Notes for Lecture 12

Carrier Equations

So far we learned the diffusion, the drift velocity, and a bit of recombination and generation. Here, we put all this together in a grand package and discuss some key physics that comes out of it.

12.1 RG processes

The recombination (R) and generation (G) occurs via the interaction of the electron with photon (Figs. T3.15a,d), phonon (Figs. T3.15b,e), or other electron (Figs. T3.15c,f). Of these, the process involving "RG centers" (Figs. T3.165) and phonons (Figs. T3.15b,e) are the most important for Si or any indirect semiconductor. For a direct semiconductor such as GaAs, the photon-mediated RG processes can occur easily, and thus they are important.

In any case, the RG process is a slow process in general, in comparison to the scattering time (i.e. the "relaxation time" $\sim ps$) of the charge transport process (LN 10). The time scale is generally longer than nano-second, can easily be on the order of a μs for indirect semiconductors, and can reach a ms or 10 ms even, if an indirect semiconductor is made very pure.

12.2 Particle Conservation

Any **"conservation law"** carries a very significant weight. Perhaps you are now into physics deep enough to know that the conservation laws are as important as the dynamical laws (such as Newton's equation or Schrödinger equation), if not more. However, one should also keep in mind when and why conservation laws may break.

The conservation law that we introduce here is the particle number conservation law. Namely, as long as particles do not get destroyed or generated, the number of particles should be conserved. It sounds plain and simple. So, this conservation law describes the "biography" of particles except their birth and death (however, see the end of this section).

Here is the law. For *any* particle (or for *any* quantum number of the particle, like charge) that is locally conserved[1],

$$\frac{\partial \mathcal{N}}{\partial t} + \nabla \cdot \vec{\mathcal{J}} = 0 \tag{12.1}$$

This equation is commonly referred to as the **continuity equation**. Here, $\mathcal{N}$ is the number density of the particle, and $\vec{\mathcal{J}}$ is the flux of the particle, the number of particles passing through a unit are per unit time. If we are talking about a quantum number of the particle, instead of the particle itself, then we must re-interpret $\mathcal{N}$ and $\vec{\mathcal{J}}$ accordingly. Indeed, in this course, we will be talking about the charge most of the time, and thus we take $\mathcal{N}$ to be the charge density (-e times the number density for electrons) and $\vec{\mathcal{J}}$ to be the electrical current.

The derivation of the above conservation law is rather easy[2]. Assume that $\vec{\mathcal{J}}$ is along a certain direction for a given arbitrary time t. We may take that direction as the positive x axis. Consider a small cylindrical volume with length dx along the x direction and a crossection A, perpendicular to the x axis. The left end of the cylinder is at x and the right end of the cylinder is at $x + dx$. We do not need to specify what the shape of the crossection is, except that A is a small area. From time t to $t + dt$, there will be particles going through the crossection A at $x + dx$, while there will be particles going through the crossection A at x. By definition the flux $\vec{\mathcal{J}}$ is a function of position and time, and so there may be net flux. By our choice of the coordinate system, the flux at $x + dx$ contributes to the loss of particles within the volume Adx, while the flux at x contributes to the gain of particles within that volume. So, the total change of the number of particles during dt is

$$-\mathcal{J}(x + dx, y, z, t)Adt + \mathcal{J}(x, y, z, t)Adt$$

[1] A local conservation law is usually what we mean by a conservation law. It means that particles can flow but they do not appear/disappear mysteriously.

[2] However, one might note that there is no need to derive/prove a law! Just as we don't need to prove Newton's law. As a matter of fact, the above law can be interpreted as a way to *define* what $\vec{\mathcal{J}}$ is. While the form of $\vec{\mathcal{J}}$ is obvious in many classical problems, it is not so in other problems, such as those in quantum mechanics. In such a case, the continuity equation can be used to *find* the proper definition of $\vec{\mathcal{J}}$ given a density function $\mathcal{N}$.

where $\mathcal{J}(x, y, z, t) = \mathcal{J}_x(x, y, z, t) \geq 0$ by our choice of the coordinate system. However, this change of the number of particles is, by definition, dt times the partial time derivative of the total number of particles within the volume Adx, i.e. it must be equal to

$$dt \frac{\partial(\mathcal{N}(x, y, z, t) A dx)}{\partial t}$$

Equating these two equations, and cancelling out the common factor, Adt, and dividing out by dt, we get

$$-\frac{\partial \mathcal{J}_x}{\partial x} = \frac{\partial \mathcal{N}}{\partial t}$$

The LHS is $-\nabla \cdot \vec{\mathcal{J}}$ by construction, and this completes the derivation.

This conservation law is of general validity, and is a recurring theme in all physics, including quantum physics. Note that this conservation law does not specify what the actual form of $\vec{\mathcal{J}}$ is. In fact, it can be viewed as the equation that defines what $\vec{\mathcal{J}}$ must be.

For our current problem, we take $\mathcal{N}$ as the charge density. What is important is that to use the above conservation law, we need to include, in the time derivative, only those processes that conserve the charge! As RG processes can generally cause particles to appear and disappear, they do not conserve charge when the electron system or the hole system is separately considered! However, DD (drift and diffusion) processes are definitely charge (or particle) conserving.

So, here is the charge conservation laws (per each system – electron or hole), or the continuity equations, as relevant to us:

$$-e \left.\frac{\partial n}{\partial t}\right|_{DD} + \nabla \cdot \vec{J}_n = 0 \tag{12.2}$$

$$e \left.\frac{\partial p}{\partial t}\right|_{DD} + \nabla \cdot \vec{J}_p = 0 \tag{12.3}$$

Here, we made use of the fact that $-en$ is the charge density of the electron and ep is the charge density of the hole. $\vec{J}_n$ or $\vec{J}_p$ is the **charge flux** associated with the electron system or the hole system, respectively. It is charge that passes per unit time and unit area. So, it is what we also call the **current density**, current per unit area.

If other non-DD process are included, then the above equations must be generalized, by noting that $\partial/\partial t = \partial/\partial t|_{DD} + \partial/\partial t|_{\text{non-}DD}$. Then, what we get is

$$-e\frac{\partial n}{\partial t} + \nabla \cdot \vec{J}_n = -e \left.\frac{\partial n}{\partial t}\right|_{\text{non-}DD} \tag{12.4}$$

$$e\frac{\partial p}{\partial t} + \nabla \cdot \vec{J}_p = e \left.\frac{\partial p}{\partial t}\right|_{\text{non-}DD} \tag{12.5}$$

The terms on the right hand side are called "source" or "sink" terms, depending on whether they generate or destroy particles/charges. In this course, we consider RG processes as main source and sink mechanisms: i.e. non-DD may be equated to RG in most cases.

12.3 Carrier Equations, General Form

At this point, it is helpful to collect all equations that we investigated so far, and establish the most general form of carrier equations.

First, the current is given by

$$\vec{J}_p = pe\vec{v}_{d,p} - eD_p\nabla p \tag{12.6}$$

$$= pe\mu_p\vec{E} - eD_p\nabla p \tag{12.7}$$

where the first term is the drift current and the second term is the diffusion current. The drift velocity is related to the electric field as $\vec{v}_{d,p} = \mu_p\vec{E}$ via the mobility μ_p, defined so. Note that, within the Drude theory (LN 10), the mobility was given as $e\tau/m_p^*$, where τ is the relaxation time due to the scattering of charge carrier. In this lecture and on, we will be talking of the minority lifetime and other time scales with the same symbol τ! Do not be confused by them, and make sure that you know what time we are talking about each time. In this lecture, we will be talking of the "lifetime" which is literally how long a carrier lives.

For the electron, the above equation becomes

$$\vec{J}_n = ne\mu_n\vec{E} + eD_n\nabla n \tag{12.8}$$

Notice the sign change in the second term, but not in the first term.

Second, how do carriers (dis-)appear can be summarized as

$$\left.\frac{\partial p}{\partial t}\right|_{RG} = G_p - R_p \tag{12.9}$$

$$\left.\frac{\partial n}{\partial t}\right|_{RG} = G_n - R_n \tag{12.10}$$

Here, G is the generation rate ("birth rate") and R is the recombination rate ("death rate"). Both are defined as positive. However, sometimes the recombination rate is defined as negative. In that case $-R_p$ or $-R_n$ is to be understood as the recombination rate (as in the textbook).

Third, we have the charge conservation equations as we investigated in the previous section.

Collecting all of these, we can set up two equations. The equation for n can be formed by collecting Eqs. 12.4 (with non-DD replaced by RG),12.8,12.10:

$$\frac{\partial n}{\partial t} = \mu_n \nabla \cdot (n\vec{E}) + D_n \nabla^2 n + G_n - R_n \tag{12.11}$$

The equation for p can be formed similarly, by collecting Eqs. 12.5 (with non-DD replaced by RG),12.7,12.9:

$$\frac{\partial p}{\partial t} = -\mu_p \nabla \cdot (p\vec{E}) + D_p \nabla^2 p + G_p - R_p \tag{12.12}$$

These are the most complete equations for the carriers as far as we are concerned at this point. These are complete and very general, but perhaps "too true to be good" for the purpose of this course. It is too big an equation to swallow for us, and we would need to gain understanding of these equations step by step by making certain realistic approximations!

First of all, let us assume that the interesting attributes of the device changes along one direction only, say along the x axis. Then, we have[3]

$$\frac{\partial n}{\partial t} = \mu_n \frac{\partial (nE)}{\partial x} + D_n \frac{\partial^2 n}{\partial x^2} + G_n - R_n \tag{12.13}$$

$$\frac{\partial p}{\partial t} = -\mu_p \frac{\partial (pE)}{\partial x} + D_p \frac{\partial^2 p}{\partial x^2} + G_p - R_p \tag{12.14}$$

Except for the one dimensional restriction, these equations remain the most general form as far as the equations for n and p go.

12.4 Frequent Assumptions for Extrinsic Semiconductors

We will for the most part assume the following, when we deal with n type or p type semiconductor. These conditions will be assumed automatically unless explicit comments are made to contradict them.

[3]The equations that were used in class had some superfluous e's in some terms of these equations. There should not be any e's in these equations.

1. The minority carrier density is much smaller than the majority carrier density.
2. The variations Δn and Δp of the charge carrier densities are much smaller than the majority carrier density.
3. The temperature is the room temperature.

The second item corresponds to the "low level injection" assumption. Note that even for a low level injection, the variation Δn or Δp can be much larger than the minority carrier density! This is quite common.

12.5 Deep Trap, Mid-Gap, and Surface, States

Before exploring the meaning of the above equations, we need to figure out what to do with *RG* terms.

To be specific, let us consider the *RG* physics due to "the *RG* center." In this case, we are considering the so-called "deep trap" states. A deep trap state is like a donor/acceptor impurity energy level, but it is much higher in energy, and appear as mid-gap states (cf. Fig. T3.16), not as shallow donor or acceptor states. Well-known examples of impurities causing deep trap states include Au in Si and N in GaP.

Deep trap states can be understood in terms of bonding and anti-bonding of atomic orbitals[4]. The fundamental cause is the increased disparity between the atomic energy levels of the impurity and the intrinsic atomic energy levels of the host crystal. Therefore, deep trap states are quite different from donor or acceptor states in that the wave functions are localized close to the impurity.

Note that the band structure of a crystal results from the repeated bonding and anti-bonding of local orbitals in a unit cell (LN 4)[5]. Each energy band corresponds to one local orbital per unit cell, and this local orbital is completely degenerate across all unit cells. The bonding and anti-bonding of this local orbital between

[4] H. P. Hjalmarson, P. Vogl, D. J. Wolford, and J. D. Dow. Theory of substitutional deep traps in covalent semiconductors. Physical Review Letters, 44(12):810, March 1980; A. Resende, R. Jones, S. berg, and P. R. Briddon. Calculations of electrical levels of deep centers: Application to Au-H and Ag-H defects in silicon. Physical Review Letters, 82(10):2111, March 1999.; E L Silva, J Coutinho, A Carvalho, V J B Torres, M Barroso, R Jones, and P R Briddon. Electronic structure of zn, cu and ni impurities in germanium. Journal of Physics: Condensed Matter, 23(6):065802, February 2011.

[5] Note that the local orbital can be an atomic orbital, as in the case of a hydrogen crytal, or it can be a *molecular* bonding (valence band) and anti-bonding (conduction band) orbital if the unit cell contains multiple atoms as in Si or GaAs.

unit cells is best described as quantum tunneling. Now, what happens when an impurity atom replaces an atom in the crystal? Of course, an impurity atom has different atomic energy levels. This means that the unit cell that contains the impurity atom will end up having somewhat different local orbitals within a unit cell. This means a discouraged quantum tunneling into and out of the unit cell containing the impurity atom. How strong this discouragement is the source of the difference between an extended impurity orbital that is easily ionized (typical donor/acceptor impurity state) and a more localized impurity orbital that ends up deep inside the energy gap ("mid-gap" state; deep trap states).

Does this mean that the Bohr atom model that we used previously for understanding the donor/acceptor impurity levels is not valid any more for deep trap states? Yes, and no. Obviously, difficulty arises if one were to describe a mid-gap state from the Bohr picture, since it is not clear where the mid-gap state should originate from: i.e. is it supposed to be a bound state of the valence band hole or a bound state of a conduction band electron? The answer is both! Thus, the formulation of the problem from this point of view is obviously not simple. An additional problem that one can note is that the "effective mass model" becomes poorly defined for mid-gap state. The reason is that the effective mass model is valid only near the bottom of the conduction band or near the top of the valence band. As the impurity orbital becomes localized to a few lattice constants (rather than a few tens of lattice constants), Δk becomes a significant fraction of the k space unit cell, and in such a large k region, the effective mass approximation becomes poor. With all these problems, though, it must be true that the general idea of the Bohr atom model must be valid in some crude sense, at least. For instance, generally the effective mass will increase, since the band structure tends to flatten away from the band gap edge, consistent with the fact the mid-gap states have higher binding energy from the Bohr atom point of view. Also, the dielectric constant will generally decrease at short distances, another qualitative feature that goes along with the higher binding energy.

Note that even without impurity states, mid-gap states can occur on the *surface* of a semiconductor. The reason is that at the surface the atomic potential energy is not the only potential energy. There is also the potential energy step that keeps the electrons contained within the crystal. Such a confining potential is the origin of the so-called "work function" of electron, which you might be familiar with if you ever learned about the photoelectric effect experiment. While this confining potential is typically thought of as a "step function," the "step" cannot be much sharper than an atomic layer, because in the end the confining potential is all due to the electrons themselves[6]! The consequence is that this potential energy step affects the potential

[6]The microscopic theory of this confining potential, and subsequently the work function, is famously very complicated! A part of it is due to a thin dipole layer formed by the leakage of an electron towards the vacuum side. Another part is due to strong and "highly correlated" interac-

energy within couple of layers of atoms on the surface. This can generally lead to a surface state that ends up mid-gap. Just as impurity states are Bohr like states with exponentially damped (radially) wave functions, surface states are those states whose wave functions are exponentially damped perpendicular to the surface. However, parallel to the surface, surface states can be thought of as band states for a two dimensional crystal (which the surface is!).

12.6 Minority Carrier Lifetime

The minority carrier lifetime is a crucial number of a semiconductor device.

Let us first take the case of RG centers, and consider its microscopic origin. Here, we consider a deep trap state with energy E_T lying mid-gap (cf. Figs. T3.16 and T3.21). Here, we consider the density of the deep trap impurities $N_T \ll$ the majority carrier density, in addition to all the usual assumptions of Section 12.4 of page 5.

The RG center mediated process is called "indirect thermal" process, since the interaction between n and p is indirect, mediated by the RG center, which is an impurity with a deep trap state. It is called thermal since the electron energy difference in the process is supplied from or dissipated to local lattice vibrations. It can be thought of as heat supplied to or taken away from the electronic system by the system of phonons, quanta of lattice vibrations. These processes involve local states (deep trap states) and local vibrations, and thus states involved do not have sharply defined momentum values. In contrast, direct photon-assisted process (cf. Fig. T3.18) involves extended band states with sharp momentum quantum numbers. Thus, the momentum conservation is of crucial importance in photon-assisted process. As the photon momentum is very small, the transition occurs vertically in the energy-momentum diagram of the electronic energy band diagram, which is why such a process is impossible at the band edge for an indirect semiconductor[7].

For definiteness, we consider an n type semiconductor, and here is the summary of our definitions and assumptions.

$$n = n_0 + \Delta n \tag{12.15}$$

$$p = p_0 + \Delta p \tag{12.16}$$

$$n_0 \gg p_0, \Delta n, \Delta p, N_T \tag{12.17}$$

tions between electrons themselves.

[7]In this case, such a direct transition requires involvement of phonon as well as photon, making the transition rate very slow

where n_0, p_0 are carrier densities in equilibrium. We assume, as typical, that at room temperature $n_0 \approx N_D$ and $p_0 \approx N_A$.

12.6.1 Zero Temperature Case

It is helpful to think of the $T = 0$ case first, where the following is true in equilibrium.

1. The valence band is completely full and the conduction band is completely empty (as always!).
2. The trap states are completely full, since E_T is much less than the donor impurity energy level, and $N_T \ll N_D$.
3. The Fermi level (E_F) is pinned to the donor impurity level, which is slightly below E_c.

Consider applying a perturbation to this system, by which Δp and Δn become non-zero with all other things being like in the above equilibrium state. We focus on the question, what would happen to Δp? (We will consider the question of the dynamics of the majority carrier density Δn in the next lecture.) Within the *RG* center model, there are two processes that will affect Δp. First, the electron that is trapped at E_T can fall down to E_v since now there are holes to fill in! This rate will depend on how many electrons we have at E_T and how many holes we have at Δp.

$$\left.\frac{\partial \Delta p}{\partial t}\right|_{R:RG-center} = -c_p N_T \Delta p \tag{12.18}$$

Here, c_p is the so-called "capture coefficient." This process may be viewed as the hole being captured/bound by the deep trap impurity. Second, the reverse process of a hole being released by the trap is possible also. For this process to occur, though, there must be holes at E_T, but there are no such holes by our assumption! Therefore, the above rate is in fact the total rate

$$\left.\frac{\partial \Delta p}{\partial t}\right|_{RG-center} = -\frac{\Delta p}{\tau_p} \tag{12.19}$$

where

$$\tau_p = \frac{1}{N_T c_p} \tag{12.20}$$

is the *minority carrier lifetime.*

12.6.2 Room Temperature Case

In this case, essentially the same argument can be applied. The only difference is that (1) a very small portion of N_T is empty in the equilibrium state, and (2) a very small portion of the valence band is empty (p_0 is now finite, but very small). Now the total number of holes is $p = p_0 + \Delta p$, as p_0 is not zero.

Thus, we have, for recombination due to the RG-center,

$$\left.\frac{\partial p}{\partial t}\right|_{R:RG-center} = -c_p N_T p \tag{12.21}$$

How about the generation? The generate rate will depend on how many electrons in the valence band can go up to fill the holes at E_T. The key thing is that the number of electrons, i.e. the total number of electrons in the valence band, hardly changed due to Δp, relative to that of the equilibrium state (or, in fact, due to p for that matter, since both p and Δp are very small fractions of the total number of states available in the valence band). Therefore, we expect that the generate rate will be independent of Δp. Namely, the generation rate will be equal to that of the equilibrium state. On the other hand, in equilibrium, the generation and the recombination should cancel each other perfectly. And, so, applying the above formula with $p = p_0$ and changing sign, we get

$$\left.\frac{\partial p}{\partial t}\right|_{G:RG-center} = c_p N_T p_0 \tag{12.22}$$

The net effect is the same as that at the zero temperature!

$$\left.\frac{\partial p}{\partial t}\right|_{RG-center} = -c_p N_T (p - p_0) = -c_p N_T \Delta p \tag{12.23}$$

Noting that the time derivative of p_0 is zero, by definition, the LHS is equal to $\partial \Delta p/\partial t$, and so we get, just as at zero temperature,

$$\left.\frac{\partial \Delta p_n}{\partial t}\right|_{RG-center} = -\frac{\Delta p_n}{\tau_p} \tag{12.24}$$

where

$$\tau_p = \frac{1}{N_T c_p} \tag{12.25}$$

is the *minority carrier lifetime*. For a p type semiconductor, we would get, similarly,

$$\left.\frac{\partial \Delta n_p}{\partial t}\right|_{RG-center} = -\frac{\Delta n_p}{\tau_n} \tag{12.26}$$

where

$$\tau_n = \frac{1}{N_T c_n} \tag{12.27}$$

is the *minority carrier lifetime* for the electron. Here, to avoid confusion, n or p is written with a subscript n or p, corresponding to the nature of the majority carrier.

12.6.3 Consideration of Majority Carriers

In the RG-center driven processes, the above arguments can apply to the *majority* carriers as well as to the minority carriers, yielding the result that τ_n applies to the case of majority carriers in an n type semiconductor and τ_p applies to the case of minority carriers in a p type semiconductor. However, we continue to call τ_n and τ_p as minority carrier lifetimes, since the majority carrier has an additional shorter time-scale (next lecture). Then, on a time scale which is the greater of the two time scales τ_n and τ_p (which one would expect to be of similar magnitude), the system will come to an equilibrium, making each of n, p and the number of holes in the deep trap level come to its equilibrium value.

Note that if τ_n and τ_p are different, then the number of holes at E_T will be time-dependent, and this may affect τ_n and τ_p themselves, making them somewhat time-dependent. However, this would be a self-balancing type of time-dependence: i.e. the difference between τ_n and τ_p will become smaller.

In the photon-driven direct processes, the rate at which the majority carrier and the minority carrier are generated or annihilated are *exactly* the same. And so, there is only one time scale in that case.

As we will see in the next lecture (and HW 7), these are slow processes, while the majority carriers have a separate fast dynamics which screens out any charge inhomogeneity of the crystal.

12.7 Carrier Equations

With minority carrier lifetimes τ_n and τ_p as defined above (which can be defined similarly for any other pair generation and recombination processes – see HW 7.1), and assuming that n_0 and p_0 are independent of x, we can write our carrier equations as

$$\frac{\partial \Delta n}{\partial t} = \mu_n \frac{\partial (nE)}{\partial x} + D_n \frac{\partial^2 \Delta n}{\partial x^2} - \frac{\Delta n}{\tau_n} \tag{12.28}$$

$$\frac{\partial \Delta p}{\partial t} = -\mu_p \frac{\partial (pE)}{\partial x} + D_p \frac{\partial^2 \Delta p}{\partial x^2} - \frac{\Delta p}{\tau_p} \tag{12.29}$$

These equations are of general validity, under the assumptions of small carrier injections (which make it possible simplify the RG terms) and the spatial independence of n_0 and p_0. As we shall see, both these assumptions are not appropriate within the "depletion region" of a pn junction. However, these assumptions are valid outside the "depletion region" – in the so-called "quasi-neutral" region (LN 15).

Note that the above equations do not consider any other RG mechanisms than those that exist in the equilibrium situation. Namely, the minority lifetime term includes the interaction of the electron-hole system with the trap impurities, or with the ambient gas of photons, or within themselves (Auger process). If the system is however subjected to a stream of energy, which results in a net generation rate of electron-hole pairs, then such interaction will lead to an additional term G_{net} to the RHS of each of the above equations.

In the next lecture, we will consider the minority carrier equation and the majority carrier equation separately.