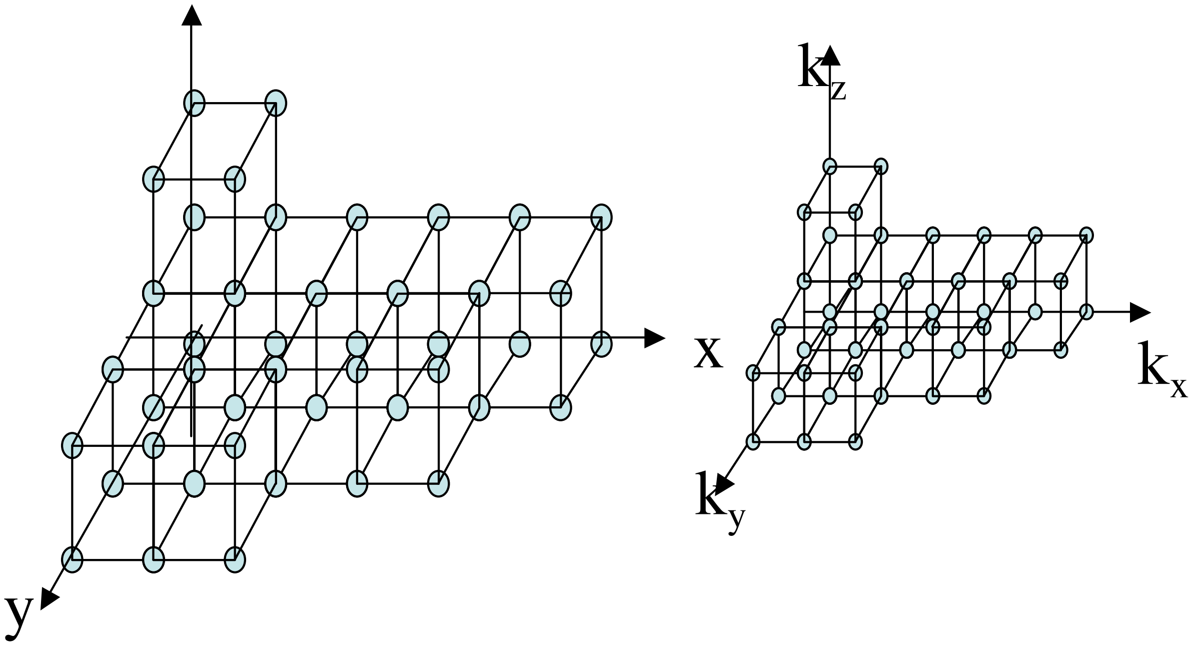
Appendix 1: Proofs and Derivations

This appendix includes major derivations omitted from the main text. It is by no means a complete set of derivations for all the relationships used. It is a selection that the authors feel is most illustrative of the underlying physical concepts as they relate to the methods described.

1. Derivation of the Density-of-States Expression.

We know from quantum mechanics that particle energies are not continuously variable in constrained systems. The famous example is that of atomic energy levels. Simply confining an electron in a “box” also creates discrete, allowed energy states. Energy and momentum are related through some “dispersion relationship” – an E-k Diagram. Thus, the allowed states of a system can be “indexed” in terms of discrete energy, E_n, or an associated momentum, h$\mathbf{k}_n$. The constant h is the Planck constant. Note the vector nature of **k** as signified by the bold print.

The individual $\mathbf{k}_n$s represent quantized solutions to the Schrödinger equation. They may be displayed as a three-dimensional array of lattice points, as shown below (Fig.1A.1.1). This lattice is a kind of Fourier transformation of the array of crystal lattice points as they appear in real space. The spacing of points in “momentum-space” is reciprocal to that of the spacing of real-space lattice points. For this reason, the momentum-space representation is frequently called a “reciprocal Lattice.”



a. Real-space lattice representation **b. Reciprocal (momentum-space) representation**

Fig. A1. Real and momentum-space lattice representations of the crystal lattice. Note the Inverse relationship between lattice constants in each representation space.

From this representation we may derive a “density of states.” The density of states (DOS) is defined as the number of allowed states in the crystal per unit of energy (at some energy) per unit volume of material. Consider a material with a cubic crystal structure. The reciprocal lattice of a cubic lattice is also a cubic lattice. Allowed momentum “solutions” of the Schrödinger equation are evenly spaced through momentum space such that:

$$k_x = n_x \pi/a \qquad k_y = n_y \pi/b \qquad k_z = n_x \pi/c$$

The n’s are integers and a,b and c are the real-space lattice constants. A unit volume of k-space is just:

$$\pi^3/abc$$

A single cube in k-space contains 8 vertices. Each k-space lattice point is shared between 8 unit cells. This there is one-eighth of a state associated with each vertex or 1 state per unit cell of the cubic reciprocal lattice. The number of solutions per unit volume of k-space is the reciprocal of this – one exception. Each discrete state can be occupied by two electrons (one spin up, the other spin down.) This doubles the electrons occupying these states. And so we have the fact that there are:

$$2\frac{abc}{\pi^3}$$

states per unit volume of k-space.

As stated above, energy is related to momentum. In quantum mechanics, the relationship is:

$$\mathbf{p} = h\,\mathbf{k}.$$

Note, that this relationship ties a wave property (wave-number, k) with a particle property (momentum, p). Just as in classical mechanics, we have:

$$E = p^2/2m$$

Or:

$$E = h^2\, k^2/2m.$$

E is energy and m is the effective mass of the particle as it moves through the solid. From this, we derive:

$$k = \frac{\sqrt{2Em}}{h}$$

Here, k may be regarded as the magnitude of a vector from the origin to any lattice point in k-space. We can use this relationship to relate intervals of energy, dE, to intervals in k-space, dk:

$$dk = \frac{1}{2h}\sqrt{\frac{2m}{E}}dE$$

The volume of a spherical shell centered at the origin, with shell thickness dk is:

$$dV_k = 4\pi k^2 dk = 4\pi \frac{2Em}{h^2}\frac{1}{2h}\sqrt{\frac{2m}{E}}dE = \frac{4\pi m^{3/2}E^{1/2}dE}{\sqrt{2}h^3}$$

In this volume there are:

$$dn = \frac{abc}{\pi^3}\frac{8\pi m^{3/2}dE}{\sqrt{2}h^3 E^{1/2}}$$

states. Of these states, only those in the "positive orthant" (the positive octant of the 3-dimensional k-space) represent unique states. Thus, the above formula for dn over-counts the exact number by 8. Dividing through by 8 on the right hand side of the above formula, and dividing both sides by dE and the volume (abc) gives us the formula for the density of states for a cubic lattice:

$$DOS = \frac{1}{V_r}\frac{dn}{dE} = \frac{1}{\pi^2}\frac{m^{3/2}E^{1/2}}{\sqrt{2}h^3}.$$

The volume subscript "r" indicates that this is the volume of the real-space unit cell. Note that the density of states increases as the square root of E. this is known as a "parabolic" density of states. In semiconductor physics, the zero of energy is taken as the band edge. The DOS increases parabolically at energies higher than the band edge.

2. Band Occupancy

Consider allowed states in an energy interval dE. The number of occupied states in that interval is the number of states in that interval, DOS(E) dE, multiplied by the probability that those states are occupied, p(E) the probability that those states are occupied. We can write:

$$n_{occ} = \int_{E_b}^{E_t} dn_{occ}(E) = \int_{E_b}^{E_t} DOS(E)p(E)dE.$$

We are assuming, here, that dE is so small that p(E) and DOS(E) are effectively constant over that energy interval. The total number of occupied states in a band is:

$$\int_{E_b}^{E_t} dn_{occ}(E) = \int_{E_b}^{E_t} DOS(E)\,p(E)\,dE \quad .$$

E_b is the energy at the "bottom" of the band and E_t is the energy at the top. As discussed in the text, electrons are Fermi particles obeying Fermi occupancy statistics. Thus

$$p(E) = \frac{1}{1 + \exp(\frac{E - E_f}{kT})} \, ,$$

where E_f is the Fermi level of the system, k is the Boltzmann constant and T is the temperature..

The resulting integral can be evaluated numerically. However, analytic results can be obtained in approximation. The basis of the most common approximation, the Boltzmann approximation, holds that the Fermi level is relatively "deep" in the bandgap and at room temperature, the Fermi-Dirac function, p(E) varies steeply around the Fermi level. Thus, we need only concern ourselves with the "tail" of the distribution function, for which the exponential in its denominator is much less than 1. For this case, the distribution is about equal to:

$$p(E) = \exp(-\frac{(E - E_f)}{kT})$$

and the occupancy integral is:

$$n_{occ} = \frac{1}{\pi^2}\frac{m^{3/2}}{\sqrt{2}h^3}\int_{E_c}^{\infty} E^{1/2}\exp(-\frac{(E - E_f)}{kT})dE$$

The E_c is the bottom of the conduction band energy and we take the integral to infinity, assuming the probability of high-energy occupancy of the band is zero. The integral evaluates to:

$$n_{occ} = 2\left(\frac{2\pi m_e^{*} kT}{h^2}\right)^{3/2} \exp\left(-\frac{(E_c - E_F)}{kT}\right)$$

where m_e* is the effective mass of the electron in the lattice.

The pre-factor to the exponential has the units "number per unit energy" and is frequently called the "effective density of conduction states", $N_{c.}$, giving:

$$n_{occ} = N_c \exp\left(-\frac{(E_c - E_F)}{kT}\right)$$

In order to calculate the valence-band occupancy, we must realize that a "hole" is the absence of an electron. We handle this in the occupancy statistics. Rather than asking "what is the probability a state is occupied," we ask, "what is the probability it is unoccupied." This is just: 1-p(E):

$$1 - p(E) = \frac{1}{1 + \exp(E_f - E/_{kT})}$$

Proceeding as above, we find:

$$p_{occ} = 2\left(\frac{2\pi m_p^* kT}{h^2}\right)^{3/2} \exp\left(-\frac{(E_f - E_v)}{kT}\right)$$

or:

$$p_{occ} = N_v \exp\left(-\frac{(E_f - E_v)}{kT}\right)$$

Again, m_p* is the effective mass of the hole.

Hand calculations can be greatly simplified as follows. In undoped material, there are as many electrons in the conduction band as there are holes in the valence band. This number, n_i, is the intrinsic carrier concentration (1.45 x 10^{10}/cm^3 in room-temperature silicon). Thus, the Fermi level lies midway between the bands at energy, E_i, known as the "intrinsic" or "mid-band" level. From out derivations above:

$$n_{occ}p_{occ} = N_c N_v \exp\left(-\frac{E_g}{kT}\right) = n_i^2$$

and:

$$n_i = \sqrt{(N_c N_v)} \exp\left(-\frac{E_g}{2kT}\right)$$

If we take Nc and Nv as approximately equal, we can re-write the mobile electron and hole occupancy expressions as:

$$n_{occ} = n_i \exp\left(\frac{E_f - E_i}{kT}\right)$$

$$p_{occ} = n_i \exp\left(\frac{E_i - E_f}{kT}\right)$$

Frequently, the "occ" subscripts are dropped, and we use n and p as the mobile electron and hole occupancies of the bands. The separation of the Fermi level from the intrinsic level is known as the "bulk potential), ϕ_b. For electrons and for holes, we have:

$$\phi_{bn} = E_f - E_i$$

$$\phi_{bp} = E_i - E_f$$

As an example, consider an n-type semiconductor. We assume that the background doping "overwhelms" the intrinsic carriers. This would imply:

$$n = N_d$$

And we can calculate the position of the Fermi level:

$$\phi_{bn} = \left(\frac{kT}{q}\right) Ln\left(\tfrac{N_d}{n_i}\right)$$

(kT/q) is just 0.026 volts at room temperature. Thus, if N_d is $10^{15}/cm^3$, the bulk electron potential is 0.34V.

3. The Depletion Approximation

The most straightforward path to hand calculation of relevant component parameters is through the depletion approximation. This approach allows us to rapidly compute the physical extent of space charge layers in bulk silicon. Here's how it works.

First, consider two separate blocks of semiconductor (Fig 1A.3.1) of different doping polarity (one p-type the other n-type). When the blocks are brought together, the mobile carriers intermingle by diffusion – holes migrate to the n-type block, electrons to the p-type block. What stops this process? Why don't the carriers just keep diffusion until a homogeneous mixture of electrons and holes is achieved? These questions are answered through electrostatic arguments. When a mobile carrier diffuses it leaves a fixes charge of opposite sign behind. For example, when an electron diffuses from the n-region to the p, a positively charged donor core is left behind. This gives rise to an electric field that "bucks out" this tendency to diffuse.

The overall tendency toward charge migration can be quantified through a "chemical potential drop." Chemical potential drops are specified in terms of Fermi-level differences encountered on moving through the solid semiconductor. More will be said on this when we deal with semiconductor transport. But for now, suffice it to say that in equilibrium, the affect of chemical potential drops nullified by electrostatic drops.

Another way of saying this is that in equilibrium, electrostatic forces cancel diffusive tendencies for mobile charges to migrate. This cancellation is purely local, and one does not have to resort to Kirchoff's circuit laws (or any factors extending through the whole system) to see how it occurs.

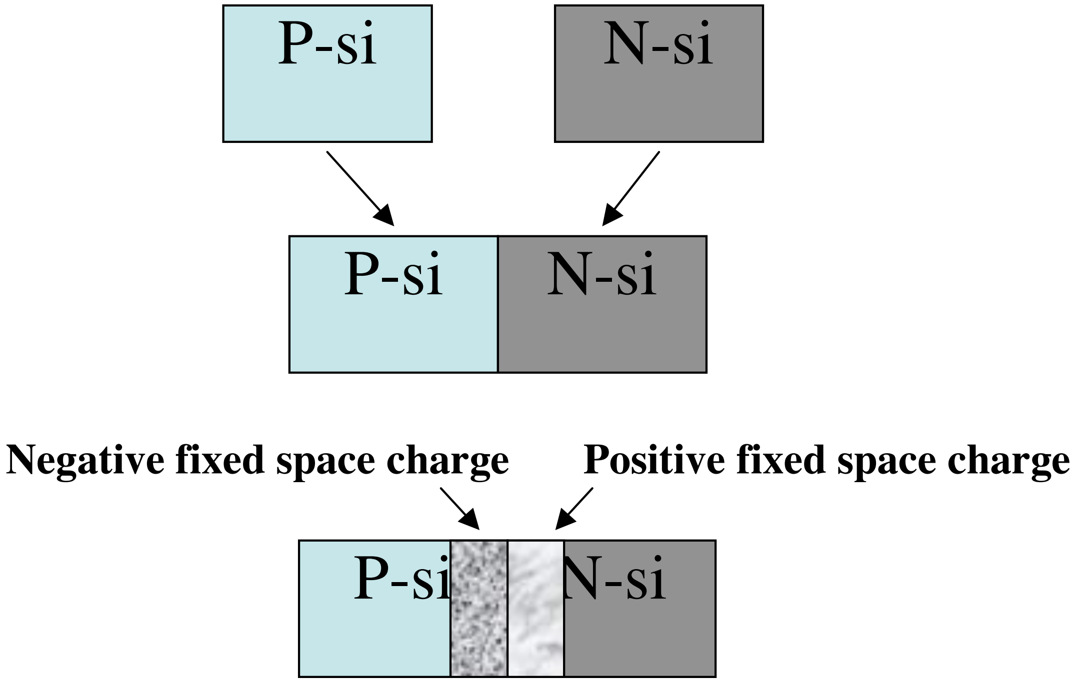


Fig 1A.3.1. Junction formation by diffusive transport between dissimilarly doped materials.

The result is a "space charge," a region of space over which the charge is not net neutral. Actually, for the case shown above, the space charge comes in two parts: one positive one negative. Both parts are "fixed" (immobile) charge distributions made up of charged dopant atom cores. As a result of the cancellation process discussed above, the total drop across both space charges is just the sum of the magnitudes of the bulk potentials of the p and n sides. Usually, we assume that the dopant-related charge is constant just up to the metallurgical junction. This is called the "abrupt junction" approximation. Furthermore, we assume that the space charge terminates abruptly at its edges and that only the fixed space charge plays a role in determining the electrostatic potential and fields. As the space charge is "depleted or mobile (or "majority") charge, this is called the depletion approximation, and the fixed-charge space charge is called the depletion region.

As we know the total drop across the space charge:

$$\Delta V = |\phi_{bn}| + |\phi_{bp}|$$

we can use the Poisson equation to find the physical dimensions of the space charge. The 1-D Poisson equation is:

$$\frac{d^2V}{dx^2} = -\frac{\rho(x)}{\varepsilon_{si}}$$

Of course, since the negative of the positional derivative of potential is electric field, we also have:

$$\frac{dE}{dx} = \frac{\rho}{\varepsilon_{si}}$$

The "first integral" of the Poisson equation (the solution to the above equation) can be obtained from physical considerations alone (with some help from Gauss' Law). Consider Fig.1A.3.2. In particular, consider the force on a positive test charge located at position x. It feels an attractive force from the sheet of charge extending to the left of the metallurgical junction. The Force this slab exerts is, by Gauss' Law, independent of the distance, x, and is given by:

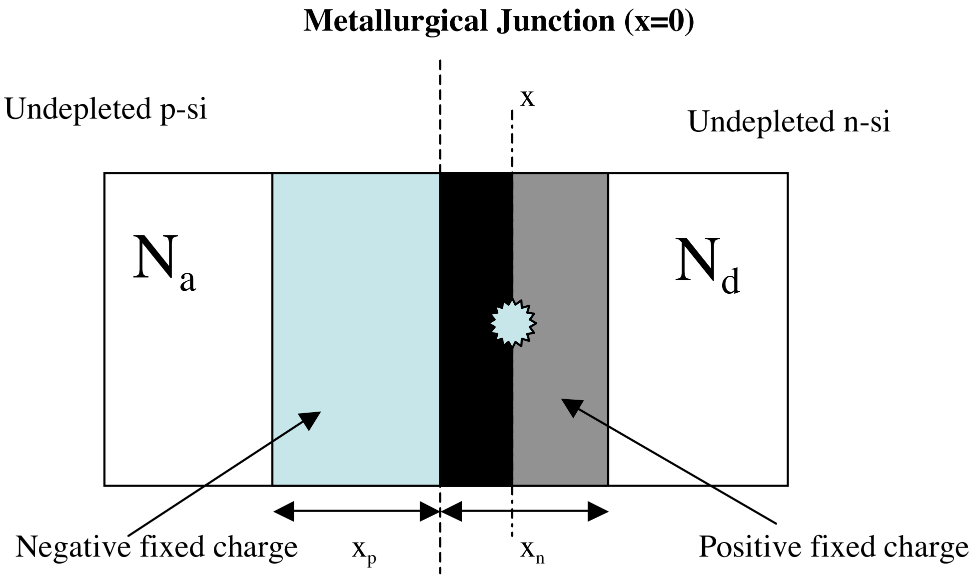


Fig.1A.3.2. A test charge in a depletion region.

$$F_1 = -\frac{\sigma}{2\varepsilon_{si}}$$

where q is the charge on the electron and σ is the charge-per-unit-area in the sheet. Note that the sign of the force is negative: it points to the negative half-space. For a constant doping, abrupt junction model:

$$\sigma = qN_a x_p$$

where x_p is the extent of the space charge into the p-region. And so:

$$F_1 = -\frac{qN_a x_p}{2\varepsilon_{si}}$$

The magnitude of this force is reduced by the positive charge sheet extending a distance x into the n-type silicon. The force reduction is given by:

$$F_2 = \frac{qN_d x}{2\varepsilon_{si}}$$

Finally, the third slab, extending to the right of x, has thickness x_n-x, and it exerts a negative force on the test charge:

$$F_3 = -\frac{qN_d\left(x_n - x\right)}{2\varepsilon_{si}}$$

The total force is the sum of the three forces:

$$F_{tot} = -\left(\frac{qN_a}{2\varepsilon_{si}}\right)x_p - \left(\frac{qN_d}{2\varepsilon_{si}}\right)\left(x_n - 2x\right)$$

where x_{sc} is the full thickness of the space charge layer. Clearly, when x is zero, the force is greatest. Thus, the maximum field is found at the metallurgical junction. It should be noted that a similar equation works for negative x as well. Using the same logic as that which produced the equation above, we have:

$$F_{tot} = -\left(\frac{qN_d}{2\varepsilon_{si}}\right)x_n - \left(\frac{qN_a}{2\varepsilon_{si}}\right)\left(x_p - 2x\right)$$

In this case, the force still pulls the test charge to the left, and still, the force decreases as the charge approaches the space charge boundary. At the space charge boundary, the force goes to zero, and remains zero throughout the undepleted region. Field lines originating in the space charge terminate in the space charge. In this one-dimensional model, no field lines extend outside of the depletion. Since:

$$F = qE = -q\frac{dV}{dx}$$

We can perform the “second integral” to get potential as a function of position. Take The metallurgical junction to be the zero of potential. First, let us integrate to the right (into the) donor doped side:

$$\int_0^{V(x)} dV = V(x) = \int_0^x \left(-\left(\frac{qN_a}{2\varepsilon_{si}}\right) x_p - \left(\frac{qN_d}{2\varepsilon_{si}}\right)(x_n - 2x) \right) dx$$

or:

$$V(x) = -\left(\frac{qN_a}{2\varepsilon_{si}}\right) x_p x - \left(\frac{qN_d}{2\varepsilon_{si}}\right)(x_n x) + \left(\frac{qN_d}{2\varepsilon_{si}}\right) x^2$$

If we carry the integration out to the space charge edge, we have the potential drop across the positively charged depletion:

$$V(x_n) = -\left(\frac{qN_a}{2\varepsilon_{si}}\right) x_p x_n$$

The negative sign is correct, since work is don on the test charge when it is moved in the positive x direction. Similarly, an integral on the acceptor-doped side yields:

$$V(x_p) = \left(\frac{qN_d}{2\varepsilon_{si}}\right) x_p x_n$$

The total potential drop across the space charge is the magnitude of the sum of these two terms:

$$V_{sc} = \left(\frac{q}{2\varepsilon_{si}}\right) x_p x_n (N_a + N_d)$$

This can be re-expressed in terms of the total extent of the space charge, x_{sc}. To see this, we must realize that:

$$x_{sc} = x_n + x_p$$

and:

$$x_n N_d = x_p N_a .$$

There is some physics to the last relationship. It states that every positive charge in the donor space charge throws out a field line that terminates on a charged acceptor in the p-region. That is implies that during the initial diffusion process, which set up the space charge to begin with, a mobile electron migrated to the p-side where it was trapped by an acceptor becoming a fixed charge. Similarly, just as many holes drifted to the n-side to become charged donors and the sum of all the charges on both sides of the metallurgical junction was zero.

From these last two relationships, we have:

$$x_n x_p = x_{sc}^2 \left(\frac{N_a N_d}{\left(N_a + N_d\right)^2} \right)$$

and:

$$V_{sc} = \left(\frac{q}{2\varepsilon_{si}} \right) x_{sc}^2 \left(\frac{N_a N_d}{N_a + N_d} \right)$$

From this we deduce the total width of the space charge, for a given space charge drop, V_{sc}:

$$x_{sc} = \sqrt{\left(\frac{2\varepsilon_{si}}{q} \left(\frac{N_a + N_d}{N_a N_d} \right) V_{sc} \right)}$$