Physics of Semiconductors (3)

Shingo Katsumoto

Institute for Solid State Physics, University of Tokyo

April 23, 2013

3.6 Quantum states at impurities

In the beginning of this lecture, the importance of structural sensitivities in semiconductors is emphasized. A representative in structural sensitivities is the phenomenon that the transport property is dramatically changed with a tiny amount of impurities which produce carriers. The change also extends to optical and chemical properties. In this section we consider "shallow" impurity states in semiconductors. What "shallow" or "deep" means for impurities? This is semi-quantitative description how far is the localized energy level is from the bottom of the conduction band or from the top of the valence band in a semiconductor. We begin with introduction of a slow-varying potential modulation to a crystal potential.

3.6.1 Envelope function and effective mass approximation

We write a single electron Hamiltonian of a periodic lattice as $\mathscr{H}_0 = \hat{T} + V(\boldsymbol{r})$ and add a potential $U(\boldsymbol{r})$ which varies slowly in space in comparison with $V(\boldsymbol{r})$. Eigenfunctions $\eta(\boldsymbol{r})$ satisfy

$$[\mathscr{H}_0 + U(\boldsymbol{r})]\eta(\boldsymbol{r}) = E\eta(\boldsymbol{r}). \tag{3.1}$$

The system has now lost the translational symmetry and $\eta(\boldsymbol{r})$ cannot be written in the form of Bloch functions $|n, \boldsymbol{k}\rangle$. We then expand $\eta(\boldsymbol{r})$ with a series of Bloch functions, which is complete.

$$\eta(\boldsymbol{r}) = \sum_{n,\boldsymbol{k}} f(n, \boldsymbol{k})|n, \boldsymbol{k}\rangle. \tag{3.2}$$

Substituting (3.2) to (3.1) we take the inner product with $\langle n', \boldsymbol{k}'|$ to obtain,

$$[E_0(n', \boldsymbol{k}') - E]f(n', \boldsymbol{k}') + \sum_{n,\boldsymbol{k}} \langle n', \boldsymbol{k}'|U|n, \boldsymbol{k}\rangle f(n, \boldsymbol{k}) = 0. \tag{3.3}$$

The second term in the left hand side represents the scattering $|n, \boldsymbol{k}\rangle \to |n', \boldsymbol{k}'\rangle$. From (C.5) in Appendix C, this term is estimated as $U_{\boldsymbol{k}'-\boldsymbol{k}}\delta_{n'n}$. Here $U_{\boldsymbol{q}}$ is defined in (C.3) as the Fourier transform of $U(\boldsymbol{r})$. This estimation can be understood as follows. The factor $\delta_{n'n}$ means there is no interband transition. Because $U(\boldsymbol{r})$ is a slowly varying function in space, the perturbation to $u_{n\boldsymbol{k}}(\boldsymbol{r})$, which has the lattice periodicity in the Bloch function (1.5) and strong spatial variation, is small. Hence from the orthogonality of $u_{n\boldsymbol{k}}$, the interband scattering should be negligible. The other factor $U_{\boldsymbol{k}'-\boldsymbol{k}}$ indicates that only the plane wave part $e^{i\boldsymbol{k}\boldsymbol{r}}$ in (1.5) is taken into account as the $\boldsymbol{k}$-dependence. In other words, we have ignored so called Umklapp scattering, which is introduced through the lattice periodic part $u_{n\boldsymbol{k}}$ and causes jumps of a reciprocal lattice vector. The latter approximation is also reasonable from the small spatial derivative of $U(\boldsymbol{r})$. Therefore

$$[E_0(\boldsymbol{k}') - E]f(n', \boldsymbol{k}') + \sum_{\boldsymbol{k}} U_{\boldsymbol{k}'-\boldsymbol{k}}f(n', \boldsymbol{k}) = 0. \tag{3.4}$$

In the summation in (3.2) we only pick up slowly varying term around $k \sim 0$. From the form of eq. (1.4) we see that u_k does not depend on k strongly. Then we replace u_k in $|n, \boldsymbol{k}\rangle = e^{i\boldsymbol{k}\boldsymbol{r}}u_{\boldsymbol{k}}(\boldsymbol{r})$ with $u_0(\boldsymbol{r})$ and put it in front of the summation. Also we ignore the interexchange between energy bands and write the equations for each band as

$$\eta_n(\boldsymbol{r}) = u_{n0} \sum_{\boldsymbol{k}} f(n, \boldsymbol{k})e^{i\boldsymbol{k}\boldsymbol{r}} = u_{n0}f_n(\boldsymbol{r}). \tag{3.5}$$

Namely $\eta_n(\boldsymbol{r})$ is written as a product of a lattice periodic function u_{n0} and $f_n(\boldsymbol{r})$, which is the inverse Fourier transform of the expansion coefficient of Bloch wavefunction $f(n, \boldsymbol{k})$ and is gradual in space in comparison with the lattice period.

$f_n(\boldsymbol{r})$ are called as **envelope function**, which is a generalization of the plane wave part of Bloch functions $e^{i\boldsymbol{k}\boldsymbol{r}}$. Just like the wavepacket in the Bloch oscillation simulation in Sec.3.2, with taking summation on $\boldsymbol{k}$, we can form some localized states, which is gradually varies in real space.

Now In (3.4) we omit n and assume the effective mass m^* is isotropic. Then we obtain

$$\frac{\hbar^2 k'^2}{2m^*} f(\boldsymbol{k}') + \sum_{\boldsymbol{k}} U_{\boldsymbol{k}'-\boldsymbol{k}} f(\boldsymbol{k}') = E f(\boldsymbol{k}'). \tag{3.6}$$

In the inverse Fourier transformation, the second term in the LHS becomes a convolution. Then the equation for real space is given as

$$\left[-\frac{\hbar^2 \nabla^2}{2m^*} + U(\boldsymbol{r}) \right] f(\boldsymbol{r}) = E f(\boldsymbol{r}). \tag{3.7}$$

This is a Schrödinger equation for a particle with the mass m^* in the potential $U(\boldsymbol{r})$.

The above derivation indicates that the picture of a particle with an effective mass holds even under introduction of a potential which breaks the spatial translational symmetry of the lattice, The approximation in which we treat the envelope function as a wavefunction of a particle with an effective mass, is called **effective mass approximation**. This is natural if we remember that the effective mass is valid for spatially localized wavepacket in the simulation of Bloch oscillation. In the field of device application of semiconductors we often adopt so called **rigid band model**, in which electric field inside semiconductors is taken into account by slanting the band structure without changing any parameters. The above result also suggests that if the slanting by the field is gradual compared with lattice potential, the rigid band model is a good approximation.

3.6.2 Donors and acceptors

Imagine that we replace a Si atom in a Si crystal, which is a carbon group elementary semiconductor, with a P atom, which is in group VA (or 15 in IUPAC). Si crystal is formed from covalent binding through sp^3 hybrid orbitals. And if the P atom obeys this crystallization rule, the nucleus has an extra $+e$ positive charge. Or the P atom has an excess electron in the outermost shell. Hence the P atom works as a $+e$ charged center. From conduction electrons it introduces localized potential $U(\boldsymbol{r})$, and if this varies slowly

III	IV	V
$_5$B $2s^2 2p$	$_6$C $2s^2 2p^2$	$_7$N $2s^2 2p^3$
$_{13}$Al $3s^2 3p$	$_{14}$Si $3s^2 3p^2$	$_{15}$P $3s^2 3p^3$

enough, the approximation in (3.7) holds and a comparatively shallow bound state is expected.[1] We call such an impurity which has a bound state just below the conduction band, a **donor**. A donor is easily ionized with thermal energy and emits an electron into the conduction band.

On the other hand when we replace a Si atom in a Si crystal with a B atom, the B atom has deficiency of an electron and if it forms the covalent bond, the total charge around it becomes $-e$. Hence a single electron extracted state, that is a hole, forms a localized state just above the valence band. This also easily absorbs an electron from the valence band and emits a hole into the valence band. Such an impurity is called an **acceptor**.

Semiconductors generally have comparatively large dielectric constant due to the electric polarization of covalent electrons. Hence the impurity potential usually becomes much weaker than that of $\pm e$ potential in the vacuum. Accordingly the binding energy at impurity potentials is much smaller than that of H-atom, the localized states extend over several neighboring unit cells. Thus the effective mass approximation is good for many cases. Actually, the accuracy of the approximation depends on the value of effective mass itself and also varies with matrix semiconductors and species of impurities.
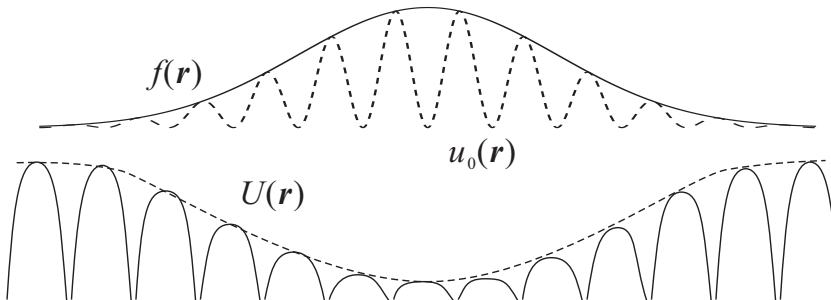

Figure 3.1: Schematics of an image of envelope function. The lower part illustrates the overlapping of a slowly varying potential $U(\boldsymbol{r})$ onto a lattice periodic potential. The upper shows a lattice periodic function $u_0(\boldsymbol{r})$ is modulated by the potential and localized in the form of an envelope wavefunction $f(\boldsymbol{r})$.

---

[1] In the case of Si:P (we use this notation as P doped Si), the measured ionization energy is 45meV, which is not so small and significantly larger than the theoretical prediction of 29meV based on the effective mass approximation. In other words, the effective mass approximation is not very good for impurities in Si.

With taking the origin at the position of the impurity, $U(\boldsymbol{r})$ can be written as $U(r) = -e^2/4\pi\epsilon_0\epsilon r$ and the effective mass equation becomes

$$\left[-\frac{\hbar^2\nabla^2}{2m^*} - \frac{e^2}{4\pi\epsilon_0\epsilon r} \right] f(\boldsymbol{r}) = Ef(\boldsymbol{r}). \tag{3.8}$$

This equation has the same form as that for a hydrogen atom other than the effective mass m^* and dielectric constant ϵ. The results for a hydrogen atom can thus be applicable.

$$Ry^* = \frac{e^2 m^*}{2(4\pi\epsilon\epsilon_0)^2\hbar^2} = \frac{m^*}{m}\frac{1}{\epsilon^2}Ry, \quad a_{\text{B}}^* = \frac{4\pi\epsilon\epsilon_0\hbar^2}{m^*e^2} = \frac{m}{m^*}\epsilon a_{\text{B}}, \tag{3.9}$$

are **effective Rydberg constant** and **effective Bohr radius** respectively.

As noted above the accuracy of the effective mass approximation largely depend on the species of semiconductors. In the case of Si, the effective masses both for conduction band and valence band are comparatively large and six bottoms of the conduction band exist in the first Brillouin zone. These make the approximation not so good for Si. On the other hand in the case of GaAs conduction band, the bottom exists at Γ point ($\boldsymbol{k} = 0$ in the reciprocal lattice space) having $\epsilon \approx 11.5$, $m^* \approx 0.067m$, which gives $a_{\text{B}}^* = 172a_{\text{B}} = 91$Åmuch larger the the lattice constant 5.65Å. Consequently the effective mass approximation very well holds giving the binding energy 6.9 meV for $1s$ state, which is close to the experimental observation of about 6meV.

3.7   Doping and carrier distribution

3.7.1   Chemical doping and types of doped semiconductors

Intentional introduction of impurities into crystals is called **doping** while we call the crystals matrices or hosts. Let us consider the case we dope donors uniformly with the density N_{D}. At absolute zero all the electrons emitted from the donors are bound to the donors. [2] At finite temperatures some of them are excited to the conduction band and can carry electric charges. We call them "carriers" or "electrons". Let n be the density of such electrons and n_{D} be the density of electrons bounded at the donors. From the charge neutrality condition we get $n + n_{\text{D}} = N_{\text{D}}$.

Now we estimate Helmholtz free energy $F = U - TS$ by considering the number of cases W for assigning n_{D} electrons to N_{D} states. From $S = k_{\text{B}}\ln W$,

$$F = E_{\text{D}}n_{\text{D}} - k_{\text{B}}T\ln\left[2^{n_{\text{D}}}\frac{N_{\text{D}}!}{n_{\text{D}}!(N_{\text{D}} - n_{\text{D}})!} \right].$$

E_{D} is the position of the bound state measured from the bottom of the conduction band and $2^{n_{\text{D}}}$ is due to the spin degeneracy. We assume that the Coulomb repulsion prevents double occupation of a localized state with two electrons. According to Starling approximation $\ln N! \sim N\ln N - N$, the chemical potential (Fermi energy) is given as

$$\mu = E_{\text{F}} = \frac{\partial F}{\partial n_{\text{D}}} = E_{\text{D}} - k_{\text{B}}T\ln\left[\frac{2(N_{\text{D}} - n_{\text{D}})}{n_{\text{D}}} \right]. \tag{3.10}$$

And from this

$$n_{\text{D}} = N_{\text{D}}\left[1 + \frac{1}{2}\exp\left(\frac{E_{\text{D}} - E_{\text{F}}}{k_{\text{B}}T} \right) \right]^{-1} \tag{3.11}$$

is obtained. The factor 1/2 on the exponential function is due to the spin degeneracy.

Similarly, for uniform doping of acceptors with density N_{A}, the density of electrons bounded to the acceptors n_{A} is

$$n_{\text{A}} = N_{\text{A}}\left[1 + 2\exp\left(\frac{E_{\text{A}} - E_{\text{F}}}{k_{\text{B}}T} \right) \right]^{-1}. \tag{3.12}$$

Here we have a factor 2 instead of 1/2 but the density of holes bounded to the acceptors is $p_{\text{A}} = N_{\text{A}} - n_{\text{A}}$ and symmetrical with n_{D} having a factor 1/2.

From (3.10), if we dope only "shallow" donors, for which the effective mass approximation holds, E_{F} comes to E_{D} at $T \to 0$. E_{D} should be much smaller than E_{g}. Accordingly from (3.11), the electron concentration n becomes much higher than that of the intrinsic semiconductor at finite temperatures. This type of semiconductors are called **n-type**. Similarly doping of acceptors enhances the hole concentration p. We call them **p-type**.

When donors and acceptors co-exist, the semiconductor becomes n-type for $N_{\text{D}} \gg N_{\text{A}}$ and p-type for $N_{\text{D}} \ll N_{\text{A}}$. In the former, some of the electrons emitted from donors are captured to acceptors and almost all the acceptors are ionized. In the latter, the other way around. In both cases we say such semiconductors are **compensated**.

---

[2]In so called degenerate semiconductors the following discussion does not hold.
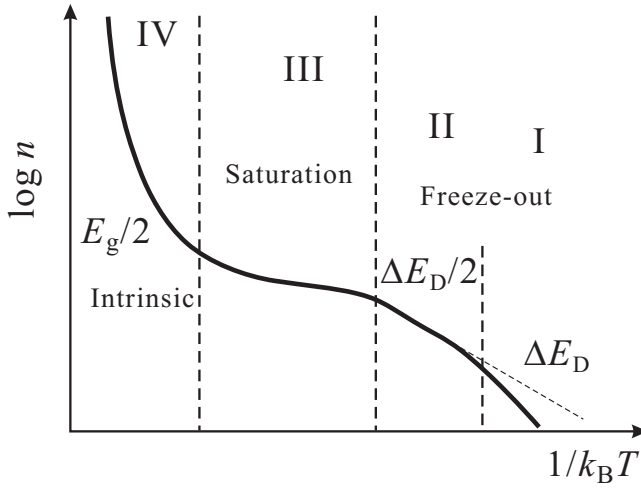
Figure 3.2: Characteristic four temperature regions of an n-type semiconductor with compensation. Schematical temperature dependence of carrier concentration n is plotted versus $1/T$ in semi-log scale.

Remember the semiconductor equation (2.23) then the product np does not depend on the doping. If one of n , p increases with doping, then the other decreases. In the case of n-type semiconductor under $N_D \gg N_A$, n is much higher than p by many orders, and we call the electrons **majority carriers** and the holes **minority carriers**. The other way around in the case of p-type semiconductors.

3.7.2 Majority carriers, minority carriers

Even in the presence of donors and acceptors eq. (2.22a,b) hold and simultaneous satisfaction of them gives n, p and E_F. To obtain E_F with knowledge of n p approximate expressions

$$E_F \approx E_C + k_B T \left[\ln\left(\frac{n}{N_C}\right) + 2^{-3/2}\left(\frac{n}{N_C}\right) \right], \tag{3.13a}$$

$$E_F \approx E_V - k_B T \left[\ln\left(\frac{p}{N_V}\right) + 2^{-3/2}\left(\frac{p}{N_V}\right) \right] \tag{3.13b}$$

are convenient. In the region where (3.11), (3.12) hold the last term can be omitted.

In an n-type semiconductor with compensation, p,n_A can be ignored and the electrically neutral condition is

$$n + N_A = N_D - n_D. \tag{3.14}$$

Substitution of eq.(3.11) gives

$$\frac{n + N_A}{N_D - N_A - n} = \frac{1}{2}\exp\left(\frac{E_D - E_F}{k_F T}\right). \tag{3.15}$$

Equation (2.22) holds for the case of doped semiconductors with shifts of E_F, multiplication of each side of the equation results in

$$\frac{n(n + N_A)}{N_D - N_A - n} = \frac{1}{2}N_c \exp\left(-\frac{\Delta E_D}{k_B T}\right), \quad \Delta E_D \equiv E_c - E_D. \tag{3.16}$$

3.7.3 Temperature dependence of carrier concentration

The temperature dependence of carrier concentration n described by eq.(3.16) has the following four characteristic regions:

1. Impurity (Freeze-out) region I: At very low temperatures and the case of $n \ll N_A \ll N_D$,

$$n \approx \frac{N_D N_c}{2N_A}\exp\left(-\frac{\Delta E_D}{k_B T}\right), \tag{3.17}$$

   where n decreases with lowering the temperature in an Arrhenius type with an activation energy of ΔE_D.

2. Impurity (Freeze-out) region II: In middle temperature range, in the case of $N_A \ll n \ll N_D$,

$$n \approx \left(\frac{N_c N_D}{2}\right)^{1/2}\exp\left(-\frac{\Delta E_D}{2k_B T}\right), \tag{3.18}$$

3-4

where the temperature dependence shows again an Arrhenius type but with a different activation energy, which is a half of that in the impurity region I.

3. Exhaustion (Saturation) region: Temperature is higher than ΔE_D ($k_B T > \Delta E_D$). The exponential function in eq.(3.16) is now almost a constant (~ 1) and

$$n \approx N_D - N_A. \tag{3.19}$$

Electrons once captured in donors are "exhaustively" excited to the conduction band and work as carriers.

4. Intrinsic region: At higher temperatures where direct thermal excitation for the valence band to conduction band cannot be ignored in comparison with N_D, the temperature dependence of the carrier concentration asymptotically approaches to that in an intrinsic semiconductor described as eq.(2.20), (2.24).

4 Electric transport

Electric transport is a most important response of materials together with optical responses. They are responses to some external perturbations and the treatments should be more or less on non-equilibrium physics. Here we have a look at very elementary part of linear response of classical electric transport. We will visit quantum transport in later sections.

4.1 Classical electric transport

We have already seen the Bloch oscillation as a motion of carrier under the application of external electric field, in which the electrons have been treated as coherent quantum waves without scattering from some imperfection or from electromagnetic interaction with vacuum. In real semiconductors, the carriers suffered from frequent scattering by various mechanisms and in a snapshot the sizes of wavefunctions for carriers are very small in real space, which fact allows us to treat the carriers as classical charged particle. Here we consider the electric conductions in which we can consider the carriers as classical particles other than that they obey the Fermi-Dirac statistics.

4.1.1 Boltzmann equation and Drude conductivity

Let us express a dynamic system consisted of classical particles of the same kind as a set of points in a six dimensional **phase space** of r (real coordinate) and p (momentum). We write the ratio of particles that are contained in an infinitesimal volume $drdp$ at the position of (r, p) at time t, as $f(r, p, t)drdp$. f is called a **distribution function** and according to the definition, the integral over the whole phase space should be 1. In the equilibrium or a steady state, the time derivative of f is zero, in other cases non-zero. We put it as $\partial f/\partial t$. When a uniform external force F is applied to each particle (*e.g.* through an electric field),

$$dr/dt = p/m, \qquad F = dp/dt,$$

where m is the (effective) mass of the particles. First we ignore the scatterings which causes jumps in the phase space and the particles in the infinitesimal volume $drdp$ at (r, p) obey the above equation of motion and because they have almost the same spatial positions and momentums initially, they move in the phase space almost in parallel. Hence if we fix our eyes on them also in parallel, there is no extra variation in f due to F.

$$f(r + (p/m)dt, p + Fdt, t + dt) = f(r, p, t).$$

(There is a time evolution due to the movements of particles in real space but that is accounted in the above in the first dt term and to the first order in dt, this equality holds.) Then we consider scatterings, which introduce in/out of particles to this volume and result in variations of f. Let us write the variation as $(\partial f/\partial t)_c dt$, thus

$$f\left(r + \frac{p}{m}dt, p + Fdt, t + dt\right) - f(r, p, t) = \left(\frac{\partial f}{\partial t}\right)_c dt.$$

We obtain the following **Boltzmann equation** with taking the terms to the first order of dt as

$$\frac{\partial f}{\partial t} + \frac{p}{m} \cdot \frac{\partial f}{\partial r} + F \cdot \frac{\partial f}{\partial p} = \left(\frac{\partial f}{\partial t}\right)_c. \tag{3.20}$$

$(\partial f/\partial t)_c dt$ is called **collision term**, the simplest approximation of which is the **relaxation time approximation**. Here we consider a relaxation time τ, with which the system relaxes to the equilibrium distribution f_0. Namely the collision term is written as
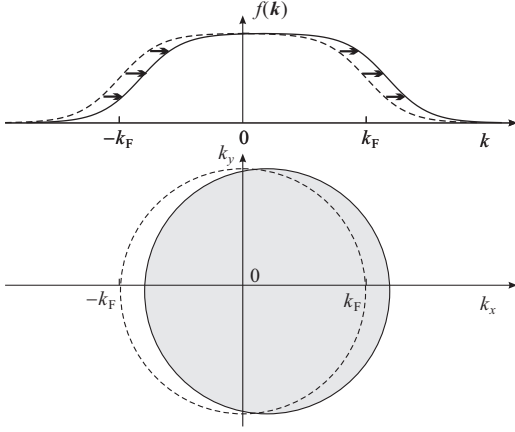
$$(\partial f/\partial t)_c = -(f - f_0)/\tau. \tag{3.21}$$

Figure 3.3: Schematic diagram illustrating a Fermi degenerate electron gas is accelerated in the wavenumber space and the distribution $f(\boldsymbol{k})$ shifts from the equilibrium with the small arrow in the figure. The lower panel shows the shift of "Fermi circle" in the case of two-dimensional system. The shift is very much exaggerated from a realistic sketch.

Macroscopic "currents" in the system are caused as a result of some shifts in f from f_0 in the $\boldsymbol{p}$ space or $\boldsymbol{r}$ space. We call a current driven by the former shift **drift current** whereas the latter **diffusion current**.

First we consider the drift current in a uniform system under uniform external electric field $\mathbf{E}$ and drop $\boldsymbol{r}$ dependence as

$$-e\mathbf{E} \cdot \frac{\partial f}{\partial \boldsymbol{p}} = -\frac{f - f_0}{\tau}, \quad \text{namely} \quad f = f_0 + e\tau\mathbf{E} \cdot \frac{\partial f}{\partial \boldsymbol{p}}.$$

As a first order approximation, we replace f in the second term of RHS with f_0 to get

$$f \approx f_0 + e\tau\mathbf{E} \cdot \frac{\partial f_0}{\partial \boldsymbol{p}} \approx f_0(\boldsymbol{p} + e\tau\mathbf{E}). \tag{3.22}$$

This form of f is obtained with the parallel translation in the $\boldsymbol{p}$ space as $-e\tau\mathbf{E}$. If the system in isotropic in $\boldsymbol{p}$ space and has a simple Fermi sphere, the distribution with f becomes a shifted Fermi sphere as illustrated in Fig.3.3.

In order to obtain the current due to the shift, we need to calculate the integral of $\boldsymbol{v}(\boldsymbol{k})f(\boldsymbol{k})$ over the $\boldsymbol{k}$ space. Without loosing generality we write $\mathbf{E} = (\mathcal{E}_x, 0, 0)$. Then within the above approximation of shifted Fermi sphere, only x-component of the drift current survives and others vanish with symmetrical integration of functions with the odd parity. Now the integration results in

$$\int \frac{d^3k}{(2\pi)^3} \boldsymbol{v}(\boldsymbol{k})\left(f_0 + e\tau\mathbf{E} \cdot \frac{\partial f_0}{\hbar\partial \boldsymbol{k}}\right) = \int \frac{d^3k}{(2\pi)^3} \frac{\hbar k_x}{m} e\tau\mathcal{E}_x \frac{\partial f_0}{\hbar\partial k_x} = \frac{e\mathcal{E}_x}{m} \int \mathscr{D}(E)\tau(E)\frac{\hbar^2 k_x^2}{m}\frac{\partial f_0}{\partial E}dE. \tag{3.23}$$

Here we assume the relaxation time τ only depends on the energy E. $\hbar^2 k_x^2/2m$ is the kinetic energy along x-direction and is $E/3$ due to the equivalent partition law.

If the system is a metal and Fermi-degenerated, $\partial f_0/\partial E$ can be approximated as $-\delta(E - E_\mathrm{F})$. From eq.(2.19a) $\mathscr{D}(E) = A\sqrt{E}$ (A is a material dependent constant and we take the origin of energy at the band edge) and (3.23) is rewritten as

$$\langle v_x \rangle = -A\frac{e\mathcal{E}_x}{m}\frac{2\tau(E_\mathrm{F})}{3}E_\mathrm{F}^{3/2}.$$

The total particle concentration n is given by

$$n = \int_0^{E_\mathrm{F}} \mathscr{D}(E)dE = A\frac{2}{3}E_\mathrm{F}^{3/2},$$

and hence we obtain the electric conductivity $\sigma = j/E = -e\langle v_x \rangle/\mathcal{E}_x$ as

$$\sigma = \frac{e^2 n\tau(E_\mathrm{F})}{m}. \tag{3.24}$$

Equation (3.24) is the well known expression of **Drude conductivity**.

Appendix B: Reciprocal lattice

Consider a one-dimensional periodic potential $V(x) = V(x + Na)$. Because this is a periodic function, we can expand it in a Fourier series with a discrete set of wavenumber as

$$V(x) = \sum_n V_n e^{ik_n x}, \quad k_n = n\frac{2\pi}{a}.$$

These discrete wavenumbers k_n form a one-dimensional **reciprocal lattice**.

Extension to a three dimensional lattice is written as

$$V(\boldsymbol{r}) = \sum_{\boldsymbol{G}} V_{\boldsymbol{G}} e^{i\boldsymbol{G}\boldsymbol{r}}. \tag{B.1}$$

A three dimensional crystal potential is invariant under parallel transformations by lattice vectors. Thus, writing basic lattice vectors as $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$,

$$V(\boldsymbol{r}) = V(\boldsymbol{r} + \boldsymbol{r_n}), \quad \boldsymbol{r_n}(\text{lattice vector}) = n_1\boldsymbol{a}_1 + n_2\boldsymbol{a}_2 + n_3\boldsymbol{a}_3, \tag{B.2}$$

where $\boldsymbol{n} = (n_1, n_2, n_3)$ is an arbitrary set of three integers. The condition for (B.1) to fulfill (B.2) is as follows;

$$\boldsymbol{G} \cdot \boldsymbol{r_n} = 2\pi m \quad (m \text{ is an integer}). \tag{B.3}$$

Now let us try to express $\boldsymbol{G}$ in the form of linear combination of basis vectors $\boldsymbol{g}_1 \quad \boldsymbol{g}_2 \quad \boldsymbol{g}_3$ as

$$\boldsymbol{G} = h_1\boldsymbol{g}_1 + h_2\boldsymbol{g}_2 + h_3\boldsymbol{g}_3, \tag{B.4}$$

with a set of integers (h_1, h_2, h_3) just like the lattice vectors (or in analogy with one-dimension). For this to hold (B.1), we need to take $\boldsymbol{g}_i$ to satisfy

$$\boldsymbol{g}_i \cdot \boldsymbol{a}_j = 2\pi \delta_{ij}. \tag{B.5}$$

Then it is clear that $\boldsymbol{G}$ in (B.4) satisfy the condition (B.3).

Let V_a be the volume of the parallelepiped formed by $\boldsymbol{a}_j$, *i.e.*, $|\boldsymbol{a}_1\boldsymbol{a}_2\boldsymbol{a}_3| = \boldsymbol{a}_1 \cdot (\boldsymbol{a}_2 \times \boldsymbol{a}_3)$ and write as

$$\boldsymbol{g}_1 = 2\pi V_a^{-1} \boldsymbol{a}_2 \times \boldsymbol{a}_3, \tag{B.6}$$

this $\boldsymbol{g}_1$ actually satisfies (B.5) and circulations between 1, 2, 3 give other two.

Appendix C: Scattering matrix elements

Let us estimate the second term in the left hand side of (3.3). By using the form in eq.(1.5) for $|n, \boldsymbol{k}\rangle$, $u_{n'\boldsymbol{k}'}^*(\boldsymbol{r})u_{n\boldsymbol{k}}(\boldsymbol{r})$ is a lattice periodic function and thus can be expanded just like (B.1) with reciprocal lattice vectors $\boldsymbol{G}$ as

$$u_{n'\boldsymbol{k}'}^*(\boldsymbol{r})u_{n\boldsymbol{k}}(\boldsymbol{r}) = \sum_{\boldsymbol{G}} h_{n'\boldsymbol{k}'n\boldsymbol{k}}(\boldsymbol{G}) e^{i\boldsymbol{G}\boldsymbol{r}}. \tag{C.1}$$

We express the coefficient $h_{n'\boldsymbol{k}'n\boldsymbol{k}}$ as

$$h_{n'\boldsymbol{k}'n\boldsymbol{k}}(\boldsymbol{G}) = \int_{\Omega_u} \frac{d^3r}{V_u} e^{-i\boldsymbol{G}\boldsymbol{r}} u_{n'\boldsymbol{k}'}^*(\boldsymbol{r})u_{n\boldsymbol{k}}(\boldsymbol{r}), \tag{C.2}$$

with inverse Fourier transform of (C.1). Ω_u indicates that the integration is performed in a unit cell and V_u is the volume of a unit cell.

On the other hand because $U(\boldsymbol{r})$ does not have the lattice periodicity, we expand it on a continuous variable as

$$U(\boldsymbol{r}) = \int d\boldsymbol{q} U_{\boldsymbol{q}} e^{-i\boldsymbol{q}\boldsymbol{r}}. \tag{C.3}$$

From (C.1) and (C.3), the matrix element can be transformed into

$$\langle n', \boldsymbol{k}' | U | n, \boldsymbol{k} \rangle = \int d\boldsymbol{q} U_{\boldsymbol{q}} \sum_{\boldsymbol{G}} h_{n'\boldsymbol{k}'n\boldsymbol{k}}(\boldsymbol{G}) \int d\boldsymbol{r} e^{i(\boldsymbol{k}-\boldsymbol{k}'+\boldsymbol{q}+\boldsymbol{G})\boldsymbol{r}}$$

$$= \int d\boldsymbol{q} U_{\boldsymbol{q}} \sum_{\boldsymbol{G}} h_{n'\boldsymbol{k}'n\boldsymbol{k}}(\boldsymbol{G})(2\pi)^3 \delta(\boldsymbol{k} - \boldsymbol{k}' + \boldsymbol{q} + \boldsymbol{G}) = (2\pi)^3 \sum_{\boldsymbol{G}} U_{\boldsymbol{k}'-\boldsymbol{k}-\boldsymbol{G}} h_{n'\boldsymbol{k}'n\boldsymbol{k}}(\boldsymbol{G}). \tag{C.4}$$

Because we have assumed that U is much more gradual function than the lattice periodic potential, the Fourier component $\boldsymbol{q}$ of U should not contain a vector as large as a reciprocal lattice vector $\boldsymbol{G}$. That is $\boldsymbol{k}' - \boldsymbol{k} - \boldsymbol{G} \ll \boldsymbol{G}$. Furthermore the change in the momentum due to the scattering by $U(\boldsymbol{r})$ is smaller than that by lattice potential. As a result in the expansion only $\boldsymbol{G} = \boldsymbol{0}$ is allowed. From (C.2),

$$\langle n', \boldsymbol{k}' | U | n, \boldsymbol{k} \rangle = U_{\boldsymbol{k}'-\boldsymbol{k}} h_{n'\boldsymbol{k}'n\boldsymbol{k}}(\boldsymbol{0}) = U_{\boldsymbol{k}'-\boldsymbol{k}} \int_{\Omega_u} \frac{d^3r}{V_u} u_{n'\boldsymbol{k}'}^*(\boldsymbol{r})u_{n\boldsymbol{k}}(\boldsymbol{r}) \approx U_{\boldsymbol{k}'-\boldsymbol{k}} \delta_{n'n}. \tag{C.5}$$

At the last line, we have used the approximation that the interband mixing is ignorable. This is justified because u_n are the solutions of (1.4) and hence orthogonal to each other, normalized in the unit cell and $\boldsymbol{k} \approx \boldsymbol{k}'$.

Appendix D: Wannier function

In Sec.3.6.1, we have introduced the effective mass approximation with expanding the eigenstates with the basis of Bloch wave functions. The same can be done with **Wannier functions**, which are given from the tight-binding approach. Here we roughly draw the senario.

Wannier functions $w_n(\boldsymbol{r} - \boldsymbol{R}_i)$ (n is the band index, $\boldsymbol{R}_i$ are the lattice vector) is defined as follows and related with Bloch functions with the Fourier transform.

$$w_n(\boldsymbol{r} - \boldsymbol{R}_i) = N^{-1/2} \sum_{\boldsymbol{k}} e^{-i\boldsymbol{k}\boldsymbol{R}_i} \psi_{n\boldsymbol{k}}(\boldsymbol{r}), \tag{D.1a}$$

$$\psi_{n\boldsymbol{k}}(\boldsymbol{r}) = N^{-1/2} \sum_{\boldsymbol{R}_i} e^{i\boldsymbol{k}\boldsymbol{R}_i} w_n(\boldsymbol{r} - \boldsymbol{R}_i). \tag{D.1b}$$

They have the following orthogonality (the proof is skipped).

$$\int d^3r w_{n'}^*(\boldsymbol{r} - \boldsymbol{R}_j') w_n(\boldsymbol{r} - \boldsymbol{R}_j) = \delta_{jj'} \delta_{nn'}. \tag{D.2}$$

$w_n(\boldsymbol{r} - \boldsymbol{R}_i)$ is spatially localied at the site $\boldsymbol{R}_i$, which can be confirmed as follows. We approximate the Bloch function in (D.1a) with neglecting the $\boldsymbol{k}$-dependence of the lattice periodic part $u_{n\boldsymbol{k}}(\boldsymbol{r})$. Then

$$w_n(\boldsymbol{r} - \boldsymbol{R}_i) = N^{-1/2} u_{n0}(\boldsymbol{r}) \sum_{\boldsymbol{k}} e^{i\boldsymbol{k}(\boldsymbol{r} - \boldsymbol{R}_i)}.$$

The summation over $\boldsymbol{k}$ clearly indicate the localization around $\boldsymbol{r} = \boldsymbol{R}_i$.

Let us consider the problem in (3.1). The expansion of $\eta(\boldsymbol{r})$ with Wannier functions is written as

$$\eta(\boldsymbol{r}) = N^{-1/2} \sum_j F(\boldsymbol{R}_j) w(\boldsymbol{r} - \boldsymbol{R}_j), \tag{D.3}$$

where the band index is dropped for simplicity. Substituting the above to (3.1) and taking the inner product with $w(\boldsymbol{r} - \boldsymbol{R}_l)$ just as in the case of Bloch function expansion, we obtain

$$\sum_j \int d^3r w^*(\boldsymbol{r} - \boldsymbol{R}_l)[\mathscr{H}_0 + U(\boldsymbol{r}) - E]F(\boldsymbol{R}_j) w(\boldsymbol{r} - \boldsymbol{R}_j). \tag{D.4}$$

Due to the localized nature of $w(\boldsymbol{r} - \boldsymbol{R}_l)$ the integration region can be done just around $\boldsymbol{R}_l$. Because we have assumed the slow variation of $U(\boldsymbol{r})$ in space, we replace it with $U(\boldsymbol{R}_l)$ and put it out from [ ] together with the third term from the integral. They give the term $[U(\boldsymbol{R}_l) - E]F(\boldsymbol{R}_l)$. Now we evaluate the term of $\mathscr{H}_0$ by using (D.1a) as

$$\sum_j \int d^3r w^*(\boldsymbol{r} - \boldsymbol{R}_l)\mathscr{H}_0 F(\boldsymbol{R}_j) w(\boldsymbol{r} - \boldsymbol{R}_j) = \sum_{j,\boldsymbol{k},\boldsymbol{k}'} N^{-1} F(\boldsymbol{R}_j) e^{i(\boldsymbol{k}'\boldsymbol{R}_l - \boldsymbol{k}\boldsymbol{R}_j)} \int \psi_{\boldsymbol{k}'}^*(\boldsymbol{r}) \mathscr{H}_0 \psi_{\boldsymbol{k}}(\boldsymbol{r})$$

$$= \sum_{j,\boldsymbol{k}} N^{-1} F(\boldsymbol{R}_j) E_0(\boldsymbol{k}) e^{i\boldsymbol{k}(\boldsymbol{R}_l - \boldsymbol{R}_j)} = \sum_{m,\boldsymbol{k}} N^{-1} F(\boldsymbol{R}_l - \boldsymbol{R}_m) E_0(\boldsymbol{k}) e^{i\boldsymbol{k}\boldsymbol{R}_m}. \tag{D.5}$$

The energy band dispersion relation $E_0(\boldsymbol{k})$ is a reciprocal lattice periodic function and can be Fourier expanded with the lattice vectors $\boldsymbol{R}_j$ as

$$E_0(\boldsymbol{k}) = \sum_j A(\boldsymbol{R}_j) e^{-i\boldsymbol{k}\boldsymbol{R}_j}. \tag{D.6}$$

Substitution of the above to (D.5) results in

$$(D.5) = \sum_{m,j,\boldsymbol{k}} N^{-1} F(\boldsymbol{R}_l - \boldsymbol{R}_m) A(\boldsymbol{R}_j) e^{i\boldsymbol{k}(\boldsymbol{R}_m - \boldsymbol{R}_j)}$$

$$= \sum_{\boldsymbol{k},j} N^{-1} F(\boldsymbol{R}_l - \boldsymbol{R}_m) A(\boldsymbol{R}_j) N \delta_{\boldsymbol{R}_m \boldsymbol{R}_j} = \sum_m A(\boldsymbol{R}_m) F(\boldsymbol{R}_l - \boldsymbol{R}_m).$$

$\exp(\boldsymbol{r} \cdot \nabla)$ is a creation operator of spatial translation by $\boldsymbol{r}$ and hence $F(\boldsymbol{R}_l - \boldsymbol{R}_m) = \exp(-\boldsymbol{R}_m \cdot \nabla)F(\boldsymbol{R}_l)$. Substitution of this into the above equation results in

$$(D.5) = \sum_m A(\boldsymbol{R}_m) \exp(-\boldsymbol{R}_m \cdot \nabla) F(\boldsymbol{R}_l). \tag{D.7}$$

The operator part on $F(\boldsymbol{R}_l)$ in this equation is formally the same as (D.6) if we replace $\boldsymbol{k} \to -i\nabla$. Hence we can write

$$(\text{D.5}) = E_0(-i\nabla)F(\boldsymbol{R}_l). \qquad (\text{D.8})$$

With the above results eq.(D.4) is expressed as

$$[E_0(-i\nabla) + U(\boldsymbol{R}_l)]F(\boldsymbol{R}_l) = EF(\boldsymbol{R}_l). \qquad (\text{D.9})$$

We approximate E_0 with $\hbar^2 \boldsymbol{k}^2/2m^*$ by using the effective mass m^* and replace the lattice vector $\boldsymbol{R}_l$ with a continuous variable $\boldsymbol{r}$. Finally

$$\left[-\frac{\hbar^2 \nabla^2}{2m^*} + U(\boldsymbol{r}) \right] F(\boldsymbol{r}) = EF(\boldsymbol{r}) \qquad (\text{D.10})$$

is obtained, which is an Schrödinger type equation in the same form of (3.7).