Part IV: Scattering and Interactions

- Generalities about scattering in semiconductors
- Electron-phonon interaction
- Scattering with ionized impurities
- Coulomb interactions among free carriers
- Radiative processes



Generalities about scattering in semiconductors

We have so far considered ideal crystals at equilibrium and the elementary excitations associated with the ionic degrees of freedom (phonons) and the electronic degrees of freedom (plasmons). But we have never considered the possible interactions between these degrees of freedom. If we recall the discussion we had when dealing with the crystal Hamiltonian (see pages 46-47 of the Notes), we have so far considered only electronic states computed assuming the ions frozen in their equilibrium positions. We have later considered the ionic states associated with the excitation of ions away from equilibrium (pages 150-156), but we have not considered the effects of this displacement on the structure of the electronic states. Using first-order perturbation theory (or, equivalently, a linear response approximation), we shall describe these effects as 'collisions' between electrons and phonons.

Collisions are indeed crucial phenomena in crystals which allow us to understand most of their properties at finite temperatures. We shall be mostly interested in charge transport. In this context, collisions constitute the factors limiting the conductivity of the crystals. In addition to electron-phonon collisions we shall also consider the 'scattering' suffered by electrons in the field of an ionized impurity (such a dopant), the 'transitions' (more than scattering or collisions) induced by electromagnetic fields (electron-photon or 'radiative' processes), and electron-electron collisions which result from flucutations of the (Hartree or Hartree-Fock) mean field.

Fermi Golden rule will be our main tool to evaluate the frequency at which these processes occur. As we saw before, if $\mathbf{H}'(\mathbf{r},t) \sim H_{\mathbf{q}} e^{i\mathbf{q}\cdot\mathbf{r}} e^{i\omega t}$ is the perturbation, then the frequency at which this perturbation induces a transition from a Bloch state $|n\mathbf{k}\rangle$ in band n to a Bloch state $|n'\mathbf{k}'\rangle$ in band n' is given by:

$$\frac{1}{\tau(n\mathbf{k})} = \frac{2\pi}{\hbar} \sum_{\mathbf{k}'} |\langle n'\mathbf{k}'|\mathbf{H}'|n\mathbf{k}\rangle|^2 \,\delta[E_n(\mathbf{k}) - E_{n'}(\mathbf{k}') + \hbar\omega] \,. \tag{540}$$

Let's consider the matrix element $< n' \mathbf{k}' | \mathbf{H}' | n \mathbf{k} >$, since this is different from the 'usual' matrix element between plane waves. Recalling the expression for Bloch waves

$$<\mathbf{r}|n\mathbf{k}> = \frac{1}{V^{1/2}} e^{i\mathbf{k}\cdot\mathbf{r}} u_{\mathbf{k}}^{(n)}(\mathbf{r}) = \frac{1}{V^{1/2}} e^{i\mathbf{k}\cdot\mathbf{r}} \sum_{\mathbf{G}} u_{\mathbf{k}+\mathbf{G}}^{(n)} e^{i\mathbf{G}\cdot\mathbf{r}},$$
 (541)

we can write:

$$< n'\mathbf{k}' |\mathbf{H}'| n\mathbf{k} > = \frac{1}{V} \int d\mathbf{r} \sum_{\mathbf{q}} e^{i(\mathbf{k} - \mathbf{k}' + \mathbf{q}) \cdot \mathbf{r}} H_{\mathbf{q}} u_{\mathbf{k}'}^{(n')*}(\mathbf{r}) u_{\mathbf{k}}^{(n)}(\mathbf{r}) =$$

$$\frac{1}{V} \sum_{\mathbf{q}} \sum_{i(\mathbf{k} - \mathbf{k}' + \mathbf{q}) \cdot \mathbf{R}_{i}} u_{\mathbf{k}} \int d\mathbf{r} \sum_{\mathbf{q}} e^{i(\mathbf{k} - \mathbf{k}' + \mathbf{q}) \cdot \mathbf{\rho}} (n') e^{i(\mathbf{k} - \mathbf{k}' + \mathbf{q}$$

$$= \frac{1}{V} \sum_{\mathbf{q}} \sum_{l} e^{i(\mathbf{k}-\mathbf{k}'+\mathbf{q})\cdot\mathbf{R}_{l}} H_{\mathbf{q}} \int_{\Omega_{cell}} d\rho \ e^{i(\mathbf{k}-\mathbf{k}'+\mathbf{q})\cdot\rho} u_{\mathbf{k}'}^{(n')*}(\rho) \ u_{\mathbf{k}}^{(n)}(\rho) , \qquad (542)$$

having set $\mathbf{r} = \rho + \mathbf{R}_l$. The sum over cells indices l yields (see page 101):

$$\sum_{l} e^{i(\mathbf{k}-\mathbf{k}'+\mathbf{q})\cdot\mathbf{R}_{l}} = N_{cell} \sum_{\mathbf{G}} \delta(\mathbf{k}-\mathbf{k}'+\mathbf{q}+\mathbf{G}) , \qquad (543)$$

so that:

$$< n'\mathbf{k}'|\mathbf{H}'|n\mathbf{k} > = \sum_{\mathbf{G}} H_{\mathbf{k}-\mathbf{k}'+\mathbf{G}} \frac{1}{\Omega_{cell}} \int_{\Omega_{cell}} d\rho \ e^{i\mathbf{G}\cdot\rho} \ u_{\mathbf{k}'}^{(n')*}(\rho) \ u_{\mathbf{k}}^{(n)}(\rho) =$$

$$\sum_{\mathbf{G},\mathbf{n}'} H_{\mathbf{k}-\mathbf{k}'+\mathbf{G}} \ \mathcal{I}_{n',\mathbf{k}';n,\mathbf{k}} , \qquad (544)$$

having used the procedure followed at page 101 to reach Eq. (173) in getting the last expression for the integral involving Bloch functions. Notice that the matrix element conserves momentum – as for electrons in free space – but only up to a wavevector \mathbf{G} of the reciprocal lattice, consistent with the fact that \mathbf{k} is the 'crystal' momentum. The factor \mathcal{I} is called the 'overlap factor' since it is the overlap integral over the WS cell of the initial and final Bloch components. For perturbation of very long wavelength it is easy to see that $\mathcal{I} \sim 1$, so it can be ignored. It is also customary to separate the contributions with $\mathbf{G} = 0$ in Eq. (544) from those with $\mathbf{G} \neq 0$. The former are called *Normal* processes, the latter (which involve a momentum transfer $\mathbf{q} = \mathbf{k}' - \mathbf{k}$ outside the first BZ, so that if must be remapped into the zone by adding a \mathbf{G} vector) are called *Umklapp* processes. Also, processes for which n = n' are called *intraband* transitions, while *interband* processes are, quite obvioulsy, those for which

 $n \neq n'$. Finally, in many-valley semiconductors (for example, think of the 6 minima of the conduction band in Si), intraband processes such that the initial and final states, $|\mathbf{k}\rangle$ and $|\mathbf{k}'\rangle$ are close to the same minimum ('in the same valley') are said to be *intra-valley* processes, while when $|\mathbf{k}\rangle$ and $|\mathbf{k}'\rangle$ belong to different valleys the process is said to be an *inter-valley* process. Static perturbations ($\omega = 0$) indude *elastic* transitions in which the energies of the initial and final states, $E_n(\mathbf{k})$ and $E_{n'}(\mathbf{k}')$, are equal. In inelastic process, instaed, energy is lost or gained during the scattering process. Scattering with the Coulomb field of an ionized impurity is a typical example of elastic process, emission of a phonon a typical inelastic process.

We have so far considered Eq. (540) in the context of 'conventional' Quantum Mechanics. However, Fermi Golden Rule applies equally to the many-body situation tackled by second quatization. The only difference will be the nature of the final and initial states which will be vectors in Fock space, as we shall see below.



Electron-Phonon interactions

We now consider the interaction between electrons and phonons. Basically there are two types of interactions: The first, called 'deformation potential' interaction, is due to the effect which displacement of the ionic potential causes on the electron energy. It is also called a 'nonpolar' interaction, because the electrons do not feel any long-range electric (dipole) field, but feel simply the perturbation caused by the fact that the ions in the electronic Hamiltonian are not frozen in their equilibrium positions $\mathbf{R}_{l\gamma}$. Both acoustic and optical phonons cause this interaction, although there are subtle differences between the two. A second type of interaction is, instead, of a 'polar' nature: Phonons, especially optical phonons, cause a long-range dipole field which we can be treated much in the same way as we treated plasmons. Electrons are scattered by this field which oscillates in time and space as $e^{i\mathbf{q}\cdot\mathbf{r}}e^{i\omega\mathbf{q}t}$, where $\hbar\omega_{\mathbf{q}}$ is the dispersion of the optical phonons cau also induce an electric field in piezoelectric materials (such as the quartz form of crystalline SiO₂, as is well known from quartz watches). Therefore, this is called the 'piezoelectroc interaction'. We shall discuss below in some detail the nonpolar interaction between electrons and acoustic phonons and polar Frölich scattering between electrons and longitudinal optical phonons.

• Nonpolar (deformation potential) interaction

We have seen before that the electronic component of the electronic Hamiltonian, Eq. (88) of the Notes, Part 1, can be written as (see Eqns. (353) and (364) of the Notes, Part 3):

$$\mathbf{H}_{el} = \int d\mathbf{r} \left[\left(-\frac{i\hbar}{2m} \right) \nabla \pi(\mathbf{r}) \cdot \nabla \psi(\mathbf{r}) - \frac{1}{\hbar} \pi(\mathbf{r}) V(\mathbf{r}) \psi(\mathbf{r}) \right] = \sum_{n\mathbf{k}} E_{n\mathbf{k}} c_{n\mathbf{k}}^{\dagger} c_{n\mathbf{k}} , \qquad (545)$$

while the ionic component takes the form (see Eq. (121) and (132) of the Notes, Part 3):

$$\mathbf{H}_{ion} = \frac{1}{2} \sum_{l\gamma} \left[\frac{P_{l\gamma}^2}{M_{\gamma}} + \sum_{l' \neq l, \gamma'\gamma} \sum_{ij} \frac{\partial^2 V_{\gamma'}^{(lat)} (\mathbf{R}_{l\gamma}^{(0)} - \mathbf{R}_{l'\gamma'}^{(0)})}{\partial R_{l\gamma i} \partial R_{l'\gamma' j}} \, \delta R_{l\gamma i} \, \delta R_{l'\gamma' j} \right] = \sum_{\eta \mathbf{q}} \, \hbar \omega_{\mathbf{q}\eta} \, b_{\mathbf{q}\eta}^{\dagger} b_{\mathbf{q}\eta} \, A_{\mathbf{q}\eta} \, \delta R_{l'\gamma' j}$$
(546)

We must now find the effect of the displacement of the ions away from their equilibrium position on the electron energy.

Proceeding as we have done in dealing with phonons, we consider the change of the ionic potential up to terms which depend linearly on the ionic displacement:

$$\delta V(\mathbf{r}) \approx \sum_{l\gamma} \nabla V_{\gamma}(\mathbf{r} - \mathbf{R}_{l\gamma}) \cdot \delta \mathbf{R}_{l\gamma} .$$
 (547)

Note that here we are assuming that, when displaced, the ionic potential does not change but simply shifts rigidly. This is known as the 'rigid ion' approximation and it ignores the fact that valence electrons will re-distribute themselves as the ions move away from their equilibrium positions (which is in essence dielectric screening). We may consider this as a higher-order effect ignored in our linear-response picture. But, clearly, this remains an approximation. Let's express the displacement of the ion γ in cell l in terms of the creation and annihilation operators of phonons of momentum q and branch η via Eqns. (358) and (362) of the Notes, Part 3 (see also Eq. (402)):

$$\delta V(\mathbf{r}) \approx \sum_{l\gamma} \frac{1}{(NM_{\gamma})^{1/2}} \nabla V_{\gamma}(\mathbf{r} - \mathbf{R}_{l\gamma}) \cdot \sum_{\mathbf{q}\eta} \mathbf{e}_{\mathbf{q}\gamma}^{(\eta)} Q_{\mathbf{q}\gamma} e^{i\mathbf{q}\cdot\mathbf{R}_{l}} =$$

$$= \sum_{\mathbf{q}\eta} \sum_{l\gamma} \frac{1}{(NM_{\gamma})^{1/2}} \left(\frac{\hbar}{2\omega_{\mathbf{q}\eta}}\right)^{1/2} (b_{\mathbf{q}\eta} + b_{\mathbf{q}\eta}^{\dagger}) \mathbf{e}_{\mathbf{q}\gamma}^{(\eta)} \cdot \nabla V_{\gamma}(\mathbf{r} - \mathbf{R}_{l\gamma}) e^{i\mathbf{q}\cdot\mathbf{R}_{l}}.$$
(548)

Let's now express the ionic potential in terms of its Fourier components:

$$V_{\gamma}(\mathbf{r} - \mathbf{R}_{l\gamma}) = V_{\gamma}(\mathbf{r} - \mathbf{R}_{l} - \tau_{\gamma}) = \sum_{\kappa} V_{\kappa\gamma} e^{i\kappa \cdot (\mathbf{r} + \mathbf{R}_{l} + \tau_{\gamma})}, \qquad (549)$$

so that

$$\nabla V_{\gamma}(\mathbf{r} - \mathbf{R}_{l} - \tau_{\gamma}) = i \sum_{\kappa} \kappa V_{\kappa,\gamma} e^{i\kappa \cdot (\mathbf{r} + \mathbf{R}_{l})}, \qquad (550)$$

having absorbed the factor $e^{i\kappa\cdot au\gamma}$ inside the Fourier component $V_{\kappa,\gamma}$.

The correction to the energy of the electrons due to the displacement Eq. (548) results simply from the electrostatic energy of the electron density $\rho(\mathbf{r}) = \psi^{\dagger}(\mathbf{r})\psi(\mathbf{r})$ interacting with the potential energy $\delta V(\mathbf{r})$. Thus:

$$\mathbf{H}_{ep} = \int d\mathbf{r} \ \psi^{\dagger}(\mathbf{r}) \ \delta V(\mathbf{r}) \ \psi(\mathbf{r}) , \qquad (551)$$

which should be regarded as the perturbation term due to the electron-photon coupling. Using the Fourier expansions

$$\psi(\mathbf{r}) = \frac{1}{V^{1/2}} \sum_{n\mathbf{k}} c_{n\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{r}} u_{\mathbf{k}}(\mathbf{r}) , \qquad (552)$$

and

$$\psi^{\dagger}(\mathbf{r}) = \frac{1}{V^{1/2}} \sum_{n\mathbf{k}} c_{n\mathbf{k}}^{\dagger} e^{-i\mathbf{k}\cdot\mathbf{r}} u_{\mathbf{k}}^{*}(\mathbf{r}) , \qquad (553)$$

we have (omitting the band-index n for simplicity):

$$\mathbf{H}_{ep} = \frac{1}{V} \sum_{\mathbf{k}\mathbf{k}'} \sum_{\mathbf{q}\eta} \sum_{l\gamma} \frac{1}{(NM_{\gamma})^{1/2}} \left(\frac{\hbar}{2\omega_{\mathbf{q}\eta}}\right)^{1/2} e^{i\mathbf{q}\cdot\mathbf{R}_l} c_{\mathbf{k}'}^{\dagger} \left(b_{\mathbf{q}\eta} + b_{\mathbf{q}\eta}^{\dagger}\right) c_{\mathbf{k}} \mathbf{e}_{\mathbf{q}\gamma}^{(\eta)} \cdot (i\kappa) V_{\kappa,\gamma} \times \int d\mathbf{r} \ e^{i\kappa\cdot(\mathbf{r}-\mathbf{R}_l)} u_{\mathbf{k}'}^*(\mathbf{r}) u_{\mathbf{k}}(\mathbf{r}) \ e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}} .$$
(554)

Now notice that, similarly to Eq. (543):

$$\sum_{l} e^{i(\mathbf{q}-\kappa)\cdot\mathbf{R}_{l}} = N_{cell} \sum_{\mathbf{G}} \delta(\mathbf{q}-\kappa+\mathbf{G}) , \qquad (555)$$

while the integral involving Bloch functions vanishes unless $\mathbf{k}' - \mathbf{k} + \kappa = \mathbf{G}$ (see Eq. (173), page 101 of the Notes). Thus:

$$\mathbf{H}_{ep} = \sum_{\mathbf{kqG}} \left(\frac{N}{M_{cell}}\right)^{1/2} \sum_{\gamma\eta} \left(\frac{M_{cell}}{M_{\gamma}}\right)^{1/2} \left(\frac{\hbar}{2\omega_{\mathbf{q}\eta}}\right)^{1/2} i(\mathbf{q} + \mathbf{G}) \cdot \mathbf{e}_{\mathbf{q}\gamma}^{(\eta)} V_{\mathbf{q}+\mathbf{G},\gamma} \mathcal{I}_{\mathbf{k}+\mathbf{q}+\mathbf{G},\mathbf{k}} \times c_{\mathbf{k}+\mathbf{q}+\mathbf{G}}^{\dagger} (b_{\mathbf{q}\eta} + b_{\mathbf{q}\eta}^{\dagger}) c_{\mathbf{k}}.$$
(556)

In order to simplify the notation, we shall make the convention that all phonon wavevectors must be mapped into the first BZ, whenever necessary. Let's define the electron-phonon coupling constant as:

$$\mathcal{A}_{\mathbf{q}\eta} = \sum_{\mathbf{G}\gamma} \left(\frac{\hbar}{2\rho_x \omega_{\mathbf{q}\eta}}\right)^{1/2} i(\mathbf{q} + \mathbf{G}) \cdot \mathbf{e}_{\mathbf{q}\gamma}^{(\eta)} V_{\mathbf{q}+\mathbf{G},\gamma} , \qquad (557)$$

having absorbed the factor $(M_{cell}/M_{\gamma})^{1/2}$ as a different normalization of the Fourier transform of the ionic potential (so that it's normalized to the atomic volume rather than to the cell volume) and having converted the wavefunction (or field) normalization from a box-normalization to the infinite-volume normalization, so that the crystal mass density ρ_x appears in this expression. Thus,

$$\mathbf{H}_{ep} = \sum_{\mathbf{kq}\eta} \mathcal{A}_{\mathbf{q}\eta} \mathcal{I}_{\mathbf{k}+\mathbf{q},\mathbf{k}} c^{\dagger}_{\mathbf{k}+\mathbf{q}} \left(b_{\mathbf{q}\eta} + b^{\dagger}_{\mathbf{q}\eta} \right) c_{\mathbf{k}} .$$
(558)

Note the physical meaning of Eq. (558): It represents the annihilation of an electron of momentum \mathbf{k} , the creation (the term $b_{\mathbf{q}\eta}^{\dagger}$) or annihilation (the term $b_{\mathbf{q}\eta}$) of a phonon of branch η and momentum \mathbf{q} , and, finally, the creation of an electron with momentum $\mathbf{k} + \mathbf{q}$. Thus, it can be viewed as the transfer of momentum \mathbf{q} from or to a phonons to/from an electron. Note that the only non-zero matrix elements of the Hamiltonian Eq. (558) are those which differ by one phonon and having an electron state $|\mathbf{k}\rangle$ swapped with a state $|\mathbf{k} + \mathbf{q}\rangle$. For absorption processes only terms of the following form will be nonzero (ignoring the branch-index η to simplify the notation):

$$\mathcal{A}_{\mathbf{q}} < N_{q} - 1 |b_{\mathbf{q}}| N_{\mathbf{q}} > < n_{\mathbf{k}} + \mathbf{q} + 1; n_{\mathbf{k}} - 1 |c_{\mathbf{k}+\mathbf{q}}^{\dagger} c_{\mathbf{k}}| n_{\mathbf{k}+\mathbf{q}}, n_{\mathbf{k}} > = \mathcal{A}_{\mathbf{q}} \sqrt{(1 - n_{\mathbf{k}+\mathbf{q}}) n_{\mathbf{k}} N_{\mathbf{q}}},$$
(559)

while for emission processes the only non-vanishing matrix elements will be of the form:

$$\mathcal{A}_{\mathbf{q}} < N_{q} + 1|b_{\mathbf{q}}^{\dagger}|N_{\mathbf{q}} > < n_{\mathbf{k}} + \mathbf{q} + 1; n_{\mathbf{k}} - 1|c_{\mathbf{k}+\mathbf{q}}^{\dagger}c_{\mathbf{k}}|n_{\mathbf{k}+\mathbf{q}}, n_{\mathbf{k}} > = \mathcal{A}_{\mathbf{q}} \sqrt{(1 - n_{\mathbf{k}+\mathbf{q}})n_{\mathbf{k}}(1 + N_{\mathbf{q}})}$$
(560)

having indicated with $n_{\mathbf{k}}$ the electron number and with $N_{\mathbf{q}}$ the phonon number.

Regarding the normalization of Fock states for bosons: Consider the state containing two phonons, $|2\rangle = b^{\dagger}b^{\dagger}|0\rangle$, having omitted the index q. Let's compute its norm using the commutation rule $bb^{\dagger} = 1 + b^{\dagger}b$:

$$<2|2> = <0|bbb^{\dagger}b^{\dagger}|0> = <0|b(1+b^{\dagger}b)b^{\dagger}|0> < 0|bb^{\dagger}|0> + <0|bb^{\dagger}bb^{\dagger}|0> = = 1 + <0|bb^{\dagger}(1+b^{\dagger}b)|0> = 1 + <0|bb^{\dagger}|0> = 2.$$
(561)

In general one can follow the same procedure to find (try it with $\langle 3|3 \rangle$):

$$\langle N|N \rangle = N! . \tag{562}$$

Now consider, for example, the term < N - 1|b|N > in Eq. (559) above:

$$< N - 1|b|N > = \frac{1}{[(N-1)!N!]^{1/2}} < 0|b^{N-1}bb^{\dagger N}|0 > = \frac{1}{[(N-1)!N!]^{1/2}} < 0|b^Nb^{\dagger N}|0 > =$$

$$= \left(\frac{N!N!}{(N-1)!N!}\right)^{1/2} < N|N\rangle = N^{1/2} .$$
(563)

Note that the terms involving the electron number simply represent the effect of 'Pauli blocking' (we can create an electron with wavevector $\mathbf{k} + \mathbf{q}$ only if that state is available (that is: $n_{\mathbf{k}} + \mathbf{q} = 0$). More interesting is the appearance of the phonon numbers (which arise from the normalization mentioned at page 147 of the Notes): When squared inside Fermi Golden Rule, they give raise to the 'Einstein' coefficients $N_{\mathbf{q}}$ for absorption, $N_{\mathbf{q}} + 1$ for emission: Stimulated emission is simply a result of the commutation rules!

• Deformation potential and Harrison interaction.

The problem remains of calculating the 'coupling constant' $\mathcal{A}_{\mathbf{q}}$. In principle, it may be computed from the known ionic (pseudo)potentials V_{γ} . The task is not easy and these calculations have been performed numerically only in the past two decades, thanks to the increased computing power invailable. However, Bardeen and Shockley in 1955 have proposed an elegant method to estimate the electron-phonon matrix element when electrons are close to the bottom of the conduction band (or holes close to the top of the valence band) from known parameters known as 'deformation potentials' (hence the name of this type of interaction).

When we compress or dilate a crystal, say by changing the lattice constant $a \rightarrow a + u$, the band structure changes. In particular, the energy of the conduction-band minimum moves by an amount

$$\Delta E_c = E_c(a+u) - E_c(a) \approx \frac{dE_c}{da} u .$$

In 3D, ΔE_c is proportional to the change of the volume ΔV which, for a lattice displacement \mathbf{u} is given by $\Delta V = V \nabla \cdot \mathbf{u}$, so that

$$\Delta E_c = V \frac{dE_c}{dV} \nabla \cdot \mathbf{u} \sim q \Delta_{ac} .$$
(564)

The constant $V dE_c/dV$ is called 'deformation potential' for the conduction band. Let's denote it here by Δ_{ac} , although, depending on the particular valley and valence/conduction band considered, several other symbols are

used in the literature (such as E_1 or Ξ).

If we consider phonons of wavelength sufficiently long (that is, spanning many cells), we can approximate the effect of the density-wave associated with acoustic phonons (consisting of alternating crests/troughs of compression/dilation/compression/dilation....) as having locally the same effect of a uniform, global compression or dilation of the crystal. So we can identify the uniform displacement \mathbf{u} above with the ionic displacement $\delta \mathbf{R}_{l\gamma}$ and we can express the local shift of the CB-edge by the polarization $\xi_{\mathbf{q}}$. This 'fluctuating' CB-edge is what causes the electrons to scatter. Thus, the squared matrix element for electron/acoustic-phonon scattering has the form:

$$|<\mathbf{k}'|H_{ac}|\mathbf{k}>|^2 = \frac{\hbar\Delta_{ac}^2q^2}{2V\rho\omega_{\mathbf{q}}}\left(n_{\mathbf{q}}+\frac{1}{2}\pm\frac{1}{2}\right)$$
,

where ρ is the density of the crystal, the 'plus' sign refers to emissions, the 'minus' sign to absorption processes. Note that the frequency of acoustic phonons (in Si there are 2 transverse modes – TA – and a single longitudinal mode – LA) can be well approximated by a linear relation $\omega_{\mathbf{q}} \approx c_s q$, where c_s is the sound velocity. Therefore the matrix element grows with momentum transfer and we see that scattering with acoustic phonons is a 'large-angle scattering', very effective in 'randomizing' the electron direction and momentum. This has a great effect on the mobility. A more correct expression for the deformation potential interaction in the ellipsoidal valleys in Si and Ge has been given by Herring and Vogt in 1955. In terms of the uniaxial shear and dilation deformation potentials Ξ_u and Ξ_d respectively, the isotropic deformation potential Δ_{ac} appearing in Eq. (564) above is replaced by:

$$\Delta_{LA} = \Xi_d + \Xi_u \cos^2 \theta_q , \qquad (565)$$

for collisions assited by LA phonons and

$$\Delta_{LA} = \Xi_u \cos \theta_q \sin \theta_q , \qquad (566)$$

where θ_q is the angle between the longitudinal axis of the ellipsoid and the phonon momentum \mathbf{q} .

A similar expression, due to Harrison, holds for optical phonons simply replacing the frequency $\omega_{\mathbf{q}}$ and deformation potential Δ_{ac} with their 'optical' counterparts ω_{op} and Δ_{op} . Note that the main difference lies in the fact while for acoustic strain the energy-change depends on the gradient of the displacement (the strain $\propto \nabla \cdot \mathbf{u}$), for optical distortions of the lattice the energy change is affected mainly by the displacement itself, so

the q-dependence in Eq. (564) above disappears. In the literature usually one finds many 'optical deformation potentials'. Usually the notation $(D_t K)_{op}$ is used, defined as $\Delta_{op}(\omega_{op}/c_s)$, so that:

$$|<\mathbf{k}'|H_{op}|\mathbf{k}>|^{2} = \frac{(D_{t}K)_{op}^{2}\hbar}{2V\rho\omega_{op}}\left(n_{op}+\frac{1}{2}\pm\frac{1}{2}\right)$$

The calculation of the scattering rates for emission and absorption of optical phonons is relatively simple, since the frequency of optical phonons is roughly independent of q, and so in the occupation number n_{op} . Thus, with simple integrations:

$$\frac{1}{\tau_{op}(E)} = \frac{(D_t K)_{op}^2 m^{*3/2}}{2^{1/2} \pi \hbar^3 \rho \omega_{op}} \left[n_{op} (E + \hbar \omega_{op})^{1/2} + (1 + n_{op}) (E - \hbar \omega_{op})^{1/2} \right] ,$$

where, of course, emission should be included only if $E > \hbar \omega_{op}$.

For acoustic phonons things are more complicated, since the phonon energy depends on the phonon wavevector q and so does the Bose factor n_q . A common approximation embraced to simplify the problem is to assume that the phonon energy is smaller that the thermal energy $k_B T$ (at sufficiently low T and for sufficiently small q), so that $n_q \approx k_B T/(\hbar \omega) = k_B T/(\hbar c_s q)$ and to ignore also the energy lost or gained in a collision, if the phonon energy is much smaller then the electron energy. These two approximations are called the 'equipartition' and the 'elastic' approximations, respectively. Having made these approximations one has:

$$\frac{1}{\tau_{ac}(E)} = \frac{2^{1/2} \Delta_{ac}^2 m^{*3/2} k_B T}{\pi \hbar^4 \rho c_s^2} E^{1/2}$$

Nonpolar scattering with acoustic phonons controls the low-field electron mobility in nonpolar semicondecutors, such as Si and Ge. Also important is intervalley scattering assisted by both acoutic and optical phonons. Note, however, that by symmetry considerations Harrison scattering in the six Si X-valleys vanishes. The figure below shows the total electron-phonon and hole-phonon scattering rates at 300 K in Si calculated using the nonlocal empirical pseudopotential band structure.

214



• Fröhlich interaction.

In polar semiconductors optical phonons do generate a long-range dipole fields with which electrons interact. In order to evaluate the interaction Hamiltonian, first let's recall that only longitudinal modes give raise to a dipole field (see page 169). The dipole field associated with LO phonons will onviously be proportional to the displacement,

$$\mathbf{P} = \mathcal{B}_{\mathbf{q}} \left(b_{\mathbf{q}} + b_{\mathbf{q}}^{\dagger} \right) \,, \tag{567}$$

having omitted the index η since we deal only with the LO branch (there's only one LO branch in the semiconductors we are interested in, all having 2 ions in each unit cell).

In order to evaluate the coupling constant \mathcal{B}_{q} we can proceed as follows (see the analogous procedure at page 148 of the Notes). Let's write the electrostatic potential due to the dipole field caused by the LO phonons as well as the (plasma) response of the carriers, if any:

$$\phi_{\mathbf{q}}(\mathbf{r},t) = \phi_{\mathbf{q}}^{(0)} \cos(\mathbf{q} \cdot \mathbf{r} - \omega_{LO} t) .$$
(568)

Let's now consider the energy associated with this potential. Since phonons and plasmons in the harmonic and linear-response approximations, respectively, are represented as harmonic oscillations, the time-averaged total energy associated with these excitations is simply twice the time-averaged potential energy, $\langle U_{\mathbf{q}} \rangle$. This, in turn, is the electrostatic (self)energy of the polarization charge density $\rho_{\mathbf{q}}(\mathbf{r},t)$ in the presence of the potential $\phi_{\mathbf{q}}(\mathbf{r},t)$ caused by the polarization charge itself. We may express this potential energy in two alternative equivalent ways: From the expression (568) for the potential, the density of the polarization charge associated with the LO mode can be obtained from the Poisson equation

$$\rho_{\mathbf{q}}(\mathbf{r},t) = \epsilon \nabla^2 \phi_{\mathbf{q}}(\mathbf{r},t) , \qquad (569)$$

so that over a volume Ω :

$$\langle W_{\mathbf{q}} \rangle = 2 \langle U_{\mathbf{q}} \rangle = \frac{2}{\Omega} \int_{\Omega} d\mathbf{r} \, \phi_{\mathbf{q}}(\mathbf{r}, t) \rho_{\mathbf{q}}(\mathbf{r}, t) = \frac{2}{\Omega} \epsilon q^2 \phi_{\mathbf{q}}^{(0)2} \int_{\Omega} d\mathbf{r} \, \cos^2(\mathbf{q} \cdot \mathbf{r}) = \epsilon q^2 |\phi_{\mathbf{q}}^{(0)}|^2 \,.$$
(570)

Alternatively, we can express $\langle W_{\mathbf{q}} \rangle$ in terms of the electrostatic energy of the field $\mathbf{E}_{\mathbf{q}} = -\nabla \phi_{\mathbf{q}}$:

$$\langle W_{\mathbf{q}} \rangle = \frac{2}{\Omega} \left\langle \int_{\Omega} d\mathbf{r} \ \epsilon \left| \mathbf{E}_{\mathbf{q}}(\mathbf{R}, t) \right|^2 \right\rangle ,$$
 (571)

which yields the same result of Eq. (570) above. We now use a semiclassical argument as our final step: We set the quantity $\langle W_{\mathbf{q}} \rangle$ equal to the zero-point energy, $\hbar \omega_{LO}/2$, of the quantized excitation. Thus:

$$\phi_{\mathbf{q}}^2 = \frac{\hbar\omega_{LO}}{2q^2\epsilon} \,. \tag{572}$$

Before identifying this with the coupling constant $\mathcal{B}_{\mathbf{q}}$ we must recall that the field $\phi_{\mathbf{q}}$ is associated not only to the LO phonons, but it includes also the possible response of free carriers and of additional ionic modes. Whatever we include is determined by our choice of ϵ . In order to isolate the contribution of the LO modes, note that if we use in Eq. (572) the static dielectric function $\epsilon_{low} = \epsilon(\omega_{low})$ evaluated at a frequency $\omega_{low} << \omega_{LO}$, then we allow the LO phonons to respond fully. If, on the other hand, we set $\epsilon_{hi} = \epsilon(\omega_{hi})$ evaluated at a frequency $\omega_{hi} >> \omega_{LO}$, then we prevent the LO phonons from responding. Therefore,

$$\phi_{\mathbf{q}}^2 = e^2 \frac{\hbar \omega_{LO}}{2q^2} \left(\frac{1}{\epsilon_{hi}} - \frac{1}{\epsilon_{low}} \right) , \qquad (573)$$

accounts only for the fraction of the potential energy due to the LO phonons alone. Therefore, identifying the expression for ϕ_q above with the coupling constant \mathcal{B}_q in Eq. (567), we arrive at the following expression for the interaction Hamiltonian:

$$\mathbf{H}_{ep}^{(LO)} = e^2 \sum_{\mathbf{kq}} \left[\frac{\hbar \omega_{LO}}{2q^2} \left(\frac{1}{\epsilon_{hi}} - \frac{1}{\epsilon_{low}} \right) \right]^{1/2} c_{\mathbf{k}+\mathbf{q}}^{\dagger} (b_{\mathbf{q}} + b_{\mathbf{q}}^{\dagger}) c_{\mathbf{k}} .$$
 (574)

When only a single LO mode is present and in the absence of free carriers able to screen the interaction, typically one sets $\epsilon_{hi} = \epsilon_{\infty}$ and $\epsilon_{low} = \epsilon_0$.

The calculation of the scattering rate for this interaction proceeds in the usual way. Considering first emission processes, we have for the rate at with an electron with wavevector \mathbf{k} emits any LO phonon:

$$\frac{1}{\tau^{(em)}(\mathbf{k})} = \frac{2\pi}{\hbar} \int \frac{d\mathbf{q}}{(2\pi)^3} | < N_{\mathbf{q}} - 1; n_{\mathbf{k}+\mathbf{q}}, 1 - n_{\mathbf{k}} | \mathbf{H}_{ep}^{(LO)} | N_{\mathbf{q}}; 1 - n_{\mathbf{k}+\mathbf{q}}, n_{\mathbf{k}} > |^2 \delta[E(\mathbf{k}) - E(\mathbf{k}+\mathbf{q}) - \hbar\omega_{LO}]$$
(575)

having considered the only matrix elements which will give a nonvanishing contribution and having assumed dispersionless LO phonons. Thus, inserting Eq. (574) into (575) we have:

$$\frac{1}{\tau^{(em)}(\mathbf{k})} = \frac{2\pi}{\hbar} \frac{e^2 \hbar \omega_{LO}}{2} \left(\frac{1}{\epsilon_{hi}} - \frac{1}{\epsilon_{low}} \right) (1 + N_{LO}) \int \frac{d\mathbf{q}}{(2\pi)^3} \frac{1}{q^2} \,\delta[E(\mathbf{k}) - E(\mathbf{k} + \mathbf{q}) - \hbar \omega_{LO}] \,.$$
(576)

Now let's express q in polar coordinates with the z-axis along k. Thus, in the simple case of parabolic, spherical bands with effective mass m^* , we can write the argument of the delta-function as:

$$E(\mathbf{k}) - E(\mathbf{k} + \mathbf{q}) = -\frac{\hbar^2 q^2}{2m^*} - \frac{\hbar^2 k q}{m^*} \cos \theta , \qquad (577)$$

so that, after the trivial integration over the azimuthal angle ϕ we have:

$$\frac{1}{\tau^{(em)}(\mathbf{k})} = \frac{e^2 \omega_{LO}(1+N_{LO})}{4\pi} \left(\frac{1}{\epsilon_{hi}} - \frac{1}{\epsilon_{low}}\right) \int_0^\infty dq \int_0^\pi d\theta \,\delta \left[\frac{\hbar^2 q^2}{2m^*} + \frac{\hbar^2 kq}{m^*}\cos\theta + \hbar\omega_{LO}\right]$$
(578)

Now consider the integral over the polar angle: Changing the integration variable θ to $x = (\hbar^2 kq/m^*) \cos \theta$, this integral becomes:

$$\int_{0}^{\pi} d\theta \,\delta \left[\frac{\hbar^{2} q^{2}}{2m^{*}} + \frac{\hbar^{2} k q}{m^{*}} \cos \theta + \hbar \omega_{LO} \right] = \frac{1}{a} \int_{-a}^{a} dx \,\delta(x+b) , \qquad (579)$$

with $a = \hbar^2 kq/m^*$ and $b = \hbar^2 q^2/(2m^*) + \hbar \omega_{LO}$. This integral yields $1/a = m^*/(\hbar^2 kq)$ if the argument of the delta-function vanishes for some x in the interval [-a, +a], that is for

$$|a| \geq |b| \rightarrow \frac{\hbar^2 q^2}{2m^*} + \hbar \omega_{LO} \leq \frac{\hbar^2 k q}{m^*},$$

$$q_{-} \leq q \leq q_{+}, \qquad (580)$$

with

or

$$q_{\pm} = k \pm \left(k^2 - \frac{2m^*\omega_{LO}}{\hbar}\right)^{1/2} = \left(\frac{2m^*}{\hbar^2}\right)^{1/2} \left[E^{1/2} \pm \left(E - \hbar\omega_{LO}\right)^{1/2}\right], \quad (581)$$

provided $E > \hbar \omega_{LO}$ (otherwise the integral vanishes, expressing the fact that the electron must have a kinetic energy larger than $\hbar \omega_{LO}$ to emit a phonon). Thus, finally, for $E > \hbar \omega_{LO}$:

$$\frac{1}{\tau^{(em)}(\mathbf{k})} = \theta(E - \hbar\omega_{LO}) \frac{e^2 m^* \omega_{LO}(1 + N_{LO})}{4\pi\hbar^2 k} \left(\frac{1}{\epsilon_{hi}} - \frac{1}{\epsilon_{low}}\right) \int_{q_-}^{q_+} dq \frac{1}{q} = \\ = \theta(E - \hbar\omega_{LO}) \frac{e^2 m^* \omega_{LO}(1 + N_{LO})}{4\pi\hbar^2 k} \left(\frac{1}{\epsilon_{hi}} - \frac{1}{\epsilon_{low}}\right) \ln\left(\frac{\sqrt{E} + \sqrt{E - \hbar\omega_{LO}}}{\sqrt{E} - \sqrt{E - \hbar\omega_{LO}}}\right),$$
(582)

where the function $\theta(x)$ is the Heavyside step-function, $\theta(x) = 1$ for x > 0, $\theta(x) = 0$ for $x \le 0$. For absorption processes we find, following identical steps:

$$\frac{1}{\tau^{(abs)}(\mathbf{k})} = \frac{e^2 m^* \omega_{LO} N_{LO}}{4\pi \hbar^2 k} \left(\frac{1}{\epsilon_{hi}} - \frac{1}{\epsilon_{low}}\right) \ln\left(\frac{\sqrt{E + \hbar\omega_{LO}} + \sqrt{E}}{\sqrt{E + \hbar\omega_{LO}} - \sqrt{E}}\right) .$$
(583)

Note the typical behavior of Coulomb processes: The scattering rate increases at first with the density of states for small E, but at larger kinetic energy it decays as $1/k \approx 1/\sqrt{E}$. The reason is quite simple: At increasingly large kinetic energies any Coulomb potential appears to be an increasingly smaller relative perturbation, thus affecting the electron to an increasingly smaller extent.

Fröhlich scattering, obviously absent in covalent materials like Si and Ge (for which the absence of ionic polarization implies $\epsilon_{low} = \epsilon_{hi}$), is the dominant scattering process affecting the mobility of III-V compound semiconductors. It also enters quite heavily in the dielectric breakdown of insulators (usually ionic materials – often amorphous) at very high electric fields: As long as the electric field is low enough, LO-scattering keeps the electrons at relatively low kinetic energies. In this low-energy range the scattering rates $1/\tau(E)$ increases with increasing E. But if the electric field increases above some critical value, electrons will be accelerated to the range of higher kinetic energies in which $1/\tau(E)$ decreases with increasing E: As they reach higher E, electrons will scatter less frequently, thus losing a smaller fraction of their kinetic energy to LO-phonons, thus gaining from the field even more kinetic energy,... etc. This diverging process, called *electron run-away*, could trigger dielectric breakdown in the absence of additional inelastic processes.

Finally, recall that we have so far ignored completely the overlap integral $\mathcal{I}_{\mathbf{k}+\mathbf{q}+\mathbf{G},\mathbf{k}}$, having taken it as equal to unity. Just to show that this approximation may be less than satisfactory in some important case, it is worth stating – without proof – a result due to Fawcett: For polar materials characterized by a nonparabolic conduction band with nonparabolicity parameter α , the scattering rate give by Eq. (583) above should be replaced by the following expression:

$$\frac{1}{\tau(\mathbf{k})} = \theta(E') \frac{e^2 \omega_{LO} m^{*1/2}}{2^{5/2} \pi \hbar} \left(\frac{1}{\epsilon_{hi}} - \frac{1}{\epsilon_{low}}\right) \frac{1 - 2\alpha E'}{\gamma(E)} c_{po} \times$$

$$\times \left\{ a_{po} \ln \left[\frac{|\gamma(E)^{1/2} + \gamma(E')^{1/2}|}{|\gamma(E)^{1/2} - \gamma(E')^{1/2}|} \right] + b_{LO} \right\} (N_{LO} + 1/2 \pm 1/2) \gamma(E')^{1/2} (1 + 2\alpha E')$$
(584)

where $E' = E(\mathbf{k}) \mp \hbar \omega_{LO}$ and $\gamma(E)$ is simply the parabolic dispersion $\hbar^2 k^2 / (2m^*)$. The effect of the overlap integral is seen in the terms:

$$a_{po} = 2(1 - \alpha E')(1 - \alpha E) - \alpha [\gamma(E) + \gamma(E')]^2$$

$$b_{po} = 2\alpha [\gamma(E)\gamma(E')]^{1/2} [4(1 - \alpha E')(1 - \alpha E) - \alpha [\gamma(E) + \gamma(E')],$$

and

$$c_{po} = [4(1 - \alpha E')(1 - \alpha E) (1 - 2\alpha E')(1 - 2\alpha E)]^{-1}.$$

The step function $\theta(E')$ is obviously always unity for absorption processes, but it accounts for the energy threshold for emission. This expression is rigorously valid for spherical valleys, while it can be derived for arbitrary ellipsoidal valleys by approximating

$$q \equiv |\mathbf{k} - \mathbf{k}'| \simeq \left(\frac{m_d}{m_{el}}\right)^{1/2} q^*,$$

in the denominator of the matrix element given in Eq. (574).

• Dielectric screening and coupled plasmon/LO-phonon modes.

Consider now a polar material (say, as usual, GaAs, to fix the ideas) in which electrons are affected by polar Fröhlich scattering. Suppose that the semiconductor is n-type doped, so that there is a uniform density of n electrons per cubic centimeter. Clearly, the interaction between electrons and the LO phonons will be screened by these free electrons, since the electrons feel the dipole field associated with the LO phonons and respond to it by screening it. However, things are a little complicated. For very small electrons densities, the interaction



can be assumed to be essentially unscreened: The plasma frequency is very small (as it vanishes as $n^{1/2}$ as $n \to 0$), so that the free electrons cannot respond to the oscillating dipole field of the LO phonons. But if we consider increasing densities, we soon reach a situation where the plasma frequency, $\omega_P = [e^2 n/(\epsilon_{\infty} m^*)]^{1/2}$ and the frequency ω_{LO} of the phonons become comparable. What this means physically is that the field due to the LO phonons excites the plasma of the free electrons. This plasma responds near resonance causing, in turn, the phonons to oscillate differently under the action of the longitudinal field of the plasma oscillations. In other words: plasmons affect phonons and phonons affect plasmons. The two modes are not decoupled (or 'independent') any longer. Rather, the system will exhibit oscillations at some new frequency and we cannot talk any more of LO phonons and plasmons as independent entities. The excitations of the systems are now 'coupled phonon-plasmon modes'.

In order to evaluate the frequency and the nature of these coupled modes we must go back at the basic definition of longitudinal excitations. We must recall that when we discussed the dielectric response of the ions, we saw that longitudinal excitations are those for which the dielectric function vanishes. In this case, indeed, we can have a non-zero electric field \mathbf{E} with a vanishing macroscopic displacement field \mathbf{D} . The full dielectric function of the system, accounting for both the ionic and the electronic response can be obtained from Eqns. (191) and (231) of the Notes, Part III, in the long-wavelength limit:

$$\epsilon(\omega) = \epsilon_{\infty} \left(1 - \frac{\omega_P^2}{\omega^2} \right) + (\epsilon_0 - \epsilon_{\infty}) \frac{\omega_{TO}^2}{\omega_{TO}^2 - \omega^2} .$$
(585)

In order to calculate the frequency of the longitudinal modes we find that $\epsilon(\omega) = 0$ for $\omega = \omega_{\pm}$, where ω_{\pm} are the two solutions of the algebraic equation:

$$\omega^{4} - \omega^{2}(\omega_{P}^{2} + \omega_{LO}^{2}) + \omega_{P}^{2}\omega_{TO}^{2} = 0,$$
(586)

or:

$$\omega_{\pm}^{2} = \frac{1}{2} \left\{ (\omega_{P}^{2} + \omega_{LO}^{2}) \pm \left[(\omega_{P}^{2} + \omega_{LO}^{2})^{2} - 4\omega_{P}^{2} \omega_{TO}^{2} \right]^{1/2} \right\} .$$
(587)

Note that for $n \rightarrow 0$ we have:

$$\omega_{\pm}^2 \to (1/2)(\omega_{LO}^2 \pm \omega_{LO}^2) ,$$
 (588)

while in the opposite limit $n \to \infty$ we have:

$$\omega_{\pm}^2 \to (1/2)[\omega_P^2 \pm (\omega_P^2 - 2\omega_{TO}^2]) ,$$
 (589)

so that the lower-frequency mode has a dispersion ω_{-} vanishing in the limit of small density and asymptotically approaching the unscreened phonon frequency ω_{TO} in the limit of very large densities. In the same limits the dispersion of the high-frequency approaches ω_{LO} and ω_{P} , respectively. The top-left frame of the figure at page 226 shows the dispersion of the two modes as a function of electron density for the case of GaAs.

The calculation of the rate at which electrons scatter with these modes can proceed along the same path we have followed to reach Eq. (583), but a few comments and modifications are necessary. First, neither of the two modes is purely a phonon or a plasmons. One must sort out the plasmon and phonons content of each coupled mode in order to evaluate its coupling with the electrons. The text by Ridley (page 332) discusses this issue, as well as the scattering strength, from a vary nice 'mechanical' perspective. Here we follow a more empirical approach. It may be shown (but we'll state the results without a rigorous proof) that the phonon content of each mode is given by:

$$P(\omega_{\pm}) = \frac{\omega_{\pm}^2 - \omega_P^2}{\omega_{\pm}^2 - \omega_{-}^2} .$$
(590)

The result should be at least 'intuitively' correct: If the dispersion tracks exactly the plasma dispersion, the excitations is indeed a pure plasmon and its phonon content vanishes. Moreover, if we define similarly the plasmon content as

$$Q(\omega_{\pm}) = \frac{\omega_{\pm}^2 - \omega_{LO}^2}{\omega_{\pm}^2 - \omega_{-}^2}, \qquad (591)$$

we have $P(\omega_{\pm}) = 1 - Q(\omega_{\pm})$ from Eq. (589), which shows that the plasmon and phonon content of each mode do indeed add up to unity, as it should be. The top-right frame of the figure at page 226 shows that the high-frequency mode is mostly phonon-like at small densities, but it becomes mostly a plasmon at large

densities. The opposite is true for the low-frequency mode.

The scattering strength can be estimated following the semiclassical approach at pages 216 and 217: We may calculate the total energy of the field associated with each mode and set it equal to the sero-point energy $\hbar \omega_{\pm}/2$ of the mode. We can then isolate the contribution of the phonon-component of each mode by considering the dielectric response when the phonon responds fully, the response when the phonon is assumed to be 'frozen', and, finally, take the difference between these two fields and assign it exclusively to the phonon contribution to the field. Thus, the 'bare' scattering strength

$$\hbar\omega_{LO}\left(\frac{1}{\epsilon_{hi}} - \frac{1}{\epsilon_{low}}\right) \tag{592}$$

of the 'bare' (*i.e.*, unscreened) LO-modes is replaced by:

$$S(\omega_{\pm}) = P(\omega_{\pm}) \hbar \omega_{\pm} \left(\frac{1}{\epsilon_{hi}(\omega_{\pm})} + \frac{1}{\epsilon_{low}(\omega_{\pm})} \right) , \qquad (593)$$

where

$$\epsilon_{hi}(\omega_{\pm}) = \epsilon_{\infty} \left(1 - \frac{\omega_P^2}{\omega_{\pm}^2} \right) ,$$
 (594)

(the limit $\omega \to \infty$ has been taken in the ionic part of Eq. (585) to account for the fact that the phonon does not respond) and

$$\epsilon_{low}(\omega_{\pm}) = \epsilon_{\infty} \left(1 - \frac{\omega_P^2}{\omega_{\pm}^2} \right) + (\epsilon_0 - \epsilon_{\infty}) = \epsilon_0 - \epsilon_{\infty} \frac{\omega_P^2}{\omega_{\pm}^2}, \qquad (595)$$

having taken the limit $\omega \to 0$ in the ionic part of Eq. (585) to account for the fact that the phonon responds fully. The last (bottom) frame of the following figure shows the scattering strength for both modes. Of interest is the comparison with the bare scattering strength, Eq. (592) (yellow line): The 'screened' scattering strength

is larger than the unscreened one! This phenomenon – as far as I know, first noted by Ridley himself, but I am sure someone else must have noted it before – is called 'anti-screening', for obvious reasons.





Scattering with Ionized Impurities

Semiconductors are obviously important because of the possibility of rendering them *p*-type or *n*-type conductors when doping them. Thus, in most applications electrons and holes must carry current in the presence of a large concentration of dopants. By definitions, we want these impurities to be ionized (so that the associated free carriers contribute to the current carrying process). Thus, to each impurity there will be associated a Coulomb field which, in the absence of screening, is long-range. Charge carriers will obviously be affected by this field. Here, we evaluate the scattering rate between electrons and the field due to these ionized impurities.

Let's assume that we have N_D ionized impurites in a volume V and that they are located at positions \mathbf{r}_i for $i = 1, N_D$. If $V(\mathbf{r} - \mathbf{r}_i)$ is the potential due to impurity i, then the Hamiltonian expressing the electron-impurity interactions will be, obviously,

$$\mathbf{H}_{imp} = \frac{e}{V} \sum_{i=1}^{N_D} \int d\mathbf{r} \ \psi^{\dagger}(\mathbf{r}) \ V(\mathbf{r} - \mathbf{r}_i) \ \psi(\mathbf{r}) \ .$$
(596)

Before proceeding we must pause and discuss a major difference between electron-phonon and electron-impurity scattering. Scattering with phonons – as with any other excitation with internal degrees of freedom – is inherently 'phase-breaking'. In other words, it is irreversible in time, as a single collision removes from the system information about its phase: We assume that the thermal bath to which phonons belong is so large (that is, it has so many degrees of freedom) that we cannot possibly keep track of the information carried by a single phonons. This information is lost – either because of our inability to record this information, or because of the fast 'decoherence' or loss of information intrinsic to any large system – so that the system behaves irreversibly, like any thermodynamic system. Impurities, on the other hand, do not carry any internal degree of freedom: The momentum transferred by an electron to the impurity is so tiny (recall: The impurity is attached to the whole lattice, so that the momentum transferred to the lattice is of the order of $m_{el}/M_{lattice} \approx 0$) that we can consider it as a fixed (in space and time) potential. In principle, we could solve the exact Schrödinger equation in the presence of the potential due to

all of the impurities and we would have a perfectly reversible, non dissipative system. In other words, we could lump the Hamiltonian (596) into the unperturbed Hamiltonian and stop talking about impurity scattering altogether. The effect of the presence of the impurities in transport would be that of modiying the wavefunctions, and so the transmission amplitude across the sample, via the coherent interaction between the electron wavefunction and the impurity potential.

However, some sort of loss of information is likely to arise from another consideration. In writing Eq. (596) we have assumed that the positions of all the impurities are precisely known. In experimental situations this is clearly not true: One must regard experiments on charge transport in a given sample as an ensemble-average over many configurations α , each configuration corresponding to a different set of random (but still yielding the same average impurity density) impurity positions, $\{\mathbf{r}_i\}_{\alpha}$. One may regard this averaging either over many different, equally 'prepared' samples, or as an average over different (microscopically large but macroscopically small) parts of the sample. In doing this average, all interference among partial waves scattered from different impurities is lost (or 'averaged out'), as in the usual random-phase approximation. Thus we are brought to an average expression for Eq. (596):

$$\langle \mathbf{H}_{imp} \rangle_{ensemble} = \frac{eN_D}{V} \int d\mathbf{r} \ \psi^{\dagger}(\mathbf{r}) \ V(\mathbf{r}) \ \psi(\mathbf{r}) = n_D \int d\mathbf{r} \ \psi^{\dagger}(\mathbf{r}) \ eV(\mathbf{r}) \ \psi(\mathbf{r}) , \qquad (597)$$

where n_D is the average impurity concentration.

To start, let's consider the impurity potential $V(\mathbf{r})$ as the bare Coulomb potential

$$eV(r) = \frac{Ze^2}{4\pi\epsilon_0 r}, \qquad (598)$$

where Z is the ionization of the impurity (typically Z=1) and ϵ_0 is the static, valence band dielectric constant of the semiconductor. The Fourier components of V are (with the 'regularization trick' we discussed before, since the 'actual' Fourier components do not exist):

$$eV_q = \frac{Ze^2}{\epsilon_0 q^2} \,. \tag{599}$$

Now we can follow the path we've followed from Eq. (542) through (544) to obtain (dropping the symbol of ensemble-average, we shall assume as implicitly taken in the following):

$$\mathbf{H}_{imp} = e \ n_D \ \sum_{\mathbf{k}\mathbf{q}\mathbf{G}} \ \mathcal{I}_{\mathbf{k}+\mathbf{q}+\mathbf{G};\mathbf{k}} \ c_{\mathbf{k}+\mathbf{q}+\mathbf{G}}^{\dagger} \ V_q \ c_{\mathbf{k}} , \qquad (600)$$

having assumed that only intra-band process are significant. This is indeed true, since the Coulomb potential not sufficiently strong to trigger inter-band processes. Assuming Normal processes and ignoring the overlap factor, we may write more simply:

$$\mathbf{H}_{imp} \approx e n_D \sum_{\mathbf{kq}} c^{\dagger}_{\mathbf{k+q}} V_q c_{\mathbf{k}} .$$
(601)

We are now ready to calculate the scattering rate:

$$\frac{1}{\tau(\mathbf{k})} = n_D \frac{2\pi}{\hbar} \sum_{\mathbf{q}} |\langle \mathbf{k} + \mathbf{q} | \mathbf{H}_{imp} | \mathbf{k} \rangle|^2 \delta[E(\mathbf{k} + \mathbf{q}) - E(\mathbf{k})] =$$

$$\frac{Z^2 e^4 n_D}{4\pi^2 \epsilon_0^2 \hbar} \int d\mathbf{q} \frac{\delta [E(\mathbf{k} + \mathbf{q}) - E(\mathbf{k})]}{q^4} = \frac{Z^2 e^4 n_D}{2\pi \epsilon_0^2 \hbar} \int_0^\infty \frac{dq}{q^2} \int_0^\pi d\theta \sin \theta \,\delta \left(\frac{\hbar^2 q^2}{2m^*} + \frac{\hbar^2 kq}{m^*} \cos \theta\right) \,. \tag{602}$$

Now, with the usual change of integration variable $x = (\hbar^2 kq/m^*) \cos \theta$ inside the angular integral we have:

$$\int_0^\infty \frac{dq}{q^2} \int_0^\pi d\theta \sin\theta \,\delta\left(\frac{\hbar^2 q^2}{2m^*} + \frac{\hbar^2 kq}{m^*}\cos\theta\right) = \frac{m^*}{\hbar^2 k} \int_0^{2k} \frac{dq}{q^3},\tag{603}$$

(the upper integration limit 2k comes from the fact that we require the argument of the δ -function to vanish in order to have a non-zero contribution and this happens only when $\cos \theta \leq 1$ or $q \leq 2k$), so that:

$$\frac{1}{\tau(\mathbf{k})} = \frac{Z^2 e^4 m^* n_D}{2\pi \epsilon_0^2 \hbar^3 k} \int_0^{2k} \frac{dq}{q^3} = \frac{Z^2 e^4 m^{*1/2} n_D}{2^{3/2} \pi \hbar^2 \epsilon_0^2} \frac{1}{E^{1/2}} \int_0^{2k} \frac{dq}{q^3} .$$
(604)

The obvious problem with this expression lies in the fact that the integral diverges! This is a well-known result: The quantum-mechanical cross section for scattering with a bare Coulomb potential (in first-order perturbation theory, also known as the Born approximation) is infinity. Clearly, we must consider the effect of screening on the bare potential of the impurities. Several models have been proposed in the past, each with its own merits and especially accurate in a limited range of impurity density.

• Brooks-Herring model.

The first model we consider is based on the simple assumption that the impurity potential, Eq. (598), should be screened statically by the static free-carrier dielectric function with screening wavevctor β given by Eq. (449) or (451) (Notes, Part 3; see also Eqns. (498) and (499)). Thus:

$$eV_q = \frac{Ze^2}{\epsilon_0(q^2 + \beta^2)} . \tag{605}$$

Equation (602) then becomes:

$$\frac{1}{\tau_{BH}(\mathbf{k})} = \frac{Z^2 e^4 n_D}{2\pi \epsilon_0^2 \hbar} \int_0^\infty dq \, \frac{q^2}{[q^2 + \beta^2]^2} \int_0^\pi d\theta \sin\theta \, \delta\left(\frac{\hbar^2 q^2}{2m^*} + \frac{\hbar^2 k q}{m^*} \cos\theta\right) = \frac{2Z^2 e^4 m^* n_D}{\pi \hbar^3 \epsilon_0^2} \, \frac{k}{\beta^2 (\beta^2 + 4k^2)} \,. \tag{606}$$

This expression was first derived by Brooks and Herring, hence the subscript BH in the formula above. Note that screening, via the term β , removes the divergence of the integral at $q \rightarrow 0$, that is, at long distances. Indeed screening makes the Coulomb potential a 'short range' potential. The source of the divergence we have met before was due to the fact that electrons feel the potential of the impurity at all distances: Screening avoids this problem.

• Conwell-Weisskopf model.

Mahan, however, has noticed an inconsistency in the way screening is employed in the Brooks-Herring approach: On the one hand we assume that electrons are plane wave in deriving the scattering rate (606): Indeed our initial and final states are pure plane waves and all electrons participate equally in the scattering process, all of remaining pure plane waves. On the other hand, we also assume that all electrons participate fully in the screening process. We are clearly double-counting the electrons. He went on and did some variational calculations trying to see what the actual electron wavefunctions look like in a homogeneous electron system in the presence of ionized impurities. He found that at small densities electrons do indeed 'pile up' around the attractrive potentials of donors (or away from the repulsive potential of acceptors), as implied by Debye-Hückel screening. But at larger densities (when the inter-impurity separations, the inter-electron separation, and the screening length β^{-1} become comparable, about 1-3 nm in Si at densities of the order of 10^{18} to 10^{19} electrons or impurities per cubic centimeters), he found that the electron wavefunctions are spread almost uniformly, without 'piling up' around the impurities: The average distance between the impurities is now smaller than the electron wavelength, so that they are not able to screen the potential. (This is accounted somewhat by the qdependence of β in the better model of Eq. (463) at page 180 of the Notes, Part 3). Thus, at large densities it makes more sense to go back to the original model proposed by Conwell and Weisskopf. They assumed that the Coulomb potential is unscreened. However, electrons can interact only with the nearest impurity, assumption consistent with our early assumption of independent collisions with impurities implicit in Eq. (597). So, let's define the average inter-impurity separation r_0 such that

$$n_D = \frac{1}{4\pi r_0^3/3} \longrightarrow r_0 = \left(\frac{3}{4\pi n_D}\right)^{1/3}$$
 (607)

This will be the maximum 'impact parameter' of the collision. The Fourier components of the Coulomb potential so 'chopped' at large distances will be:

$$V_q^{(CW)} = \frac{Ze^2}{\epsilon_0 q} \int_0^{r_0} dr \, \sin(qr) = \frac{Ze^2}{\epsilon_0 q^2} \left[1 - \cos(qr_0)\right]. \tag{608}$$

Thus:

$$\frac{1}{\tau_{CW}(\mathbf{k})} = \frac{Z^2 e^4}{2\pi\hbar\epsilon_0^2} \int_0^\infty dq \, \frac{1}{q^2} \left[1 - \cos(qr_0)\right]^2 \int_0^\pi d\theta \sin\theta \, \delta\left(\frac{\hbar^2 q^2}{2m^*} + \frac{\hbar^2 kq}{m^*}\cos\theta\right) \,. \tag{609}$$

Proceeding as usual with the azimuthal integral, we finally have:

$$\frac{1}{\tau_{CW}(\mathbf{k})} = \frac{Z^2 e^4 m^* n_D r_0^3}{2\pi \hbar^3 \epsilon_0^2} \,\mathcal{F}(2kr_0) \,, \tag{610}$$

where

$$\mathcal{F}(x) = \frac{1}{x} \int_0^x dt \, \frac{(1 - \cos t)^2}{t^3} \, .$$

Note that the scattering rate (610) does not depend on the impurity density n_D , as can be seen from Eq. (607). There are several variations of this expressions (such as still using the bare Coulomb potential at infinity, but 'chopping' the integration over q above), but the essence of the method remains unaltered: Rather than relying on screening to avoid the divergence at $q \rightarrow 0$ in Eq. (603), the cut-off at small q is obtained from the maximum impact parameter allowed. If the impact parameter exceeds r_0 (however this might be defined), then another impurity will exhibit a smaller impact parameter and scattering will occur via that second impurity center.

• Ridley's statistical screening model.

Ridley (see the text, pages 148-152) has attempted to find a model reconciling the BH and CW approaches. He notices that in the CW approach the impurity with the smallest impact parameter is always chosen, even if it happens to be farther away than another impurity with a larger impact parameter. But he thinks this is inconsistent with a true two-body, single-scattering-center collision. Thus, he is led to consider the probability that another scattering center may be closer, even if it happens to be associated with a larger impact parameter.



This method (which we shall not discuss furher and is known as 'statistical screening') results in a correction to the BH scattering rate as follows:

$$\frac{1}{\tau_R(\mathbf{k})} = \frac{\upsilon(\mathbf{k})}{d} \left\{ 1 - \exp\left[\frac{d}{\tau_{BH}(\mathbf{k})\upsilon(\mathbf{k})}\right] \right\} , \qquad (611)$$

where $v(\mathbf{k})$ is the group velocity and $d = (2\pi n_D)^{-1/3}$ is the average inter-impurity distance.

• Additional corrections to the impurity potential.

There are several even more sophisticated models which attempt to find a better approximation to the potential of the substitutional impurity. An article by Chattopadhaya and Qeisser (Rev. Mod. Phys.) discusses the state of the art. Since impurity scattering is important in determining the carrier mobility in heavily-doped semiconductors, and since the average doping of semiconductors increases with the shrinking dimensions of the devices, it is easy to understand why there is such an interest.

A first correction is an attempt to account for wavelength-dependent screening by valence electrons. So far, screening by valence electrons has been treated simply via the static dielectric constant ϵ_0 , but some dependence on the wavelength may be included. Let's follow Nara and Morita. They consider the bare core potential, $V(Z_I, n; \mathbf{r})$, of a substitutional donor *n*-fold ionized, taken isotropic since only core electrons are considered. They express this as:

$$V(Z_I, n; \mathbf{r}) = -\frac{e^2}{4\pi\epsilon_{vac}r} \left[n + (Z_I - n) e^{-\sigma_I r} \right], \qquad (612)$$

where ϵ_{vac} is the permittivity of vacuum and the atomic radii $1/\sigma_I$ are obtained from a fit to the Hartree-Fock potential. Their values are of the order of 0.01 nm for the species usually considered $(1/\sigma_I = 0.0124 \text{ nm for Si}^{4+}, 0.0111 \text{ nm for P}^{5+}, 0.0148 \text{ nm for As}^{5+}, \text{ and } 0.0156 \text{ nm for Sb}^{5+})$, consistent with the use of the vacuum permittivity ϵ_{vac} in Eq. (612).

Valence electrons are now brought in first as frozen in their anisotropic, perfect-lattice distribution. Thus, the

net potential of the substitutional impurity, $v_i(\mathbf{r})$, is obtained by replacing the Si⁴⁺ ion with the *n*-fold ionized donor, that is:

$$v_{imp}(\mathbf{r}) = V(Z_I, n; \mathbf{r}) - V(Z_{Si}, 4; \mathbf{r})$$
 (613)

Finally, in the linear response approximation, the valence electrons are allowed to readjust and screen $v_{imp}(\mathbf{r})$ via the static valence dielectric function of Si, $\epsilon_v(\mathbf{q})$. A possible model for $\epsilon_v(\mathbf{q})$ – by no means unique, but approximated by a convenient analytic expression – has the following isotropic form:

$$\frac{\epsilon_{vac}}{\epsilon_v(q)} = \frac{Aq^2}{q^2 + a^2} + \frac{Bq^2}{q^2 + b^2} + \frac{Cc^2}{q^2 + c^2}, \qquad (614)$$

where A = 1.175, B = -0.175, $C = \epsilon_0/\epsilon_{sc} \approx 0.08547$, a = 0.7572, b = 0.3123, and c = 2.044, a, b, and c being in atomic units. Finally, the Fourier components of the screened isotropic impurity potential, $v_{imp}(q)$, become:

$$v_{imp}(q) = \frac{e^2}{\epsilon_v(q)} \frac{H_I(q)}{q^2}, \qquad (615)$$

where

$$H_I(q) = (n-4) + (Z_I - n) \frac{q^2}{q^2 + \sigma_I^2} - (Z_{Si} - 4) \frac{q^2}{q^2 + \sigma_{Si}^2}, \qquad (616)$$

where n = 3 for acceptors, 5 for donors. Equation (615) should replace the Fourier transform of the impurity potential we have used before in Eq. (605) or, appropriately 'chopped' at large distances, in Eq. (608). In practice, this corrections do not matter too much quantitatively. Only at very large energies, when the electron approach the impurities at very short distances (large q) one sees some differences between ddifferent donor or acceptor species, because of some differences in the core potentials of different ions. These are known as 'central cell' corrections.

Finally, Friedel's sum rule as well as partial-waves approaches can tackle another interesting problem. The Born approximation we have used so far does not distinguish between repulsive and attractive potentials. Yet, we expect some differences: In the case of attractive potentials (electron-donors) the electronic wavefunction gets

closer to the impurity core than in the case of repulsive potentials (electron-acceptors). Partial-wave analysis (which we saw above dealing with Friedel's sum rule, pages 191-195 of the Notes), which go beyond the first-order Born approximation, can account for this difference. The practical interest lies in the context of transport of electrons (holes) in the p-type (n-type) base of npn (pnp) bipolar transistors. These so-called 'minority carrier mobilities' are indeed experimetally observed to be different (usually larger, as expected) than the majority carrier mobilities.

Coulomb interactions among free carriers

In the first part of these Notes (pages 50 and following) we have considered a 'mean field' approximation to treat the Coulomb interactions among valence electrons. The same approximation is usually embraced also when dealing with free carriers (say: electrons in the conducxtion band of an *n*-type doped semiconductor). This approximation is obtained by replacing the 'fluctuating' inter-particle Coulomb potential with a smoother Hartree potential obtained by averaging out (spatially and temporally) the instantaneous and local potential over all electron configurations. The Hartree-Fock approximation is similar, since it only adds exchange effects, but it employes the same averaging procedure. In order to re-capture the 'oscillating' nature of the interaction, we must consider the fluctuations of the electron positions and the associated fluctuations of the potential.

• Electron-electron Hamiltonian.

Let's consider the free-particle case (no additional potential) and ignore spin for now. The associated Hamiltonian has the 'obvious' form:

$$\mathbf{H}_{ee} = \frac{1}{2} \int d\mathbf{r} \int d\mathbf{r}' \,\psi^{\dagger}(\mathbf{r}',t) \,\psi(\mathbf{r}',t) \,V(\mathbf{r}'-\mathbf{r}) \,\psi^{\dagger}(\mathbf{r},t) \,\psi(\mathbf{r},t) , \qquad (617)$$

where $V(\mathbf{r}) = e^2/(4\pi\epsilon_{\infty}r)$, since $\psi^{\dagger}(\mathbf{r},t) \psi(\mathbf{r},t)$ is the density operator which satifies the expected property $\int d\mathbf{r} \psi^{\dagger}(\mathbf{r},t) \psi(\mathbf{r},t) = \sum_{\mathbf{k}} c_{\mathbf{k}}^{\dagger}(t) c_{\mathbf{k}}(t)$. Using Eqns. (291) and (292) (page 134), the Hamiltonian (617) can be written as:

$$\mathbf{H}_{ee} = \frac{1}{2} \sum_{\mathbf{k},\mathbf{p},\mathbf{q}} V_q c^{\dagger}_{\mathbf{k}+\mathbf{q}} c^{\dagger}_{\mathbf{p}-\mathbf{q}} c_{\mathbf{k}} c_{\mathbf{p}} , \qquad (618)$$

where $V_q = e^2/(\epsilon_{\infty}q^2)$ is the Fourier transform of the Coulomb potential. We use the high-frequency dielectric response ϵ_{∞} for now. We shall worry about screening later on. This expression has a simple interpretation:

The Coulomb Hamiltonian destroys two electrons in the states k and p and creates two new electrons in states k + q and p - q via the exchange of momentum $\hbar q$.

In taking the Hartree or Hartree-Fock approximation we would look at a single electron of wavevector k and replace the effect of all 'other' electrons by considering only the limit $q \rightarrow 0$ (that is, a spatial average over large distances), so that we would retain only the 'diagonal' part of $c_{\mathbf{p}-\mathbf{q}}^{\dagger}c_{\mathbf{p}}$, that is, the number operator $n_{\mathbf{p}} = c_{\mathbf{p}}^{\dagger}c_{\mathbf{p}}$. We would then take a suitable ensemble average, $< n_{\mathbf{p}} >$, over the electron ensemble, obtaining just the average charge density $\rho_{\mathbf{p}} = en_{\mathbf{p}}$. This would give raise to the average 'mean field' felt by each electron. On the contrary, now we want to isolate deviations from this mean field. Thus, we shall assume that the Hartree (or Hartree-Fock) potential is substracted and isolated into another term $eV_{Hartree}(\mathbf{r})$ in the Hamiltonian and assume that Eq. (618) represents only deviations from the mean field. Thus, we should consider in the Hamiltonian above the terms $c_{\mathbf{p}-\mathbf{q}}^{\dagger}c_{\mathbf{p}}$ as representing the fluctuating part $c_{\mathbf{p}-\mathbf{q}}^{\dagger}c_{\mathbf{p}} - < c_{\mathbf{p}-\mathbf{q}}^{\dagger}c_{\mathbf{p}} >$, and similarly for $c_{\mathbf{k}+\mathbf{q}}^{\dagger}c_{\mathbf{k}}$ as representing the fluctuating part $c_{\mathbf{k}+\mathbf{q}}^{\dagger}c_{\mathbf{k}} >$. How exactly this is done accounting for the Hartree and the exchange component we may see (if time allows) when we shall derive the Boltzmann Transport Equation from the quantum transport equation, the so-called Liouville-von Neumann equation.

In a series of articles published in the Physical Review in the 1950's David Bohm and his student David Pines have studied the effects of Eq. (618) and have reached the following general conclusions:

- When the magnitude q of the momentum transfer \mathbf{q} is small, the interaction of an electron with all of the other electrons occurs via 'collective excitations': The 'initial' electron with wavevector \mathbf{k} triggers (via emission or absorption) a plasma wave in the surrounding gas of free carriers. The interaction is an electron-plasmon collision. It may be characterized by the same procedure we have followed to study the Fröhlich interaction, simply replacing the coupling constant $(\hbar\omega_{LO}/2)[1/\epsilon_{hi}-1/\epsilon_{low}]$ with $(\hbar\omega_P/2)(1/\epsilon_{\infty})$ (see Eqns. (348) of Part 3 and (573)). Note, however, that in this process momentum is not lost by the electron gas: It is simply re-distributed among electrons.
- When the magnitude q of the momentum transfer grows beyond a certain limit, plasmons cease to be good

excitations: They decay into single-particle excitations. Think of a surfer extracting energy from the wave as the surfer's velocity matches the phase velocity of the wave. Similarly, when the wavelength of the plasmon matches the Fermi wavelength, electrons at the Fermi surface will 'surf' the plasmon, thus absorbing energy from it and damping it. This damping occurs via excitations of electrons from the 'Fermi sea' which gain the plasma energy This process is called 'Landau damping'. Energy and momentum conservation allow this process to happen whenever the plasmon wavevector \mathbf{q} satisfies the condition

$$E(\mathbf{k}_F + \mathbf{q}) - E(\mathbf{k}_F) = \hbar \omega_P , \qquad (619)$$

or, for parabolic and spherical bands,

$$\frac{\hbar^2}{2m^*}(q^2 - 2k_F q) \leq \hbar\omega_P \leq \frac{\hbar^2}{2m^*}(q^2 + 2k_F q) .$$
 (620)

For q in this range of values, the interaction between an electron and all other free electrons occurs via single-particle (two-body) collisions, as in the case of a classical collision between two charged particles. The figure below shows the region in the $(\hbar \omega, q)$ -space in which plasmons are damped.



R)

• Collision rate: General expression.

We can calculate the scattering rate for electron-electron collisions in the single-particle picture as follows. We consider as usual Fermi Golden rule (the first order perturbation, more appropriately known as the first Born approximation for an electron gas which is diluted enough), looking at the matrix element of the dynamically screened Coulomb potential between antisymmetrized wavefunction of spin-1/2 particles. As briefly mentioned below, we must apply some approximated phase-shift corrections in order to account for the difference between repulsive and attractive interactions. Thus, for a 'primary' particle of crystal momentum \mathbf{k} in band ν , the Coulomb scattering rate at location \mathbf{r} at time t is approximated by:

$$\frac{1}{\tau_{ee}(\mathbf{k},\mathbf{r},\nu)} = \frac{2\pi}{\hbar} \sum_{\mathbf{G},\mu,\nu',\mu'} \int \frac{d\mathbf{p}}{(2\pi)^3} f(\mathbf{p},\mu,\mathbf{r},t) \int \frac{d\mathbf{k}'}{(2\pi)^3} |M(\mathbf{k}\nu,\mathbf{p}\mu;\mathbf{k}'\nu',\mathbf{p}'\mu';\mathbf{G})|^2 \times \\ \delta[E_{\nu}(\mathbf{k}) + E_{\mu}(\mathbf{p}) - E_{\nu'}(\mathbf{k}') - E_{\mu'}(\mathbf{p}')][1 - f(\mathbf{k}',\nu',\mathbf{r},t)][1 - f(\mathbf{p}',\mu',\mathbf{r},t)].$$
(621)

Let's recall here the meaning of all terms occurring in this equation:

- $\mathbf{p}' = \mathbf{k} + \mathbf{p} \mathbf{k}'$ is the final state of the scattering partner of the 'primary' particle with crystal momentum \mathbf{k} . Momentum is of course conserved.
- $f(\mathbf{k}, \mu, \mathbf{r}, t)$ is the distribution function, *i.e.*, the phase-space density at the phase-space location (\mathbf{r}, \mathbf{p}) in band μ , at time t, normalized so that the real-space density at \mathbf{r} is:

$$n(\mathbf{r},t) = 2\sum_{\mu} \int \frac{d\mathbf{k}}{(2\pi)^3} f(\mathbf{k},\mu,\mathbf{r},t)$$

Note that in Eq. (621) the integral over the states of the 'partner' electrons of crystal momentum \mathbf{p} is done over 1/2 of them, since the sum over spin states is already included in the matrix element (with the associated overlap factors).

- The G's are the reciprocal-lattice wavevectors.

The figures above show the ratio of the dynamic and static screening parameter to the Thomas-Fermi approximation β as a function of momentum transfer (left) and the interparticle scattering rate as a function of electron energy in a variety of approximations: classical particles (no spin), spin-1/2 but no phase-shift corrections, and spin-1/2 with phase-shift corrections for attractive and repulsive potentials, the latter being of course appropriate to e-e scattering.



- Phase-shift corrections obtained via the partial-wave analysis we have seen before dealing with Friedel's sum rule – can be implemented and some of the following figures show the effect of these corrections (labeled as 'phase shift') They do indeed yield a different scattering rate for repulsive (weaker scattering) and attractive (stronger) interactions.
- $M(\mathbf{k}\nu,\mathbf{p}\mu;\mathbf{k}'\nu',\mathbf{p}'\mu';\mathbf{G})$ is the Coulomb matrix element. In the case of scattering between two

indistinguishable particles (electron-electron or hole-hole), it is the antisymmetrized Coulomb matrix element obtained by summing over all spin states: If

$$M_d = \frac{e^2}{\epsilon_s} \frac{\mathcal{I}_{\mathbf{k}\nu,\mathbf{k}'\nu'} \mathcal{I}_{\mathbf{p}\mu,\mathbf{p}'\mu'}}{|\mathbf{k} - \mathbf{k}' + \mathbf{G}|^2 + \beta^2(q_d,\omega_d)},$$
(622)

where $\beta(q, \omega)$ is defined below, is the matrix element for the 'direct' process $(\mathbf{k}\mu) \rightarrow (\mathbf{k}'\mu')$, $(\mathbf{p}\nu) \rightarrow (\mathbf{p}'\nu')$, and

$$M_x = \frac{e^2}{\epsilon_s} \frac{\mathcal{I}_{\mathbf{k}\nu,\mathbf{p}'\mu'} \mathcal{I}_{\mathbf{p}\mu,\mathbf{k}'\nu'}}{|\mathbf{k} - \mathbf{p}' + \mathbf{G}|^2 + \beta^2(q_x,\omega_x)}$$
(623)

is the matrix element for the 'exchange' process $(\mathbf{k}\mu) \rightarrow (\mathbf{p}'\nu')$, $(\mathbf{p}\nu) \rightarrow (\mathbf{k}'\mu')$, the matrix element for scattering of spin-1/2 particles is given by the sum of the scattering probability in the singlet state and the scattering probability in the triplet state, *i.e.* :

$$|M|^{2} = \frac{1}{4}|M_{d} + M_{x}|^{2} + \frac{3}{4}|M_{d} - M_{x}|^{2} = |M_{d}|^{2} + |M_{x}|^{2} - \frac{1}{2}|M_{d}^{*}M_{x} + M_{d}M_{x}^{*}|.$$
(624)

The last 'interference term', the 'exchange term', is responsible for depressing the cross section at small relative energies, $\mathbf{k} - \mathbf{p} \ll \beta$.

In the case of scattering between distinguishable particles (electron-hole scattering), these terms do not appear. Since, after integration, $|M_d|^2$ and $|M_x|^2$ yield the same result, one could obtain an equivalent result by ignoring spin altogether and replacing the integral over \mathbf{p} above with twice that, *i.e.*, by summing over all spin states. The figure on page 241 (right frame) shows the effect of (anti)symmetrization on the short-range scattering rate for a simple parabolic band of Si at 300 K.

At low relative energies, Goodnik and Lugli, in dealing with scattering between indistinguishable particles, have assumed that only the direct term $|M_d|^2$ matters (since the interaction between electron of parallel spins is reduced by exclusion) thus effectively obtaining a scattering rate equal to half the rate for indistinguishable



particles.

- $\mathcal{I}_{\mathbf{k}\nu,\mathbf{k}'\nu'}$ is the overlap integral between the Bloch states, accounting for the sum over spin states.
- $\beta(q, \omega)$ is the dynamic screening parameter, *i.e.* the quantity such that the dielectric function can be written as $\epsilon(q, \omega)/\epsilon_s = 1 + \beta(q, \omega)^2/q^2$. It must be evaluated at the crystal momentum transfer $q_d = |\mathbf{k} \mathbf{k}' + \mathbf{G}|$, and at the energy transfer $\hbar \omega_d = E_{\nu}(\mathbf{k}) E_{\nu'}(\mathbf{k}')$ for the direct process, at $q_x = |\mathbf{k} \mathbf{p}' + \mathbf{G}|$, and at $\hbar \omega_x = E_{\nu}(\mathbf{k}) E_{\mu'}(\mathbf{p}')$ for the exchange process.



The figure above shows the effect of dynamic screening on the short-range inter-particle scattering rate for a parabolic band model of Si at 300 K. Three cases are shown: the dotted line is for the case of static Debye-Hückel screening, the dashed line for a wavevector dependent static screening parameter, corresponding to a 'head-on' collision. Finally, the solid line is relative to the case

of a collision in which one electron is at rest in the frame of the electron gas.

• Some common approximations.

- The effect of degeneracy, represented by the factors (1 - f) in Eq. (621) is often included by replacing the distribution functions f with their equilibrium values,

$$f(\mathbf{k}, \mu, \mathbf{r}, t) \simeq f_0[\mathbf{k}, \mu, T_p, E_F],$$
(625)

where $f_0(\mathbf{k}, \mu, T, E_F) = \{1 + \exp[(E_\mu(\mathbf{k}) - E_F)/(k_B T)]\}^{-1}$ is the Fermi-Dirac function at temperature T for band μ and Fermi energy E_F , T_p is the equivalent particle temperature. Note that both this degeneracy correction as well as the appearance of the distribution function in the screening corrections render the transport problem non-linear: Typically, we use scattering rates to solve the transport problem in which f is the unknown. But when the scattering rates depend on f itself, either iterative or self-consistent schemes are devised, or the dependence on f of the rates must be approximated. In typical cases, as we have seen here, a Fermi expression with an equivalent electron temperature is used.

- Accordingly, dynamic screening may be treated in the high temperature, nondegenerate limit valid for a single parabolic and spherical valley, as we have seen in the Notes, Part 3, Eqns. (232) and (233), page 180 and 181. We have seen in a figure above (left frame, page 241) the ratio of the squared dynamic screening parameter to the squared Debye-Hückel parameter for the case of a 'parabolic' relative energy transfer $\omega = \hbar q^2/(2m_d)$ in Si. The same figure also shows the behavior of the static screening parameter $\beta^2(q, 0)$.



The figures above show – at the left – the short-range electron-electron scattering as a function of electron energy in Si at 300 K at a uniform field of 200 kV/cm, averaged over the distribution function during Monte Carlo runs at the two densities indicated. At the right we see the short-range electron-electron scattering as a function of electron energy in Si at 300 K at a density of 10^{18} cm⁻³ at a field of 100 kV/cm. The effect of spin and dynamic screening is also shown. The figure below shows the same information, but at a field of 200 kV/cm and at a density of 10^{20} cm⁻³.



The figure above shows the effect of the phase-shift corrections on the short-range, (anti)symmetrized electronelectron scattering rate for the 'usual' parabolic band model for Si at 300 K and static screening.

• Electron-electron scattering in the parabolic-band approximation.

In the simple case in which a parabolic approximation is used, the electron-electron scattering rate between two electrons can be evaluated analytically when ignoring degeneracy effects, Umklapp processes, and taking the overlap integrals as unity. Note the words in italic: Equation (621) expresses the collision rate for a given 'primary' electron of wavevector k in band ν with any other electron in the system. Here we consider the collision rate between our primary electron and one particular 'partner' electron with wavevector p in band μ , lumping the effect of all other electrons into the electron density $n(\mathbf{r}, t)$ appearing as a multiplicative factor. In



other words, we approximate $f(\mathbf{p}', \mu', \mathbf{r}, t)$ with $n(\mathbf{r}, t)\delta_{\mu\mu'}\delta(\mathbf{p} - \mathbf{p}')$.

Thus, Eq. (621) simplifies to:

$$\frac{1}{\tau_{ee}(\mathbf{k}\nu,\mathbf{p}\mu,\mathbf{r},t)} = \frac{2\pi}{\hbar} \frac{e^4 n(\mathbf{r},t)}{4\epsilon_{\infty}^2} \sum_{\mathbf{k}'} \delta[E(\mathbf{k}) + E(\mathbf{p}) - E(\mathbf{k}') - E(\mathbf{p}')] \times$$

$$\times \left\{ \frac{1}{(|\mathbf{k}' - \mathbf{k}|^2 + \beta_d^2)^2} + \frac{1}{(|\mathbf{k}' - \mathbf{p}|^2 + \beta_x^2)^2} - \frac{2}{(|\mathbf{k}' - \mathbf{k}|^2 + \beta_d^2)(|\mathbf{k}' - \mathbf{p}|^2 + \beta_x^2)} \right\}, \quad (626)$$

where β_d is avaluated at $\mathbf{q} = \mathbf{k}' - \mathbf{k}$ and $\hbar \omega_d = E(\mathbf{k}') - E(\mathbf{k})$ while β_x is avaluated at $\mathbf{q} = \mathbf{k}' - \mathbf{p}$ and $\hbar \omega_x = E(\mathbf{k}') - E(\mathbf{p})$. Additional approximations are required in order to reach an analytic expression:

 Screening must be taken in the long-wavelength static limit. The justification lies in the fact that dynamic screening corrections are typically very small. Note, however, that this is not true in general: In cases wher a 'hot' electron interacts with a pool of 'cold' electrons (such as an electron entering the drain region of a highly-biased MOSFET), the large energy transfer renders either the direct or the exchange matrix element essentially unscreened, thus rendering the interaction stronger than what static screening would suggest. When static screening is considered, the screening parameter takes the form:

$$\beta_{pm}^2 \simeq \beta_s^2 g_1(q\lambda),$$

where

$$\begin{split} \lambda &= \left(\frac{2\pi\hbar^2}{m_{red}k_BT}\right)^{1/2}\\ g_1(x) &= \frac{2\pi^{1/2}}{x} \Phi\left(\frac{x}{4\pi^{1/2}}\right)\,, \end{split}$$

At the long wavelengths which dominate scattering when screening is not too strong, $\beta_{\pm} \simeq \beta_s$.

2. Intervalley processes must also be ignored, since the valley separation would appear to the fourth power in the denominators of the Coulomb matrix elements, thus depressing their contribution.

Using the approximations above, Eq. (626) can be integrated analytically. Go to the center-of-mass frame by putting:

$$\mathbf{g} = \mathbf{k} - \mathbf{p}$$
, $\mathbf{g}' = \mathbf{k}' - \mathbf{p}'$,

so that:

$$|\mathbf{k} - \mathbf{k'}| = \frac{1}{2}|\mathbf{g} - \mathbf{g'}|,$$

keeping also in mind that $\mathbf{p}'=\mathbf{k}+\mathbf{p}-\mathbf{k}'.$ Then:

$$E(\mathbf{k}) + E(\mathbf{p}) - E(\mathbf{k}') - E(\mathbf{p}') = \frac{\hbar^2 g^2}{2m^*} - \frac{\hbar^2 g'^2}{2m^*}$$

Consider now just the first (direct) term inside the integral in Eq. (626):

$$\frac{\pi e^4 n(\mathbf{r}, t)}{2\hbar\epsilon_{\infty}^2} \int \frac{d\mathbf{g}'}{(2\pi)^3} \frac{\delta[E(\mathbf{g}) - E(\mathbf{g}')]}{[|\mathbf{g} - \mathbf{g}'|^2/4 + \beta^2]^2} \,.$$
(627)

This expression is identical to the integration we had to perform in order to calculate the electron-impurity scattering rate and we find:

$$\frac{e^4 n(\mathbf{r}, t)m^*}{2\pi\hbar^3\epsilon_{\infty}^3} \frac{g}{\beta^2(g^2 + \beta^2)} .$$
(628)

Accounting for the other two terms, we finally obtain:

$$\frac{1}{\tau_{ee, self}(g)} \simeq \frac{e^4 m^* g \, n(\mathbf{r}, t)}{2\pi \epsilon_s^2 \hbar^3} \left[\frac{1}{\beta_s^2 (\beta_s^2 + 4g^2)} - \frac{\ln(1 + 4g^2/\beta_s^2)}{8g^2 (\beta_s^2 + 2g^2)} \right] \,. \tag{629}$$

Note that the cross section, $\sigma=2m^*/(\tau n\hbar g)$, corresponding to this process equals the result reported by

Ridley (Eq. (4.136) of the text). Replacing 2g with $k_r = |\mathbf{k} - \mathbf{p}|$, Eq. (629) takes the form:

$$\frac{1}{\tau_{ee}(k_r)} \simeq \frac{e^4 m^* k_r n(\mathbf{r}, t)}{4\pi \epsilon_s^2 \hbar^3} \left[\frac{1}{\beta_s^2 (\beta_s^2 + k_r^2)} - \frac{\ln(1 + k_r^2 / \beta_s^2)}{2k_r^2 (\beta_s^2 + k_r^2 / 2)} \right] .$$
(630)

For the case of spinless, distinguishable particles (*i.e.*, classical statistics, unsymmetrized wavefunctions) and using the first Born approximation, the equation above reduces to the (familiar?) form:

$$\frac{1}{\tau_{ee}(k_r)} \simeq \frac{e^4 m^* k_r n(\mathbf{r}, t)}{4\pi \epsilon_s^2 \hbar^3} \frac{1}{\beta_s^2 (\beta_s^2 + k_r^2)},\tag{631}$$

since only the first term in Eq. (630) contributes. This is result of 'classical mechanics'. This limit is reached in the limit of high relative energies: In this limit, $k_r/\beta_s \gg 1$, exchange does not play a dominant role, because particles can approach one another even if they have parallel spins, since the large relative energy pushes them sufficiently close. Not so in the opposite limit of small relative energies, $k_r/\beta_s \ll 1$. In this case Eqn (630) approaches half of the classical result,

$$\frac{1}{\tau_{ee, self}(k_r)} \to \frac{e^4 m^* k_r n(\mathbf{r}, t)}{8\pi \epsilon_s^2 \hbar^3 \beta_s^4} \qquad (\frac{k_r}{\beta_s} \to 0),$$

since particles with parallel spins remain separated by the exclusion principle and their interaction vanishes. This is the limit considered by Goodnik and Lugli.

• Long-range Coulomb interactions and band-gap narrowing.

The subject of Coulomb interactions in an electron gas has been discussed for decades. Even for the 'simple' case of a homogeneous electron gas the calculations are daunting, because of the complex nature of the many-body interactions with exchange and correlation effects, as well as of the interactions between the electrons and the

'jellium' of the background positive charge required to keep the system at charge neutrality, without 'exploding' as a result of Coulomb repulsion. Here, let's see the basic features of this system in a classical picture, mentioning later the main quantum-mechanical corrections.

- Classical Model.

Consider a homogeneous electron gas of density $n = N/\Omega$, where N is the number of electrons and Ω the volume occupied by the gas. Ignoring Coulomb interactions, the total energy W per electron is purely the Fermi kinetic energy K_F :

$$W = K_F = \frac{g}{n} \int \frac{d\mathbf{k}}{(2\pi)^3} E_0(\mathbf{k}) f_{FD}[E_0(\mathbf{k})] , \qquad (632)$$

where $f_{FD}(E) = \{1 + \exp[(E_0 - E_F)/(k_B T)]\}^{-1}$, is the equilibrium Fermi function at temperature T, k_B is the Boltzmann constant, $E_0(\mathbf{k})$ is the g-fold degenerate electron dispersion, and the Fermi energy E_F is determined so that

$$n = g \int \frac{d\mathbf{k}}{(2\pi)^3} f_{FD}[E_0(\mathbf{k})] .$$
 (633)

In the simplest possible picture, the Hartree model, the spatial distribution of both electrons and background positive charge (donors), required by charge neutrality, is assumed to be homogenous and uniform. In this model the Coulomb interaction between electrons and donors vanishes. The first correction to this model consists in what Kittel calls the 'modified Hartree' approximation. In this model one assumes that the electrons still constitute a uniform distribution of charge, but they are now embedded in a lattice of background, positively charged, pointlike donors. Customarily, one thinks of the donor potential as statically screened by the electrons according to the Debye-Hückel theory, resulting in a Yukawa-like potential $e^{-\beta_s r}/(4\pi\epsilon_s^{\infty}r)$, where $\beta_s = [(e^2/\epsilon_s^{\infty})(\partial n/\partial E_F)]^{1/2}$ is the screening parameter, ϵ_s^{∞} being the static (valence) permittivity of Si, and *e* the magnitude of the electron charge. However, as we already saw when dealing with the potential of an ionized impurity, Mahan, employing quantum-mechanical variational calculations, has shown that this picture is correct only at moderate densities ($\sim 10^{19}$ cm⁻³ or below in Si), but at higher densities the assumption of a uniform electron distribution gives better agreement with the variational results. This is because the

Yukawa-model overcounts the electrons: On the one hand, they are assumed to completely screen the donors, and so to be fully localized around each dopant ion. On the other hand, they are assumed to be uniformly distributed.

In order to compute the Coulomb energy of this model system, in principle one should follow a procedure similar to what is required to compute the Madelung energy in crystals and derive the positions of the donors by minimizing their total energy. Here we assume the impurities frozen in a cubic lattice, and consider spherical volumes of radius $r_0 = [3/(4\pi n)]^{1/3}$ around each donor. Thus, the total energy per dopant (*i.e.*, in a volume $(4/3)\pi r_0^3$ around each impurity) of the electron gas and background pointlike donors will be lowered by the electron-impurity Coulomb attraction,

$$\delta W_{ed} \approx -\frac{e^2}{(4/3)\pi r_0^3} \int_0^{r_0} dr \frac{r}{\epsilon_s^\infty} = -\frac{3}{2} \frac{e^2}{4\pi \epsilon_s^\infty r_0} = -\frac{3}{2} \left(\frac{4\pi}{3}\right)^{1/3} \frac{e^2 n^{1/3}}{4\pi \epsilon_s^\infty} \approx -2.418 \ \kappa \ n^{1/3} ,$$
(634)

where $\kappa = e^2/(4\pi\epsilon_s^{\infty})$, and raised by the Coulomb repulsion of the uniform electron distribution with itself within an elemental spherical volume ,

$$\delta W_{ee} \approx \frac{e^2}{[(4/3)\pi r_0^3]^2} \int_0^{r_0} dr \frac{4}{3}\pi r^3 4\pi r^2 \frac{1}{4\pi\epsilon_s^\infty r}$$
$$= \frac{3}{5} \frac{e^2}{4\pi\epsilon_s^\infty r_0} = \frac{3}{5} \left(\frac{4\pi}{3}\right)^{1/3} \frac{e^2 n^{1/3}}{4\pi\epsilon_s^\infty} \approx 0.967 \ \kappa \ n^{1/3} \ . \tag{635}$$

Note that the mutual donor-donor repulsion vanishes in this model, since the impurities in different volume elements are viewed as fully screened by the surrounding electrons when frozen in their minimum-energy configuration. Therefore, the total energy per impurity atom due to the Coulomb interaction will be:

$$\delta W_C = \delta W_{ed} + \delta W_{ee} \approx -1.451 \kappa n^{1/3} .$$
(636)

Depending on the configuration of the impurities, one may actually obtain for the 'Madelung coefficient' 1.451

above a slightly different value of 1.444 for close-packed fcc, hcp, or bcc lattices. Here, we shall not worry about these numerically small differences.

At densities low enough for the standard Debye-Hückel screening to be valid, an expression given by Lanyon and Tuft may more appropriately replace Eq. (636),

$$\delta W_C = -\frac{3}{4} \kappa \beta_s . \tag{637}$$

As discussed by Mahan, the numerical values provided by Eqns. (636) and (637) are very similar in the range of densities of interest here, since the three characteristic lengths one can consider, the inverse wavevector at the Fermi surface k_F , the screening length β_s^{-1} , and the average donor separations r_0 , are not too different in Si.

In the presence of random motion, induced by the Fermi kinetic energy and, at nonzero temperatures, by the electron-phonon interactions, the electrons will move around their 'frozen' uniform configuration and set-up a fluctuating potential whose spatial Fourier components are identified with plasmons. In this case the total energy δW_C will redistribute into a potential energy, δU_C , and a kinetic energy component δK_C . The virial theorem can be used to estimate the time-average of the kinetic energy δK_C once δW_C is known. Since for N particles interacting via a central potential V(r), a function only of the interparticle distance r,

$$<\delta K_C> = \frac{1}{2} < \sum_{i\neq j} \nabla_i V(\mathbf{r}_i - \mathbf{r}_j) \cdot \mathbf{r}_i > ,$$
 (638)

where the brackets < ... > denote time-average, for a purely Coulomb system

$$<\delta K_C> = -\frac{1}{2} < \delta U_C>, \qquad (639)$$

so that (dropping the time-average brackets here and in the following, since we shall deal exclusively with time- or ensemble-average, equivalent by the ergodic theorem)

$$\delta K_C \approx 1.451 \kappa n^{1/3} . \tag{640}$$

The final quantity to consider is the shift of the chemical potential, $\delta\mu$, (in principle different from the Fermi energy E_F for interacting particles, as discussed below) induced by the Coulomb interactions among the N electrons and the N_D donors. At zero temperature or, more generally, in degenerate situations, $\delta\mu$ gives the electron contribution to the narrowing of the energy gap. In general the chemical potential μ of a grand canonical ensemble is defined as (see any text on Statistical Mechanics, such as Kubo's 1964 book):

$$\mu = \frac{1}{\Omega} \frac{\partial F}{\partial n} , \qquad (641)$$

at constant temperature and volume Ω , where $F = W_{tot} - TS$ is the (Helmholtz) free energy of the system, W_{tot} being its total energy and S its entropy. Therefore, at zero temperature the change of the chemical potential caused by the Coulomb interactions will be:

$$\delta\mu = \frac{1}{\Omega} \frac{\partial \delta W_{C,tot}}{\partial n} , \qquad (642)$$

where $\delta W_{C,tot} = N_D \delta W_C$ is the total Coulomb energy of the system. Keeping the volume Ω constant means that the identity $N_D = N$ should be used only after having taken the derivative. Thus

$$\delta \mu = \frac{\partial}{\partial N} (N_D \ \delta W_C) = -0.481 \ \kappa \ n^{1/3} \ . \tag{643}$$

Ignoring band-tailing and renormalization of the electron dispersion, since the total number of particles remains constant when accounting for Coulomb interactions, this shift of the chemical potential must be accompanied by an equal downward shift of the reference energy, that is, the bottom of the conduction band, $\delta E_{CB} = \delta \mu$.

Quantum corrections

The modified Hartree model considered so far ignores the effect of the electron spin: Since Pauli's principle prevents electrons with parallel spins from getting too close to each other, their mutual Coulomb repulsion will be reduced by an amount known as exchange energy, δW_x . Following Haas and Mahan, for Si the exchange

energy per particle is given by

$$\delta W_x = -\frac{3}{4} \frac{e^2 \Lambda k_F}{4\pi^2 \epsilon_s^\infty} \approx -0.3861 \ \kappa \ n^{1/3} , \qquad (644)$$

where $\Lambda = (m_l/m_t)^{1/3} (\tan^{-1} \delta/\delta)$, $\delta = [(m_l/m_t) - 1]^{1/2}$, $m_{t/l}$ denoting the transverse/longitudinal mass in the six ellipsoidal valleys of the Si conduction band. This second correction amounts to the Hartree-Fock results. The difference between this result and the energy of the ground state one would obtain performing an 'exact' calculation is called 'correlation' energy: The electrons are not uniformly distributed, but the energy of the system is lowered when accounting for a correlation of electron wavefunctions (or positions) which minimizes their mutual Coulomb repulsion. In the range of densities considered here, an expression for the correlation energy, δW_{corr} has been given by Gell-Mann and Brueckner:

$$\delta W_{corr} = -0.096 + 0.00622 \ln r_s \tag{645}$$

while Nozières and Pines give:

$$\delta W_{corr} = -0.115 + 0.0031 \ln r_s , \qquad (646)$$

where the unit of energy is the Rydberg (\approx 32.1 meV in Si) and r_s is the average separation r_0 in units of the Bohr radius (\approx 1.92 nm in Si), so that $r_s \approx 3.231 \ (n/10^{18})^{-1/3}$, where the electron density n is measured in cm⁻³. Re-expressing Eq. (636) in atomic units, we have

$$\delta W_C = -\frac{1.798}{r_s} \,, \tag{647}$$

while the exchange term, Eq. (644), is

$$\delta W_x = -\frac{0.479}{r_s} \,, \tag{648}$$

so we see that the correlation energy can be neglected, while ignoring the exchange term, Eq. (644), amounts to underestimating the Coulomb energy by about 26%. Correspondingly, the kinetic energy of the system will be modified only by a small term, usually called 'correlation kinetic energy'.

Finally, note that these corrections to the total energy of the electron gas at high densities are some of the causes of the well-known band-gap narrowing in semiconductors: We have already noticed above how the Coulomb interactions in a classical three-dimensional electron gas cause a shift $\delta E_{CB} = -0.481 \kappa n^{1/3}$ of the bottom of the conduction band. Mahan has evaluated the shifts of the chemical potential $\delta \mu_x = [e^2/(4\pi^2\epsilon_s^\infty)]k_F\Lambda \approx -0.515 \kappa n^{1/3}$ and $\delta \mu_{corr}$ associated to exchange and correlation effects, and so the corresponding shifts δE_{CB} . When accounting also for the contribution to the Coulomb energies due to the (minority) hole in the valence band (hole-donor and hole-electron interaction), Mahan has accounted for all of the contributions to the band-gap narrowing, with the exception of band-tailing, not discussed here. Clearly, the semiclassical model errs in ignoring exchange and correlation. As for its small contribution to the total energy, correlation can be safely ignored. Not so for exchange terms: Since the shift $\delta \mu$ due to the electron-donor and the hole-donor interaction cancel, exchange is now a dominant contribution.

• Impact Ionization.

Another example of electron-electron collision is 'impact ionization'. This is essentially a two-body collision, identical to what we have discussed above, but it involves *interband* transitions: Picture an electron traveling in the conduction band with a large kinetic energy. If this energy is sufficiently large, the 'primary' electron can collide with an electron in the valence band, 'kicking' this 'partner' electron into the conduction band, leaving a 'secondary' hole in the valence band. This process, which is the inverse of the Auger recombination mechanism, is responsible for avalanche breakdown phenomena in semiconductors and semiconductor devices.

The formal expression for the ionization rate can be obtained in a way very similar to the path we have followed above to calculate the collision rate for short-range electron-electron collisions. Within the Born approximation, the rate, $1/\tau_{ii}(\mathbf{k}n)$, at which an electron of quasi-momentum $\hbar \mathbf{k}$ in band n creates an electron-hole pair is given by:

$$\frac{1}{\tau_{ii}(\mathbf{k}n)} = \frac{2\pi}{\hbar} \sum_{\mathbf{GG'}} \sum_{n_v n'_c n''_c} \int \frac{d\mathbf{k'}}{(2\pi)^3} \int \frac{d\mathbf{p'}}{(2\pi)^3} |M_{\mathbf{GG'}}(\mathbf{k}n, \mathbf{p}n_v; \mathbf{k'}n'_c, \mathbf{p'}n''_c)|^2 \times \delta[E_n(\mathbf{k}) + E_{n_v}(\mathbf{p}) - E_{n'_c}(\mathbf{k'}) - E_{n''_c}(\mathbf{p'})]$$
(649)

where $E_n(\mathbf{k})$ the energy of the state \mathbf{k} in band n, and M is the (anti-)symmetrized screened Coulomb matrix element obtained by adding the scattering probabilities in the singlet and triplet states (as we saw before):

$$|M_{\mathbf{GG'}}|^2 = |M_d|^2 + |M_x|^2 - \frac{1}{2}|M_d^*M_x + M_x^*M_d|$$
(650)

 M_d and M_x being the 'direct' and 'exchange' matrix elements,

$$M_d = e^2 \epsilon_{\mathbf{G}\mathbf{G}'}^{-1}(\mathbf{q}_d, \omega_d) \frac{I_{\mathbf{G}}(\mathbf{k}n; \mathbf{k}'n_c') I_{\mathbf{G}'}(\mathbf{p}n_v; \mathbf{p}'n_c'')}{|\mathbf{q}_d + \mathbf{G}|^2},$$
(651)

$$M_x = e^2 \epsilon_{\mathbf{G}\mathbf{G}'}^{-1}(\mathbf{q}_d, \omega_x) \frac{I_{\mathbf{G}}(\mathbf{k}n; \mathbf{p}'n_c'')I_{\mathbf{G}'}(\mathbf{p}n_v; \mathbf{k}'n_c')}{|\mathbf{q}_x + \mathbf{G}|^2}.$$
 (652)

Here q_d , $\hbar \omega_d$ and q_x , $\hbar \omega_x$ are the direct and exchange quasi-momentum and energy transfers, respectively, defined as:

$$\mathbf{q}_d = \mathbf{k} - \mathbf{k}' \qquad \hbar \omega_d = E_n(\mathbf{k}) - E_{n'_c}(\mathbf{k}')$$
(653)

$$\mathbf{q}_x = \mathbf{k} - \mathbf{p}' \qquad \hbar \omega_x = E_n(\mathbf{k}) - E_{n_c''}(\mathbf{p}').$$
(654)

The vectors \mathbf{k}' , \mathbf{p} , and \mathbf{p}' are the quasi-momenta of the final state of the primary particle and of the initial and final states of the 'partner' particle, respectively. By momentum conservation they are related via $\mathbf{p}' = \mathbf{k} + \mathbf{p} - \mathbf{k}'$, up to a vector of the reciprocal lattice, already accounted for in the sums entering Eq. (649). The functions $I_{\mathbf{G}}(\mathbf{k}n;\mathbf{k}'n')$ are overlap integrals between Bloch states, summed over spin states $\sigma = \pm$:

$$\left|I_{\mathbf{G}}(\mathbf{k}n;\mathbf{k}'n')\right|^{2} = \frac{1}{2} \sum_{\sigma\sigma'} \left|I_{\mathbf{G}}(\mathbf{k}n\sigma;\mathbf{k}'n'\sigma')\right|^{2},\tag{655}$$

where:

$$I_{\mathbf{G}}(\mathbf{k}n\sigma;\mathbf{k}'n'\sigma') = \frac{1}{\Omega_c} \int_{cell} d\mathbf{r} \ u_{\mathbf{k}'n'}^{(\sigma')}(\mathbf{r})^* \ u_{\mathbf{k}n}^{(\sigma)}(\mathbf{r}) \ e^{i\mathbf{G}\cdot\mathbf{r}}, \tag{656}$$

where Ω_c is the cell volume and $u_{\mathbf{k}n}^{(\sigma)}(\mathbf{r}) = \sum_{\mathbf{G}} u_{\mathbf{G}}^{(\sigma)}(\mathbf{k}n) e^{i\mathbf{G}\cdot\mathbf{r}}$ is the periodic part of the (pseudo)wavefunction. Finally, the sum over reciprocal-lattive wavevectors \mathbf{G} in Eq. (649) includes Normal ($\mathbf{G} = \mathbf{0}$) as well as Umklapp ($\mathbf{G} \neq \mathbf{0}$) processes (referred to as N- and U-processes in the following). A similar expression, with the obvious changes, describes the rate for hole-initiated processes.

The difficulties in the evaluation of Eq. (649) are of both a physical and numerical nature, since, given the large energies involved, any form of approximation for the electron dispersion is going to be less than satisfactory (except for very few cases, like small band-gap materials). The full band-structure must be used if we want to get realistic results and this requires intensive state-of-the-art numerical work.

Physically, the major difficulty lies in choosing an accurate yet manageable expression for the (dynamic) dielectric matrix $\epsilon_{\mathbf{GG'}}(\mathbf{q}, \omega)$. Within the random phase approximation (RPA), we saw that the dielectric function is given by the 'usual' Lindhard expression:

$$\epsilon_{\mathbf{G}\mathbf{G}'}(\mathbf{q},\omega) = \epsilon_0 \, \delta_{\mathbf{G}\mathbf{G}'} - \frac{e^2}{|\mathbf{q} + \mathbf{G}|^2} \sum_{nn'\mathbf{k}} \frac{f_{n'}(\mathbf{k} + \mathbf{q}) - f_n(\mathbf{k})}{E_{n'}(\mathbf{k} + \mathbf{q}) - E_n(\mathbf{k}) + \hbar\omega}$$
$$\times \langle \mathbf{k}, n | e^{-i(\mathbf{q} + \mathbf{G}) \cdot \mathbf{r}} | \mathbf{k} + \mathbf{q}, n' \rangle \langle \mathbf{k} + \mathbf{q}, n' | e^{i(\mathbf{q} + \mathbf{G}') \cdot \mathbf{r}} | \mathbf{k}, n \rangle,$$
(657)

where ϵ_0 is the permittivity of vacuum and $f_n(\mathbf{k})$ the occupation of the state $|\mathbf{k}, n >$ of quasi-momentum \mathbf{k} in band n. It is customary to consider only the longitudinal response (*i.e.*, the diagonal terms of the dielectric function) and it is convenient to separate Eq. (657) into the two terms representing the response of the valence electrons $\epsilon_{\mathbf{GG}}^{(v)}$ (simply written now as $\epsilon_{\mathbf{G}}^{(v)}$) and of the free carriers, by setting $f_n(\mathbf{k}) = f_n^0(\mathbf{k}) + p_n(\mathbf{k})$, $f_n^0(\mathbf{k})$ being unity for all states in the valence bands, zero otherwise, and $p_n(\mathbf{k})$ being the population of the free carriers. Thus, Eq. (657) can be recast as follows:

$$\epsilon_{\mathbf{G}}(\mathbf{q},\omega) = \epsilon_{\mathbf{G}}^{(v)}(\mathbf{q},\omega) \left[1 + \epsilon_{\mathbf{G}}^{(v)-1}(\mathbf{q},\omega) \frac{\epsilon_0 \beta^2 (\mathbf{q} + \mathbf{G},\omega)}{|\mathbf{q} + \mathbf{G}|^2} \right],$$
(658)

having made use of a free-carrier screening parameter $\beta^2(\mathbf{q},\omega)$ defined as:

$$\beta^{2}(\mathbf{q},\omega) = -\frac{e^{2}}{\epsilon_{0}} \sum_{nn'\mathbf{k}} \frac{p_{n'}(\mathbf{k}+\mathbf{q}) - p_{n}(\mathbf{k})}{E_{n'}(\mathbf{k}+\mathbf{q}) - E_{n}(\mathbf{k}) + \hbar\omega} | < \mathbf{k}, n | e^{-i\mathbf{q}\cdot\mathbf{r}} |\mathbf{k}+\mathbf{q}, n' > |^{2}.$$
(659)

An approximation usually tacitly made consists in considering only the G-vector which minimizes the denominator of the Coulomb matrix elements, because of their small contribution (as $|\mathbf{q} + \mathbf{G}|^{-4}$) at short wavelengths in Eq. (649). Thus, one can drop the subscript **G** and write simply $\epsilon(\mathbf{q}, \omega)$, understanding that all vectors in k-space must be mapped to the first BZ. The calculation can proceed by ignoring the contribution of the free carriers, and employing the RPA and the pseudopotential band structure to evaluate $\epsilon^{(v)}(\mathbf{q}, \omega)$ while retaining its frequency dependence. In general, though, in treating the response of the carriers in the valence bands, the static approximation (*i.e.*, setting $\omega = 0$ in Eq. (657)) is a satisfactory simplification, since the energy exchanged in a pair-production process in realistic situations is relatively small (< 3 eV or so), so that the static value is adequate. So, several approximations have been made, ranging from the use of the static, wavevector dependent expression for $\epsilon^{(v)}(\mathbf{q})$ to approximating $\epsilon_0\beta^2(\mathbf{q})/\epsilon^{(v)}(\mathbf{q})$ with β_s^2 , the 'usual' Thomas-Fermi screening parameter $(e^2/\epsilon_s)\partial n_f/\partial E_F$, where E_F is the Fermi level , $\epsilon_s = \epsilon^{(v)}(\mathbf{q} = 0)$ is the static, long-wavelength dielectric function of the semiconductor, and n_f the density of free carriers, so that:

$$\epsilon(\mathbf{q}) \simeq \epsilon^{(v)}(\mathbf{q}) \left[1 + \frac{\beta_s^2}{q^2} \right],$$
(660)

This expression approaches correctly $\epsilon^{(v)}(\mathbf{q})$ at short-wavelengths, and is generally valid at long wavelengths, independent of the form of $\epsilon^{(v)}(\mathbf{q})$. Note also that dynamic corrections – in principle important in treating the response of the free carriers whose plasma energy is much lower than the typical energy exchanged in ionization events – become neglible at long wavelengths, so that the use of a static $\beta(\mathbf{q}, \omega)$ appears justified in this

context. Finally, in an attempt to simplify the calculation as much as possible, one may consider the static, long wavelength limit for the response of the valence carriers, by setting $\epsilon_v(\mathbf{q}) = \epsilon_s$ at all wavelengths, despite its questionable validity, and the Thomas-Fermi approximation for the free carriers. One possibility is the use of Eq. (660), using an expression proposed by Nara and Morita (see page 234, Eq. (614)) for $\epsilon^{(v)}(\mathbf{q})$, and assuming a small density of free electrons (say, $n_f = 10^{16} \text{ cm}^{-3}$), so that the Thomas-Fermi expression for β_s^2 reduces to its non-degenerate Debye-Hückel form $e^2 n_f / (\epsilon_s k_B T)$.

From a numerical point of view, local (Cohen-Bergstresser) or nonlocal (Cohen-Chelikowsky) empirical pseudopotentials have been used to evaluate the band structure and the pseudo-wavefunctions for 152 points in the irreducible wedge of the first Brillouin zone (BZ). The evaluation of Eq. (649) provides an anisotropic ionization rate, function of the quasi-momentum \mathbf{k} of the ionizing carrier and of the band index n.

In addition to the 'correct' evaluation of the anisotropic ionization rate expressed by Eq. (649), other useful approximations can be employed. The first one, the 'constant matrix element' (CME) approximation, emphasizes the role played (or, more likely, *not* played) by the matrix element in determining the energy dependence of the ionization rate. Accordingly, one assumes that the Coulomb matrix element M in Eq. (649) is a constant, equal to its average value over all **k**-points in the whole BZ, over all bands, and over both N and U-processes. Thus:

$$|M_{\mathbf{GG'}}|^2 \simeq \langle M^2 \rangle = \frac{e^4 a_0^4}{(2\pi)^4 \epsilon_s^2} m^2 ,$$
 (661)

where m is a number of order 1 and:

$$\frac{1}{\tau_{ii}^{(CME)}(\mathbf{k}n)} \simeq \frac{2\pi}{\hbar} < M^2 > \sum_{\mathbf{G}} \sum_{n_v n'_c n''_c} \int \frac{d\mathbf{p}}{(2\pi)^3} \,\delta(\mathbf{k} + \mathbf{p} - \mathbf{k}' - \mathbf{p}' + \mathbf{G}) \\ \times \int \frac{d\mathbf{k}'}{(2\pi)^3} \,\int \frac{d\mathbf{p}'}{(2\pi)^3} \,\delta[E_n(\mathbf{k}) + E_{n_v}(\mathbf{p}) - E_{n'_c}(\mathbf{k}') - E_{n''_c}(\mathbf{p}')].$$
(662)

Note that we have explicitly re-introduced the delta-function expressing momentum conservation in the first (trivial) integral in the equation above, in order to make it clear in the following how the CME and random-**k** approximations can be related. The actual value of the average matrix element $\langle M^2 \rangle$ can be determined, whenever Eqns. (649) and (662) exhibit the same isotropic energy dependence, by adjusting its value in Eq. (662) in order to match the result of Eq. (649).

An even stronger approximation, motivated by its success in the case of electron-initiated ionizations in Si (Kane, 1967), consists in assuming that the kinematics of the ionization process is not controlled by momentum conservation, but mainly by the (joint) density of states available to the final (recoil and ionized) particles. This is expected to be approximately true when momentum randomization via U-processes is the most probable pair-production channel. In this case, many nonzero **G**-vectors contribute in the sum in Eq. (662) above. Thus, it is sensible to make the approximation, known as the 'random-k approximation' (E. O. Kane, 1967), of replacing the the many 'spikes' of the momentum-conserving delta function with a uniform sampling of the entire Brillouin Zone, thus rewriting Eq. (662) as:

$$\frac{1}{\tau_{ii}^{(CME \to relx)}(\mathbf{k}n)} = \frac{2\pi}{\hbar} < M^2 > \sum_{n_v n'_c n''_c} \int_{BZ} \frac{d\mathbf{p}}{\Omega_{BZ}}$$
$$\times \int \frac{d\mathbf{k}'}{(2\pi)^3} \int \frac{d\mathbf{p}'}{(2\pi)^3} \,\delta[E_n(\mathbf{k}) + E_{n_v}(\mathbf{p}) - E_{n'_c}(\mathbf{k}') - E_{n''_c}(\mathbf{p}')] \,, \tag{663}$$

where now, in place of the sum over G-vectors, p runs over the Brillouin Zone, restricted only by energy conservation. Let's now recall that the volume of the Brillouin Zone, Ω_{BZ} , is given by $(4\pi/a)^3/2 =$

 $(2\pi)^3/\Omega_c$, and employ the identity:

$$\int \, dE \, \, \rho_{c/v}(E) = 2 \; \sum_{n_{c/v}} \; \int \; \frac{d{\bf k}}{(2\pi)^3} \; , \label{eq:rho}$$

where ρ_c and ρ_v are the densities of states in the conduction and valence bands, respectively, accounting for spin degeneracy and always measuring energies from the band extrema with (kinetic hole) energies taken as positive in the valence bands. Thus, the ionization rate obtained within the CME approximation *and* having relaxed momentum conservation reduces to:

$$\frac{1}{\tau_{ii}^{(rk)}(E)} = \frac{2\pi \Omega_c}{\hbar 8} < M^2 > \int_0^{E-Eg} dE_c \int_0^{E-Eg-Ec} dE_v \ \rho_c(E_c)\rho_v(E_v)\rho_c(E-E_g-E_c-E_v),$$
(664)

where E_g is the energy gap and noting that $(2\pi)^3/(8 \ \Omega_{BZ}) = \Omega_c/8$ and having considered throughout electron-initiated processes, to fix the ideas. The quality of this approximation can be judged by whether the value of $\langle M^2 \rangle$ in Eq. (664) required to fit quantitatively the CME results matches the value used in Eq. (662).

Finally, we report here the popular Keldysh form of the ionization rate,

$$\frac{1}{\tau_{ii}(\mathbf{k}n)} = \frac{P}{\tau_{ep}(E_{th})} \left(\frac{E - E_{th}}{E_{th}}\right)^2,\tag{665}$$

where E_{th} is some ionization-threshold energy, P a coupling parameter related to the Coulomb matrix element, and $\tau_{ep}(E_{th})$ is either the carrier-phonon scattering rate at threshold, according the Keldysh's original formulation (Keldysh, 1960), or the energy relaxation rate at threshold, according to Ridley's lucky drift theory

(Ridley, 1987). This expression has been used extensively in the past, treating the parameters P and E_{th} as fitting quantities.

In the figures on the next page we see the ionization rates for electrons in Si as well as the so-called 'ionization coefficient', the number of pairs created per unit length in the presence of a uniform electric field F:

$$\alpha(F) = \frac{1}{v_d(F)} \int \frac{d\mathbf{k}}{(2\pi)^3} \frac{f_F(\mathbf{k})}{\tau_{ii}(\mathbf{k})} , \qquad (666)$$

where $v_d(F)$ is the average electron drift velocity in the field F and f_F is the electron distribution. An empirical epression commonly used is:

$$\alpha(F) = \frac{eF_{th}}{E_{th}} e^{-F_{th}/F} ,$$

where F_{th} is some threshold field and E_{th} the energy threshold for pair generation.





Radiative Transitions

So far we have considered processes in which electrons either did not gain or lose energy (elastic processes, as impurity scattering), or in which energy is exchanged within the lattice (with ions, as in electron-phonon interactions, or with other electrons). These are called 'non-radiative' processes, since energy is not transferred to radiation (that is, electromagnetic waves, light). Here let's consider processes in which electrons interact with light. We shall focus on light absorption across the band-gap and mention only the existence of other processes.

• Electron-photon Hamiltonian.

We have already used before (see Notes, Part 1, page 108, Eq. (194)) the fact that the Hamiltonian of an electron in the presence of an electromagnetic field can be obtained from the Hamiltonian in the absence of the field but replacing the electron momentum $\mathbf{p} = -i\hbar\nabla$ with $\mathbf{p} - e\mathbf{A}$, where \mathbf{A} is the vector potential. Therefore, the single-electron Schrödinger equation reads:

$$i\hbar\frac{\partial\psi}{\partial t} = \frac{1}{2m}(\mathbf{p} - e\mathbf{A})^2\psi - e\varphi(\mathbf{r})\psi. \qquad (667)$$

Assuming that the electrostatic potential φ vanishes (so, assuming a purely electromagnetic field, such as in interactions with light) and retaining only terms first-order in the 'perturbation' **A**, we have:

$$i\hbar\frac{\partial\psi}{\partial t} = -\frac{\hbar^2}{2m}\nabla^2\psi + i\hbar\frac{e}{m}\mathbf{A}\cdot\nabla\psi.$$
(668)

The Lagrangian density yielding Eq. (668) via the Euler-Lagrange equations is (see Eq. (282) on page 132 of the Notes):

$$\mathcal{L}(\psi, \dot{\psi}, t) = i\hbar\psi^* \frac{\partial\psi}{\partial t} - \frac{\hbar^2}{2m} \nabla\psi^* \cdot \nabla\psi - i\hbar \frac{e}{m} \psi^* \mathbf{A} \cdot \nabla\psi , \qquad (669)$$

so that we can identify the perturbation Lagrangian with the last term. Therefore the perturbation Hamiltonian density will be:

$$\mathcal{H}_{em}(\psi,\pi,t) = \frac{e}{m}\pi \mathbf{A} \cdot \nabla \psi = i\hbar \frac{e}{m} \psi^* \mathbf{A} \cdot \nabla \psi .$$
(670)

Using the second-quantized version of the vector potential (Eq. (399), page 162 of the Notes) in terms of the photon creation and annihilation operators, we can write the Hamiltonian as (using a finite-volume normalization):

$$\mathbf{H}_{em} = \int d\mathbf{r} \,\mathcal{H}_{em}(\psi, \pi, t) = i\hbar \frac{e}{m} \int d\mathbf{r} \,\psi^* \mathbf{A} \cdot \nabla \psi = \sum_{\mathbf{k}\mathbf{k}'\mathbf{q}\lambda nn'} i\hbar \frac{e}{m} \left(\frac{\hbar}{2\epsilon_{\infty}\omega_q}\right)^{1/2} c^{\dagger}_{\mathbf{k}'}(b_{\mathbf{q}\lambda} + b^{\dagger}_{\mathbf{q}\lambda})c_{\mathbf{k}} \,\mathbf{e}_{\mathbf{q}\lambda} \cdot \frac{1}{V} \int d\mathbf{r} \,e^{-i\mathbf{k}'\cdot\mathbf{r}} \,u^*_{\mathbf{k}'n'}(\mathbf{r}) \,e^{-i\mathbf{q}\cdot\mathbf{r}} \,\nabla \,e^{i\mathbf{k}\cdot\mathbf{r}} \,u_{\mathbf{k}n}(\mathbf{r}) \,,$$
(671)

where n and n' are band indices. This expression is identical to the expression we have derived before for the electron-phonon interaction. The only nonzero matrix elements are those in which the final state differs from the initial by the presence of one fewer or additional photon of momentum \mathbf{q} and polarization λ (for absorption and emission processes, respectively associated to the term $b_{\mathbf{q}\lambda}$ and $b_{\mathbf{q}\lambda}^{\dagger}$) and by the conversion of the wavevector of one electron from \mathbf{k} to \mathbf{k}' . Such a matrix element will have the form:

$$<\mathbf{k}', n'|\mathbf{H}_{em}|\mathbf{k}, n> = \frac{ie\hbar}{m} \sum_{\mathbf{q}\lambda} \left\{ \begin{array}{c} N_{\mathbf{q}\lambda}^{1/2} \\ (1+N_{\mathbf{q}\lambda})^{1/2} \end{array} \right\} \left(\frac{\hbar}{2\epsilon_{\infty}\omega_{q}} \right)^{1/2} \times$$

$$\times \mathbf{e}_{\mathbf{q}\lambda} \cdot \frac{1}{V} \int d\mathbf{r} \ e^{-i\mathbf{k}' \cdot \mathbf{r}} \ u_{\mathbf{k}'n'}^*(\mathbf{r}) \ e^{-i\mathbf{q} \cdot \mathbf{r}} \ \nabla \ e^{i\mathbf{k} \cdot \mathbf{r}} \ u_{\mathbf{k}n}(\mathbf{r}) \ , \tag{672}$$

the factor $N_{q\lambda}$ being valid for absorption, $1 + N_{q\lambda}$ for emission. If the wavelength of the photon is much longer than the electron wavelength (which is always the case up into the visible/UV portion of the spectrum),

we can set $e^{-i{f q}\cdot{f r}}pprox 1$ and

$$<\mathbf{k}', n'|\mathbf{H}_{em}|\mathbf{k}, n>\approx \sum_{\mathbf{q}\lambda} \left\{ \begin{array}{c} N_{\mathbf{q}\lambda} \\ 1+N_{\mathbf{q}\lambda} \end{array} \right\} \frac{ie\hbar}{m} \left(\frac{\hbar}{2\epsilon_{\infty}\omega_{q}} \right)^{1/2} \mathbf{e}_{\mathbf{q}\lambda} \cdot \frac{1}{V} \int d\mathbf{r} \, e^{-i\mathbf{k}'\cdot\mathbf{r}} \, u_{\mathbf{k}'n'}^{*}(\mathbf{r}) \, \nabla \, e^{i\mathbf{k}\cdot\mathbf{r}} \, u_{\mathbf{k}n}(\mathbf{r}) \,.$$

$$\tag{673}$$

This is called the 'dipole approximation'. Now:

$$\int d\mathbf{r} \ e^{-i\mathbf{k'}\cdot\mathbf{r}} \ u^*_{\mathbf{k'}n'}(\mathbf{r}) \ \nabla \ e^{i\mathbf{k}\cdot\mathbf{r}} \ u_{\mathbf{k}n}(\mathbf{r}) =$$

$$= \int d\mathbf{r} \ e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}} \ u^*_{\mathbf{k}'n'}(\mathbf{r}) \ \nabla \ u_{\mathbf{k}n}(\mathbf{r}) \ + \ \int \ d\mathbf{r} \ e^{-i\mathbf{k}'\cdot\mathbf{r}} \ u^*_{\mathbf{k}'n'}(\mathbf{r}) \ u_{\mathbf{k}n}(\mathbf{r}) \nabla e^{i\mathbf{k}\cdot\mathbf{r}} \ .$$
(674)

The second term contains an integral which is significantly different from zero only when $\mathbf{k}' \approx \mathbf{k}$. In this case the Bloch functions $u_{\mathbf{k}'n'}(\mathbf{r})$ and $u_{\mathbf{k}n}(\mathbf{r})$ are orthogonal and the exponential is slowly varying. Therefore the second integral is usually neglected. The first term can be treated in the 'usual' way, by splitting the integral over the entire volume into a sum of integrals over individual cells by setting $\mathbf{r} \to \mathbf{R}_l + \mathbf{r}$:

$$<\mathbf{k}'n'|\mathbf{H}_{em}|\mathbf{k}n>\approx \frac{ie\hbar}{m}\sum_{\mathbf{q}\lambda} \left\{ \begin{array}{c} N_{\mathbf{q}\lambda}^{1/2} \\ (1+N_{\mathbf{q}\lambda})^{1/2} \end{array} \right\} \left(\frac{\hbar}{2\epsilon_{\infty}\omega_{q}} \right)^{1/2} \mathbf{e}_{\mathbf{q}\lambda} \cdot \sum_{l} e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{R}_{l}} \frac{1}{V} \int_{\Omega} d\mathbf{r} \, u_{\mathbf{k}'n'}^{*}(\mathbf{r}) \nabla \, u_{\mathbf{k}n}(\mathbf{r}) \left(\frac{\hbar}{2\epsilon_{\infty}\omega_{q}} \right)^{1/2} \mathbf{e}_{\mathbf{q}\lambda} \cdot \sum_{l} e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{R}_{l}} \frac{1}{V} \int_{\Omega} d\mathbf{r} \, u_{\mathbf{k}'n'}^{*}(\mathbf{r}) \nabla \, u_{\mathbf{k}n}(\mathbf{r}) \left(\frac{\hbar}{2\epsilon_{\infty}\omega_{q}} \right)^{1/2} \mathbf{e}_{\mathbf{q}\lambda} \cdot \sum_{l} e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{R}_{l}} \frac{1}{V} \int_{\Omega} d\mathbf{r} \, u_{\mathbf{k}'n'}^{*}(\mathbf{r}) \nabla \, u_{\mathbf{k}n}(\mathbf{r}) \left(\frac{\hbar}{2\epsilon_{\infty}\omega_{q}} \right)^{1/2} \mathbf{e}_{\mathbf{q}\lambda} \cdot \sum_{l} e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{R}_{l}} \frac{1}{V} \int_{\Omega} d\mathbf{r} \, u_{\mathbf{k}'n'}^{*}(\mathbf{r}) \nabla \, u_{\mathbf{k}n}(\mathbf{r}) \left(\frac{\hbar}{2\epsilon_{\infty}\omega_{q}} \right)^{1/2} \mathbf{e}_{\mathbf{q}\lambda} \cdot \sum_{l} e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{R}_{l}} \frac{1}{V} \int_{\Omega} d\mathbf{r} \, u_{\mathbf{k}'n'}^{*}(\mathbf{r}) \nabla \, u_{\mathbf{k}n}(\mathbf{r}) \left(\frac{\hbar}{2\epsilon_{\infty}\omega_{q}} \right)^{1/2} \mathbf{e}_{\mathbf{q}\lambda} \cdot \sum_{l} e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{R}_{l}} \frac{1}{V} \int_{\Omega} d\mathbf{r} \, u_{\mathbf{k}'n'}^{*}(\mathbf{r}) \nabla \, u_{\mathbf{k}n}(\mathbf{r}) \left(\frac{\hbar}{2\epsilon_{\infty}\omega_{q}} \right)^{1/2} \mathbf{e}_{\mathbf{q}\lambda} \cdot \sum_{l} e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{R}_{l}} \frac{1}{V} \int_{\Omega} d\mathbf{r} \, u_{\mathbf{k}'n'}^{*}(\mathbf{r}) \nabla \, u_{\mathbf{k}n'}(\mathbf{r}) \nabla \, u_{\mathbf{k}n'}(\mathbf{r}) \left(\frac{\hbar}{2\epsilon_{\infty}\omega_{q}} \right)^{1/2} \mathbf{e}_{\mathbf{k}'n'} \left(\frac{\hbar}{2\epsilon_{\infty}\omega_{q}} \right)^{1/2} \mathbf{e}_{\mathbf{k}'n'}(\mathbf{r}) \nabla \, u_{\mathbf{k}'n'}(\mathbf{r}) \nabla \, u_{\mathbf{k}'n'}(\mathbf{r})$$

$$= \frac{ie\hbar}{m} \sum_{\mathbf{q}\lambda} \left\{ \begin{array}{c} N_{\mathbf{q}\lambda}^{1/2} \\ (1+N_{\mathbf{q}\lambda})^{1/2} \end{array} \right\} \left(\frac{\hbar}{2\epsilon_{\infty}\omega_{q}} \right)^{1/2} \mathbf{e}_{\mathbf{q}\lambda} \cdot \frac{1}{\Omega} \int_{\Omega} d\mathbf{r} \ u_{\mathbf{k}n'}^{*}(\mathbf{r}) \nabla \ u_{\mathbf{k}n}(\mathbf{r}) \ \delta_{\mathbf{k}'\mathbf{k}} , \quad (675)$$

since $N_{cells}/V = 1/\Omega$. Thus, the optical matrix element vanishes unless $\mathbf{k}' = \mathbf{k}$. Only 'vertical (or direct) transitions' can happen. Transitions which involve a change in (crystal) momentum can only happen at higher orders in perturbation theory, requiring extra momentum which must be provided by phonons, traps, etc.

Now let's define the 'momentum matrix element' between two Bloch functions:

$$< n'|\mathbf{p}|n> = -\frac{i\hbar}{\Omega} \int_{\Omega} d\mathbf{r} \ u_{\mathbf{k}n'}^*(\mathbf{r}) \nabla \ u_{\mathbf{k}n}(\mathbf{r}) \ .$$
 (676)

Then:

$$<\mathbf{k}'n'|\mathbf{H}_{em}|\mathbf{k}n>\approx -\frac{e}{m}\sum_{\mathbf{q}\lambda} \left\{ \begin{array}{c} N_{\mathbf{q}\lambda}^{1/2} \\ (1+N_{\mathbf{q}\lambda})^{1/2} \end{array} \right\} \left(\frac{\hbar}{2\epsilon_{\infty}\omega_{q}}\right)^{1/2} \mathbf{e}_{\mathbf{q}\lambda} \cdot < n'|\mathbf{p}|n> .$$
(677)

Note that this is the most notable example of a type of transition for which the form of the Bloch functions matters. In particular, well-known selection rules apply. For example, the the momentum matrix between atomic orbitals is nonzero only when the azimuthal quantum number m_J of the final and initial orbitals differs by zero or one. Similarly in a crystal, photon absorption can occur only between bands whose orbital character satisfies this selection rule. In addition, the knowledge of the Bloch functions is required to calculate the matrix element (677).

• Transition rates.

The rate at which a photon of momentum q and polarization λ is absorbed by an electron of wavevector k in band n can be calculated from Fermi Golden Rule:

$$W_{abs,n} = \frac{2\pi}{\hbar} \int \frac{d\mathbf{k}}{(2\pi)^3} \sum_{n'} \frac{\hbar e^2}{2\epsilon_{\infty}\omega_q m^2} N_{\mathbf{q}\lambda} |\mathbf{e}_{\mathbf{q}\lambda} \cdot \langle n'|\mathbf{p}|n \rangle |^2 \delta[E_n(\mathbf{k}) - E_{n'}(\mathbf{k}) - \hbar\omega_{\mathbf{q}\lambda}] =$$

$$= \frac{\pi e^2}{m^2 \omega \epsilon_{\infty}} N_{\omega} \sum_{n'} |\mathbf{e}_{\mathbf{q}\lambda} \cdot \langle n' |\mathbf{p}|n \rangle|^2 \rho_{n'}(E + \hbar \omega) , \qquad (678)$$

where $\rho_{n'}(E)$ is the density of (final) electron states in band n'.

On the contrary, for the rate at which an electron of wavevector ${f k}$ in band n emits a photon, we have (assuming

spontaneous emission only) will be:

$$W_{em} = \frac{2\pi}{\hbar} \int \frac{d\mathbf{q}}{(2\pi)^3} \sum_{n'} \frac{\hbar e^2}{2\epsilon_{\infty}\omega_q m^2} |\mathbf{e}_{\mathbf{q}\lambda} \cdot \langle n'|\mathbf{p}|n \rangle|^2 \delta[E(\mathbf{k}+\mathbf{q}) - E(\mathbf{k}) - \hbar\omega_{\mathbf{q}\lambda}] =$$

$$= \frac{\pi e^2}{m^2 \epsilon_{\infty}} | \langle n' | \mathbf{p} | n \rangle |_{ave}^2 \int \frac{q^2 dq}{8\pi^3 \omega} = \frac{\pi e^2}{m^2 \epsilon_{\infty}} | \langle n' | \mathbf{p} | n \rangle |_{ave}^2 \int d(\hbar\omega) \rho(\hbar\omega) , \quad (679)$$

where $| \langle n' | \mathbf{p} | n \rangle |_{ave}$ is the dipole matrix element summed over all bands and averaged over the photon polarizations and $\rho(\hbar\omega) = \omega^2/(2\pi^2 v_{\omega}^3)$ is the photon density of states expressed as a function of the phase velocity $v_{\omega} = \omega/q$.

• Absorption spectrum.

In the particular case in which n' labels the conduction band and n the valence band, the matrix element for the process in which a photon of momentum q and polarization λ is absorbed by the solid triggering the excitation of an electron from the valence to the conduction band will be:

$$<\mathbf{k}c|\mathbf{H}_{em}|\mathbf{k}v>\approx -\frac{e}{m}N_{\mathbf{q}\lambda}\left(\frac{\hbar}{2\epsilon_{\infty}\omega_{q}}\right)^{1/2}\mathbf{e}_{\mathbf{q}\lambda}\cdot < c|\mathbf{p}|v>$$
 (680)

The absorption rate will be, according to Eq. (678):

$$I(\hbar\omega) = \frac{\pi e^2}{m^2 \epsilon_{\infty} \omega} N_{\omega} \int dE \left| \mathbf{e} \cdot \langle c | \mathbf{p} | v \rangle \right|^2 \rho_c (E + \hbar\omega - E_{gap}) \left[1 - f_c (E + \hbar\omega - E_{gap}) \right] \rho_v(E) f_v(E) ,$$
(681)

where ρ_v and ρ_c are the DOS in the valence and conduction bands, respectively, and f_v and f_c the corresponding

Fermi occupation factors. Assuming a completely full VB and empty CB,

$$I(\hbar\omega) \approx \frac{\pi e^2}{m^2 \epsilon_{\infty} \omega} N_{\omega} \int dE |\mathbf{e} < c|\mathbf{p}|v > |^2 \rho_c(E + \hbar\omega - E_{gap}) \rho_v(E) .$$
 (682)

The factor $\rho_c(E + \hbar\omega - E_{gap}) \rho_v(E)$ is the 'joint density of states' of the valence and conduction bands. The absorption spectrum of a direct-gap semiconductor will show a sharp threshold at $\hbar\omega = E_{gap}$. For indirect-gap materials, instead, a weaker and softer threshold occurs at the gap, since the help of a phonon is necessary to conserve momentum and the matrix element in Eq. (681) is replaced by a second-order electron-phonon-photon matrix element.

One important feature of the absorption spectrum of several semiconductors is the appearance of 'exciton peaks'. When an electron-hole pair is formed by the absorption of a photon, the pair can remain weakly bound, as in an hydrogen atom. The binding energy of the pair will be exactly the binding energy of the H atom in its ground state, corrected for the different dielectric constant and for the reduced mass, $1/m_{red} = 1/m_e + 1/m_h$:

$$\Delta E = \frac{e^2}{(4\pi\epsilon_{\infty})^2} \frac{m_{red}}{2\hbar^2} .$$
(683)

Thus, the absorption spectrum shows a peak at a photon frequency $\hbar \omega = E_{gap} - \Delta E$ (see Fig. 5.4 in Ridley's book).

• A summary of radiative and non-radiative transitions.

The theoretical scheme we have employed so far can be used in a variety of different situations. Electronic transitions can occur within the conduction band, from valence to conduction band, from valence or conduction band to localized defect or impurity states (traps). They can be radiative (when assisted by photons) or non-radiative (when assisted by interparticle Coulomb forces, also called 'Auger processes', by phonons or by a more complicated lattice rearrangement). We have seen the cases of photon, phonon, and Coulomb matrix elements at first order. The general transition rate can be evaluated using first-order perturbation theory (Fermi Golden Rule or Born approximation for direct transitions). The major cases we have not tretaed are transitions

involving traps (which requires the knowledge of the localized wavefunction, often hard to estimate) while non-radiative transitions involving large energy exchange (such as an conduction electron being trapped into a localized state energetically located aroung the middle of the gap) may require multi-phonon processes (such as emission of many phonons or processes relying on an-harmonic corrections to the phonon Hamiltonian) or even more complicated processes in which the lattice re-arranges itself. Ridley's text describes in Chapter 5 radiative transitions of types different from the VB-to-CB direct absorption we have considered above: Free carrier absorption (that is: absorption of a photon by a free electron in the CB, involving an analysis of the requirements posed by energy and momentum conservation), photoionization of localized states of exciton absorption (requiring some discussion of the localized or excitonic wavefunctions), indirect transitions (assisted by phonons and so requiring second-order perturbation theory) and light scattering by free electrons (which depends on the second-order term $e^2A^2/(2m)$ we have so far neglected in Eq. (667)). In Chapter 6 Ridley discusses non-radiative transitions. While the details (mainly dictated by energy/momentum conservation and by the particular form of the initial/final wavefunctions involved) change significantly from one case to the next, the general scheme employed is always the same. We shall not discuss these issues in detail, as the examples we have seen so far should be sufficient to give the gist of the matter.