!}{(g_i - 1)!}$$

ways of doing this. However, the **order** in which the dots were assigned does not matter since the bosons are indistinquishable. Hence, the number of **distinquishable** ways is

$$W_i(N_i, g_i)\ =\ \frac{(N_i + g_i - 1)!}{N_i!\,(g_i - 1)!}\ . \qquad (6.1.2.2)$$

Thus the total number of ways of distributing the particles amongst the levels is

$$P\ =\ \prod_i W_i(N_i, g_i)\ =\ \prod_i \frac{(N_i + g_i - 1)!}{N_i!\,(g_i - 1)!}\ , \qquad (6.1.2.3)$$

with $E\ =\ \sum_i N_i E_i\ ,\quad N\ =\ \sum_i N_i\ .$

As before, maximize P with respect to the $\{N_i\}$ accounting for the constraints using Lagrange multipliers. We maximize $f \;=\; \log P \;-\; \beta E \;-\; \alpha N$ with

$$\log P \;=\; \sum_i \left[(N_i + g_i - 1)\log(N_i + g_i - 1) \;-\; (g_i - 1)\log(g_i - 1) \;-\; N_i \log N_i\right] \tag{6.1.2.4}$$

If $\{N_i\} = \{\overline{N}_i\}$ at the stationary point we have

$$\frac{\partial f}{\partial N_i} \;=\; \log(\overline{N}_i + g_i - 1) \;-\; \log \overline{N}_i \;-\; \beta E_i \;-\; \alpha \;=\; 0\ . \tag{6.1.2.5}$$

It is reasonable to assume that $g_i \gg 1$ (remember $g(\overline{E})$ in the calculation of equipartition is huge), and then we can ignore the odd “1” lying around to get

$$\frac{\overline{N}_i}{\overline{N}_i + g_i} \;=\; e^{-\beta E_i - \alpha} \tag{6.1.2.6}$$

or

$$\boxed{\overline{N}_i \;=\; \frac{g_i}{e^{\beta(E_i - \mu)} \;-\; 1} \qquad \textbf{Bose-Einstein distribution.} \qquad (6.1.2.7)}$$

Here $\alpha \;=\; -\beta\mu$ and $\beta \;=\; 1/kT$. Only if $\beta(E - \mu) \gg 1$, i.e., $\overline{N}_i \ll g_i$, does this become the Boltzmann distribution. In this case, the occupation numbers are so low that indistinquishability is not important. The parameter μ is the **chemical potential** which must be chosen to ensure $\sum_i \overline{N}_i = N$. Note that $\mu < min(E_i)$ since the $\{\overline{N}_i\}$ must be positive for all E_i.

The Bose-Einstein distribution is the one appropriate for photons.

6.2.2 The Fermi-Dirac distribution

We consider the same system as in the previous section but now we must impose that no more than one fermion can occupy any one given energy level because of the exclusion principle.

As before we assign N_i fermions to the g_i levels which have common energy E_i. The number of distinct ways of doing this ($N_i \leq g_i$) is

$$W_i(N_i, g_i) \;=\; \frac{g_i\,(g_i - 1)\,\ldots\,(g_i - N_i + 1)}{N_i!} \;=\; \frac{g_i!}{(g_i - N_i)!\,N_i!}\ . \tag{6.2.2.1}$$

For equilibrium and the most probable distribution maximise

$$P \;=\; \prod_i \frac{g_i!}{(g_i - N_i)!\,N_i!}\ , \tag{6.2.2.2}$$

with

$$E \;=\; \sum_i N_i E_i \qquad \text{and} \qquad N \;=\; \sum_i N_i\ . \tag{6.2.2.3}$$

As before, in fact, we take account of these constraints with Lagrange multipliers and maximize $f \;=\; \log P \;-\; \beta E \;-\; \alpha N$ to give

$$\frac{\partial}{\partial N_i} \sum_j \left[g_j \log g_j \;-\; N_j \log N_j \;-\; (g_j - N_j)\log(g_j - N_j) \;-\; \alpha N_j \;-\; \beta N_j E_j\right] \;=\; 0\ . \tag{6.2.2.4}$$

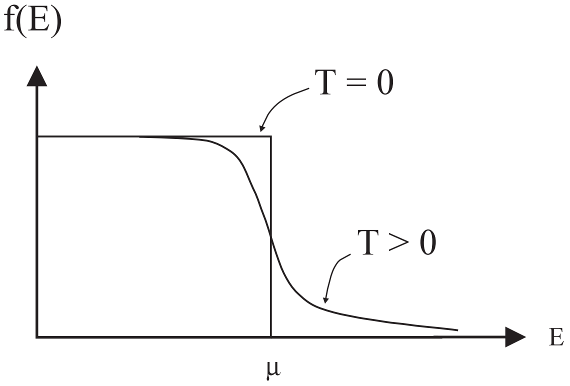
Then the most probable distribution, $\overline{N}_i$, satisfies

$$-\log \overline{N}_i \;+\; \log(g_i - \overline{N}_i) \;=\; \alpha \;+\; \beta E_i \;, \tag{6.2.2.5}$$

or

$$\boxed{\begin{aligned} \overline{N_i} \;&=\; g_i\, f(E_i) \;, && (6.2.2.6) \\ f(E) \;&=\; \frac{1}{e^{\beta(E-\mu)} \;+\; 1} \qquad \textbf{Fermi-Dirac distribution.} && (6.2.2.7) \end{aligned}}$$

Here, again, the chemical potential μ has been defined by $\alpha = -\beta\mu$ and $\beta = 1/kT$. In the context of the Fermi-Dirac distribution μ is also called the **Fermi energy** and can be denoted E_F instead. For $E_i = \mu$ we have from eqn. (6.2.2.7) that $\overline{N}_i = g_i/2$, i.e., such a level has a 50% probability of occupation. The distribution $f(E)$ looks like:



(i) We have the mean energy, $\overline{E}$ given by

$$\overline{E} \;=\; \sum_i \frac{g_i E_i}{e^{\beta(E_i-\mu)} \;+\; 1} \;\;=\;\; -\frac{\partial}{\partial\beta}\log Z + \mu N \;, \tag{6.2.2.8}$$

where the partition function, Z, is defined by

$$Z \;=\; \prod_i \Big(1 \;+\; e^{-\beta(E_i-\mu)}\Big)^{g_i} \;\;. \tag{6.2.2.9}$$

(ii) The number of particles, N, is given by

$$N \;=\; \sum_i \frac{g_i}{e^{\beta(E_i-\mu)} \;+\; 1} \;\;=\;\; -\frac{\partial}{\partial\mu}\log Z \;. \tag{6.2.2.10}$$

Since N is given, this equation determines the value of μ (remember μ is basically the Lagrange multiplier α introduced to ensure that the number of particles is fixed at N). μ is thus a function of N and T in general.

(iii) At $T = 0$ from the diagram we see that μ is the increase in the energy of the system when the particle number is increased by one.

The Fermi-Dirac distribution is applicable to electrons and will be important when we discuss semi-conductor physics.

6.3 Black-Body Radiation

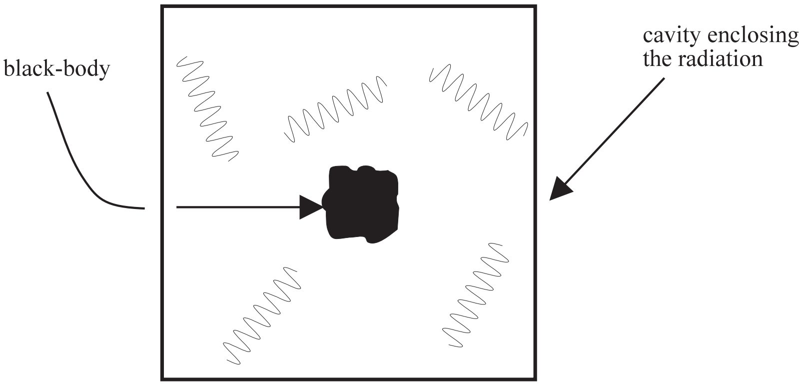
Any body with temperature $T > 0$ emits radiation in the form of photons. For example:

(i) a light bulb filament is heated by the electric current to a high temperature and it glows;

(ii) iron heated in a forge glows red hot meaning that it emits visible red light. As the iron gets hotter the colour changes toward the blue end of the spectrum, i.e., there are more photons of shorter wavelength than before, which means they have higher energy;

(iii) bodies at room temperature typically emit infra-red radiation, e.g., people. This is often referred to as "heat" but really it is only distinquished by being of longer wavelength than visible light.

Consider radiation trapped in a box with perfectly reflecting walls. The different energy states for the photons will each be occupied by different numbers of photons. Suppose, a body of temperature T is introduced into the box.



This body will emit and absorb photons and so change the occupation numbers of the levels. The box will thus contain radiation in the form of many photons of varying frequency which are being continually absorbed and emitted by the body, and these photons will come into equilibrium with it and so will conform to the **Bose-Einstein distribution**. Now, the number of photons is not constrained since they are continually being emitted from (or absorbed by) the body, and hence there is **no need** for the Lagrange multiplier α in the derivation in section (6.2.1) above. The effect is to set $\mu = 0$ for photons.

The body is a **black-body** if it can absorb and emit radiation of **all** frequencies (i.e. energies) across a wide spectral range. If it is not a black-body then photons of some frequencies **cannot** be absorbed or emitted by the body and so their number will be fixed at some non-equilibrium value. The corollary is that if radiation is in contact with a black-body at temperature T, then photons of **all** frequencies will conform to the Bose-Einstein distribution with $\mu = 0$. Thus:

Radiation in equilibrium with a **black-body** at temperature T is characterized **only** by the temperature T. Such radiation is called **black-body radiation**.

6.3.1 The Black-Body Distribution

We will calculate the occupation number density for photons of frequency ω and energy $E = \hbar\omega$ for black-body radiation.

The photon modes satisfy the standard wave equation:

$$c^2 \frac{\partial^2}{\partial t^2} \boldsymbol{F}(\boldsymbol{x}, t) \;=\; \nabla^2 \boldsymbol{F}(\boldsymbol{x}, t) \; . \tag{6.1.3.1}$$

We impose the boundary conditions that the solution is periodic in the 1-,2-, and 3-directions with period L, i.e.,

$$\boldsymbol{F}(x_1{+}L, x_2, x_3, t) \;=\; \boldsymbol{F}(x_1, x_2{+}L, x_3, t) \;=\; \boldsymbol{F}(x_1, x_2, x_3{+}L, t) \;=\; \;=\; \boldsymbol{F}(x_1, x_2, x_3, t) \; . \tag{6.1.3.2}$$

This is different from the case of the perfect gas where we imposed that the wavefunction vanished on the boundary. However, in the limit $L \to \infty$ the final result does **not depend** on the details of the boundary condition but it is a device that helps us calculate: the volume V does not appear in the expressions for physical observables.

The general normal-mode solution is the complex separable one:

$$\boldsymbol{E}(\boldsymbol{x}, t) \;=\; \boldsymbol{E}_0 \, \exp\left\{-i\left(\omega t \;-\; \frac{2\pi}{L}(l_1 x_1 + l_2 x_2 + l_3 x_3)\right)\right\} \; . \tag{6.1.3.3}$$

with

$$\frac{4\pi^2}{L^2}(l_1^2 + l_2^2 + l_3^2) \;=\; \frac{\omega^2}{c^2} \; , \qquad l_i \quad \text{integers} \; . \tag{6.1.3.4}$$

This is a travelling wave with energy $E \;=\; \hbar\omega$ and momentum $\boldsymbol{p}$

$$\boldsymbol{p} \;=\; \left(\frac{2\pi\hbar}{L} l_1, \frac{2\pi\hbar}{L} l_2, \frac{2\pi\hbar}{L} l_3\right) \; . \tag{6.1.3.5}$$

The mode is a vector function, $\boldsymbol{E}$, because there are **two** polarization states per given energy mode, and $\boldsymbol{E}_0$ encodes this internal degree of freedom. From the earlier discussion we have

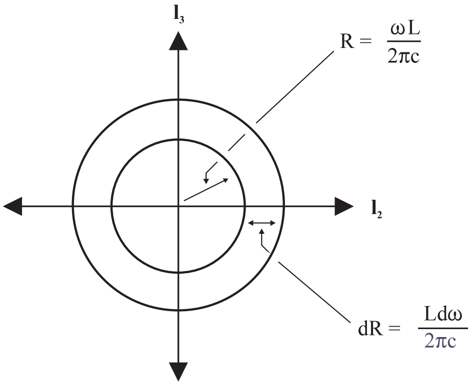
$$\boldsymbol{E}_0 \cdot \boldsymbol{p} \;=\; 0 \; . \tag{6.1.3.6}$$

As before $\boldsymbol{l} = (l_1, l_2, l_3)$ is treated as a continuous vector and from eqn. (6.1.3.4)

$$|\boldsymbol{l}|^2 \;=\; \frac{\omega^2 L^2}{4\pi^2 c^2} \; . \tag{6.1.3.7}$$

Hence,

$$\begin{aligned} & \text{The number of modes in the interval } \; (\omega, \omega + d\omega) \\ = \quad & \text{Volume of spherical shell of radius } \frac{\omega L}{2\pi c} \text{ and thickness } \frac{L d\omega}{2\pi c} \\ = \quad & \frac{4\pi\omega^2 L^3 d\omega}{8\pi^3 c^3} \; . \end{aligned}$$



Each photon has two polarizations and so the total number of states in $(\omega, \omega + d\omega)$ is twice this, namely:

$$g(\omega) d\omega \;=\; \frac{\omega^2 V}{\pi^2 c^3} d\omega \; . \tag{6.1.3.8}$$

where $g(\omega)$ is the density of states.

Thus the number density of photons **per unit volume**, $n(\omega)$, is given using the Bose-Einstein distribution to be ($\mu = 0$):

$$n(\omega) \;=\; \frac{g(\omega)}{V} \frac{1}{e^{\beta E} \;-\; 1} \; ,$$

which, using eqn. (6.1.3.8) and $E = \hbar\omega$, becomes

$$\boxed{n(\omega) \;=\; \frac{\omega^2}{\pi^2 c^3} \frac{1}{e^{\beta\hbar\omega} - 1} \qquad \textbf{Black-body radiation distribution}. \qquad (6.1.3.9)}$$

(i) The **energy density** (per unit volume), $\epsilon(\omega)$, is given by

$$\begin{aligned} \epsilon(\omega) \;&=\; \hbar\omega \, n(\omega) \\ &=\; \frac{\hbar\omega^3}{\pi^2 c^3} \frac{1}{e^{\beta\hbar\omega} - 1} \end{aligned} \tag{6.1.3.10}$$

This is the **Planck radiation law**.

(ii) The total energy per unit volume is then

$$\begin{aligned} E(T) \;&=\; \int_0^\infty \epsilon(\omega) d\omega \\ &=\; \frac{1}{\hbar^3 c^3 \pi^2} \int_0^\infty (\hbar\omega)^3 \frac{1}{e^{\beta\hbar\omega} - 1} d(\hbar\omega) \; . \end{aligned} \tag{6.1.3.11}$$

Setting $x = \beta\hbar\omega$ we get

$$E(T) \;=\; \frac{1}{\beta^4} \frac{1}{\hbar^3 c^3 \pi^2} \int_0^\infty x^3 \frac{1}{e^x - 1} dx \; .$$

Hence, (we look up the integral, it is $\pi^4/15$)

$$E(T) \;=\; CT^4 \;, \qquad C \;=\; \frac{\pi^2 k}{15}\left(\frac{k}{\hbar c}\right)^3 \;. \tag{6.1.3.12}$$

(iii) The energy flux, $\mathcal{E}(T)$, radiated from a black-body is defined as the energy per second leaving a small hole of unit area in the wall of the box, assuming no inward flux. The speed of the photons is c, and the number in interval $(\omega, \omega + d\omega)$ crossing unit area per second is

$$df(\omega) \;=\; \frac{c}{4} n(\omega)\, d\omega \;,$$

where the factor 1/4 comes from an angular integration over velocity directions. Thus,

$$\mathcal{E}(T) \;=\; \frac{1}{4} E(T) c \;,$$

or

$$\boxed{\mathcal{E}(T) \;=\; \sigma T^4 \;, \qquad \sigma \;=\; \frac{\pi^2 kc}{60}\left(\frac{k}{\hbar c}\right)^3 \qquad \textbf{Stefan-Boltzmann law}. \qquad (6.1.3.13)}$$

Here $\sigma = 5.67 \cdot 10^{-8} J s^{-1} m^{-2} K^{-4}$ is Stefan's constant.

6.4 Atomic Transitions

The atoms in the black-body have many quantum states. In particular consider a pair of states, m and n, with energies E_m and E_n, respectively, with $\hbar\omega \;=\; E_n - E_m$. There are three kinds of quantum transitions between these levels:

(i) **Spontaneous emission.** An atom in state n may spontaneously decay to state m emitting a photon of frequency ω. Let A_{nm} be the probability for this to happen per second.

(ii) **Absorption.** An atom in state m may absorb a quantum of radiation of frequency ω and jump to state n. The rate is proportional to the **energy density**, $\epsilon(\omega)$, of radiation present. Let this rate be $B_{mn}\,\epsilon(\omega)$.

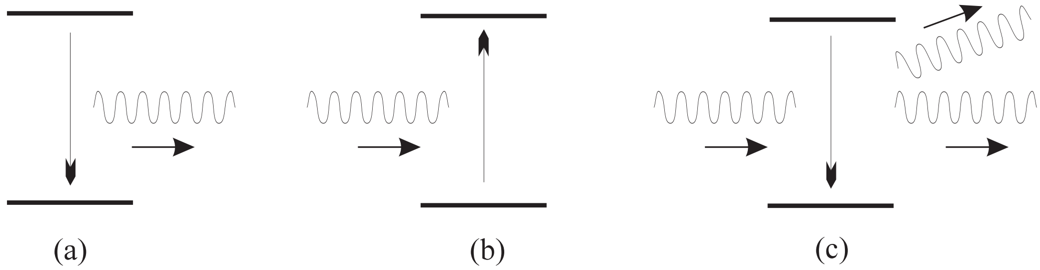
(iii) **Stimulated emission.** The presence of radiation **induces** an atom in state n to drop to state m and emit a quantum of frequency ω. This is unlikely unless the stimulating radiation has frequency very close to ω. The rate is proportional to the energy density and we write it as $B_{nm}\,\epsilon(\omega)$.

In equilibrium a steady state obtains: the rate of transitions from $m \to n$ balances the rate from $n \to m$. If N_m and N_n are the respective equilibrium populations of the two levels then we must have

$$N_m \cdot B_{mn}\,\epsilon(\omega) \;=\; N_n \cdot [A_{nm} \;+\; B_{nm}\,\epsilon(\omega)] \;.$$

$$\Longrightarrow$$

$$\epsilon(\omega) \;=\; \frac{A_{nm}}{(N_m/N_n) \cdot B_{mn} \;-\; B_{nm}} \tag{6.4.1}$$



The Boltzmann distribution for the atomic levels gives

$$\frac{N_m}{N_n} = e^{-\beta(E_m - E_n)} = e^{\beta\hbar\omega} .$$

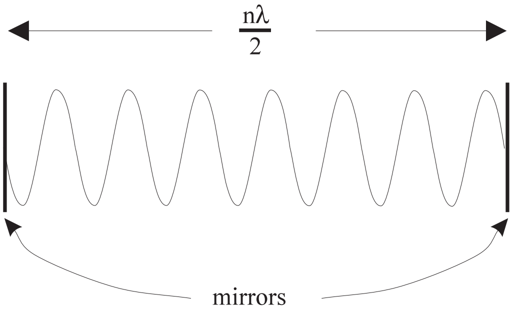
These last two equations must be consistent with the known black-body expression for $\epsilon(\omega)$ in eqn. (6.1.3.10). Comparing the two alternative expressions for $\epsilon(\omega)$ we find

$$\boxed{\left.\begin{aligned} B_{nm} &= B_{mn} \\ A_{nm} &= (\hbar\omega^3/\pi^2 c^3) \cdot B_{nm} \end{aligned}\right\} \quad \textbf{Einstein A and B coefficients.} \qquad (6.4.2)}$$

This allows the rate for spontaneous emission to be calculated from the rate for stimulated emission. The latter is much easier to calculate – spontaneous emission calculations require the full apparatus of quantizing the electric field.

7 Lasers

We describe the Helium-Neon laser as an example and paradigm. This laser consists of a mixture of helium and neon in a long tube with parallel mirrors at each end forming a tuned cavity which is designed to sustain a **standing wave** of light of a certain wavelength λ :



It is a bit like a fluorescent tube in construction.

(i) As in a fluorescent tube, a discharge is set up which consists of accelerating electrons between high voltage electrodes at opposite ends of the cavity. The electrons collide with the helium atoms which absorb energy in the collision by being excited to levels above their ground state.

(ii) The helium atoms in turn collide with the neon atoms giving up energy in the collision and exciting the neon atoms to a **meta-stable** excited state. This is a state for which the **spontaneous** emission rate is small. This is the **upper** laser state, n .

(iii) The **lower** laser state, m, ($E_m < E_n$), is a short-lived state. It decays through **spontaneous** emission and thus will always depopulate very quickly.

Thus

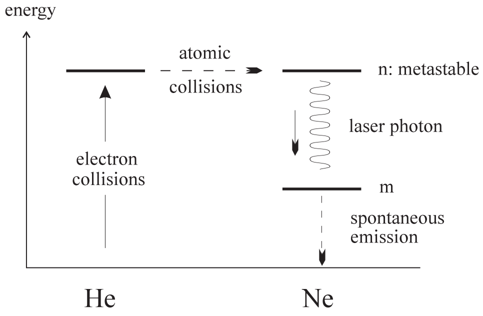
(a) N_n is large because n is meta-stable;

(b) N_m is small because m decays quickly.

Hence $N_n/N_m \gg 1$. Note that the usual equilibrium conditions give

$$\frac{N_n}{N_m} \;=\; e^{(E_n-E_m)/kT} \;\ll\; 1\ .$$

The population is **inverted** – like negative temperature.



Suppose there is a lot of energy in the cavity field. This is tuned by mirrors to have frequency $\omega = 2\pi/\lambda = (E_n - E_m)/\hbar$.

(i) If the energy density is high enough then **stimulated** emission from $n \rightarrow m$ will occur at a much higher rate than incoherent spontaneous emission.

(ii) The photon emitted has energy $\hbar\omega$ and is in **phase** with the stimulating field being sustained by the cavity. This latter fact is not easy to show but is plausible since energy has been released into the cavity field which therefore must increase in intensity – the emitted photon must hence reinforce the field.

(iii) The cavity field becomes **very** intense under these circumstances and it is **coherent** which means that all the photons constituting it are in step. A definite phase can be assigned to the electric standing wave:

$$\boldsymbol{E}(z,t) \;=\; \boldsymbol{E}_0 \cos\omega t \sin\left(\frac{\omega}{c}\right) z\ . \tag{7.1}$$

(iv) If one of the mirrors is not quite 100% reflecting then a few % of the radiation can be transmitted and used.

Why must the lower level m depopulate quickly? Consider the Einstein relation relating absorption to stimulated emission eqns. (6.4.2). These occur at the **same** rate, namely $\epsilon(\omega)B$. Hence, if the lower level becomes appreciably populated it will make the **inverse** transition to n, **absorbing** radiation from the cavity which will compete strongly with the wanted stimulated emission. This will only not be a problem if m depopulates quickly by decaying fast to even lower levels by other emissions etc.

The major properties of such laser light are that it is

(a) very intense – can produce high energy densities;

(b) totally coherent – can be used for holography;

(c) highly collimated – the divergence of the beam is close to the theoretical minimum value.

For general laser action require:

(1) Two active levels of the medium, n and m , $E_n > E_m$. The laser transition is $n \rightarrow m$, $\hbar\omega = E_n - E_m$.

(2) A **pumping mechanism** which will create and sustain a population inversion $N_n > N_m$.

(3) A tuned cavity which will sustain standing waves of frequency ω parallel to the tube axis.

Then

(1) Pumping $\longrightarrow$ population inversion;

(2) Stimulated emission $n \rightarrow m$ caused by standing wave. Emitted photons are in phase with standing wave thus reinforcing it if emitted sufficiently parallel to tube axis, otherwise they are lost in the walls;

(3) Useful radiation is allowed to pass out through a mirror which is not 100% reflecting;

(4) Fluctuations and losses are caused by fluctuations in mirrors, spontaneous emissions, losses in walls and leakage of radiation for use.

The laser is a new state of radiation in the presence of matter. An ordinary flash tube or light bulb is **not** a laser but an ordinary light source. Such devices produce only incoherent light dominated by **spontaneous** emission from excited states. Each atom emits independently of any other and consequently the photons are not phase-locked and are incoherent and uncollimated. A laser below the lasing threshold is the same.

7.1 A model for the laser

Let the number of photons in the cavity field be r . Then

$$\dot{r} \;=\; \text{gain} \;-\; \text{loss}\ . \tag{7.1.1}$$

The gain is from stimulated emission and the loss is due to mainly to the finite lifetime of the photons in the cavity field. We write $N = N_n - N_m$ and then have the reasonable parametrization:

$$\text{gain} \;=\; GNr\ , \qquad \text{loss} \;=\; \chi\, r\ . \tag{7.1.2}$$

Suppose that **without** laser action the pumping mechanism would maintain the population inversion at $N = N_0$. Then **with** laser action we must have

$$N \;=\; N_0 - \alpha r\ . \tag{7.1.3}$$

The decrease in N is due to stimulated emission which is proportional to r . Then

$$\begin{aligned} & \dot{r} &=& \quad G(N_0 - \alpha r)r - \chi\, r \\ \text{or} \qquad & \dot{r} &=& \quad -(\chi - N_0 G)r - \alpha G r^2 \\ & &\equiv& \quad -ar - br^2\ . \end{aligned} \tag{7.1.4}$$

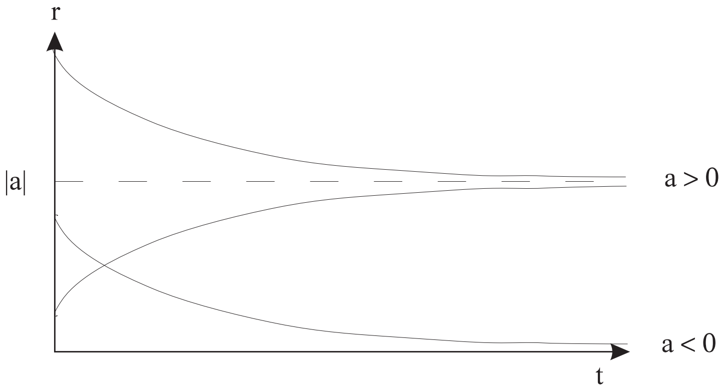
The steady state is $\dot{r} = 0$:

$a \;>\; 0$	$t \to \infty$	$r \to 0$	no laser action
$a \;<\; 0$	$t \to \infty$	$r \to \dfrac{\lvert a\rvert}{b}$	laser action

Consider $b = 1$ for simplicity. The solution to the eqn. (7.1.4) is ($r(0) = r_0$)

$$\begin{aligned} a > 0 \qquad r \;&=\; \frac{aCe^{-at}}{1 - Ce^{-at}}\ , \qquad C \;=\; \frac{r_0}{a + r_0} \\ a < 0 \qquad r \;&=\; \frac{\lvert a\rvert}{1 + Ce^{-\lvert a\rvert t}}\ , \qquad C \;=\; \frac{\lvert a\rvert - r_0}{r_0} \end{aligned} \tag{7.1.5}$$

This looks like:



The two cases are $N_0 < \chi/G$, $N_0 > \chi/G$. The pumping strength given by N_0 precipitates the **phase transition** to the laser phase as it passes through $N_c = \chi/G$.

Because the photon has a finite lifetime, $t_0 = 1/\chi$, there is an uncertainty in the frequency, $\delta\omega$, of the observed radiation. This is given by

$$\delta\omega \;\approx\; \frac{2\pi}{t_0} \;=\; 2\pi\,\chi\ . \tag{7.1.6}$$

7.2 Laser statistics

Another way of investigating the laser is to apply the same "most probable" counting method that we used to study quantum statistics. Consider the following system

(i) A cavity which sustains **one** particular frequency mode. This is like picking just one mode out of all the possible modes in a box. This will be occupied by photons with momentum parallel to the tube axis and each with energy $\hbar\omega$. This means there is just **one** state for the photons.

(ii) A gas of N atoms ($N \gg 1$). We are only interested in the two laser states n and m, separated by energy $E_n - E_m = \hbar\omega$. These atoms are treated as distinquishable particles: they each have their own pair of laser states, so there is no way in which one atom can be excited to the upper state of another and so no confusion about indistinquishability.

(iii) The system contains energy $E = R\hbar\omega$ above its ground state, where R is an integer.

Let the energy be distributed with r photons in the cavity mode and $(R-r)$ atoms in the upper state, n. The number of ways of achieving this division is

$$P(r) \;=\; \frac{N!}{(N-R+r)!(R-r)!}\ . \tag{7.2.1}$$

Maximise $P(r)$ w.r.t r to find the most probable configuration.

$$\begin{aligned}
\log P \;&\approx\; N\log N \;-\; (N-R+r)\log(N-R+r) \;-\; (R-r)\log(R-r)\ ,\\
\frac{d}{dr}\log P\bigg|_{r=r_0} \;&=\; 0 \qquad \Rightarrow \qquad (N-R+r) \;=\; (R-r)\ ,\\
\text{or} \qquad r_0 \;&=\; R-\frac{N}{2}\ .
\end{aligned}\tag{7.2.2}$$

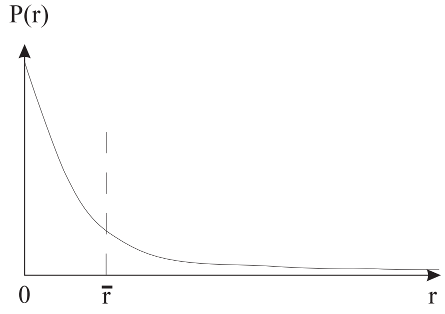
7.2.1 $R \ll N/2$

Since must have $r \geq 0$, the **most likely** configuration is $r = 0$, i.e., **no** photons in the cavity at all. Consider now the fluctuations in r . What is the probability of finding r photons in the cavity?

$$\log P(r) \;=\; \log P(0) + r\frac{d}{dr}\log P(r)\bigg|_0 \;+\; \ldots .$$

Now, from eqn. (7.2.1)

$$\begin{aligned}
\frac{d}{dr}\log P(r) \;&=\; \log\left(\frac{R-r}{N-R+r}\right)\\
\Longrightarrow \qquad \log P(r) \;&=\; \log P(0) \;+\; r\log\left(\frac{R}{N-R}\right) \;+\; \ldots\\
\Longrightarrow \qquad P(r) \;&=\; P(0)\left(\frac{R}{N-R}\right)^r\ .
\end{aligned}\tag{7.1.2.1}$$



Let $p = R/(N-R)$. Then the average value, $\overline{r}$, for r is given by

$$\overline{r} \;=\; \frac{\sum_0^\infty \; rp^r}{\sum_0^\infty \; p^r} \;,$$

giving

$$\overline{r} \;=\; \frac{p}{1-p} \;. \tag{7.1.2.2}$$

How do we interpret this result? If we **define** β' by setting $p = \exp(-\beta'\hbar\omega)$ we find

$$\overline{r} \;=\; \frac{1}{e^{\beta'\hbar\omega}-1} \;, \tag{7.1.2.3}$$

i.e., the Bose-Einstein distribution if we interpret $\beta' = 1/kT'$. We have really re-derived the BE distribution in this special case. The cavity mode exchanges energy with the large **heat-bath** consisting of the N gas atoms and we get the expected functional form for the mean BE occupation number. This is an example of what happens when a subsystem (here just the cavity mode and its photons) is brought into contact with (i.e., can exchange energy with) a reservoir of energy (called a heat bath). The temperature is then T' .

Now, β' is clearly a function of R and hence of the energy, E, of the whole system. T' increases with E as we should expect. We can go further, however. Suppose, now, that the gas itself is allowed to exchange energy with the walls which we will assume are at temperature T. Then the atoms occupy their levels with a probability given by the **Boltzmann** distribution. In particular

$$\frac{\text{number in level } n}{\text{number in level } m} \;=\; \frac{R}{N-R} \;=\; e^{-\hbar\omega/kT} \;. \tag{7.1.2.4}$$

However, from above

$$p \;=\; \frac{R}{N-R} \;=\; e^{-\hbar\omega/kT'} \;,$$

and hence $T' = T$ and our interpretation of T' as a temperature is consistent.

The real message is that systems in equilibrium are characterized by a few special functional forms for the important quantities and, in particular, by the values of the parameters parametrizing those functions, such as temperature and chemical potential. The power of the method is that the description works so well for a wide range of systems.

The discussion of this section shows that the system is not a laser since it is characterised by the BE distribution and there is not much energy in the cavity mode. This state is characteristic of an ordinary light source. Can it be different?

7.2.2 $R > N/2$

Let $R = N/2 + k$. From above the **most likely** configuration is

$$r_0 = R - N/2 = k \ .$$

Thus there are now k photons in the coherent cavity mode giving laser action. Here $R > N - R$ which corresponds to having so much energy in the system that there must be a population inversion. Consider the fluctuations again. Let $r = k + \Delta k$. Then

$$P(k+\Delta k) \ = \ \frac{N!}{(N/2+\Delta k)!(N/2-\Delta k)!} \ . \qquad (7.2.2.1)$$

Then, using Stirling's formula

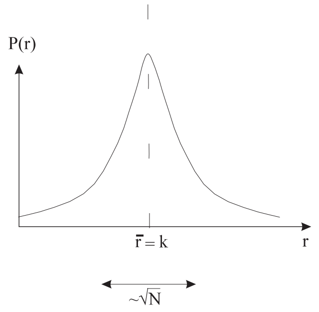
$$\begin{aligned} &\log P \ \approx \\ &N\log N \ - \ \left(\frac{N}{2}+\Delta k\right)\log\left(\frac{N}{2}+\Delta k\right) \ - \ \left(\frac{N}{2}-\Delta k\right)\log\left(\frac{N}{2}-\Delta k\right) \ . \end{aligned} \qquad (7.2.2.2)$$

Let $P_0 = P(k)$ and from eqn. (7.2.2.2) we find $P_0 \approx 2^N$. For small Δk we get

$$\begin{aligned} \log P \ &\approx \ N\log N \ - \ \left(\frac{N}{2}+\Delta k\right)\left(\log\frac{N}{2}+\frac{2\Delta k}{N}\right) \ - \ \left(\frac{N}{2}-\Delta k\right)\left(\log\frac{N}{2}-\frac{2\Delta k}{N}\right) \\ &= \ \log P_0 \ - \ \frac{2(\Delta k)^2}{N} \ . \end{aligned} \qquad (7.2.2.3)$$

Or

$$P(k+\Delta k) \ \approx \ P_0\, e^{-2(\Delta k)^2/N} \ . \qquad (7.2.2.4)$$



Hence, the fluctuations obey a **normal** distribution with variance $\langle(\Delta k)^2\rangle = N/8$, and so the fluctations are proportional to $\sqrt{N}$.

For example, consider $k = a\sqrt{N}$ with $a \gg 1$. Then $\overline{r} = k \gg 1$ and

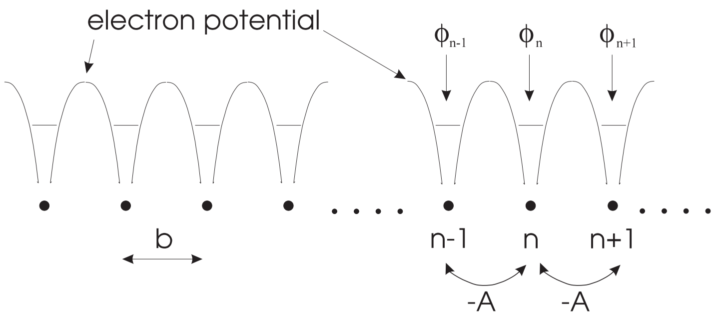
$$\begin{aligned} \frac{\overline{r}}{N} \ &= \ \frac{a}{\sqrt{N}} \qquad \longrightarrow 0 \qquad \text{as} \qquad N \longrightarrow \infty \ , \\ \frac{\sqrt{\langle r-\overline{r}\rangle^2}}{\overline{r}} \ &\sim \ \frac{1}{a} \ \ll \ 1 \ . \end{aligned} \qquad (7.2.2.5)$$

Hence, the cavity field intensity, $\overline{r}$, is very large with very small fluctuations. To establish $R > N/2$ we require a population inversion, and so laser action sets in if the pumping mechanism is strong enough to generate a reasonable population inversion $R - N/2 \sim a\sqrt{N}$ at least. This agrees with the simple model without fluctuations discussed earlier.

Note here that we have produced a **different** distribution from BE. There is consequently no concept of "temperature" for the system in this state: the lasing state. It is simply a new state of the system. We have a large number, N, of atoms as is usual in statistical physics but the amount of energy in the system is huge compared with systems we have discussed before: $E \ll N\hbar\omega$ gives usual statistics as $N \to \infty$, but $E \sim N\hbar\omega$ gives new statistics as $N \to \infty$.

8 Band Structure

Consider a 1D array of N atoms and concentrate on the **same** atomic level in each atom. Consider an electron in that level for a given atom. How is it that, although this electron is bound to the site, there is a probability that given enough time it will be found near a far-away site? What happens is that the electron tunnels through the barrier between the atoms and can thus migrate arbitrarily far from the original site. A simple picture is one where tunnelling to nearest-neighbour sites dominates all other processes:



The lattice spacing is b . The electron on the n-th atom is represented by the state ϕ_n . The Hamiltonian is $H = H_0 + V$ where

$$\begin{aligned} H_0\phi &= E_0\phi \ , \\ (\phi_n, V\phi_n) = \alpha \ , &\qquad (\phi_n, V\phi_{n\pm1}) = -A \ , \\ (\phi_n, V\phi_m) &= 0 \ , \quad m \neq n, n\pm 1 \ . \end{aligned} \tag{8.1}$$

So H_0 is the hamiltonian for the atom in isolation and ϕ_n is the wavefunction for the **chosen** level on the n-th atom. Note that

$$\phi_n(x) = \phi_0(x - nb) \ , \tag{8.2}$$

i.e., all wavefunctions are related by translation of the origin – it is the **same** level in each atom. We will also make the approximation that $(\phi_n, \phi_m) = \delta_{mn}$. From eqn. (8.2) this is not strictly true but is correct up to exponentially small corrections.

The potential V is due soley to the presence of the neighbouring atoms and the matrix element $-A$ is the "tunnelling amplitude" between neighbouring sites. We want to solve the full problem:

$$H\Phi \;=\; E\Phi \;, \tag{8.3}$$

and as usual we write

$$\Phi \;=\; \sum_p c_p\phi_p \;, \tag{8.4}$$

using the $\{\phi_n\}$ as a complete basis. Then using (8.3)

$$\begin{aligned} H\sum_p c_n\phi_p \;&=\; (H_0+V)\textstyle\sum_p c_p\phi_p \;=\; E\displaystyle\sum_p c_p\phi_p \\ \Rightarrow \qquad \sum_p c_p(E_0+V)\phi_p \;&=\; \sum_p c_pE\phi_p \;. \end{aligned} \tag{8.5}$$

Take the inner product of both sides with ϕ_n and use eqns. (8.1):

$$Ec_n \;=\; \underbrace{(E_0+\alpha)}_{E_0'}c_n \;-\; A(c_{n-1}+c_{n+1}) \;. \tag{8.6}$$

[This kind of equation occurs in the classical dynamics of a set of equally spaced particles in 1D interconnected by springs. The solution to that problem is that waves can travel freely up and down the chain. As a classical model of a 1D crystal such waves are simply sound waves.]

We need to solve this difference equation for the $\{c_n\}$ and so try the usual substitution:

$$\begin{aligned} c_n \;&=\; e^{iknb} \;, \qquad \text{normal mode solution} \\ \Rightarrow \qquad c_{n\pm1} \;&=\; e^{\pm ikb}c_n \;, \end{aligned} \tag{8.7}$$

where we remember that b is the spacing and that the n-th atom is situated at $x_n = nb$. Substituting is eqn. (8.6) we get

$$\boxed{E \;\equiv\; E(k) \;=\; E_0' \;-\; 2A\cos kb \;.} \tag{8.8}$$

The N atomic levels which were originally **degenerate** at E_0 are replaced by N new levels which are spread out in energy according to eqn. (8.8). The old state labelling, n, was by atomic position. The new labelling is by the **wavevector**, k, which contains no reference to position at all: for **each** value of k there is an eigenstate with energy $E(k)$.

To get the counting of states right we must think about boundary conditions at the ends of the chain of atoms. We impose periodicity, i.e., we assume the atoms form a ring. Then we have

$$e^{ikNb} \;=\; 1 \;. \tag{8.9}$$

What happens at the ends cannot affect the bulk properties (c.f., the boundary conditions in a box that we used). Define $L = Nb$, the chain length. Taking N even, this implies that

$$k \;=\; 2\pi p/L \;, \qquad p \;\text{ an integer}: \quad -\frac{N}{2} \;<\; p \;\le\; \frac{N}{2} \;. \tag{8.10}$$

So k is discrete, taking one of N values which is consistent with starting with having N states in the first place.

The important features so far are

(i) the eigenstates are no longer degenerate in energy but are labelled by a wavevector k and have energy $E(k)$ which lies in a **band** of allowed energies determined by eqn. (8.8) called the dispersion relation. The band has width $4A$. This structure will occur for **each** atomic level. For the lower, more tightly bound, levels A is **small** and the bands are **narrow**. The higher bands are **wider** and may even overlap in energy;

(ii) the electron is no longer **localized** on a given site but has a non-zero probability of being found anywhere, i.e., they are **extended**:

$$P_k(x) \;\propto\; |\Phi_k(x)|^2 \;=\; |\sum_n e^{iknb}\phi_n(x)\,|^2 \;\neq\; 0 \qquad \forall\, x\ .$$

The number of states per band is then $2N$ (2 for electron spin) ⇒ band is **filled** by $2N$ electrons (unless bands overlap). What is the density of states ($N \to \infty$):

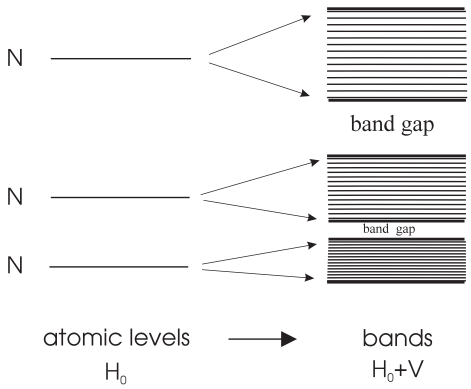
$$\frac{dp}{dk} \;=\; \frac{L}{2\pi} \;\Rightarrow\; \frac{dp}{dE} \;=\; \frac{L}{2\pi}\left|\frac{dk}{dE}\right|\ . \qquad (8.11)$$

The level density per unit length, $g(E)$, is

$$g(E) \;=\; 2\frac{2}{L}\frac{dp}{dE} \;=\; \frac{2}{\pi}\left|\frac{dk}{dE}\right|\ . \qquad (8.12)$$

Factors of 2 for $\pm p$ and 2 for spin. This is a **general** result in 1D and is not special to this model.

Schematically, the physical outcome of what we have solved is shown in the figure.



8.1 Bloch Waves

The stationary state is thus

$$\Phi_k(x) \;=\; C\sum_p e^{ikpb}\phi_p(x)\ . \qquad (8.1.1)$$

Using eqn. (8.2), consider $x \to x + b$:

$$\begin{aligned}\Phi_k(x+b) &= C\sum_p e^{ikpb}\phi_p(x+b) \;=\; C\sum_p e^{ikpb}\phi_{p-1}(x)\ ,\\ &= e^{ikb}\,\Phi_k(x)\ . \qquad (8.1.2)\end{aligned}$$

Notice we have labelled the state by k, the wave vector, and the momentum is $\hbar k$. Thus we can always write

$$\Phi_k(x) \;=\; u_k(x)\, e^{ikx} \;, \tag{8.1.3}$$

where the function $u_k(x)$ is **defined** by this relation and satisfies

$$u_k(x+b) \;=\; u_k(x) \;. \tag{8.1.4}$$

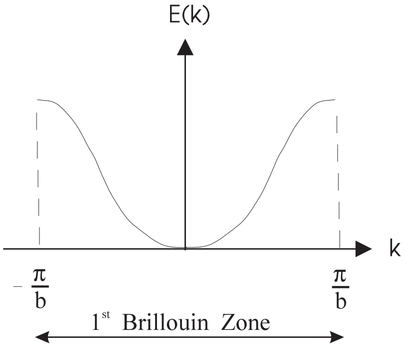
I.e., $u_k(x)$ is **periodic** on the lattice. Of course, the explicit expression for u_k can be deduced from eqn. (8.1.1) but it is the **form** of the solution that I am emphasizing. With time-dependence the stationary state is

$$\boxed{\Psi(x,t) \;=\; u_k(x)\, e^{i(kx-\omega t)} \qquad \textbf{Bloch Wave Solution.}} \tag{8.1.5}$$

I have defined $\omega = E(k)/\hbar$. This solution can be generalized to 3D and represents the basic form of the stationary state for all problems involving periodic potentials and especially crystal lattices. It shows that the energy eigenstates are **travelling waves** which are **extended**. In particular, the electron can now carry charge from one end to the other in the form of an electric current.

8.2 Brillouin Zones

Clearly, the range of k is $-\pi/b < k \leq \pi/b$. This is called the **first Brillouin zone** in k space. The n-th Brillouin zone corresponds to $k_n = k + 2n\pi/b$. A graph of the dispersion relation (energy E versus k) is shown for the first Brillouin zone in the figure.



9 Insulators and Conductors

From the fermi distribution we learn that $E_F \equiv \mu$ for a free gas of N electrons is determined by:

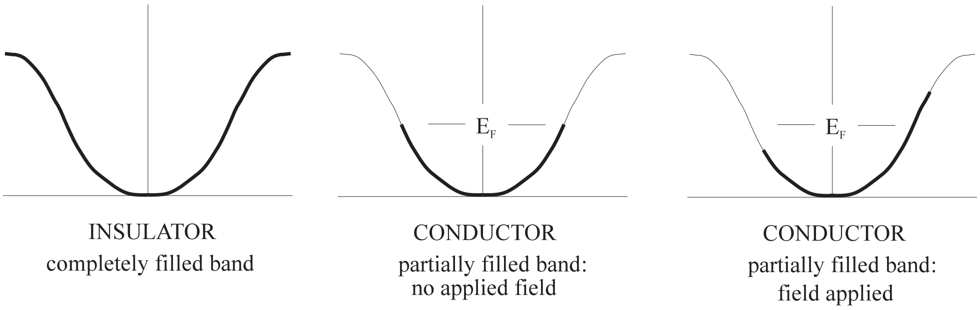
$$\frac{N}{V} \;=\; \int_0^\infty \frac{g(E)dE}{e^{\beta(E-E_F)}+1} \;, \tag{9.1}$$

where $g(E)$ is the density of states per unit volume. As $T \to 0$ all levels with $E < E_F$ are occupied and those with $E > E_F$ are unoccupied. At $T = 0$ we find (see Q12 on example sheet 1)

$$E_F \;=\; \frac{\hbar^2}{2m}\left(\frac{N}{V}\right)^{2/3} \;\approx\; 4\cdot 10^{-19}\text{eV m}^2\left(\frac{N}{V}\right)^{2/3} \quad . \tag{9.2}$$

In many materials such as metals, a typical value for N/V is 10^{28} $\text{m}^{-3} \Rightarrow E_F \approx 2\,\text{eV}$. At room temperature $T \approx 300\ {}^0K$, $kT \approx 3\cdot 10^{-2}\,\text{eV} \Rightarrow kT \ll E_F$ and the fermi distribution in this case differs **very little** from that at $T = 0$, i.e., there is still a sharp decrease at $E \sim E_F$. Thus, at relevant temperatures all levels with $E < E_F$ are filled and those with $E > E_F$ are empty. Only for $E_F - kT < E < E + kT$ is this not true but $kT \ll E_F$.

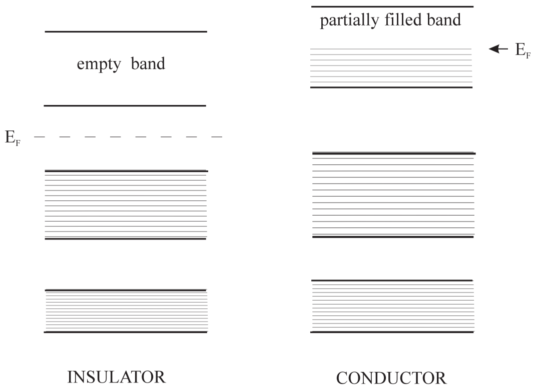
Consider filling up the energy levels of an atom with electrons using the Exclusion Principle. The set of levels with a **given** energy is called a **shell** and when it is full it is known as a **closed shell**: the next electron must go into a state of **next** higher energy. From the discussion of band structure each shell corresponds to an energy band in the crystal made from the atoms: the N copies of a given shell combine because of tunnelling into a band of states with an energy spread labelled by k . A **closed shell** will give rise to a band which is also **completely filled** with electrons. This idea of a **closed shell** giving rise to completely filled bands allows insulators and conductors to be distinquished from each other.



(i) **Insulators** contain atoms in which the electrons occupy **closed shells**. The atoms in an insulator have just the right number of electrons to fill all the shells with no electrons left over. The corresponding bands are completely filled.

(ii) **Conductors** contain atoms in which the last (i.e., highest energy) occupied shell is **not** closed. The lower energy shells are closed but there are not enough electrons to completely fill the highest of the occupied shells. Correspondingly, the low-lying bands are **full** but the **top** band is only partially filled.

Why is this? When there is **no** applied electric field the electrons move equally in all directions and there is no nett transport of charge. When an electric field is applied it must accelerate some electrons and deccelerate others to generate a flow asymmetry. This means that electrons must vacate some states to occupy other, originally vacant, states (see figure). However, if the band is **full** there can be **no** such asymmetry under any conditions since all states are occupied at all times. Such a system is in the **same** state whether or not an electric field is applied and so there can never be nett charge



transport: the system is an **insulator**. Conversely, if at least one band is only partially filled there are vacant states nearby in energy to which an an electron can move and the state of the system changes when a field is applied (see figure): the system is a **conductor**. In the next section this argument is considered in more detail.

9.1 Conduction

9.1.1 group velocity

When a particle has plane–wave stationary states it has equal probability of being found anywhere. To localize it we must superpose the waves. This is wave–particle duality and forms one way of looking at the Uncertainty Principle. Consider superposing two plane waves with momenta $\hbar(k \pm \Delta k)$ with energy dispersion $E(k) = \hbar\omega(k)$. Let $\Delta\omega = \frac{d\omega}{dk}\Delta k$ and then the wavefunction is:

$$\psi(x,t) \;=\; \sqrt{\frac{1}{2V}}\,\left(e^{(i(k+\Delta k)x - i(\omega+\Delta\omega)t)} \;+\; e^{(i(k-\Delta k)x - i(\omega-\Delta\omega)t)}\right) \;. \tag{9.1.1.1}$$

The **observable** quantity is $|\psi(x,t)|^2$ which is clearly given by

$$|\psi(x,t)|^2 \;=\; \frac{1}{V}\,\left(1 \;+\; \cos\left(\Delta k\, x - \Delta\omega\, t\right)\right) \;. \tag{9.1.1.2}$$

The velocity of this wave–packet is given by following points of constant $|\psi(x,t)|^2$ as t changes. From above this means $\Delta k\, x - \Delta\omega\, t \;=\;$ constant, i.e., keep the phase constant. Thus the velocity is

$$v \;=\; \frac{dx}{dt} \;=\; \frac{\Delta\omega}{\Delta k} \;=\; \frac{d\omega(k)}{dk} \;. \tag{9.1.1.3}$$

Thus

$$\boxed{v \;=\; \frac{d\omega(k)}{dk}\;, \qquad \textbf{the group velocity.}} \tag{9.1.1.4}$$

The **group velocity** is the **observable** velocity of particles. This analysis applies equally well to signals carried by waves, e.g., radio waves and waves in waveguides.

Remember the gaussian wavepacket for free particles where $\omega(k) = \hbar k^2/2m$. The group velocity is $v = d\omega/dk = \hbar k/m$ where $\hbar k$ is the momentum. This is the **correct** classical result, namely "velocity = momentum/mass ". Equation (9.1.1.4) is very important generally and we shall use it below.

9.1.2 Dynamics

In a band, the group velocity of the electron in state labelled by wavenumber k is

$$v \;=\; \frac{d\omega(k)}{dk} \;=\; \frac{1}{\hbar}\frac{dE(k)}{dk}\;. \tag{9.2.1.1}$$

This is the velocity at which the charge carried by the electron is transported.

Consider an applied electric field $\mathcal{E}$. The force on the electron of charge q is $F = q\mathcal{E}$, and so the change in energy of the electron is

$$dE \;=\; F dx \;=\; q\mathcal{E} v\, dt\;. \tag{9.2.1.2}$$

Substituting the group velocity for v, we get

$$dE \;=\; \frac{q}{\hbar}\mathcal{E}\,\frac{dE(k)}{dk}\,dt\;. \tag{9.2.1.3}$$

We can reinterpret this result in terms of a time-dependent k :

$$dE \;=\; \frac{dE(k)}{dk}\,dk \;\;\Rightarrow\;\; \boxed{\frac{dk}{dt} \;=\; \frac{q}{\hbar}\mathcal{E}\;.} \tag{9.2.1.4}$$

This is what we expect **classically** since the momentum is $\hbar k$ and

$$\frac{d\ \text{momentum}}{dt} \;=\; \text{applied force}\;.$$

We see, however, that since k changes with time the electron **must** be continually making the transition to states of more positive k . From above

$$\dot{v} \;=\; \frac{1}{\hbar}\frac{d}{dt}\frac{dE}{dk} \;=\; \frac{1}{\hbar}\left(\dot{k}\frac{d}{dk}\right)\frac{dE}{dk} \;=\; \frac{1}{\hbar}\dot{k}\,\frac{d^2E}{dk^2}. \tag{9.2.1.5}$$

Substituting for $\dot{k}$ from eqn. 9.2.1.4

$$\dot{v} \;=\; \frac{q}{\hbar^2}\frac{d^2E}{dk^2}\,\mathcal{E}\;. \tag{9.2.1.6}$$

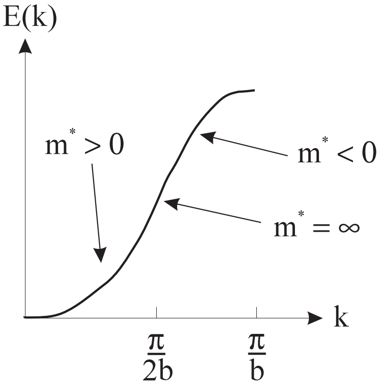
Comparing this with Newton's law which would be of the form $m\dot{v} = q\mathcal{E}$, we see that we can write this result as

$$\boxed{\begin{aligned} m^*\dot{v} \;&=\; q\mathcal{E}\;, \\ \text{with} \\ m^* \;&=\; \hbar^2\left(\frac{d^2E}{dk^2}\right)^{-1}\;, \qquad \textbf{the effective mass.}\end{aligned}} \tag{9.2.1.7}$$

Thus we can treat many of the subsequent calculations as classical as long as we use an **effective mass**, m^*, for the electron. The major features are

(i) If the band is **narrow** d^2E/dk^2 is small (small A above), and m^* is large: the electron is hard to move.

(ii) For **any** band:



(a) near the bottom $d^2E/dk^2 > 0 \;\Rightarrow\; m^* > 0$ there;

(b) near the top $d^2E/dk^2 < 0 \;\Rightarrow\; m^* < 0$ there;

(c) at the point of **inflexion**, $k = \pm\pi/2b$, $m^* = \infty$.

(d) at the top $k = \pm\pi/b$ and at the bottom, $k = 0$, $dE/dk = 0$.

(iii) Under an applied field the velocity of an electron starting from $k = 0,\ v = 0$ **increases** to its maximum at $k = \pi/b$, and then it **decreases** again reaching $v = 0$ at the zone boundary. By periodicity, k **re-enters** the zone at $k = -\pi/b$. In general, **symmetry** requires $v = dE/dk = 0$ at the top and bottom of the band.

In a perfect crystal there is **no** steady state for electrons in a constant applied electric field. They accelerate through the Brillouin zone and their velocity oscillates thus giving an oscillating charge flow. However, this effect can only be seen in very large and perfect crystals at low temperature. The usual state of affairs is different because of impurities and imperfections in the crystal lattice.

Consider for the moment the case of the perfect crystal. When the field $\mathcal{E}$ is applied the distribution of electrons per unit length, $n(k,t)$, is time-dependent and not the equilibrium distribution $n_0(k)$ appropriate to $\mathcal{E} = 0$. We now use conservation of particle number to get information about $n(k,t)$. This tells us that

$$n(k,t) \;=\; n(k+dk, t+dt) \; . \qquad (9.2.1.8)$$

The particles that have wavevector $k + dk$ at time $t + dt$ are the **same ones** that had wavevector k at time t: they just moved from k to $k + dk$ (in $k-$space) in time dt. Then we have

$$\left(\frac{\partial n(k,t)}{\partial t}\right)_{\text{fields}} \;+\; \dot{k}\,\frac{\partial n(k,t)}{\partial k} \;=\; 0 \; . \qquad (9.2.1.9)$$

[This is just the 1D version of the usual conservation equation

$$\frac{\partial n(k,t)}{\partial t} \;+\; \nabla_k \cdot \mathbf{j}(k,t) = 0 \; ,$$

with $\mathbf{j}(\mathbf{k},\mathbf{t}) = n(k,t)\dot{k}$. Note, also, that $\dot{k}$ is constant.]

Then, using $\dot{k} = q\mathcal{E}/\hbar$,

$$\left(\frac{\partial n(k,t)}{\partial t}\right)_{\text{fields}} \;=\; -\,\frac{q\mathcal{E}}{\hbar}\,\frac{\partial n(k,t)}{\partial k} \; . \qquad (9.2.1.10)$$

This is the effect on $n(k,t)$ of the applied fields **alone**.

Suppose scattering off imperfections causes the electron to jump randomly between states. This means that its velocity (or state) **after** such a scattering is **independent** of the original velocity (or state). This introduces a "friction" term into the dynamics. Assume τ is the time between collisions and that after a collision the electron velocity is random. Then, by definition,

$$\left(\frac{\partial n(k,t)}{\partial t}\right)_{\text{collisions}} \;=\; -\frac{1}{\tau}(n(k,t)-n_0(k)) \;. \tag{9.2.1.11}$$

Here, τ could be k-dependent but we ignore this from now on. Note, that in absence of fields this means that $n(k,t) \to n_0(k)$ as $t \to \infty$, i.e., the systems equilibriates on characteristic time scale τ. The steady state is defined by a t-independent density $n(k)$, and so must have

$$\frac{dn}{dt} \;=\; \left(\frac{\partial n}{\partial t}\right)_{\text{collisions}} + \left(\frac{\partial n}{\partial t}\right)_{\text{fields}} \;=\; 0 \;, \tag{9.2.1.12}$$

for the steady state distribution $n(k)$. Thus

$$n(k) \;=\; n_0 \;+\; \tau\left(\frac{\partial n}{\partial t}\right)_{\text{fields}} \;=\; n_0 \;-\; \frac{q\mathcal{E}\tau}{\hbar}\frac{dn}{dk} \;. \tag{9.2.1.13}$$

The electric current I is defined by

$$I \;=\; q\int_{-\pi/b}^{\pi/b} n(k)v_k\, dk \;=\; \frac{q}{\hbar}\int_{-\pi/b}^{\pi/b} n(k)\frac{dE}{dk}\, dk \;, \tag{9.2.1.14}$$

where we have used eqn. (9.2.1.1), i.e., that $\hbar v_k = dE/dk$. Using the expression for $n(k)$ we get

$$I \;=\; -\frac{q^2\mathcal{E}\tau}{\hbar^2}\int_{-\pi/b}^{\pi/b} \frac{dn(k)}{dk}\frac{dE}{dk}\, dk \;. \tag{9.2.1.15}$$

Note, the first term $n_0(k)$ gives zero contribution to I by symmetry (it gives the current for $\mathcal{E} = 0$, which is zero). Because $\mathcal{E}$ is, in fact, a weak field we approximate

$$n(k) \;\approx\; n_0(k) \;=\; \frac{1}{\pi}f(k) \;,$$

where $f(k)$ is the Fermi-Dirac distribution. Remember, $n(k)$ is number density per unit length and the $1/\pi$ is the number states per unit length in $(k, k+dk)$. Then

$$I \;=\; -2\frac{q^2\mathcal{E}\tau}{\pi\hbar^2}\int_0^{\pi/b} \frac{df(k)}{dk}\frac{dE}{dk}\, dk \;. \tag{9.2.1.16}$$

The conductivity is defined by Ohm's law which says $V = IR$. In 1D for a conductor of length L the conductivity is $\sigma = L/R$. Also $\mathcal{E} = V/L$ is the *definition* of $\mathcal{E}$. Hence, Ohm's law becomes $\sigma\mathcal{E} = I$, and we can read off σ from (9.2.1.16) to get

$$\sigma \;=\; -\frac{2q^2\tau}{\pi\hbar^2}\int_0^{\infty} dk\,\frac{df(k)}{dk}\frac{dE(k)}{dk} \;, \tag{9.2.1.17}$$

But $f(k)$ varies only very slowly except at $k = k_F$ where it approximates a step function, and so

$$\begin{aligned} \frac{df}{dk} \;&=\; -\,\delta(k-k_F) \\ \Rightarrow \qquad & \\ \sigma \;&\approx\; \frac{2q^2\tau}{\pi\hbar^2}\left.\frac{dE}{dk}\right|_{k_F} \;. \end{aligned} \tag{9.2.1.18}$$

Clearly $\sigma = 0$ if the band is full since the upper limit is then $k = \pi/b$ (E vanishes for higher k), and $dE/dk = 0$ there.

Note that the last part of this discussion applies strictly only at $T = 0\ {}^0K$. It explains why metals conduct at all temperatures and why insulators do not conduct at low$-T$. Non-zero conductivity in semiconductors (see later) and insulators depends strongly on temperature.

We **define**

$$n_{eff} \;=\; \frac{2m}{\pi\hbar^2}\,\frac{dE}{dk}\bigg|_{k_F} \;, \qquad (9.2.1.19)$$

where m is the **usual** electron mass. For **free** particles for which $E = \hbar^2 k/2m$, n_{eff} is the **actual** number density. Then

$$\sigma \;=\; \frac{n_{eff}\, q^2 \tau}{m} \;. \qquad (9.2.1.20)$$

This way of writing σ in terms of **effective** parameters is useful and will occur again later.

9.1.3 Impurities

Lattice imperfections are due to

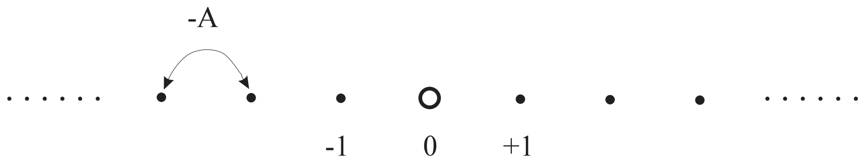
(i) impurity atoms in the lattice;

(ii) defects. Atoms may be missing at some sites or out of place;

(iii) thermal motion of atoms from their equilibrium positions: this destroys the exact periodicity of the crystal.

Items (ii) and (iii) become less important as $T \to 0$.

We now study what happens when an impurity atom replaces one of the regular atoms in the lattice. We redo the band structure calculation but substitute the impurity atom for the usual one at $n = 0$.



Let the wavefunction for the electron bound to the **impurity** atom be $\chi(x)$ with $H_0\chi(x) = (E_0 + F)\chi(x)$. Remember, for $n \neq 0$ we have $H_0\phi(x) = E_0\phi(x)$. Then we write the eigenstate of the **full** Hamiltonian, $H_0 + V$ as

$$\Phi \;=\; \sum_{n\neq 0} a_n\phi_n \;+\; a_0\chi \;, \qquad (9.3.1.1)$$

with

$$H\Phi \;=\; E\Phi \;. \qquad (9.3.1.2)$$

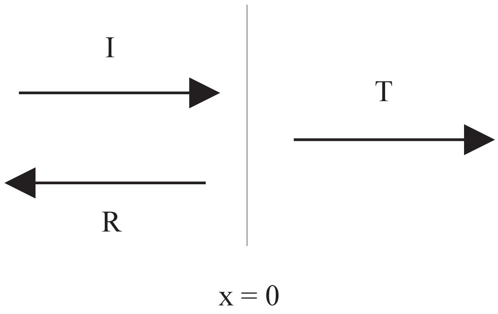
As before we have

$$\begin{aligned}
(\phi_n, V\phi_n) &= (\chi, V\chi) \;=\; \alpha \;, \\
(\phi_n, V\phi_{n\pm1}) &= (\chi, V\phi_{\pm1}) \;=\; -\,A \;, \\
(\phi_n, V\phi_m) &= 0 \;, \qquad n \neq m, m\pm 1 \\
(\chi, \phi_n) &= 0 \qquad n \neq 0 \;.
\end{aligned} \tag{9.3.1.3}$$

All states are assumed orthonormal. Multiplying in turn both sides of eqn. (9.3.1.2) by ϕ_n^* and χ^* we find the coupled difference equations:

$$\begin{aligned}
Ea_n &= E_0'\,a_n \;-\; A\,(a_{n-1} + a_{n+1}) \qquad\qquad n \neq 0 \\
Ea_0 &= (E_0' + F)\,a_0 \;-\; A\,(a_{-1} + a_{+1}) \;,
\end{aligned} \tag{9.3.1.4}$$

with $E_0' = E_0 + \alpha$. Far to the left and right of $n = 0$ we have the usual travelling wave solutions which means that the energy dispersion formula, eqn (8.8) is **unchanged**. However, the impurity at $n = 0$ will reflect part of an incident wave, transmitting the rest. It is the same phenomenon as in quantum mechanics where waves are reflected and transmitted by a barrier, or as in waves in a string where a massive particle fixed to the string also causes reflection of waves.



To see this in more detail consider a steady-state of the form:

$$a_n \;=\; \begin{cases} e^{iknb} \;+\; \beta\, e^{-iknb} & n < 0 \\ \gamma\, e^{iknb} & n > 0 \end{cases} \tag{9.3.1.5}$$

This solves the equations which do **not** include a_0 . From these equations we find, as before, that $E \equiv E(k) = E_0' - 2A\cos kb$: the energy dispersion relation is the same as before. The three equations which include a_0 are

$$\begin{aligned}
&\boldsymbol{n = -1:} && (E - E_0')\,(e^{-ikb} + \beta\, e^{ikb}) = -\,A\,(e^{-2ikb} + \beta\, e^{2ikb} + a_0) \\
&\boldsymbol{n = 0:} && (E - E_0' - F)\,a_0 = -\,A\,(e^{-ikb} + (\beta + \gamma)\, e^{ikb}) \\
&\boldsymbol{n = 1:} && (E - E_0')\,\gamma\, e^{ikb} = -\,A\,(a_0 + \gamma\, e^{2ikb}) \;.
\end{aligned} \tag{9.3.1.6}$$

Substituting the expression for E in terms of k gives

$$\boldsymbol{n = -1:} \quad a_0 \;=\; 1 + \beta \qquad\qquad \boldsymbol{n = 1:} \quad a_0 \;=\; \gamma \;.$$

Hence $\gamma \;=\; 1+\beta$, and from $\boldsymbol{n=0}$ we find

$$\beta \;=\; \frac{-F}{F-2iA\sin kb}\;. \tag{9.3.1.7}$$

Note that there is **no** change in the energy but just a change in direction: there is a reflected wave and a transmitted wave.

There is another solution which corresponds to the **trapping** of the electron. This means that there is an eigenstate which is not extended but localized and represents the electron **bound** to the impurity. To see this look for a solution of eqns. (9.3.1.6) of the form

$$\begin{aligned} a_n &= c\,e^{\kappa nb} && n<0 \\ a_n &= c\,e^{-\kappa nb} && n>0\;. \end{aligned}$$

If a solution exists with $\kappa > 0$ then the electron is localized at $n=0$. Substituting into the equations **not** including a_0, we find similar algebra as before and get the energy dispersion:

$$E(\kappa) \;=\; E_0' \;-\; 2A\cosh\kappa b\;. \tag{9.3.1.8}$$

The equations for a_1 and a_{-1} are also satisfied if $a_0=c$, and that for a_0 determines the value of κ . This equation gives

$$Ec \;=\; (E_0'+F)c \;-\; Ac\,(e^{-\kappa b}+e^{-\kappa b})\;,$$

and, on substituting for E from eqn. (9.3.1.8), we find

$$-2A\sinh\kappa b \;=\; F\;. \tag{9.3.1.9}$$

Note also that since $\kappa > 0$ we must have that F and A have **opposite** signs for a localized state to appear.

If we write $z=e^{\kappa b}$ we can write eqn. (9.3.1.9) as a quadratic in z :

$$\begin{aligned} & z^2+\frac{F}{A}z \;-\; 1 \;=\; 0\;, \\ \Rightarrow \quad & \\ & z=\frac{1}{2}\left(-\frac{F}{A} \;\pm\; \sqrt{4+\frac{F^2}{A^2}}\right)\;. \end{aligned}$$

Since clearly $z>0$, we must choose the “+” sign above. Using this result for κ and substituting into eqn. (9.3.1.8) we get

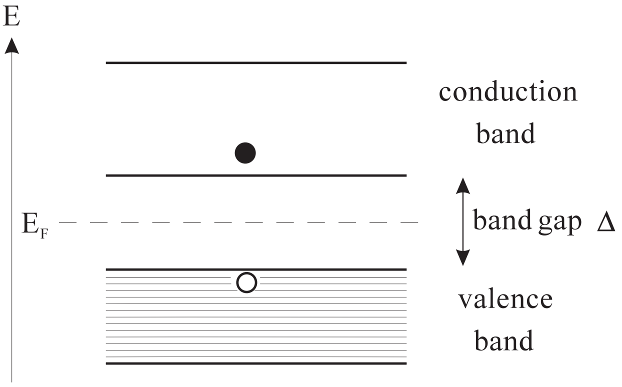
$$E \;=\; E_0' \;-\; A\sqrt{4+\frac{F^2}{A^2}}\;. \tag{9.3.1.10}$$

The result is that a localized state can exist with energy given by eqn. (9.3.1.10) which lies in the band (i.e., “forbidden”) gap either **above**, (sign(A) $=-1$), or **below**, (sign(A) $=+1$), the band in question. This result is very important for the physics of semiconductors.

9.2 Semiconductors

A semiconductor has filled bands and so is an insulator at $T = 0$. However, the band-gap, Δ, between the top-most filled band, called the **valence** band, and the next empty one, called the **conduction** band, is small: $\Delta \leq 1$eV. At room temperatures ($\sim 300\,^0K$) electrons, total number density denoted by "n", are thermally excited from the valence band to states in the conduction band. Conversely, these excited electrons leave behind vacant states in the valence band which are called "holes", total number density denoted by "p". Clearly, $n = p$. Semiconductors for which $n = p$ are called **intrinsic** semiconductors. The values of n and p are **strongly** temperature dependent. Conduction is due to both the electrons in the conduction band and the holes in the valence band. To explain this fact we now study the dynamics of holes. The most useful semiconductor is silicon but germanium was originally used in the early days.

9.2.1 Holes



The figure shows a semiconductor at non-zero temperature. The fermi energy E_F lies near the middle of the band gap. We will verify this fact later. The vacated site will always be near the top of the filled band. Why? The probability that a state is occupied is $f(E)$, the Fermi-Dirac distribution. Hence the number of **unoccupied** states in $(E, E + dE)$ is $p(E)$:

$$p(E)dE \;=\; g(E)(1 \;-\; f(E))dE \;, \qquad (9.1.2.1)$$

where $g(E)$ is the density of states. But $f(E)$ approaches unity exponentially rapidly as E decreases below E_F, and hence only near the top of the valence band is $p(E)$ different from zero.

Now, in the absence of a field the current in a filled band is zero. I.e.,

$$\boldsymbol{I} \;=\; q\sum_i \boldsymbol{v}_i \;=\; q\left(\boldsymbol{v}_j \;+\; \sum_{i\neq j}\boldsymbol{v}_i\right) \;=\; 0\;. \qquad (9.1.2.2)$$

If just "j" were missing then the current would be

$$\boldsymbol{I}' \;=\; q\sum_{i\neq j}\boldsymbol{v}_i \;=\; -\,q\boldsymbol{v}_j\;. \qquad (9.1.2.3)$$

Thus the vacancy or "hole" behaves like a charge $-q$ of velocity $\boldsymbol{v}_j$. Apply an external field $\boldsymbol{\mathcal{E}}$, then

$$\frac{d\boldsymbol{I}'}{dt} = -q\frac{d\boldsymbol{v}_j}{dt} = -\frac{q^2}{m^*}\boldsymbol{\mathcal{E}} \ . \qquad (9.1.2.4)$$

(Remember $\dot{\boldsymbol{v}} = q\boldsymbol{\mathcal{E}}/m^*$.) However, the missing electron comes from the **top** of the band where $m^* < 0$, and so

$$\frac{d\boldsymbol{I}'}{dt} = \frac{q^2}{(-m^*)}\boldsymbol{\mathcal{E}} \ . \qquad (9.1.2.5)$$

This is similar to an electron of charge q with mass $|m^*|$ near the **bottom** of a band. We see that a band in which an electron is **missing** from the top behaves dynamically like a single carrier of charge $-q$ and mass $|m^*|$. I.e., the **absence** of a particle of charge q and mass m^* behaves in most details like a carrier of charge $-q$ and mass $-m^*$. Since $m^* < 0$ the carrier has positive mass. Such a carrier is treated as a particle and is called a "**hole**".

We must include holes in our conductivity calculation. Since σ only depends on the charge squared, carriers of both charges $\pm q$ contribute positively to σ . The conductivity analysis is repeated for holes which have a characteristic relaxation time τ_p . Denoting the relaxation time for electrons as τ_n we can write

$$\begin{aligned} \boldsymbol{I} &= \left(\frac{nq^2\tau_n}{m_n^*} + \frac{pq^2\tau_p}{m_p^*}\right)\boldsymbol{\mathcal{E}} \\ \Rightarrow \qquad \sigma &= q^2\left(\frac{n\tau_n}{m_n^*} + \frac{p\tau_p}{m_p^*}\right) \ . \qquad (9.1.2.6) \end{aligned}$$

Where

(i) n is the number density of charge $+q$ carriers (electrons) in the **conduction** band;

(ii) p is the number density of charge $-q$ carriers (holes) in the **valence** band;

(iii) it is assumed that $m_n^*, m_p^*, \tau_n, \tau_p$ are constant for the states of interest: those occupied in the conduction band and those vacated in the valence band.

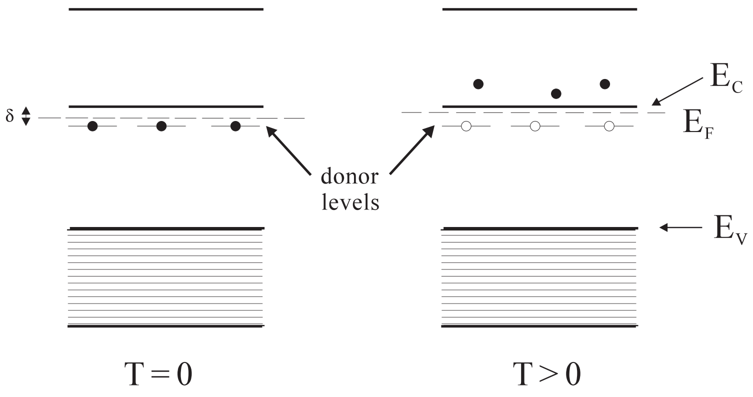
Clearly, σ will be strongly dependent on T unlike the case of metals.

9.2.2 Doping

The carrier density can be increased by impregnating the semiconductor with impurities. For $T \neq 0$ it is possible to enhance the number of electrons (n) in the conduction band **or** the number of holes (p) in the valence band. From section (9.1.3) we saw that an impurity atom creates a state in the band gap. An electron in such a state is **localized** and does **not** contribute to the conductivity. Conversely, if the state is occupied by a hole (i.e., it is empty!) the hole is **localized** and it does not contribute to the conductivity either. Bearing these ideas in mind we now discuss the two important cases of electron carrier and hole carrier enhancement respectively.

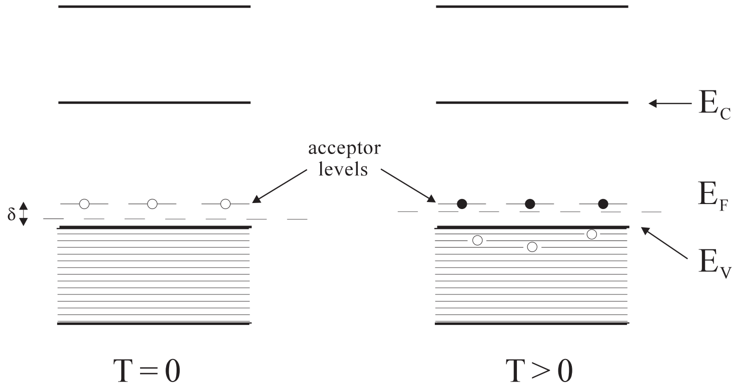
Silicon atoms have four outermost electrons which exactly fill the valence band at $T = 0$: these are the valence electrons which give the band its name. There are two kinds of important impurities which are introduced:

(i) **Donor impurites.** The impurity in this case has five valence electrons, i.e., one **more** than silicon. An example is phosphorous. Each impurity atom gives rise to a localized state just **below** the bottom of the conduction band. At $T = 0$ this state is occupied by the extra, fifth, electron:



Notice that the Fermi energy moves up towards the top of the gap. This is because **more** electrons are present. The localized states are very close to the bottom of the conduction band, a gap of only δ, and for $T > 0$ most of the electrons trapped in these states at $T = 0$ are thermally excited to states in the conduction band. All that is needed is that $kT \sim \delta$. Typically $\delta = 0.04\,\mathrm{eV}$ and at room temperature $kT \sim 0.025\,\mathrm{eV}$. The number of states available in the conduction band is very much reater than those due to the impurities and so it is overwhelmingly likely that the electron leaves the localized impurity state and occupies an extended state in the conduction band.

(ii) **Acceptor impurities.** The impurity in this case has three valence electrons, i.e., one **less** than silicon. An example is boron. Each impurity atom gives rise to a localized state just **above** the top of the valence band. At $T = 0$ this state is unoccupied since there is one electron lacking compared with the state of pure silicon:



In this case the Fermi energy moves down towards the bottom of the gap since there are states in this region which are unoccupied, thus reducing the energy at which occupation is 50% probable. For $T > 0$, as above, most of these localized empty states (i.e., "occupied" by holes) become filled with electrons thermally excited from the top part of the valence band. This can be thought of as holes localized in the impurity states being thermally excited to occupy hole states (states associated with **positive** effective mass) in the top of the valence band. As before this is overwhelmingly likley since there are very many more extended hole states in the valence band than localized impurity hole states.

The outcome is that

(i) silicon doped with **donor** impurities has, for $T > 0$, very many electrons in the conduction band and so the conductivity increases because of n–type carriers. Silicon doped in this way is called n-type silicon.

(ii) silicon doped with **acceptor** impurities has, for $T > 0$, very many holes in the valence band and so the conductivity increases because of p–type carriers. Silicon doped in this way is called p-type silicon.

Notation: in either type silicon the number density of electron carriers is denoted n, and the number density of hole carriers is denoted p. In intrinsic silicon $n = p \equiv n_i$. We shall now show that at a given temperature $np \;=\; n_i^2$. A consequence is that in n–type $n \gg p$ and in p–type $p \gg n$.

The number density of electrons in the conduction band is

$$n \;=\; \int_{E_c}^{\infty} \frac{g(E)}{e^{(E-E_F)/kT}+1}\, dE \; . \tag{9.2.2.1}$$

In this integral the range of E is such that $E - E_F \gg kT$ gives the major contribution and so we can approximate the Fermi distribution by the Boltzmann distribution:

$$n \;=\; \int_{E_c}^{\infty} g(E) e^{-(E-E_F)/kT}\, dE \; . \tag{9.2.2.2}$$

The density of states per unit volume, $g(E)$, depends on $E - E_c$ (i.e., the distance from the bottom of the band). In practice the electrons behave very much like free particles with an effective mass and we can use the form derived much earlier for a free gas, eqn. (6.1.1.5):

$$g(E) \;\sim\; A(E-E_c)^{\frac{1}{2}} \; . \tag{9.2.2.3}$$

This is the form appropriate for three dimensions. Then we can rearrange the expression above to be

$$n \;=\; e^{-(E_c-E_F)/kT} \int_{E_c}^{\infty} A(E-E_c)^{\frac{1}{2}}\, e^{-(E-E_c)/kT} dE \; . \tag{9.2.2.4}$$

Change variables in the integral to $x = (E - E_c)/kT$ and find

$$n \;=\; BT^{\frac{3}{2}}\, e^{-(E_c-E_F)/kT} \; , \tag{9.2.2.5}$$

with

$$B \;=\; \frac{A}{2}(k^3\pi)^{\frac{1}{2}} \; .$$

Similarly the number density for holes is

$$p \;=\; \int_{-\infty}^{E_v} g'(E)\, h(E)\, dE \;,$$

where

$$h(E) \;=\; 1 \;-\; \frac{1}{e^{(E-E_F)/kT}+1} \;\approx\; e^{(E-E_F)/kT} \;. \qquad (9.2.2.6)$$

Writing $g'(E) = A'(E_v - E)^{\frac{1}{2}}$ we find in a similar manner

$$p \;=\; B'T^{\frac{3}{2}}\, e^{-(E_F-E_v)/kT} \;. \qquad (9.2.2.7)$$

E_F is determined by the values of n and p.

(i) **Intrinsic semiconductor:** $n = p$ gives

$$E_F \;=\; \frac{E_c \;+\; E_v}{2} \;+\; kT \log \frac{B'}{B} \;. \qquad (9.2.2.8)$$

A result of this kind holds for a general density of states. In this case E_F is in the **middle** of the band gap up to a generally small correction which reflects the difference between the properties of the two bands.

(ii) **Non-intrinsic semiconductor:** To get $n \gg p$ or $p \gg n$ the position of E_F changes accordingly. If $n \gg p$ then from eqn. (9.2.2.6) we see $E_F \sim E_c$ and E_F is located near the top of the band gap. Similarly from eqn. (9.2.2.7), when $p \gg n$ get $E_F \sim E_v$ and E_F is near the bottom of the gap. In either case from eqns. (9.2.2.6, 9.2.2.7) we get

$$np \;=\; BB'T^3\, e^{-(E_c-E_v)/kT} \;\equiv\; CT^3\, e^{-\Delta/kT} \;, \qquad (9.2.2.9)$$

where Δ is the band gap and C is a material-dependent constant. This is a general result. So np is independent of E_F and hence of the impurity concentrations. Thus if n_i is the electron density in intrinsic semiconductor at temperature T then

$$np \;=\; n_i^2 \;. \qquad (9.2.2.10)$$

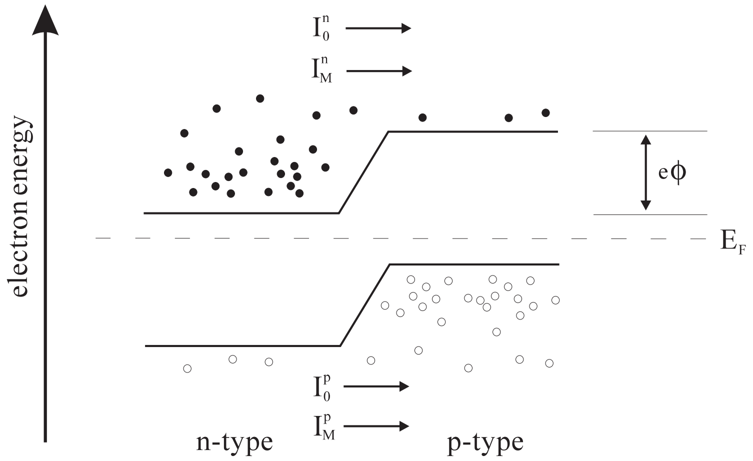
(Remember, $n_i = p_i$.)

> In n–type the electrons are the **majority** carriers and the holes are the **minority** carriers.
> In p–type the holes are the **majority** carriers and the electrons are the **minority** carriers.

10 Semiconductor Devices

10.1 The *pn* Junction

(i) n and p type materials are brought into contact and a **contact** potential ϕ is built up across the junction. This general happens when two dissimilar materials are brought into close contact.



(ii) The contact potential occurs because n–type has an excess of electrons and p–type an excess of holes. Electrons flow across the junction from n to p and holes from p to n. The n–type becomes **positive** and the p–type **negative**. Eventually, the charges build up enough for it to become energetically unfavourable for any more electrons or holes to change sides: the electrons are impeded by the −ve charge on the p–type and the holes by the +ve charge on the n–type.

(iii) At equilibrium the electrons crossing from n– to p–type balance those crossing from p– to n–type and similarly for the holes. The condition for equilibrium is that the Fermi energies on **both** sides of the junction are the **same**. This is a very general condition that can be justified for all systems. Remember, two systems with different temperatures brought into contact exchange energy until there is one **common** temperature. If they have two different Fermi energies (or chemical potentials) they will exchange **particles** until there is one **common** Fermi energy. Both these conditions result from the idea of particle distributions being most-probable.

[Notation: the direction of **all** currents is from left to right, i.e., from n– to p–type. The sign of the current determines the direction in which charge actually flows.]

(iv) The current due to n crossing from n–type to p–type is called the **majority** n current.
The current due to n crossing from p–type to n–type is called the **minority** n current.

Likewise

The current due to p crossing from p–type to n–type is called the **majority** p current.
The current due to p crossing from n–type to p–type is called the **minority** p current.

In other words, in equilibrium the barrier inhibits **majority** carrier flow of electrons from n– to p–type and holes from p– to n–type. In contrast, the **minority** carriers in each material are pulled over junction as fast as they arrive there by the electric field due to the potential ϕ . Equilibrium occurs when the **majority** current is balanced by the **minority** current.

(v) Concentrate on just the electrons.

(a) **All minority** n in the p–type which reach the junction ($x = 0$) are attracted across by the junction field into the n–type. The rate is independent of ϕ: it depends only on how fast they get to $x = 0$ in the first place. This gives the minority n current I_0^n across the junction.

(b) The **majority** n approaching the junction from within the n–type must have enough energy to surmount the energy barrier of height $q\phi$. Those that get across give rise to the **majority** current I_M due to electrons crossing from n– to p–type. Calculate I_M by reference to the figure:

$$I_M^n \;=\; qC \int_{E_c+q\phi}^{\infty} dE \; g(E)\, e^{-(E-E_F)/kT}\, v(E) \;, \qquad (10.1.1)$$

where $g(E)$ is the density of states, $v(E)$ is the velocity and C is a constant which is hard to calculate in practice but which accounts for the fact that electrons impinge obliquely on the junction etc. In equilibrium $I_M^n + I_0^n = 0$.

(c) The junction region extends a distance L into the material on either side. (The magnitude of L depends very strongly on the doping strengths but typically $L \sim 10^{-6 \text{ to } -9}$ metres.) Both kinds of majority carrier occur in the junction which combine (i.e., the electrons occupy the hole states) giving a **depletion** of carriers and a consequent **decrease** in conductivity. This region is called the **depletion layer**.

(d) Suppose an external voltage V is applied across the junction, say by a battery contacting on the left and the right. Because the depletion layer has a high resistance the extra potential drop of V occurs entirely across this region and not across the bulk of either semiconductor. The effect is to change the barrier height from $q\phi$ to $q(\phi + V)$. What is the current $I(V)$ flowing through the device as a function of V?

I_0^n is independent of the barrier height and hence of V, but now

$$I_M^n(V) \;=\; qC \int_{E_c+q(\phi+V)}^{\infty} dE \; g(E)\, e^{-(E-E_F)/kT}\, v(E) \;. \qquad (10.1.2)$$

Since $g(E)$ and $v(E)$ vary only slowly with E compared with the exponential we have

$$I_M^n(V) \;\approx\; I_M^n(0)\, e^{-qV/kT} \;. \qquad (10.1.3)$$

But $I_M^n(0) + I_0^n = 0$ and so the **total** current due to electrons is

$$\begin{aligned} I^n(V) \;&=\; I_M^n(V) \;+\; I_0^n \\ \Rightarrow \qquad & \\ I^n(V) \;&=\; I_0^n \left(1 \;-\; e^{-qV/kT}\right) \;. \end{aligned} \qquad (10.1.4)$$

I_0^n is due to the minority of electrons in the p–type and is thus proportional to their density which we denote $\bar{n}_p$. Hence

$$I^n(V) \;=\; D^n \bar{n}_p \left(1 \;-\; e^{-qV/kT}\right) \;. \qquad (10.1.5)$$

(vi) The calculations for holes are along similar lines. The notation is the same as above with n replaced by p. Then

$$I_M^p(V) \;=\; -\,(-q)C\int_{-\infty}^{E_v-q(\phi+V)} dE\; g(E)\, e^{(E-E_F)/kT}\, v(E)\;, \tag{10.1.6}$$

where $g(E)$, $v(E)$ and C are appropriate for the valence band and the charge on the hole is $-e$. The extra $-$ sign is because the majority holes cross from right to left which by our convention carries a $-$ sign. Then, as before, we find

$$I_M^p(V) \;=\; I_M^p(0)\, e^{-qV/kT}\;. \tag{10.1.7}$$

The equilibrium condition is $I_M^p(0) + I_0^p = 0$ and so the hole contribution to the total current is

$$I^p \;=\; I_0^p\left(1 \;-\; e^{-qV/kT}\right)\;. \tag{10.1.8}$$

Since I_0^p is proportional to $\bar{p}_n$, the density of the minority p carriers in the n–type:

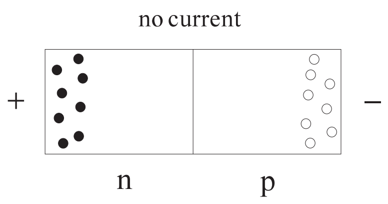
$$I^p \;=\; D^p\bar{p}_n\left(1 \;-\; e^{-qV/kT}\right)\;. \tag{10.1.9}$$

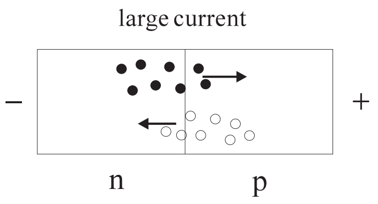
The total current is then

$$\boxed{I \;=\; \underbrace{(D^p\bar{p}_n \;+\; D^n\bar{n}_p)}_{I_0}\left(1 \;-\; e^{-qV/kT}\right)\;.} \tag{10.1.10}$$

Then

$$\begin{array}{llll} qV > 0 & \Rightarrow & I < I_0 & I \text{ is small,} \\ qV < 0 & \Rightarrow & I = -I_0\, e^{|qV|/kT} & |I| \text{ is large.} \end{array}$$





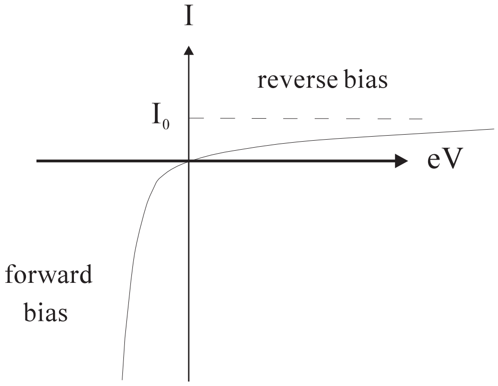
The np junction **rectifies**: it only passes current **one way**. Pictorially, this can be seen in the next figure. In the n–type only $-$ve charges carry the current and in the p–type the current is carried by only +ve charges. The two kinds of carrier swap rôles in the depletion layer. If

(i) a $-$ve voltage is applied to the p–type and hence a +ve voltage to the n–type both kinds of carrier are **attracted** to their respective terminals and **away** from the junction. **No** carrier of either charge **crosses** the junction and so **no** nett current flows;

(ii) a +ve voltage is applied to the p–type and hence a $-$ve voltage to the n–type both kinds of carrier are **repelled** by their respective terminals and **attracted** to the terminal on the other side. Both kinds of carrier flow across the junction and a large current is possible.

(iii) When the contact potential is reduced and a large current flows the junction is **forward biased** by the applied voltage. When the contact potential is increased and only the minority current flows occur the junction is **reverse biased** by the applied voltage.

We have calculated the form of the current characteristic, eqn (10.1.10) which looks like



Typically, $I_0 \sim 100 \mu a$ at $T = 300\,^0K$.

10.2 The *npn* Transistor

The transistor is a **current** amplifier. In operation the input is a small fluctuating current (e.g., from a microphone) and the output is a current proportional to the input but much larger. In practice the proportionality is not exact unless circuits are designed properly.

The *npn* transistor consists of a sandwich of a very thin p–type layer, of thickness W, between two pieces of n–type silicon.

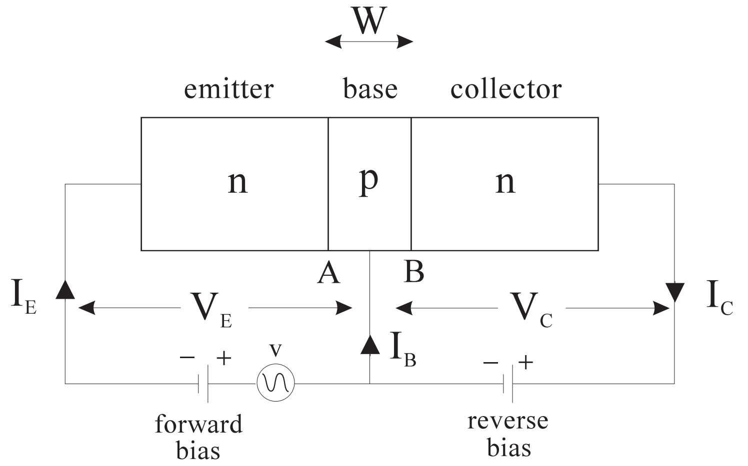
(i) The system can be though of as two back-to-back pn junctions, the first at A and the second at B.

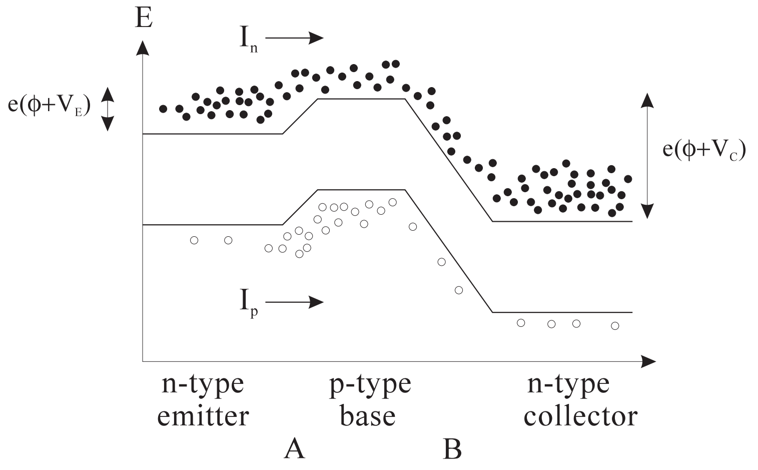
(ii) Two DC potentials, represented by the batteries, are applied which **forward bias** junction A and **reverse bias** junction B. The n–type of the forward biased junction is called the **emitter** and that of the reversed biased junction is called the **collector**. The p–type in the sandwich is the **base**.

The energy diagram with the various internal carrier currents is thus

(iii) Consider the junction A. Because it is forward biased I_E is large and is controlled by V_E. Across A there is an electron current I_n and a hole current I_p.

- (a) The electron current is due to electrons being injected or **emitted** into the base where they are **minority** carriers.





(b) The hole current is due to majority holes moving across A from p– to n–type.

(c) The electrons emitted into the base diffuse across the base but also combine with **majority** holes. However, if the base is **thin** then **all** the electrons diffuse to B before recombination can occur. At B the junction is biased in their favour and they are pulled across into the collector. Note: in the base they are masquerading as minority carriers which are helped across B by the contact potential enhanced by the reverse bias. They then generate I_C the external collector current.

(d) Things are different for the holes responsible for I_p. Where do they come from? Either they must come from the collector across B or are supplied by the base current I_B. In the collector holes are minority carriers so there can only be a very small minority hole current from collector to base – this can be ignored. Thus they must be due to I_B only. It will turn out that I_p and hence I_B is small and so, since $I_E = I_n + I_p$, we have $I_E \approx I_n$.

The outcome is that I_E is controlled by V_E according to the characteristic derived for the pn junction and that from (c) and (d) we expect $I_E \approx I_C$. V_E is modulated by the signal voltage $v(t)$ and so setting $V_E = V_E^0 + v(t)$:

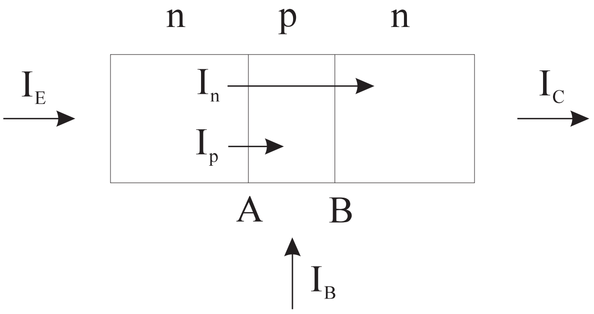
$$\begin{aligned} I_E \;&\approx\; I_C \;\approx\; -\,I_0\,e^{-q(V_E^0+v(t))/kT} \\ &\approx\; I_C^0\left(1 \;-\; \frac{qv(t)}{kT}\right) \;, \end{aligned} \tag{10.2.1}$$

where $I_C^0 = I_0 \exp(-qV_E^0/kT)$. At $T = 300\,{}^0K$ $kT/e \approx 0.025$ volts and so

$$I_E \;\approx\; I_C^0\;(1 \;+\; v(t)/0.025) \;, \qquad \text{note: } q \;=\; -\,e\;. \tag{10.2.2}$$

Clearly, a small signal voltage change causes a similar change in I_C. Small signals of a few milli-volts will cause large changes in I_C because of the small denominator. This is the amplification process which we shall examine more carefully below.

First we must look at the current balance more carefully to justify the statement that I_B is small, i.e., $I_B \ll I_E$. The flow diagram is



Then we have

$$I_E \;=\; I_n \;+\; I_p \;, \qquad I_C \;=\; I_n \;, \qquad I_B \;=\; -\,I_p\;. \tag{10.2.3}$$

Notation:

$\bar{n}_p$ ($\bar{p}_n$) is the equilibrium **minority** carrier density in p–type (n–type);
$\bar{p}_p$ ($\bar{n}_n$) is the equilibrium **majority** carrier density in p–type (n–type);

Consider the transistor **current gain** parameter β defined by $\beta \;=\; |I_C/I_B|$. From above

$$\beta \;=\; \left|\frac{I_n}{I_p}\right| \;=\; C\frac{\bar{n}_p}{\bar{p}_n}\;. \tag{10.2.4}$$

Here C is a constant and we have used the result from pn junction theory, eqns. (10.1.5, 10.1.9), that $I_n \propto \bar{n}_p$ and $I_p \propto \bar{p}_n$: the currents are controlled by the respective **minority** carrier densities. From eqn. (9.2.2.10) we have that

$$\bar{n}_p \;=\; \frac{C_p}{\bar{p}_p}\;, \qquad \bar{p}_n \;=\; \frac{C_n}{\bar{n}_n}\;, \tag{10.2.5}$$

and from eqn. (9.1.2.6) we see that the n– and p–type conductivities are proportional to the respective **majority** carrier densities:

$$\sigma_n \;\propto\; \bar{n}_n\;, \qquad \sigma_p \;\propto\; \bar{p}_p\;. \tag{10.2.6}$$

Thus

$$\beta \;=\; C\frac{\bar{n}_p}{\bar{p}_n} \;=\; C\frac{C_p\bar{n}_n}{C_n\bar{p}_p} \;=\; C'\frac{\sigma_n}{\sigma_p}\ . \tag{10.2.7}$$

The constant C' is built from many factors, some of which have been explicitly stated earlier. At fixed temperature these factors are mainly properties of silicon itself and we have exposed here the dependence on the important factors under experimental control. However, C' does depend on the width W of the base region and we just state here the result:

$$\beta \;=\; D\frac{L_R}{W}\frac{\sigma_n}{\sigma_p}\ , \tag{10.2.8}$$

Where L_R is a characteristic length in silicon and is the mean distance that an electron emitted into the base diffuses **before** it recombines with a majority hole.

In order to ensure $I_B \ll I_C$ the transistor is constructed so that β is large. This means

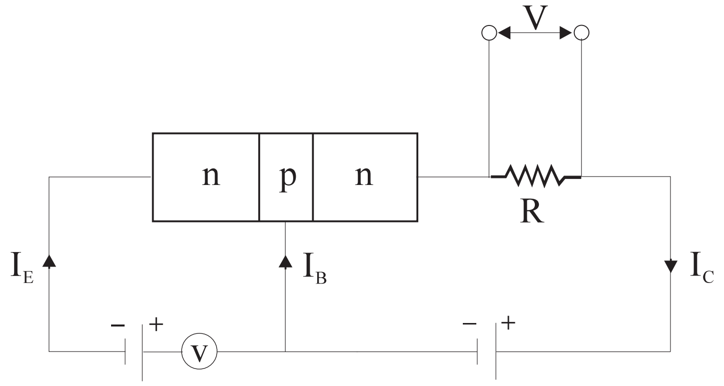
(i) $\sigma_p \ll \sigma_n$. This is achieved by using only lightly doped p–type for the base;

(ii) making W as small as possible.

Since β is a constant we can trivially write

$$\left|\frac{dI_C}{dI_B}\right| \;=\; \beta\ . \tag{10.2.9}$$

(i) A small fluctuation in I_B is replicated in I_C but amplified by β. The transistor is a **current** amplifier.



(ii) For **constant** V_E, I_C is **not** affected by changes in V_C. This is because B is reverse biased and from pn-junction theory the current is insensitive to the amount of reverse bias. Thus the emitter acts as a **constant** current source (c.f. a battery is a **constant** voltage source) controlled by V_E.

(iii) The effective resistance of the emitter–base diode is

$$R_E \;=\; \frac{dV_E}{d(-I_E)} \;\equiv\; \frac{dv}{d(-I_E)}\ , \quad \text{with} \quad I_E \;\approx\; Ce^{-qV_E/kT}\ . \tag{10.2.10}$$

(The “$-$” sign is because the positive current flows from the “+” terminal to the “$-$” terminal: the battery exposes the sign on the electron charge. I.e., $q < 0 \quad \Rightarrow \quad I_E < 0$.)

Thus

$$R_E \;=\; \frac{kT}{eI_E} \;=\; \frac{0.025}{|I_E|} \;=\; 25\Omega \;, \tag{10.2.11}$$

for $T = 300\,^0K$ and $I_E = 1$ milliamp (a typical value).

In the arrangement shown the **input** voltage is $v(t)$ and the **output** is $V(t)$. Then

$$\frac{dV}{dv} \;=\; \frac{dV}{d(-I_C)}\,\frac{d(-I_C)}{dv} \;. \tag{10.2.12}$$

But $I_C \approx I_E$ and so

$$\frac{d(-I_C)}{dv} \;\approx\; \frac{1}{R_E} \qquad \Rightarrow \qquad \frac{dV}{dv} \;=\; \frac{R}{R_E} \;\approx\; 40 \;, \tag{10.2.13}$$

for $R = 1K\Omega,\; I_C = 1\,\text{ma},\; T = 300\,^0K$. This is **voltage** amplification and the **voltage gain** is R/R_E .