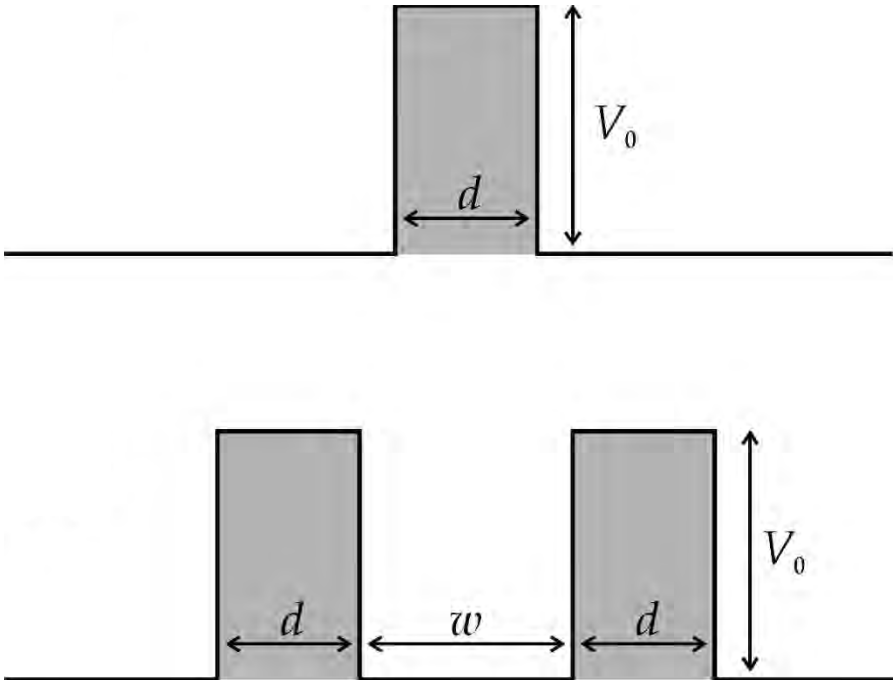
Electric, Optical & Magnetic Properties of Nanostructures

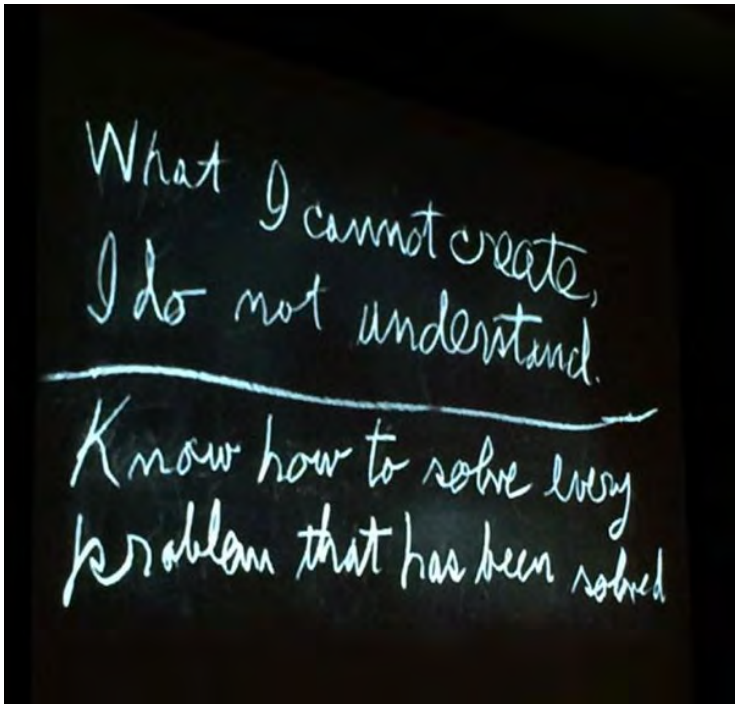


Thomas Garm Pedersen

Aalborg University

2022

Table of Contents



Richard Feynman's office black board at the time of his death.

Introduction

Properties are essentially about cause and effect. A material left completely undisturbed doesn't display its properties. However, if we probe something the response will reveal the characteristic properties of the object. Here, probing should be understood in its broadest sense. If we look at something it is probed by the light reflected or transmitted by the object. If we put an object on a table it is probed by contact forces. If we place the object on a hot plate it is probed by the heat. An endless list of external "perturbations" serve as probes of properties that, taken together, characterize the object. At a qualitative level we might label the object as shiny, hard, heavy etc. More accurately, however, we should specify its reflectance, hardness, heat capacity and so on as precisely defined properties that can be measured in a specific setup. This also allows us to describe theoretically the phenomenon in a precise manner. Hence, to this end we must construct a model that encompasses the object, the perturbation and the response.

The present set of lecture notes deals only with a small part of all the properties imaginable. We restrict ourselves to the following perturbations: electric potentials, light and magnetic fields. The responses to these stimuli define together the electric, optical and magnetic properties of an object. Moreover, with a few exception, we exclude cases, where any two of these perturbations are present at the same time. Finally, we restrict the analysis to weak disturbances that don't perturb significantly the object. This might seem as rather severe restrictions but, actually, they are not. They will allow us to understand a wide range of phenomena such as the color of an object. And within this restricted analysis we'll still be able to explain several important features of advanced devices such as diodes, lasers and magnetic hard discs.

What is characteristic of electric, optical and magnetic properties? To answer this question, we should specify some appropriate experimental setup, in which we wish to measure a given property. Focusing on the perturbations, we might think of something like the sceneries shown in the Fig. I.1. Here, the bar represents a sample that we want to characterize. The electrical setup in the upper panel is probably the most familiar type of setup. In this case, the perturbation is the voltage applied between the end points. The response is the current, which flows in the circuit. Hence, perturbation and response can be determined experimentally by means of a voltmeter and amp meter, respectively. Really, such a measurement is tricky because not only the sample plays a role in determining the current. Resistances in the wires and power supply in general cannot be ignored and so more advanced methods (such as four-point measurements) may be needed. But what is the role of the sample? Well, for a voltage V the current I is limited by the electrical resistance $R = V / I$ or, if we like, we could measure the conductance $G = I / V = 1 / R$. The point here is that these quantities depend on both the *material* in the sample and *shape* of the sample. For a large homogeneous bar of length L and cross section A the resistance is expected to

vary as $R = \rho L / A$, where ρ is the specific resistance of the material. In turn, ρ is a characteristic of the material and depends only on external parameters such as temperature and pressure and so on.

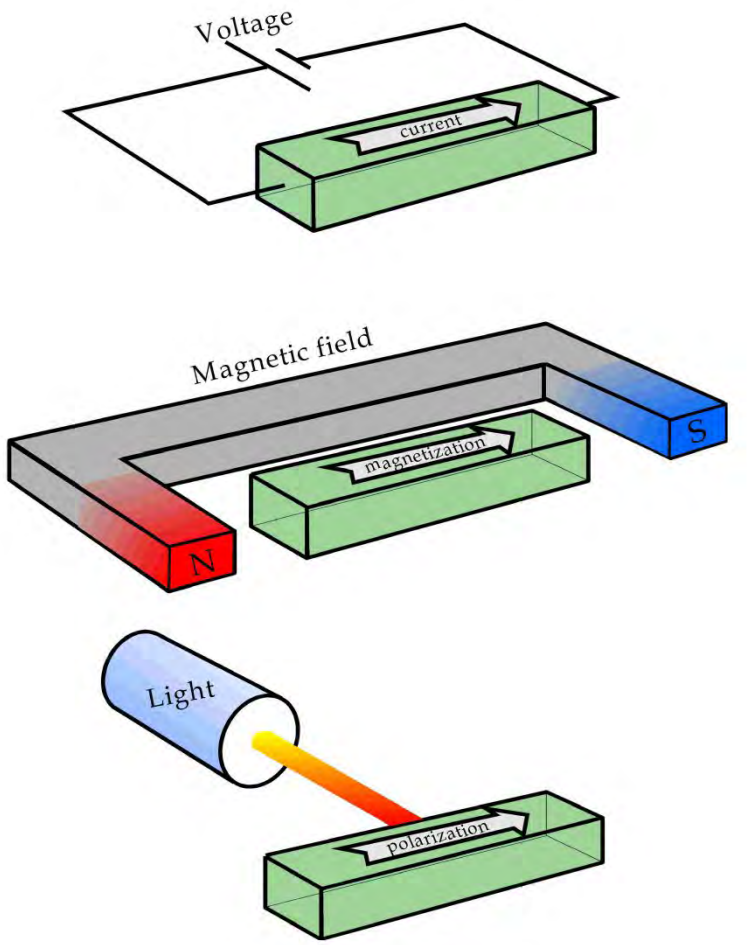


Figure I.1. Schematic illustration of electric, magnetic and optical perturbations of sample.

Next, we turn to the magnetic response. As shown in the middle panel of the figure, we could imagine placing a sample between the poles of a magnet. Hence, the perturbation in this case is the magnetic field penetrating the sample. What happens inside the sample? The answer to this question depends on the nature of the sample. We consider first the case of so-called non-magnetic materials. Imagine the external magnetic field intensity $\mathcal{H}$ penetrating the sample. Inside the sample, the electrons act as tiny bar magnets themselves. A characteristic property of non-magnetic materials is that these tiny bar magnets point in all directions with equal probability if the external magnetic field is switched off. However, when the external field $\mathcal{H}$ is applied they will try to align with this field. The result is that the material becomes *magnetized* and we say that a certain *magnetization* $\mathcal{M}$ has been induced. Since the magnetization is induced by the magnetic field we expect a linear dependence if the field is not too large: $\mathcal{M} = \chi_M \mathcal{H}$, where χ_M is the magnetic susceptibility. Hence, in this case, χ_M is the important material quantity, which we wish to describe.

Measuring χ_M is less simple than measuring e.g. the resistance described above. One possibility is to use a flux meter to record the flux of the total magnetic field $\mathcal{B} = \mu_0(\mathcal{H} + \mathcal{M}) = \mu_0(1+\chi_M)\mathcal{H}$ through a wire loop as the external field $\mathcal{H}$ is turned on.

The exception to this behavior of non-magnetic materials is found in *magnetic materials*. These are substances, in which a finite magnetization can exist even without an external field. The challenge in this case is to understand this phenomenon and to see how such a material will be influenced by an external field. An extremely important aspect of this case is the change in the *direction* of the magnetization that can be induced by an external field. This forms the basis for magnetic storage devices.

The optical response is, in fact, quite similar to the magnetic one. In this case, a high-frequency electric (and magnetic) field $\mathcal{E}$ is incident on our sample and this field displaces the charges in the material. In the optical regime, the relevant charges are electrons as the nuclei respond primarily to fields of much lower frequency. The displaced electron charges are described by the *polarization* $\mathcal{P}$ and in perfect analogy with the magnetic case we expect a linear relation $\mathcal{P} = \varepsilon_0 \chi \mathcal{E}$, where χ is now the electric susceptibility. In turn, χ determines the dielectric constant of the material $\varepsilon = 1 + \chi$ and also the refractive index $n = \sqrt{1+\chi}$. Hence, these quantities are the relevant material properties. Measuring the refractive index is in itself a complicated task. The imaginary part reveals itself in absorption measurements, while the real part is responsible for refraction and interference effects. The technique known as ellipsometry is developed for the purpose of accurately measuring these quantities.

So far, the discussion has been quite general and applicable to many circumstances. The focus of these notes, however, will be on nanostructures. The reason is simply that our ability to design and fabricate structures on a nanometer scale has expanded the range of electric, optical and magnetic phenomena in a truly amazing manner. There are several reasons for this. Most importantly, the rules of physics are different on the nanoscale. Pronounced quantization effects appear and, hence, scaling bulk results such as $R = \rho L / A$ into the nanoscale simply doesn't work if the cross section A becomes sufficiently small. A dramatic consequence in this particular case is the appearance of quantized conductance, in which individual quantum levels determine the resistance. Similarly, electric and magnetic susceptibilities are entirely different in low-dimensional structures. This opens the window to novel phenomena as well as brand new applications. Even if the material properties are almost identical to the bulk values, new effects can appear if the size of the sample is in the nanoscale. An important example is found in nanooptics. Here, optical components of a size comparable to the wavelength are used and this dramatically changes the way light propagates, diffracts and so on. Actually, completely new *metamaterials* can be fabricated in this fashion.

Through these lecture notes I hope to display the differences between the properties of bulk materials and nanostructures. I do this by describing both so that a comparison can be made. Our approach consists in the formulation of a very general framework for the response to external perturbations in the bulk as well as in nanostructures. We then turn to all the different electric, optical and magnetic applications of this general framework in the subsequent chapters. The notes are intended for students in the final year of undergraduate study having already established a solid knowledge in quantum mechanics, statistical mechanics, solid state physics as well as optics and electromagnetism. Several people made helpful comments on the manuscript and, in particular, I wish to thank Thomas Bastholm Lynge, Kjeld Pedersen, Mads Lund Trolle, Jesper Jung, Jonas Have, Thomas Søndergaard and Alireza Taghizadeh.

1. Time-Dependent Perturbation Theory

In order to describe the properties of a system we need to describe the response of the system to external perturbations such as an electromagnetic field. The external perturbation may be time-dependent so what is needed is *time-dependent perturbation theory*. In general, the response of the system can be very complicated. In the present discussion, however, we restrict ourselves to *linear response theory*, i.e. we only consider changes to the wave function of the system that are linear in the perturbation.

The starting point of the discussion is the unperturbed system. The unperturbed system is described by a time-independent Hamiltonian $\hat{H}_0$. We assume that all eigenstates of $\hat{H}_0$ are known, i.e. that we know the solutions of the stationary Schrödinger equation

$$\hat{H}_0\varphi_n = E_n\varphi_n\,, \tag{1.1}$$

where φ_n is the stationary wave function, which depends on all space coordinates but not on time, and E_n is the corresponding energy eigenvalue. In an energy eigenstate, the full time-dependent wave function is given by $\varphi_n \exp(-iE_n t/\hbar)$. We now introduce the perturbation. We assume that the time-dependence is harmonic, i.e. characterized by a single frequency. In fact, the linear response to perturbations with more complicated time-dependence can be constructed from a series of harmonic perturbations using Fourier analysis. The Hamiltonian therefore changes from $\hat{H}_0$ to $\hat{H}_0 + \frac{1}{2}\hat{H}_1 e^{-i\omega t} + \frac{1}{2}\hat{H}_1^\dagger e^{i\omega t}$, where $\hat{H}_1$ contains the spatial part of the perturbation, ω is the frequency and the conjugation in the last term ($\dagger$) is in the operator sense ("Hermitian conjugation"). Typical perturbations are electric or magnetic fields that interact with electrons via their charge or spin.

1.1 Linear Response Theory

The key to describing the response to the perturbation is trying to solve the time-dependent Schrödinger equation

$$i\hbar\frac{\partial\psi}{\partial t} = \left\{\hat{H}_0 + \tfrac{1}{2}\hat{H}_1 e^{-i\omega t} + \tfrac{1}{2}\hat{H}_1^\dagger e^{i\omega t}\right\}\psi\,.$$

This is generally an impossible task but we're helped by the fact that we're only looking for the linear change to the wave function. Our knowledge of the φ_n's can be used if we write the unknown ψ in the following form

$$\psi = \sum_n a_n \varphi_n e^{-iE_n t/\hbar}\,, \tag{1.2}$$

where a_n is an unknown time-dependent coefficient. Any ψ can be written in this form since the φ_n's constitute a complete set. We next insert this expression in the time-dependent Schrödinger equation

$$i\hbar \frac{\partial}{\partial t} \sum_n a_n \varphi_n e^{-iE_n t/\hbar} = \left\{ \hat{H}_0 + \tfrac{1}{2}\hat{H}_1 e^{-i\omega t} + \tfrac{1}{2}\hat{H}_1^\dagger e^{i\omega t} \right\} \sum_n a_n \varphi_n e^{-iE_n t/\hbar}$$

$$\Downarrow$$

$$\sum_n \left\{ a_n E_n \varphi_n + i\hbar \frac{\partial a_n}{\partial t} \varphi_n \right\} e^{-iE_n t/\hbar} = \sum_n a_n \left\{ \hat{H}_0 \varphi_n + \tfrac{1}{2}\hat{H}_1 \varphi_n e^{-i\omega t} + \tfrac{1}{2}\hat{H}_1^\dagger \varphi_n e^{i\omega t} \right\} e^{-iE_n t/\hbar}.$$

Due to the eigenstate condition Eq.(1.1), the first terms on the right-hand and left-hand sides cancel. We consequently find

$$\sum_n \frac{\partial a_n}{\partial t} \varphi_n e^{-iE_n t/\hbar} = \frac{1}{2i\hbar} \sum_n a_n \left\{ \hat{H}_1 e^{-i\omega t} + \hat{H}_1^\dagger e^{i\omega t} \right\} \varphi_n \, e^{-iE_n t/\hbar}. \tag{1.3}$$

Next, we exploit the orthogonality between the eigenstates

$$\langle \varphi_m | \varphi_n \rangle = \delta_{nm} = \begin{cases} 1 & \text{if } n = m \\ 0 & \text{otherwise} \end{cases},$$

where $\langle \varphi_m | \varphi_n \rangle$ is the shorthand Dirac notation for the integral

$$\langle \varphi_m | \varphi_n \rangle = \int \varphi_m^* \varphi_n d^3 r_1 ... d^3 r_N$$

over all the coordinates of the N electrons. If the electron spin is considered, the integration is over the spin variables as well. Hence, we multiply Eq.(1.3) by φ_m^* and integrate to find

$$\sum_n \frac{\partial a_n}{\partial t} \delta_{nm} e^{-iE_n t/\hbar} = \frac{1}{2i\hbar} \sum_n a_n \left\{ \left\langle \varphi_m \middle| \hat{H}_1 \middle| \varphi_n \right\rangle e^{-i\omega t} + \left\langle \varphi_m \middle| \hat{H}_1^\dagger \middle| \varphi_n \right\rangle e^{i\omega t} \right\} e^{-iE_n t/\hbar}$$

$$\Downarrow \tag{1.4}$$

$$\frac{\partial a_m}{\partial t} = \frac{1}{2i\hbar} \sum_n a_n \left\{ \left\langle \varphi_m \middle| \hat{H}_1 \middle| \varphi_n \right\rangle e^{-i\omega t} + \left\langle \varphi_m \middle| \hat{H}_1^\dagger \middle| \varphi_n \right\rangle e^{i\omega t} \right\} e^{iE_{mn} t/\hbar},$$

where $E_{mn} = E_m - E_n$. Again, the Dirac notation is used for matrix elements such as

$$\left\langle \varphi_m \middle| \hat{H}_1 \middle| \varphi_n \right\rangle = \int \varphi_m^* \hat{H}_1 \varphi_n d^3 r_1 ... d^3 r_N \,.$$

To proceed, we now need the important expansion of the unknown coefficients a_n in powers of the perturbation. For the electromagnetic perturbations considered in the present discussion, the perturbing Hamiltonian is proportional to the field strength, i.e. $\hat{H}_1 \propto \mathcal{E}$ and $\hat{H}_1 \propto \mathcal{B}$ for electric and magnetic perturbations, respectively. Hence, any coefficient a_n may be considered a function of the field strength, e.g. $a_n = a_n(\mathcal{E})$ in the electric case. We may consequently make a Taylor expansion in field strength, i.e.

$$a_n = \underbrace{a_n^{(0)}}_{0\text{'th order}} + \underbrace{a_n^{(1)}}_{1\text{'st order}} + \ldots,$$

where the superscript indicates the power of the perturbation. We now utilize a familiar theorem from polynomial series:

$$\text{If, for all } x, \sum_p b_p x^p = \sum_p c_p x^p, \text{ then } b_p = c_p .$$

Used in Eq.(1.4) this implies that the powers of the perturbation on both sides must be equal, i.e.

$$\frac{\partial a_m^{(p)}}{\partial t} = \frac{1}{2i\hbar}\sum_n a_n^{(p-1)} \left\{ \left\langle \varphi_m \middle| \hat{H}_1 \middle| \varphi_n \right\rangle e^{-i\omega t} + \left\langle \varphi_m \middle| \hat{H}_1^\dagger \middle| \varphi_n \right\rangle e^{i\omega t} \right\} e^{iE_{mn}t/\hbar}, \qquad (1.5)$$

since the matrix elements $\left\langle \varphi_m \middle| \hat{H}_1 \middle| \varphi_n \right\rangle$ and $\left\langle \varphi_m \middle| \hat{H}_1^\dagger \middle| \varphi_n \right\rangle$ on the right-hand side already contain one power of the perturbation. The technique now consists in starting by setting $p = 0$ in this expression. Subsequently, we set $p = 1$ and so on. In fact, for the linear response we need not go beyond $p = 1$. The so-called nonlinear response can be found by repeating this exercise to higher orders [1]. Setting $p = 0$, we find

$$\frac{\partial a_m^{(0)}}{\partial t} = 0,$$

i.e. the zero'th order coefficients are constant. This is quite obviously correct since zero'th order means that the perturbation is ignored altogether. Having established this simple result, we can immediately proceed to $p = 1$ for which Eq.(1.5) yields

$$\frac{\partial a_m^{(1)}}{\partial t} = \frac{1}{2i\hbar}\sum_n a_n^{(0)} \left\{ \left\langle \varphi_m \middle| \hat{H}_1 \middle| \varphi_n \right\rangle e^{-i\omega t} + \left\langle \varphi_m \middle| \hat{H}_1^\dagger \middle| \varphi_n \right\rangle e^{i\omega t} \right\} e^{iE_{mn}t/\hbar}.$$

We integrate this expression to find

$$
\begin{aligned}
a_m^{(1)} &= \frac{1}{2i\hbar}\sum_n a_n^{(0)} \int \left\{ \left\langle \varphi_m \middle| \hat{H}_1 \middle| \varphi_n \right\rangle e^{-i\omega t} e^{iE_{mn}t/\hbar} + \left\langle \varphi_m \middle| \hat{H}_1^\dagger \middle| \varphi_n \right\rangle e^{i\omega t} e^{iE_{mn}t/\hbar} \right\} dt \\
&= -\frac{1}{2}\sum_n a_n^{(0)} \left\{ \left\langle \varphi_m \middle| \hat{H}_1 \middle| \varphi_n \right\rangle \frac{e^{-i\omega t} e^{iE_{mn}t/\hbar}}{E_{mn} - \hbar\omega} + \left\langle \varphi_m \middle| \hat{H}_1^\dagger \middle| \varphi_n \right\rangle \frac{e^{i\omega t} e^{iE_{mn}t/\hbar}}{E_{mn} + \hbar\omega} \right\}.
\end{aligned}
\tag{1.6}
$$

The lower limit of the integral does not contribute if we assume that the perturbation was turned off in the infinite past. Physically, the result Eq.(1.6) is an idealization because we have ignored losses that tend to de-excite the system, i.e. make the system decay back to the ground state. To incorporate these effects, we may introduce a certain damping $\hbar\Gamma$ in the expression above

$$
a_m^{(1)} = -\frac{1}{2}\sum_n a_n^{(0)} \left\{ \frac{\left\langle \varphi_m \middle| \hat{H}_1 \middle| \varphi_n \right\rangle e^{-i\omega t} e^{iE_{mn}t/\hbar}}{E_{mn} - \hbar\omega - i\hbar\Gamma} + \frac{\left\langle \varphi_m \middle| \hat{H}_1^\dagger \middle| \varphi_n \right\rangle e^{i\omega t} e^{iE_{mn}t/\hbar}}{E_{mn} + \hbar\omega - i\hbar\Gamma} \right\}. \tag{1.7}
$$

We are now in a position to calculate the expectation value of any time-independent operator corresponding to some measurable quantity. We denote this operator (response observable) by $\hat{X}$ so that what we want is $\left\langle \psi \middle| \hat{X} \middle| \psi \right\rangle$. We therefore use Eq.(1.2) and keep only terms up to linear order in the perturbation

$$
\left\langle \psi \middle| \hat{X} \middle| \psi \right\rangle \approx \sum_{m,n} \left\{ a_n^{(0)*} a_m^{(0)} + a_n^{(0)*} a_m^{(1)} + a_n^{(1)*} a_m^{(0)} \right\} \left\langle \varphi_n \middle| \hat{X} \middle| \varphi_m \right\rangle e^{iE_{nm}t/\hbar}.
$$

To interpret this result, we see what normalization of the total wave function tells us. Using Eq.(1.2) we see that

$$
1 = \left\langle \psi \middle| \psi \right\rangle = \sum_n \left| a_n \right|^2,
$$

using the orthogonality of the φ_n's. Thus, if the perturbation is absent and $a_n = a_n^{(0)}$ we have

$$
\sum_n \left| a_n^{(0)} \right|^2 = 1 \ \text{(no perturbation)}.
$$

As usual, we interpret $\left| a_n^{(0)} \right|^2$ as the probability that the unperturbed system is in the state φ_n. In thermal equilibrium, this means that $\left| a_n^{(0)} \right|^2 = f(E_n)$, where f is the probability distribution, which depends on the energy of the state only. We use this property to postulate the following:

$$a_n^{(0)*}a_m^{(0)} = \begin{cases} f(E_n) & \text{if } n=m \\ 0 & \text{otherwise} \end{cases}.$$

This postulate can be made more rigorous using the so-called density matrix formulation of perturbation theory [1]. We will simply accept it as a reasonable postulate here. We subsequently apply Eq.(1.7) to show that

$$\begin{aligned}\left\langle\psi\middle|\hat{X}\middle|\psi\right\rangle &\approx \sum_n f(E_n)\left\langle\varphi_n\middle|\hat{X}\middle|\varphi_n\right\rangle \\ -\frac{1}{2}\sum_{m,n} f(E_n)\left\langle\varphi_n\middle|\hat{X}\middle|\varphi_m\right\rangle &\left\{\frac{\left\langle\varphi_m\middle|\hat{H}_1\middle|\varphi_n\right\rangle e^{-i\omega t}}{E_{mn}-\hbar\omega-i\hbar\Gamma}+\frac{\left\langle\varphi_m\middle|\hat{H}_1^\dagger\middle|\varphi_n\right\rangle e^{i\omega t}}{E_{mn}+\hbar\omega-i\hbar\Gamma}\right\} \\ -\frac{1}{2}\sum_{m,n} f(E_m)\left\langle\varphi_n\middle|\hat{X}\middle|\varphi_m\right\rangle &\left\{\frac{\left\langle\varphi_m\middle|\hat{H}_1^\dagger\middle|\varphi_n\right\rangle e^{i\omega t}}{E_{nm}-\hbar\omega+i\hbar\Gamma}+\frac{\left\langle\varphi_m\middle|\hat{H}_1\middle|\varphi_n\right\rangle e^{-i\omega t}}{E_{nm}+\hbar\omega+i\hbar\Gamma}\right\}.\end{aligned}$$

If this result is Fourier decomposed into frequency components it is found that

$$\left\langle\psi\middle|\hat{X}\middle|\psi\right\rangle = \sum_n f(E_n)\left\langle\varphi_n\middle|\hat{X}\middle|\varphi_n\right\rangle + \frac{1}{2}X(\omega)e^{-i\omega t} + \frac{1}{2}X^*(\omega)e^{i\omega t}. \tag{1.8}$$

Here, the time-independent first term is the permanent contribution, which exists in the absence of the perturbation, while comparison with the previous result shows that

$$X(\omega) = -\sum_{m,n} f_{nm}\frac{\left\langle\varphi_m\middle|\hat{H}_1\middle|\varphi_n\right\rangle\left\langle\varphi_n\middle|\hat{X}\middle|\varphi_m\right\rangle}{E_{mn}-\hbar\omega-i\hbar\Gamma}, \tag{1.9}$$

where $f_{nm} \equiv f(E_n) - f(E_m)$. This is the time-dependent induced response due to the perturbation. It is noted that the form Eq.(1.8) ensures that the response is real-valued. At zero temperature, the ground state is known with certainty to be occupied while all other states are empty. Hence, the only non-vanishing f_{nm} terms are $f_{0m} = 1$ and $f_{n0} = -1$, where n and m are both larger than 0. For this reason we find

$$X(\omega) = -\sum_{n>0}\left\{\frac{\left\langle\varphi_n\middle|\hat{H}_1\middle|\varphi_0\right\rangle\left\langle\varphi_0\middle|\hat{X}\middle|\varphi_n\right\rangle}{E_{n0}-\hbar\omega-i\hbar\Gamma}+\frac{\left\langle\varphi_0\middle|\hat{H}_1\middle|\varphi_n\right\rangle\left\langle\varphi_n\middle|\hat{X}\middle|\varphi_0\right\rangle}{E_{n0}+\hbar\omega+i\hbar\Gamma}\right\} \quad (T=0) \tag{1.10}$$

at zero temperature. The expressions in Eqs.(1.9) and (1.10) are the fundamental results of linear response theory. They constitute the basis upon which all subsequent results are built.

Exercise: High-frequency limit

In this exercise, we will investigate the response of a nanostructure subjected to a perturbation with a very high frequency. If $\hbar\omega \gg \hbar\Gamma$ and we change the order of certain terms, we may, first of all, write Eq.(1.10) as

$$X(\omega) \approx \sum_{n>0} \left\{ \frac{\left\langle \varphi_0 \left| \hat{X} \right| \varphi_n \right\rangle \left\langle \varphi_n \left| \hat{H}_1 \right| \varphi_0 \right\rangle}{\hbar\omega - E_{n0}} - \frac{\left\langle \varphi_0 \left| \hat{H}_1 \right| \varphi_n \right\rangle \left\langle \varphi_n \left| \hat{X} \right| \varphi_0 \right\rangle}{\hbar\omega + E_{n0}} \right\}.$$

Secondly, we will assume that $\hbar\omega \gg E_{n0}$ for all the important transitions in the sum.

a) Show that using a geometric series

$$\frac{1}{\hbar\omega \pm E_{n0}} = \frac{1}{\hbar\omega} \mp \frac{E_{n0}}{(\hbar\omega)^2} + \ldots$$

We also need the completeness property of the states, which means that $\sum_n |\varphi_n\rangle\langle\varphi_n| = 1$. If the ground state n = 0 is excluded from the summation it follows that $\sum_{n>0} |\varphi_n\rangle\langle\varphi_n| = 1 - |\varphi_0\rangle\langle\varphi_0|$.

b) If only the first term of the geometric series is retained, show that

$$X(\omega) \approx \frac{1}{\hbar\omega} \left\langle \varphi_0 \left| \left[\hat{X}, \hat{H}_1 \right] \right| \varphi_0 \right\rangle,$$

where $\left[\hat{X}, \hat{H}_1 \right]$ is the commutator between the two operators.

References

[1] R.W. Boyd *Nonlinear Optics* (Academic Press, San Diego, 1992).

2. Preliminaries

Before we start seriously applying the results of perturbation theory to real nanostructures or bulk materials it is useful to consider some general features of perturbations and responses. The most common perturbations and, also, the ones treated in this work are:

- Static or low-frequency electric fields
- Static or low-frequency magnetic fields
- High-frequency electric fields (light)

In addition, two or more of these perturbations may be present simultaneously. Examples of perturbations that are not considered here are mechanical ones such as pressure or mechanical stress. The most important combinations of perturbation and response for solids are shown in the table below.

Perturbation→

Response→	Electric field	Magnetic field	Electric + magnetic field
Electric current	Electric conductivity		Hall effect
Polarization	Electric susceptibility		Faraday effect
Spin magnetization		Spin-magnetic susceptibility	
Orbital magnetization		Orbital-magnetic susceptibility	

The electric and magnetic perturbations are intimately related to the electric and magnetic dipole moments. Classically, the changes in energy associated with an electric field $\vec{\mathcal{E}}$ or magnetic field $\vec{\mathcal{B}}$ are $-\vec{\mu}_e \cdot \vec{\mathcal{E}}$ and $-\vec{\mu}_m \cdot \vec{\mathcal{B}}$, where $\vec{\mu}_e$ and $\vec{\mu}_m$ are electric and magnetic dipole moments, respectively. For a quantum mechanical system of electrons, the electric dipole operator is

$$\vec{\mu}_e = -e\vec{r}, \quad \vec{r} = \sum_i \vec{r}_i \,.$$

Here, the sum is over all the electrons. The magnetic interaction is a little more complicated because both orbital and spin angular momentum contributes. Hence, $\vec{\mu}_m = \vec{\mu}_{orb} + \vec{\mu}_{spin}$ with

$$\vec{\mu}_{orb} = -\mu_B \vec{l}, \quad \vec{\mu}_{spin} = -2\mu_B \vec{s}, \quad \vec{l} = \sum_i \vec{l}_i, \quad \vec{s} = \sum_i \vec{s}_i,$$

where $\mu_B = \hbar e / 2m$ is the Bohr magneton. Here, $\vec{l}$ and $\vec{s}$ are the total angular and spin angular momentum, respectively. The factor 2 in the expression for $\vec{\mu}_{spin}$ is actually the g-factor of the electron and should be replaced by $g_e \approx 2.0023$ if quantum-electrodynamic effects are included.

The response observables, in turn, are also related to the dipole moments. Thus, polarization $\vec{\mathcal{P}}(\omega)$ is nothing more that the induced electric dipole moment per volume and magnetization $\vec{\mathcal{M}}(\omega)$ is the induced magnetic dipole moment per volume. It follows that the associated observables are

$$\begin{aligned}
\text{Polarization } \vec{\mathcal{P}}(\omega): &\quad \hat{\vec{X}} = \vec{\mu}_e / \Omega = -e\vec{r} / \Omega \\
\text{Orbital magnetization } \vec{\mathcal{M}}_{orb}(\omega): &\quad \hat{\vec{X}} = -\mu_B \vec{l} / \Omega \\
\text{Spin magnetization } \vec{\mathcal{M}}_{spin}(\omega): &\quad \hat{\vec{X}} = -2\mu_B \vec{s} / \Omega,
\end{aligned}$$

where Ω is the crystal volume. The only additions needed to this list are those related to electric currents. In this case, the perturbation is the action of the electrostatic potential $-e\sum_i V(\vec{r}_i)$ and the observable is the current density operator is $\hat{\vec{J}} = -e\hat{\vec{p}} / \Omega m$, where $\hat{\vec{p}} = \sum_i \hat{\vec{p}}_i$ is the total momentum operator. As we shall see, however, this case is also described by the electric case described above.

The actual calculation of response functions always relies on the computation of matrix elements such as $\langle \varphi_n | \hat{o} | \varphi_m \rangle$, where $\hat{o}$ could be any of the Hermetian operators encountered above. Many apparently different matrix elements are closely related as can be proved by a clever trick that goes as follows: We start from the matrix element of the commutator between the Hamiltonian $\hat{H}_0$ and the operator $\hat{o}$

$$\begin{aligned}
\left\langle \varphi_m \left| \left[\hat{H}_0, \hat{o} \right] \right| \varphi_n \right\rangle &= \left\langle \varphi_m \left| \hat{H}_0 \hat{o} - \hat{o} \hat{H}_0 \right| \varphi_n \right\rangle \\
&= \left\langle \varphi_m \left| \hat{H}_0 \hat{o} \right| \varphi_n \right\rangle - \left\langle \varphi_m \left| \hat{o} \hat{H}_0 \right| \varphi_n \right\rangle \\
&= \left\langle \varphi_n \left| \hat{o} \hat{H}_0 \right| \varphi_m \right\rangle^* - \left\langle \varphi_m \left| \hat{o} \hat{H}_0 \right| \varphi_n \right\rangle,
\end{aligned}$$

where the last equality follows from the fact that both $\hat{H}_0$ and $\hat{o}$ are Hermitian. We then utilize the action of $\hat{H}_0$ on the eigenstates

$$\begin{aligned}\left\langle \varphi_m \left| \left[\hat{H}_0, \hat{o}\right] \right| \varphi_n \right\rangle &= E_m \left\langle \varphi_n \left| \hat{o} \right| \varphi_m \right\rangle^* - E_n \left\langle \varphi_m \left| \hat{o} \right| \varphi_n \right\rangle \\ &= E_m \left\langle \varphi_m \left| \hat{o} \right| \varphi_n \right\rangle - E_n \left\langle \varphi_m \left| \hat{o} \right| \varphi_n \right\rangle \\ &= E_{mn} \left\langle \varphi_m \left| \hat{o} \right| \varphi_n \right\rangle .\end{aligned} \tag{2.1}$$

On the other hand, many commutators can be calculated explicitly. We may consider just a single electron and for the z component of the position we find

$$\begin{aligned}\left[\hat{H}_0, z\right]\varphi &= -\frac{\hbar^2}{2m}\left[\frac{\partial^2}{\partial z^2}, z\right]\varphi \\ &= -\frac{\hbar^2}{2m}\left(\varphi\frac{\partial^2 z}{\partial z^2} + 2\frac{\partial z}{\partial z}\frac{\partial}{\partial z}\varphi + z\frac{\partial^2}{\partial z^2}\varphi - z\frac{\partial^2}{\partial z^2}\varphi\right) \\ &= -\frac{\hbar^2}{m}\frac{\partial}{\partial z}\varphi \\ &= \frac{\hbar}{im}\hat{p}_z\varphi .\end{aligned}$$

Thus, in fact $\left[\hat{H}_0, z\right] = (\hbar / im)\hat{p}_z$. Similarly, for the electrostatic potential

$$\begin{aligned}\left[\hat{H}_0, V\right]\varphi &= -\frac{\hbar^2}{2m}\left[\nabla^2, V\right]\varphi \\ &= -\frac{\hbar^2}{2m}\left(\varphi\nabla^2 V + 2\nabla V\cdot\nabla\varphi + V\nabla^2\varphi - V\nabla^2\varphi\right) \\ &= -\frac{\hbar^2}{2m}\left(\varphi\nabla^2 V + 2\nabla V\cdot\nabla\varphi\right).\end{aligned}$$

Now, $\vec{\mathcal{E}} = -\nabla V$ and if the electric field is assumed constant in space then $\nabla^2 V = 0$. Hence,

$$\left[\hat{H}_0, V\right]\varphi = \frac{\hbar^2}{m}\vec{\mathcal{E}}\cdot\nabla\varphi = \frac{i\hbar}{m}\vec{\mathcal{E}}\cdot\hat{\vec{p}}\varphi .$$

As a consequence, only momentum matrix elements are really needed in both cases. The results can be summarized as follows

$$\begin{aligned}\frac{\hbar}{im}\left\langle \varphi_m \left| \hat{p}_z \right| \varphi_n \right\rangle &= E_{mn}\left\langle \varphi_m \left| z \right| \varphi_n \right\rangle , \\ \frac{i\hbar}{m}\left\langle \varphi_m \left| \vec{\mathcal{E}}\cdot\hat{\vec{p}} \right| \varphi_n \right\rangle &= E_{mn}\left\langle \varphi_m \left| V \right| \varphi_n \right\rangle .\end{aligned} \tag{2.2}$$

2.1 A First Example: Electric Polarizability

We will now illustrate some of the concepts above by working our way through a simple example: the electric polarizability of a spherical quantum well with infinite barrier confinement. This model and the first few eigenstates are illustrated in the figure below. In this situation, the perturbation is the electric field $\vec{\mathcal{E}}$ and the response is the induced dipole moment $\vec{d} = -e\langle\psi|\vec{r}|\psi\rangle$. For simplicity, we assume that only a single electron is in the system. Also, an isotropic system is assumed, which means that an electric field along the z-axis $\vec{\mathcal{E}} = \vec{e}_z\mathcal{E}$ induces a dipole moment in the same direction $\vec{d} = \vec{e}_z d$. From the general framework presented in the previous chapter it then follows that

$$\begin{aligned} d(\omega) &= -\sum_{m,n} f_{nm} \frac{\langle\varphi_m|e\mathcal{E}z|\varphi_n\rangle\langle\varphi_n|-ez|\varphi_m\rangle}{E_{mn} - \hbar\omega - i\hbar\Gamma} \\ &= e^2\mathcal{E}\sum_{m,n} f_{nm} \frac{\left|\langle\varphi_m|z|\varphi_n\rangle\right|^2}{E_{mn} - \hbar\omega - i\hbar\Gamma}. \end{aligned}$$

The constant of proportionality between d and $\mathcal{E}$ is the polarizability $\alpha(\omega)$ given by

$$\alpha(\omega) = e^2\sum_{m,n} f_{nm} \frac{\left|\langle\varphi_m|z|\varphi_n\rangle\right|^2}{E_{mn} - \hbar\omega - i\hbar\Gamma}.$$

At low temperature, we can similarly to Eq.(1.10) utilize the fact that the only non-zero f_{nm} terms are $f_{0m} = 1$ and $f_{n0} = -1$, where n and m are both larger than 0 and after re-labeling

$$\begin{aligned} \alpha(\omega) &= e^2\sum_{n>0}\left\{\frac{\left|\langle\varphi_n|z|\varphi_0\rangle\right|^2}{E_{n0} - \hbar\omega - i\hbar\Gamma} - \frac{\left|\langle\varphi_0|z|\varphi_n\rangle\right|^2}{E_{0n} - \hbar\omega - i\hbar\Gamma}\right\} \\ &= 2e^2\sum_{n>0}\frac{\left|\langle\varphi_n|z|\varphi_0\rangle\right|^2 E_{n0}}{E_{n0}^2 - \hbar^2(\omega + i\Gamma)^2}. \end{aligned}$$

This expression obviously leads to resonances whenever $\hbar\omega = E_{n0}$, i.e. whenever the photon energy matches the energy difference between the ground state and an excited state. The dimensionless quantity

$$g_{n0} = \frac{2m}{\hbar^2}\left|\langle\varphi_n|z|\varphi_0\rangle\right|^2 E_{n0}$$

is known as the oscillator strength of the $|\varphi_0\rangle \rightarrow |\varphi_n\rangle$ transition. Thus,

$$\alpha(\omega) = \frac{\hbar^2 e^2}{m} \sum_n \frac{g_{n0}}{E_{n0}^2 - \hbar^2(\omega + i\Gamma)^2}.$$

Note that we removed the restriction on n in the sum because $g_{00} = 0$. We now wish to investigate the high and low frequency limits. In the former case, we may assume that $\hbar\omega \gg E_{n0}$ for all important transitions and so

$$\alpha(\omega) \approx -\frac{e^2}{m\omega^2} \sum_n g_{n0}.$$

To evaluate the sum over oscillator strengths we utilize Eq.(2.2)

$$\begin{aligned}
\sum_n g_{n0} &= \frac{2m}{\hbar^2} \sum_n \left|\langle \varphi_0 | z | \varphi_n \rangle\right|^2 E_{n0} \\
&= \frac{m}{\hbar^2} \sum_n \left\{ \langle \varphi_0 | z | \varphi_n \rangle \langle \varphi_n | z | \varphi_0 \rangle E_{n0} - E_{0n} \langle \varphi_0 | z | \varphi_n \rangle \langle \varphi_n | z | \varphi_0 \rangle \right\} \\
&= \frac{1}{i\hbar} \sum_n \left\{ \langle \varphi_0 | z | \varphi_n \rangle \langle \varphi_n | \hat{p}_z | \varphi_0 \rangle - \langle \varphi_0 | \hat{p}_z | \varphi_n \rangle \langle \varphi_n | z | \varphi_0 \rangle \right\}.
\end{aligned}$$

Now, applying the completeness of the states $\sum_n |\varphi_n\rangle\langle\varphi_n| = 1$ and the commutator relation $[\hat{p}_z, z] = N\hbar / i$ (for N-electron operators), we see that

$$\sum_n g_{n0} = \frac{i}{\hbar} \langle \varphi_0 | [\hat{p}_z, z] | \varphi_n \rangle = N.$$

This general result is known as the Thomas-Reiche-Kuhn sum rule. In the present example, $N = 1$ and $g_{00} = 0$ and so

$$\alpha(\omega) \approx -\frac{e^2}{m\omega^2} \quad \text{(high frequency limit).}$$

Conversely, in the low frequency limit

$$\alpha(\omega) \approx 2e^2 \sum_{n>0} \frac{\left|\langle \varphi_n | z | \varphi_0 \rangle\right|^2}{E_{n0}} \quad \text{(low frequency limit).}$$

In any spherically symmetric system (an atom or spherical nanoparticle), the eigenstates are of the form $\varphi_{nlm}(\vec{r}) = Y_{lm}(\theta, \phi) R_{nl}(r)$ with the angular dependence in

the form of a spherical harmonic. The ground state is $\varphi_{100}(\vec{r}) = Y_{00}(\theta,\phi)R_{10}(r)$. Moreover, z equals $r\cos\theta$ in polar coordinates that, in turn, can be written as $rY_{10}(\theta,\phi)\sqrt{4\pi/3}$. Using $Y_{00}(\theta,\phi) = 1/\sqrt{4\pi}$ we therefore readily see that

$$\langle \varphi_{nlm} | z | \varphi_{100} \rangle = \frac{1}{\sqrt{3}} \int_0^\infty R_{n1}(r) R_{10}(r) r^3 dr \, \delta_{l1}\delta_{m0}.$$

The Kronecker deltas follow from the orthogonality of the spherical harmonics $\langle Y_{lm} | Y_{10} \rangle = \delta_{l1}\delta_{m0}$. Thus, the ground state only couples to *p*-type excited states. For the spherical quantum well with infinite barriers and radius a (see Appendix 1)

$$R_{10}(r) = \frac{\sqrt{2}}{r\sqrt{a}} \sin\frac{\pi r}{a}, \quad R_{n1}(r) = \frac{\sqrt{2a}}{r^2 \beta_n |\sin\beta_n|} \left\{ \sin\frac{\beta_n r}{a} - \frac{\beta_n r}{a} \cos\frac{\beta_n r}{a} \right\}.$$

Here, β_n is the n'th root of the 1'st spherical Bessel function, i.e. $\sin\beta_n - \beta_n \cos\beta_n = 0$. Numerically, the first few are $\beta_n \in \{4.49341, 7.72525, 10.9041, ...\}$. The ground state and the first two *p*-type excited states are shown in Fig. 2.1.

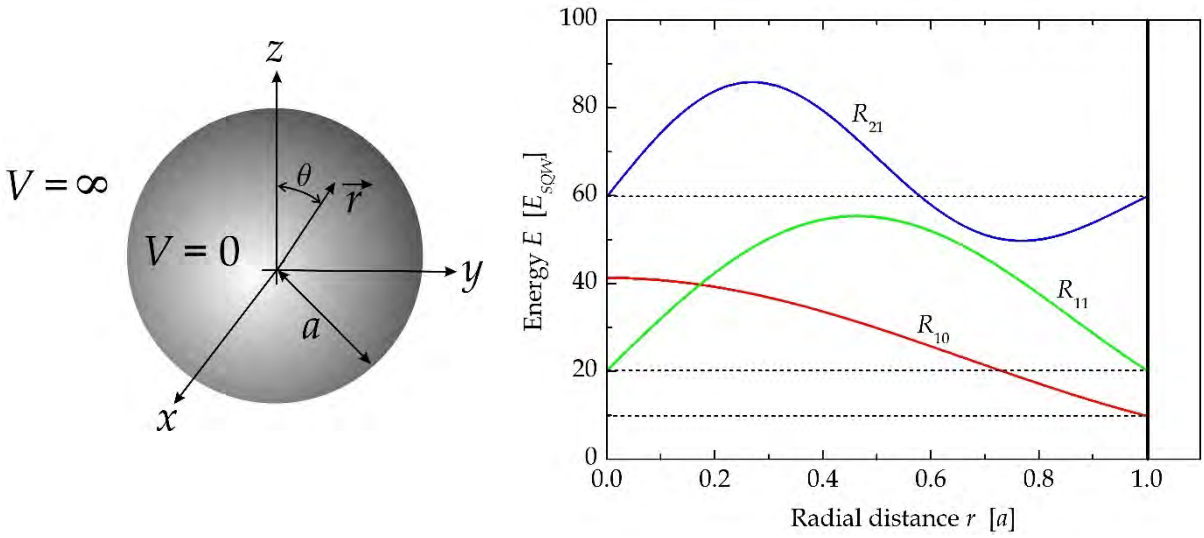


Figure 2.1. Model of a spherical quantum well with infinite barriers (left) and some of the first eigenstates displaced vertically by the energy (right).

From these states, the dipole matrix elements are easily calculated

$$\frac{1}{\sqrt{3}} \int_0^a R_{n1}(r) R_{10}(r) r^3 dr = -\frac{4\pi a \beta_n}{\sqrt{3}\left(\beta_n^2 - \pi^2\right)^2}.$$

Also, the energies are $E_{100} = \hbar^2\pi^2/2ma^2$ and $E_{n10} = \hbar^2\beta_n^2/2ma^2$. We consequently find

$$g_{n0} = \frac{16\pi^2\beta_n^2}{3\left(\beta_n^2 - \pi^2\right)^3}.$$

Summed over the first 50 states, we find $\sum_n g_{n0} = 0.9999986$, which illustrates the Thomas-Reiche-Kuhn sum rule. Similarly, the static polarizability becomes

$$\alpha(0) = \frac{e^2ma^4}{\hbar^2}\sum_n \frac{64\pi^2\beta_n^2}{3\left(\beta_n^2 - \pi^2\right)^5} \approx 0.0363\frac{e^2ma^4}{\hbar^2}.$$

More advanced versions of perturbation theory (see Appendix 11) can be used to show that the numerical factor is really $(3+4\pi^2)/(12\pi^4)$. In terms of the characteristic energy of the spherical quantum well $E_{SQW} = \hbar^2/2ma^2$, the resonances $\hbar\omega_n = E_{n10} - E_{100}$ are found at $\hbar\omega_n / E_{SQW} = \beta_n^2 - \pi^2 \in \{10.32, 49.81, 109.03, ...\}$. In the plot below illustrating the general result, these resonances are clearly noticeable.

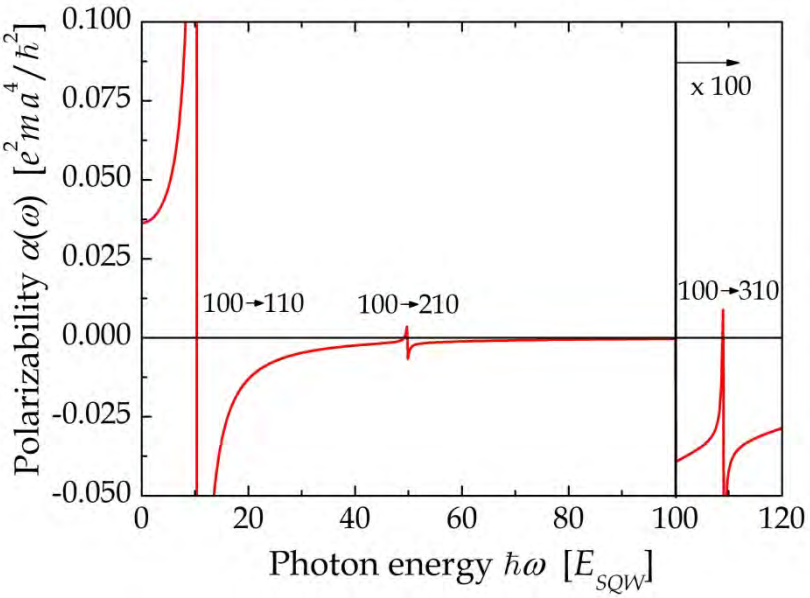


Figure 2.2. The frequency dependent polarizability of a spherical quantum well. Three resonances are found in the depicted energy range.

Exercise: One-dimensional quantum well

In a simple quantum well stretching between $z = \pm L/2$ with infinite barriers, the eigenstates are of the form (see Appendix 1)

$$\varphi_n(z) = \sqrt{\frac{2}{L}} \cdot \begin{cases} \cos(nz\pi/L) & n \text{ odd} \\ \sin(nz\pi/L) & n \text{ even} \end{cases},$$

and the energies are $E_n = \hbar^2 n^2 \pi^2 / 2mL^2$.

a) Verify that

$$\langle \varphi_1 | z | \varphi_1 \rangle = 0, \quad \langle \varphi_1 | z | \varphi_2 \rangle = \frac{16L}{9\pi^2}.$$

The general result is

$$\langle \varphi_1 | z | \varphi_{2n+1} \rangle = 0, \quad \langle \varphi_1 | z | \varphi_{2n} \rangle = \frac{16nL(-1)^{n+1}}{\pi^2(4n^2-1)^2}.$$

b) Show that, accordingly, the oscillator strength is

$$g_{2n,1} = \frac{256n^2}{\pi^2(4n^2-1)^3}$$

c) By direct calculation, show that $g_{2,1} \approx 0.961$ and $g_{4,1} \approx 0.031$. This shows that the sum rule for oscillator strengths is very nearly satisfied with only a few terms.

d) Use the above results to demonstrate that

$$\alpha(0) = \frac{1024e^2mL^4}{\pi^6\hbar^2}\sum_{n=1}^{\infty}\frac{n^2}{(4n^2-1)^5}$$

The sum may actually be evaluated exactly with the result $\sum_{n=1}^{\infty}\frac{n^2}{(4n^2-1)^5} = \frac{\pi^2(15-\pi^2)}{12288}$. The net polarizability is therefore

$$\alpha(0) = \frac{(15-\pi^2)e^2mL^4}{12\pi^4\hbar^2}.$$

3. Bulk Response of Metals

In the present chapter, we restrict the attention to metals. Metals are generally divided into noble metals, transition metals and alkali metals. Technologically, the first two groups are the most important ones. The noble metals include copper (Cu) and gold (Au), that are of great importance for electronics and conductors. Transition metals count materials such as iron (Fe) and titanium (Ti) of importance for electric motors, steel constructions and high strength alloys. A notable addition to these main classes is aluminum (Al), which is a so-called free-electron metal, and also of immense importance in many different areas. The band structure of copper is shown in Fig. 3.1. This picture is characteristic of noble metals in that it features some relatively flat *d*-bands well below the Fermi level and a single *sp*-band crossing the Fermi level.

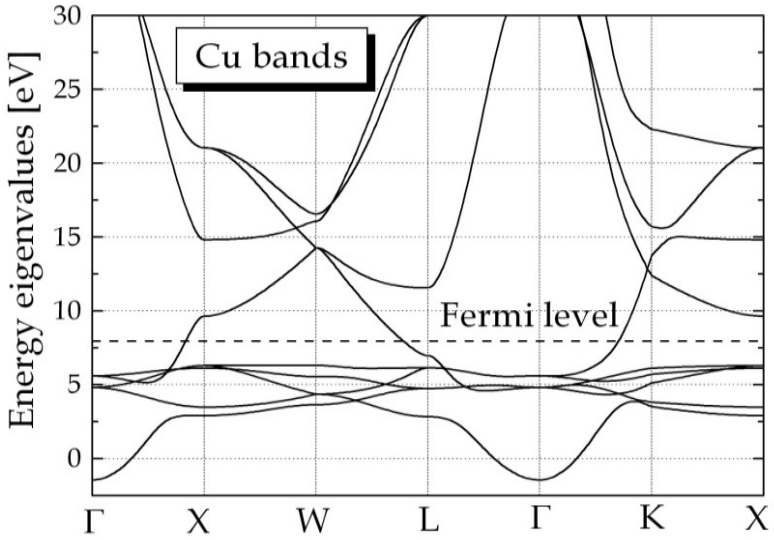


Figure 3.1. Band structure of copper with the Fermi level indicated by the dashed line.

An external perturbation connects an occupied to an empty level. Thus, if the typical energy of the perturbation is rather small, only states in the vicinity of the Fermi level are of importance. Examples of this are static electric and magnetic fields of modest field strength as well as low-frequency (infrared) electromagnetic fields. For these cases, hence, we only have to consider the single *sp*-band crossing the Fermi level in order to describe the response.

3.1 Spin Magnetization

The simplest example of perturbation theory in bulk metals is the spin magnetization. In this case, we consider as the perturbation the coupling between a magnetic field $\mathcal{B}$ and the magnetic moment $-2\mu_B\hat{s}_z$ of an electron

$$\hat{H}_1 = 2\mu_B \mathcal{B}\,\hat{s}_z\,,$$

where $\mathcal{B}$ is the magnetic field strength, and $\hat{s}_z$ is the z-component of the electron spin. The response observable we're looking for is $\hat{X} = -2\mu_B \hat{s}_z / \Omega$, where Ω is the crystal volume. This operator describes the average magnetic moment per volume. The response itself is the spin magnetization for which we use the symbol $\mathcal{M}(\omega)$. Before actually calculating the response we need to consider the following point: The perturbation and response introduced in Chapter 1 were for the entire system containing many particles. On the other hand, the present expressions for $\hat{H}_1$ and $\hat{X}$ are for a single electron only. It can be shown, however, that if all the electrons are independent then the response expression

$$X(\omega) = -\sum_{m,n} f_{nm} \frac{\left\langle \varphi_m \middle| \hat{H}_1 \middle| \varphi_n \right\rangle \left\langle \varphi_n \middle| \hat{X} \middle| \varphi_m \right\rangle}{E_{mn} - \hbar\omega - i\hbar\Gamma}$$

in fact still holds but with a new meaning: Now that $\hat{H}_1$ and $\hat{X}$ are single-particle operators, φ_n and E_n should be taken as single-particle eigenfunctions and eigenvalues. Also, several single-particle eigenstates rather than just the many-particle ground state are now occupied (below the Fermi level).

Since this is the first application of linear response theory, we keep things simple and restrict ourselves to the special case of zero frequency and zero damping, i.e. $\omega = \Gamma = 0$. Thus, we wish to calculate the static response

$$X(0) = -\sum_{m,n} \frac{f_{nm}}{E_{mn}} \left\langle \varphi_m \middle| \hat{H}_1 \middle| \varphi_n \right\rangle \left\langle \varphi_n \middle| \hat{X} \middle| \varphi_m \right\rangle .$$

As mentioned above, the perturbation energy is very small for a static magnetic field and, hence, we may approximate

$$\frac{f_{nm}}{E_{mn}} = \frac{f(E_n) - f(E_n + E_{mn})}{E_{mn}} \approx -f'(E_n), \quad f'(E) = \frac{\partial f(E)}{\partial E}. \tag{3.1}$$

so that

$$X(0) = \sum_{m,n} f'(E_n) \left\langle \varphi_m \middle| \hat{H}_1 \middle| \varphi_n \right\rangle \left\langle \varphi_n \middle| \hat{X} \middle| \varphi_m \right\rangle . \tag{3.2}$$

The present example is simplified further by the fact that both $\hat{H}_1$ and $\hat{X}$ only operate on the spin part of the wave function. For an infinite, periodic solid the eigenstates can be labeled by three quantum numbers: The band index *c*, the Bloch wave vector $\vec{k}$ and the spin σ. Hence, the state index n above is replaced by $\{c\vec{k}\sigma\}$. The eigenstates themselves are Bloch states of the form

$$\left|c\vec{k}\sigma\right\rangle = \varphi_{c\vec{k}}(\vec{r})\left|\sigma\right\rangle, \quad \varphi_{c\vec{k}}(\vec{r}) = u_{c\vec{k}}(\vec{r})e^{i\vec{k}\cdot\vec{r}}, \tag{3.3}$$

where $u_{c\vec{k}}$ is the lattice-periodic part and $\left|\sigma\right\rangle$ contains the spin part. We consider only a single band and so we begin by evaluating the $\hat{H}_1$ matrix element as follows

$$\begin{aligned}\left\langle c\vec{k}'\sigma'\left|\hat{H}_1\right|c\vec{k}\sigma\right\rangle &= 2\mu_B\mathcal{B}\int\varphi^*_{c\vec{k}'}(\vec{r})\varphi_{c\vec{k}}(\vec{r})d^3r\left\langle\sigma'\left|\hat{s}_z\right|\sigma\right\rangle \\ &= 2\mu_B\mathcal{B}\delta_{\vec{k}',\vec{k}}\delta_{\sigma'\sigma}\times\begin{cases}1/2 & \sigma=\uparrow\\ -1/2 & \sigma=\downarrow\end{cases}.\end{aligned}$$

In exactly the same manner,

$$\begin{aligned}\left\langle c\vec{k}'\sigma'\left|\hat{X}\right|c\vec{k}\sigma\right\rangle &= -\frac{2\mu_B}{\Omega}\int\varphi^*_{c\vec{k}'}(\vec{r})\varphi_{c\vec{k}}(\vec{r})d^3r\left\langle\sigma'\left|\hat{s}_z\right|\sigma\right\rangle \\ &= -\frac{2\mu_B}{\Omega}\delta_{\vec{k}',\vec{k}}\delta_{\sigma'\sigma}\times\begin{cases}1/2 & \sigma=\uparrow\\ -1/2 & \sigma=\downarrow\end{cases}.\end{aligned}$$

In the static response Eq.(3.2), the summations over *n* and *m* are now over $\{\vec{k}\sigma\}$ and $\{\vec{k}'\sigma'\}$. Each sum over wave vectors is restricted to the Brillouin zone and the spin summations cover the two possibilities $\{\uparrow,\downarrow\}$. Three sums can be carried out immediately and the static magnetization is

$$\begin{aligned}\mathcal{M}(0) &= \sum_{\vec{k}\sigma,\vec{k}'\sigma'} f'(E_{c\vec{k}})\left\langle c\vec{k}'\sigma'\left|\hat{H}_1\right|c\vec{k}\sigma\right\rangle\left\langle c\vec{k}\sigma\left|\hat{X}\right|c\vec{k}'\sigma'\right\rangle \\ &= -\sum_{\vec{k}\sigma,\vec{k}'\sigma'} f'(E_{c\vec{k}})2\mu_B\mathcal{B}\delta_{\vec{k}',\vec{k}}\delta_{\sigma'\sigma}\frac{2\mu_B}{\Omega}\delta_{\vec{k}',\vec{k}}\delta_{\sigma'\sigma}\frac{1}{4} \\ &= -\frac{2\mu_B^2\mathcal{B}}{\Omega}\sum_{\vec{k}} f'(E_{c\vec{k}}).\end{aligned}$$

At sufficiently low temperatures, the derivative of the Fermi function is approximately a delta function $f'(E_{c\vec{k}}) \approx -\delta(E_{c\vec{k}} - E_F)$ and therefore

$$\mathcal{M}(0) \approx \mu_B^2\mathcal{B}D(E_F),$$

where $D(E_F) = 2\Omega^{-1}\sum_{\vec{k}}\delta(E_{c\vec{k}} - E_F)$ is the density of states at the Fermi level. This result is known as Pauli paramagnetism. We finally introduce the susceptibility χ_M via

$$\chi_M = \mu_0 \mathcal{M}(0)/\mathcal{B} = \mu_0 \mu_B^2 D(E_F),$$

which is known as the Pauli paramagnetic susceptibility.

3.2 Electric Current

The conductivity of electrons in an electric field can be handled much like the previous case. The electric field is taken to be polarized in the z-direction $\vec{\mathcal{E}} = \vec{e}_z \mathcal{E}$. The perturbation on the electrons is given by $-eV(\vec{r})$, where V is the electric potential related to the field via $\vec{\mathcal{E}} = -\nabla V$. This means that the current response is given by

$$J(\omega) = -\sum_{m,n} f_{nm} \frac{\langle \varphi_m | -eV | \varphi_n \rangle \langle \varphi_n | \hat{J} | \varphi_m \rangle}{E_{mn} - \hbar\omega - i\hbar\Gamma}, \tag{3.4}$$

where the current density operator is $\hat{J} = -e\hat{p}_z/\Omega m$, i.e. equal to charge times velocity. The matrix elements of this operator is, therefore, given by

$$\langle \varphi_n | \hat{J} | \varphi_m \rangle = \frac{-e}{m\Omega} \langle \varphi_n | \hat{p}_z | \varphi_m \rangle .$$

Using Eq. (2.2) to evaluate matrix elements of the potential, the expression for the current becomes

$$J(\omega) = \frac{-i\hbar e^2 \mathcal{E}}{m^2 \Omega} \sum_{m,n} \frac{f_{nm}}{E_{mn}} \cdot \frac{\left| \langle \varphi_m | \hat{p}_z | \varphi_n \rangle \right|^2}{E_{mn} - \hbar\omega - i\hbar\Gamma}. \tag{3.5}$$

We now again apply Eq.(3.1) due to the smallness of relevant E_{mn} and, in addition, we neglect E_{mn} in the frequency dependent denominator above so that

$$J(\omega) \approx -\frac{ie^2 \mathcal{E}}{m^2 \Omega(\omega + i\Gamma)} \sum_{m,n} f'(E_n) \left| \langle \varphi_m | \hat{p}_z | \varphi_n \rangle \right|^2 .$$

Next, we need the matrix elements of the momentum operator. Due to the Bloch form of the eigenstates Eq.(3.3) it follows that

$$\langle c\vec{k}'\sigma' | \hat{p}_z | c\vec{k}\sigma \rangle = \delta_{\sigma'\sigma} \delta_{\vec{k}',\vec{k}} \left\{ \hbar k_z + \langle u_{c\vec{k}} | \hat{p}_z | u_{c\vec{k}} \rangle \right\}. \tag{3.6}$$

These matrix elements, however, can be related to the energy using a few tricks. First, we note that, as usual, our single-particle states are eigenstates of a Hamiltonian $\hat{H}_0 = \hat{p}^2/2m + U$ so that

$$\left\{\frac{\hat{p}^2}{2m}+U\right\}u_{c\vec{k}}(\vec{r})e^{i\vec{k}\cdot\vec{r}} = E_{c\vec{k}}u_{c\vec{k}}(\vec{r})e^{i\vec{k}\cdot\vec{r}}$$

$$\Downarrow$$

$$\hat{h}_0(\vec{k})u_{c\vec{k}}(\vec{r}) = E_{c\vec{k}}u_{c\vec{k}}(\vec{r}),$$

where

$$\hat{h}_0(\vec{k}) = \frac{\hat{p}^2}{2m}+U+\frac{\hbar^2k^2}{2m}+\frac{\hbar\vec{k}\cdot\hat{\vec{p}}}{m}. \tag{3.7}$$

This *k*-dependent Hamiltonian is obtained by carrying through the phase factor $e^{i\vec{k}\cdot\vec{r}}$. Hence, multiplying on both sides by $u^*_{c\vec{k}}$ and integrating demonstrates that $E_{c\vec{k}} = \left\langle u_{c\vec{k}}\middle|\hat{h}_0(\vec{k})\middle|u_{c\vec{k}}\right\rangle$. We now wish to see how this energy varies with k_z. Hence, we take the derivative

$$\frac{\partial E_{c\vec{k}}}{\partial k_z} = \left\{\frac{\partial}{\partial k_z}\left\langle u_{c\vec{k}}\right|\right\}\hat{h}_0(\vec{k})\left|u_{c\vec{k}}\right\rangle + \left\langle u_{c\vec{k}}\middle|\frac{\partial\hat{h}_0(\vec{k})}{\partial k_z}\middle|u_{c\vec{k}}\right\rangle + \left\langle u_{c\vec{k}}\right|\hat{h}_0(\vec{k})\left\{\frac{\partial}{\partial k_z}\middle|u_{c\vec{k}}\right\rangle\right\}.$$

By means of the eigenvalue condition, this can be re-expressed as

$$\begin{aligned}\frac{\partial E_{c\vec{k}}}{\partial k_z} &= E_{c\vec{k}}\frac{\partial}{\partial k_z}\left\langle u_{c\vec{k}}\middle|u_{c\vec{k}}\right\rangle + \left\langle u_{c\vec{k}}\middle|\frac{\partial\hat{h}_0(\vec{k})}{\partial k_z}\middle|u_{c\vec{k}}\right\rangle \\ &= \left\langle u_{c\vec{k}}\middle|\frac{\partial\hat{h}_0(\vec{k})}{\partial k_z}\middle|u_{c\vec{k}}\right\rangle.\end{aligned}$$

The first contribution vanishes since $\left\langle u_{c\vec{k}}\middle|u_{c\vec{k}}\right\rangle = 1$. This result is an example of the famous Hellmann-Feynman theorem. It follows using Eq.(3.7) that

$$\frac{\partial E_{c\vec{k}}}{\partial k_z} = \frac{\hbar}{m}\left\{\hbar k_z + \left\langle u_{c\vec{k}}\middle|\hat{p}_z\middle|u_{c\vec{k}}\right\rangle\right\}. \tag{3.8}$$

This result allows us the rewrite the general matrix element Eq.(3.6) as

$$\left\langle c\vec{k}'\sigma'\middle|\hat{p}_z\middle|c\vec{k}\sigma\right\rangle = \delta_{\sigma'\sigma}\delta_{\vec{k}',\vec{k}}\,mv_c(\vec{k}), \quad v_c(\vec{k}) \equiv \frac{1}{\hbar}\frac{\partial E_{c\vec{k}}}{\partial k_z}. \tag{3.9}$$

This expression shows that the momentum matrix element is mass times "band velocity" $v_c(\vec{k})$. It should be emphasized, however, that this result is only valid for transitions within a single band. For transitions between bands, so-called *interband* transitions, the momentum matrix elements are more complicated. Precisely as in the magnetization example, the summations over n and m are replaced by sums over $\{\vec{k}\sigma\}$ and $\{\vec{k}'\sigma'\}$ and upon performing the sums over σ, σ' and $\vec{k}'$ the current is given by

$$J(\omega) = -\frac{2ie^2}{\hbar^2\Omega(\omega + i\Gamma)}\mathcal{E}\sum_{\vec{k}}\left(\frac{\partial E_{c\vec{k}}}{\partial k_z}\right)^2 f'(E_{c\vec{k}}).$$

Again, at low enough temperatures

$$J(\omega) \approx \frac{2ie^2}{\hbar^2\Omega(\omega + i\Gamma)}\mathcal{E}\sum_{\vec{k}}\left(\frac{\partial E_{c\vec{k}}}{\partial k_z}\right)^2 \delta(E_{c\vec{k}} - E_F).$$

This result may be written in a briefer form by introducing the plasma frequency ω_p via

$$\omega_p^2 \equiv \frac{2e^2}{\hbar^2\varepsilon_0\Omega}\sum_{\vec{k}}\left(\frac{\partial E_{c\vec{k}}}{\partial k_z}\right)^2 \delta(E_{c\vec{k}} - E_F). \tag{3.10}$$

Thus, if the conductivity is defined via $J(\omega) = \sigma(\omega)\mathcal{E}$ then

$$\sigma(\omega) = \frac{i\varepsilon_0\omega_p^2}{\omega + i\Gamma}. \tag{3.11}$$

This is the famous Drude form of the electric conductivity. It relates the induced current density to the applied time-dependent electric field. In a similar optical setting, a piece of metal might also be subjected to a time-dependent electric field due to the light source. In that case, we would look for the induced polarization as the response. Thus, one could expect that there is a connection between current and polarization $\vec{\mathcal{P}}$ and, indeed, there is. The connection follows from the fact that polarization is related to charge density $\rho(t)$ via $\rho(t) = -\nabla\cdot\vec{\mathcal{P}}(t)$. Moreover, the continuity relation $\partial\rho(t)/\partial t = -\nabla\cdot\vec{J}(t)$ relates charge and current densities. Put together, $\vec{J}(t) = \partial\vec{\mathcal{P}}(t)/\partial t$ or, for harmonic time variation, $\vec{J}(\omega) = -i\omega\vec{\mathcal{P}}(\omega)$. Now, just as $J(\omega) = \sigma(\omega)\mathcal{E}$ we can write $\mathcal{P}(\omega) = \varepsilon_0\chi(\omega)\mathcal{E}$, where $\chi(\omega)$ is the susceptibility. In turn, $\chi(\omega)$ is related to the dielectric constant $\varepsilon(\omega)$ via $\varepsilon(\omega) = 1 + \chi(\omega)$. Putting it all together, it follows that

$$\varepsilon(\omega) = 1 + \frac{i}{\varepsilon_0 \omega}\sigma(\omega).$$

In particular, for the intraband response of a metal described by the Drude expression we find

$$\varepsilon_{intra}(\omega) = 1 - \frac{\omega_p^2}{\omega(\omega + i\Gamma)}.$$

As an approximation the *interband* contribution may be taken to be roughly independent of frequency and so a simplified expression for the total dielectric constant is given by

$$\varepsilon(\omega) = \varepsilon_\infty - \frac{\omega_p^2}{\omega(\omega + i\Gamma)}, \tag{3.12}$$

where ε_∞ is a constant representing the high-frequency interband response. In Fig. 3.2, the optical response of bulk silver is considered [1]. For this material the appropriate parameters in the Drude model are $\hbar\omega_p = 9.3$ eV, $\hbar\Gamma = 0.03$ eV and $\varepsilon_\infty = 5$. As the plot demonstrates, the Drude form is quite accurate at low frequencies.

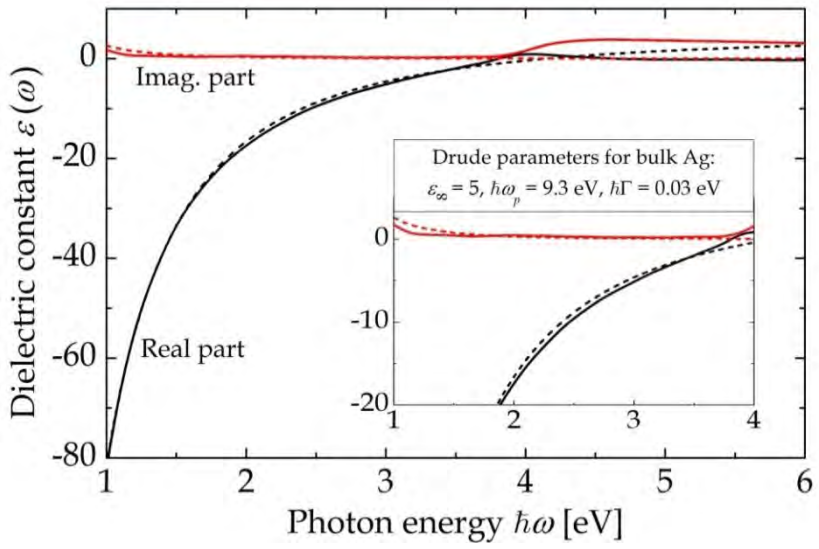


Figure 3.2. Comparison of experimental [1] and Drude model dielectric constant of Ag.

Exercise: Properties of free-electron metals

For a true free-electron metal show that

a) the density of states is $D(E_F) = \frac{3n}{2E_F}$, where n is the electron density, and therefore

$$\chi_M \approx \frac{3n\mu_0\mu_B^2}{2E_F},$$

b) the squared plasma frequency is $\omega_p^2 = \frac{e^2 n}{\varepsilon_0 m}$.

References

[1] P.B. Johnson and R.W. Christy, Phys. Rev. B6, 4370 (1972).

4. Electric Currents in Nanostructures

The electric properties and in particular electric currents are greatly influenced by quantum confinement effects in nanostructures and entirely new features arise. One of the most striking features is quantization of conductance so that under ideal conditions, the conductivity becomes an integer multiple of the fundamental conductivity quantum $2e^2/h$. In addition, important effects arise when the distance traveled by charge carriers becomes comparable to or less than the mean free path, i.e. the average distance traveled between scattering events due to e.g. phonons or impurities. In this "ballistic" regime, coherence of the electronic wave function is maintained and new features arise from interference effects.

We begin our study of conductivity in nanostructures by considering the general setup illustrated in Fig. 4.1. This figure shows a "system" through which an electric current is passing from left to right. The system, which might be a molecule, a semiconductor slab and so on, is connected to ideal wires or "leads" on both sides. Through these leads, an electric potential is applied so that an electric field exists inside the system. We suppose that we are in the ballistic regime so that an electron can be faithfully represented as a quantum mechanical wave incident on the system. This wave is then partly reflected and partly transmitted by the system. Hence, the important characteristics of the system lie in the reflection and transmission coefficients.

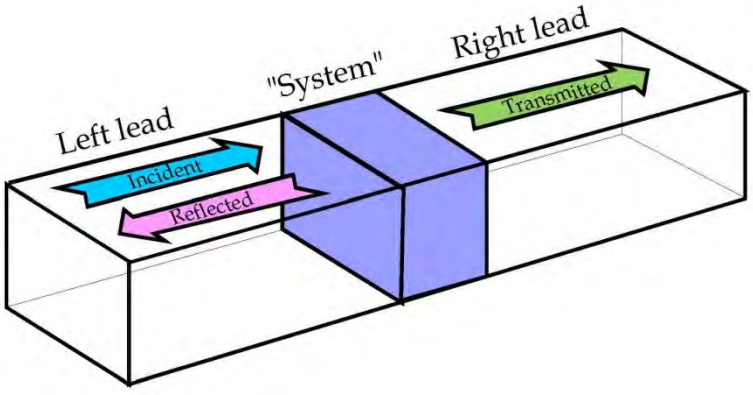


Figure 4.1. Illustration of the nanoscopic conductor consisting of a "system" connected to leads on both sides.

As in the previous chapter, the electric current along the z-axis is calculated as the response to the applied electric potential. Hence, the response expression is given by Eq.(3.5)

$$J(\omega) = \frac{-i\hbar e^2 \mathcal{E}}{m^2 \Omega} \sum_{m,n} \frac{f_{nm}}{E_{mn}} \cdot \frac{\left| \langle \varphi_m | \hat{p}_z | \varphi_n \rangle \right|^2}{E_{mn} - \hbar\omega - i\hbar\Gamma}.$$

In the present chapter, we'll restrict the discussion to the DC limit $\omega = 0$. Using Eq.(1.8) it is clear that the total DC current J_{DC} is $J_{DC} = \{J(0) + J^*(0)\}/2 = \mathrm{Re}\{J(0)\}$. Hence, we need to extract the following real part

$$\mathrm{Re}\left\{\frac{-i}{E_{mn} - i\hbar\Gamma}\right\} = \mathrm{Re}\left\{\frac{-i(E_{mn} + i\hbar\Gamma)}{E_{mn}^2 + \hbar^2\Gamma^2}\right\} = \frac{\hbar\Gamma}{E_{mn}^2 + \hbar^2\Gamma^2}.$$

As we are in the ballistic regime, we should take the limit $\Gamma \to 0$. Using the identity

$$\lim_{\varepsilon \to 0} \frac{\varepsilon}{x^2 + \varepsilon^2} = \pi\delta(x),$$

we have

$$\lim_{\Gamma \to 0} \mathrm{Re}\left\{\frac{-i}{E_{mn} - i\hbar\Gamma}\right\} = \pi\delta(E_{mn}).$$

so that

$$J_{DC} = \frac{\pi\hbar e^2 \mathcal{E}}{m^2 \Omega} \sum_{m,n} \frac{f_{nm}}{E_{mn}} \left|\left\langle \varphi_m \left| \hat{p}_z \right| \varphi_n \right\rangle\right|^2 \delta(E_{mn}). \tag{4.1}$$

The first factor in the sum can now be found by taking the limit as the energy difference E_{mn} vanishes

$$\lim_{E_{mn} \to 0} \frac{f_{nm}}{E_{mn}} = \lim_{E_{mn} \to 0} \frac{f(E_n) - f(E_n + E_{mn})}{E_{mn}} = -f'(E_n), \quad f'(E) = \frac{\partial f(E)}{\partial E}.$$

We therefore finally get

$$J_{DC} = -\frac{\pi\hbar e^2 \mathcal{E}}{m^2 \Omega} \sum_{m,n} f'(E_n) \left|\left\langle \varphi_m \left| \hat{p}_z \right| \varphi_n \right\rangle\right|^2 \delta(E_{mn}). \tag{4.2}$$

We now need to investigate in detail the relevant electronic states that carry the current. In any of the two leads, the eigenstates are of the form $\varphi_n(\vec{r}) = \phi_\nu(\vec{r}) \exp(\pm i k_n z)/\sqrt{L_i}$, where L_i is the length of the i'th lead and ϕ_ν describes the behavior the state in the directions perpendicular to the lead-direction. In general, ϕ_ν also depends on z and has the lattice periodicity in this direction. The energy eigenvalue corresponding to this state is denoted $E_n = E_\nu(k_n)$.

In the present quasi-one dimensional problem, it is obviously important to understand how the summations in Eq.(4.2) should be performed. Primarily,

summing over n means summing over a composite index $n=\{\nu k_n\sigma\}$, with σ is the spin index, and similarly for $m=\{\mu k_m\sigma'\}$. The spin summation is readily performed and simply produces a factor of two. Hence,

$$J_{DC}=-\frac{2\pi\hbar e^2\mathcal{E}}{m^2\Omega}\sum_{\mu,\nu,k_n,k_m} f'(E_\nu(k_n))\left|\left\langle\varphi_{\mu k_m}\left|\hat{p}_z\right|\varphi_{\nu k_n}\right\rangle\right|^2\delta\left(E_\mu(k_m)-E_\nu(k_n)\right).$$

Now, the sum over k_m can be converted to an integral using $\sum_k = L/2\pi\int dk$. In addition, we use the delta-function property

$$\int G(k)\delta\left(F(k)\right)dk=\frac{G(k_0)}{|F'(k_0)|},$$

where k_0 is the root of F, i.e. $F(k_0)=0$. This identity holds under the assumption that only a single root lies within the integration interval. The k derivative of the energy is basically the band velocity, c.f. Eq. (3.9)

$$v_\nu(k_n)\equiv\frac{1}{\hbar}\frac{\partial E_\nu(k_n)}{\partial k_n}.$$

We consequently find

$$J_{DC}=-\frac{e^2\mathcal{E}L}{m^2\Omega}\sum_{\mu,\nu,k_n}\frac{f'(E_\nu(k_n))}{|v_\nu(k_n)|}\left|\left\langle\varphi_{\mu k_m}\left|\hat{p}_z\right|\varphi_{\nu k_n}\right\rangle\right|^2. \tag{4.3}$$

It is important to note that in this expression k_m is understood to be determined by the condition $E_\mu(k_m)=E_\nu(k_n)$.

4.1 Matrix Elements

We now see what happens when the "system" region is inserted between the two leads. If an electron is propagating to the right in the left lead it encounters the system and two things can happen: either it is reflected back into the left lead OR it is transmitted into the right lead. In both cases, there is a chance that if the incident electron is in the n'th mode it will be reflected/transmitted into a different mode, say the l'th mode. Hence, the total wave function is given by

$$\varphi_{\nu k_n} = \begin{cases} N_n \left\{ \phi_\nu(\vec{r}) \frac{e^{ik_n z}}{\sqrt{L_1}} + \sum_\lambda \rho_{\nu\to\lambda} \phi_\lambda(\vec{r}) \frac{e^{-ik_l z}}{\sqrt{L_1}} \right\} & z \text{ in 1'st lead} \\ N_n \sum_\lambda \tau_{\nu\to\lambda} \phi_\lambda(\vec{r}) \frac{e^{ik_l z}}{\sqrt{L_2}} & z \text{ in 2'nd lead} \end{cases}, \qquad (4.4)$$

where N_n is a normalization constant and ρ and τ are reflection and transmission coefficients, respectively. The requirement that Eq.(4.4) be an eigenstate of the entire structure means that k_l and k_n are related via $E_\lambda(k_l) = E_\nu(k_n)$.

The three parts of the wave function (incident, reflected and transmitted) correspond to three currents. As usual, they can be calculated as the matrix element of the current operator $\hat{J} = -e\hat{p}_z / \Omega m$ taken over the appropriate volumes, i.e. the 1'st lead for incident and reflected and 2'nd lead for the transmitted current. This procedure yields

$$J_I = -|N_n|^2 \frac{e}{\Omega} v_\nu(k_n), \quad J_R = |N_n|^2 \frac{e}{\Omega} \sum_\lambda |\rho_{\nu\to\lambda}|^2 v_\lambda(k_l) \text{ and } J_T = -|N_n|^2 \frac{e}{\Omega} \sum_\lambda |\tau_{\nu\to\lambda}|^2 v_\lambda(k_l),$$

where we, again, need the velocities

$$v_\nu(k_n) = \frac{1}{mL_1} \left\langle \phi_\nu(\vec{r}) e^{ik_n z} \middle| \hat{p}_z \middle| \phi_\nu(\vec{r}) e^{ik_n z} \right\rangle = \frac{1}{\hbar} \frac{\partial E_\nu(k_n)}{\partial k_n}.$$

We can then calculate the reflectance $R_{\nu\to\lambda}$ as the ratio between the (absolute) current reflected into the *l*'th mode and the incident current and, similarly, the transmittance $T_{\nu\to\lambda}$ as the ratio between the current transmitted into the *l*'th mode and the incident current. These ratios are

$$R_{\nu\to\lambda} = |\rho_{\nu\to\lambda}|^2 \frac{v_\lambda(k_l)}{v_\nu(k_n)}, \quad T_{\nu\to\lambda} = |\tau_{\nu\to\lambda}|^2 \frac{v_\lambda(k_l)}{v_\nu(k_n)}.$$

Conservation of current implies that

$$\sum_l R_{\nu\to\lambda} + \sum_l T_{\nu\to\lambda} = 1.$$

In addition, we can express the normalization constant N_n in terms of these factors by integrating the square of the wave function. In the integral, we should include a contribution from the system region but if this is assumed much smaller than the leads we may safely ignore it and so

$$N_n = \left\{ 1 + \sum_l \frac{v_\nu(k_n)}{v_\lambda(k_l)} R_{\nu\to\lambda} + \sum_l \frac{v_\nu(k_n)}{v_\lambda(k_l)} T_{\nu\to\lambda} \right\}^{-1/2}.$$

What we really need is the momentum matrix elements to be used in Eq.(4.3). We therefore use Eq.(4.4) and a similar expression for φ_m to show that

$$\left\langle \varphi_{\mu k_m} \left| \hat{p}_z \right| \varphi_{\nu k_n} \right\rangle = m N_m^* N_n \left\{ v_\nu(k_n)\delta_{\mu\nu} - \sum_\lambda v_\lambda(k_l)\rho^*_{\mu\to\lambda}\rho_{\nu\to\lambda} + \sum_\lambda v_\lambda(k_l)\tau^*_{\mu\to\lambda}\tau_{\nu\to\lambda} \right\}.$$

Now, in order to evaluate Eq.(4.3) we should sum over *all* eigenstates, not only those given by Eq.(4.4). In fact, Eq.(4.4) only provides the right-travelling half the eigenstates and there is another left-travelling half given by

$$\tilde{\varphi}_n = \begin{cases} N_n \sum_l \tau^*_{\nu\to\lambda} \phi_\lambda(\vec{r}) \dfrac{e^{-ik_l z}}{\sqrt{L_1}} & z \text{ in 1'st lead} \\ N_n \left\{ \phi_\nu(\vec{r}) \dfrac{e^{-ik_n z}}{\sqrt{L_2}} - \sum_l \rho^*_{\nu\to\lambda} \phi_\lambda(\vec{r}) \dfrac{e^{ik_l z}}{\sqrt{L_2}} \right\} & z \text{ in 2'nd lead} \end{cases}.$$

These new states are easily shown to be orthogonal to the old ones. We subsequently need new momentum matrix elements and they turn out to be

$$\left\langle \tilde{\varphi}_m \left| \hat{p}_z \right| \tilde{\varphi}_n \right\rangle = -m N_m^* N_n \left\{ v_\nu(k_n)\delta_{\mu\nu} - \sum_l v_\lambda(k_l)\rho_{\mu\to\lambda}\rho^*_{\nu\to\lambda} + \sum_l v_\lambda(k_l)\tau_{\mu\to\lambda}\tau^*_{\nu\to\lambda} \right\}$$

$$\left\langle \tilde{\varphi}_m \left| \hat{p}_z \right| \varphi_n \right\rangle = -m N_m^* N_n \left\{ \sum_l v_\lambda(k_l)\tau_{\mu\to\lambda}\rho_{\nu\to\lambda} + \sum_l v_\lambda(k_l)\rho_{\mu\to\lambda}\tau_{\nu\to\lambda} \right\}$$

$$\left\langle \varphi_m \left| \hat{p}_z \right| \tilde{\varphi}_n \right\rangle = -m N_m^* N_n \left\{ \sum_l v_\lambda(k_l)\tau^*_{\mu\to\lambda}\rho^*_{\nu\to\lambda} + \sum_l v_\lambda(k_l)\rho^*_{\mu\to\lambda}\tau^*_{\nu\to\lambda} \right\}.$$

These are very general results but, unfortunately, rather complicated in further use. We consequently introduce a useful simplification.

4.2 Simplification: Decoupled Channels

In the expressions above, we see that electrons incident in the *n*'th mode may scatter into the *l*'th mode upon reflection or transmission. In this connection, we speak of the modes as "channels" in which electrons are transported. Obviously, these channels are coupled and an electron that is initially in a particular channel may scatter into another as a result of the coupling. We now make the simplifying assumption that coupling can be ignored so that among all the reflectances and transmittances, only $R_{\nu\to\nu}$ and $T_{\nu\to\nu}$ remain. This greatly simplifies the results above and we have

$$R_{\nu\to\nu}+T_{\nu\to\nu}=1$$
$$N_n=\left\{1+R_{\nu\to\nu}+T_{\nu\to\nu}\right\}^{-1/2}=2^{-1/2}$$
$$\left\langle\varphi_m\left|\hat{p}_z\right|\varphi_n\right\rangle=\frac{mv_\nu(k_n)}{2}\left\{1-R_{\nu\to\nu}+T_{\nu\to\nu}\right\}\delta_{\mu\nu}=mv_\nu(k_n)T_{\nu\to\nu}\delta_{\mu\nu}$$
$$\left\langle\tilde{\varphi}_m\left|\hat{p}_z\right|\tilde{\varphi}_n\right\rangle=-mv_\nu(k_n)T_{\nu\to\nu}\delta_{\mu\nu}$$
$$\left\langle\tilde{\varphi}_m\left|\hat{p}_z\right|\varphi_n\right\rangle=-mv_\nu(k_n)\tau_{\nu\to\nu}\rho_{\nu\to\nu}\delta_{\mu\nu},\quad\left\langle\varphi_m\left|\hat{p}_z\right|\tilde{\varphi}_n\right\rangle=-mv_\nu(k_n)\tau^*_{\nu\to\nu}\rho^*_{\nu\to\nu}\delta_{\mu\nu}.$$

After summing, this yields a very simple current density expression

$$J_{DC}=-\frac{e^2\mathcal{E}L}{\Omega}\sum_{\nu,k_n}f'(E_\nu(k_n))\left|v_\nu(k_n)\right|\left\{2T^2_{\nu\to\nu}+2T_{\nu\to\nu}R_{\nu\to\nu}\right\},$$

that can be reduced even further using the conservation relation $R_{\nu\to\nu}+T_{\nu\to\nu}=1$ so that

$$J_{DC}=-\frac{2e^2\mathcal{E}L}{\Omega}\sum_{\nu,k_n}f'(E_\nu(k_n))\left|v_\nu(k_n)\right|T_{\nu\to\nu}.$$

At this point, we can convert the remaining k summation to an integral that, in turn, can be turned into an energy integral

$$\begin{aligned}J_{DC}&=-\frac{e^2\mathcal{E}L^2}{\pi\Omega}\int\sum_\nu f'(E_\nu(k_n))\left|v_\nu(k_n)\right|T_{\nu\to\nu}dk_n\\&=-\frac{e^2\mathcal{E}L^2}{\pi\hbar\Omega}\int\sum_\nu f'(E)T_{\nu\to\nu}dE.\end{aligned}$$

To relate this result to directly measurable quantities we introduce the total potential difference (voltage) $V=\mathcal{E}L$ between the two leads and the total current equal to current density times cross sectional area of the leads, i.e. $I=AJ_{DC}$. Applying, in addition, the relations $\Omega=AL$ and $h=2\pi\hbar$ we find the celebrated *Landauer formula* for the quantized current

$$I=-\frac{2e^2V}{h}\int\sum_\nu f'(E)T_{\nu\to\nu}dE. \tag{4.5}$$

In the low-temperature limit, we then find

$$I=\frac{2e^2V}{h}\sum_\nu T_{\nu\to\nu}(E_F). \tag{4.6}$$

It shows that the conductance

$$G = \frac{I}{V} = \frac{2e^2}{h} \sum_{\nu} T_{\nu\to\nu}(E_F). \tag{4.7}$$

is quantized. As an example, we may assume effective mass dispersion $E_\nu(k_n) = E_\nu^\perp + \hbar^2 k_n^2 / 2m_e$, where $E_\nu^\perp = E_\nu(k_n = 0)$ is the quantization energy. Then, in the idealized case of vanishing reflection, the transmittance equals unity whenever the Fermi level exceed the band edge, i.e. $T_{\nu\to\nu}(E_F) = \theta(E_F - E_\nu^\perp)$ and the conductance is simply $G = NG_0$, where $G_0 = 2e^2/h = (12.90641\ \mathrm{k\Omega})^{-1}$ is the fundamental conductance quantum and N is the number of channels below the Fermi level. If the assumption of completely ballistic transport is not fully applicable because of scattering losses, the conductance formula should be convoluted with the broadening function $\hbar\Gamma / (E^2 + \hbar^2\Gamma^2) / \pi$. This produces the broadened expression

$$G = \frac{2e^2}{h} \sum_{\nu} \left\{ \frac{1}{\pi} \tan^{-1}\left(\frac{E_F - E_\nu^\perp}{\hbar\Gamma} \right) + \frac{1}{2} \right\}. \tag{4.8}$$

Figure 4.2 illustrates the quantized conductance as a function of Fermi energy for a fictitious system in which $E_\nu^\perp = \nu \cdot 1\ \mathrm{eV}$, $\nu = 1, 2, 3, \ldots$ and so on. The ideal case $\hbar\Gamma = 0$ as well as a case of finite broadening $\hbar\Gamma = 0.1\ \mathrm{eV}$ is shown in the two graphs.

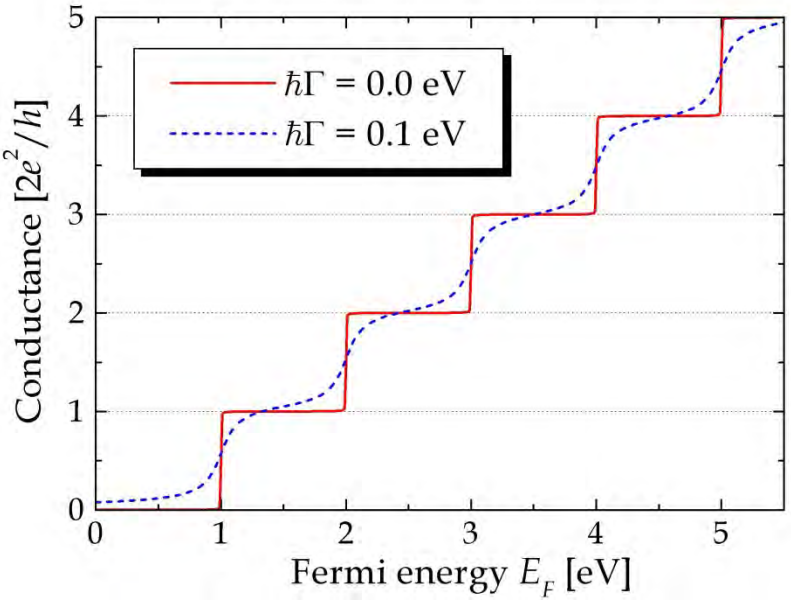


Figure 4.2. Illustration of conductance quantization for the ideal and broadened cases.

The figure 4.3 below shows experimental conductance plots for a GaAs/AlGaAs junction onto which an extremely narrow conducting channel is formed by metallic point contacts [1]. Tunneling occurs through a central metallic contact connected to leads on the left and right via tunneling regions. The tunneling current is recorded as a function of a gate potential applied to the central contact.

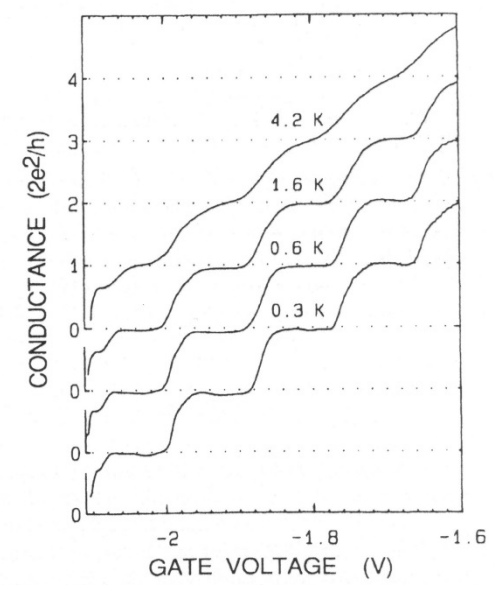


Figure 4.3. Experimental conductance traces for a metallic point contact as a function of gate voltage. Taken from [1].

Exercise: Mode counting and minimum resistance

Even for an ideal conductor having all transmittances $T_{\nu\to\nu}=1$ there is a maximal conductance given by Eq. (4.7)

$$G_{\max}=\frac{2e^2}{h}\cdot\#\text{modes}.$$

Here, $\#\text{modes}$ should be understood as the number of occupied electron channels, i.e. channels for which $E_\nu^\perp \le E_F$ so that the band crosses the Fermi level. For a conductor with a square profile defined by side lengths L_x and L_y and infinite surface barriers, the mode index ν is actually two-dimensional $\nu\to pq$ because the energies depend on two quantum numbers (see Appendix 1)

$$E_{pq}^\perp=\frac{\hbar^2\pi^2}{2m_e}\left(\frac{p^2}{L_x^2}+\frac{q^2}{L_y^2}\right).$$

We first assume an ideal quantum well, i.e. take L_y so small that only a single y-mode contributes.

a) Show that by setting $E_{pq}^\perp=E_F$ the maximal occupied p-level and, hence, the number of occupied modes becomes

$$\#\,\text{modes} = \sqrt{\frac{2m_e L_x^2}{\pi^2\hbar^2}\left(E_F - E_{01}^{\perp}\right)}\,.$$

Next, we set $L_x = L_y \equiv L$ and consider a square quantum wire. In this case, setting $E_{pq}^{\perp} = E_F$ means that $(p^2+q^2)_{\max} = E_F 2m_e L^2/(\pi^2\hbar^2)$. The problem of mode counting therefore reduces to finding the number of non-negative integers within a circle of radius $\sqrt{E_F 2m_e}L/(\pi\hbar)$.

b) By assuming this radius much larger than unity, show that

$$\#\,\text{modes} = E_F m_e L^2/(2\pi\hbar^2)\,.$$

We now assume that all bands are parabolic, i.e. that the full energy dispersion is $E_{pq} = E_{pq}^{\perp} + \hbar^2 k^2/(2m_e)$. Thus, for a sufficiently thick wire we know that the electron density is $n = (2E_F m_e/\hbar^2)^{3/2}/3\pi^2$.

c) Show that the minimum resistance becomes

$$R_{\min} = \left(\frac{\pi}{3n}\right)^{2/3}\frac{4\hbar}{e^2L^2}\,.$$

References

[1] B.J. de Wees *et al.* Phys. Rev. B43, 12431 (1991).

5. Electron Transmission and Reflection

In the previous chapter, the general formulas for the electric current were developed in terms of electron transmittances. Here, we will continue by actually computing the transmittance for a range of illustrative examples. To simplify matters, we restrict ourselves to a one-dimensional description, i.e. we ignore what happens in the directions parallel to barriers and other potential steps. Throughout, we rely on an effective mass picture, in which electrons behave as free electrons but with a modified mass m_e rather than the free electron mass (explained further in the next chapter). We also assume the potential energy to be piecewise constant but, as will be shown below, any true profile can be approximated by a sequence of constant potential steps via the "staircasing" approach. Thus, very general cases can be treated using the methods developed here.

Electrons with energy E moving through a region with constant potential energy V are described by the wave function

$$\psi(x) = Ae^{ikx} + Be^{-ikx}, \quad k = \sqrt{\frac{2m_e}{\hbar^2}(E-V)}.$$

If we evaluate the same state at a shifted position $x+d$ we obviously have $\psi(x+d) = Ae^{ik(x+d)} + Be^{-ik(x+d)}$. Hence, if we write this state as $\psi(x+d) = Ce^{ikx} + De^{-ikx}$, we can write

$$\begin{pmatrix} A \\ B \end{pmatrix} = \overleftrightarrow{M} \cdot \begin{pmatrix} C \\ D \end{pmatrix}, \quad \overleftrightarrow{M} = \begin{pmatrix} e^{-ikd} & 0 \\ 0 & e^{ikd} \end{pmatrix}. \tag{5.1}$$

Similarly, the electron may pass from a material having potential energy V_1 to a position with potential energy V_2. In this case, the wave functions for the positions 1 and 2 (located immediately left and right of the interface, respectively) are

$$\psi_1(x) = Ae^{ik_1 x} + Be^{-ik_1 x}, \quad k_1 = \sqrt{\frac{2m_e}{\hbar^2}(E-V_1)}$$
$$\psi_2(x) = Ce^{ik_2 x} + De^{-ik_2 x}, \quad k_2 = \sqrt{\frac{2m_e}{\hbar^2}(E-V_2)}.$$

To relate the coefficients in this situation, we need to match the wave function and its derivative at the junction

$$A + B = C + D$$
$$k_1(A-B) = k_2(C-D).$$

By adding and subtracting these equations, it follows that we may write the relations as the matrix expression

$$\begin{pmatrix} A \\ B \end{pmatrix} = \overleftrightarrow{M} \cdot \begin{pmatrix} C \\ D \end{pmatrix}, \quad \overleftrightarrow{M} = \frac{1}{2} \begin{pmatrix} 1+\frac{k_2}{k_1} & 1-\frac{k_2}{k_1} \\ 1-\frac{k_2}{k_1} & 1+\frac{k_2}{k_1} \end{pmatrix}. \tag{5.2}$$

In this manner, transmission through any potential energy profile can be modeled by a series of matrix multiplications. As simple examples, we consider tunneling through single and double barriers as illustrated in Fig. 5.1.

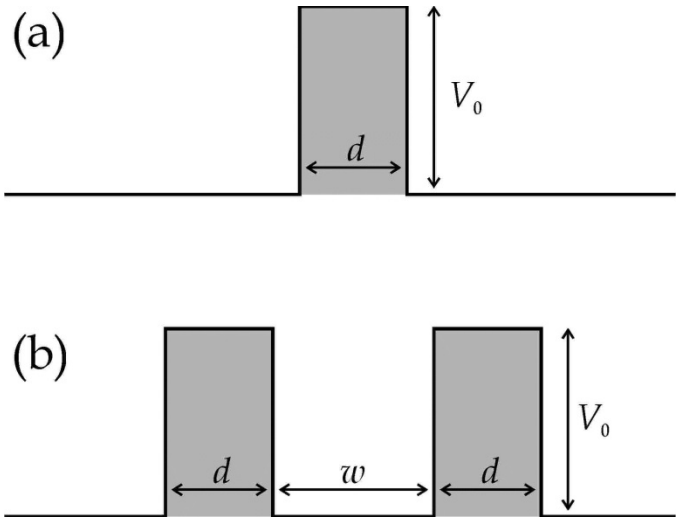


Figure 5.1. Single and double tunneling barriers.

For the single barrier, transmission involves (1) entering the barrier, (2) transversing the barrier, and (3) exiting the barrier. If we choose the potential energy outside the barrier as the zero-point, the wave number in this region is simply $k = k_1 = \sqrt{2m_e E/\hbar^2}$. Inside the barrier the potential energy is V_0 and, correspondingly, the wave number is $k_2 = i\beta, \quad \beta = \sqrt{2m_e(V_0 - E)/\hbar^2}$. The combined effect of the three processes is described by the "system" matrix

$$\overleftrightarrow{M} = \overleftrightarrow{M}_{in} \cdot \overleftrightarrow{M}_B \cdot \overleftrightarrow{M}_{out}.$$

Now, in a reflection-transmission experiment with an incoming wave from the left, the (un-normalized) wave functions far to the left and right are, respectively,

$$\psi_L(x) = e^{ikx} + re^{-ikx}, \quad \psi_R(x) = te^{ikx},$$

where r and t are reflection and transmission coefficients. In terms of the system matrix, we therefore have

$$\begin{pmatrix} 1 \\ r \end{pmatrix} = \overleftrightarrow{M} \cdot \begin{pmatrix} t \\ 0 \end{pmatrix}$$

so that $t = 1/M_{11}$ and $r = M_{21}/M_{11}$. Finally, the transmission and reflection probabilities can be computed from $T = |t|^2 = 1/|M_{11}|^2$ and $R = |r|^2 = |M_{21}/M_{11}|^2$, respectively. As we will see below, quite complicated geometries can be handled using this technique.

For the particular case of a single rectangular barrier of width d, the general expressions above yield the following matrices:

$$\overleftrightarrow{M}_{in} = \frac{1}{2}\begin{pmatrix} 1+i\frac{\beta}{k} & 1-i\frac{\beta}{k} \\ 1-i\frac{\beta}{k} & 1+i\frac{\beta}{k} \end{pmatrix}, \quad \overleftrightarrow{M}_{out} = \frac{1}{2}\begin{pmatrix} 1-i\frac{k}{\beta} & 1+i\frac{k}{\beta} \\ 1+i\frac{k}{\beta} & 1-i\frac{k}{\beta} \end{pmatrix}, \quad \overleftrightarrow{M}_{B} = \begin{pmatrix} e^{\beta d} & 0 \\ 0 & e^{-\beta d} \end{pmatrix}.$$

Performing matrix multiplications, it follows that the system matrix for the single barrier ("S") becomes

$$\overleftrightarrow{M}^{(S)} = \begin{pmatrix} \cosh\beta d + i\frac{(\beta^2-k^2)\sinh\beta d}{2\beta k} & i\frac{(\beta^2+k^2)\sinh\beta d}{2\beta k} \\ -i\frac{(\beta^2+k^2)\sinh\beta d}{2\beta k} & \cosh\beta d - i\frac{(\beta^2-k^2)\sinh\beta d}{2\beta k} \end{pmatrix}.$$

And the transmittance is

$$T^{(S)} = \frac{4\beta^2 k^2}{4\beta^2 k^2 \cosh^2\beta d + (\beta^2-k^2)^2 \sinh^2\beta d}. \tag{5.3}$$

In the exercise at the end of the chapter, a plot of the single barrier transmittance is shown. For the double barrier ("D") in Fig. 5.1b we need, in addition, the matrix for translation between the barriers

$$\overleftrightarrow{M}_{W} = \begin{pmatrix} e^{-ikw} & 0 \\ 0 & e^{ikw} \end{pmatrix},$$

and the transmission matrix becomes

$$\begin{aligned}\overleftrightarrow{M}^{(D)} &= \overleftrightarrow{M}_{in}\cdot\overleftrightarrow{M}_{B}\cdot\overleftrightarrow{M}_{out}\cdot\overleftrightarrow{M}_{W}\cdot\overleftrightarrow{M}_{in}\cdot\overleftrightarrow{M}_{B}\cdot\overleftrightarrow{M}_{out}\\ &= \overleftrightarrow{M}^{(S)}\cdot\overleftrightarrow{M}_{W}\cdot\overleftrightarrow{M}^{(S)}.\end{aligned}$$

It is easily demonstrated that

$$M_{11}^{(D)} = M_{11}^{(S)}e^{-ikw}M_{11}^{(S)} + M_{12}^{(S)}e^{ikw}M_{21}^{(S)}.$$

Below, we plot the double barrier transmittance vs. energy using $V_0 = 4$ eV, d = 6Å and w = 8Å and taking m_e equal to the free electron mass. The peaks correspond to "resonant tunneling" [1] and they are located, where true bound eigenstates would appear if the barriers were infinitely wide. Note that we plot the result for $E > V_0$ also. In fact, the only modification for this energy range is that β becomes imaginary. Since the formulas above remain valid for imaginary β, the results still apply for this energy range.

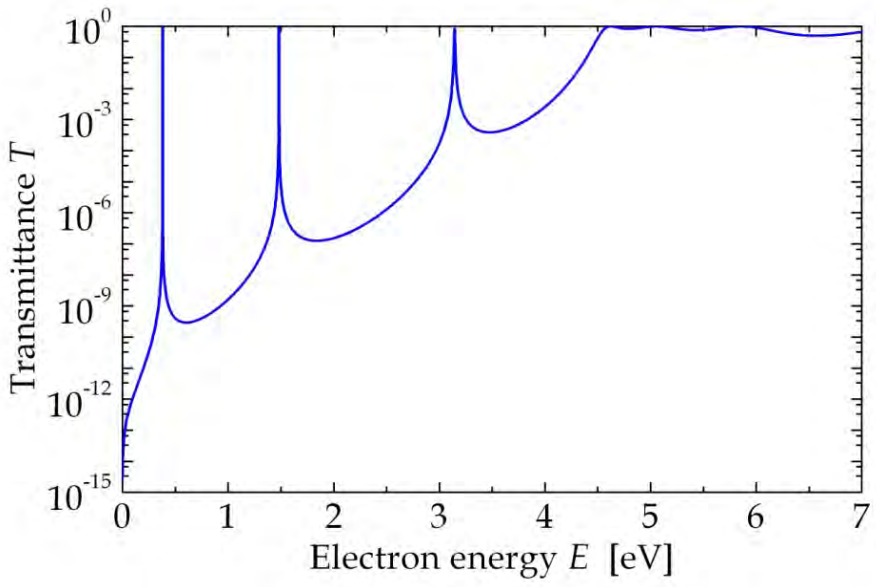


Figure 5.2. Tunneling transmittance T for the resonant tunneling structure shown in Fig. 5.1b using $V_0 = 4$ eV, d = 6Å and w = 8Å.

Any full system matrix built from multiplying the matrices above will be of the form

$$\overleftrightarrow{M} = \begin{pmatrix} M_{11} & M_{21}^{*} \\ M_{21} & M_{11}^{*} \end{pmatrix}.$$

Since $T = 1/|M_{11}|^2$ and $R = |M_{21}/M_{11}|^2$, we find for the determinant

$$\det\overleftrightarrow{M} = |M_{11}|^2 - |M_{21}|^2 = \frac{1}{T} - \frac{R}{T} = \frac{1-R}{T}.$$

As electrons don't disappear, we should have $R+T=1$ and so any valid system matrix must have unit determinant $\det \overleftrightarrow{M}=1$.

5.1 Triangular Barrier

We will now attempt to tackle a much more complicated problem: That of a triangular barrier such as the one illustrated in Fig. 5.3.

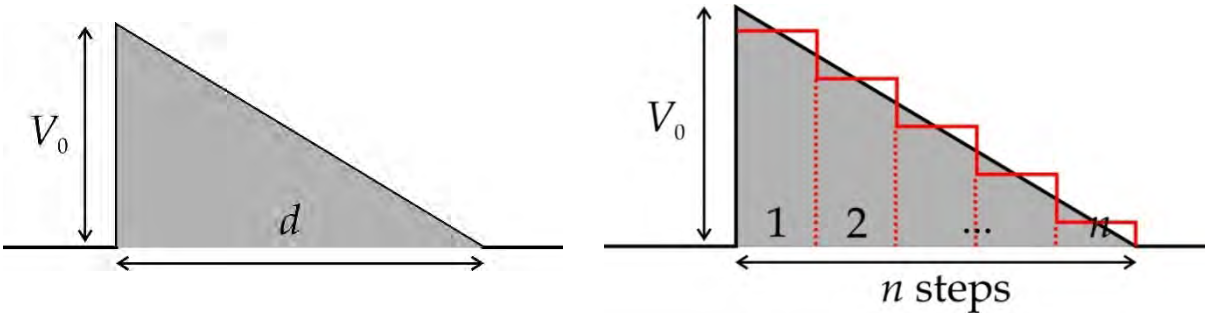


Figure 5.3. The triangular tunneling barrier (left) and the "staircased" approximation (right).

Such a barrier naturally arises in field emission structures. Here, electrons are emitted from a metal by approaching a positively biased electrode to the metal surface. For a simple planar geometry, the electric field is constant and, hence, the potential decreases linearly with the distance from the surface (actually the image charge effect modifies the potential near the surface [2] but this complication is ignored). Therefore the electrons have to overcome a triangular barrier, whose height V_0 equals the metal work function for electrons at the Fermi level. This tunneling problem can be attacked in several ways, ranging from approximate over fully numerical to analytical approaches. The first type (approximate) is called the WKB approximation after Wentzel, Kramers and Brillouin. The starting point is the $T^{(S)}$ expression Eq.(5.3) for the single barrier, derived above. If the barrier is sufficiently wide, the argument βd is large and we may approximate

$$T^{(S)} \approx \frac{16\beta^2 k^2}{4\beta^2 k^2 + (\beta^2 - k^2)^2} \exp(-2\beta d).$$

For simplicity, the prefactor can also be replaced by unity, as we are mainly interested in the thickness dependence, i.e. $T^{(S)} \approx \exp(-2\beta d)$. In this manner, we are sure that $T^{(S)} \to 1$ as $d \to 0$. Now, for this case it is clearly no approximation to write

$$\beta d = \int_0^d \sqrt{\frac{2m_e}{\hbar^2}(V_0 - E)}\, dx.$$

The WKB approximation consists in applying this expression to barriers that are not rectangular. In general, the tunneling range through the barrier is determined by the

electron energy E as illustrates in Fig. 5.4. Here, the tunneling range is between x_- and x_+ that are defined by the intersections $V(x_\pm) = E$.

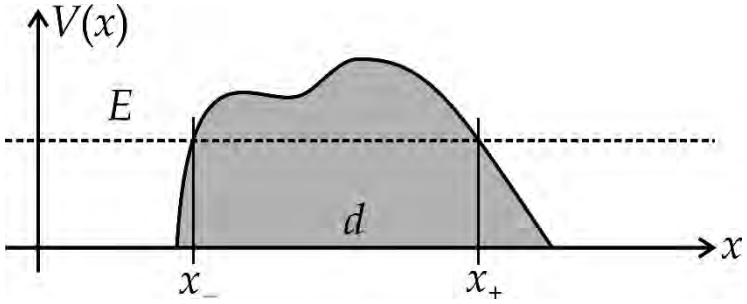


Figure 5.4. Sketch of an arbitrary tunneling barrier with classical turning points x_- and x_+.

In this manner, the general approximate transmittance becomes

$$T_{WKB} = \exp\left\{-2\int_{x_-}^{x_+}\sqrt{\frac{2m_e}{\hbar^2}\left(V(x)-E\right)}\,dx\right\}. \tag{5.4}$$

The triangular barrier is defined by the potential

$$V(x) = \begin{cases} V_0 - Fx & 0 < x < d \\ 0 & \text{otherwise} \end{cases},$$

where F is the force proportional to the electric field $F = e\mathcal{E}$ and the width d is related to F via $d = V_0 / F$. Also, as long as $0 < E < V_0$ we find $x_- = 0$ and $x_+ = (V_0 - E)/F$. In this case, an elementary integral shows that

$$T_{WKB} = \exp\left\{-\frac{4}{3F}\sqrt{\frac{2m_e}{\hbar^2}\left(V_0 - E\right)^3}\right\}. \tag{5.5}$$

For the triangular barrier, however, it is possible to find a relatively simple analytical solution. The key point is to consider the Schrödinger equation in the barrier region $0 < x < d$:

$$-\frac{\hbar^2}{2m_e}\psi''(x) + \left(V_0 - Fx\right)\psi(x) = E\psi(x).$$

Rewriting slightly, we find

$$\psi''(x) = \frac{2m_eF}{\hbar^2}\left(x_0 - x\right)\psi(x), \quad x_0 = \frac{V_0 - E}{F}.$$

This resembles Airy's differential equation

$$\psi''(z) = z\psi(z)\,.$$

In fact, if we write $z = q(x_0 - x)$, Airy's equation reads

$$\frac{d^2}{dx^2}\psi(q(x_0-x)) = q^3(x_0-x)\psi(q(x_0-x))\,.$$

This clearly fits if $q = \left(2m_e F/\hbar^2\right)^{1/3}$. The two linearly independent solutions are denoted Ai and Bi and so the full solution must be

$$\psi(x) = A\cdot \mathrm{Ai}\left(q(x_0-x)\right) + B\cdot \mathrm{Bi}\left(q(x_0-x)\right).$$

The reflection and transmission coefficients r and t along with A and B follow from the four boundary conditions

$$\begin{aligned}
1 + r &= A\cdot \mathrm{Ai}\left(qx_0\right) + B\cdot \mathrm{Bi}\left(qx_0\right)\\
ik(1-r) &= -q\left[A\cdot \mathrm{Ai}'\left(qx_0\right) + B\cdot \mathrm{Bi}'\left(qx_0\right)\right]\\
t &= A\cdot \mathrm{Ai}\left(q(x_0-d)\right) + B\cdot \mathrm{Bi}\left(q(x_0-d)\right)\\
ikt &= -q\left[A\cdot \mathrm{Ai}'\left(q(x_0-d)\right) + B\cdot \mathrm{Bi}'\left(q(x_0-d)\right)\right].
\end{aligned}$$

In this manner, utilizing the Wronskian $W[\mathrm{Ai},\mathrm{Bi}] = 1/\pi$,

$$\begin{aligned}
t &= \frac{2kq}{\pi\left[\left(ika+qa'\right)\left(k\tilde{b}+iq\tilde{b}'\right) - \left(ikb+qb'\right)\left(k\tilde{a}+iq\tilde{a}'\right)\right]}\\
a &= \mathrm{Ai}\left(qx_0\right),\ \tilde{a} = \mathrm{Ai}\left(q(x_0-d)\right),\ b = \mathrm{Bi}\left(qx_0\right),\ \tilde{b} = \mathrm{Bi}\left(q(x_0-d)\right)\\
a' &= \mathrm{Ai}'\left(qx_0\right),\ \tilde{a}' = \mathrm{Ai}'\left(q(x_0-d)\right),\ b' = \mathrm{Bi}'\left(qx_0\right),\ \tilde{b}' = \mathrm{Bi}'\left(q(x_0-d)\right).
\end{aligned} \tag{5.6}$$

The fully numerical approach to transmission through arbitrary barriers consists of applying the matrix formalism presented above. As the method only applies to potential profiles that are piecewise constant, we need to approximate the barrier using "staircasing". As an example, a staircased triangular barrier is shown in the right panel of Fig. 5.3. The idea is that the barrier is chopped into n pieces, each having a certain width and constant barrier height. When the number of steps n becomes large, the result comes close to the exact one.

In Fig. 5.5, we have compared the exact result Eq.(5.6) to the WKB approximation Eq.(5.5) and to the staircasing approach using n = 2 and 3. Here, a barrier of $V_0 = 5$ eV and a force of $F = 1$ eV/Å are assumed. Notice how well the n = 3 result approximates the exact curve.

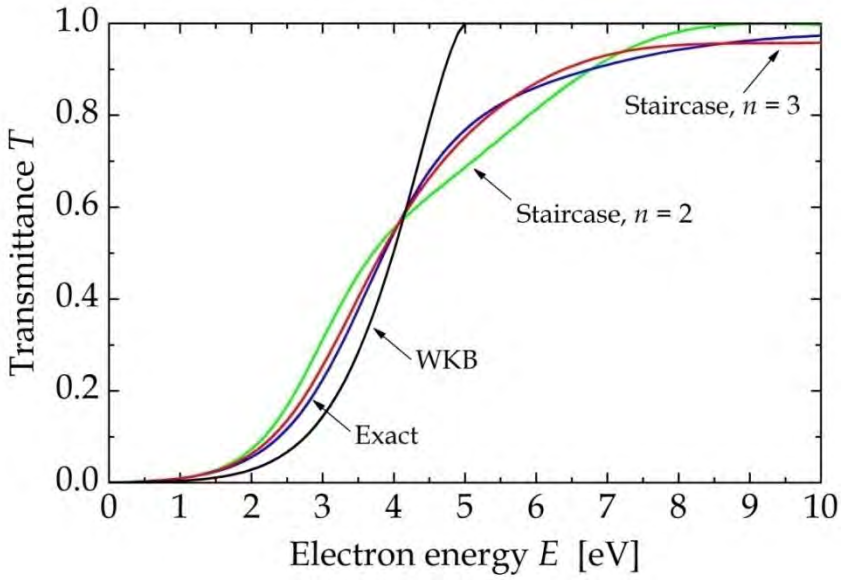


Figure 5.5. Electron transmittance for the triangular barrier of $V_0 = 5$ eV using (1) the WKB approximation, (2) the staircase method, and (3) the exact solution.

Exercise

In Fig. 5.6 below, we have plotted the transmittance of a single barrier with $V_0 = 4$ eV and d = 6Å. Similarly to the double barrier case, we note that the plot extends to the "fly-over" range $E > V_0$. It is also noticed that a unit transmittance is found for some very specific barrier thicknesses.

a) Show that in the fly-over energy range

$$\overleftrightarrow{M}_{in} = \frac{1}{2}\begin{pmatrix} 1+\frac{k_B}{k} & 1-\frac{k_B}{k} \\ 1-\frac{k_B}{k} & 1+\frac{k_B}{k} \end{pmatrix}, \quad \overleftrightarrow{M}_{out} = \frac{1}{2}\begin{pmatrix} 1+\frac{k}{k_B} & 1-\frac{k}{k_B} \\ 1-\frac{k}{k_B} & 1+\frac{k}{k_B} \end{pmatrix}, \quad \overleftrightarrow{M}_B = \begin{pmatrix} e^{-ik_B d} & 0 \\ 0 & e^{ik_B d} \end{pmatrix},$$

where $k_B = \sqrt{2m_e(E-V_0)/\hbar^2}$.

b) Show that the modified (1,1) element of the system matrix becomes

$$\overleftrightarrow{M}_{11}^{(S)} = \cos k_B d - i\frac{(k^2+k_B^2)\sin k_B d}{2kk_B}$$

and, correspondingly, the transmittance

$$T^{(S)} = \frac{1}{\cos^2 k_B d + \dfrac{(k^2 + k_B^2)^2}{4k^2 k_B^2} \sin^2 k_B d} = \frac{1}{1 + \dfrac{k^4 + k_B^4}{4k^2 k_B^2} \sin^2 k_B d}. \qquad (5.7)$$

Writing the result in the second form clearly demonstrates that $T^{(S)} \leq 1$. It is also clear that the first form agrees with Eq.(5.3) as follows from the substitutions $\beta \to -ik_B$, $\cosh\beta \to \cos k_B$ and $\sinh\beta \to -i\sin k_B$.

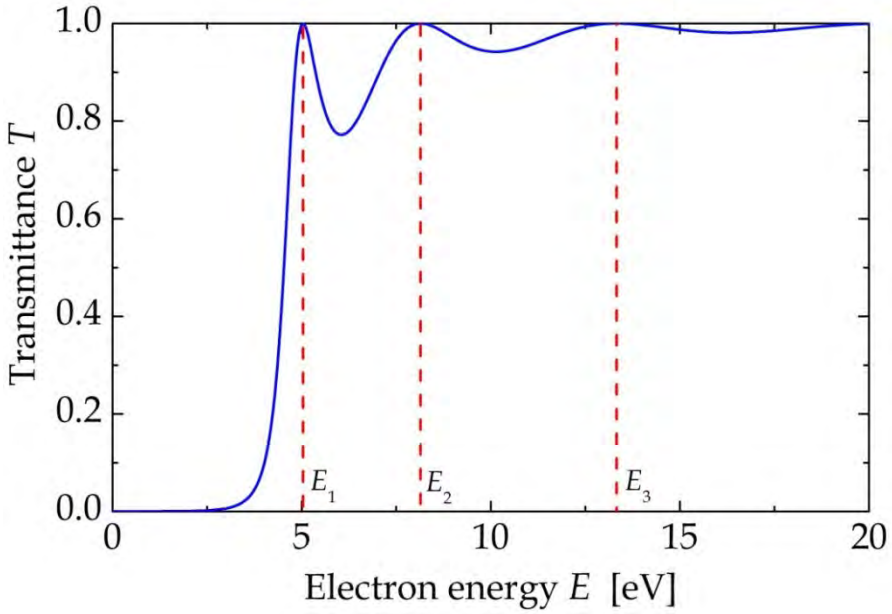


Figure 5.6.Transmission of a single rectangular barrier with energies at transmittance maxima highlighted.

c) Show that the transmission attains its maximum value of unity at particular values of the energy given by

$$E_p = V_0 + \frac{\hbar^2}{2m_e}\left(\frac{p\pi}{d}\right)^2, \quad p = 1, 2, 3, \ldots.$$

References

[1] B. Ricco and M. Ya. Azbel, Phys. Rev. B29, 1970 (1984).
[2] R.H. Fowler and L. Nordheim, Proc. R. Soc. A. 119, 173 (1928).

6. Electron Transmission in Molecules: Introduction

In this chapter, we turn to the transmission of electrons through molecules. We will therefore not use any effective mass approximation nor the rectangular barriers that were introduced in the previous chapter. Those concepts were appropriate for semiconductor nanostructures but will not suffice for individual molecules. We will start by considering a simple example that can be analyzed analytically. This geometry is illustrated in Fig. 6.1 and consists of a benzene molecule attached to simple one-dimensional leads with a single atom per unit cell [1].

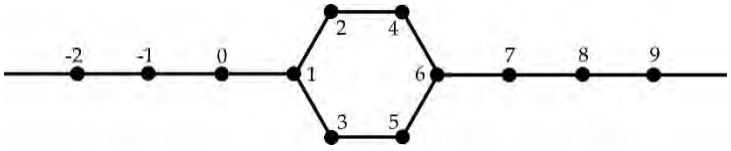


Figure 6.1. Transmission geometry consisting of a benzene ring ("system") attached to two semi-infinite mono-atomic chains ("leads").

We analyze the states using the tight-binding approach described in Appendix 2. Moreover, we will assume that only a single orbital on each atom contributes to the transport. This orbital will be denoted $|n\rangle$ if it belongs to the n'th atom. The quantum states can be written as expansions in this basis

$$|\varphi\rangle = \sum_n c_n |n\rangle,$$

where the sum is over all sites (atoms) in the structure. In our case, the structure consists of infinite leads attached to a finite molecule. Generally, a large part of the calculation is concerned with handling infinite leads. Here, as a starting point, we'll consider the isolated molecule. For the benzene ring chosen as our example, the molecule has 6 sites. For simplicity, assume that only nearest neighbor coupling exists and denote the hopping matrix element by $-\gamma$ (the minus sign is in order to keep $\gamma > 0$). Also, we'll choose the zero point of energy such that the onsite matrix elements $\langle n|\hat{H}|n\rangle = 0$. Hence, the eigenvalue problem for the isolated system reads

$$\overleftrightarrow{H}_S \cdot \vec{c} = E\vec{c}, \quad \overleftrightarrow{H}_S = -\begin{pmatrix} 0 & \gamma & \gamma & 0 & 0 & 0 \\ \gamma & 0 & 0 & \gamma & 0 & 0 \\ \gamma & 0 & 0 & 0 & \gamma & 0 \\ 0 & \gamma & 0 & 0 & 0 & \gamma \\ 0 & 0 & \gamma & 0 & 0 & \gamma \\ 0 & 0 & 0 & \gamma & \gamma & 0 \end{pmatrix}, \quad \vec{c} = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ c_4 \\ c_5 \\ c_6 \end{pmatrix}. \tag{6.1}$$

It's a simple matter to show that the eigenvalues are $\pm 2\gamma$ and $\pm\gamma$ with the last pair twice degenerate. The isolated leads are even simpler. Under the same assumptions as for the molecule, we find

$$-(\gamma c_{n-1} + \gamma c_{n+1}) = E c_n .$$

The eigenstates are simple propagating waves of the form $c_n = Ae^{iqna} + Be^{-iqna}$ and the associated energy is $E = -2\gamma\cos(qa)$, where a is the lattice constant of the chain. Hence, the energies form a continuous band with values in the range $-2\gamma \le E \le 2\gamma$.

We now aim to couple system and leads together. In doing so, we will assume that a wave of unit amplitude is incident from the left and then partially reflected by the system. Designating the reflection coefficient by r, the full wave in the left lead is thus $c_n^{(in)} = e^{iqna} + re^{-iqna}$. Similarly, in the right lead a transmitted wave $c_n^{(out)} = te^{iqna}$ propagates, where t is the transmission coefficient. Now, if we naively tried to set up the eigenvalue problem for the coupled system, an infinite matrix would result. Fortunately, the influence of the infinite leads can be accounted for via their electronic Green's function. In the present mono-atomic case, the construction is particularly simple. First, we note that $c_0 = 1 + r$ and $c_{-1} = e^{-iqa} + re^{iqa}$ and, hence, $c_{-1} = e^{-iqa} + (c_0 - 1)e^{iqa}$. Secondly, the hopping integral coupling the left lead and the system is denoted τ. Then, the Schrödinger equation for the $0'th$ site becomes $-\gamma c_{-1} + \tau c_1 = E c_0$. Combining these relations, we now see that

$$(E - H_L)c_0 = H_L^* - H_L + \tau c_1, \quad H_L = -\gamma e^{iqa} .$$

If the coupling between right lead and system is also τ we find, similarly,

$$(E - H_R)c_7 = \tau c_6, \quad H_R = -\gamma e^{iqa} .$$

We are now in a position to write the matrix equation for the full coupled problem. To keep the notation as simple as possible, we will suppress vector and matrix symbols. Thus,

$$\begin{pmatrix} g_L^{-1} & -V_L^\dagger & 0 \\ -V_L & EI - H_S & -V_R^\dagger \\ 0 & -V_R & g_R^{-1} \end{pmatrix} \cdot \begin{pmatrix} c_0 \\ c \\ c_7 \end{pmatrix} = \begin{pmatrix} H_L^* - H_L \\ 0 \\ 0 \end{pmatrix}. \tag{6.2}$$

We stress that in this expression, H_S and c are really the matrix and vector defined in Eq.(6.1) and I is the 6×6 unit matrix. Also, several pieces of new notation have been introduced:

$$V_L^\dagger = \begin{pmatrix} \tau & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad V_R = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & \tau \end{pmatrix} \tag{6.3}$$

are coupling matrices connecting the system with the left and right leads, respectively. Finally,

$$g_L = (E - H_L)^{-1}, \quad g_R = (E - H_R)^{-1} \tag{6.4}$$

are the lead Green's functions.

Now, in order to compute the transmission coefficient t we need to find $c_7 = te^{7iqa}$. The phase factor is irrelevant as we are really interested only in the transmittance $T = |t|^2 = |c_7|^2$. We could, of course, simply invert the matrix equation Eq.(6.2). However, to prepare ourselves for tackling more elaborate cases we will demonstrate how the calculation can be simplified using a few mathematical manipulations. Primarily, the full Green's function of the entire device (leads plus system) is just the inverse of the matrix in Eq.(6.2), i.e.

$$\begin{pmatrix} g_L^{-1} & -V_L^\dagger & 0 \\ -V_L & EI - H_S & -V_R^\dagger \\ 0 & -V_R & g_R^{-1} \end{pmatrix} \cdot \begin{pmatrix} G_{11} & G_{12} & G_{13} \\ G_{21} & G_{22} & G_{23} \\ G_{31} & G_{32} & G_{33} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{6.5}$$

From Eq.(6.2), it follows that $c_7 = G_{31}(H_L^* - H_L)$ and, hence, G_{31} is the only important element of the full Green's function matrix. It can be computed by simple manipulations of the first column of equations in the matrix expression above. These are

$$\begin{gathered} g_L^{-1} G_{11} - V_L^\dagger G_{21} = 1 \\ -V_L G_{11} + (EI - H_S) G_{21} - V_R^\dagger G_{31} = 0 \\ -V_R G_{21} + g_R^{-1} G_{31} = 0. \end{gathered}$$

A few mathematical operations demonstrate that

$$G_{31} = g_R V_R G V_L g_L. \tag{6.6}$$

Here, G is the device Green's function

$$\begin{gathered} G = (EI - H_S - \Sigma_L - \Sigma_R)^{-1} \\ \Sigma_L = V_L g_L V_L^\dagger, \quad \Sigma_R = V_R^\dagger g_R V_R. \end{gathered} \tag{6.7}$$

The corrections $\Sigma_{L/R}$ are the so-called self-energies that incorporate the effect of coupling the systems to the leads. Thus, $T = \left| g_R V_R G V_L g_L (H_L^* - H_L) \right|^2$. We seek to find the transmission T as a function of the electron energy E. This means that we should express the lead Green's functions in terms of E as well. However, since $H_L = H_R = -\gamma e^{iqa}$ and $E = -2\gamma \cos(qa)$ it follows that

$$H_L = H_R = \frac{E}{2} + i\left(\gamma^2 - \frac{E^2}{4}\right)^{1/2}$$

$$\Downarrow$$

$$g_L^{-1} = g_R^{-1} = \frac{E}{2} - i\left(\gamma^2 - \frac{E^2}{4}\right)^{1/2}.$$

In the present example, the only non-vanishing elements of the self-energy matrices are $\Sigma_{L,11} = \Sigma_{R,66} = \tau^2 g_{L/R}$. Taking $\tau = -\gamma$, it can actually be demonstrated analytically that

$$T = \begin{cases} 4\gamma^2(4\gamma^2 - E^2)/(5\gamma^2 - E^2)^2 & |E| < 2\gamma \\ 0 & \text{otherwise.} \end{cases},$$

This result is plotted in Fig. 6.2. The maxima of unit transmittance are found at $E = \pm\sqrt{3}\gamma$. The effect of varying τ is illustrated in the right-hand panel. It is clearly seen that transmission becomes restricted to a narrow energy range around $\pm\gamma$, which coincides with the eigenstates of the isolated molecule. When the coupling increases, so does the broadening due to the self energy until, eventually, the transmittance is high in the entire energy range allowed by the leads.

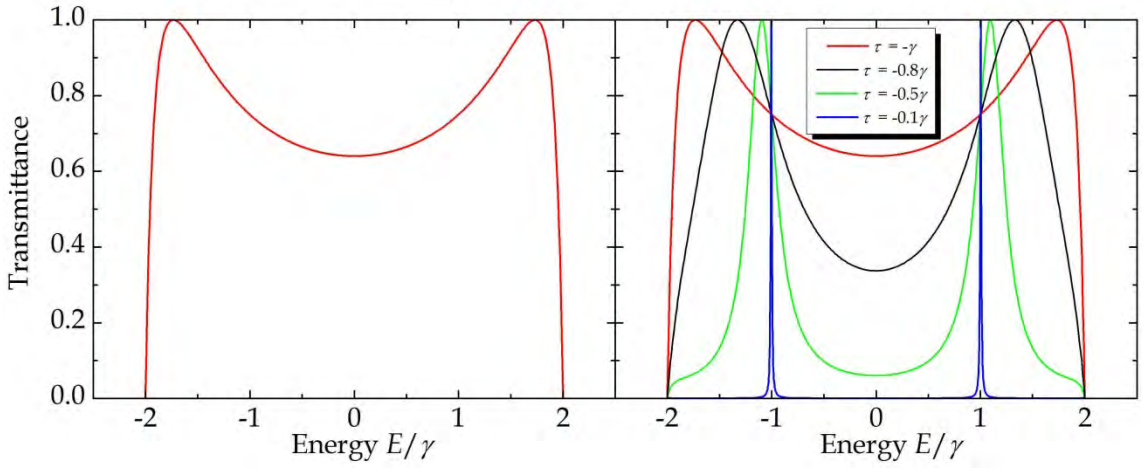


Figure 6.2. Transmission spectrum for the device illustrated in Fig. 6.1. In the right-hand panel, the lead-system coupling is varied.

6.1 General Landauer Formula

The expression for the transmittance appears rather asymmetrical in spite of the clear symmetry between left and right sides. Using some rewriting, however, the symmetry can be restored. First, since $H_L = H_R$ we find that

$$\begin{aligned} T &= (H_R - H_R^*)\left|g_R V_R G V_L g_L\right|^2 (H_L^* - H_L) \\ &= g_R^*(H_R - H_R^*)g_R \left|V_R G V_L\right|^2 g_L (H_L^* - H_L) g_L^* \\ &= -(g_R - g_R^*)\left|V_R G V_L\right|^2 (g_L - g_L^*). \end{aligned}$$

The last equality follows from the identity $g_R^*(H_R - H_R^*)g_R = g_R - g_R^*$ and similarly for the left lead. Now, bearing in mind that V_R, G and V_L are really matrices it follows that $\left(V_R G V_L\right)^* = V_L^\dagger G^\dagger V_R^\dagger$. Then, by judicious ordering of terms,

$$\begin{aligned} T &= -(g_R - g_R^*)V_R G V_L (g_L - g_L^*)V_L^\dagger G^\dagger V_R^\dagger \\ &= -(g_R - g_R^*)V_R G(\Sigma_L - \Sigma_L^\dagger)G^\dagger V_R^\dagger. \end{aligned}$$

Now, if we could just bring the last factor $V_R^\dagger$ to the front of the expression, the final result would be nice and symmetrical. This, though, cannot be right since $V_R^\dagger$ is a matrix and the final result should be a scalar. This can be remedied by a neat mathematical trick. We can take the *trace* (Tr) of the expression, i.e.

$$T = -\mathrm{Tr}\left\{(g_R - g_R^*)V_R G(\Sigma_L - \Sigma_L^\dagger)G^\dagger V_R^\dagger\right\}.$$

Taking the trace simply means summing the diagonal elements of the matrix in the argument. In our case, however, the argument is just a scalar and taking the trace doesn't do anything at all to the expression. But the trace has a very important property: it is cyclic, meaning that $\mathrm{Tr}\{ABC\} = \mathrm{Tr}\{CAB\} = \mathrm{Tr}\{BCA\}$ as can easily be verified. Thus, *under the trace* the terms can be rearranged so that

$$\begin{aligned} T &= -\mathrm{Tr}\left\{V_R^\dagger(g_R - g_R^*)V_R G(\Sigma_L - \Sigma_L^\dagger)G^\dagger\right\} \\ &= -\mathrm{Tr}\left\{(\Sigma_R - \Sigma_R^\dagger)G(\Sigma_L - \Sigma_L^\dagger)G^\dagger\right\}. \end{aligned}$$

In much of the literature [2], one introduces line width functions $\Gamma_{L/R} = i(\Sigma_{L/R} - \Sigma_{L/R}^\dagger)$ that are essentially the imaginary part of the self energy. Hence, we find the final symmetrical expression

$$T = \mathrm{Tr}\left\{\Gamma_R G \Gamma_L G^\dagger\right\}. \tag{6.8}$$

An additional advantage of the trace is that we automatically compute the entire transmission, i.e. $T = \sum_{\nu} T_{\nu \to \nu}$ summed over channels. This very handy and elegant Landauer expression can be taken as the starting point for most calculations of transport in molecular systems. It applies to asymmetric geometries, where left and right leads differ. It also applies to leads that are much more complicated that mono-atomic ones. In fact, it can even be generalized to devices that are periodic in the direction perpendicular to the transport.

Exercise: Transmission in biphenyl molecules.

The aim of this exercise is to calculate the electron transmittance through a biphenyl molecule as illustrated in Fig. 6.3. It is obviously quite similar to the benzene case discussed in the text.

a) Before specializing to biphenyl, show that Eq.(6.6) actually follows from Eq.(6.5).

b) Write down expressions for the matrices H_S, $V_L^{\dagger}$ and V_R.

c) Write a computer program to evaluate Eq.(6.8) assuming all hopping elements to be $-\gamma$. The result should look similar to the spectrum in Fig. 6.3.

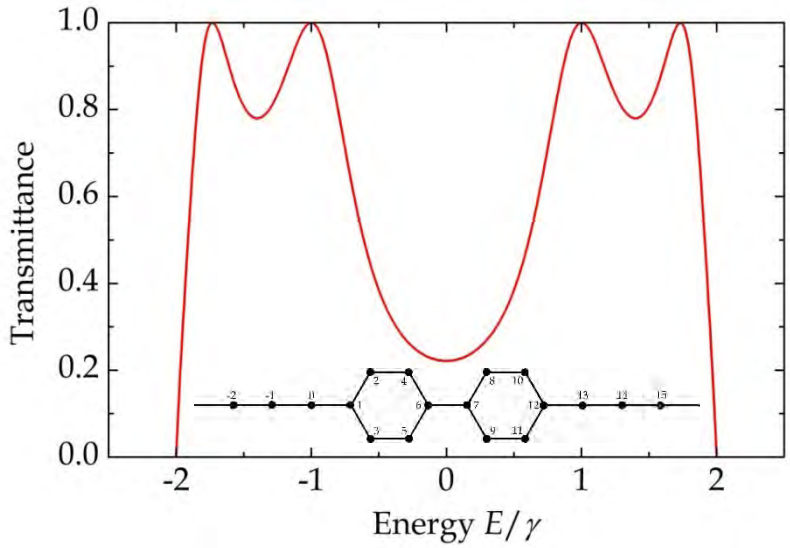


Figure 6.3. Transmission geometry and spectrum for a biphenyl molecule.

References

[1] E. Cuansing and J.S. Wang, Eur. Phys. J. B69, 505 (2009).
[2] S. Datta *Electronic Transport in Mesoscopic Systems* (Cambridge Univ. Press, Cambridge, 1997)

7. Electron Transmission in Molecules: Challenges

Based on the results of the previous chapter, we are now in a position to tackle more challenging examples of electronic transmission through molecules. Primarily, we wish to include more complicated leads with several atoms per unit cell. These leads may couple to the molecule through several bonds as well. Secondly, geometries that repeat periodically in the direction perpendicular to the transport will be studied. As a starting point, we will consider again the simple mono-atomic chain. This example can be analyzed analytically and used to illustrate the recursive approach introduced below. For this simple chain, the Hamiltonian is

$$H=-\begin{pmatrix} 0 & \gamma & 0 & 0 & \cdots \\ \gamma & 0 & \gamma & 0 & \\ 0 & \gamma & 0 & \gamma & \cdots \\ 0 & 0 & \gamma & 0 & \\ \vdots & & \vdots & & \ddots \end{pmatrix}. \tag{7.1}$$

Note that for notational simplicity we again suppress matrix symbols. The Green's matrix for this chain is then $G=(EI-H)^{-1}$. Due to the repeated structure of H we write

$$\begin{pmatrix} E & -V^{\dagger} \\ -V & EI-H \end{pmatrix}\cdot\begin{pmatrix} G_{11} & G_{12} \\ G_{12} & G_{22} \end{pmatrix}=\begin{pmatrix} 1 & 0 \\ 0 & I \end{pmatrix},\quad V^{\dagger}=\begin{pmatrix} -\gamma & 0 & 0 & \cdots \end{pmatrix}.$$

Solving the equations resulting from the first column we find

$$EG_{11}-V^{\dagger}(EI-H)^{-1}VG_{11}=1.$$

Now, $G=(EI-H)^{-1}$ and $V^{\dagger}GV=\gamma^2 G_{11}$ and so $EG_{11}-\gamma^2 G_{11}^2=1$ with the solution

$$G_{11}=\frac{E}{2\gamma^2}-\frac{i}{\gamma^2}\left(\gamma^2-\frac{E^2}{4}\right)^{1/2},$$

in complete agreement with the previous chapter. Our aim is to extend the method to more complicated one-dimensional leads. Hence, such general leads could have several, possibly different, atoms per unit cell. An example is the two-atomic lead in Fig. 7.1.

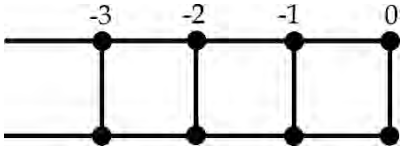


Figure 7.1. A two-atomic lead.

In this case, the wave function has components for both the upper (*u)* and lower (*l*) rows of atoms and for the *n*'th unit cell these can be collected as a vector $c_n = \left(c_n^{(u)}, c_n^{(l)}\right)^T$. In this manner, the chain Schrödinger equation can be written

$$\alpha c_{n-1} + \beta c_{n+1} + h c_n = E c_n,$$

where we introduce matrices

$$\alpha = \beta = -\begin{pmatrix} \gamma & 0 \\ 0 & \gamma \end{pmatrix}, \quad h = -\begin{pmatrix} 0 & \gamma \\ \gamma & 0 \end{pmatrix}.$$

The reason for keeping separate symbols α and β for the couplings left and right, respectively, is that in general these might be different. The Hamiltonian is now

$$H = \begin{pmatrix} h & \beta & 0 & 0 & \cdots \\ \alpha & h & \beta & 0 & \\ 0 & \alpha & h & \beta & \cdots \\ 0 & 0 & \alpha & h & \\ \vdots & & \vdots & & \ddots \end{pmatrix}. \tag{7.2}$$

Similarly to the previous chapter, we consider the first column of equations derived from the defining equation for the Green's function $(EI - H)G = 1$:

$$\begin{pmatrix} E-h & -\beta & 0 & 0 & \cdots \\ -\alpha & E-h & -\beta & 0 & \\ 0 & -\alpha & E-h & -\beta & \cdots \\ 0 & 0 & -\alpha & E-h & \\ \vdots & & \vdots & & \ddots \end{pmatrix} \cdot \begin{pmatrix} G_{00} \\ G_{10} \\ G_{20} \\ G_{30} \\ \vdots \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ \vdots \end{pmatrix}. \tag{7.3}$$

Consider the relation obtained from the second row

$$-\alpha G_{00} + (E-h)G_{10} - \beta G_{20} = 0 \quad \Rightarrow \quad G_{10} = g\left(\alpha G_{00} + \beta G_{20}\right), \quad g = (E-h)^{-1}.$$

When combined with the first row it follows that

$$(E-h-\beta g\alpha)G_{00}-\beta g\beta G_{20}=1\,.$$

The resulting system of equations can now be formulated as

$$\begin{pmatrix} E-h-\beta g\alpha & -\beta g\beta & 0 & 0 & \cdots \\ -\alpha g\alpha & E-h-\alpha g\beta-\beta g\alpha & -\beta g\beta & 0 & \cdots \\ 0 & -\alpha g\alpha & E-h-\alpha g\beta-\beta g\alpha & -\beta g\beta & \cdots \\ 0 & 0 & -\alpha g\alpha & E-h-\alpha g\beta-\beta g\alpha & \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}\cdot\begin{pmatrix} G_{00} \\ G_{20} \\ G_{40} \\ G_{60} \\ \vdots \end{pmatrix}=\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ \vdots \end{pmatrix}.$$

As is apparent, the 0'th cell now couples to the 2'nd, 4'th and so on. If iterated once more, coupling will be to cells number 4, 8, 12 etc. This recursive scheme [1] can be formulated as the following repeated sequence, initialized by setting $\lambda=\lambda_L=\lambda_R=h$:

$$\left.\begin{aligned} g &:= (E-\lambda)^{-1} \\ \lambda &:= \lambda+\alpha g\beta+\beta g\alpha \\ \lambda_L &:= \lambda_L+\beta g\alpha \\ \lambda_R &:= \lambda_R+\alpha g\beta \\ \alpha &:= \alpha g\alpha \\ \beta &:= \beta g\beta \end{aligned}\right\}\begin{matrix}\text{loop} \\ \text{until} \\ \text{convergence}\end{matrix}\quad.$$

After convergence, the left and right surface Green's functions of the leads are given by $g_L=(E-\lambda_L)^{-1}$ and $g_R=(E-\lambda_R)^{-1}$, respectively. For numerical stability, it's necessary to add a small imaginary part $i\eta$ and typically we choose $\eta\approx\gamma/1000$.

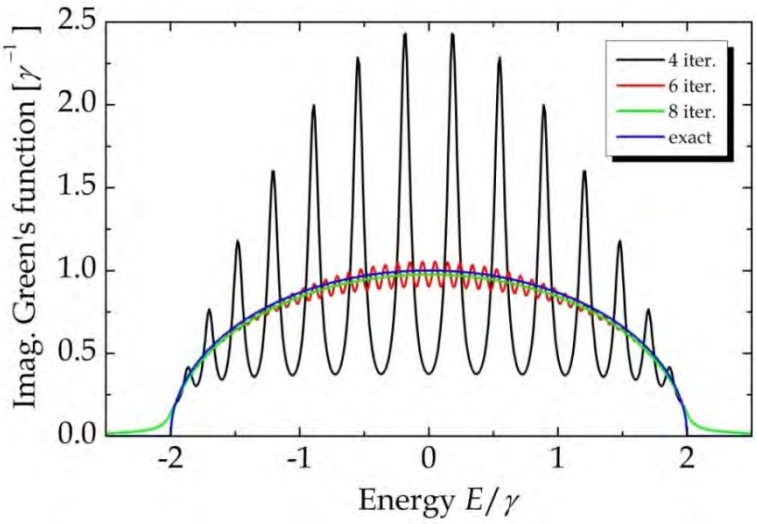


Figure 7.2. Comparison of the numerically generated mono-atomic lead Green's functions at different stages in the iteration process. The blue curve is the exact result.

As an illustration, the result for the simple mono-atomic chain is shown in Fig. 7.2. It is seen that 8 iterations produce a result very close to the exact one apart from the broadening introduced by the imaginary term $i\eta$.

For the two-atomic lead, Fig. 7.1, the Green's function is a 2×2 matrix G determined by the condition

$$(EI - h)G - \alpha G \beta G = I.$$

This equation, in fact, has an analytical solution. It turns out that $G_{22} = G_{11}$ and $G_{21} = G_{12}$. Isolating, it can be shown that

$$4\gamma^4 G_{11}^4 - 8E\gamma^2 G_{11}^3 + (5E^2 + 3\gamma^2)G_{11}^2 - E(E^2/\gamma^2 + 3)G_{11} + E^2/\gamma^2 = 0.$$

Despite appearances, this equation has a relatively simple solution and the same goes for G_{12}. The results are plotted in Fig. 7.3. The spectral features are found at $\pm 3\gamma$ and $\pm\gamma$. These features correspond to the eigenmodes of the two-atomic lead. It is easily shown that the two modes are $\pm\gamma - 2\gamma\cos(ka)$. Hence, the energy ranges of the modes are $-3\gamma \le E \le \gamma$ and $-\gamma \le E \le 3\gamma$, respectively. In the central range $-\gamma \le E \le \gamma$, the two bands overlap, which leads to the addition bump in G_{11}. These exact results can be used to check the numerical routine above in a simple example.

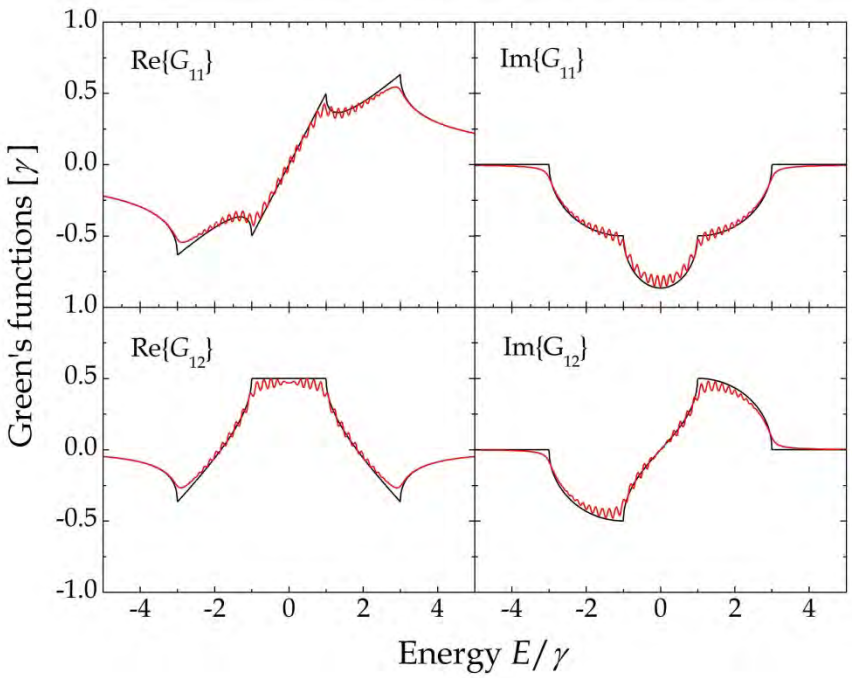


Figure 7.3. Comparison of the numerically generated two-atomic lead Green's functions after 5 iterations with the exact result.

In order to apply the two-atomic leads in an actual transport calculation, we imagine that such leads are attached to the left and right of a very simple system, as shown in Fig. 7.4, top panel.

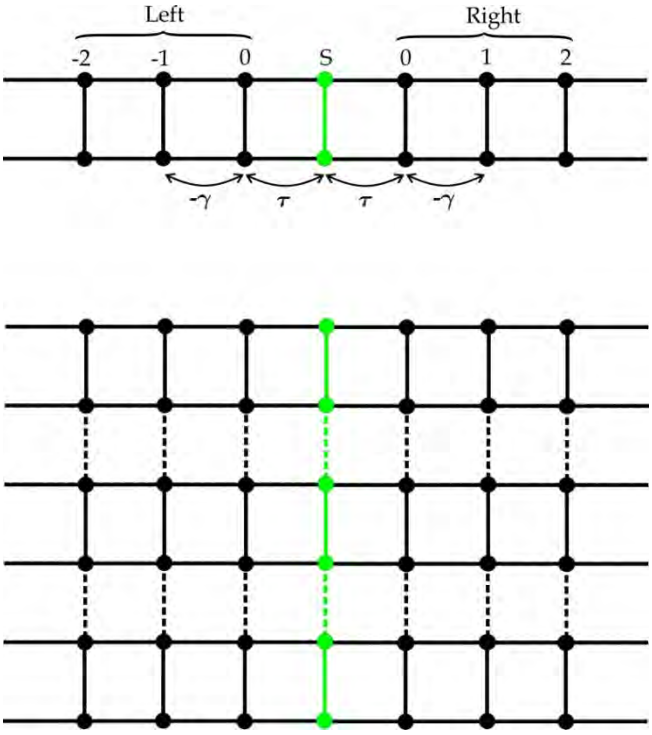


Figure 7.4. Two-atomic leads attached to a system represented by the green atoms. Some of the hopping integrals are indicated. Both an isolated 1D device (top) and periodic 2D device (bottom) is shown.

This "system" is also two atoms wide and the only difference compared to the lead atoms is that we let the device atoms be of a different sort. This implies two things: First the system Hamiltonian will be taken as

$$H_S = \begin{pmatrix} \varepsilon & -\gamma \\ -\gamma & \varepsilon \end{pmatrix},$$

where ε is the on-site potential that describes the energy of an electron on an isolated "system" atom relative to the lead atoms. Here, we also assumed that the hopping integral within the system is the same as for the leads i.e. $-\gamma$. Secondly, the coupling between leads and systems should be considered. Again, their structure is similar to the coupling matrices of the leads but if we allow for an arbitrary coupling strength we can write them as

$$V_L = V_R = \begin{pmatrix} \tau & 0 \\ 0 & \tau \end{pmatrix},$$

where τ is the lead-system coupling. The system on-site potential ε could also be the result of a local electrostatic gate, which would shift the energy of the system atoms relative to the leads. The transmittance of this simple barrier device is illustrated in Fig. 7.5. We see some expected trends: When the coupling between leads and system is reduced, transmittance drops and becomes highly peaked around $\pm\gamma$. Similarly, when the barrier potential ε is raised the transmittance decreases in a symmetric fashion

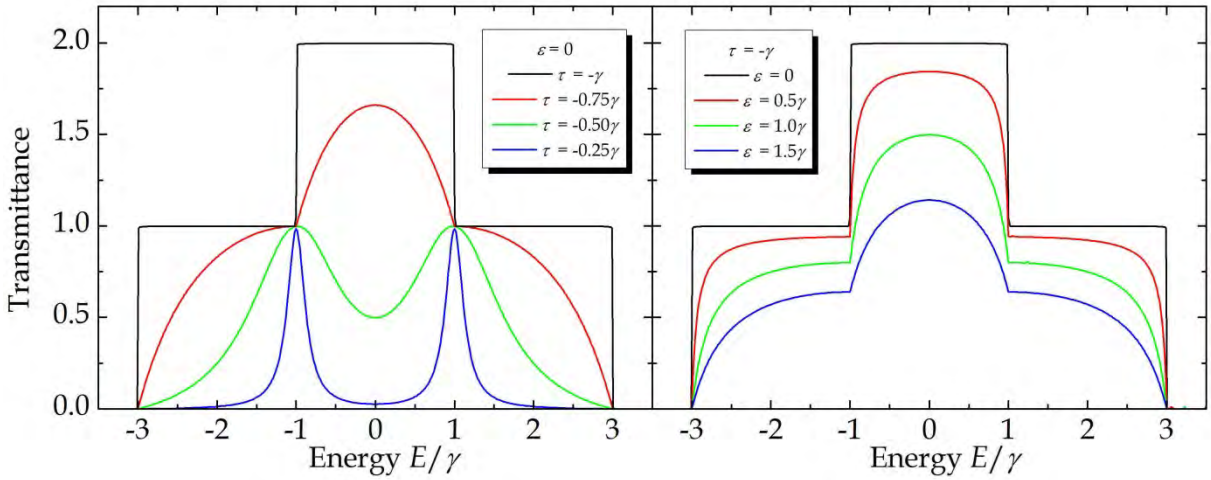


Figure 7.5. Transmittance of the isolated two-atomic lead device. In the left and right panel, the system-lead coupling and system on-site potential is varied, respectively.

7.1 Two-Dimensional Leads

So far, we have focused on one-dimensional geometries, where a small number of modes contribute to the transport. While such geometries are realistic representations of certain experimental systems, a more common case is that of much wider leads and devices. In such cases, it is more appropriate to view the system as an infinitely wide one. Moreover, the system can typically be broken down to a certain unit that is then repeated periodically along the dimension perpendicular to the transport direction. An example of such a geometry is shown in Fig. 7.4, bottom panel. We see that the system is essentially identical to the ones we have been considering so far except for the fact that periodic boundary conditions are used to couple neighboring units. As usual, whenever we encounter periodic boundary conditions, the wave functions become periodic apart from a Bloch phase factor. Hence, if the width of a single unit is w, the Bloch factor is $\exp(\pm ikw)$, where k is the wave vector restricted to the range $-\pi / w < k < \pi / w$ and the sign depends on whether coupling is upwards or downwards.

As long as we only consider nearest-neighbor coupling, only the system and lead Hamiltonians H_S and h are affected. The coupling matrices are not affected because all new coupling interactions are at least a factor of $\sqrt{2}$ further apart than nearest-neighbors. Thus, adding periodic boundaries simply means that we should use

$$H_S = \begin{pmatrix} \varepsilon & -\gamma(1+e^{ikw}) \\ -\gamma(1+e^{-ikw}) & \varepsilon \end{pmatrix}, \quad h = -\gamma \begin{pmatrix} 0 & 1+e^{ikw} \\ 1+e^{-ikw} & 0 \end{pmatrix}.$$

In practice, we do the averaging over *k*-vectors by running simulation for a range of discrete values and then simply divide the sum by the number of values N_k, i.e.

$$\langle T \rangle = \frac{1}{N_k} \sum_k T(k).$$

In Fig. 7.6, this average has been made for the structure in Fig. 7.4 (bottom) using a coarse and fine *k*-grid. In the periodic case, since each atom couples to 4 nearest neighbors, the full energy range of the leads is $|E| \leq 4\gamma$. Hence, transmission is non-zero in this range. We also see that discretization errors are washed out when using the fine grid.

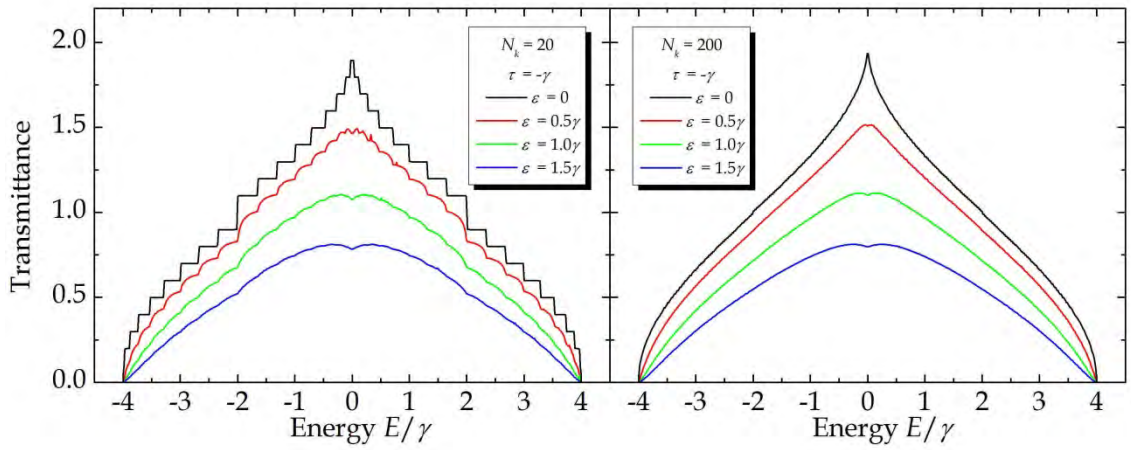


Figure 7.6. Transmittance averaged over transverse k-points using a coarse (left) and fine (right) grid.

Exercise: Transmission in graphene sheets.

In this exercise, we consider the graphene sheet shown in Fig. 7.7. As shown, it can be constructed as "armchair" chains joined periodically. We therefore model the sheet as a collection of coupled chains.

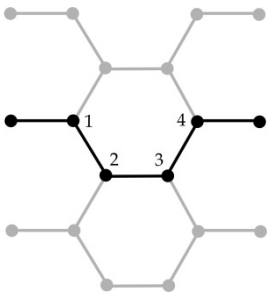


Figure 7.7. Graphene sheet built from coupled "armchair" chains.

a) Show that the lead Hamiltonian including coupling is

$$h = -\gamma \begin{pmatrix} 0 & 1+e^{ikw} & 0 & 0 \\ 1+e^{-ikw} & 0 & 1 & 0 \\ 0 & 1 & 0 & 1+e^{-ikw} \\ 0 & 0 & 1+e^{ikw} & 0 \end{pmatrix}.$$

Similarly, the only non-vanishing entries of the coupling matrices are $\alpha_{14} = \beta_{41} = -\gamma$.

b) Use the recursive scheme to construct the left and right surface Green's functions. Tip: since the *k*-points cover the interval $-\pi / w < k < \pi / w$ we can just use the full phase kw as a variable covering the range $-\pi < kw < \pi$ without worrying about the value of w.

c) Compute the *k*-averaged transmittance of the sheet. Take the system to be four central atoms shifted by an on-site potential ε and all coupling integrals to be $-\gamma$. The result should resemble the plot below.

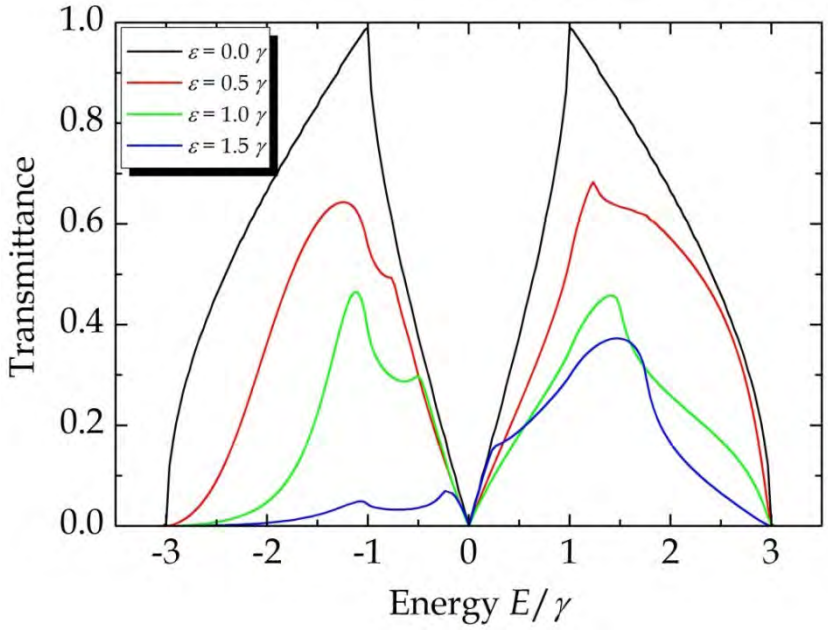


Figure 7.8. Transmittance in graphene sheet averaged over 500 transverse k-points.

References

[1] T. Markussen, R.Rurali, M. Brandbyge, and A.-P. Jauho, Phys. Rev. B74, 245313 (2006).

8. Electric Properties of Semiconductors

We now turn from metals to semiconductors. Semiconductors like Si and GaAs are the active materials in electronic and optoelectronic devices such as transistors and semiconductor lasers. While traditional applications of semiconductors use bulk materials, many recent devices rely heavily on quantum confined structures, in particular quantum wells. Also, quantum wires and dots are emerging in applications such as fluorescent nanoparticles. Finally, two thirds of all possible carbon nanotubes are semiconductors and constitute extreme examples of one-dimensional quantum structures. In the present chapter, we will look at doping in semiconductors of various dimensions in order to describe the electric conductivity.

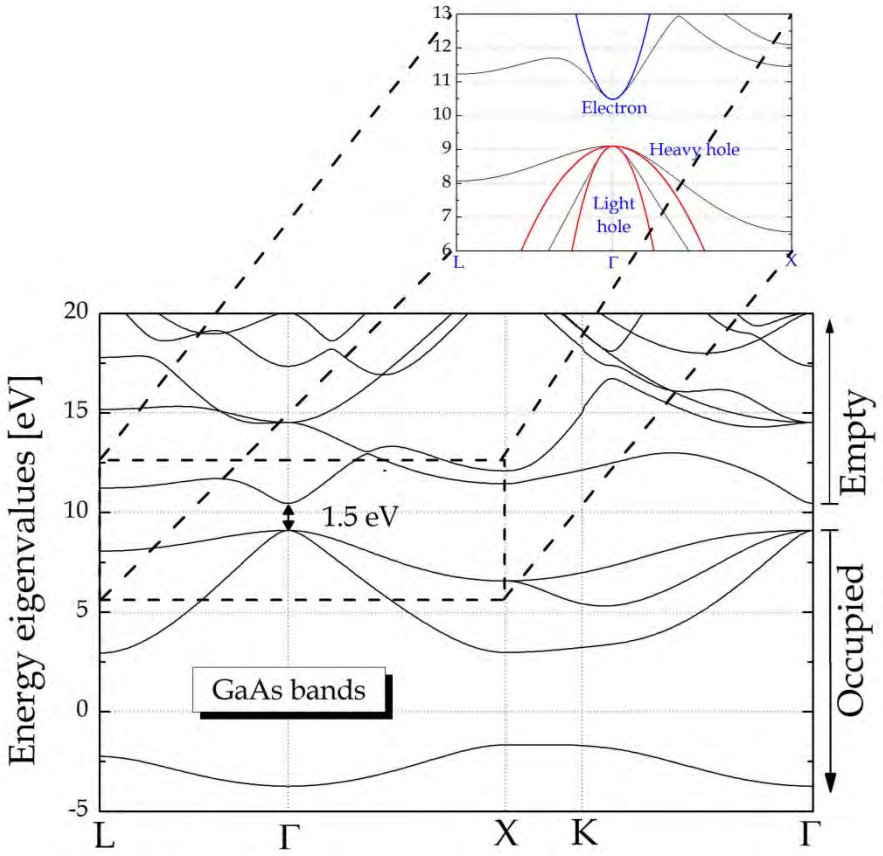


Figure 8.1. Band structure of GaAs with a zoom of the bands in the vicinity of the band gap and parabolic approximations to the bands added.

In Fig. 8.1, the band structure of the prototypical semiconductor GaAs is displayed. In the undoped case and at zero temperature, the four lower bands are completely occupied while the remaining are empty. Consequently, there is no conduction because occupied and empty states are separated by a large energy gap (roughly 1.5 eV), which is far more than a normal electric potential can surmount. However, two factors can change that: doping and temperature. Hence, it is the aim here to describe the effects of these factors on the conductivity. The mechanism of conduction can be understood on the basis of Eq.(3.11):

$$\sigma(\omega) = \frac{i\varepsilon_0 \omega_p^2}{\omega + i\Gamma}. \tag{8.1}$$

In deriving the expression for the plasma frequency in Chapter 3, we took the temperature to zero, which is perfectly alright for a metal. In the semiconducting case, this is no longer justified and we should use the more general expression

$$\omega_p^2 = -\frac{2e^2}{\hbar^2 \varepsilon_0 \Omega} \sum_{\vec{k}} \left(\frac{\partial E_{c\vec{k}}}{\partial k_z} \right)^2 \frac{\partial f(E_{c\vec{k}})}{\partial E_{c\vec{k}}}$$

instead. As a starting point, we consider the bulk case. Here, the sum over $\vec{k}$ is readily converted into an integral and we get

$$\omega_p^2 = -\frac{2e^2}{\hbar^2 \varepsilon_0 (2\pi)^3} \int \left(\frac{\partial E_{c\vec{k}}}{\partial k_z} \right)^2 \frac{\partial f(E_{c\vec{k}})}{\partial E_{c\vec{k}}} d^3k \,.$$

We next apply the chain rule

$$\frac{\partial f_{c\vec{k}}}{\partial k_z} = \frac{\partial E_{c\vec{k}}}{\partial k_z} \frac{\partial f_{c\vec{k}}}{\partial E_{c\vec{k}}}$$

to demonstrate that

$$\begin{aligned} \omega_p^2 &= -\frac{2e^2}{\hbar^2 \varepsilon_0 (2\pi)^3} \int \frac{\partial E_{c\vec{k}}}{\partial k_z} \frac{\partial f(E_{c\vec{k}})}{\partial k_z} d^3k \\ &= \frac{2e^2}{\hbar^2 \varepsilon_0 (2\pi)^3} \int \frac{\partial^2 E_{c\vec{k}}}{\partial k_z^2} f(E_{c\vec{k}})\, d^3k. \end{aligned}$$

The last equality follows by partial integration. For simplify, we will focus the attention on the conduction band. Hence, what we describe is the change in conductivity due to added electrons in this otherwise empty band. This phenomenon is known as *n*-doping. In general, an additional contribution will come from removed electrons or "holes" in the valence bands. This contribution can be described using a similar approach and added if needed. The bands generally have a complicated dependence on $\vec{k}$. However, the relatively few electrons that are added to the conduction band will occupy states very close to the bottom of the band as those are the lowest in energy. Similarly, the few electrons removed from the valence band will be taken from the very top of the band. Hence, as only a small portion of the bands near the extrema matter, we can use the parabolic approximation, which reads for the conduction band

$$E_{c\vec{k}} \approx E_c + \frac{\hbar^2 k^2}{2m_e},$$

where E_c denotes the energy at the bottom of the band and m_e is the effective electron mass deduced from the curvature of the band. When applied in the expression above it follows immediately that

$$\begin{aligned} \omega_p^2 &= \frac{2e^2}{\varepsilon_0 m_e (2\pi)^3} \int f(E_{c\vec{k}})\, d^3k \\ &= \frac{e^2 n}{\varepsilon_0 m_e}, \end{aligned} \tag{8.2}$$

where $n = 2(2\pi)^{-3} \int f(E_{c\vec{k}})\, d^3k$ is the electron density. We notice that this is precisely as for the free-electron case except that the effective electron mass has replaced the free-electron mass. As usual, we may alternatively express the electron density in terms of the density of states, i.e.

$$n = \int f(E) D(E) dE\,. \tag{8.3}$$

Now, we should also worry about confinement effects. First, the sum over $\vec{k}$ should be replaced by a general summation over all states. When we speak of a D-dimensional material (D = 0, 1, 2 or 3), we simply mean that the structure is "large" in D out of the 3 possible spatial directions. In the remaining directions, the structure is small enough for quantum effects to arise. Thus, for the quantized directions the sum over k should be replaced by a sum over an index m labeling the discrete eigenstates. The total energy in a particular state now becomes $E = \hbar^2 k^2 / 2m_e + E_m^e$, where k is the magnitude of the D-dimensional wave vector and $E_m^e > 0$ is the quantization energy of the electron. The consequence of these changes is that Eq.(8.3) should be replaced by

$$n_D = \int f(E) D_D(E) dE\,,$$

where $D_D(E)$ is the D-dimensional density of states. Equation (8.1) with $\omega_p^2 = e^2 n_D / \varepsilon_0 m_e$ then still applies for the conductivity along one of the extended directions. The density of states expressions are derived in Appendix 1 and given by

$$D_D(E) = \begin{cases} \dfrac{\sqrt{2}m_e^{3/2}}{\pi^2\hbar^3}\sqrt{E-E_c}\,\theta(E-E_c) & D=3 \\ \dfrac{m_e}{d\pi\hbar^2}\displaystyle\sum_m \theta(E-E_c-E_m^e) & D=2 \\ \dfrac{\sqrt{2}m_e^{1/2}}{A\pi\hbar}\displaystyle\sum_m \dfrac{\theta(E-E_c-E_m^e)}{\sqrt{E-E_c-E_m^e}} & D=1 \\ \dfrac{2}{\Omega}\displaystyle\sum_m \delta(E-E_c-E_m^e) & D=0. \end{cases} \tag{8.4}$$

Here, the width of the 2D system (quantum well) is denoted d, the cross sectional area of the 1D system (quantum wire) is denoted A and the volume of the 0D system (quantum dot) is Ω. Using the Fermi function we can now write

$$n_D = \int \frac{D_D(E)}{\exp\{(E-E_F)/kT\}+1}dE.$$

Here, we follow the usual convention in solid state physics and denote the temperature dependent chemical potential by E_F. For the three cases with at least one extended dimension, we find

$$n_D = \begin{cases} -2\left(\dfrac{m_e kT}{2\pi\hbar^2}\right)^{3/2}\mathrm{Li}_{3/2}\left(-\exp\{(E_F-E_c)/kT\}\right) & D=3 \\ \dfrac{m_e kT}{d\pi\hbar^2}\displaystyle\sum_m \ln\left(1+\exp\{(E_F-E_c-E_m^e)/kT\}\right) & D=2 \\ -\dfrac{1}{A}\left(\dfrac{2m_e kT}{\pi\hbar^2}\right)^{1/2}\displaystyle\sum_m \mathrm{Li}_{1/2}\left(-\exp\{(E_F-E_c-E_m^e)/kT\}\right) & D=1. \end{cases} \tag{8.5}$$

In these expressions, Li_p denotes the p'th polylogaritm defined as

$$\mathrm{Li}_p(x) \equiv \frac{x}{\Gamma(p)}\int_0^\infty \frac{t^{p-1}}{e^t-x}dt, \quad p>0,$$

where $\Gamma(p)$ is the gamma function. In most situations, E_F lies several kT below E_c and, hence, we may expand the results above using

$$\mathrm{Li}_p(x) \approx x, \quad \ln(1+x) \approx x, \quad |x| \ll 1.$$

In this case, we find

$$n_D \approx \begin{cases} 2\left(\dfrac{m_e kT}{2\pi\hbar^2}\right)^{3/2} \exp\{(E_F - E_c)/kT\} & D=3 \\ \dfrac{m_e kT}{d\pi\hbar^2}\displaystyle\sum_m \exp\{(E_F - E_c - E_m^e)/kT\} & D=2 \\ \dfrac{1}{A}\left(\dfrac{2m_e kT}{\pi\hbar^2}\right)^{1/2}\displaystyle\sum_m \exp\{(E_F - E_c - E_m^e)/kT\} & D=1. \end{cases} \tag{8.6}$$

In particular, these formulas clearly show that provided $E_F < E_c$ we have $n_D \approx 0$ at low temperatures. To add electrons, we need to introduce doping.

8.1 Doping

In order to add electrons to the conduction band, we might introduce some suitable electron donating atoms (donor impurities) to the semiconductor. To function as efficient donors, the added atoms should be characterized by an energy level just slightly below the conduction band edge. Such donors are called shallow donors. Thus, when the electron resides on the parent atom it occupies this shallow energy level. This situation is illustrated in Fig. 8.2.

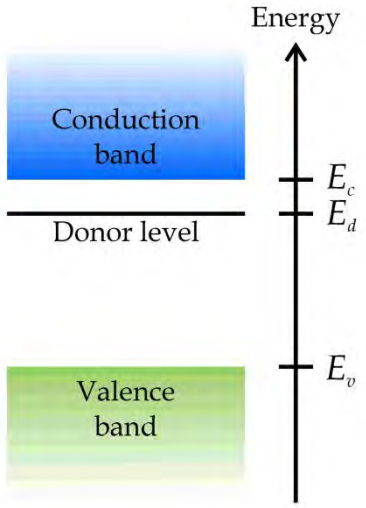


Figure 8.2. Illustration of energy levels in a semiconductor doped with donor impurities.

The occupation of the donor level is a standard exercise in statistical mechanics. There are four relevant states distinguished by the number and nature of the particles occupying the state: (1) zero electrons, (2) one spin-down electron, (3) one spin-up electron and (4) one spin-down plus one spin-up electron. No more than two electrons can occupy the state due to the exclusion principle. The energy of a single (spin-up or spin-down) electron in the state is denoted E_d. Similarly, the energy of two electrons occupying the state is denoted E_{dd}. The partition function is therefore

$$Z = 1 + 2\exp\{(E_F - E_d)/kT\} + \exp\{(2E_F - E_{dd})/kT\}.$$

We then have the mean occupancy

$$\langle n \rangle = \frac{2\exp\{(E_F - E_d)/kT\} + 2\exp\{(2E_F - E_{dd})/kT\}}{1 + 2\exp\{(E_F - E_d)/kT\} + \exp\{(2E_F - E_{dd})/kT\}}.$$

Now, it's important to realize that $E_{dd} \neq 2E_d$ and, in fact, normally $E_{dd} \gg 2E_d$. The reason is that two electrons localized on the same atom will lead to a substantial Coulomb repulsion energy. Thus, we can in fact neglect this possibility and so

$$\begin{aligned} \langle n \rangle &\approx \frac{2\exp\{(E_F - E_d)/kT\}}{1 + 2\exp\{(E_F - E_d)/kT\}} \\ &= \frac{2}{\exp\{(E_d - E_F)/kT\} + 2}. \end{aligned} \tag{8.7}$$

Next, we want to set up a balance for the number of charges. Equation (8.6) gives us the number of electrons in the conduction band and Eq.(8.7) describes the number of electrons in the donor level. In addition, we need the number of electrons removed from the valence band or, equivalently, the number of holes put into the valence band. In the simplest case of a single valence band, the hole density p_D is given by an expression identical to Eq.(8.6) except for the following changes: (1) the effective electron mass m_e is replaced by the effective hole mass m_h, (2) the conduction band edge E_c is replaced by the valence band top E_v, (3) the sign of all energies is reversed, and (4) the electron quantization energies E_m^e are replaced by the hole equivalent E_m^h. In this manner,

$$p_D \approx \begin{cases} 2\left(\dfrac{m_h kT}{2\pi\hbar^2}\right)^{3/2} \exp\{(E_v - E_F)/kT\} & D = 3 \\ \dfrac{m_h kT}{d\pi\hbar^2} \displaystyle\sum_m \exp\{(E_v + E_m^h - E_F)/kT\} & D = 2 \\ \dfrac{1}{A}\left(\dfrac{2m_h kT}{\pi\hbar^2}\right)^{1/2} \displaystyle\sum_m \exp\{(E_v + E_m^h - E_F)/kT\} & D = 1. \end{cases}$$

In analogy with the electron case, it is required that the Fermi energy E_F lies well above E_v for these expressions to be correct. We will denote the total density of impurity atoms by N_I. Thus, when multiplied by Eq.(8.7) this gives the total density of electrons residing in the donor level. It follows that the grand total electron density is $n_D + \langle n \rangle N_I$. Similarly, we should count the positive charges. These come from two

sources: the holes in the valence band and the positive impurity ions, whose density is N_I. Hence, overall charge neutrality requires that

$$n_D + \langle n \rangle N_I = p_D + N_I .$$

To solve for the Fermi energy we introduce the notation $n_D = N_D \exp\{(E_F - E_c)/kT\}$, where

$$N_3 = 2\left(\frac{m_e kT}{2\pi\hbar^2}\right)^{3/2}, \quad N_2 = \frac{m_e kT}{d\pi\hbar^2}\sum_m \exp\{-E_m^e/kT\}, \quad N_1 = \frac{1}{A}\left(\frac{2m_e kT}{\pi\hbar^2}\right)^{1/2}\sum_m \exp\{-E_m^e/kT\}.$$

Similarly, we write $p_D = P_D \exp\{(E_v - E_F)/kT\}$ with

$$P_3 = 2\left(\frac{m_h kT}{2\pi\hbar^2}\right)^{3/2}, \quad P_2 = \frac{m_h kT}{d\pi\hbar^2}\sum_m \exp\{E_m^h/kT\}, \quad P_1 = \frac{1}{A}\left(\frac{2m_h kT}{\pi\hbar^2}\right)^{1/2}\sum_m \exp\{E_m^h/kT\}.$$

Rather than actually solving for E_F, we will introduce $x \equiv \exp\{E_F/kT\}$ and solve for x. Hence, putting all the pieces together, we can reformulate the charge balance as

$$N_D x \exp\{-E_c/kT\} = P_D \frac{1}{x}\exp\{E_v/kT\} + N_I\left[1 - \frac{2}{\exp\{E_d/kT\}/x + 2}\right]$$

$$\Downarrow$$

$$\left(N_D x^2 - P_D \exp\{(E_c + E_v)/kT\}\right)\left(\exp\{E_d/kT\} + 2x\right) = N_I x \exp\{(E_c + E_d)/kT\}.$$

To simplify even further, we introduce $x_{cv} \equiv \exp\{(E_c + E_v)/2kT\}$, $x_{cd} \equiv \exp\{(E_c + E_d)/2kT\}$ and finally $x_d \equiv \exp\{E_d/kT\}$ so that

$$\left(N_D x^2 - P_D x_{cv}^2\right)\left(x_d + 2x\right) = N_I x x_{cd}^2 .$$

The solution to this seemingly innocent equation is

$$x = \sqrt[3]{R + \sqrt{Q^3 + R^2}} + \sqrt[3]{R - \sqrt{Q^3 + R^2}} - \frac{x_d}{6}$$

$$R = -x_d\frac{N_D x_d^2 + 9N_I x_{cd}^2 - 36P_D x_{cv}^2}{216N_D}, \quad Q = -\frac{N_D x_d^2 + 6N_I x_{cd}^2 + 12P_D x_{cv}^2}{36N_D}.$$

Once x is calculated, the electron density in the conduction band is found from $n_D = N_D x \exp\{-E_c/kT\}$. To illustrate the results, we take the following parameters appropriate for moderately doped GaAs:

- $m_e = 0.066\, m$, $m_h = 0.5\, m$ (heavy hole)
- $E_v = 0$, $E_c = 1.5$ eV, $E_d = E_c - 0.1$ eV
- $N_I = 10^{21}$ m^{-3}
- Quantum well ($D = 2$): $d = 2$ nm, $E_1^e = 0.15$ eV, $E_1^h = 0.02$ eV
- Quantum wire ($D = 1$): $A = 4$ nm^2, $E_1^e = 0.30$ eV, $E_1^h = 0.04$ eV

We only include the lowest quantized states for wells and wires as the higher states will be virtually empty. For this set of parameters, the Fermi energy and corresponding electron density vary with temperature as illustrates in Fig. 8.3. At low temperatures, the Fermi energy is pinned at the midpoint between the donor level and the effective conduction band edge, i.e. E_c for D = 3 and $E_c + E_1^e$ for D = 1 or 2. Physically, this corresponds to the situation in which all electrons reside on the donor atoms. As the temperature is raised, the Fermi level drops below the donor level and electrons are transferred into the conduction band. Eventually, most of the donor atoms have given up their electron to the conduction band and a plateau at $n_D \approx N_I = 10^{21}$ m^{-3} is reached for the electron density. This temperature range is called the *saturation range* whereas the low temperature range in which the electron density increases exponentially is known as the *freeze out range*. If more electrons should be added to the conduction band, they must be taken from the valence band, which is only feasible at rather high temperatures about 800 K as seen in the graph.

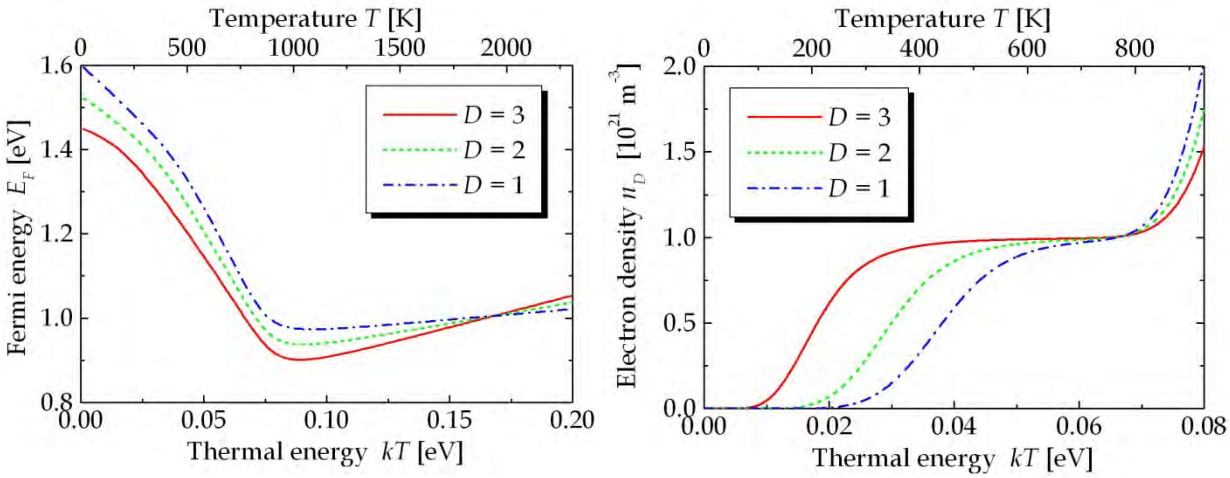


Figure 8.3. Fermi energy and electron density vs. temperature for *n* doped GaAs in 1,2 and 3 dimensional structures.

It is noted, that at relatively low temperatures the electron density in the quantized structures is less than for the bulk. This is mainly due to the increased separation between the donor level and the effective conduction band edge due to the confinement energy. However, at high temperatures (above the saturation range) the trend is reversed. The reason is that the 3*D* Fermi level increases steeply above the saturation range and much less so for the 2*D* and 1*D* cases. The origin of this

difference is the way the prefactors in the electron density Eq.(8.6) vary with temperature, i.e. the $T^{D/2}$ behavior of the factors in front of the exponentials.

Exercise: Properties of *D*-dimensional semiconductors.

Consider a material with a simple parabolic conduction band.

a) Show that the density of states expressions in Eq.(8.4) are correct.

Now, assume that we're dealing with an un-doped semiconductor, i.e. take $N_I = 0$.

b) Provided only the lowest quantized electron and hole levels are included, show that

$$E_F = \frac{1}{2}\left(E_c + E_1^e + E_v + E_1^h\right) + \frac{D}{4} kT \ln\left(\frac{m_h}{m_e}\right).$$

9. PN and PIN Junctions

In this chapter, we turn to some basic but important electronic applications of semiconductors and investigate the peculiar features that appear in the nanoscale regime. The basic element of (bipolar junction) transistors is the PN junction, which itself acts as a current rectifier and, hence, is an important device in its own right. We therefore discuss this device in detail below. As the name says, a *pn junction diode* consists of a junction between a *p*-doped and an *n*-doped semiconductor. In the interface region, an electric field arises due to unbalanced charges. The field exists in a narrow region called the *depletion layer*, which forms a barrier for the charge carriers. Here, we analyze the junction in order to compute the electrostatic potential corresponding to the field. Subsequently, we look at a related structure, i.e. the pin junction, in which an intrinsic layer is inserted between the doped regions.

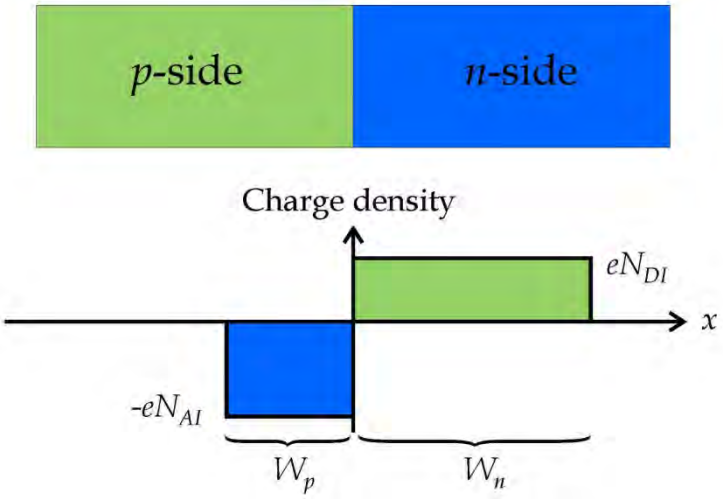


Figure 9.1. Illustration of the pn junction. The bottom part shows the space charge density.

We restrict the analysis to three dimensional structures so that carrier concentrations, electric fields etc. vary only along x. Our starting point is the pn junction illustrated in Fig. 9.1. In the last chapter, we discussed *n*-doped materials, in which a large concentration of donor impurities supplied the extra electrons. Similarly, *p*-doped materials contain a large concentration of acceptor impurities supplying additional holes to the conduction band. To distinguish between the two types of impurities, we denote their concentrations by N_{DI} and N_{AI}, respectively. The basic physics of the pn junction is rather simple: On the *n*-side there is a surplus of mobile electrons and on the *p*-side there is a surplus of mobile holes. A certain number of electrons will therefore diffuse from the *n*- into the *p*-side and vice versa for the holes. We now make the following simplifying assumption: The electrons missing from the *n*-side are taken from a layer of width W_n, which is completely depleted of free electrons. Similarly, the holes are taken from a depleted layer of width W_p on the *p*-side. We saw in the previous chapter that for temperatures in the depletion range, the electron

concentration in the *n*-doped material is approximately equal to the donor concentration $n \approx N_{DI}$. In analogy, we'll have $p \approx N_{AI}$ for a bulk *p*-doped material. Hence, after formation of the depletion layer at the interface between the two sides, the net charge density will correspond to the illustration in the lower part of Fig. 9.1. Overall charge neutrality of course demands that

$$N_{DI} W_n = N_{AI} W_p . \tag{9.1}$$

Locally, charges are obviously not balanced and so there exists a "space charge" density. This charge density will produce an electric field $\mathcal{E}(x)$ directed from the positive charge towards the negative, i.e. from right to left. This field acts to prevent further electrons from diffusing into the *p*-side and holes from diffusing into the *n*-side. Hence, one can imagine the formation of the junction as follows: Start from two separated pieces of *n*- and *p*-doped semiconductors. Upon bringing the two together, electrons will diffuse to the left and holes to the right. This continues until a sufficient electric field has been established and further diffusion is prevented. The depletion layer therefore forms a barrier for the charges and our aim is to describe the widths W_n and W_p .

9.1 Analysis of the PN Junction

In the analysis, we first solve the Poisson equation to find the field. We take $x = 0$ at the interface and so the boundary conditions for the field are that $\mathcal{E}(W_n) = \mathcal{E}(-W_p) = 0$. With the charge density in Fig. 9.1 we then find

$$\frac{d\mathcal{E}}{dx} = \frac{1}{\varepsilon\varepsilon_0}\begin{cases} eN_{DI} & 0 < x < W_n \\ -eN_{AI} & -W_p < x < 0 \end{cases}$$

$$\Downarrow$$

$$\mathcal{E}(x) = \frac{1}{\varepsilon\varepsilon_0}\begin{cases} eN_{DI}(x - W_n) & 0 < x < W_n \\ -eN_{AI}(x + W_p) & -W_p < x < 0 \end{cases}.$$

Here, ε is the relative dielectric constant. Setting the two expressions for the field equal at $x = 0$ is easily seen to lead to the charge conservation condition Eq.(9.1). Next, the electric potential V is related to $\mathcal{E}$ via $\mathcal{E} = -dV / dx$. Outside the depletion layers, the potential must be constant and on the two sides the constant values are denoted V_n and V_p, respectively. After integration, we consequently have

$$V(x) = \begin{cases} V_n - \dfrac{eN_{DI}}{2\varepsilon\varepsilon_0}(x - W_n)^2, & 0 < x < W_n \\ V_p + \dfrac{eN_{AI}}{2\varepsilon\varepsilon_0}(x + W_p)^2, & -W_p < x < 0. \end{cases}$$

Again, we require continuity and set the two expressions for the potential equal at $x = 0$. i.e.

$$V_n - \frac{eN_{DI}}{2\varepsilon\varepsilon_0} W_n^2 = V_p + \frac{eN_{AI}}{2\varepsilon\varepsilon_0} W_p^2$$
$$\Downarrow \tag{9.2}$$
$$V_D \equiv V_n - V_p = \frac{eN_{DI}}{2\varepsilon\varepsilon_0} W_n^2 + \frac{eN_{AI}}{2\varepsilon\varepsilon_0} W_p^2,$$

where V_D is the so-called diffusion potential or built-in potential. Schematically, the potential varies across the junction as illustrated by the solid line in Fig. 9.2.

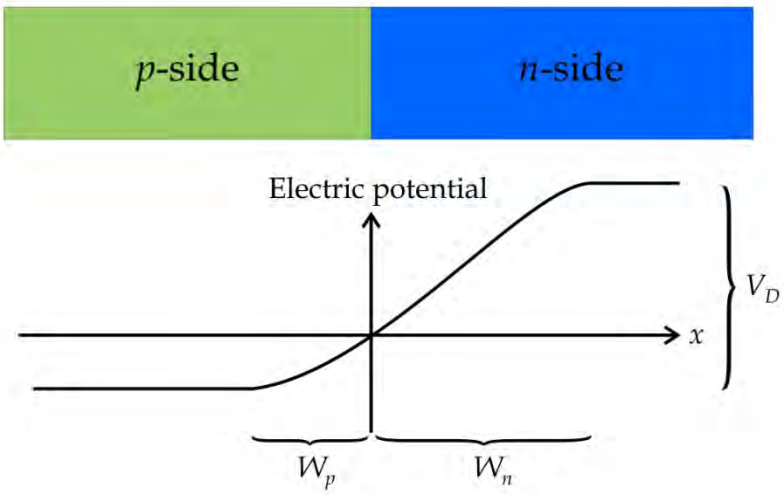


Figure 9.2. The electric potential V across the junction.

Treating V_D as a known quantity, Eqs.(9.1) and (9.2) provide two equations for the two unknowns W_n and W_p that we can solve to give

$$W_n = \left\{ \frac{2\varepsilon\varepsilon_0 N_{AI} V_D}{eN_{DI}(N_{DI} + N_{AI})} \right\}^{1/2}, \quad W_p = \left\{ \frac{2\varepsilon\varepsilon_0 N_{DI} V_D}{eN_{AI}(N_{DI} + N_{AI})} \right\}^{1/2}.$$

In addition, their sum $W = W_n + W_p$ is

$$W = \left\{ \frac{2\varepsilon\varepsilon_0 (N_{DI} + N_{AI})}{eN_{DI} N_{AI}} V_D \right\}^{1/2}. \tag{9.3}$$

To be of any use, however, we still need the value of V_D. To this end, we first establish the *law of mass action*, which is the semiconductor equivalent of the chemical equilibrium condition of statistical mechanics. It is easily derived using the density

expressions from last chapter $n = N_3 \exp\{(E_F - E_c)/kT\}$ and $p = P_3 \exp\{(E_v - E_F)/kT\}$. By forming the product it is seen that the unknown Fermi energy cancels and we get

$$np = N_3 P_3 \exp\{(E_v - E_c)/kT\} = N_3 P_3 \exp\{-E_g/kT\} \equiv n_i^2,$$

where $E_g = E_c - E_v$ is the energy gap and n_i is the density of both electrons and holes in the intrinsic (un-doped) case. The law of mass action therefore states that no matter what happens to the Fermi energy, the product of n and p remains a constant. For GaAs around room temperature, $n_i \approx 2.1 \cdot 10^{12}$ m^{-3} [1]. Now, to the far left the space charges have produced a potential V_p and to the far right the potential is V_n. The potential energy of an electron is changed by an amount $-eV$ by placing it in a potential V. On the other hand, without V the potential energy must be E_c since this is the total energy if $\vec{k} = 0$. Hence, adding V simply corresponds to replacing E_c by $E_c - eV$ for electrons. Similarly, for the valence band E_v is to be replaced by $E_v - eV$. It follows that the electron concentration must vary as $n = N_3 \exp\{(E_F - E_c + eV)/kT\}$. We now form the ratio between electron concentrations on the n- and p-sides:

$$\frac{n(n\text{-side})}{n(p\text{-side})} = \frac{N_3 \exp\{(E_F - E_c + eV_n)/kT\}}{N_3 \exp\{(E_F - E_c + eV_p)/kT\}} = \exp\{eV_D/kT\}.$$

We saw already that $n(n\text{-side}) \approx N_{DI}$ if we are far from the junction. Additionally, the law of mass action tell us that

$$n(p\text{-side}) \cdot p(p\text{-side}) = n_i^2.$$

Thus, using $p(p\text{-side}) \approx N_{AI}$ we see that

$$\exp\{eV_D/kT\} = \frac{N_{DI}N_{AI}}{n_i^2} \quad \Rightarrow \quad V_D = \frac{kT}{e}\ln\left(\frac{N_{DI}N_{AI}}{n_i^2}\right).$$

Used in Eq.(9.3), this demonstrates that the total depletion layer width W is

$$W = \left\{\frac{2\varepsilon\varepsilon_0 kT(N_{DI} + N_{AI})}{e^2 N_{DI}N_{AI}}\ln\left(\frac{N_{DI}N_{AI}}{n_i^2}\right)\right\}^{1/2}.$$

The doping dependence of the depletion width is shown in Fig. 9.3 under the assumption of equal donor and acceptor doping $N_{DI} = N_{AI}$. It is seen that W may become as small as 10 nm or even less if the doping concentration is sufficiently high.

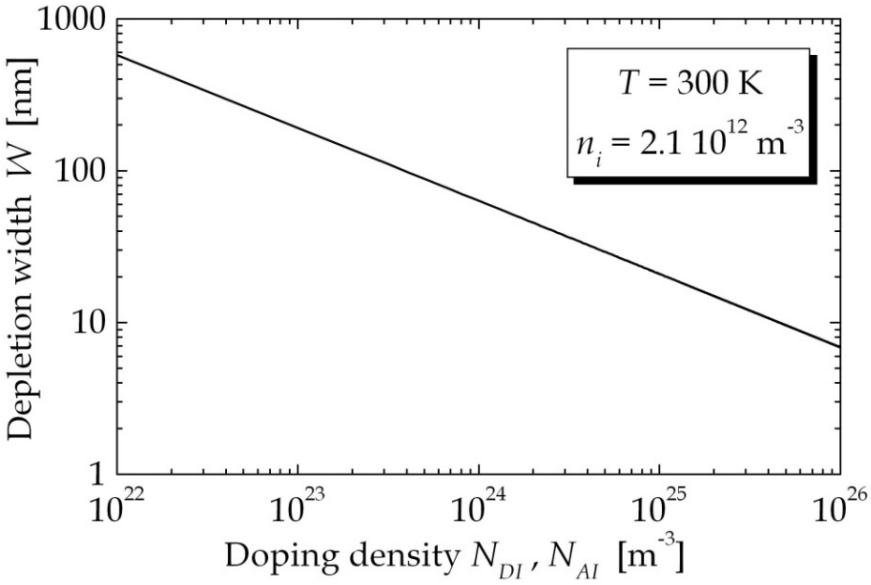


Figure 9.3. The total depletion width W as a function of doping concentrations for GaAs assuming identical donor and acceptor concentrations.

A word of caution is required here, however: At such high impurity concentrations, the assumptions used for the calculation of carrier densities in the previous chapter are likely to break down. In particular, for e.g. heavily n-doped materials the Fermi energy will lie above the band edge E_c and, similarly, E_F is likely to be below E_v on the p-side. Hence, the actual depletion width W at high doping density is expected to deviate from the plot in Fig. 9.3. We will not try to construct a more elaborate description here, however, because we expect the simple theory to be at least qualitatively correct.

9.2 PIN Junction

A pin junction is similar to a pn junction but differs by the insertion of an intrinsic (un-doped) central region, as illustrated in Fig. 9.4. The thickness of the intrinsic layer is d, and the constant electric field in this range is denoted $\mathcal{E}_I$. The analysis of the electrostatic potential follows the previous section closely, and we only list the final result

$$V(x)=\begin{cases} V_n-\dfrac{eN_{DI}}{2\varepsilon\varepsilon_0}(x-d-W_n)^2, & d<x<d+W_n \\ V_p+\dfrac{eN_{AI}}{2\varepsilon\varepsilon_0}W_p^2-\mathcal{E}_I x, & 0<x<d \\ V_p+\dfrac{eN_{AI}}{2\varepsilon\varepsilon_0}(x+W_p)^2, & -W_p<x<0. \end{cases}$$

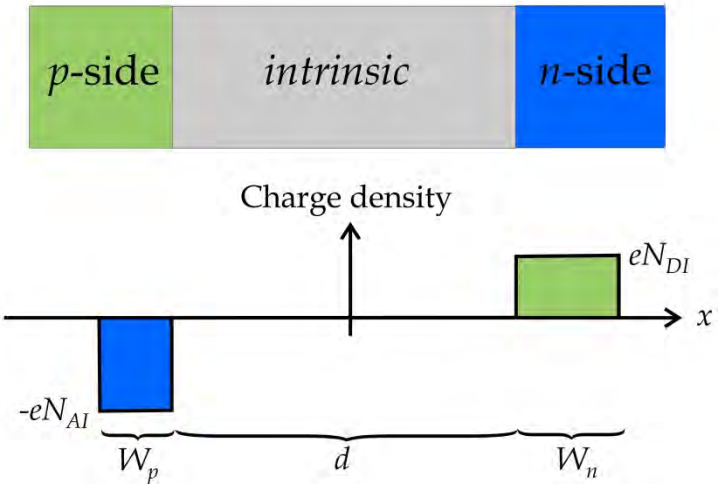


Figure 9.4. Schematic illustration of a pin junction geometry.

By construction, the potential is continuous across the boundary at $x = 0$. To ensure continuity at $x = d$ we require that

$$V_n - \frac{eN_{DI}}{2\varepsilon\varepsilon_0} W_n^2 = V_p + \frac{eN_{AI}}{2\varepsilon\varepsilon_0} W_p^2 - \mathcal{E}_I d .$$

In addition, the slope must be continuous and we find

$$\frac{eN_{DI}}{\varepsilon\varepsilon_0} W_n = -\mathcal{E}_I , \quad \frac{eN_{AI}}{\varepsilon\varepsilon_0} W_p = -\mathcal{E}_I . \tag{9.4}$$

These results just demonstrate that Eq.(9.1) is still valid, as expected. We can combine the boundary conditions to form a single equation for W_p

$$V_D = \frac{eN_{AI}^2}{2\varepsilon\varepsilon_0 N_{DI}} W_p^2 + \frac{eN_{AI}}{2\varepsilon\varepsilon_0} W_p^2 + \frac{eN_{AI}}{\varepsilon\varepsilon_0} W_p d$$

with the solution

$$W_p = \left\{ \left(\frac{N_{DI} d}{N_{DI} + N_{AI}} \right)^2 + \frac{2\varepsilon\varepsilon_0 N_{DI} V_D}{eN_{AI}\left(N_{DI} + N_{AI}\right)} \right\}^{1/2} - \frac{N_{DI} d}{N_{DI} + N_{AI}} .$$

This result is seen to agree with that of the pn junction if d = 0. Also, from Eq.(9.4) it is a simple matter to compute the built-in field. A particular case arises when the intrinsic region is wide, i.e. when $d \gg W_n, W_p$. In this case, the equation for W_p reduces to $V_D \approx eN_{AI} W_p d / (\varepsilon\varepsilon_0)$ and combined with Eq.(9.4) it follows that $\mathcal{E}_I \approx -V_D / d$. Hence, the field is simply the diffusion potential divided by the

thickness. The thickness dependence of the field in a silicon junction is compared to the $1/d$ approximation in Fig. 9.5.

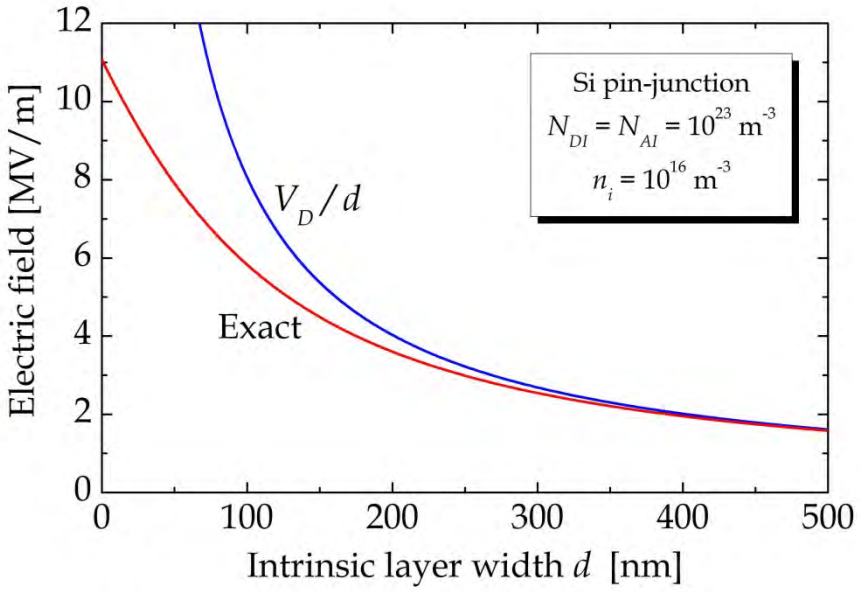


Figure 9.5. Magnitude of the built-in electric field of a pin junction in the depletion approximation.

9.3 Self-Consistent Potential and Trapped Charges

The previous analysis relied heavily on the depletion approximation. In this section, we wish to study the exact, numerical solution to the Poisson equation without invoking this approximation. Moreover, our numerical scheme allows us to include more advanced effects such as trapped charges. To keep things on a reasonably simple level, we restrict the analysis to completely symmetrical pin junctions. That is, the two halves of the junction are complete images of one another. This means that $N_{AI} = N_{DI}$ and we use identical widths of the doped regions. We choose the potential at the midpoint as our zero point and so $V_n = -V_p = V_D/2$. According to the discussion above, the electron and hole densities in the presence of an electric potential are given by

$$n = N_3 \exp\{(E_F - E_c + eV)/kT\}, \quad p = P_3 \exp\{(E_v - E_F - eV)/kT\},$$

respectively. Our assumption of complete electron-hole symmetry implies that we can write

$$n = n_i e^{\varphi}, \quad p = n_i e^{-\varphi}, \quad \varphi \equiv \frac{eV}{kT}.$$

Note that the common pre-factor must necessarily be n_i to comply with the law of mass action $np = n_i^2$.

We now turn to the self-consistent solution of the Poisson equation. The investigated geometry is shown below

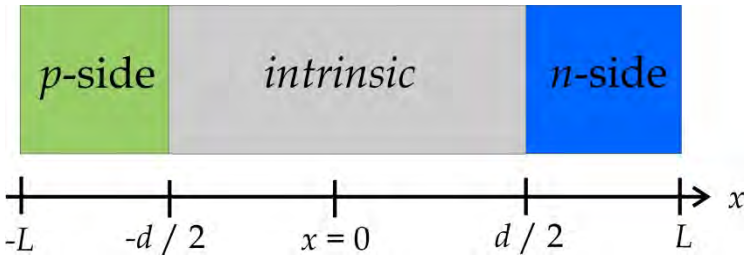


Figure 9.6. Fully symmetric pin junction geometry.

We focus on the region $x \geq 0$ for which we can write the Poisson equation

$$\frac{d^2\varphi}{dx^2} = \frac{e^2}{\varepsilon\varepsilon_0 kT}\left\{n_i e^{\varphi} - n_i e^{-\varphi} - N_{DI}\theta(x - d/2)\right\}.$$

The boundary conditions here are $\varphi(0) = 0$ and $\varphi(L) = \varphi_n \equiv \ln(N_{DI} / n_i)$. The latter is equivalent to $V_n = V_D / 2$. Before attempting a solution, we rewrite as

$$\frac{d^2\varphi}{dx^2} = \frac{2n_i e^2}{\varepsilon\varepsilon_0 kT}\left\{\sinh\varphi - n\,\theta(x - d/2)\right\}, \quad n = \frac{N_{DI}}{2n_i}.$$

Also, it is convenient to measure x in units of the screening length $l = \left(\varepsilon\varepsilon_0 kT / 2n_i e^2\right)^{1/2}$ so that

$$\frac{d^2\varphi}{dx^2} = \sinh\varphi - n\,\theta(x - d/2). \tag{9.5}$$

This differential equation is an example of a broad class of nonlinear problems that can be formulated generally as

$$\frac{d^2\varphi}{dx^2} = F(\varphi(x), x), \quad a \leq x \leq b, \quad F(\varphi(a), a) = F_a, \quad F(\varphi(b), b) = F_b.$$

The last two relations are meant to indicate that boundary conditions are specified in the form of function values at the end points. The general strategy for solving such equations is through discretization. Hence, we divide the x-range into N+1 points via

$$x_i = a + i\Delta, \quad i = 0, \ldots, N, \quad \Delta = \frac{b - a}{N}.$$

Similarly, we define $\varphi_i \equiv \varphi(x_i)$ and $F_i \equiv F(\varphi_i, x_i)$. A simple method involves discretizing the second derivative as follows

$$\frac{d^2\varphi}{dx^2} \approx \frac{\varphi(x+\Delta)+\varphi(x-\Delta)-2\varphi(x)}{\Delta^2}.$$

In this way, we can write

$$\varphi_{i+1} + \varphi_{i-1} - 2\varphi_i = \Delta^2 F_i, \quad i \neq 0, N.$$

This set of nonlinear equations can be reformulated as a matrix equation by introducing

$$\vec{\varphi} = \begin{pmatrix} \varphi_1 \\ \varphi_2 \\ \vdots \\ \varphi_{N-2} \\ \varphi_{N-1} \end{pmatrix}, \quad \vec{F} = \begin{pmatrix} \Delta^2 F_1 - \varphi_0 \\ \Delta^2 F_2 \\ \vdots \\ \Delta^2 F_{N-2} \\ \Delta^2 F_{N-1} - \varphi_N \end{pmatrix}, \quad \overleftrightarrow{H} = \begin{pmatrix} -2 & 1 & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 1 & -2 \end{pmatrix},$$

so that

$$\overleftrightarrow{H} \cdot \vec{\varphi} = \vec{F} \quad \Rightarrow \quad \vec{\varphi} = \overleftrightarrow{H}^{-1} \cdot \vec{F}.$$

It's important to remember that the F's depend on the φ's so that method must be iterated until convergence. For instance, if we write two successive iterations as $\vec{\varphi}_{\text{old}}$ and $\vec{\varphi}_{\text{new}}$ we have

$$\vec{\varphi}_{\text{new}} = \overleftrightarrow{H}^{-1} \cdot \vec{F}(\vec{\varphi}_{\text{old}}).$$

This simple approach, however, often fails to converge. A dramatically improved version is the matrix Newton method that proceeds as follows. First, we write the discretized set of equations as $\overleftrightarrow{H} \cdot \vec{\varphi} = \vec{F}(\vec{\varphi})$. Secondly, we expand around $\vec{\varphi}_{\text{old}}$, i.e.

$$\overleftrightarrow{H} \cdot (\vec{\varphi}_{\text{old}} + \delta\vec{\varphi}) \approx \vec{F}(\vec{\varphi}_{\text{old}}) + \overleftrightarrow{G}(\vec{\varphi}_{\text{old}}) \cdot \delta\vec{\varphi}, \quad \delta\vec{\varphi} = \vec{\varphi}_{\text{new}} - \vec{\varphi}_{\text{old}}.$$

Solving, we then find

$$\vec{\varphi}_{\text{new}} = \vec{\varphi}_{\text{old}} + \left(\overleftrightarrow{H} - \overleftrightarrow{G}(\vec{\varphi}_{\text{old}})\right)^{-1} \cdot \left(\vec{F}(\vec{\varphi}_{\text{old}}) - \overleftrightarrow{H} \cdot \vec{\varphi}_{\text{old}}\right).$$

In the present case, the matrix $\overleftrightarrow{G}$ is a simple diagonal matrix with elements

$$G_{ii} = \Delta^2 \left. \frac{\partial F}{\partial \varphi} \right|_{\varphi=\varphi_i} .$$

For the pin junction, $G_{ii} = \Delta^2 \cosh \varphi_i$. The converged result for Si (l = 29 μm at room temperature) is illustrated as the black curve in Fig. 9.7.

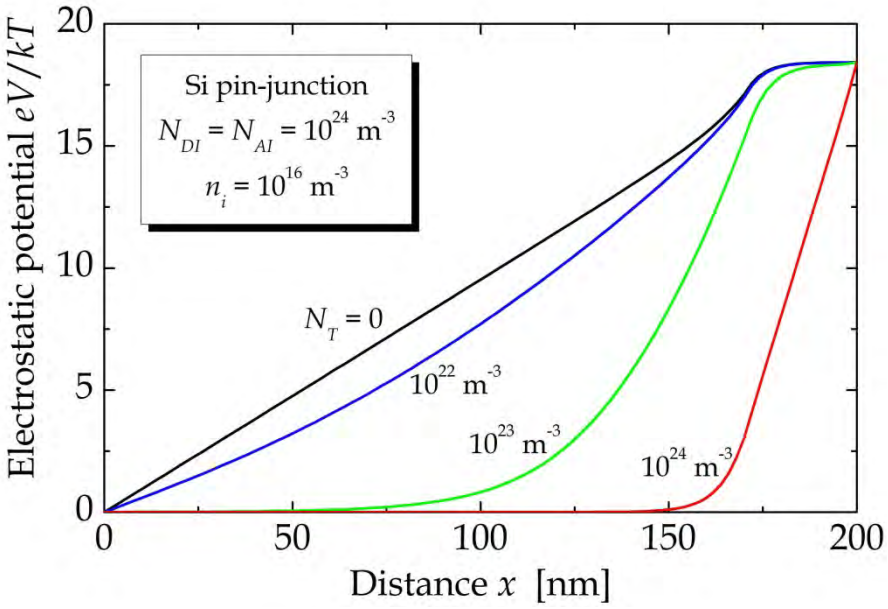


Figure 9.7. Normalized electrostatic potential of a symmetric Si pin junction with varying trap concentrations taking d = 340 nm and L = 200 nm.

We finally wish to discuss the role of traps in such a junction. Traps capture electrons and holes in localized states and prevent them from contributing to the current. Moreover, trapped charges tend to screen the built-in electric field, which may be detrimental for charge separation in a pin junction solar cell. In the previous chapter, we investigated electron donors and derived their occupancy under the assumption that two electrons in a single site is highly improbable due to the repulsive energy. In a completely electron-hole symmetric model this doesn't work, however. The trap state has three possible charge values: $0, +1,$ and -1. The neutral trap corresponds to one electron in the state and ± 1 correspond to an additional hole or electron. To maintain symmetry, it is necessary to require that the energy of the -1 state is precisely twice the energy of the 0 state. Thus, the electron occupancy of a single trap becomes

$$\langle n_T \rangle = \frac{2\exp\{(E_F - E_T)/kT\} + 2\exp\{2(E_F - E_T)/kT\}}{1 + 2\exp\{(E_F - E_T)/kT\} + \exp\{2(E_F - E_T)/kT\}} .$$

If the trap concentration is N_T it follows that the trapped charge concentration is

$$eN_T\left(1-\langle n_T\rangle\right) = eN_T\,\frac{1-\exp\{2(E_F-E_T)/kT\}}{1+2\exp\{(E_F-E_T)/kT\}+\exp\{2(E_F-E_T)/kT\}}$$
$$= eN_T\tanh\left(\frac{E_T-E_F}{2kT}\right).$$

Again, in the presence of a potential V, we must replace $E_T \to E_T - eV$. Moreover, the electron-hole symmetry requires $E_T = 0$ eV. Finally, the Fermi level (which is obviously independent of x) must be fixed at $E_F = 0$ eV as well. This implies that we can write the modified Poisson equation as

$$\frac{d^2\varphi}{dx^2} = \sinh\varphi + n_T\tanh\frac{\varphi}{2} - n\,\theta(x-d/2), \quad n_T = \frac{N_T}{2n_i}.$$

Naturally, in the Newton method, we now include the additional term and take $G_{ii} = \Delta^2\left\{\cosh\varphi_i + n_T/[2\cosh^2(\varphi_i/2)]\right\}$. Some results for various trap densities are shown in Fig. 9.7. Notice how trapped charges tend to screen the potential and thereby reduce the electric field.

Exercise: Self-consistent pn junction model.

The pn junction model used above is only approximately correct. The main problem lies in our assumption of a simple rectangular charge distribution as shown in Fig. 9.1. To remedy this, one should calculate the actual charge distribution rather than postulating it. In this exercise, we consider a pn junction under the simplifying assumption of symmetry between N- and P-sides and between electrons and holes. Hence, the starting point is Eq.(9.5) but with d = 0

$$\frac{d^2\varphi}{dx^2} = \sinh\varphi - n, \quad n = \frac{N_{DI}}{2n_i}, \quad x > 0.$$

This is a complicated, second-order nonlinear differential equation but can be reformulated as a somewhat simpler first-order equation using a mathematical trick. First, we multiply by $d\varphi/dx$:

$$\frac{d\varphi}{dx}\frac{d^2\varphi}{dx^2} = \frac{d\varphi}{dx}\{\sinh\varphi - n\}.$$

a) Show that this is equivalent to

$$\frac{d}{dx}\left\{\left(\frac{d\varphi}{dx}\right)^2\right\} = 2\frac{d}{dx}\{\cosh\varphi - n\varphi\}.$$

b) Integrate this result and use the boundary conditions to show that

$$\frac{d\varphi}{dx} = \sqrt{2}\left\{\cosh\varphi - n\varphi - \cosh\varphi_n + n\varphi_n\right\}^{1/2}, \quad \varphi_n = \ln(2n).$$

This result, which is still exact, provides an analytical expression for the normalized electric field $\mathcal{E}_{\text{norm}}(x) = -d\varphi / dx$ at the junction

$$\mathcal{E}_{\text{norm}}(0) = -\sqrt{2}\left\{1 - \cosh\varphi_n + n\varphi_n\right\}^{1/2} \approx -\sqrt{2n\left(\ln(2n) - 1\right)}.$$

The actual and normalized fields are related via $\mathcal{E}(x) = \left(2n_i kT / \varepsilon\varepsilon_0\right)^{1/2} \mathcal{E}_{\text{norm}}(x)$.

c) Show that the depletion approximation predicts a normalized field of

$$\mathcal{E}_{\text{norm}}(0) = -\sqrt{2n\ln(2n)}.$$

References

[1] M. Balkanski and R.F. Wallis *Semiconductor Physics and Applications* (Oxford University Press, New York, 2000)

10. PN Junction and Tunneling Diodes

In the previous chapter, we analyzed the pn and pin junctions without bias. Hence, we could compute electrostatic potentials and carrier concentrations based on thermal equilibrium statistical mechanics. In the present chapter, we wish to study the behavior under a bias voltage producing a current flowing through the junction. The purpose is to compute this current under general assumptions and ultimately understand the rectifying behavior of pn junction diodes. In order to fully describe the device, however, we want to generalize the results of Chapter 4. Throughout this book, we have relied on linear response theory and so the calculated current necessarily varies linearly with the applied voltage. This means that we restrict ourselves to the ohmic regime, in which current and voltage are proportional. To fully describe the rectifying behaviour of a diode, we clearly need to go beyond the linear response and investigate the non-linear relation between current I and applied voltage V_A. Our approach will rely on plausible arguments in order to generalize the results of Chapter 4 rather than rigorous analysis. We begin by recalling Eq.(4.5) for the ohmic regime

$$I = -\frac{e^2 V_A}{\pi\hbar}\sum_n \int f'(E) T_{n\to n}(E) dE\,, \tag{10.1}$$

where the assumption of decoupled channels has been made. Here, the summation is over the transverse eigenstates labeled by the index n. Each eigenstate corresponds to a conduction "channel" with transmittance $T_{n\to n}$ and transverse energy $E_n^{\perp}$. The total energy of an electron in this state is therefore $E = E_n^{\perp} + E_{\|}$, where $E_{\|}$ is the kinetic energy associated with the motion parallel to the wire axis.

To proceed, we now consider the specific geometry illustrated in Fig. 10.1. Here, the "system" again is positioned between two leads: left (L) and right (R). The right lead is grounded and the left lead is biased at a potential V_A.

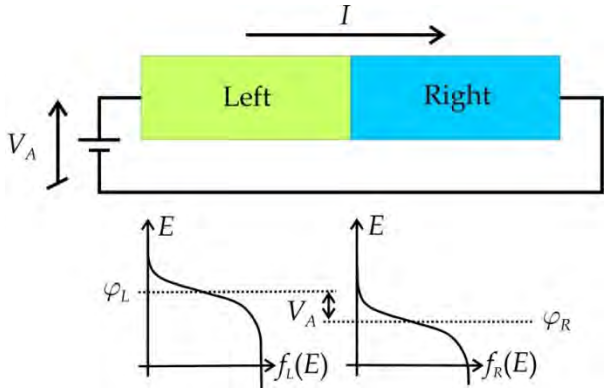


Figure 10.1. Biased junction with the right-hand side grounded. The Fermi functions with Fermi levels $\varphi_{L/R}$ for the two sides are shown below.

In the presence of a potential V, the electron energies are shifted by an amount $-eV$. Hence, in the left and right lead, the electron energies are really $E_L = E - eV_A$ and $E_R = E$, respectively. This means that the statistical distributions for the two halves can be written

$$f_L(E) = \frac{1}{\exp\{(E - eV_A - E_F)/kT\} + 1}, \quad f_R(E) = \frac{1}{\exp\{(E - E_F)/kT\} + 1}.$$

Naturally, $f_R(E)$ is simply the usual, unbiased Fermi function $f(E)$. Now, if V_A is sufficiently small it is obvious that

$$f_L(E) - f_R(E) \approx V_A \cdot \left.\frac{\partial f_L(E)}{\partial V_A}\right|_{V_A=0} = -eV_A \cdot \left.\frac{\partial f_L(E)}{\partial E}\right|_{V_A=0}.$$

On this ground, we postulate [1] that the current flowing in Fig. 10.1 under a finite bias voltage should be calculated by the following generalization of Eq.(10.1):

$$I = \frac{e}{\pi\hbar}\sum_n \int_0^\infty \{f_L(E) - f_R(E)\} T_{n\to n}(E) dE_{\|}. \tag{10.2}$$

Here, we changed integration variable from E to $E_{\|}$ since their difference $E_n^{\perp}$ is just a constant. This expression is a very general result applying equally well to low dimensional structures for which the summation is over a restricted set of discrete transverse states as to "bulk" devices for which the summation becomes an integral over a continuum of transverse states. We now focus on the latter situation.

10.1 Three-Dimensional Non-Degenerate Devices

If the transverse area of the junction A is sufficiently large, the summation over n in Eq.(10.3) becomes an integral over a continuum of transverse eigenstates. These states are labelled by the transverse wave vector $\vec{k}_\perp = (k_y, k_z)$, which replaces the subscript n. Thus, the sum is actually over (k_y, k_z) and writing $E_\perp = \hbar^2 k_\perp^2 / 2m_e$ to get $dE_\perp / dk_\perp = \hbar^2 k_\perp / m_e$ we find

$$\sum_n (...) = \sum_{k_y,k_z} (...) = \frac{A}{4\pi^2}\int (...) dk_y dk_z = \frac{Am_e}{2\pi\hbar^2}\int (...) dE_\perp .$$

By writing $E_\perp = \hbar^2 k_\perp^2 / 2m_e$ we have implicitly chosen the semiconductor conduction band edge E_c as the zero-point for the energy. Summing over all transverse states is consequently replaced by an integral over the transverse energies $E_\perp$. Also, the n-

dependence of $T_{n\to n}(E)$ now should be thought of as a dependence on the continuous parameter $E_\perp$ and, thereby, T becomes a function of the both energies $E_\parallel$ and $E_\perp$ and so Eq.(11.2) becomes

$$I = \frac{eAm_e}{2\pi^2\hbar^3}\int_0^\infty\int_0^\infty \{f_L(E) - f_R(E)\}T(E_\parallel, E_\perp)dE_\parallel dE_\perp.$$

Here, of course $E = E_\parallel + E_\perp$. Note that by integrating only over $E_\perp \geq 0$ we restrict the calculation to states with energy above the semiconductor conduction band edge. In the present section, we want to focus on "non-degenerate" situations similar to the treatment in Chapter 8. Non-degeneracy means that the mean occupancy of states in the conduction band is much less than unity. This is the case whenever the Fermi energy is far below all energies E in the band. In this case, the distribution functions can be approximated as in Chapter 8 so that

$$f_L(E) \approx \exp\{(E_F + eV_A - E)/kT\}, \quad f_R(E) \approx \exp\{(E_F - E)/kT\}. \tag{10.3}$$

This approximation is applicable to both the metal in the left lead and to the semiconductor since all energies considered lie above E_c and, hence, far above E_F. Consequently,

$$f_L(E) - f_R(E) \approx \left(e^{eV_A/kT} - 1\right)\exp\{(E_F - E)/kT\}$$

and so

$$I = \frac{eAm_e}{2\pi^2\hbar^3}\left(e^{eV_A/kT} - 1\right)e^{E_F/kT}\int_0^\infty\int_0^\infty e^{-E/kT}T(E_\parallel, E_\perp)dE_\parallel dE_\perp.$$

As we have indicated, the transmittance may, in general, depend on both components of the kinetic energy. In systems with perfect translational invariance along the interfaces, however, phase-matching requires conservation of the perpendicular component of the wave vector $\vec{k}_\perp$ and accordingly of $E_\perp = \hbar^2 k_\perp^2/2m_e$. Separation of perpendicular and parallel variables then shows that only the latter are involved in the tunnelling process. Consequently, T is independent of $E_\perp$ and integrating over this degree of freedom leads to

$$I = \frac{eAm_e kT}{2\pi^2\hbar^3}\left(e^{eV_A/kT} - 1\right)e^{E_F/kT}\int_0^\infty e^{-E_\parallel/kT}T(E_\parallel)dE_\parallel. \tag{10.4}$$

This expression is a valuable tool for the calculation of currents in various types of junctions. It is also clear that we expect a rather universal behaviour $I \propto e^{eV_A/kT} - 1$ for

different devices. As we shall see next, a fully classical analysis of the pn junction diode agrees precisely with this expectation.

10.2 PN Junction Diode

In diode applications, an external bias voltage V_A is applied across the diode. For definiteness, we take the *n*-side to be grounded so that this potential remains fixed at a value V_n. Hence, by a positive bias we understand a positive voltage applied to the *p*-side raising the potential here from V_p to $V_p + V_A$. In the biased case, the potential follows the dashed line in Fig. 10.2.

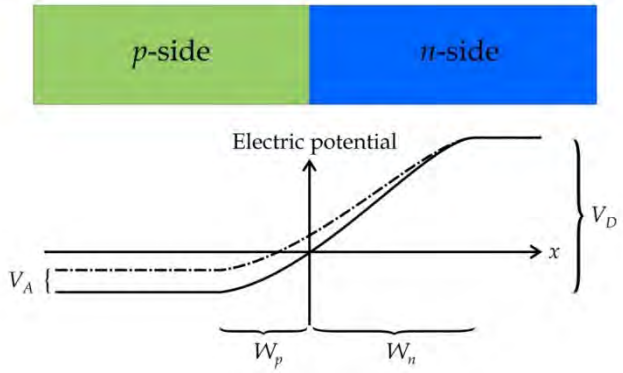


Figure 10.2. The electric potential V across the junction. The solid and dashed lines illustrate the cases without and with and external bias V_A assuming the *n*-side grounded.

The full analysis of the biased pn junction is a complicated numerical problem, even for the present one-dimensional case. It consists in solving the coupled drift-diffusion equations. The first of these is the Poisson equation for the potential

$$\frac{d^2V}{dx^2} = -\frac{1}{\varepsilon\varepsilon_0}\rho(x),$$

where ρ is the charge density. Secondly, we need the expressions for electron and hole currents

$$J_e(x) = \sigma_e\mathcal{E} + eD_e\frac{dn}{dx}, \quad J_h(x) = \sigma_h\mathcal{E} - eD_h\frac{dp}{dx}.$$

Here, $\mathcal{E}$ is the electric field whereas σ_i and D_i are conductivities and diffusivities of electrons (i = e) and holes (i = h). Under the effective mass approximation, the conductivities can be assumed proportional to carrier densities as shown in Eq.(8.2). Hence, we write $\sigma_e = e\mu_e n$ and $\sigma_h = e\mu_h p$, where μ_i is the mobility. The diffusivities are connected to the mobilities via the Einstein relation $D_i = \mu_i kT/e$ [1]. To complete the system of equations, we need to look at the way currents change with position.

The total current J flowing through the junction is a sum $J = J_e + J_h$ of electron and hole currents and, moreover, independent of position x. Hence,

$$\frac{\partial J_e}{\partial x} + \frac{\partial J_h}{\partial x} = 0 .$$

However, the individual currents *do* vary across the junction. This follows from the continuity equation, which for electrons reads as

$$\frac{\partial n}{\partial t} = \frac{1}{e}\frac{\partial J_e}{\partial x} + G_e ,$$

where G_e is the net generation rate of electrons. Similarly for holes

$$\frac{\partial p}{\partial t} = -\frac{1}{e}\frac{\partial J_h}{\partial x} + G_h ,$$

where G_h is the net generation rate of holes. In steady state, subtracting the two equations shows that $G_e = G_h \equiv G$. This makes sense, as electrons and holes are always produced in pairs. Several different mechanisms contribute to carrier generation and recombination. The most important ones are (i) band to band, (ii) trap mediated and (iii) Auger processes. We will focus on the important case of trap mediated carrier generation and recombination. In this case, an analysis of the relevant transition rates (see the exercise) leads to the result

$$G = \frac{n_i^2 - np}{\tau_e (n + n_0) + \tau_h (p + p_0)} .$$

Here, τ_e and τ_h are electron and hole lifetimes and n_0 and p_0 are factors proportional to the conduction and valence state densities subject to the condition $n_0 p_0 = n_i^2$. This is the famous Shockley-Read-Hall expression for the carrier generation rate. To simplify matters even further, we may assume that things are roughly symmetrical in that $\tau_e \approx \tau_h = \tau$ and $n_0 \approx p_0 = n_i$. In that case,

$$G = \frac{1}{\tau}\frac{n_i^2 - np}{n + p + 2n_i} . \tag{10.5}$$

When combined under steady-state conditions, the drift-diffusion equations therefore read

$$\frac{d^2V}{dx^2} = \frac{e}{\varepsilon\varepsilon_0}\{n(x) - p(x) - N_{DI}(x) + N_{AI}(x)\}$$

$$J_e(x) = e\mu_e n(x)\mathcal{E}(x) + kT\mu_e \frac{dn}{dx}, \quad J_h(x) = e\mu_h p(x)\mathcal{E}(x) - kT\mu_h \frac{dp}{dx} \qquad (10.6)$$

$$\frac{dJ_e}{dx} = -eG(x), \quad \frac{dJ_h}{dx} = eG(x).$$

These must be supplemented by the density expressions

$$n(x) = N_3 \exp\{[E_F - E_c + eV(x)]/kT\} \quad \text{and} \quad p(x) = P_3 \exp\{[E_v - eV(x) - E_F]/kT\}.$$

Solving the full set is difficult and can only be done using numerical tools. However, a simple approximation can be constructed by the following reasoning: Far to the left of the depletion region in a wide device there is practically no electric field. Hence, the current is purely diffusive. By combining the last two lines of equations, it then follows that

$$kT\mu_e \frac{d^2n}{dx^2} = -eG(x) \quad \Rightarrow \quad D_e \frac{d^2n}{dx^2} = -G(x).$$

Moreover, far to the left we may take $p \approx N_{AI}$ and so using the generation rate Eq.(10.5) and the fact that $N_{AI} \gg n, n_i$ leads to

$$D_e \frac{d^2n}{dx^2} = -\frac{n_i^2 - nN_{AI}}{\tau N_{AI}}.$$

At the extreme left, near the metal contact, thermal equilibrium is restored because all excess carriers are forced to recombine via metal and interface states. Hence, the electron density is n_i^2/N_{AI}. The simple second order equation above can now be solved under the boundary conditions $n(-\infty) = n_i^2/N_{AI}$ and $n(-W_p) = n_i^2/N_{AI}\exp\{eV_A/kT\}$ that lead to

$$n(x) = \frac{n_i^2}{N_{AI}}\left\{1 + \left(e^{eV_A/kT} - 1\right)e^{(x+W_p)/L_e}\right\},$$

where $L_e = \sqrt{\tau D_e}$ is the electron diffusion length. A similar expression is found for the hole density to the right of the depletion zone. Taking the gradients, we therefore find the currents

$$J_e(x) = \frac{eD_e n_i^2}{L_e N_{AI}}\left(e^{eV_A/kT} - 1\right)e^{(x+W_p)/L_e}, \quad x < -W_p$$
$$J_h(x) = \frac{eD_h n_i^2}{L_h N_{DI}}\left(e^{eV_A/kT} - 1\right)e^{-(x-W_n)/L_h}, \quad x > W_n.$$

The total current is $J = J_e(x) + J_h(x)$. The problem, however, with these expressions is that they cannot simply be added because they are valid in different regions. Of course, we can always write

$$J = J_h(W_n) + J_e(-W_p) + J_{scr}, \quad J_{scr} = J_e(W_n) - J_e(-W_p).$$

Here, J_{scr} is the space charge region current that accounts for the change in electron current over the depletion region. We can rewrite this quantity using the generation rate

$$J_{scr} = \int_{-W_p}^{W_n} \frac{dJ_e(x)}{dx}dx = -\frac{e}{\tau}\int_{-W_p}^{W_n} \frac{n_i^2 - np}{n + p + 2n_i}dx.$$

At both edges of the depletion region we have $np = n_i^2 \exp\{eV_A/kT\}$. Hence, as a simple estimate we will take $n \approx p \approx n_i \exp\{eV_A/2kT\}$ throughout the zone so that

$$J_{scr} \approx \frac{eWn_i}{2\tau}\frac{e^{eV_A/kT} - 1}{e^{eV_A/2kT} + 1} = \frac{eWn_i}{2\tau}\left(e^{eV_A/2kT} - 1\right).$$

As a consequence, this contribution varies as $\exp\{eV_A/2kT\}$ at large bias in contrast to the diffusive part varying as $\exp\{eV_A/kT\}$. Taken together the total current can be approximated as

$$I(V_A) = I_0\left(\exp\{eV_A/\eta kT\} - 1\right), \tag{10.7}$$

where η is an "ideality factor" and I_0 is a constant related to the electric properties of the materials.

10.3 Tunneling Current

In a low or moderately doped pn junction diode, the depletion region is sufficiently thick that carriers cannot tunnel directly between valence and conduction bands. In contrast, in highly doped structures the barrier thickness is greatly reduced. Consequently, when the barrier becomes sufficiently thin (about 10 nm) tunneling of carriers is expected to occur. This phenomenon is the basis of the *tunneling diode*, also known as the Esaki diode after its inventor. Because carriers simply tunnel through

a thin barrier, the tunnel diode is an extremely fast device capable of operation at several hundred GHz. Hence, it is an important application of nanoscale features in electronics.

It is noted that the built-in voltage V_D is reduced to $V_D - V_A$ in the presence of the bias. In turn, replacing V_D by $V_D - V_A$ in Eq.(9.3) shows that V_A tends to reduce the depletion width, i.e.

$$W(V_A) = \left\{ \frac{2\varepsilon\varepsilon_0 (N_{DI} + N_{AI})}{eN_{DI}N_{AI}} (V_D - V_A) \right\}^{1/2}, \quad V_A \le V_D . \tag{10.8}$$

This expression breaks down if $V_A > V_D$ and is, therefore, restricted to the low bias range $V_A \le V_D$. The depletion region forms a quantum mechanical barrier for electrons to tunnel directly between valence and conduction bands. We will analyze this problem using the WKB tunneling formula derived in Chapter 5

$$T(E) = T_0 \exp\left\{ -2 \int_{x_1}^{x_2} \sqrt{\frac{2m_e}{\hbar^2} (U(x) - E)} dx \right\}, \tag{10.9}$$

in which $U(x)$ is the varying potential energy and $x_{1,2}$ are the endpoints of the barrier given by the solutions of $U(x_{1,2}) - E = 0$. In the present context, the tunneling barrier can be understood from Fig. 10.3.

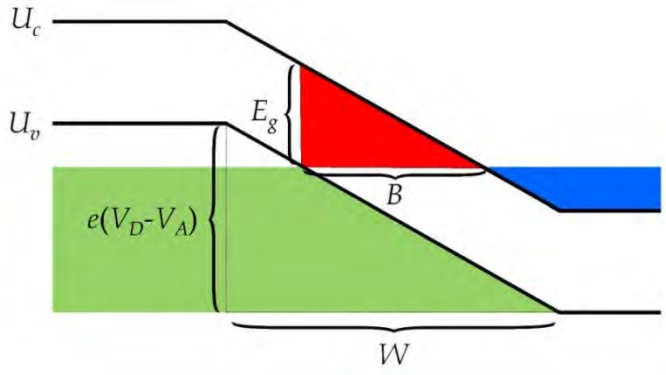


Figure 10.3. Tunneling geometry showing the potential energy of electrons in the two bands. The green and blue areas are occupied by electrons and the red triangle is the tunneling barrier.

This figure illustrates the x-dependence of the potential energy across the junction. For an electron in the conduction band, the potential energy is $U_c = E_c - eV$ and similarly for the valence band $U_v = E_v - eV$. These energies are displayed as the solid lines in the figure using the simplification that the electric potential varies roughly linearly across the barrier rather than having the actual curved shape shown in Fig.

10.2. The two bands are separated by the band gap E_g. Also, the green and blue areas illustrate occupied electron states. In general, the Fermi levels on the two sides of the junction will not be equal since equilibrium is disturbed when a current is flowing. We ignore this complication here, however, and take a single value E_F as the Fermi energy throughout the structure.

Consider an electron with energy equal to the Fermi energy moving from right to left in the *n*-side. It obviously belongs to the conduction band. At a certain point inside the depletion layer, the electron encounters a region that is classically forbidden, that is, the potential energy U_c exceeds the total energy E_F of the electron. This is indicated as the red area in the figure. The electron cannot move further to the left within the conduction band and normally would be reflected back. There is, however, a possibility that the electron tunnels through the red barrier into the valence band, where it is free to keep moving to the left. This is the origin of the tunneling current. As shown in the figure, the full height of the triangular barrier is the band gap E_g and the width is denoted *B*. This width *B* is not known but by comparing congruent triangles in the figure it is seen that

$$\frac{B}{E_g} = \frac{W}{e(V_D - V_A)} \quad \Rightarrow \quad B = \frac{W}{e(V_D - V_A)} E_g .$$

With an *x*-coordinate running from right to left, the barrier height is given by $U_c(x) - E_F = E_g x / B$. Thus, the transmittance in Eq.(10.4) is easily obtained as

$$T(E_F) = T_0 \exp\left\{ -2 \int_0^B \sqrt{\frac{2m_e}{\hbar^2} \frac{E_g}{B} x}\, dx \right\} = T_0 \exp\left\{ -\frac{4}{3} \sqrt{\frac{2m_e E_g}{\hbar^2}} B \right\}.$$

If finally the expression for the barrier width *B* is inserted we find

$$T(E_F) = T_0 \exp\left\{ -\frac{4}{3} \sqrt{\frac{2m_e E_g^3}{\hbar^2}} \frac{W}{e(V_D - V_A)} \right\}.$$

When we combine with Eq.(9.3) for the depletion width we can rewrite this expression in a simple form highlighting the bias dependence:

$$T(E_F) = T_0 \exp\left\{ -\sqrt{\frac{V_B}{V_D - V_A}} \right\}, \quad V_B = \frac{64 \varepsilon \varepsilon_0 m_e E_g^3 (N_{DI} + N_{AI})}{9 e^3 \hbar^2 N_{DI} N_{AI}}. \qquad (10.10)$$

Having determined the barrier transmittance allows us to calculate the tunneling current I_T by means of the low-bias Landauer formula Eq.(4.6):

$$I_T(V_A) = \frac{2e^2}{h} V_A T(E_F)$$

per occupied channel. In fact, we don't fulfill the conditions applied in deriving the Landauer formula in the present case. The problem is that electrons are incident in one channel (conduction band) and tunneling into a completely different one (valence band). However, we expect a similar behavior and we might therefore assume $I_T(V_A) \propto V_A T(E_F)$. Thus, we can formulate the tunneling current as

$$I_T(V_A) = G_T V_A \exp\left\{-\sqrt{\frac{V_B}{V_D - V_A}}\right\},$$

where G_T is an effective conductance related to T_0. To get a sizable tunneling current we need a small V_B. From Eq.(10.10) is it clear that this requires a small effective mass m_e and a large doping concentration. Thus, heavily doped GaAs is a good candidate and in Fig. 10.5 below, we have taken $N_{DI} = N_{AI} = 10^{26}$ m^{-3}. This choice yields $V_B \approx 300$ V and $V_D \approx 1.6$ V. The figure illustrates the tunneling current as well as the normal current given by Eq.(10.7) and the total current.

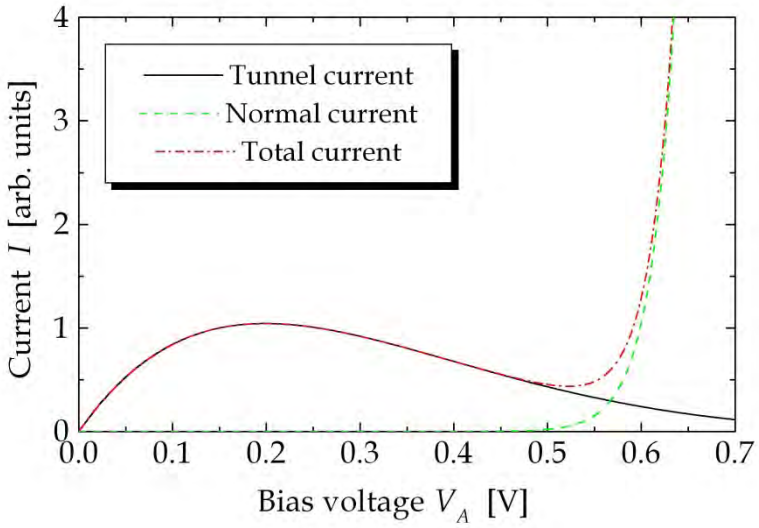


Figure 10.5. Tunnel, normal and total current vs. applied voltage for heavily doped GaAs.

Experimentally, the tunneling effect was first observed in Germanium pn junctions by L. Esaki. Eventually, Esaki was rewarded with the Nobel Prize for this work. The plot below is taken from his original publication from 1958.

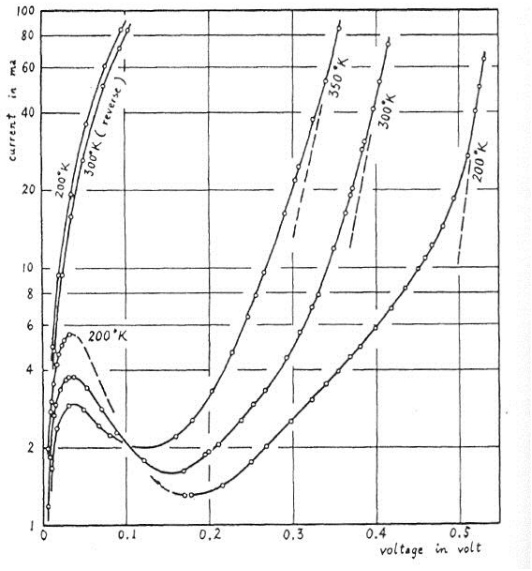


Figure 10.6. Experimental tunneling current as a function of voltage for a narrow Ge pn junction. Taken from [2].

Exercise: Trap-assisted carrier generation and recombination.

In this exercise, we analyse electron and hole generation and recombination via a trap state located in the band gap. The situation is illustrated in Fig. 10.7.

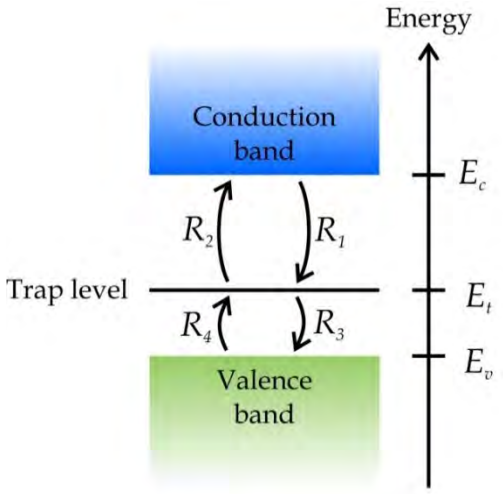


Figure 10.7. Schematic of the four processes by which carriers hop between bands and trap states.

Four processes, denoted R_1 through R_4 in the figure, contribute to the charge balance. They are given by

$$R_1 = C_n n(1-f_t), \quad R_2 = C_n n_0 f_t$$
$$R_3 = C_p p_0(1-f_t), \quad R_4 = C_p p f_t$$

In these expressions, f_t denotes the Fermi functions evaluated at the trap level, i.e. the probability that a trap is occupied. Also, C_n and C_p denote characteristic trap capture/emission coefficients that are proportional to trap density. Finally, n_0 and p_0 denote the density of available electron states in the conduction band and available hole states in the valence band, respectively. Obviously, $G_e = R_2 - R_1$ and $G_h = R_3 - R_4$.

a) Show that by equating these rates, we are lead to the balance

$$C_n n_0 f_t - C_n n(1-f_t) = C_p p_0 (1-f_t) - C_p p f_t$$
$$\Downarrow$$
$$f_t = \frac{C_n n + C_p p_0}{C_n (n+n_0) + C_p (p+p_0)}.$$

b) Show that applying the results immediately provides the carrier generation rates as

$$G_e = G_h = \frac{C_n C_p (n_0 p_0 - np)}{C_n (n+n_0) + C_p (p+p_0)} \equiv G.$$

Now, this expression should hold also in the unbiased cased for which $G_e = G_h = 0$ and $np = n_i^2$. Hence, it follows that $n_0 p_0 = n_i^2$. Also, we will introduce carrier lifetimes $\tau_e \equiv 1/C_p$ and $\tau_h \equiv 1/C_n$ so that

$$G = \frac{n_i^2 - np}{\tau_e (n+n_0) + \tau_h (p+p_0)}.$$

This famous expression for the carrier generation rate was first derived by Shockley, Read, and Hall.

References

[1] M. Balkanski and R.F. Wallis *Semiconductor Physics and Applications* (Oxford University Press, New York, 2000)
[2] L. Esaki, Phys. Rev. 109, 603 (1958).

11. Metal-Semiconductor Junctions

Junctions between metallic leads and semiconductors are of great practical importance. As we shall see, under appropriate conditions currents flowing across the junction are "rectified" in that current flow is only possible if electrons move from the semiconductor into the metal. This is the basis for "Schottky" diodes and the physical barrier preventing currents flowing in the opposite direction is known as a Schottky barrier. We will consider the specific geometry illustrated in Fig. 11.1. Here, the "system" again is positioned between two leads: left (L) and right (R). The right lead is grounded and the left lead is biased at a potential V_A.

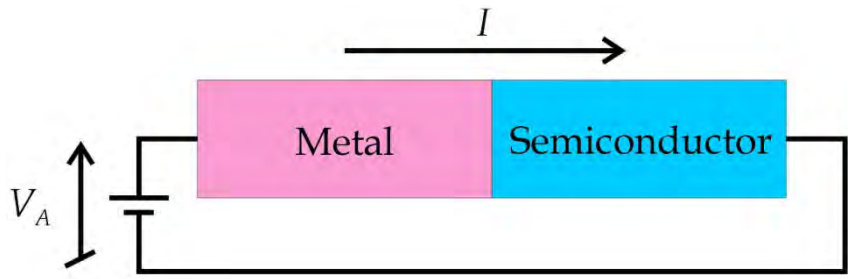


Figure 11.1. Biased metal-semiconductor junction with the semiconductor side grounded.

11.1 Schottky Diode

As in the previous chapter, we base the analysis on the general current-voltage relation for three-dimensional degenerate devices, c.f. Eq.(10.4) repeated here:

$$I = \frac{eAm_e kT}{2\pi^2\hbar^3}\left(e^{eV_A/kT}-1\right)e^{E_F/kT}\int_0^\infty e^{-E_\parallel/kT}T(E_\parallel)dE_\parallel. \qquad (11.1)$$

When a metal is brought into contact with a semiconductor, charge will be redistributed near the interface. If the semiconductor is *n*-type, electrons will flow into the metal until a balance is reached. Electrons move into the metal because low-energy states are available there. The flow soon stops, however, because a surplus of negative charge is build up on the metallic side of the junction. This negative space charge produces an electric field that counteracts the electron flow. Stated in a different way, a potential energy barrier is formed and the tunnelling transmittance T of this barrier is precisely the one needed to calculate the I/V characteristic of the Schottky diode using Eq.(11.1). We now seek to describe the barrier in the *un-biased* case, i.e. assuming vanishing applied voltage.

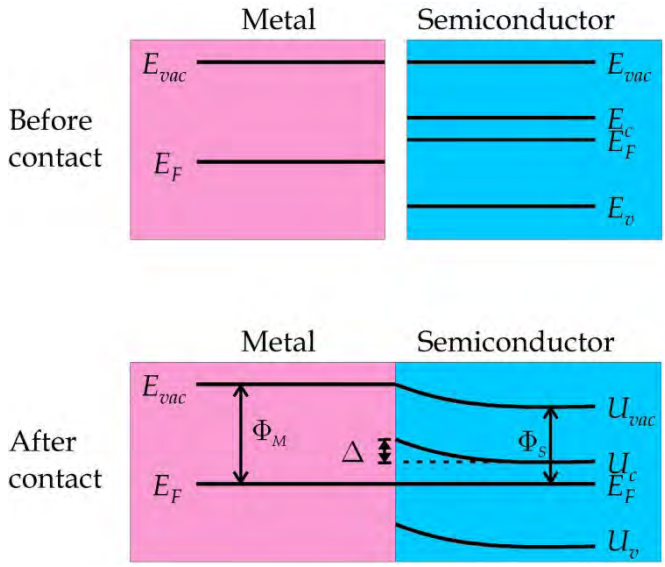


Figure 11.2. Schematic of the metal-semiconductor junction formed by contacting metal and semiconductor pieces. After contact, charge redistribution leads to band bending.

To compute shape and height of the barrier, we consider Fig. 11.2, which illustrates the situation before and after a metal is brought into contact with an *n*-type semiconductor. An important consequence of the contact is that the Fermi levels of the two materials equilibrate. This is a requirement if thermal equilibrium without charge flow is to establish. The potential curves really display the electron energies in the presence of the electrostatic potential in the junction. Hence, the conduction band energy $U_c = E_c - eV$ contains the constant band edge E_c and the electrostatic "band bending" $-eV$ of the electrons. Similarly, the "vacuum" level consists of the overall vacuum level and the band bending $U_{vac} = E_{vac} - eV$. As indicated in the figure, no band bending is assumed in the metallic region. The reason for this is the much higher density of screening charges in the metal that, as we shall see below, leads to a negligible variation of the potential inside the metal. Inside the metal, the vacuum level is separated from the Fermi level by the bulk metal work function Φ_M. Similarly, well into the semiconductor side, all bands are flat and the distance from the Fermi level to the vacuum level Φ_S is a sum of the electron affinity $EA = E_{vac} - E_c$ and the conduction band off-set $E_c - E_F$. These observations allow us to calculate the barrier height Δ if we make the following assumptions: (1) the vacuum level is completely flat inside the metal and (2) the vacuum level in continuous across the junction. In reality, assumption (2) is of course always fulfilled. However, charges may accumulate in an extremely thin layer at the interface due to defect and interface states. Such a localized charge layer will contribute to the electrostatic potential V and effectively cause all bands to appear discontinuous at the junction. Under the two assumptions, though, Fig. 11.2 immediately shows that the barrier for an electron in the semiconductor conduction band to tunnel into the metal is $\Delta = \Phi_M - \Phi_S$.

The spatial width of the barrier is of importance as well. To find it, the spatial dependence of the electrostatic potential V must be calculated. As in Chapter 10, the basic relation to be used is the Poisson equation

$$\frac{d^2V}{dx^2} = -\frac{\rho}{\varepsilon\varepsilon_0},$$

where ρ is the spatially varying charge density and ε is the relative dielectric constant of the material. As in the analysis of the pn junction in Chapter 10, we will assume total depletion of electrons in a layer of width W_n inside the *n*-type semiconductor. As we have seen, this amount to saying that the potential is a constant V_n for $x > W_n$ and follows a quadratic behaviour

$$V(x) = V_n - \frac{eN_{DI}}{2\varepsilon_S\varepsilon_0}(x - W_n)^2,$$

for $0 < x < W_n$. Here, ε_S is the relative dielectric constant of the semiconductor and N_{DI} is the density of donor impurities. On the metallic side of the junction, an estimate of the charge density is found using the Thomas-Fermi model. Here, the metal is assumed to be a simple free-electron metal with a density of states given by Eq.(8.4), i.e. $D(E) = \eta\sqrt{E}$ with $\eta = \sqrt{2}m^{3/2}/\pi^2\hbar^3$. It follows that the electron concentration is $n_0 = \frac{2}{3}\eta E_F^{3/2}$ under normal circumstances. However, if an electrostatic potential V is applied while the Fermi level is kept fixed, all energies are lowered by $-eV$, which amounts to raising the Fermi level by an amount $+eV$ just as in Eq.(10.3). Now, in a neutral bulk material the density of positive charge must be n_0 to compensate the electrons. Hence, the net charge in presence of the potential must be

$$\rho = -e\left\{\tfrac{2}{3}\eta(E_F + eV)^{3/2} - \tfrac{2}{3}\eta E_F^{3/2}\right\} \approx -e^2\eta V E_F^{1/2} = -\tfrac{3}{2}n_0e^2V/E_F, \qquad (11.2)$$

where we have Taylor-expanded to first order under the assumption that $e|V| \ll E_F$. When used in the Poisson equation, it follows that

$$\frac{d^2V}{dx^2} = \frac{3e^2n_0}{2\varepsilon_M\varepsilon_0E_F}V \equiv k_{TF}^2V, \quad k_{TF}^2 = \frac{3e^2n_0}{2\varepsilon_M\varepsilon_0E_F}, \qquad (11.3)$$

where ε_M is the relative dielectric constant of the metal and k_{TF} is the Thomas-Fermi wave number. We choose as a boundary condition $V(-\infty) = 0$, i.e. a vanishing potential in the bulk of the metal and so the solution to this equation is a simple exponential

$$V(x) = V_0 \exp(k_{TF} x),$$

where V_0 is a (yet undetermined) constant. To determine V_0 and the remaining constant W_n we now compare the charges inside the metal to that in the semiconductor. Integrating Eq.(11.2) for the metal we find a total charge per area of

$$Q_M = \int_{-\infty}^{0} \rho dx = -\tfrac{3}{2} n_0 e^2 / E_F \int_{-\infty}^{0} V_0 e^{k_{TF} x} dx = -\varepsilon_M \varepsilon_0 k_{TF} V_0 .$$

As in Chapter 9, we find for the semiconductor $Q_S = eN_{DI} W_n$. Putting $Q_M + Q_S = 0$ provides one equation for the unknowns. A second is obtained from the requirement of continuity of the potential at $x = 0$, which is equivalent to the condition $V_0 = V_n - eN_{DI} W_n^2 / 2\varepsilon_S \varepsilon_0$. Taken together, we find

$$W_n^2 + \frac{2\varepsilon_S}{\varepsilon_M k_{TF}} W_n - \frac{2\varepsilon_S \varepsilon_0}{eN_{DI}} V_n = 0$$

$$\Downarrow$$

$$W_n = -\frac{\varepsilon_S}{\varepsilon_M k_{TF}} + \left\{ \left(\frac{\varepsilon_S}{\varepsilon_M k_{TF}} \right)^2 + \frac{2\varepsilon_S \varepsilon_0}{eN_{DI}} V_n \right\}^{1/2} .$$

(the minus-solution is discarded as we're obviously looking for a positive quantity). To evaluate this result we need to compare two length scales: k_{TF}^{-1} and $(V_n / eN_{DI})^{1/2}$. Using the definition in Eq.(11.3) we see that $k_{TF}^{-1} \propto (E_F / e^2 n_0)^{1/2}$. Since for any real metal-semiconductor junction $n_0 \gg N_{DI}$, we must have $(V_n / eN_{DI})^{1/2} \gg k_{TF}^{-1}$ and so

$$W_n \approx \left\{ \frac{2\varepsilon_S \varepsilon_0}{eN_{DI}} V_n \right\}^{1/2} .$$

By the same token, $V_0 = V_n - eN_{DI} W_n^2 / 2\varepsilon_S \varepsilon_0 \approx 0$ follows by insertion. Thus, as promised above, band bending inside the metal is negligible. A direct consequence is that $eV_n = \Delta$ and so

$$W_n \approx \left\{ \frac{2\varepsilon_S \varepsilon_0}{e^2 N_{DI}} \Delta \right\}^{1/2} . \qquad (11.4)$$

Having established height and width of the tunnelling barrier we may now compute the tunnelling current using Eq.(11.1). If the barrier is approximated by a triangular one of height Δ and width W_n and the WKB-approximation is applied, we can write

$$T(E_\parallel) \approx \begin{cases} \exp\left\{-\dfrac{4W_n}{3\Delta}\sqrt{\dfrac{2m_e}{\hbar^2}}\left(\Delta - E_\parallel\right)^{3/2}\right\} & E_\parallel < \Delta \\ 1 & E_\parallel > \Delta \end{cases}.$$

Hence, the total current can be split into two parts: A genuine tunnelling contribution from states with $E_\parallel < \Delta$ and a so-called thermionic current produced by states with energy $E_\parallel > \Delta$. The latter is easily evaluated and we find

$$\begin{aligned} I_{thermionic} &= \frac{eAm_e kT}{2\pi^2\hbar^3}\left(e^{eV_A/kT}-1\right)e^{E_F/kT}\int_\Delta^\infty e^{-E_\parallel/kT}dE_\parallel \\ &= \frac{eAm_e(kT)^2}{2\pi^2\hbar^3}\left(e^{eV_A/kT}-1\right)e^{(E_F-\Delta)/kT} \\ &\equiv I^0_{thermionic}\left(e^{eV_A/kT}-1\right), \quad I^0_{thermionic} = \frac{eAm_e(kT)^2}{2\pi^2\hbar^3}e^{(E_F-\Delta)/kT}. \end{aligned} \tag{11.5}$$

The constant $R \equiv em_e(kT)^2/(2\pi^2\hbar^3)$ is known as Richardson's constant and for fixed temperature it only depends on the effective mass of the semiconductor material.

The tunnelling current is found from a similar integral

$$I_{tunnel} = \frac{eAm_e kT}{2\pi^2\hbar^3}\left(e^{eV_A/kT}-1\right)e^{E_F/kT}\int_0^\Delta e^{-E_\parallel/kT}\exp\left\{-\frac{4W_n}{3\Delta}\sqrt{\frac{2m_e}{\hbar^2}}\left(\Delta-E_\parallel\right)^{3/2}\right\}dE_\parallel.$$

To evaluate this expression, we first introduce a new variable $z = (\Delta - E_\parallel)/kT$ and a new dimensionless constant $\beta = 4\sqrt{2m_e}W_n(kT)^{3/2}/3\hbar\Delta$. Considering the typical magnitude of W_n (100-1000 nm) it is clear that normally $\beta \gg 1$. In terms of these quantities, the tunnelling current can be written as

$$I_{tunnel} = \frac{eAm_e(kT)^2}{2\pi^2\hbar^3}\left(e^{eV_A/kT}-1\right)e^{(E_F-\Delta)/kT}\int_0^{\Delta/kT} e^{z-\beta z^{3/2}}dz.$$

The integral cannot be evaluated analytically but thanks to the large value of β, the upper limit may be extended to infinity with little loss of accuracy. The resulting integral can be expressed exactly in terms of hypergeometric functions. To simplify the final result, however, a Taylor expansion in β^{-1} can be applied whenever $\beta \gg 1$, i.e. for wide barriers. In this manner,

$$\int_0^\infty e^{z-\beta z^{3/2}}dz \approx \frac{1}{3\beta^2} + \frac{2\Gamma\left(\frac{2}{3}\right)}{3\beta^{2/3}}\left(1+\frac{5}{27\beta^2}\right) + \frac{2\Gamma\left(\frac{1}{3}\right)}{9\beta^{4/3}}\left(1+\frac{7}{54\beta^2}\right).$$

The final tunnelling current then becomes

$$I_{tunnel} = I^0_{tunnel}\left(e^{eV_A/kT} - 1\right),$$
$$I^0_{tunnel} = \frac{eAm_e(kT)^2}{2\pi^2\hbar^3} e^{(E_F-\Delta)/kT}\left\{\frac{1}{3\beta^2} + \frac{2\Gamma\left(\frac{2}{3}\right)}{3\beta^{2/3}}\left(1+\frac{5}{27\beta^2}\right) + \frac{2\Gamma\left(\frac{1}{3}\right)}{9\beta^{4/3}}\left(1+\frac{7}{54\beta^2}\right)\right\}. \quad (11.6)$$

It follows that the ratio between tunnelling and thermionic current is simply

$$\frac{I_{tunnel}}{I_{thermionic}} = \frac{1}{3\beta^2} + \frac{2\Gamma\left(\frac{2}{3}\right)}{3\beta^{2/3}}\left(1+\frac{5}{27\beta^2}\right) + \frac{2\Gamma\left(\frac{1}{3}\right)}{9\beta^{4/3}}\left(1+\frac{7}{54\beta^2}\right).$$

We take the example of GaAs ($m_e = 0.066m$) to evaluate the current ratio as a function of barrier width. For room temperature and $\Delta = 1$ eV, the result is shown in Fig. 11.3. As expected, the tunnelling current dominates for a sufficiently narrow barrier ($W_n < 250$ nm). Finally, for completeness, we note that a drift-diffusion current exists as well. It also follows the form $I_{d-d} = I^0_{d-d}\left(e^{eV_A/kT} - 1\right)$ but we will not prove this.

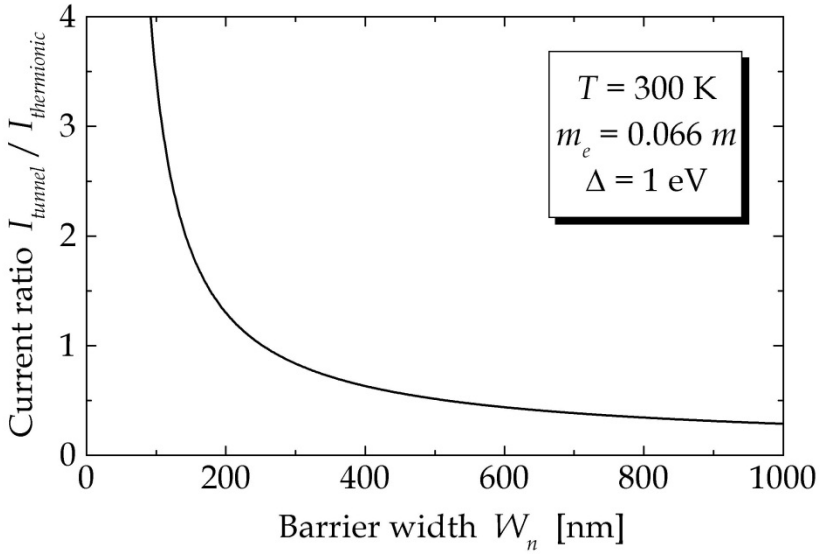


Figure 11.3. Ratio between tunneling and thermionic currents in a GaAs Schottky diode at room temperature.

According to our results, the ratio between tunnelling and thermionic currents is independent of the applied voltage. In reality, the barrier itself is influenced by the voltage and so the current ratio may depend on voltage as well.

Exercise: Nano-scale Schottky diode

This exercise is built on Ref. [2]. We consider the geometry in Fig. 11.4 consisting of a circular metallic pad on top of an n-type semiconductor. The radius a of the pad is in the 10-1000 nm range. The problem of the exercise is to calculate the electrostatic potential for this geometry in the case of vanishing bias voltage. The boundary conditions for the potential are (1) $V=0$ on the contact area between metal and semiconductor, i.e. for $\{z=0 \ \wedge \ 0\le r\le a\}$ and (2) $V=V_n$ outside the border Γ in the figure.

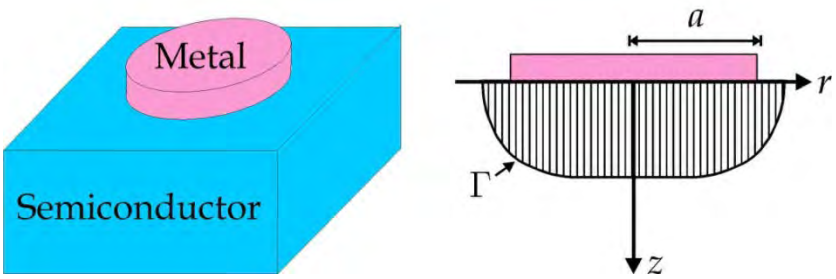


Figure 11.4. A nano-scale Schottky diode formed as a circular metallic pad on top of an n-type semiconductor. The hatched area in the right panel illustrates the extent of the depleted region with boundary curve Γ.

The Poisson equation for this problem is

$$\nabla^2 V=-\frac{eN_{DI}}{\varepsilon_S\varepsilon_0}\times\begin{cases}1 & \text{above curve } \Gamma\\ 0 & \text{below curve } \Gamma.\end{cases}$$

To simplify the mathematics, we switch to so-called oblate spheroidal coordinates (s,t) related to (r,z) via the transformations

$$\begin{gathered}r=a\sqrt{(1+s^2)(1-t^2)},\quad z=ast\\ 0\le s<\infty,\quad -1\le t\le 1\end{gathered}.$$

The Laplacian in these coordinates reads as

$$\nabla^2=\frac{1}{a^2(s^2+t^2)}\left\{\frac{\partial}{\partial s}(1+s^2)\frac{\partial}{\partial s}+\frac{\partial}{\partial t}(1-t^2)\frac{\partial}{\partial t}\right\}.$$

a) Consider first the case of vanishing doping for which $\nabla^2V=0$ throughout. Show that the solution is of the form $V=f(s)g(t)$ and that $f(s)=A+B\tan^{-1}(s)$ and $g(t)=1+C\tanh^{-1}(t)$. Hint: $d\tan^{-1}(s)/ds=(1+s^2)^{-1}$ and $d\tanh^{-1}(t)/dt=(1-t^2)^{-1}$.

b) The contact area between metal and semiconductor $\{z=0 \wedge 0\le r\le a\}$ corresponds to $\{s=0 \wedge -1\le t\le 1\}$. Also, $z\to\infty$ is equivalent to $s\to\infty$. Use the boundary conditions to show that $A=C=0$ and $B=\frac{2}{\pi}V_n$. Hence, the full solution for the un-doped case is $V=\frac{2}{\pi}V_n\tan^{-1}(s)$.

The fact that the un-doped solution is independent of t indicates that the general solution will depend only weakly on t. Hence, in the Poisson equation we will ignore $\partial V/\partial t$ and so the full problem for the depletion region (region above curve Γ) is approximately

$$\frac{\partial}{\partial s}(1+s^2)\frac{\partial V}{\partial s}=-(s^2+t^2)v,\quad v=\frac{eN_{DI}a^2}{\varepsilon_S\varepsilon_0}.$$

c) While the homogeneous solution is still of the form $V=A(t)+B(t)\tan^{-1}(s)$, show that the particular one is $V=\{(1-3t^2)\ln(1+s^2)-s^2\}v/6$. The full solution is the sum of these. Use the boundary condition for the contact area to show that $A(t)=0$.

d) On the boundary curve Γ the boundary condition is $V=V_n$. In addition, $\partial V/\partial s=0$ on the curve Γ. We denote by $s_0(t)$ the values of s found by tracing the curve Γ as t varies between -1 and 1. Show that the condition $\partial V/\partial s=0$ on Γ implies $B(t)=s_0(s_0^2+3t^2)v/3$.

The condition $V=V_n$ on Γ now means that

$$\frac{6V_n}{v}=2s_0(s_0^2+3t^2)\tan^{-1}(s_0)+(1-3t^2)\ln(1+s_0^2)-s_0^2.$$

e) Taking $V_n=0.4$ V and $v=1.5\ \text{V}/\mu\text{m}^2\cdot a^2$, which are realistic values, solve this equation numerically for $0\le t\le 1$. Use the coordinate transformation to calculate corresponding (r,z) points. In this manner, the curve Γ is computed. Take $a=1\mu\text{m}, 0.3\mu\text{m}$ and $0.1\mu\text{m}$ and plot the Γ curves that, if successful, should look like the ones in Fig. 11.5.

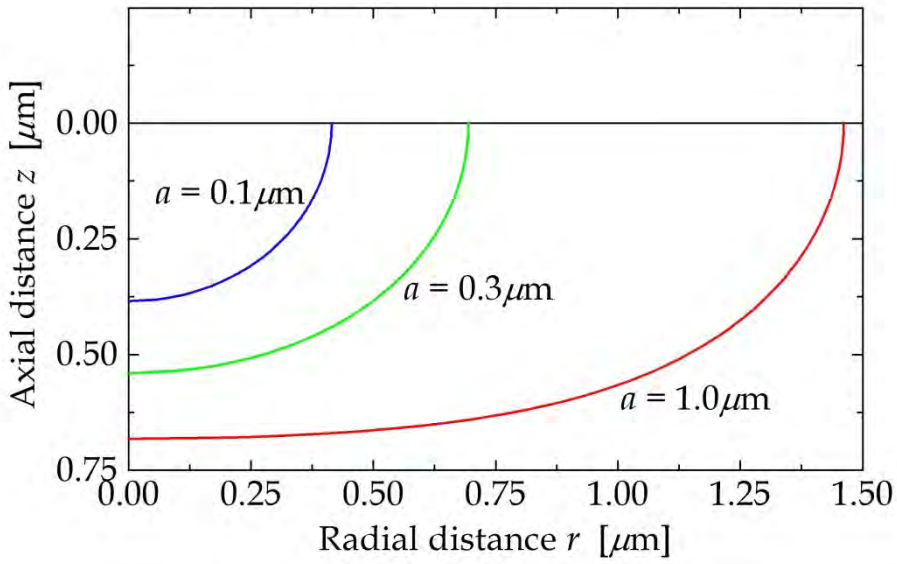


Figure 11.5. Calculated boundary curves Γ for three different radii of the metallic pad.

References

[1] S. Datta *Electronic Transport in Mesoscopic Systems* (Cambridge Univ. Press, Cambridge, 1997).
[2] C. Donolato, J. Appl. Phys. 95, 2184 (2004).

12. Semiclassical Transport

In this chapter, we try to establish the link between quantum and classical approaches to transport. As we will demonstrate, the two agree for metallic structures provided momentum and velocity are treated *semiclassically*, i.e. determined from the slope of the true energy bands. In our exposition, we will focus on a single-band metal. However, for multi-band structures, the final result should simply be summed over bands.

We start by introducing the electron distribution function g that governs the number of electrons at a given position, momentum and time:

$$dN(\vec{r},\vec{k},t) = g(\vec{r},\vec{k},t)\frac{d^3rd^3k}{4\pi^3}.$$

Here, the normalization is clearly such that the total number of electron is

$$N_{tot} = \int g(\vec{r},\vec{k},t)\frac{d^3rd^3k}{4\pi^3}.$$

In thermal equilibrium at a given position characterized by a local Fermi level $E_F(\vec{r})$ and temperature $T(\vec{r})$, the distribution function is just the Fermi-Dirac distribution

$$g(\vec{r},\vec{k},t) \to f(\vec{r},\vec{k}) = \frac{1}{\exp\left\{\frac{E(\vec{k})-E_F(\vec{r})}{kT(\vec{r})}\right\}+1}.$$

We restrict ourselves to the relaxation time approximation, in which the rate of change of the distribution function is determined by a characteristic relaxation time towards equilibrium

$$\frac{d}{dt}g(\vec{r},\vec{k},t) = -\frac{g(\vec{r},\vec{k},t)-f(\vec{r},\vec{k})}{\tau}. \tag{12.1}$$

If we momentarily ignore all but the explicit time dependence, we have

$$\frac{d}{dt}g(t) = -\frac{g(t)-f}{\tau},$$

with the simple solution $g(t) = f + [g(t_0)-f]e^{-t/\tau}$ as can easily be verified. Hence, the distribution relaxes exponentially toward thermal equilibrium. More generally, the rate of change is

$$\begin{aligned}\frac{d}{dt}g(\vec{r},\vec{k},t) &= \frac{g(\vec{r}+d\vec{r},\vec{k}+d\vec{k},t+dt)-g(\vec{r},\vec{k},t)}{dt} \\ &\approx \frac{\nabla g\cdot d\vec{r}+\nabla_{\vec{k}}g\cdot d\vec{k}+\frac{\partial g}{\partial t}dt}{dt} \\ &= \nabla g\cdot\vec{v}+\nabla_{\vec{k}}g\cdot\frac{d\vec{k}}{dt}+\frac{\partial g}{\partial t}.\end{aligned} \tag{12.2}$$

Here, we Taylor expanded in the infinitesimals in the second line and applied $\vec{v} = d\vec{r}\,/\,dt$ in the third. This differential equation for the electron distribution is known as Boltzmann's transport equation. We now return to Eq.(12.1) and rewrite as

$$g(\vec{r},\vec{k},t) = f(\vec{r},\vec{k}) - \tau\frac{d}{dt}g(\vec{r},\vec{k},t). \tag{12.3}$$

Under the assumption that we are close to equilibrium, we can write down an order-by-order expansion for g. In 0^{th} order, we clearly have

$$g_0(\vec{r},\vec{k},t) = f(\vec{r},\vec{k}).$$

To construct the 1^{st} order result, we note that by inserting the 0^{th} order term in Eq.(12.2) it follows that

$$\frac{d}{dt}g(\vec{r},\vec{k},t) \approx \nabla f\cdot\vec{v}+\nabla_{\vec{k}}f\cdot\frac{d\vec{k}}{dt}, \tag{12.4}$$

as $\partial f\,/\,\partial t = 0$. We then find

$$g_1(\vec{r},\vec{k},t) = f(\vec{r},\vec{k}) - \tau\nabla f\cdot\vec{v} - \tau\nabla_{\vec{k}}f\cdot\frac{d\vec{k}}{dt}.$$

The last term can be evaluated by appealing to semiclassical arguments. First, $\vec{p} = \hbar\vec{k}$ is the momentum, which in an electric field $\vec{\mathcal{E}}$ is governed by Newton's law

$$\frac{d\vec{p}}{dt} = -e\vec{\mathcal{E}}.$$

Hence, we find

$$g_1(\vec{r},\vec{k},t) = f(\vec{r},\vec{k}) - \tau\nabla f\cdot\vec{v} + \frac{e\tau}{\hbar}\nabla_{\vec{k}}f\cdot\vec{\mathcal{E}}. \tag{12.5}$$

We will apply this expression as an approximation for the general distribution function. Since f is known explicitly, all derivatives are easily calculated. Focusing on the x-direction we have

$$\begin{aligned}\frac{\partial f}{\partial x} &= \frac{\partial}{\partial x}\left(\frac{E-E_F}{kT}\right)kT\frac{\partial f}{\partial E} \\ &= -\frac{\partial E_F}{\partial x}\frac{\partial f}{\partial E} - \left(\frac{E-E_F}{T}\right)\frac{\partial T}{\partial x}\frac{\partial f}{\partial E}.\end{aligned}$$

Similarly, for the k-derivative

$$\frac{\partial f}{\partial k_x} = \frac{\partial E}{\partial k_x}\frac{\partial f}{\partial E}.$$

Semiclassically, $E = p^2/2m$ with $\vec{p} = \hbar\vec{k}$ so that $\partial E/\partial k_x = \hbar p_x/m = \hbar v_x$. Putting it all together, we then find

$$g_1(\vec{r},\vec{k},t) = f + \tau\frac{\partial f}{\partial E}\left\{e\vec{\mathcal{E}} + \nabla E_F + \left(\frac{E-E_F}{T}\right)\nabla T\right\}\cdot\vec{v}\,.$$

The first two driving terms may be grouped into an effective electric field $\vec{\mathcal{E}}_{eff} = \vec{\mathcal{E}} + \nabla E_F/e$ and so

$$g_1(\vec{r},\vec{k},t) = f + \tau\frac{\partial f}{\partial E}\left\{e\vec{\mathcal{E}}_{eff} + \left(\frac{E-E_F}{T}\right)\nabla T\right\}\cdot\vec{v}\,. \tag{12.6}$$

12.1 Transport Coefficients

We now want to apply the distribution function to evaluate certain transport coefficients. To simplify matters, we will assume transport in the x-direction only so that $\vec{\mathcal{E}}_{eff} = \mathcal{E}_{eff}\,\vec{e}_x$ and $\vec{v} = v_x\vec{e}_x$ and so

$$g_1(\vec{r},\vec{k},t) = f + \tau\frac{\partial f}{\partial E}\left\{e\mathcal{E}_{eff} + \left(\frac{E-E_F}{T}\right)\frac{\partial T}{\partial x}\right\}v_x\,. \tag{12.7}$$

Generally, both charge and energy flows. Thus, we can introduce the (familiar) electric current as well as a new energy (or heat) current density given by

$$J_e = -e\int v_x g(\vec{r},\vec{k},t)\frac{d^3k}{4\pi^3}$$
$$J_Q = \int [E-E_F] v_x g(\vec{r},\vec{k},t)\frac{d^3k}{4\pi^3}.$$

The reason that E_F is subtracted in the last expression is that the heat current is really the flow of free energy [1]. Approximating *g* by Eq.(12.7) we first find a contribution from $g_0 = f$. This contribution clearly vanishes as it represents the current flow in the unperturbed state. For the remaining contributions we can generally write

$$J_e = L_{11}\mathcal{E}_{eff} - L_{12}\frac{\partial T}{\partial x}$$
$$J_Q = L_{21}\mathcal{E}_{eff} - L_{22}\frac{\partial T}{\partial x}.$$

The coefficients L_{ij} are the co-called transport coefficients and by insertion it follows that

$$L_{11} = -\frac{e^2}{4\pi^3}\int \tau v_x^2 \frac{\partial f}{\partial E} d^3k, \quad L_{12} = \frac{e}{4\pi^3 T}\int \tau v_x^2 [E-E_F]\frac{\partial f}{\partial E} d^3k,$$
$$L_{21} = TL_{12}, \quad L_{22} = -\frac{1}{4\pi^3 T}\int \tau v_x^2 [E-E_F]^2 \frac{\partial f}{\partial E} d^3k. \tag{12.8}$$

Throughout, the velocity should be evaluated from $v_x = \hbar^{-1}\partial E/\partial k_x$. Among these coefficients, L_{11} is recognized as the usual intraband conductivity. Also, the ratio $Q = L_{12}/L_{11}$ is often called the thermopower or Seebeck coefficient and $\Pi = L_{21}/L_{11}$ is the Peltier coefficient. Note that $\Pi = TQ$.

We will now evaluate the transport coefficients for a free electron metal. Such a material is characterized by a uniform electron density *n* given by

$$n = \frac{1}{4\pi^3}\int f(E) d^3k = \frac{k_F^3}{3\pi^2}, \tag{12.9}$$

where $k_F = \sqrt{2mE_F}/\hbar$ is the Fermi wave number. The density of states is simply $D(E) = 3nE^{1/2}/(2E_F^{3/2})$. Finally, for this isotropic material we make the substitution $v_x^2 \to \overline{v}^2 \equiv (v_x^2 + v_y^2 + v_z^2)/3$. As $E = m(v_x^2 + v_y^2 + v_z^2)/2$ it follows that $\overline{v}^2 = 2E/3m$. Thus, converting the integrations from *k* to *E* means that

$$L_{11} = -e^2\int_0^\infty \tau \frac{2E}{3m}\frac{\partial f}{\partial E} D(E) dE = -\frac{e^2 n}{mE_F^{3/2}}\int_0^\infty \tau \frac{\partial f}{\partial E} E^{3/2} dE.$$

If we now (1) assume τ independent of energy, and (2) take the low temperature limit $\partial f / \partial E \approx -\delta(E - E_F)$ we see that $L_{11} = \tau e^2 n / m$ in agreement with the quantum result. For L_{12} we find by analogy

$$L_{12} = \frac{en}{mE_F^{3/2}T} \int_0^\infty \tau [E - E_F] \frac{\partial f}{\partial E} E^{3/2} dE.$$

This integral is trickier because the energy factor $E - E_F$ cancels the naïve low temperature contribution. However, using the so-called Sommerfeld expansion it can be shown that [1]

$$\int_0^\infty h(E) \frac{\partial f}{\partial E} dE \approx -\int_0^{E_F} \frac{\partial h(E)}{\partial E} dE - \frac{\pi^2 k^2 T^2}{6} \frac{\partial^2 h(E)}{\partial E^2}\bigg|_{E=E_F}.$$

For τ independent of energy this yields

$$L_{12} \approx \frac{e\tau n}{mE_F^{3/2}T} \left(-\frac{\pi^2 k^2 T^2}{6} \right) 3E_F^{1/2} = -\frac{\pi^2 e\tau n k^2 T}{2mE_F},$$

which means $Q \approx -\pi^2 k^2 T / (2eE_F)$. As the figure shows, this is an excellent approximation under realistic conditions.

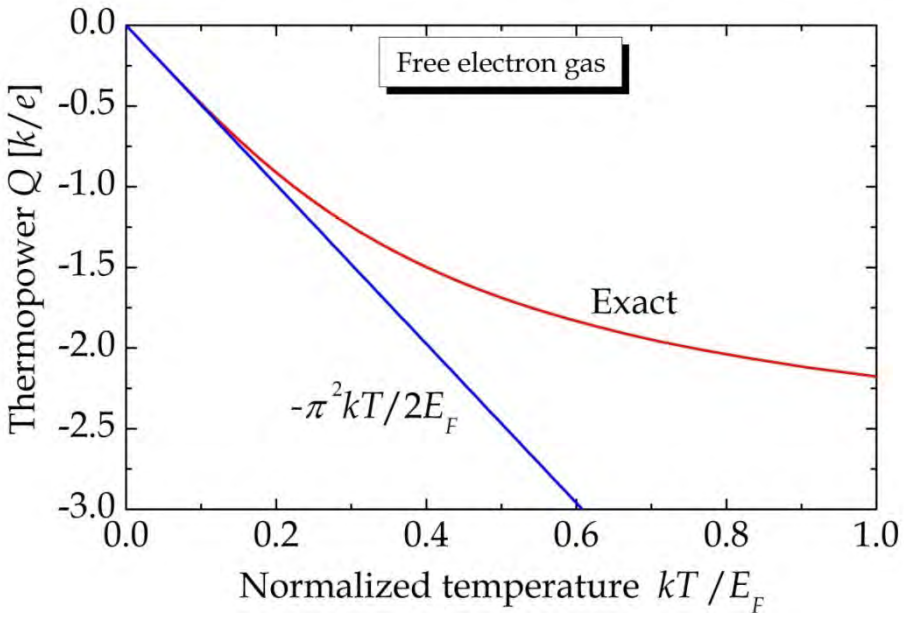


Figure 12.1. Comparison of the numerically evaluated thermopower to the low-temperature approximation for a free electron gas.

Exercise: Transport in graphene

Graphene is a two-dimensional material and so it is more appropriate to talk of sheet transport coefficients given by (with the notation $f'(E) = \partial f / \partial E$)

$$L_{11} = -\frac{e^2}{2\pi^2}\int \tau v_x^2 f'(E) d^2k, \quad L_{12} = \frac{e}{2\pi^2 T}\int \tau v_x^2 [E - E_F] f'(E) d^2k, \\ L_{21} = TL_{12}, \quad L_{22} = -\frac{1}{2\pi^2 T}\int \tau v_x^2 [E - E_F]^2 f'(E) d^2k. \tag{12.10}$$

Moreover, the energy spectrum is especially simple in the so-called Dirac approximation, where the actual band structure is replaced by two Dirac cones with the dispersion $E = \pm\varepsilon, \varepsilon = \hbar v_F k$, where $v_F \approx 10^6$ m/s is the Fermi speed (see Chapter 17 for details). To account for the two cones, all coefficients should simply be multiplied by two. Finally, the $\pm$ solutions for the energy should be summed over.

We take the graphene sheet to lie in the (*x*,*y*) plane and for such an isotropic material we may make the replacement $v_x^2 \to \overline{v}^2 \equiv (v_x^2 + v_y^2)/2$.

a) Show that $\overline{v}^2 = v_F^2/2$.

Utilizing this result and the fact that $v_F^2 d^2k = v_F^2 2\pi k dk = 2\pi\varepsilon d\varepsilon / \hbar^2$ means that

$$L_{11} = -\frac{e^2}{\pi\hbar^2}\int_0^\infty \tau\{f'(\varepsilon) + f'(-\varepsilon)\}\varepsilon d\varepsilon .$$

To express the results we introduce $\varphi_F \equiv E_F / kT$. Also, to evaluate the required integrals you will need

$$\int_0^\infty f'(\pm\varepsilon)\varepsilon d\varepsilon = -kT\ln[1 + \exp(\pm\varphi_F)]$$

$$\int_0^\infty (\pm\varepsilon - E_F) f'(\pm\varepsilon)\varepsilon d\varepsilon = (kT)^2\{\varphi_F \ln[1 + \exp(\pm\varphi_F)] \pm \mathrm{Li}_2[-\exp(\pm\varphi_F)]\},$$

where Li_2 is a so-called polylogarithm.

b) Assume τ independent of energy to show that

$$L_{11} = \frac{2e^2\tau kT}{\pi\hbar^2}\ln[2\cosh(\varphi_F/2)].$$

Quite similarly, we find

$$L_{12} = \frac{e\tau}{\pi\hbar^2 T}\int_0^\infty \{(\varepsilon - E_F) f'(\varepsilon) - (\varepsilon + E_F) f'(-\varepsilon)\}\varepsilon d\varepsilon \,.$$

c) Show that

$$L_{12} = \frac{2e\tau k}{\pi\hbar^2}\left\{E_F \ln\left[2\cosh(\varphi_F/2)\right] + kT\mathrm{Li}_2\left[-\exp(\varphi_F)\right] - kT\mathrm{Li}_2\left[-\exp(-\varphi_F)\right]\right\}.$$

For positive Fermi energy and moderately low temperatures $\varphi_F \gg 1$. Thus, we may expand using

$$\ln\left[2\cosh(\varphi_F/2)\right] \approx \frac{\varphi_F}{2}, \quad \mathrm{Li}_2\left[-\exp(\varphi_F)\right] \approx -\frac{\varphi_F^2}{2} - \frac{\pi^2}{6}, \quad \mathrm{Li}_2\left[-\exp(-\varphi_F)\right] \approx 0.$$

d) Use the expansion to demonstrate that at low temperature

$$L_{11} \approx \frac{e^2\tau E_F}{\pi\hbar^2}, \quad L_{12} \approx -\frac{\pi e\tau k^2 T}{3\hbar^2}, \quad Q \approx -\frac{\pi^2 k^2 T}{3eE_F}.$$

These results could also have been obtained directly using the Sommerfeld expansion.

References

[1] N.W. Ashcroft and N.D. Mermin *Solid State Physics* (Saunders College, Philadelphia, 1976)

13. Field Effect Transistors

Field effect transistors are arguably the most important semiconductor devices ever fabricated. They form the basis of logical circuits and largely paved the way for the computer revolution. This chapter is aimed at providing a physical understanding of field effect transistors and, in particular, the dependence of current on gate and drain voltages. Both three-dimensional (MOSFETs) and two-dimensional (MODFETs) transistors will be studied. The MOSFET geometry is illustrated in Fig. 13.1.

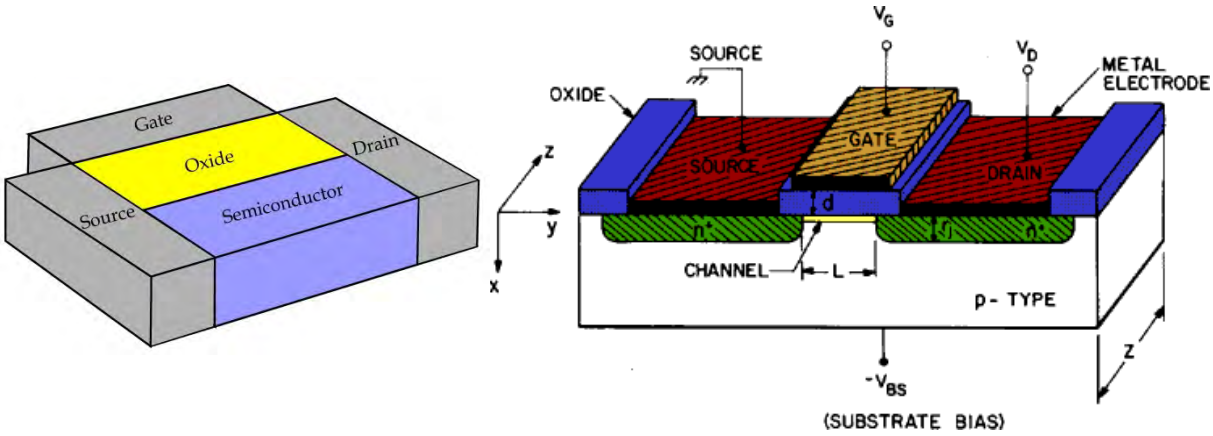


Figure 13.1. Schematic of the MOSFET. Left: The central part is the channel connecting source and drain. The channel is separated from the metallic gate by a thin oxide. Taken from [1]. Right: Simplified model.

The source electrode and substrate are grounded but gate and drain are biased. We need to calculate the electrostatic potential inside the semiconductor. To this end, we consider the potential diagram in Fig. 13.2.

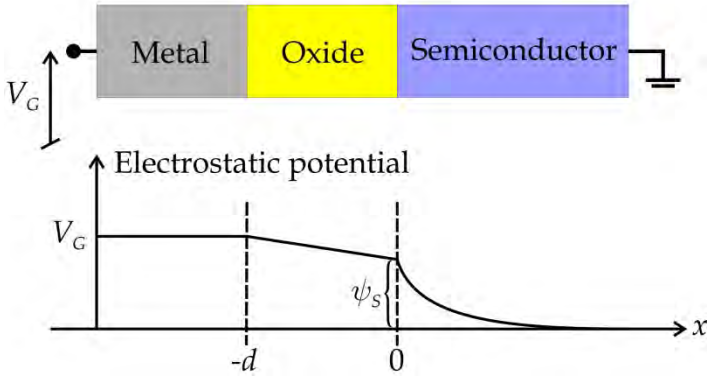


Figure 13.2. Variation of the electrostatic potential across metal, oxide and semiconductor.

The spatial variation of the electrostatic potential ψ inside the semiconductor is governed by the Poisson equation

$$\frac{d^2\psi}{dx^2} = -\frac{\rho}{\varepsilon\varepsilon_0},$$

where ρ is the spatially varying charge density and ε is the relative dielectric constant of the material. The boundary conditions are $\psi(0)=\psi_S$ (the surface potential) and $\psi(\infty)=0$. We consider a *p*-type semiconductor so that the charge density is given as

$$\rho=-e\{n-p+N_A\}, \tag{13.1}$$

where N_A is the density of acceptor impurities. The hole density follows the potential according to $p=p_0\exp(-e\psi/kT)$. In the absence of a source-drain voltage, the electron density would simply vary as $n=n_0\exp(e\psi/kT)$. By requiring charge neutrality at infinity in this situation, it follows from the boundary condition $\psi(\infty)=0$ that $n_0-p_0+N_A=0$. However, with the source grounded and the drain at a potential $V_D\neq 0$, a drain current flows and we no longer have thermal equilibrium. As the name says, the drain contact *drains* electrons from the semiconductor. As a consequence, the electron density at the source n_S is still given by $n_S=n_0\exp(e\psi/kT)$. At the drain, however, the effective Fermi level (or quasi-Fermi level) is lowered by eV_D and so the electron density is *reduced* to $n_D=n_0\exp(e(\psi-V_D)/kT)$. Between source and drain, the density varies according to $n=n_0\exp(e(\psi-V)/kT)$, where the counter potential V varies between 0 at the source and V_D at the drain (holes are not affected due to the $n+$ regions that block hole transport c.f. Fig. 13.1). Hence, writing $N_A=p_0-n_0$ the charge density is

$$\rho=-e\left\{n_0\left[\exp\left(\frac{e(\psi-V)}{kT}\right)-1\right]-p_0\left[\exp\left(-\frac{e\psi}{kT}\right)-1\right]\right\}. \tag{13.2}$$

At this point it is highly convenient to introduce normalized potentials $\varphi=e\psi/kT$, $\varphi_S=e\psi_S/kT$, $v=eV/kT$ and $v_G=eV_G/kT$. Hence, the Poisson equation yields

$$\frac{d^2\varphi}{dx^2}=\frac{e^2}{\varepsilon\varepsilon_0 kT}\left\{n_0\left[e^{\varphi-v}-1\right]-p_0\left[e^{-\varphi}-1\right]\right\}.$$

This differential equation is not solvable but a first integral can be obtained by the standard procedure. Thus,

$$\frac{d\varphi}{dx}=-\left(\frac{2e^2n_0}{\varepsilon\varepsilon_0 kT}\right)^{1/2} g(\varphi,v),\quad g(\varphi,v)\equiv\left\{e^{\varphi-v}-\varphi-e^{-v}+p_0/n_0\left[e^{-\varphi}+\varphi-1\right]\right\}^{1/2}. \tag{13.3}$$

Here and throughout, it is understood that the positive sign of the square root in *g* is to be used if $\psi_S>0$ and the negative if $\psi_S<0$. The boundary condition relating the potentials in the oxide and semiconductor is

$$\varepsilon_{OX}\left.\frac{d\varphi}{dx}\right|_{x=0_-} = \varepsilon\left.\frac{d\varphi}{dx}\right|_{x=0_+}.$$

The slope of the normalized potential inside the oxide is simply $(\varphi_S - v_G)/d$. Hence, $\varepsilon_{OX}\varepsilon_0\, d\varphi/dx\big|_{x=0_-} = C_{OX}(\varphi_S - v_G)$, where $C_{OX} = \varepsilon_{OX}\varepsilon_0/d$ is the capacitance per area of the oxide layer. The right-hand side is given by Eq.(13.3) and in combination we get

$$v_G = \varphi_S + \gamma g(\varphi_S, v), \tag{13.4}$$

where $\gamma = \left(2\varepsilon\varepsilon_0 n_0/kT\right)^{1/2} e/C_{OX}$. For given parameters, Eq.(13.4) must be solved numerically in order to determine the surface potential. We now want to determine the electron fraction of the total charge. At any given y point, the electron charge per area Q_n is given by

$$Q_n = -e\int_0^\infty n dx = -en_0\int_0^\infty e^{\varphi - v} dx.$$

By substitution, this result can be rewritten as

$$Q_n = -en_0\int_{\varphi_S}^{0} e^{\varphi - v}\frac{dx}{d\varphi}d\varphi = -\left(\frac{\varepsilon\varepsilon_0 n_0 kT}{2}\right)^{1/2}\int_0^{\varphi_S}\frac{e^{\varphi - v}}{g(\varphi, v)}d\varphi. \tag{13.5}$$

In Fig. 13.3, the solution of Eq.(13.4) is plotted together with the electron charge Eq.(13.5) assuming room temperature (kT = 0.026 eV) and using silicon parameters: ε = 11.9, $p_0 = 10^{20}$ m^{-3}, $n_0 = 10^{12}$ m^{-3} and $\gamma = 2\cdot 10^{-6}$.

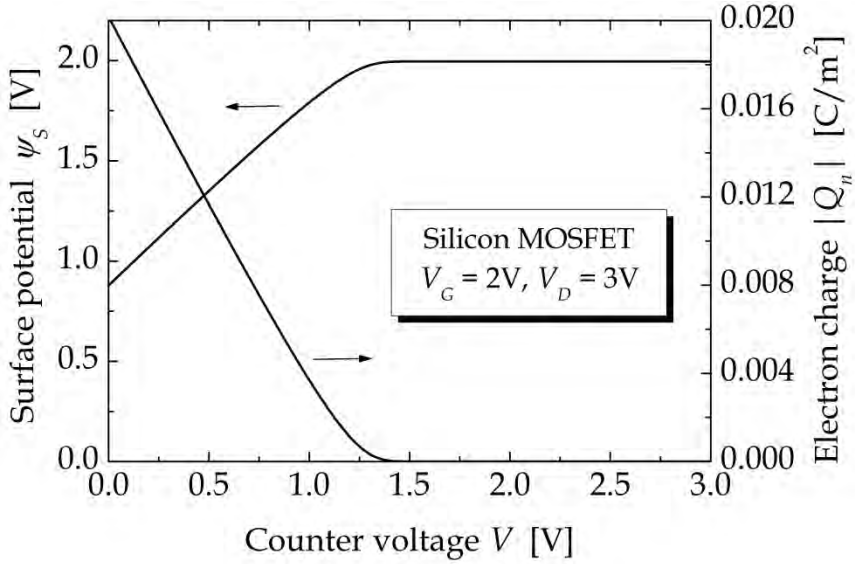


Figure 13.3. Surface potential and electron charge vs. counter potential for a Si MOSFET.

13.1 MOSFET I/V Characteristic

At any given point, the y component of the current density J_y consists of a drift and a diffusion part

$$\begin{aligned} J_y &= -e\mu_e n\frac{d\psi}{dy} + \mu_e kT\frac{dn}{dy} \\ &= -e\mu_e n\frac{dV}{dy}, \end{aligned}$$

using $n = n_0\exp(e(\psi - V)/kT)$. Hence, for a device of thickness Z in the z-direction we find that the drain current follows as

$$I_D = -Z\int_0^\infty J_y dx = -\mu_e Z Q_n\frac{dV}{dy} = -\frac{\mu_e ZkT}{e}Q_n\frac{dv}{dy}.$$

Our problem here is that we don't know how v depends on y. Following Pao and Sah [2], one possible way around this obstacle is to integrate over the length of the channel L

$$\begin{aligned} I_D &= -\frac{\mu_e ZkT}{eL}\int_0^L Q_n\frac{dv}{dy}dy = -\frac{\mu_e ZkT}{eL}\int_0^{v_D} Q_n dv \\ &= \frac{\mu_e ZkT}{eL}\left(\frac{\varepsilon\varepsilon_0 n_0 kT}{2}\right)^{1/2}\int_0^{v_D}\int_0^{\varphi_S}\frac{e^{\varphi - v}}{g(\varphi, v)}d\varphi dv. \end{aligned}$$

Note, that this expression is exact because I_D is independent of y and so "averaging" over y doesn't change the result. Thus, we can finally write

$$I_D = I_{MOSFET}\int_0^{v_D}\int_0^{\varphi_S}\frac{e^{\varphi - v}}{g(\varphi, v)}d\varphi dv, \quad I_{MOSFET} \equiv \frac{\mu_e Z}{L}\left(\frac{\varepsilon\varepsilon_0 n_0 (kT)^3}{2e^2}\right)^{1/2}. \tag{13.6}$$

Hence, to compute the current in Eq.(13.6) we first need to solve Eq.(13.4) for a given v_G to obtain φ_S as a function of v. This method provides a very accurate expression for the current but due to the two integrals that have to be done numerically, this method is also rather cumbersome. An alternative approximate expression can be found by observing that the total charge Q is the sum of the electron charge Q_n and a depletion charge Q_d. The latter is approximately $Q_d \approx -\left(2e\varepsilon\varepsilon_0 p_0\right)^{1/2}\sqrt{\psi_S}$. Thus,

$$I_D \approx -\frac{\mu_e Z}{L}\left\{\int_0^{V_D} Q dV + \left(2e\varepsilon\varepsilon_0 p_0\right)^{1/2}\int_0^{V_D}\sqrt{\psi_S}dV\right\}.$$

Now, Q is given as $Q = C_{OX}(\psi_S - V_G)$. We still don't know the dependence of ψ_S on V, however. Consequently, as a simplification we'll assume that the surface potential varies in direct proportion to the counter-potential, i.e. $d\psi_S / dV \approx 1$. This implies that we change integration variable from V to ψ_S, provided integration limits are also changed. Hence, as V varies between 0 and V_D, the surface potential varies between ψ_{S0} and ψ_{SL}. It follows that the current integral becomes

$$I_D \approx -\frac{\mu_e Z}{L}\left\{C_{OX}\int_{\psi_{S0}}^{\psi_{SL}}(\psi_S - V_G)d\psi_S + \left(2e\varepsilon\varepsilon_0 p_0\right)^{1/2}\int_{\psi_{S0}}^{\psi_{SL}}\sqrt{\psi_s}\,d\psi_S\right\}$$

$$= -\frac{\mu_e Z C_{OX}}{L}\left\{\frac{1}{2}\left(\psi_{SL}^2 - \psi_{S0}^2\right) - V_G\left(\psi_{SL} - \psi_{S0}\right) + \frac{2\left(2e\varepsilon\varepsilon_0 p_0\right)^{1/2}}{3C_{OX}}\left(\psi_{SL}^{3/2} - \psi_{S0}^{3/2}\right)\right\}.$$

In terms of I_{MOSFET} defined in Eq.(13.6), the current finally becomes

$$I_D \approx \frac{2I_{MOSFET}}{\gamma}\left\{v_G\left(\varphi_{SL} - \varphi_{S0}\right) - \frac{1}{2}\left(\varphi_{SL}^2 - \varphi_{S0}^2\right) - \frac{2}{3}\gamma\left(\frac{p_0}{n_0}\right)^{1/2}\left(\varphi_{SL}^{3/2} - \varphi_{S0}^{3/2}\right)\right\}. \quad (13.7)$$

It is clear that to evaluate this expression only the surface potentials at the ends of the channel are needed. A comparison of full and approx. expressions is shown in Fig. 13.4.

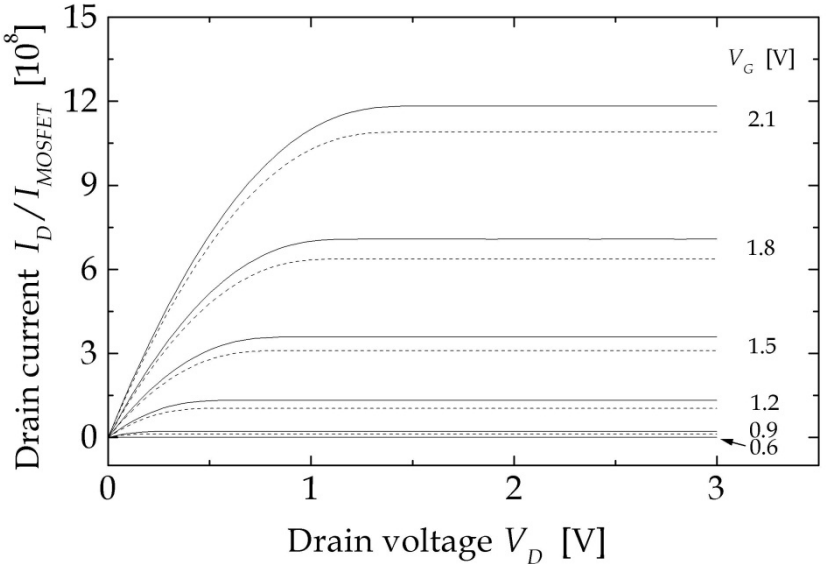


Figure 13.4. Drain current versus drain voltage for different values of the gate potential. Solid and dashed curves represent full and approximate calculations, respectively.

Parameters are $\gamma = 2 \cdot 10^{-6}$ and $p_0 / n_0 = 10^8$.

It is clear from Fig. 13.4 that a threshold condition for the gate voltage exists. In the figure, the threshold apparently is about 0.6 V. The requirement for inversion is that

$e^{\varphi} \geq p_0 / n_0 \equiv e^{2\varphi_F}$, where $\varphi_F = \frac{1}{2}\ln(p_0 / n_0) = \ln(N_A / n_i)$ is the bulk Fermi level. Following Eq.(13.4), the normalized gate voltage needed to achieve inversion is

$$v_G \geq v_{GT} \equiv 2\varphi_F + \gamma g(2\varphi_F, 0) \approx 2\varphi_F + e^{\varphi_F}\gamma\sqrt{2\varphi_F}\,.$$

In fact, a slightly different threshold of $v_{GT} = 2\varphi_F + 2e^{\varphi_F}\gamma\sqrt{\varphi_F}$ is closer to the numerical value. For the parameters of Fig. 13.3 we find $v_{GT} \approx 25$ or $V_{GT} \approx 0.64$ V. Above threshold, the saturation value of the drain current $I_{D,Sat}$ is quite closely matched by the parabolic approximation $I_{D,Sat} \approx I_{MOSFET} \times \frac{0.8}{\gamma}(v_G - v_{GT})^2$.

13.2 Modulation Doped Field Effect Transistors

Modulation doped field effect transistors (MODFETs) also go by the names HEMT (High Electron Mobility Transistor) and HFET (Heterostructure FET). Several differences between MOSFETs and MODFETs are worth noting:

1. The oxide is replaced by an *n*-doped $Al_xGa_{1-x}As$ layer
2. The semiconductor is undoped GaAs
3. Band bending is sufficiently strong that electrons donated by the $Al_xGa_{1-x}As$ layer accumulate in a very thin layer on the GaAs side of the junction.

The geometry of the structure is illustrated in Fig. 13.5.

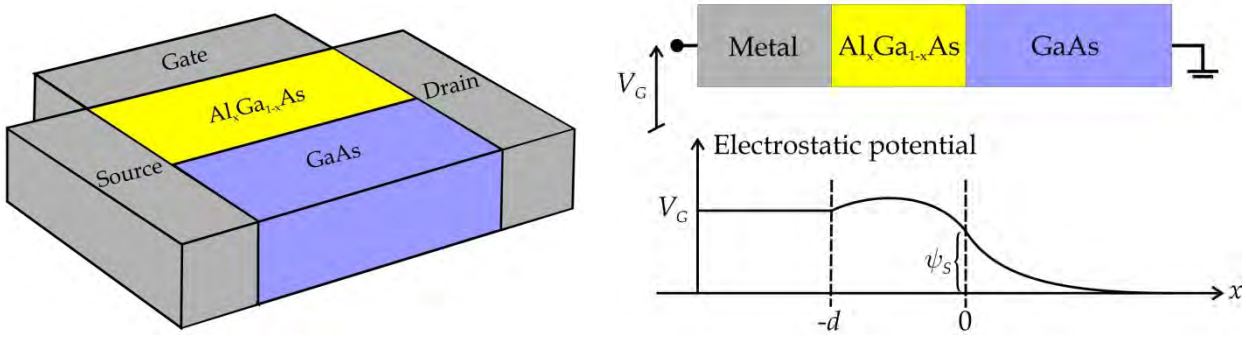


Figure 13.5. Schematic structure of a MODFET. The right-hand diagram illustrates the profile of the electrostatic potential

The donor density in the $Al_xGa_{1-x}As$ layer is denoted N_D and we will assume that all donors are ionized and the electrons transferred to the GaAs side. For simplicity, we will assume that all electrons occupy a single state with quantization energy E_0. Also, we will ignore the variation of the potential across the quantum well and simply take it to be ψ_S, the surface potential. In this case, the electron density per area is

$$n = \frac{m_e kT}{\pi\hbar^2}\ln\left\{1+\exp\left(\frac{e\psi_S + E_F - E_0}{kT}\right)\right\}$$
$$= \frac{m_e kT}{\pi\hbar^2}\ln\left\{1+e^{\varphi_S+\varphi_{F0}}\right\},\quad \varphi_{F0} \equiv \frac{E_F - E_0}{kT}.$$

A consequence of the $Al_xGa_{1-x}As$ doping is that the layer becomes an "imperfect" capacitor. In a normal capacitor, the potential drops linearly across the layer but in the present case the Poisson equation for the layer (using the normalized potential) becomes

$$\frac{d^2\varphi}{dx^2} = -\frac{e^2 N_D}{\varepsilon\varepsilon_0 kT}.$$

Here, we have ignored the small difference in dielectric constant between the two semiconductors and denoted the common value by ε. The general solution is obviously a parabola $\varphi(x) = a + bx + cx^2$. With the boundary conditions $\varphi(-d) = v_G$ and $\varphi(0) = \varphi_S$ it follows that

$$\varphi(x) = v_G + (\varphi_S - v_G)\frac{x+d}{d} - \frac{e^2 N_D}{2\varepsilon\varepsilon_0 kT}x(x+d).$$

As the dielectric constant is assumed independent of position, it follows that the derivative of the potential is continuous across the boundary. Hence, just inside the GaAs side, the potential derivative $\varphi_S' \equiv \varphi'(0_+)$ is given by

$$\varphi_S' = \frac{\varphi_S - v_G}{d} - \frac{\varphi_D}{d},\quad \varphi_D \equiv \frac{e^2 N_D d^2}{2\varepsilon\varepsilon_0 kT}. \tag{13.8}$$

If a drain voltage is applied, charge neutrality is broken. In this case, the counter voltage V varies between 0 and V_D at source and drain, respectively. Accordingly, the electron charge is now

$$Q_n = -\frac{e m_e kT}{\pi\hbar^2}\ln\left\{1+e^{\varphi_S - v + \varphi_{F0}}\right\}, \tag{13.9}$$

This charge is the total charge per area in the GaAs layer. We may consequently apply Gauss' law to a box enclosing the GaAs layer and write $\varphi_S' = eQ_n / kT\varepsilon\varepsilon_0$. Combining the results above we finally find

$$\frac{\varphi_S - v_G}{d} - \frac{\varphi_D}{d} = -\frac{e^2 m_e}{\pi\hbar^2\varepsilon\varepsilon_0}\ln\left\{1+e^{\varphi_S - v + \varphi_{F0}}\right\}.$$

Introducing the auxiliary potential $\varphi_e \equiv \frac{e^2 m_e d}{\pi\hbar^2 \varepsilon\varepsilon_0}$ the balance equation can be written

$$\varphi_S - v_G - \varphi_D = -\varphi_e \ln\left\{1 + e^{\varphi_S - v + \varphi_{F0}}\right\}. \tag{13.10}$$

This is the equation we need to solve to compute φ_S as a function of v. First, however, the location of the Fermi level must be determined. This is a complicated problem related to "pinning" by the $Al_xGa_{1-x}As$ donors. For simplicity, therefore, the Fermi energy can be determined by requiring charge balance whenever the gate and drain biases are absent. In this case, electron charges must be balanced by donor charges in a $Al_xGa_{1-x}As$ layer of thickness d, and so we have $dN_D = n$, which means that

$$dN_D = \frac{m_e kT}{\pi\hbar^2} \ln\left\{1 + e^{\varphi_{S0} + \varphi_{F0}}\right\}, \tag{13.11}$$

where φ_{S0} is the surface potential in the absence of biases. In terms of the auxiliary potentials defined above, this means that $2\varphi_D / \varphi_e = \ln\left\{1 + e^{\varphi_{S0} + \varphi_{F0}}\right\}$. On the other hand, Eq.(13.10) with $v_G = v = 0$ shows that $\varphi_{S0} - \varphi_D = -\varphi_e \ln\left\{1 + e^{\varphi_{S0} + \varphi_{F0}}\right\}$. Taken together, it follows that $\varphi_{S0} = -\varphi_D$ and so the balance equation finally becomes

$$\varphi_S - v_G - \varphi_D = -\varphi_e \ln\left\{1 + e^{\varphi_S - v + \varphi_D}\left(e^{2\varphi_D/\varphi_e} - 1\right)\right\}. \tag{13.12}$$

An example of the surface potential profile is shown in Fig. 13.6. Here, the following parameters representative of GaAs MODFETs have been applied:

- $m_e = 0.067\ m$
- $N_D = 10^{23}\ \mathrm{m}^{-3}$
- ε = 12.9
- $Al_xGa_{1-x}As$ layer thickness: $d = 30$ nm

These values correspond to $\varphi_D = 2.43$ and $\varphi_e = 11.87$.

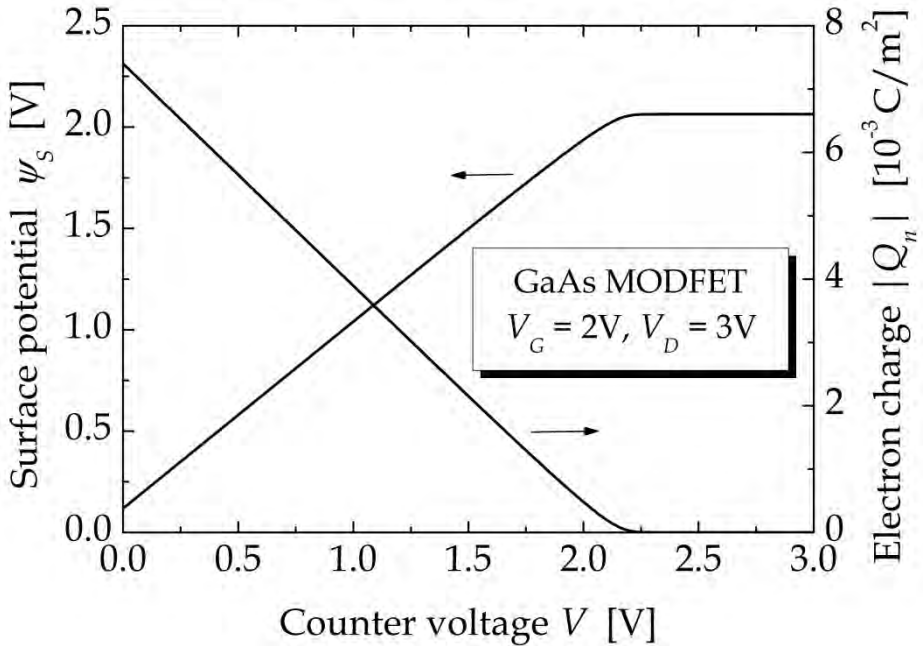


Figure 13.6. Surface potential and electron charge vs. counter potential for a GaAs MODFET.

13.3 MODFET I/V Characteristic

Similarly to the MOSFET, the drain current is given as $I_D = -\mu_e Z Q_n kT / e \cdot dv / dy$. To avoid computing dv / dy we again integrate over y and find

$$I_D = -\frac{\mu_e Z k T}{eL} \int_0^L Q_n \frac{dv}{dy} dy = -\frac{\mu_e Z k T}{eL} \int_0^{v_D} Q_n dv .$$

Now, using $\varphi_S' = eQ_n / kT\varepsilon\varepsilon_0$ as well as Eq.(13.8) we can write

$$I_D = I_{MODFET} \int_0^{v_D} \{v_G + \varphi_D - \varphi_S\} dv, \quad I_{MODFET} \equiv \frac{\varepsilon\varepsilon_0 \mu_e Z (kT)^2}{e^2 L d}. \tag{13.13}$$

In Fig. 13.7, the I/V characteristic for the GaAs MODFET is illustrated.

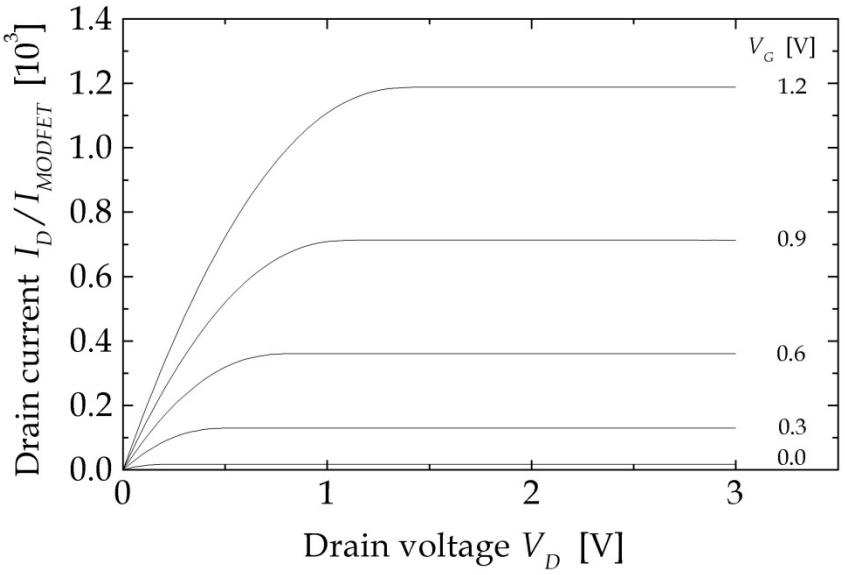


Figure 13.7. Drain current versus drain voltage for different values of the gate potential in a GaAs MODFET at room temperature.

Again, a pronounced saturation of the current is found as the drain voltage increases. As for the MOSFET, an analytical estimate of the saturation current can be given. First, we note from Fig. 13.6 that to a good approximation the surface potential profile is piecewise linear. Ignoring the factor "1" inside the curly brackets of Eq.(13.10) we therefore find

$$\varphi_S \approx \begin{cases} \dfrac{\varphi_e(v-\varphi_{F0})+v_G}{1+\varphi_e}, & v < v_G+\varphi_{F0} \\ v_G \qquad\qquad\quad , & v \geq v_G+\varphi_{F0} \end{cases} .$$

If this approximation is used and $v_D \geq v_G+\varphi_{F0}$ is assumed, we find for the saturation current $I_{D,Sat} \approx \frac{1}{2} I_{MODFET}(v_G+\varphi_{F0})^2 \varphi_e/(1+\varphi_e)$ as can easily be verified. Hence, just as for the MOSFET, the saturation current varies quadratically with gate voltage. Comparing Figs. 13.4 and 13.7 it would appear that the MOSFET current is much larger than the MODFET current. However, it is easily shown that

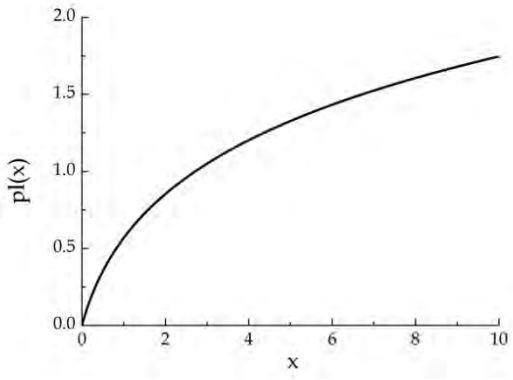
$$\frac{I_{MODFET}}{I_{MOSFET}} = \left(\frac{N_D}{n_0\varphi_D}\right)^{1/2} .$$

Using the same parameters as above, this ratio equals $2\cdot 10^5$ and so the two currents are, in fact, roughly equal.

13.4 Analytical MOSFET/MODFET Model

The models of the last sections are rather accurate and general but, also, mathematically complicated and involve solving balance equations numerically. We now try to establish a very simple, approximate model that can be applied to illustrate the essential physics of the devices without sophisticated mathematics. The model relies on three simplifications: (1) The conducting channel is infinitely thin, (2) the carrier statistics are "non-degenerate", and (3) hole contributions to charge and current can be ignored. Non-degenerate statistics means that we replace the Fermi distribution by $f(E) \approx \exp(-(E-E_F)/kT)$. The combination of these assumptions leads to the fact that $n = n_0 \exp(\varphi_S - v)$ is the total carrier concentration with φ_S the value of the potential in the thin channel. We will assume a gate oxide with no net charge and it then follows that the charge balance for the gate capacitor must be $C_{OX}(v_G - \varphi_S) = e^2 n / kT = e^2 n_0 / kT \exp(\varphi_S - v)$. This is a rather simple problem with the solution

$$\varphi_S = v_G - \mathrm{pl}\left(\frac{e^{v_G - v} e^2 n_0}{C_{OX} kT}\right).$$



Here, pl is the "product logarithm" defined as the solution to the equation $w\exp(w) = x$, i.e. $w = \mathrm{pl}(x)$. Also, using the results of the previous section in the non-degenerate limit $n_0 \approx m_e kT e^{\varphi_{F0}} / (\pi\hbar^2)$, it is easily shown that $e^2 n_0 / (C_{OX} kT) = 2\varphi_D e^{\varphi_D}$ with φ_D given by Eq.(13.8). An important property of the product logarithm is that $\int e^{-x} \exp\{-\mathrm{pl}(ce^{-x})\} dx = -\frac{1}{2c}\mathrm{pl}(ce^{-x})\{\mathrm{pl}(ce^{-x}) + 2\}$. We apply this formula to compute the drain current

$$I_D = \frac{\mu_e Z kT}{L} \int_0^{v_D} n dv$$
$$= \frac{1}{2} I_{MODFET} \left\{\mathrm{pl}\left(2\varphi_D e^{\varphi_D + v_G}\right) - \mathrm{pl}\left(2\varphi_D e^{\varphi_D + v_G - v_D}\right)\right\}\left\{2 + \mathrm{pl}\left(2\varphi_D e^{\varphi_D + v_G}\right) + \mathrm{pl}\left(2\varphi_D e^{\varphi_D + v_G - v_D}\right)\right\}.$$

with I_{MODFET} given by Eq.(13.13). The comparison between the full numerical model and the simplified, analytical one is shown in Fig. 13.8. We see that the results agree to a very high degree of precision.

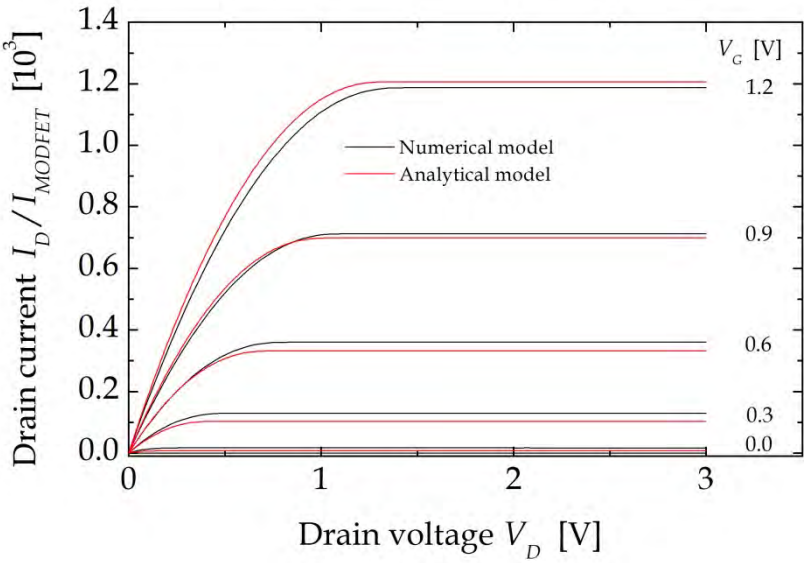


Figure 13.8. Comparison of numerical and analytical MODFET models.

13.5 Transit Time

For both MOSFETs and MODFETs, the transit time τ of an electron traveling between source and drain can be computed as the integral

$$\tau = \int_0^L \frac{1}{u} dy = \int_0^{V_D} \frac{1}{u} \cdot \frac{dy}{dV} dV .$$

Here, u is the electron velocity $u = \mu_e \cdot dV / dy$. Using $I_D = -\mu_e Z Q_n \cdot dV / dy$, it then follows that

$$\tau = \frac{\mu_e Z^2}{I_D^2} \int_0^{V_D} Q_n^2 dV .$$

We wish to reformulate in terms of normalized potentials and dimensionless quantities. It turns out that for both MOSFETs and MODFETs we can write

$$\tau = \frac{eL^2}{\mu_e kT} \int_0^{v_D} q_n^2 dv \Big/ \left(\int_0^{v_D} q_n dv \right)^2 \tag{13.14}$$

where

$$q_n = \begin{cases} \displaystyle\int_0^{\varphi_S} \frac{e^{\varphi - v}}{g(\varphi, v)} d\varphi, & \text{MOSFET} \\ v_G + \varphi_D - \varphi_S, & \text{MODFET.} \end{cases}$$

To illustrate the results we again consider a Si MOSFET ($\mu_e = 0.145\ \mathrm{m}^2/\mathrm{Vs}$) and a GaAs MODFET ($\mu_e = 0.85\ \mathrm{m}^2/\mathrm{Vs}$). We take in both cases the device length to be $L = 1\ \mu\mathrm{m}$ and otherwise use parameters as above. The results are shown in Figs. 13.9 and 13.10 below. It should be noted that the transit time provides the ultimate physical limit for the cut-off frequency of the devices given by $f_{cut-off} = 1/(2\pi\tau)$. Hence, for the two devices we find approximate cut-offs of 16 GHz and 160 GHz, respectively. In a real device, the value will be somewhat lower due to capacitive effects.

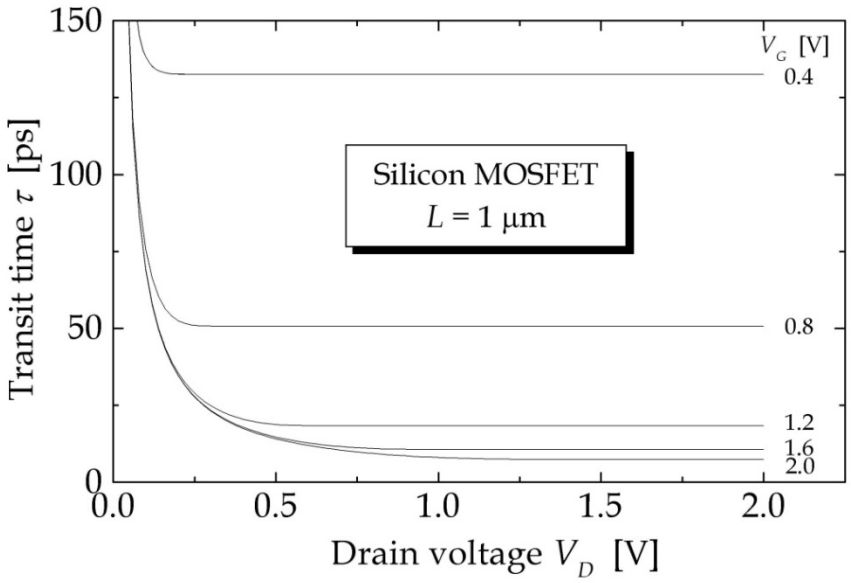


Figure 13.9. Transit time vs. drain voltage for different values of the gate potential in a Si MOSFET.

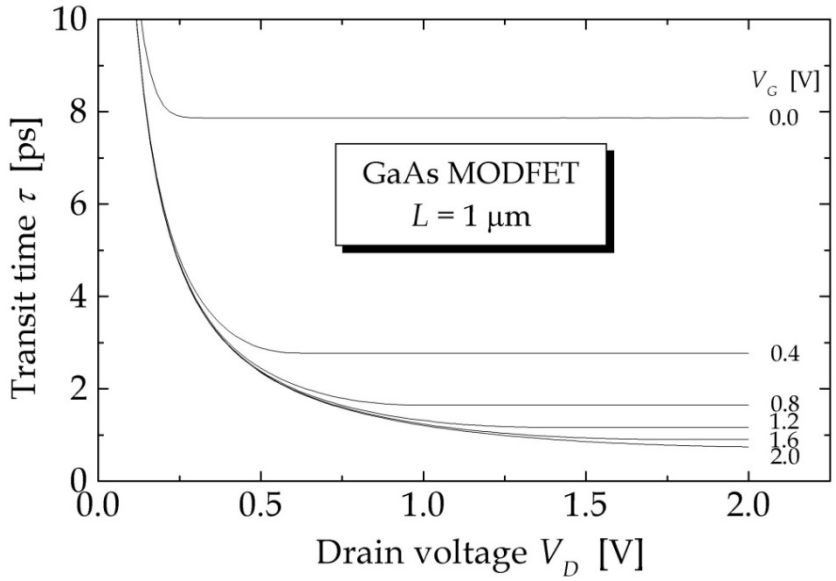


Figure 13.10. Transit time vs. drain voltage for different values of the gate potential in a GaAs MODFET.

Exercise: Variation of the counter potential

An unsolved problem is the y-dependence of the counter potential. Here, we will show how this may be computed.

a) Show that for both MOSFETs and MODFETs $I_D = I_{MOSFET/MODFET} \cdot L q_n \frac{dv}{dy}$, where

$$q_n = \begin{cases} \int_0^{\varphi_S} \frac{e^{\varphi - v}}{g(\varphi, v)} d\varphi, & \text{MOSFET} \\ v_G + \varphi_D - \varphi_S, & \text{MODFET.} \end{cases}$$

b) Show by rearrangement and integration that

$$\frac{y}{L} = \int_0^v q_n dv \Big/ \int_0^{v_D} q_n dv \,.$$

c) Integrate numerically to find $V(y)$. If successful, you should find two highly similar curves as illustrated in Fig. 13.11.

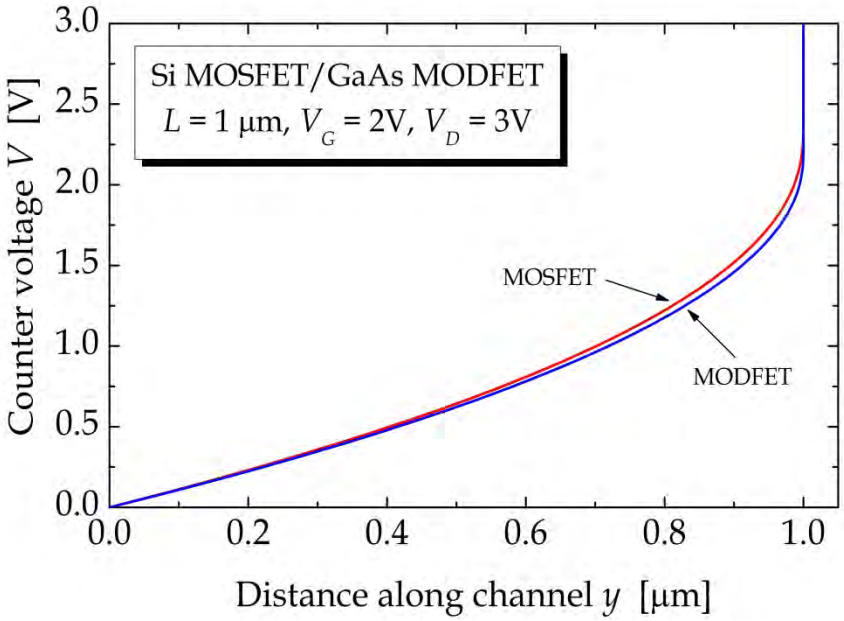


Figure 13.11. Counter voltage vs. y-position for both MOSFET and MODFET.

References

[1] S.M. Sze *Physics of Semiconductor Devices* (Wiley, New York 1981).
[2] H.C. Pao and C.T. Sah, Solid State Electron. 9, 927 (1966).

14. Nanowire MOSFETs

The geometry of a cylindrical nanowire MOSFET is shown in Fig. 14.1. Source and drain contact are attached at the ends and along the nanowire an oxide isolates the semiconductor core from the surrounding metallic gate.

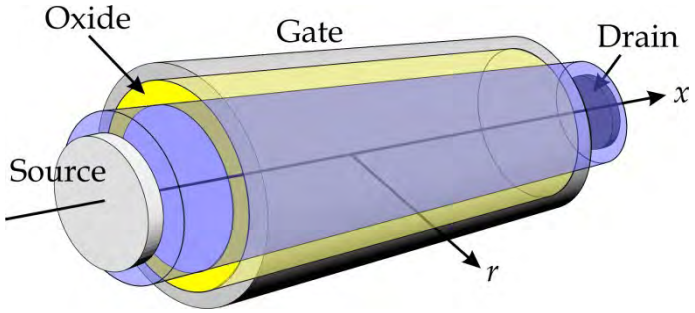


Figure 14.1. Schematic of the nanowire MOSFET. The semiconductor core is surrounded by a thin oxide and a cylindrical metallic gate.

Our analysis will be based on certain assumptions about the structure and the mode of operation: (1) The semiconductor core is undoped and (2) the gate is positively biased so that (practically) only electrons need to be considered and (3) the non-degenerate limit is applicable. In this case, the coupled Poisson-Boltzmann equations read

$$\frac{d^2\psi}{dr^2}+\frac{1}{r}\frac{d\psi}{dr}=\frac{en_0}{\varepsilon\varepsilon_0}\exp\left(\frac{e(\psi-V)}{kT}\right).$$

The boundary conditions for the potential $\psi(r)$ are $\psi'(0)=0$ and $\psi(0)=\psi_0$, where the center value ψ_0 will be determined later. As usual we introduce normalized potentials $\varphi=e\psi/kT$ and $v=eV/kT$ so that we obtain the reduced Poisson-Boltzmann equation

$$\frac{d^2\varphi}{dr^2}+\frac{1}{r}\frac{d\varphi}{dr}=\frac{e^2n_0}{\varepsilon\varepsilon_0 kT}e^{\varphi-v}\equiv N_0e^{\varphi-v},\quad N_0=\frac{e^2n_0}{\varepsilon\varepsilon_0 kT} \tag{14.1}$$
$$\text{Boundary conditions:}\quad \varphi'(0)=0,\quad \varphi(0)=\varphi_0.$$

It can be shown by inspection that an analytical solution to this equation is given by

$$\varphi(r)=\varphi_0-2\ln\left(1-\tfrac{1}{8}N_0e^{\varphi_0-v}r^2\right). \tag{14.2}$$

We also need the potential inside the oxide. We denote radius of the semiconductor by R and the oxide thickness by d. Thus, we require $\varphi(R)=\varphi_S$ and $\varphi(R+d)=v_G$. The potential inside the (neutral) oxide is given by the homogenous solution to the

Poisson equation $\varphi(r) = a \ln r + b$. Subtracting the boundary conditions for the two faces of the oxide then means that $a = (v_G - \varphi_S)/\ln(1 + d/R)$. In addition, the slope of the oxide potential is $d\varphi/dr = a/r$ and we can therefore relate slopes on both sides of the oxide-semiconductor interface via

$$\varepsilon_{OX} \left.\frac{d\varphi}{dr}\right|_{r=R_+} = \varepsilon \left.\frac{d\varphi}{dr}\right|_{r=R_-},$$

where ε_{OX} and ε are dielectric constants of the oxide and semiconductor, respectively. This, in turn, means that

$$\frac{\varepsilon_{OX}(v_G - \varphi_S)}{R \ln(1 + d/R)} = \varepsilon \left.\frac{d\varphi}{dr}\right|_{r=R_-}.$$

The potential derivative inside the semiconductor is found from Eq.(14.2) and equals

$$\frac{d\varphi}{dr} = \frac{4N_0 e^{\varphi_0 - v} r}{8 - N_0 e^{\varphi_0 - v} r^2}. \tag{14.3}$$

On the other hand, it also follows from Eq.(14.2) that

$$\varphi_S = \varphi_0 - 2\ln\left(1 - \tfrac{1}{8} N_0 e^{\varphi_0 - v} R^2\right).$$

Introducing the oxide capacitance $C_{OX} = \varepsilon_{OX}\varepsilon_0 / R \ln(1 + d/R)$ and putting everything together we finally obtain an equation for the centre potential

$$v_G = \varphi_0 - 2\ln\left(1 - \tfrac{1}{8} N_0 e^{\varphi_0 - v} R^2\right) + \frac{\varepsilon\varepsilon_0}{C_{OX}} \frac{4N_0 e^{\varphi_0 - v} R}{8 - N_0 e^{\varphi_0 - v} R^2}. \tag{14.4}$$

This is the equation to solve to find the centre potential profile along the nanowire as a function of the counter voltage v that varies between 0 and v_D.

We need, in addition, the variation of the electron charge Q_n along the wire. Rather than integrating the electron density over the cross section, we use Gauss' theorem for a surface bounding a small slab of the cylindrical core:

$$\left.\frac{d\varphi}{dr}\right|_{r=R_-} = -\frac{e}{\varepsilon\varepsilon_0 kT} Q_n \equiv \frac{e}{\varepsilon\varepsilon_0 kT} Q.$$

Here, Q_n is the charge per area and $Q \equiv -Q_n$. Along with Eq.(14.3) this means that

$$Q = Q_0 \frac{N_0 e^{\varphi_0 - v} R^2}{8 - N_0 e^{\varphi_0 - v} R^2} \quad \Rightarrow \quad N_0 e^{\varphi_0 - v} R^2 = \frac{8Q}{Q + Q_0},$$

where $Q_0 = 4\varepsilon\varepsilon_0 kT/(eR)$. It follows that

$$\varphi_0 = v + \ln\left\{\frac{8Q}{N_0 R^2 (Q + Q_0)}\right\}. \tag{14.5}$$

This means that the potential balance Eq.(14.4) can be reformulated as a charge balance

$$v_G = v + \ln\left\{\frac{8Q(Q + Q_0)}{N_0 R^2 Q_0^2}\right\} + \frac{eQ}{kTC_{OX}}. \tag{14.6}$$

From a numerical stand point, it is advantageous to solve Eq.(14.6) and subsequently use Eq.(14.5) to compute the potential profile. For later purposes, we note that differentiating this relation leads to

$$\frac{dv}{dQ} = -\frac{2Q + Q_0}{(Q + Q_0)Q} - \frac{e}{kTC_{OX}}. \tag{14.7}$$

An example of an actual example is shown in Fig. 14.2. Here, the following Si/SiO_2 nanowire parameters have been used: ε=11.9, $n_0 = 10^{16}$ m^{-3}, R = 25 nm, d = 5 nm and $\varepsilon_{OX} = 3.9$ leading to $C_{OX} = 7.57 \cdot 10^{-3}$ F/m^2.

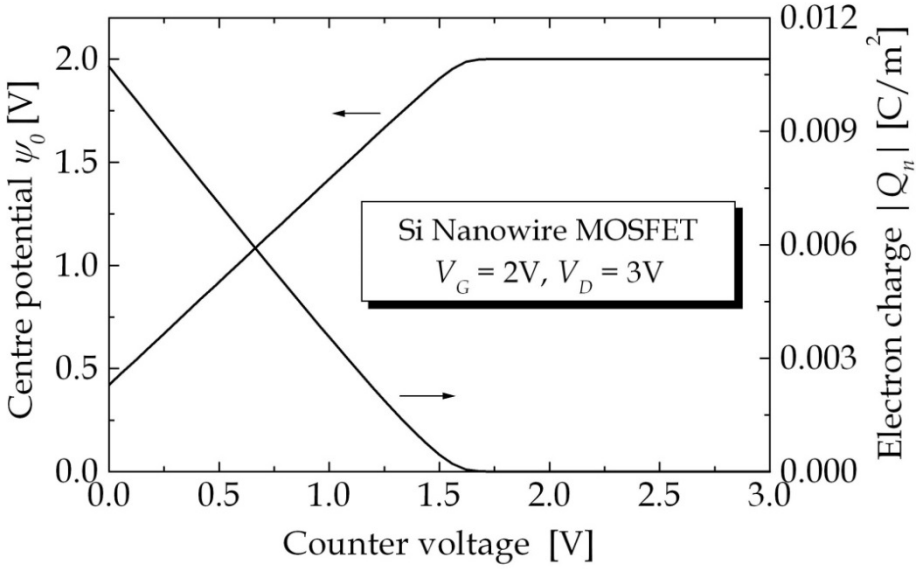


Figure 14.2. The charge and potential profile of a nanowire MOSFET.

We now proceed to calculate the drain current. Similarly to the MOSFET and MODFET cases we write

$$I_D = -\frac{2\pi R\mu_e}{L}\int_0^{V_D} Q_n dV = -\frac{2\pi R\mu_e kT}{eL}\int_0^{v_D} Q_n dv\,.$$

It turns out that integrations are easier if we use charge rather that counter potential as integration variable [1]. Hence, we utilize Eq.(14.7) and write

$$\begin{aligned} I_D &= \frac{2\pi R\mu_e kT}{eL}\int_{Q_S}^{Q_D} Q\frac{dv}{dQ}dQ \\ &= \frac{2\pi R\mu_e kT}{eL}\int_{Q_D}^{Q_S}\left\{\frac{2Q+Q_0}{Q+Q_0}+\frac{eQ}{kTC_{OX}}\right\}dQ. \end{aligned}$$

Here, Q_S and Q_D are (positive) charges at source and drain, respectively. After simple integrations we finally find

$$I_D = \frac{2\pi R\mu_e kT}{eL}\left\{2(Q_S-Q_D)-Q_0\ln\left(\frac{Q_S+Q_0}{Q_D+Q_0}\right)+\frac{e\left(Q_S^2-Q_D^2\right)}{2kTC_{OX}}\right\}. \qquad (14.8)$$

This result is illustrated in Fig. 14.3 below using $\mu_e = 0.145\ \mathrm{m}^2/\mathrm{Vs}$ and $L = 1\ \mu\mathrm{m}$.

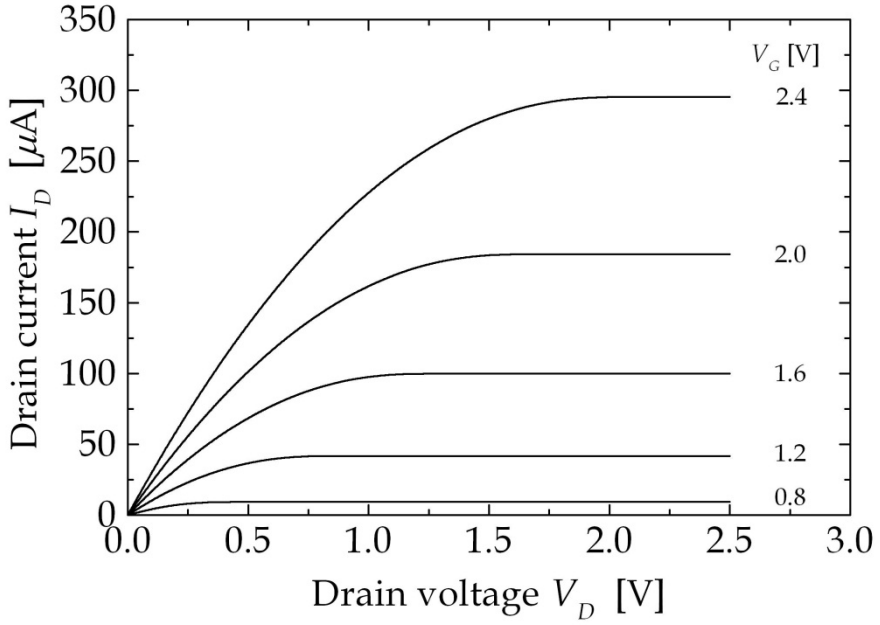


Figure 14.3.Drain I/V characteristic of a 1 micron nanowire MOSFET.

We end this chapter by calculating the transit time τ. In a manner completely analogous to the MOSFET and MODFET cases it turns out that

$$\tau = \frac{eL^2}{\mu_e kT} \int_0^{v_D} Q_n^2 dv \Big/ \left(\int_0^{v_D} Q_n dv \right)^2 .$$

The integral in the numerator is easily calculated using the technique explained above. As a result, we find

$$\tau = \frac{eL^2}{\mu_e kT} \frac{Q_S^2 - Q_D^2 - Q_0(Q_S - Q_D) + Q_0^2 \ln\left(\frac{Q_S + Q_0}{Q_D + Q_0}\right) + \frac{e\left(Q_S^3 - Q_D^3\right)}{3kTC_{OX}}}{\left\{ 2(Q_S - Q_D) - Q_0 \ln\left(\frac{Q_S + Q_0}{Q_D + Q_0}\right) + \frac{e\left(Q_S^2 - Q_D^2\right)}{2kTC_{OX}} \right\}^2} . \qquad (14.9)$$

For a 1 μm device, the result is as shown below.

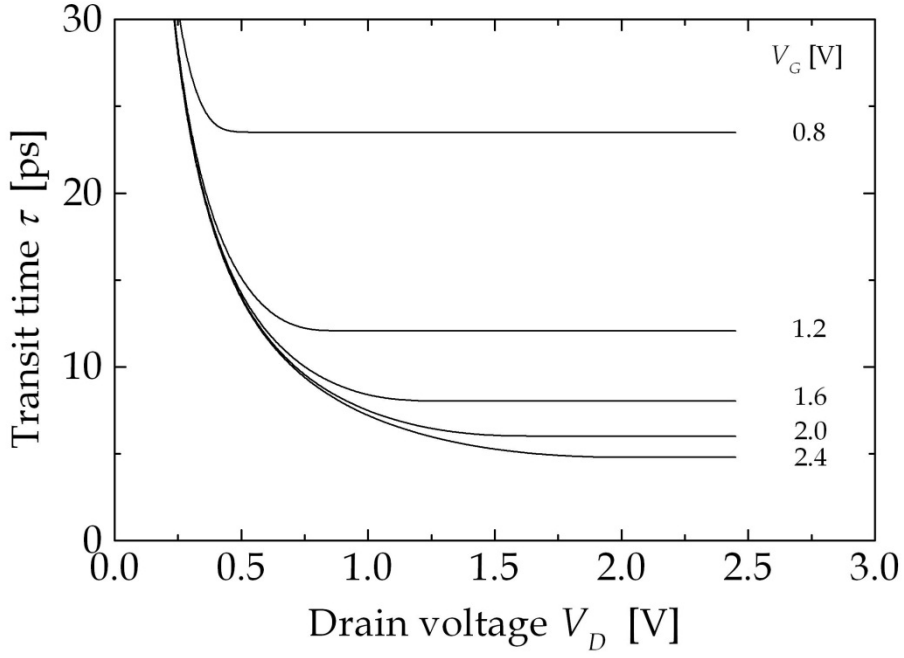


Figure 14.4.Drain and gate dependence of the nanowire MOSFET transit time.

Exercise: InAs nanowire transistors

The Physics Group at Lund University produces nanowire transistors made from InAs (ε = 14.5), see Ref. [2]. The "oxide" in their structures is actually silicon nitride (ε_{OX} = 6.3) with a thickness d = 50 nm. Also, radius and effective length of the nanowires are approximately 40 nm and 1 μm, respectively. We will take the mobility to be $\mu_e \approx 2\ \mathrm{m}^2/\mathrm{Vs}$ and the unbiased electron density to be $n_0 = 10^{23}\ \mathrm{m}^{-3}$. Also, measurements are made on 40 transistors in parallel.

a) To quantify the device sensitivity one sometimes measures the saturation current (the current at high drain voltage) as a function of gate voltage. We expect a quadratic dependence $I_{D,sat} \propto V_G^2$, so that plotting $\sqrt{I_{D,sat}}$ versus V_G should produce a straight line. Use the theory of this chapter to make the plot below by fixing the drain voltage at $V_D = 0.4$ V and calculating for 40 nanowires in parallel. Compare to the experimental plot.

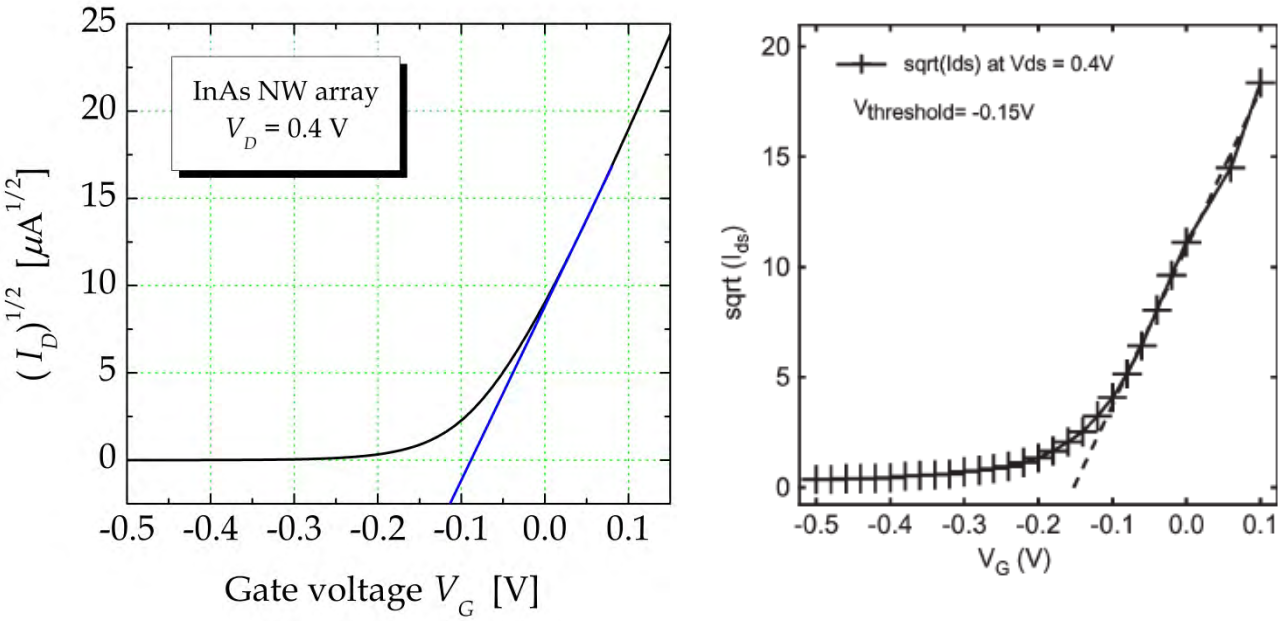


Figure 14.5.Calculated (left) and experimental (right, taken from [2]) sensitivity plot for a 40-nanowire array.

References

[1] B. Iñíguez *et al.*, IEEE Trans. Electron Devices 52, 1868 (2005).
[2] T. Bryllert *et al.*, IEEE Electron Devices Lett. 27, 323 (2006).

15. Optical Properties of Semiconductors

The optical properties of semiconductors are applied in an impressive list of devices: lasers, light-emitting diodes, CCD cameras, solar cells to name a few. Recently, the list has been extended to optical semiconductor devices that actively utilize quantum confinement: quantum well lasers, fluorescent quantum dots, solar cells and photocatalysts based on semiconducting TiO_2 nanoparticles and so on. Also, the traditional inorganic semiconductors such as Si, GaAs and GaP have seen competition from organic or carbon based ones, most notably conjugated polymers and carbon nanotubes. These novel materials are one-dimensional semiconductors with properties that deviate significantly from bulk inorganic semiconductors. Devices such as displays and light-emitting diodes based on these materials are emerging now and may well play an important role in future applications due to reduced cost and possible molecular design. In this chapter, we investigate the optical response of bulk and low-dimensional semiconductors in order to display their differences.

By now, the fundamental approach should be familiar: We need the perturbation and the response observable to calculate the induced response. The perturbation is the well-known interaction between the optical electric field $\vec{\mathcal{E}} = \vec{e}_z \mathcal{E}$ oscillating at a frequency ω and the electric dipole moment $-e\vec{r}$ given by $\hat{H}_1 = e\vec{\mathcal{E}}\cdot\vec{r} = e\mathcal{E}z$. Also, the response observable is the dipole moment density $-ez/\Omega$, where, as always, Ω is the volume. We only consider the z component of the dipole moment, i.e. we restrict the discussion to co-linear cases where the induced dipole moment is parallel to the incident field. The measurable response is the *polarization* $\mathcal{P}(\omega)$ that we consequently find from

$$\begin{aligned}\mathcal{P}(\omega) &= -\frac{1}{\Omega}\sum_{m,n} f_{nm} \frac{\langle\varphi_m|e\mathcal{E}z|\varphi_n\rangle\langle\varphi_n|-ez|\varphi_m\rangle}{E_{mn}-\hbar\omega-i\hbar\Gamma} \\ &= \frac{e^2}{\Omega}\mathcal{E}\sum_{m,n} f_{nm} \frac{\left|\langle\varphi_m|z|\varphi_n\rangle\right|^2}{E_{mn}-\hbar\omega-i\hbar\Gamma}.\end{aligned}$$

The ratio between the polarization and the electric field is the electric susceptibility $\chi(\omega)$ multiplied by ε_0 and so

$$\chi(\omega) = \frac{e^2}{\varepsilon_0\Omega}\sum_{m,n} f_{nm} \frac{\left|\langle\varphi_m|z|\varphi_n\rangle\right|^2}{E_{mn}-\hbar\omega-i\hbar\Gamma}. \tag{15.1}$$

We wish to reformulate this in a form that is suitable for both doped and intrinsic semiconducting periodic structures. To get there, we need a few tricks. First, the fact that we sum over m and n allows us to interchange these indices under the

summation sign. By averaging Eq.(15.1) with an expression having m and n interchanged and using $f_{mn} = -f_{nm}$ and $E_{nm} = -E_{mn}$ we find

$$\chi(\omega) = \frac{e^2}{\varepsilon_0 \Omega} \sum_{m,n} f_{nm} \frac{E_{mn} \left| \langle \varphi_m | z | \varphi_n \rangle \right|^2}{E_{mn}^2 - \hbar^2 (\omega + i\Gamma)^2}.$$

The operator z is difficult to handle in extended systems and we therefore prefer to reformulate the susceptibility expression. To this end, we apply the commutator trick derived in Chapter 2, Eq.(2.2):

$$\langle \varphi_m | z | \varphi_n \rangle = \frac{\hbar}{imE_{mn}} \langle \varphi_m | \hat{p}_z | \varphi_n \rangle .$$

This is obviously dangerous in the case $m = n$, but we will deal with this later. For now, we will simple write

$$\chi(\omega) = \frac{e^2 \hbar^2}{\varepsilon_0 m^2 \Omega} \sum_{m,n} f_{nm} \frac{\left| \langle \varphi_m | \hat{p}_z | \varphi_n \rangle \right|^2}{E_{mn} \left[E_{mn}^2 - \hbar^2 (\omega + i\Gamma)^2 \right]}. \tag{15.2}$$

The case $m \neq n$ is referred to as the *interband* case when dealing with solids. Hence, there are contributions, for which $m > n$, and others with $n > m$. Interchanging m and n in the second half shows that

$$\chi(\omega) = \frac{2e^2 \hbar^2}{\varepsilon_0 m^2 \Omega} \sum_{m>n} f_{nm} \frac{\left| \langle \varphi_m | \hat{p}_z | \varphi_n \rangle \right|^2}{E_{mn} \left[E_{mn}^2 - \hbar^2 (\omega + i\Gamma)^2 \right]}.$$

We will now specialize to the description of intrinsic semiconductors, in which the empty states are separated from the occupied ones by a large energy gap E_g. The Fermi level in this case lies close to the middle of the gap and provided $E_g \gg kT$ we may safely take the temperature to zero. We denote collectively the occupied states by v (for valence) and the empty ones by c (for conduction). In the low temperature limit, $f_{nm} \approx 1$ as long as $n \in v$ and $m \in c$ whereas $f_{nm} \approx 0$ in all other cases having $m > n$. Hence, in the expression above, the sums over m and n are over the occupied and empty states, respectively. Thus, the response can be written

$$\chi(\omega) = \frac{2e^2 \hbar^2}{\varepsilon_0 m^2 \Omega} \sum_{m \in c} \sum_{n \in v} \frac{\left| \langle \varphi_m | \hat{p}_z | \varphi_n \rangle \right|^2}{E_{mn} \left[E_{mn}^2 - \hbar^2 (\omega + i\Gamma)^2 \right]}. \tag{15.3}$$

This result is quite general and can be applied in accurate numerical calculations if energies and momentum matrix elements are calculated from a set of eigenstates

obtained from e.g. *ab initio* or empirical quantum methods. In a three-dimensional crystal, we can write general occupied and empty eigenstates as

$$\left|\varphi_{n\vec{k}}\right\rangle = \frac{1}{\sqrt{\Omega}} u_{v\vec{k}}(\vec{r}) e^{i\vec{k}\cdot\vec{r}} \left|\sigma\right\rangle, \quad \left|\varphi_{m\vec{k}}\right\rangle = \frac{1}{\sqrt{\Omega}} u_{c\vec{k}}(\vec{r}) e^{i\vec{k}\cdot\vec{r}} \left|\sigma\right\rangle,$$

where, again, $u_{v\vec{k}}(\vec{r})$ and $u_{c\vec{k}}(\vec{r})$ are the lattice-periodic parts and $\left|\sigma\right\rangle$ is the spin part. Summing over occupied states then really means summing over band v and integrating over k-vectors. Similar reasoning applies to the empty states. However, in the dipole approximation, only transitions between band states at identical k-vectors survive. Hence, summing over spins and converting to k-integration, we find

$$\chi(\omega) = \frac{4e^2\hbar^2}{\varepsilon_0 m^2 (2\pi)^3} \sum_{m\in c} \sum_{n\in v} \int \frac{\left|\left\langle \varphi_{m\vec{k}} \left| \hat{p}_z \right| \varphi_{n\vec{k}} \right\rangle\right|^2}{E_{mn}(\vec{k}) \left[E_{mn}^2(\vec{k}) - \hbar^2(\omega + i\Gamma)^2 \right]} d^3k.$$

In general, this expression is hard to evaluate. In the special case, for which only a single pair of bands needs to be considered, and if the momentum matrix element in the numerator can be assumed independent of k-vector and approximated by a constant p_{cv}, one finds

$$\chi(\omega) \approx \frac{4e^2\hbar^2 \left| p_{cv} \right|^2}{\varepsilon_0 m^2 (2\pi)^3} \int \frac{1}{E_{cv}(\vec{k}) \left[E_{cv}^2(\vec{k}) - \hbar^2(\omega + i\Gamma)^2 \right]} d^3k.$$

Moreover, in the limit $\Gamma \to 0$ the imaginary part is easily extracted

$$\mathrm{Im}\,\chi(\omega) \approx \frac{\pi e^2 \left| p_{cv} \right|^2}{\varepsilon_0 m^2 \omega^2} J(\omega), \quad J(\omega) = \frac{2}{(2\pi)^3} \int \delta\left(E_{cv}(\vec{k}) - \hbar\omega \right) d^3k.$$

Here, J is the so-called joint density of states, which counts the number of vertical (k-vector preserving) transitions across the band gap. Mathematically, it is obviously very similar to the usual density of states discussed in Appendix 1.

Let us now return to the slightly more tricky $m = n$ case in Eq.(15.2). In reality, this case arises for excitations within a single band, so-called *intraband* excitations. The way to deal with them is by realizing that one must carefully consider the momentum transfer $\hbar\vec{q}$ from the photon to the electrons. Hence, if the initial state is $\{n, \vec{k}\}$, the final one is actually $\{n, \vec{k} + \vec{q}\}$. We then find from Eq.(15.2) after spin summation

$$\chi_{intra}(\omega)=\frac{2e^2\hbar^2}{\varepsilon_0 m^2(2\pi)^3}\sum_n\int\left\{f(E_{n,\vec{k}})-f(E_{n,\vec{k}+\vec{q}})\right\}\frac{\left|\left\langle\varphi_{n,\vec{k}+\vec{q}}\left|\hat{p}_z\right|\varphi_{n,\vec{k}}\right\rangle\right|^2}{(E_{n,\vec{k}+\vec{q}}-E_{n,\vec{k}})\left[(E_{n,\vec{k}+\vec{q}}-E_{n,\vec{k}})^2-\hbar^2(\omega+i\Gamma)^2\right]}d^3k.$$

Now, the limit $\vec{q}\to 0$ can be taken and will lead to a finite, well-behaved expression. If we write $\Delta E=E_{n,\vec{k}+\vec{q}}-E_{n,\vec{k}}$, we find to first order $f(E_{n,\vec{k}+\vec{q}})\approx f(E_{n,\vec{k}})+f'(E_{n,\vec{k}})\Delta E$ with the prime denoting energy derivative. Moreover, we know from Eq.(3.9) that $\left\langle\varphi_{n,\vec{k}}\left|\hat{p}_z\right|\varphi_{n,\vec{k}}\right\rangle=(m/\hbar)\partial E_{n,\vec{k}}/\partial k_z$. It is then a simple task to show that as $\vec{q}\to 0$

$$\chi_{intra}(\omega)=-\frac{\omega_p^2}{(\omega+i\Gamma)^2}$$

with

$$\omega_p^2=-\frac{2e^2}{\varepsilon_0\hbar^2(2\pi)^3}\sum_n\int f'(E_{n,\vec{k}})\left(\frac{\partial E_{n,\vec{k}}}{\partial k_z}\right)^2 d^3k$$

in agreement with chapter 8. At this point, we wish to gain some basic insight into the optical properties of semiconductors in various dimensions. For this purpose we'll need some approximations to make the calculation tractable.

15.1 Two-Band and Envelope Approximations

The first simplification we can make is that only two bands are considered: one occupied valence band (v) and one empty conduction band (c). Now, even in low-dimensional geometries we may write the eigenstates in the form

$$\left|\varphi_n\right\rangle=u_{v\vec{k}}(\vec{r})F^{(D)}_{\nu\vec{k}}(\vec{r})\left|\sigma\right\rangle,\quad\left|\varphi_m\right\rangle=u_{c\vec{k}}(\vec{r})F^{(D)}_{\mu\vec{k}}(\vec{r})\left|\sigma\right\rangle.$$

The functions $F^{(D)}_{\nu\vec{k}}$ and $F^{(D)}_{\mu\vec{k}}$ that replace $e^{i\vec{k}\cdot\vec{r}}$ are now labeled by composite indices $\{\nu\vec{k}\}$ and $\{\mu\vec{k}\}$. This notation is to be understood as follows: $\vec{k}$ is now the D dimensional wave vector that contains the quantum numbers for the D extended dimensions and $\nu=1,2,3,...$ and $\mu=1,2,3,...$ label the quantized eigenstates for the reduced dimensions. As an example, the empty states in a quantum well are labeled by $\vec{k}=(k_x,k_y)$ for the in-plane motions and $\nu=1,2,3,...$ is the quantum number for the quantized states perpendicular to the quantum well plane. The *envelope approximation* consists in taking for $F^{(D)}_{\nu\vec{k}}$ and $F^{(D)}_{\mu\vec{k}}$ the wave functions determined exclusively by the quantum confinement, i.e. ignoring the lattice-periodic part of the potential energy.

Hence, we are assuming that the total wave function can be factored into a rapidly varying lattice-periodic part and a slowly varying *envelope function*. This makes sense if the confining potential is slowly varying compared to the lattice constant. The envelope function is a mix of running waves for the extended directions and standing waves for the confined ones. As the simplest possible case, we may take the confinement to be a rectangular potential well with zero potential energy inside the well and infinite outside (see Appendix 1). For semiconductors in 0, 1, 2 and 3 dimensions this corresponds to a cubic box, a rod with square cross section, a slab, and infinite space, respectively. Hence, if the width of the potential well is d in all cases, the occupied envelope functions in D dimensions are

$$F_{\nu\vec{k}}^{(D)}(\vec{r}) = \frac{1}{\sqrt{\Omega}} \begin{cases} 2^{3/2} \sin(\frac{\nu_x \pi x}{d}) \sin(\frac{\nu_y \pi y}{d}) \sin(\frac{\nu_z \pi z}{d}) & D = 0 \\ 2 \sin(\frac{\nu_x \pi x}{d}) \sin(\frac{\nu_y \pi y}{d}) e^{ik_z z} & D = 1 \\ \sqrt{2} \sin(\frac{\nu \pi x}{d}) e^{i(k_y y + k_z z)} & D = 2 \\ e^{i\vec{k}\cdot\vec{r}} & D = 3. \end{cases}$$

Note that we have chosen to normalize the envelope functions within the total volume Ω in each case. We will next assume that the lattice-periodic parts do not depend too strongly on wave vector in the relevant part of the Brillouin zone so that we may take $u_{v\vec{k}}(\vec{r}) \approx u_{v0}(\vec{r})$ and $u_{c\vec{k}}(\vec{r}) \approx u_{c0}(\vec{r})$. For transitions between states with identical spin, this means that in the general case

$$\begin{aligned} \langle \varphi_m | \hat{p}_z | \varphi_n \rangle &= \frac{\hbar}{i} \int u_{c0}^*(\vec{r}) F_{\mu\vec{k}}^{(D)*}(\vec{r}) \frac{\partial}{\partial z} u_{v0}(\vec{r}) F_{\nu\vec{k}}^{(D)}(\vec{r}) d^3 r \\ &= \frac{\hbar}{i} \int u_{c0}^*(\vec{r}) u_{v0}(\vec{r}) F_{\mu\vec{k}}^{(D)*}(\vec{r}) \frac{\partial}{\partial z} F_{\nu\vec{k}}^{(D)}(\vec{r}) d^3 r + \frac{\hbar}{i} \int F_{\mu\vec{k}}^{(D)*}(\vec{r}) F_{\nu\vec{k}}^{(D)}(\vec{r}) u_{c0}^*(\vec{r}) \frac{\partial}{\partial z} u_{v0}(\vec{r}) d^3 r. \end{aligned}$$

The integrand here is comprised of a rapidly varying lattice-periodic part and a slowly varying envelope part. To see how such integrals are evaluated we consider the product of a rapid lattice-periodic function $R(x)$ and a slow function $S(x)$ in one dimension. If we are to integrate this product over N unit cells each of a size a we can split the integral as follows

$$\int_0^{Na} R(x)S(x)dx = \sum_{l=1}^{N} \int_{(l-1)a}^{la} R(x)S(x)dx.$$

Next, as the slow function barely varies across a single unit cell we replace it by the value found at the midpoint $x = (l - \frac{1}{2})a$ in each unit cell:

$$\sum_{l=1}^{N}\int_{(l-1)a}^{la} R(x)S(x)dx \approx \sum_{l=1}^{N} S((l-\tfrac{1}{2})a)\int_{(l-1)a}^{la} R(x)dx$$
$$= \int_{0}^{a} R(x)dx \sum_{l=1}^{N} S((l-\tfrac{1}{2})a),$$

where the periodicity of $R(x)$ is used in the last line. Finally, running the argument backwards, the sum over l can be approximated by an integral

$$\sum_{l=1}^{N} S((l-\tfrac{1}{2})a) \approx \frac{1}{a}\sum_{l=1}^{N}\int_{(l-1)a}^{la} S(x)dx$$
$$= \frac{1}{a}\int_{0}^{Na} S(x)dx.$$

Putting things together, we see that

$$\int_{0}^{Na} R(x)S(x)dx \approx \frac{1}{a}\int_{0}^{a} R(x)dx \int_{0}^{Na} S(x)dx.$$

The essence of this result is that an integral of a product between a rapid lattice-periodic part and a slow envelope part is approximately equal to the average of the rapid part over the unit cell times the integral of the slow part over the entire volume of integration. If we apply this to the momentum matrix element we find

$$\left\langle \varphi_m \middle| \hat{p}_z \middle| \varphi_n \right\rangle \approx \frac{1}{\Omega_{UC}}\int_{UC} u_{c0}^{*}(\vec{r})u_{v0}(\vec{r})d^3r \int F_{\mu\vec{k}}^{(D)*}(\vec{r})\hat{p}_z F_{\nu\vec{k}}^{(D)}(\vec{r})d^3r$$
$$+\frac{1}{\Omega_{UC}}\int_{UC} u_{c0}^{*}(\vec{r})\hat{p}_z u_{v0}(\vec{r})d^3r \int F_{\mu\vec{k}}^{(D)*}(\vec{r})F_{\nu\vec{k}}^{(D)}(\vec{r})d^3r,$$

where Ω_{UC} is the volume of the unit cell and the first integral in each term is over Ω_{UC}. The orthogonality between $u_{v0}(\vec{r})$ and $u_{c0}(\vec{r})$ means that the first term vanishes and so

$$\left\langle \varphi_m \middle| \hat{p}_z \middle| \varphi_n \right\rangle \approx p_{cv} S_{\mu\nu}^{(D)},$$

where

$$p_{cv} = \frac{1}{\Omega_{UC}}\int_{UC} u_{c0}^{*}(\vec{r})\hat{p}_z u_{v0}(\vec{r})d^3r, \quad S_{\mu\nu}^{(D)} = \int F_{\mu\vec{k}}^{(D)*}(\vec{r})F_{\nu\vec{k}}^{(D)}(\vec{r})d^3r,$$

are the interband momentum matrix element and envelope overlap, respectively. This notation allows us to reformulate Eq.(15.3) as

$$\chi(\omega) = \frac{4e^2\hbar^2 \left|p_{cv}\right|^2}{\varepsilon_0 m^2 \Omega} \sum_{\mu,\nu,\vec{k}} \frac{\left|S_{\mu\nu}^{(D)}\right|^2}{E_{\mu\nu}(\vec{k})\left[E_{\mu\nu}^2(\vec{k}) - \hbar^2(\omega + i\Gamma)^2\right]}. \tag{15.4}$$

Here, the extra factor of 2 is from summation over spin. The excitation energy $E_{\mu\nu}(\vec{k})$ is the difference between the energy of the empty state $E_{\mu\vec{k}}$ and the energy of the occupied state $E_{\nu\vec{k}}$ participating in the transition. These quantities both include a quantization contribution in addition to the usual $\vec{k}$ -dependent kinetic part. They are given by

$$E_{\mu\vec{k}} = E_c + E_\mu^e + \frac{\hbar^2 k^2}{2m_e}, \quad E_{\nu\vec{k}} = E_v + E_\nu^h - \frac{\hbar^2 k^2}{2m_h},$$

where the quantization energies of electrons and holes are denoted E_μ^e and E_ν^h, respectively. In turn, their difference is

$$\begin{aligned} E_{\mu\nu}(\vec{k}) &= E_{\mu\vec{k}} - E_{\nu\vec{k}} = E_c + E_\mu^e + \frac{\hbar^2 k^2}{2m_e} - E_v - E_\nu^h + \frac{\hbar^2 k^2}{2m_h} \\ &\equiv E_{g\mu\nu} + \frac{\hbar^2 k^2}{2m_{eh}}. \end{aligned}$$

Here, $E_{g\mu\nu} = E_c + E_\mu^e - E_v - E_\nu^h$ is the effective band gap and $m_{eh} = m_e m_h / (m_e + m_h)$ is the reduced mass of an electron-hole pair. In the next chapter, we successively treat 3, 2, 1 and 0 dimensional semiconductors and for these cases evaluate Eq.(15.4) for the optical susceptibility.

Exercise: Envelope functions in parabolic confinement

Suppose that the quantizing potential of a quantum well is parabolic. The curvature of the confinement may be different for electrons and holes just as they will have different effective masses. This means that the envelope functions will be different as well. From basic quantum mechanics we know that the first two eigenfunctions in a parabolic potential are

$$\varphi_0(z) = \left(\frac{\beta}{\pi}\right)^{1/4} \exp\{-\tfrac{1}{2}\beta z^2\}, \quad \varphi_1(z) = \left(\frac{4\beta^3}{\pi}\right)^{1/4} z \exp\{-\tfrac{1}{2}\beta z^2\},$$

where $\beta = m\omega_0 / \hbar$ depends on mass and curvature. Electrons and holes will consequently be described by these functions but different expressions for β should be used: $\beta_e = m_e\omega_e / \hbar$ for electrons and $\beta_h = m_h\omega_h / \hbar$ for holes.

a) Calculate the partial envelope overlaps given by

$$S_{\mu\nu}^{(2)} = \int_{-\infty}^{\infty} \varphi_\mu(z)\varphi_\nu(z)dz\,,$$

in terms of β_e and β_h. Help:

$$\int_{-\infty}^{\infty} e^{-ax^2}dx = \sqrt{\frac{\pi}{a}}, \quad \int_{-\infty}^{\infty} x^2 e^{-ax^2}dx = \frac{1}{2a}\sqrt{\frac{\pi}{a}}\,.$$

b) What happens if $\beta_e = \beta_h$?

16. Optics of Bulk and Low-Dimensional Semiconductors

We now want to apply the general theory of the optical properties of semiconductors developed in the previous chapter to some interesting cases. The general theory applies equally well to bulk and low-dimensional materials and by investigating their optical responses we can highlight their differences. As will become apparent, the mathematical techniques for dealing with the different cases are highly similar. Physically, the difference between bulk and low-dimensional structures originates in quantization effects. The quantized motion of carriers influences both selection rules, strength of transitions and position of resonances. Most obviously, quantization tends to blue-shift the absorption edge, i.e. the photon energy threshold for absorption, due to the added quantization energy. Hence, the effective band gap is increased in a quantized geometry. As we will see below, however, by going from 3D to 0D materials the shape of the spectra changes dramatically as well. For a bulk semiconductor, the absorption strength is a smooth function of frequency above the band gap. When quantization increases, a strong absorption feature develops directly above the effective band gap. Another important feature is the appearance of multiple resonances in the spectra due to transitions between many subbands.

We now systematically study the optical response of 3, 2, 1 and 0-dimensional semiconductors using the two-band model and the envelope approximation described in Chapter 15. The general starting point is Eq.(15.4), which we repeat here:

$$\chi(\omega) = \frac{4e^2\hbar^2 \left|p_{cv}\right|^2}{\varepsilon_0 m^2 \Omega} \sum_{\mu,\nu,\vec{k}} \frac{\left|S_{\mu\nu}^{(D)}\right|^2}{E_{\mu\nu}(\vec{k})\left[E_{\mu\nu}^2(\vec{k}) - \hbar^2(\omega + i\Gamma)^2\right]}. \tag{16.1}$$

The summations here are over a D-dimensional k-vector $\vec{k}$ and the indices of the occupied (ν) and empty (μ) quantized states. The summation over $\vec{k}$ can in each case be performed analytically leaving us with a summation of the quantization indices. For a bulk or 3-dimensional semiconductor there are no quantization effects and, hence, the summation in Eq.(16.1) is, in fact, only over $\vec{k}$. As always, we convert the $\vec{k}$ summation into an integral and so

$$\chi(\omega) = \frac{4e^2\hbar^2 \left|p_{cv}\right|^2}{\varepsilon_0 m^2 (2\pi)^3} \int \frac{d^3k}{E_{\mu\nu}(\vec{k})\left[E_{\mu\nu}^2(\vec{k}) - \hbar^2(\omega + i\Gamma)^2\right]}.$$

The excitation energy for transitions between the two parabolic bands is

$$E_{\mu\nu}(\vec{k}) = E_g + \frac{\hbar^2 k^2}{2m_e} + \frac{\hbar^2 k^2}{2m_h}$$
$$= E_g + \frac{\hbar^2 k^2}{2m_{eh}},$$

where $m_{eh} = m_e m_h / (m_e + m_h)$ is the reduced mass of an electron-hole pair. We evaluate the integral by introducing $x = \hbar^2 k^2 / 2m_{eh}$, which means that

$$d^3k = 4\pi k^2 dk = 2\pi k d(k^2) = 2\pi \sqrt{\frac{2m_{eh}x}{\hbar^2}} d\left(\frac{2m_{eh}x}{\hbar^2}\right) = 2\pi \left(\frac{2m_{eh}}{\hbar^2}\right)^{3/2} \sqrt{x}\, dx\,.$$

In this manner, the susceptibility integral becomes

$$\chi(\omega) = \frac{4e^2\hbar^2 |p_{cv}|^2}{\varepsilon_0 m^2 (2\pi)^2} \left(\frac{2m_{eh}}{\hbar^2}\right)^{3/2} \int_0^\infty \frac{\sqrt{x}\, dx}{(E_g + x)\left[(E_g + x)^2 - \hbar^2(\omega + i\Gamma)^2\right]}.$$

The x-integral can now be evaluated. To this end, we introduce the complex frequency $w \equiv \omega + i\Gamma$ and the normalized 3D susceptibility function $X_3(w)$ as

$$\begin{aligned} X_3(w) &\equiv \int_0^\infty \frac{E_g^{3/2}\sqrt{x}\, dx}{(E_g + x)\left[(E_g + x)^2 - \hbar^2 w^2\right]} \\ &= \frac{\pi E_g^{3/2}}{2\hbar^2 w^2}\left\{2\sqrt{E_g} - \sqrt{E_g - \hbar w} - \sqrt{E_g + \hbar w}\right\}. \end{aligned} \tag{16.2}$$

This leads to the result

$$\chi(\omega) = \frac{e^2\hbar^2 |p_{cv}|^2}{\pi^2 \varepsilon_0 m^2 E_g^{3/2}} \left(\frac{2m_{eh}}{\hbar^2}\right)^{3/2} X_3(w).$$

The limiting value of the susceptibility function is $X_3(0) = \pi/8 \approx 0.393$. The plot in Fig. 16.1 illustrates this result using the values $E_g = 1.5$ eV and $\hbar\Gamma = 0.025$ eV that are representative of GaAs. A few things are worth pointing out about this figure. First, the absorption is determined by the imaginary part of the refractive index $n_i(\omega)$. Since $2n_r n_i = \chi_i$, we have $n_i = \chi_i / 2n_r$ so that the absorption is directly proportional to the imaginary part of the susceptibility. From the figure, it is obvious that the absorption threshold is located at the energy gap, i.e. if a frequency scan is made, strong absorption will set in at $\hbar\omega = E_g$.

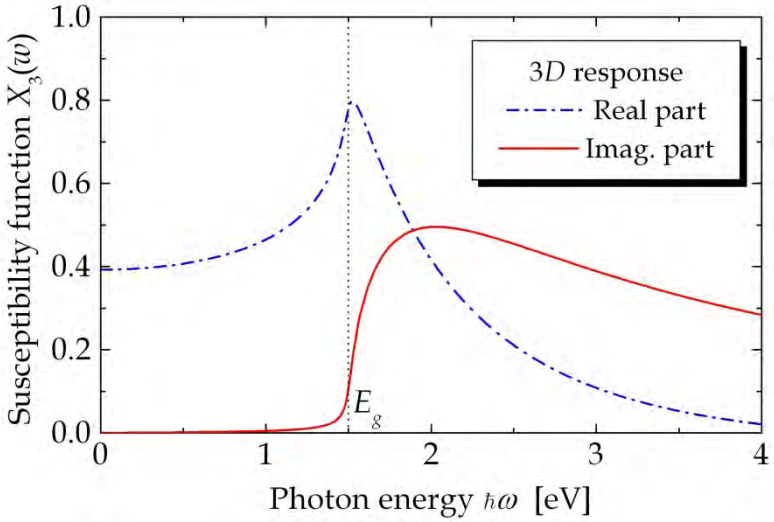


Figure 16.1. Complex susceptibility function of a 3D semiconductor with a band gap of 1.5 eV.

A second point is that the susceptibility at low photon energy (well below the energy gap) is purely real-valued and corresponds to the usual "refractive index" n_r of the material via $n_r^2 = 1 + \chi_r$ so that $n_r = \sqrt{1+\chi_r}$.

16.1 Semiconductor Quantum Wells

A quantum well is a semiconductor slab having a smaller band gap than the surroundings and therefore being able to confine carriers to the slab. The confinement produces a series of subbands offset by the quantization energy. Hence, for the $\mu\nu$ 'th subband the excitation energy is

$$E_{\mu\nu}(\vec{k}) = E_{g\mu\nu} + \frac{\hbar^2 k^2}{2m_{eh}},$$

where $\vec{k}$ is the 2 dimensional wave vector in the plane and $E_{g\mu\nu}$ is the effective energy gap including the quantization energy. Hence, if infinite square-well confinement with a width d is assumed

$$E_{g\mu\nu} = E_g + \frac{\mu^2\hbar^2\pi^2}{2m_e d^2} + \frac{\nu^2\hbar^2\pi^2}{2m_h d^2}.$$

In analogy with the 3D case, we now convert the 2D $\vec{k}$ - summation into an integral and subsequently write the susceptibility as

$$\chi(\omega) = \frac{4e^2\hbar^2\left|p_{cv}\right|^2}{\varepsilon_0 m^2 (2\pi)^2 d}\sum_{\mu,\nu}\left|S_{\mu\nu}^{(2)}\right|^2 \int \frac{d^2k}{E_{\mu\nu}(\vec{k})\left[E_{\mu\nu}^2(\vec{k}) - \hbar^2(\omega + i\Gamma)^2\right]},$$

using $\Omega = Ad$, where A is the area of the slab. Again, we use $x = \hbar^2 k^2 / 2m_{eh}$ so that

$$d^2k = 2\pi k dk = \pi d(k^2) = \pi \frac{2m_{eh}}{\hbar^2} dx \,.$$

This allows us to write the susceptibility as the integral

$$\chi(\omega) = \frac{4e^2\hbar^2 |p_{cv}|^2}{\varepsilon_0 m^2 (2\pi)^2 d} \frac{2\pi m_{eh}}{\hbar^2} \sum_{\mu,\nu} \left|S^{(2)}_{\mu\nu}\right|^2 \int_0^\infty \frac{dx}{(E_{g\mu\nu}+x)\left[(E_{g\mu\nu}+x)^2 - \hbar^2(\omega+i\Gamma)^2\right]}.$$

Hence, we naturally introduce the 2D susceptibility function $X_2(w)$ as the frequency dependent part

$$\begin{aligned} X_2(w) &\equiv \int_0^\infty \frac{E^2_{g\mu\nu} dx}{(E_{g\mu\nu}+x)\left[(E_{g\mu\nu}+x)^2 - \hbar^2 w^2\right]} \\ &= \frac{E^2_{g\mu\nu}}{2\hbar^2 w^2} \ln\left(\frac{E^2_{g\mu\nu}}{E^2_{g\mu\nu} - \hbar^2 w^2}\right), \end{aligned} \tag{16.3}$$

with the limiting behavior $X_2(0) = 1/2$. It follows that the 2D susceptibility is

$$\chi(\omega) = \frac{2e^2 |p_{cv}|^2 m_{eh}}{\pi \varepsilon_0 m^2 d} \sum_{\mu,\nu} \frac{\left|S^{(2)}_{\mu\nu}\right|^2}{E^2_{g\mu\nu}} X_2(w).$$

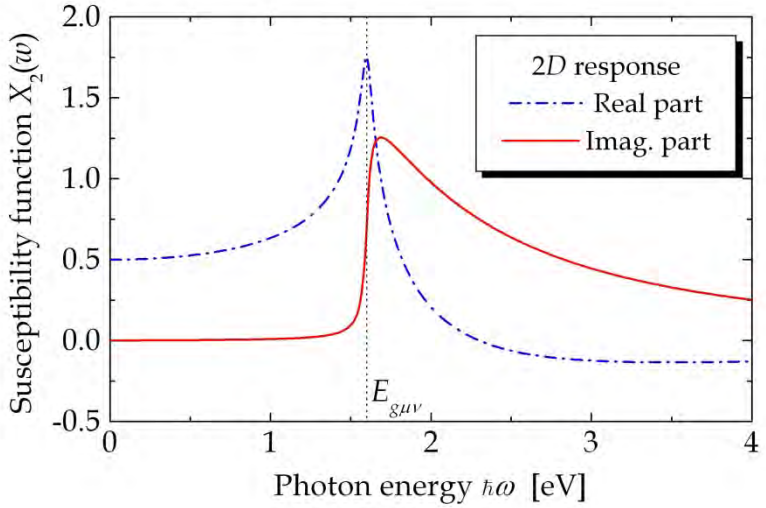


Figure 16.2. Complex susceptibility function of a quantum well with an effective band gap of 1.6 eV.

This new susceptibility function is depicted in Fig. 16.2. Here we have again taken data for GaAs and assumed a quantization energy of 0.1 eV so that the effective gap

is $E_{g\mu\nu} = 1.6$ eV. By comparison to Fig. 16.1 it is seen that the band gap feature is significantly steeper in the quantum well case.

16.2 Semiconductor Quantum Wires and Dots

Quantum wires are one-dimensional in the sense that their extension along one axis is much greater than the other two, which we assume to have nanoscale dimensions. The calculation of the susceptibility is highly similar to the quantum well case, except that now the *k*-vector is one-dimensional and the sum over $\mu\nu$ covers the two-dimensionally quantized stated. We begin by noting that

$$dk = \frac{1}{2k} d(k^2) = \frac{1}{2}\left(\frac{\hbar^2}{2m_{eh}x}\right)^{1/2} d\left(\frac{2m_{eh}x}{\hbar^2}\right) = \left(\frac{m_{eh}}{2\hbar^2}\right)^{1/2} \frac{1}{\sqrt{x}} dx.$$

This, in turn, yields a susceptibility given by (remembering a factor of 2 from $\pm k$)

$$\begin{aligned}\chi(\omega) &= \frac{8e^2\hbar^2 \left|p_{cv}\right|^2}{\varepsilon_0 m^2 2\pi d^2}\left(\frac{m_{eh}}{2\hbar^2}\right)^{1/2} \sum_{\mu,\nu}\left|S_{\mu\nu}^{(1)}\right|^2 \int_0^\infty \frac{dx}{\sqrt{x}(E_{g\mu\nu}+x)\left[(E_{g\mu\nu}+x)^2 - \hbar^2(\omega+i\Gamma)^2\right]} \\ &= \frac{4e^2\hbar^2 \left|p_{cv}\right|^2}{\pi\varepsilon_0 m^2 d^2}\left(\frac{m_{eh}}{2\hbar^2}\right)^{1/2} \sum_{\mu,\nu}\frac{\left|S_{\mu\nu}^{(1)}\right|^2}{E_{g\mu\nu}^{5/2}} X_1(w).\end{aligned}$$

In this case, the corresponding susceptibility function is

$$\begin{aligned}X_1(w) &\equiv \int_0^\infty \frac{E_{g\mu\nu}^{5/2} dx}{\sqrt{x}(E_{g\mu\nu}+x)\left[(E_{g\mu\nu}+x)^2 - \hbar^2 w^2\right]} \\ &= \frac{\pi E_{g\mu\nu}^{5/2}}{2\hbar^2 w^2}\left\{-\frac{2}{\sqrt{E_{g\mu\nu}}} + \frac{1}{\sqrt{E_{g\mu\nu}-\hbar w}} + \frac{1}{\sqrt{E_{g\mu\nu}+\hbar w}}\right\}.\end{aligned} \tag{16.4}$$

In this case, the DC limit is $X_1(0) = 3\pi/8 \approx 1.178$. The result is illustrated below assuming $E_{g\mu\nu} = 1.8$ eV. The trend from the quantum well case is continued and now the absorption peak at the band edge is further sharpened. In fact, if the broadening $\hbar\Gamma$ goes to zero the imaginary part of the susceptibility function will diverge. This is readily seen from Eq.(16.4). When $\Gamma = 0$, the imaginary part can only come from the second term in the curly bracket. This term will be purely imaginary if $\hbar\omega > E_g$ and consequently

$$\lim_{\Gamma\to 0} \mathrm{Im}\{X_1(w)\} = \frac{\pi E_{g\mu\nu}^{5/2}}{2\hbar^2\omega^2} \frac{\theta\left(\hbar\omega - E_{g\mu\nu}\right)}{\sqrt{\hbar\omega - E_{g\mu\nu}}}.$$

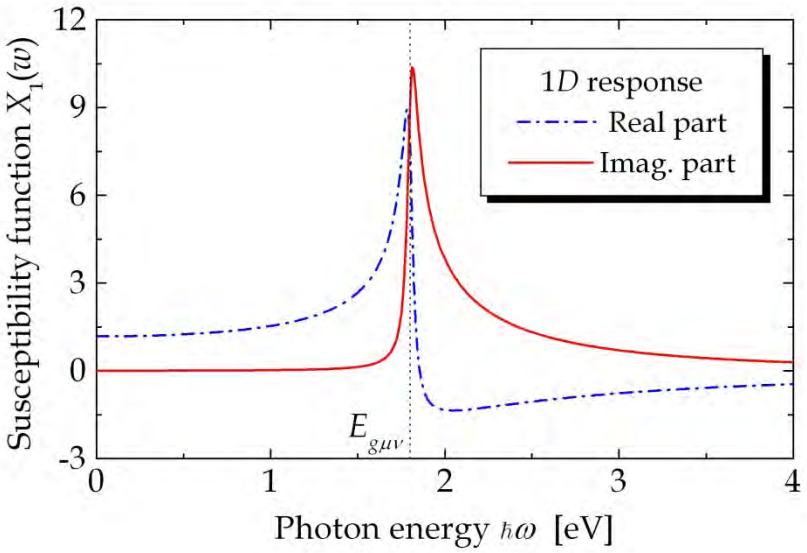


Figure 16.3. Complex susceptibility function of a quantum wire with an effective band gap of 1.8 eV.

It follows that as the absorption edge is approached from above, the imaginary part will diverge as an inverse square root of the frequency.

Quantum dots are semiconductor nanoparticles with no extended dimensions. Hence, the spectrum of eigenvalues is purely discrete. If a box shape with side length d is assumed for the particle, the excitation energies are

$$E_{g\mu\nu} = E_g + \frac{\mu^2\hbar^2\pi^2}{2m_e d^2} + \frac{\nu^2\hbar^2\pi^2}{2m_h d^2}.$$

Hence, there is no k-integration is this case and we simply find

$$\chi(\omega) = \frac{4e^2\hbar^2\left|p_{cv}\right|^2}{\varepsilon_0 m^2 d^3}\sum_{\mu,\nu}\frac{\left|S_{\mu\nu}^{(0)}\right|^2}{E_{g\mu\nu}^3}X_0(w),$$

where

$$X_0(w) = \frac{E_{g\mu\nu}^2}{E_{g\mu\nu}^2 - \hbar^2 w^2} \tag{16.5}$$

is the 0-dimensional susceptibility function with the limit $X_0(0) = 1$ as illustrated in Fig. 16.4 below. Here, $E_{g\mu\nu} = 2.0$ eV is assumed in order to follow the trend of increased quantization energy as we go through wells, wires and now dots. The response of the 0D system is an isolated resonance at the effective gap. It is noted that the imaginary part is completely symmetric in contrast to the higher-dimensional structures.

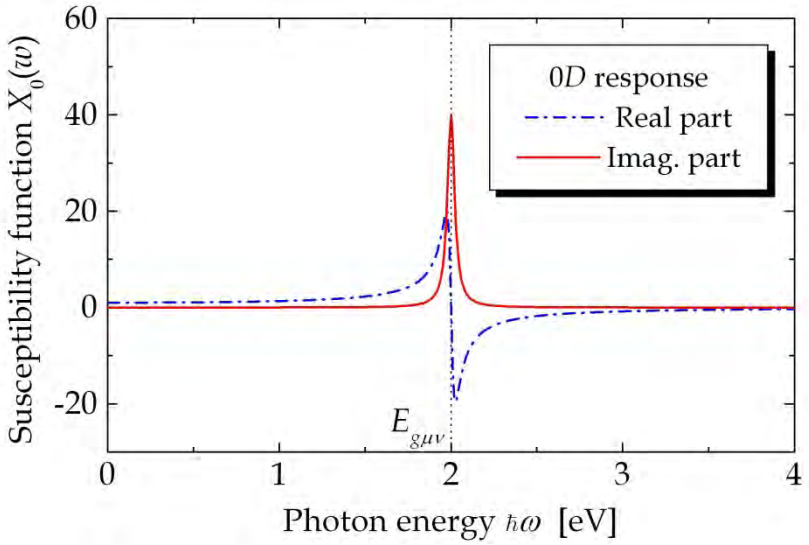


Figure 16.4. Susceptibility function of a quantum dot with an effective band gap of 2.0 eV.

In the plots above, we have displayed the contribution from a single resonance. In general, the spectrum will consist of a series of resonances due to transitions between multiple subbands. The weight of each transition is given by the envelope overlap $S_{\mu\nu}^{(D)}$ in Eq.(16.1). If the envelope functions of conduction and valence band states were eigenfunctions of the *same* Hamiltonian the overlap would simply be $S_{\mu\nu}^{(D)} = \delta_{\mu\nu}$ because of orthogonality. However, the Hamiltonians of the two bands are generally different due to different effective masses and differently shaped potential wells. A special case is the infinite rectangular well, for which $S_{\mu\nu}^{(D)} = \delta_{\mu\nu}$ still holds. The reason is that, in this case, the eigenstates are simple standing waves that do not depend on effective masses anywhere. Hence, the wave functions remain orthogonal. As an example of several subbands contributing to the total response we consider a quantum wire with three allowed transitions $\nu = 1 \to \mu = 1$, $\nu = 2 \to \mu = 2$ and $\nu = 3 \to \mu = 3$. We take $E_{g11} = 1.8$ eV, $E_{g22} = 2.5$ eV and $E_{g33} = 3.5$ eV and keep $\hbar\Gamma = 0.025$ eV. The resulting spectrum for the imaginary part of the susceptibility function $X_1(w)$ is illustrated in Fig. 16.5. Since the separation between the resonances is far greater than the broadening of the individual peaks, the three resonances are well resolved. If the spacing between the peaks becomes comparable to $\hbar\Gamma$ the peaks will overlap and be hard to distinguish.

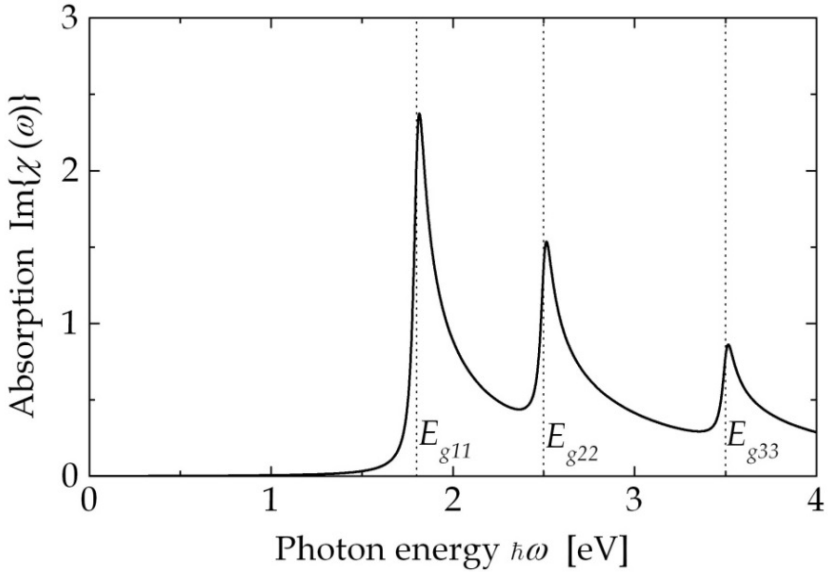


Figure 16.5. Absorption spectrum of a quantum wire with three allowed subband transitions.

Exercise: Limits of the susceptibility functions

a) Show that the susceptibility functions have the DC limits stated in the text, that is, show that

$$X_3(0) = \pi/8$$
$$X_2(0) = 1/2$$
$$X_1(0) = 3\pi/8$$
$$X_0(0) = 1.$$

b) Find limiting expression for the imaginary parts of the susceptibility functions as the broadening $\hbar\Gamma$ goes to zero.

17. Electronic and Optical Properties of Graphene

Graphene is a novel wonder material that has stimulated a tremendous amount of experimental and theoretical work. For theoretical nanoscience, it is a great material because of its extreme simplicity but, nonetheless, intriguing properties. A lot of this fascination derives from the fact that carriers in graphene near the intrinsic Fermi level behave similarly to massless relativistic particles, albeit moving with an effective speed of light of roughly $c/300$.

We will start this chapter by a brief review of the electronic structure of this material and subsequently consider the specific properties derived from the electronic eigenstates. More details are given in Appendix 2. Graphene is a truly two-dimensional material consisting of a honeycomb lattice of carbon atoms. A piece of the lattice is shown in the left panel of Fig. 17.1.

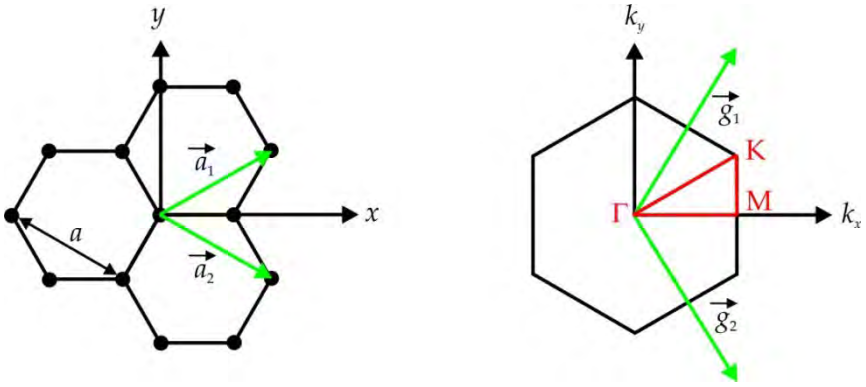


Figure 17.1. Honeycomb lattice incl. lattice vectors (left) and Brillouin zone incl. reciprocal lattice vectors (right).

The lattice constant of graphene is $a = 2.46\,\text{Å}$. Correspondingly, the elementary lattice vectors are

$$\vec{a}_1 = \frac{a}{2}\begin{pmatrix}\sqrt{3}\\ 1\end{pmatrix}, \quad \vec{a}_2 = \frac{a}{2}\begin{pmatrix}\sqrt{3}\\ -1\end{pmatrix}.$$

Clearly, there are two atoms per unit cell. Also, the reciprocal lattice vectors are easily found from the requirement $\vec{g}_i \cdot \vec{a}_j = 2\pi\delta_{ij}$ to be

$$\vec{g}_1 = \frac{2\pi}{\sqrt{3}a}\begin{pmatrix}1\\ \sqrt{3}\end{pmatrix}, \quad \vec{g}_2 = \frac{2\pi}{\sqrt{3}a}\begin{pmatrix}1\\ -\sqrt{3}\end{pmatrix}.$$

The Brillouin zone spanned by these vectors is shown in the right panel of Fig. 17.1. Here, the so-called irreducible Brillouin zone is also highlighted as the red triangle along with high symmetry points.

The simplicity of graphene derives from the fact that each carbon atom comes with 4 valence orbitals: $s, p_x, p_y,$ and p_z. Among these, p_z has odd parity with respect the reflections in the (*x*,*y*) plane whereas the remaining ones are even. Thus, p_z - states completely decouple from the rest. As a consequence, extended Bloch states formed by coupling p_z - orbitals (called π - states) also decouple from those formed by the remaining orbitals (σ - states). The π - states are loosely bound and responsible for all electronic and optical properties in the usual low-energy range. Conversely, the σ - states are largely responsible for holding the material together but are very difficult to excite and, therefore, irrelevant for most response properties.

The simplest, realistic electronic model is constructed by assuming that each p_z - orbital is coupled to its nearest neighbors only. As a basis for extended Bloch eigenstates of the full lattice, we take Bloch sums formed by summing p_z - orbitals belonging to the two atoms *A* and *B* in the unit cell (so-called sublattices), separately. In a unit cell at position $\vec{R}$, the position of the atoms *A* and *B* will be denoted $\vec{R}_A$ and $\vec{R}_B$, respectively. To simplify the notation, we will designate the p_z - orbitals belonging to the two sublattices by $\left|\pi_{\vec{R}}^A\right\rangle \equiv \left|p_z(\vec{r}-\vec{R}_A)\right\rangle$ and $\left|\pi_{\vec{R}}^B\right\rangle \equiv \left|p_z(\vec{r}-\vec{R}_B)\right\rangle$, respectively. Forming the two Bloch sums $|\alpha\rangle$ and $|\beta\rangle$ for the *A* and *B* sublattices, we find

$$|\alpha\rangle = \frac{1}{\sqrt{N}}\sum_{\vec{R}} e^{i\vec{k}\cdot\vec{R}_A}\left|\pi_{\vec{R}}^A\right\rangle, \quad |\beta\rangle = \frac{1}{\sqrt{N}}\sum_{\vec{R}} e^{i\vec{k}\cdot\vec{R}_B}\left|\pi_{\vec{R}}^B\right\rangle,$$

where *N* is formally the number of unit cells that is eventually taken to infinity. The Hamiltonian in this basis is given by

$$\overleftrightarrow{H} = \begin{pmatrix} \langle\alpha|\hat{H}|\alpha\rangle & \langle\alpha|\hat{H}|\beta\rangle \\ \langle\beta|\hat{H}|\alpha\rangle & \langle\beta|\hat{H}|\beta\rangle \end{pmatrix}.$$

Assuming only nearest neighbor coupling, the diagonal elements both equal the on-site energy $\left\langle\pi_{\vec{R}}^A\left|\hat{H}\right|\pi_{\vec{R}}^A\right\rangle$ that we will set to zero by choosing it as our energy zero point. The off-diagonal elements are non-zero, however. The nearest neighbor coupling, sometimes denoted the hopping integral or tunnel coupling, is $\gamma \equiv -\left\langle\pi_{\vec{R}}^A\left|\hat{H}\right|\pi_{\vec{R}}^B\right\rangle$ (the minus sign is introduced for numerical convenience as the matrix element itself is actually negative). Hence, it is straightforward to show that

$$\langle\alpha|\hat{H}|\beta\rangle = -\gamma h(\vec{k}), \quad h(\vec{k}) = e^{ik_x a/\sqrt{3}} + 2e^{-ik_x a/2\sqrt{3}}\cos\left(k_y a/2\right), \tag{17.1}$$

and $\langle\beta|\hat{H}|\alpha\rangle = \langle\alpha|\hat{H}|\beta\rangle^*$. At this point, it is then clear that the eigenvalues are $E_\pm = \pm\gamma|h(\vec{k})|$, i.e.

$$E_\pm = \pm\gamma\sqrt{1+4\cos^2\left(\frac{k_y a}{2}\right)+4\cos\left(\frac{k_y a}{2}\right)\cos\left(\frac{\sqrt{3}k_x a}{2}\right)}. \qquad (17.2)$$

The band structure is illustrated in Fig. 17.2. Importantly, the bands touch at the K point $\vec{K} = 2\pi/a\left(1/\sqrt{3}, 1/3\right)^T$. In fact, near the K point, the bands form a cone. This is readily seen if we expand the eigenvalues around K, i.e. by writing $\vec{k} = \vec{K} + \vec{q}$ and subsequently expand in $\vec{q}$. In this manner, we find that $h(\vec{k}) \approx \sqrt{3}a/2\left(-q_x + iq_y\right)e^{i\pi/6}$. The constant phase can be absorbed by redefining $|\alpha\rangle \to |\alpha\rangle e^{i\pi/12}$ and $|\beta\rangle \to |\beta\rangle e^{-i\pi/12}$. This means that the Hamiltonian becomes

$$\overleftrightarrow{H} = \hbar v_F \begin{pmatrix} 0 & q_x - iq_y \\ q_x + iq_y & 0 \end{pmatrix}, \quad v_F = \frac{\sqrt{3}a\gamma}{2\hbar}.$$

Hence, the eigenvalues are of the form $E_\pm = \pm\hbar v_F q$. The quantity v_F is the Fermi velocity. The hopping integral is found to be around $\gamma \approx 3$ eV implying that $v_F \approx 10^6$ m/s. Thus, carriers really resemble massless relativistic ones (since the energy dispersion is linear in the momentum $\hbar\vec{q}$) moving with a velocity $v_F \approx c/300$. This linearized approximation is usually called the Dirac model of graphene. Each carbon atom supplies precisely one electron to the π- bands (the other 3 are bound in chemical bonds with the neighbor atoms). Thus, for intrinsic graphene, the lower band is filled while the upper is empty at low temperatures. This means that the Fermi level coincides with the so-called Dirac point, at which the bands cross. Note that the eigenvectors are simply $\vec{v}_\pm = 2^{-1/2}\left(\pm e^{-i\varphi}, 1\right)^T$ with φ the polar angle of $\vec{q}$.

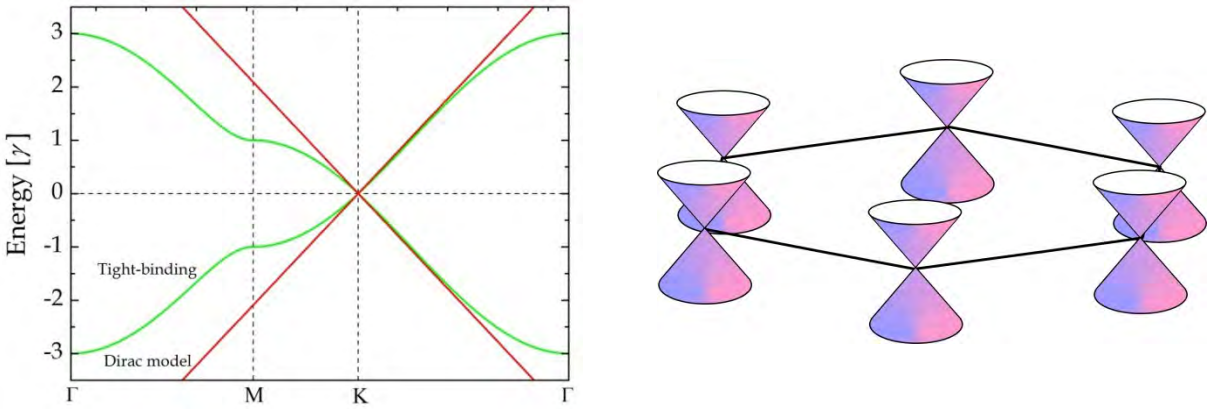


Figure 17.2. Full tight-binding band structure compared to the Dirac model (left). In the linearized model, a Dirac cone sits in each corner of the Brillouin zone (right).

17.1 Electronic and Optical Properties

In the following, we will focus on the Dirac model because the simplicity allows us to compute many quantities analytically. At various points, we can then compare to more complicated results obtained for the full tight-binding theory. First, we will consider the density of states. In order to perform the computation, we note that the contribution from a single Dirac cone should be multiplied by a factor of 4: first there are two cones per unit cell (valley degeneracy) and secondly there is spin degeneracy. Hence, density of states per area for an area A and energy $E > 0$ is

$$D(E) = \frac{4}{A}\sum_{\vec{q}} \delta\left(E - \hbar v_F q\right) = \frac{1}{\pi^2}\int \delta\left(E - \hbar v_F q\right) d^2 q = \frac{2}{\pi}\int_0^\infty \delta\left(E - \hbar v_F q\right) q dq = \frac{2E}{\pi \hbar^2 v_F^2} \quad (17.3)$$

A similar result is found for $E < 0$ and generally we can write $D(E) = 2|E|/\pi\hbar^2 v_F^2$. In a more elaborate treatment using the full tight-binding model one finds [1, 2]

$$D(E) = \frac{\gamma\sqrt{3\Omega}}{\pi^2 \hbar^2 v_F^2}\,\mathrm{Re}\left[K\left(\frac{(3-\Omega)(1+\Omega)^3}{16\Omega}\right)\right], \quad (17.4)$$

where K is the complete elliptic integral of the first kind and we have introduced normalized energies $\Omega = |E|/\gamma$. Using the limit $\lim_{z\to\infty} \mathrm{Re}\left[K(z)\right] = \pi/2\sqrt{z}$ shows that the two expressions agree in the low-energy range. The results are compared in the figure below.

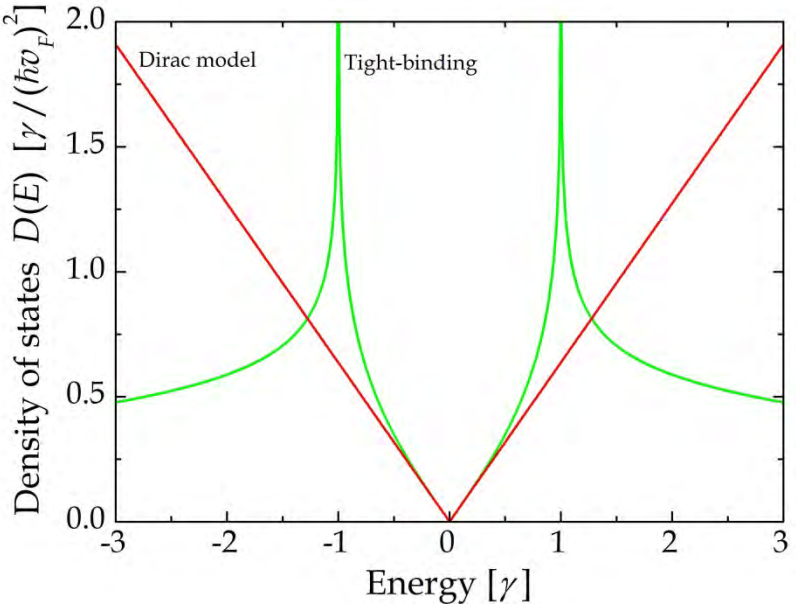


Figure 17.3. Comparison of the tight-binding and Dirac model results for the density of states.

Because graphene has a vanishing band gap but, at the same time, a vanishing density of states at the Dirac point, it can be classified as a material in between

semiconductors and metals, i.e. a so-called semimetal. The obtained density of states enables us to compute the excess electron density. As mentioned above, intrinsic graphene is characterized by a Fermi level located at the energy zero point. We therefore find the excess carriers, that is the number of additional electrons per area, by the integral

$$n = \int_{-\infty}^{\infty} D(E)\{f(E) - f_0(E)\}\, dE\,.$$

Here, $f(E) = (\exp\{(E - E_F)/kT\} + 1)^{-1}$ is the Fermi-Dirac distribution and $f_0(E) = (\exp\{E/kT\} + 1)^{-1}$ is the reference for the intrinsic case, $E_F = 0$. Inserting and reformulating, it turns out that

$$\begin{aligned} n &= \frac{2}{\pi}\left(\frac{kT}{\hbar v_F}\right)^2 \sinh\{E_F/kT\} \int_0^{\infty} \frac{x}{\cosh x + \cosh\{E_F/kT\}}\, dx \\ &= \frac{2}{\pi}\left(\frac{kT}{\hbar v_F}\right)^2 \left\{\mathrm{Li}_2\left(-\exp\{-E_F/kT\}\right) - \mathrm{Li}_2\left(-\exp\{E_F/kT\}\right)\right\}. \end{aligned} \tag{17.5}$$

Here, Li_2 is the second polylogarithm. At low and high doping levels, we find

$$n \approx \frac{E_F}{\pi(\hbar v_F)^2} \begin{cases} 4\ln 2\, kT & , \ |E_F|/kT \ll 1 \\ E_F & , \ |E_F|/kT \gg 1 \end{cases}.$$

Next, we go on to calculate the conductivity of graphene. First, however, a word about the current response of 2D systems. In the usual 3D case, we write the response to an electric field $\vec{\mathcal{E}}$ as $\vec{j} = \sigma_{3D}\vec{\mathcal{E}}$, where σ_{3D} is the 3D conductivity and $\vec{j}$ is the 3D current density defined as current per area. For a very thin 2D sheet, we may integrate over the thickness assuming that the field does not vary across the sheet. Hence, we define a *sheet conductivity* $\sigma = \int \sigma_{3D} dz$ that relates the sheet current density (current per length) to the electric field. In this manner, we need a sheet plasma frequency

$$\omega_p^2 = -\frac{4e^2}{\hbar^2 \varepsilon_0 A} \sum_{\vec{q}} \left\{ \left(\frac{\partial E_+}{\partial q_x}\right)^2 f'(E_+) + \left(\frac{\partial E_-}{\partial q_x}\right)^2 f'(E_-) \right\}.$$

Again, the factor of 4 is due to spin- and valley-degeneracy. Also, we summed over the two bands. Converting to an integral, it is found that

$$\omega_p^2 = -\frac{e^2}{\pi\hbar^2\varepsilon_0}\int_0^\infty \{f'(E)+f'(-E)\}EdE$$
$$= \frac{2e^2kT}{\pi\hbar^2\varepsilon_0}\ln\left(2\cosh\{E_F/2kT\}\right) \approx \frac{e^2E_F}{\pi\hbar^2\varepsilon_0}, \quad \text{if } E_F \gg kT.$$

The optical response generally consists of an intraband σ_{intra} contribution and an interband contribution σ_{inter}. Given the plasma frequency, the former is easily evaluated, c.f. Eq.(3.11)

$$\sigma_{intra}(\omega) = \frac{i\varepsilon_0\omega_p^2}{\omega + i\Gamma},$$

where Γ is a phenomenological broadening parameter. The interband term is significantly more complicated. In the limit of vanishing broadening, the general result for the real part after summation over spin for a two-dimensional material is

$$\mathrm{Re}\,\sigma_{inter}(\omega) = \frac{e^2}{2\pi m^2\omega}\sum_{v,c}\int \{f(E_v)-f(E_c)\}\left|P_{vc}\right|^2 \delta\left(E_c - E_v - \hbar\omega\right)d^2k\,. \qquad (17.6)$$

Here, the sum is over occupied valence bands (v) and empty conduction bands (c). Also, P_{vc} is the momentum matrix element. We evaluate this matrix element using the commutator $\hat{p}_x = im/\hbar[\hat{H},x]$ and the completeness of the p_z - states, i.e.

$$1 = \sum_{\vec{R}}\left|\pi_{\vec{R}}^A\right\rangle\left\langle\pi_{\vec{R}}^A\right| + \sum_{\vec{R}}\left|\pi_{\vec{R}}^B\right\rangle\left\langle\pi_{\vec{R}}^B\right|.$$

Thus,

$$\left\langle\alpha\left|\hat{p}_x\right|\beta\right\rangle = \frac{im}{\hbar}\left\langle\alpha\left|[\hat{H},x]\right|\beta\right\rangle = \frac{im}{\hbar}\sum_{\vec{R}}\left\{\left\langle\alpha\left|\hat{H}\right|\pi_{\vec{R}}^A\right\rangle\left\langle\pi_{\vec{R}}^A\right|x\left|\beta\right\rangle - \left\langle\alpha\left|x\right|\pi_{\vec{R}}^A\right\rangle\left\langle\pi_{\vec{R}}^A\right|\hat{H}\left|\beta\right\rangle\right\}$$
$$+\frac{im}{\hbar}\sum_{\vec{R}}\left\{\left\langle\alpha\left|\hat{H}\right|\pi_{\vec{R}}^B\right\rangle\left\langle\pi_{\vec{R}}^B\right|x\left|\beta\right\rangle - \left\langle\alpha\left|x\right|\pi_{\vec{R}}^B\right\rangle\left\langle\pi_{\vec{R}}^B\right|\hat{H}\left|\beta\right\rangle\right\}.$$

We continue assuming that our orbitals are sufficiently localized that only on-site terms matter. Hence, $\left\langle\pi_{\vec{R}}^A\left|x\right|\beta\right\rangle = 0$ and $\left\langle\alpha\left|x\right|\pi_{\vec{R}}^A\right\rangle = X_A e^{-i\vec{k}\cdot\vec{R}_A}/\sqrt{N}$ etc. We therefore see that

$$\left\langle\alpha\left|\hat{p}_x\right|\beta\right\rangle = \frac{im}{\hbar\sqrt{N}}\sum_{\vec{R}}\left\{-X_A e^{-i\vec{k}\cdot\vec{R}_A}\left\langle\pi_{\vec{R}}^A\left|\hat{H}\right|\beta\right\rangle + \left\langle\alpha\left|\hat{H}\right|\pi_{\vec{R}}^B\right\rangle X_B e^{i\vec{k}\cdot\vec{R}_B}\right\}.$$

An important point, now, is that $-X_A e^{-i\vec{k}\cdot\vec{R}_A} = -i\partial/\partial k_x e^{-i\vec{k}\cdot\vec{R}_A}$ and $X_B e^{i\vec{k}\cdot\vec{R}_B} = -i\partial/\partial k_x e^{i\vec{k}\cdot\vec{R}_B}$. Thus,

$$\begin{aligned}\langle\alpha|\hat{p}_x|\beta\rangle &= \frac{m}{\hbar}\frac{1}{\sqrt{N}}\sum_{\vec{R}}\left\{\frac{\partial}{\partial k_x}\left[e^{-i\vec{k}\cdot\vec{R}_A}\right]\langle\pi^A_{\vec{R}}|\hat{H}|\beta\rangle + \langle\alpha|\hat{H}|\pi^B_{\vec{R}}\rangle\frac{\partial}{\partial k_x}\left[e^{i\vec{k}\cdot\vec{R}_B}\right]\right\} \\ &= \frac{m}{\hbar}\frac{\partial}{\partial k_x}\langle\alpha|\hat{H}|\beta\rangle.\end{aligned}$$

As a consequence, the interband momentum operator can be replaced by $(m/\hbar)\partial\overleftrightarrow{H}/\partial k_x$. For the Dirac model, this implies that

$$\overleftrightarrow{p}_x = mv_F\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}, \quad \overleftrightarrow{p}_y = mv_F\begin{pmatrix}0 & -i\\ i & 0\end{pmatrix}.$$

The matrix elements, therefore, are exceedingly simple

$$\langle+|\hat{p}_x|-\rangle = \frac{mv_F}{2}\left(e^{i\varphi},1\right)\cdot\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}\cdot\begin{pmatrix}-e^{-i\varphi}\\ 1\end{pmatrix} = imv_F\sin\varphi, \quad \langle+|\hat{p}_y|-\rangle = -imv_F\cos\varphi\,.$$

When this result is used in the interband expression we see that (remembering to include valley degeneracy)

$$\begin{aligned}\mathrm{Re}\,\sigma_{inter}(\omega) &= \frac{e^2v_F^2}{\omega}\int_0^\infty\left\{f(E_-)-f(E_+)\right\}\delta\left(E_+-E_--\hbar\omega\right)qdq \\ &= \frac{e^2}{\hbar^2\omega}\int_0^\infty\left\{f(-E)-f(E)\right\}\delta\left(2E-\hbar\omega\right)EdE \\ &= \frac{e^2}{4\hbar}\left\{f(-\hbar\omega/2)-f(\hbar\omega/2)\right\} \\ &= \frac{e^2}{4\hbar}\frac{\sinh\{\hbar\omega/2kT\}}{\cosh\{\hbar\omega/2kT\}+\cosh\{E_F/kT\}}.\end{aligned} \tag{17.7}$$

In the low-temperature limit, the full interband conductivity becomes

$$\sigma_{inter}(\omega) = \frac{e^2}{4\hbar}\left\{\theta(\hbar\omega-2E_F)+\frac{i}{\pi}\ln\left|\frac{\hbar\omega-2E_F}{\hbar\omega+2E_F}\right|\right\}. \tag{17.8}$$

If the full tight-binding model is considered, an analytical calculation shows that in the low-temperature, intrinsic limit $T, E_F \to 0$ [1, 2]

$$\mathrm{Re}\,\sigma_{inter}(\omega)=\frac{e^2}{\sqrt{24}\pi\hbar\Omega^{3/2}}\mathrm{Re}\left[\frac{144-12\Omega+\Omega^3}{24}K\left(\frac{(6-\Omega)(2+\Omega)^3}{128\Omega}\right)-12E\left(\frac{(6-\Omega)(2+\Omega)^3}{128\Omega}\right)\right],$$

where $\Omega=\hbar\omega/\gamma$ and K and E are elliptic integrals. It can be shown that taking the zero frequency limit of the above expression leads to a minimum graphene conductivity of $\sigma_0\equiv e^2/4\hbar$. A comparison of the two models for $T, E_F\to 0$ is shown in Fig. 17.4. We also include a plot of real and imaginary parts in doped graphene (right panel), with the tight-binding imaginary part found using the Kramers-Kronig relation.

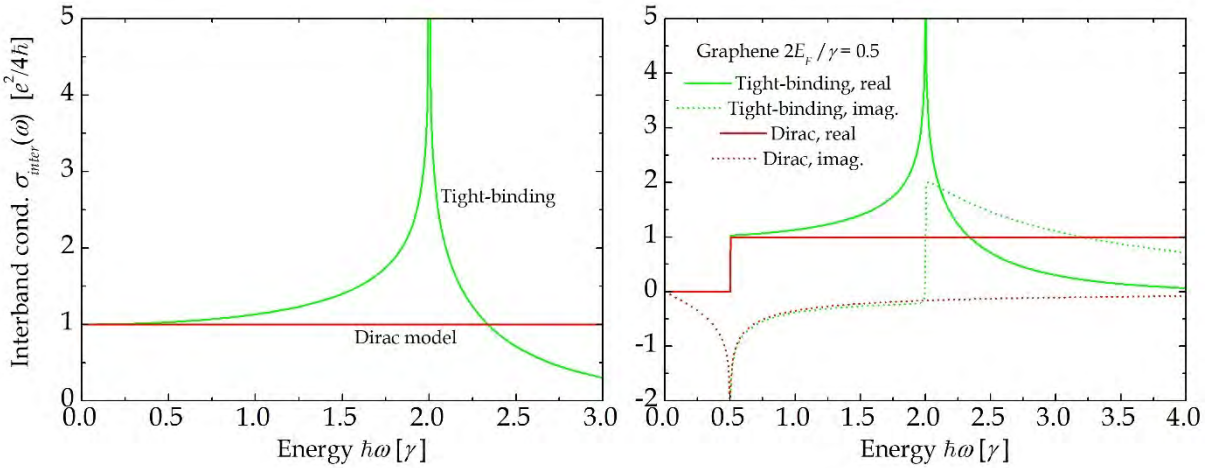


Figure 17.4. Comparison of the tight-binding and Dirac model results for the interband conductivity. Left: real part with doping. Right: real and imaginary parts of doped graphene.

In actual measurements, only the sum of intra- and interband terms is measured. As an example, this sum (real part) is illustrated in Fig. 17.5 for some reasonable values of the relevant parameters (compare to experiments in Fig. 17.6).

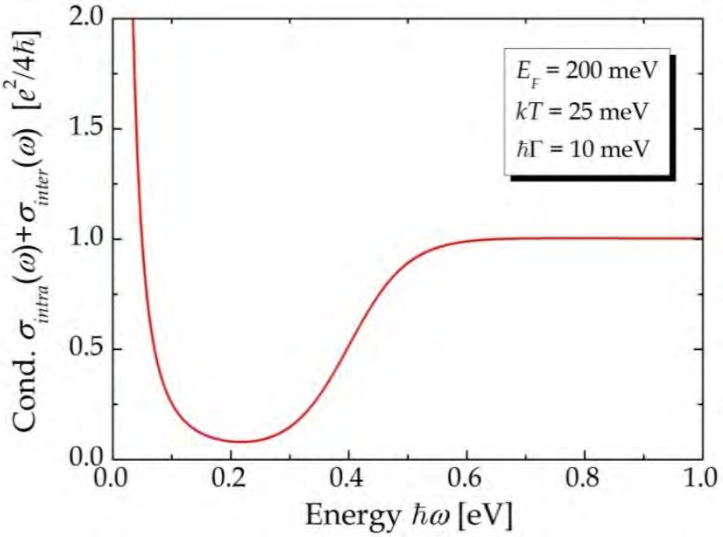


Figure 17.5. Full conductivity (real part) for a graphene sample doped to E_F = 200 meV.

In a transmission experiment, the transmittance of light with different wavelengths is measured. Hence, a measured spectrum can be related to our calculations in the following manner. First, for a graphene slab of thickness d, the dielectric constant is $\varepsilon = 1 + i\sigma/(d\varepsilon_0\omega)$. Secondly, the transmittance is $T = \exp\{-2\frac{\omega}{c} d \operatorname{Im} n\}$, where $\operatorname{Im} n$ is the imaginary part of the refractive index $n = \sqrt{\varepsilon}$. Assuming weak absorption, we can approximate $n \approx 1 + i\sigma/(2d\varepsilon_0\omega)$ and so $\operatorname{Im} n \approx \operatorname{Re}\sigma/(2d\varepsilon_0\omega)$. Approximating further, we find that

$$T = \exp\{-2\tfrac{\omega}{c} d \operatorname{Im} n\} \approx 1 - 2\tfrac{\omega}{c} d \operatorname{Im} n \approx 1 - \operatorname{Re}\sigma/(\varepsilon_0 c). \qquad (17.9)$$

Thus, for graphene in the low-energy interband range, where $\sigma \approx e^2/4\hbar$, we find $T \approx 1 - e^2/(4\varepsilon_0 c\hbar) = 1 - \pi\alpha$, where $\alpha \equiv e^2/(4\pi\varepsilon_0 c\hbar) \approx 1/137$ is the fine-structure constant. It follows that the absorption loss for monolayer graphene is roughly 2.3%. In Fig. 17.6, this is seen to be confirmed experimentally. Note also the absorption peak around 250 nm that corresponds to the tight-binding resonance in Fig. 17.4.

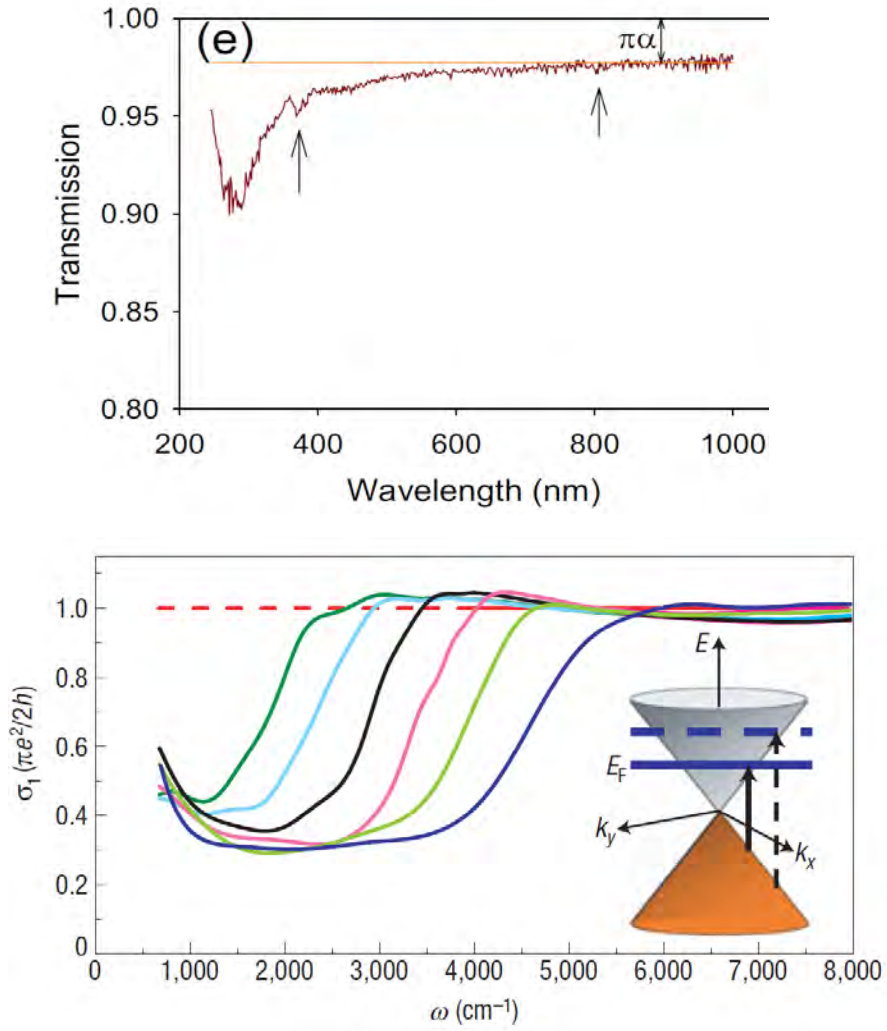


Figure 17.6. Experimental optical spectra in the UV/VIS range [V.G. Kravets *et al.*, Phys. Rev. B81, 155413 (2010)] (top) and infrared [Z.Q. Li *et al.*, Nat. Phys. 4, 532 (2008)] (bottom).

Exercise: Gapped graphene

Many ideas have been suggested to provide graphene with a band gap, so that a true semiconductor would result. Notable ones include interactions with substrates, adding electric fields to multilayer structures and antidot lattices [3]. A common phenomenological model of such "gapped" graphene is obtained with the Hamiltonian (with valley index $\eta = \pm 1$)

$$\overleftrightarrow{H} = \begin{pmatrix} \alpha & \hbar v_F(\eta q_x - i q_y) \\ \hbar v_F(\eta q_x + i q_y) & -\alpha \end{pmatrix}.$$

a) Show that for this model the eigenvalues are $E_\pm = \pm\sqrt{(\hbar v_F q)^2 + \alpha^2}$ so that a band gap of 2α results. The corresponding eigenvectors are

$$\vec{v}_+ = \frac{1}{\sqrt{2}}\begin{pmatrix} \eta e^{-i\eta\varphi}\sqrt{(E_+ + \alpha)/E_+} \\ \sqrt{(E_+ - \alpha)/E_+} \end{pmatrix}, \quad \vec{v}_- = \frac{1}{\sqrt{2}}\begin{pmatrix} -\eta e^{-i\eta\varphi}\sqrt{(E_- + \alpha)/E_-} \\ \sqrt{(E_- - \alpha)/E_-} \end{pmatrix}.$$

b) Using the new eigenvectors, demonstrate that

$$\langle + | \hat{p}_x | - \rangle = m v_F \left(\frac{\alpha}{E}\cos\varphi + i\eta \sin\varphi \right), \quad \langle + | \hat{p}_y | - \rangle = m v_F \left(\frac{\alpha}{E}\sin\varphi - i\eta\cos\varphi \right)$$

so that

$$\left|\langle + | \hat{p}_x | - \rangle\right|^2 + \left|\langle + | \hat{p}_y | - \rangle\right|^2 = m^2 v_F^2 \frac{E^2 + \alpha^2}{E^2}.$$

c) Show that the interband conductivity is given by the integral

$$\begin{aligned} \mathrm{Re}\,\sigma_{inter}(\omega) &= \frac{e^2}{\hbar^2\omega}\int_\alpha^\infty \{f(-E) - f(E)\}\frac{E^2 + \alpha^2}{E^2}\delta(2E - \hbar\omega)\,E dE \\ &= \frac{e^2}{4\hbar}\{f(-\hbar\omega/2) - f(\hbar\omega/2)\}\frac{\hbar^2\omega^2 + 4\alpha^2}{\hbar^2\omega^2}\theta(\hbar\omega - 2\alpha) \\ &= \frac{e^2}{4\hbar}\frac{\sinh\{\hbar\omega/2kT\}}{\cosh\{\hbar\omega/2kT\} + \cosh\{E_F/kT\}}\frac{\hbar^2\omega^2 + 4\alpha^2}{\hbar^2\omega^2}\theta(\hbar\omega - 2\alpha). \end{aligned} \tag{17.10}$$

References

[1] T. G. Pedersen, Phys. Rev. B67, 113106 (2003).
[2] T. G. Pedersen, A-P. Jauho, and K. Pedersen, Phys. Rev. B79, 113406 (2009).
[3] T. G. Pedersen, C. Flindt, J. Pedersen, A-P. Jauho, N.A. Mortensen and K. Pedersen, Phys. Rev. Lett. 100, 136804 (2008).

18. Models of Excitons

At this point, one might wonder about the accuracy of the optical response calculated in the previous chapters. So far, we have been treating the electrons as independent particles and the question is to what extent this is sufficient. For bulk semiconductors, the single-particle calculations predict an absorption edge that is essentially a square root, c.f. Fig. 16.1. As a classic example of the failure of this prediction, Fig. 18.1 shows a comparison between experimental spectra [1] and theoretical single-particle spectra computed from Eq.(15.3) for the wurtzite semiconductor ZnO.

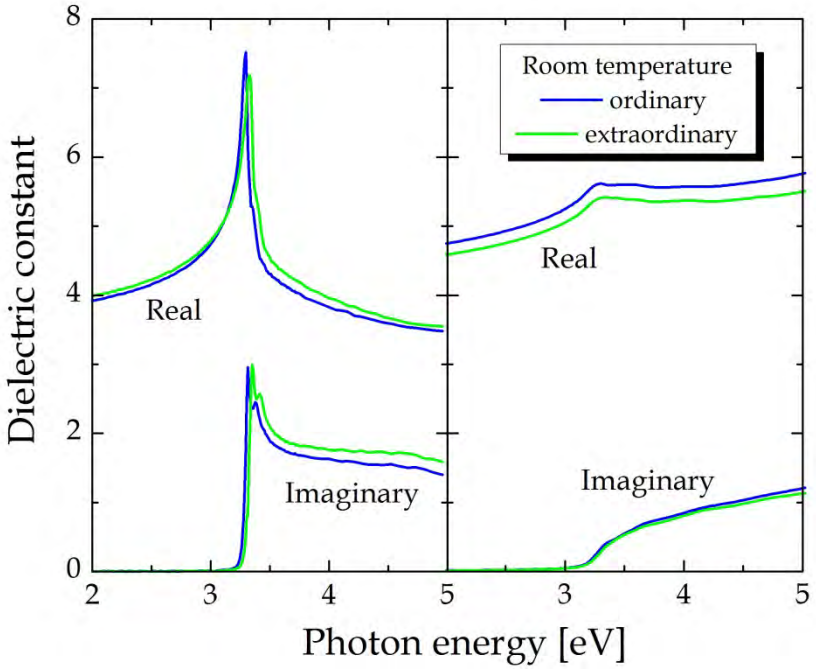


Figure 18.1. Measured spectra (left panel) and calculated single-particle spectra (right panel) for ZnO. The ordinary and extraordinary spectra correspond to light polarized perpendicular and parallel to the crystal *c*-axis, respectively.

It is obvious that single-particle theory fails miserably in this case. In most materials, however, the discrepancy is less pronounced but still noticeable. The aim of this chapter is to describe a method for the inclusion of effects beyond the single-particle response. It involves a much more accurate calculation of many-body excited states usually referred to as excitons. In subsequent chapters, the effects of excitons in low-dimensional semiconductors will be investigated. We will demonstrate that excitons are even more important for those cases.

Applying the single-particle approximation means, in effect, approximating all-electron wave functions by Slater determinants. To demonstrate this fact, we turn to the more general expression for the optical susceptibility

$$\chi(\omega) = \frac{2e^2\hbar^2}{\varepsilon_0 m^2 \Omega} \sum_{exc} \frac{\left| \left\langle 0 \left| \hat{P}_z \right| exc \right\rangle \right|^2}{E_{exc}\left[E_{exc}^2 - \hbar^2(\omega + i\Gamma)^2 \right]}. \tag{18.1}$$

This expression differs from the single-particle result Eq.(15.3) in that the sum is over all excited states $|exc\rangle$ with excitation energy E_{exc}, i.e. energy measured relative to the ground state $|0\rangle$. The prefactor of 2 (rather than 4) is used because the summation also covers spin. Also, the operator $\hat{P}_z$ is the many-body momentum operator, which for a system with 2*N* electrons is given as the sum of single-electron operators

$$\hat{P}_z = \sum_{n=1}^{2N} \hat{p}_{z,n},$$

where $\hat{p}_{z,n} = -i\hbar\, d/dz_n$ operates on the *n*'th electron coordinate only. The ground state is a Slater determinant

$$|0\rangle = |(v_1\uparrow),(v_1\downarrow),.....(v_N\uparrow),(v_N\downarrow)|$$

with all single-electron valence states occupied by spin-up and –down electrons. The total spin of the ground state is zero and since optical excitations don't flip spins we look for excited states with vanishing spin. These are so-called singlet states. To construct them, we first examine two types of singly-excited states

$$\left|(v_i\uparrow) \to (c_j\uparrow)\right\rangle = |.....,(c_j\uparrow),(v_i\downarrow),.....|$$
$$\left|(v_i\downarrow) \to (c_j\downarrow)\right\rangle = |.....,(v_i\uparrow),(c_j\downarrow),.....|$$

in which a single occupied spin-up or –down orbital is replaced by unoccupied (conduction) states with similar spin. Neither of these states have definite *total* spin. However, the combination $\left|v_i \to c_j\right\rangle \equiv \left\{ \left|(v_i\uparrow) \to (c_j\uparrow)\right\rangle + \left|(v_i\downarrow) \to (c_j\downarrow)\right\rangle \right\}/\sqrt{2}$ is a singlet with total spin S = 0. We now use the rules for matrix elements between Slater determinants [2] to calculate for the momentum

$$\langle 0 | \hat{P}_z \left| v_i \to c_j \right\rangle = \sqrt{2} \langle v_i | \hat{p}_z | c_j \rangle .$$

Moreover, the energy difference between the singlet and the ground state is simply $E_{c_j} - E_{v_i}$. Hence, Eq.(18.1) reduces exactly to Eq.(15.3) in this case.

We now wish to be somewhat more accurate. To this end, we write the excited states as linear combinations of the singlets above, i.e.

$$|exc\rangle = \sum_{ij} \Psi_{ij} \left| v_i \to c_j \right\rangle,$$

where Ψ_{ij} are unknown expansion factors. The problem is how to find matrix elements of the total Hamiltonian for any two singlets

$$H_{ij,kl} = \left\langle v_i \to c_j \right| \hat{H} \left| v_k \to c_l \right\rangle.$$

The total Hamiltonian is given by

$$\hat{H} = \sum_{n=1}^{2N} \hat{h}_n + \sum_{n<m}^{2N} V(\vec{r}_n - \vec{r}_m), \quad V(\vec{r}) = \frac{e^2}{4\pi\varepsilon_0 r},$$

where $\hat{h}_n$ is the single-electron Hamiltonian. As a start we look at the energy of the ground state

$$\langle 0|\hat{H}|0\rangle = 2\sum_{n=1}^{N} \langle v_n | \hat{h} | v_n \rangle + \sum_{n,m}^{N} \left\{ 2\langle v_n v_m | V | v_n v_m \rangle - \langle v_n v_m | V | v_m v_n \rangle \right\}.$$

Note that the additional factors of 2 appear because the spin-summation has already been performed. Next, we look at the diagonal elements for the state $\left|(v_i\uparrow) \to (c_j\uparrow)\right\rangle$. Compared to the ground state, v_i should be replaced by c_j. It follows that

$$\left\langle (v_i\uparrow) \to (c_j\uparrow) \right| \hat{H} \left| (v_i\uparrow) \to (c_j\uparrow) \right\rangle = \langle 0|\hat{H}|0\rangle + \left\langle c_j \left| \hat{h} \right| c_j \right\rangle - \langle v_i | \hat{h} | v_i \rangle$$
$$+ \sum_{n\neq i} \left\{ 2\left\langle v_n c_j \left| V \right| v_n c_j \right\rangle - \left\langle v_n c_j \left| V \right| c_j v_n \right\rangle \right\} + \left\langle v_i c_j \left| V \right| v_i c_j \right\rangle - \sum_{n} \left\{ 2\langle v_n v_i | V | v_n v_i \rangle - \langle v_n v_i | V | v_i v_n \rangle \right\}.$$

We now introduce the *quasi-particle* energies

$$\tilde{E}_{c_j} \equiv \left\langle c_j \left| \hat{h} \right| c_j \right\rangle + \sum_{n=1}^{N} \left\{ 2\left\langle v_n c_j \left| V \right| v_n c_j \right\rangle - \left\langle v_n c_j \left| V \right| c_j v_n \right\rangle \right\}$$
$$\tilde{E}_{v_i} \equiv \langle v_i | \hat{h} | v_i \rangle + \sum_{n=1}^{N} \left\{ 2\langle v_n v_i | V | v_n v_i \rangle - \langle v_n v_i | V | v_i v_n \rangle \right\}.$$

Note that there is no restriction on the summations, i.e. all valence states are summed. In terms of these quantities we have

$$\left\langle (v_i\uparrow) \to (c_j\uparrow) \right| \hat{H} \left| (v_i\uparrow) \to (c_j\uparrow) \right\rangle = \langle 0|\hat{H}|0\rangle + \tilde{E}_{c_j} - \tilde{E}_{v_i} - \left\langle v_i c_j \left| V \right| v_i c_j \right\rangle + \left\langle v_i c_j \left| V \right| c_j v_i \right\rangle,$$

where the last two terms serve to correct the unrestricted summations in the quasi-particle energies. The exact same expression is obtained if the spin-down Slater determinant is considered. The cross-term yields

$$\left\langle (v_i\uparrow)\to(c_j\uparrow)\middle|\hat{H}\middle|(v_i\downarrow)\to(c_j\downarrow)\right\rangle = \left\langle v_i c_j\middle|V\middle|c_j v_i\right\rangle.$$

Combining, we find the full diagonal matrix element for the singlet excitation

$$\left\langle v_i\to c_j\middle|\hat{H}\middle|v_i\to c_j\right\rangle = \langle 0|\hat{H}|0\rangle + \tilde{E}_{c_j} - \tilde{E}_{v_i} - \left\langle v_i c_j\middle|V\middle|v_i c_j\right\rangle + 2\left\langle v_i c_j\middle|V\middle|c_j v_i\right\rangle.$$

It can be shown that coupling between singly excited states and the ground state is identically zero, i.e. that $\langle 0|\hat{H}\left|v_i\to c_j\right\rangle = 0$ [3]. The non-zero off-diagonal terms follow in much the same style as the diagonal ones

$$\left\langle (v_i\uparrow)\to(c_j\uparrow)\middle|\hat{H}\middle|(v_k\uparrow)\to(c_l\uparrow)\right\rangle = \left\langle c_j v_k\middle|V\middle|v_i c_l\right\rangle - \left\langle c_j v_k\middle|V\middle|c_l v_i\right\rangle$$
$$\left\langle (v_i\uparrow)\to(c_j\uparrow)\middle|\hat{H}\middle|(v_k\downarrow)\to(c_l\downarrow)\right\rangle = \left\langle c_j v_k\middle|V\middle|v_i c_l\right\rangle.$$

As $\langle 0|\hat{H}|0\rangle$ is the ground state energy, which we use as a zero-point of energy, we finally find

$$H_{ij,kl} = \left[\tilde{E}_{c_j} - \tilde{E}_{v_i}\right]\delta_{ik}\delta_{jl} - \left\langle c_j v_k\middle|V\middle|c_l v_i\right\rangle + 2\left\langle c_j v_k\middle|V\middle|v_i c_l\right\rangle. \tag{18.2}$$

The matrix problem then reads as

$$\sum_{kl} H_{ij,kl}\Psi_{kl} = E_{exc}\Psi_{ij},$$

from which exciton wave functions and energies are computed. In turn, the exciton momentum matrix elements become

$$\left\langle 0\middle|\hat{P}_z\middle|exc\right\rangle = \sqrt{2}\sum_{ij}\Psi_{ij}\left\langle v_i\middle|\hat{p}_z\middle|c_j\right\rangle. \tag{18.3}$$

We now specialize to periodic solids for which orbitals are labeled by a band index (v or c) and a wave vector $\vec{k}$. In an optical process, the only relevant excitations are those that preserve $\vec{k}$ (neglecting the small momentum lost/gained by the photon). Thus, the singlets are of the type $\left|vc\vec{k}\right\rangle \equiv \left|v\vec{k}\to c\vec{k}\right\rangle$. In turn, the sought matrix elements are

$$H_{vc\vec{k},v'c'\vec{k}'} = \left[\tilde{E}_{c\vec{k}} - \tilde{E}_{v\vec{k}}\right]\delta_{vv'}\delta_{cc'}\delta_{\vec{k}\vec{k}'} - \left\langle vc\vec{k}\middle|V_C\middle|v'c'\vec{k}'\right\rangle + 2\left\langle vc\vec{k}\middle|V_x\middle|v'c'\vec{k}'\right\rangle, \tag{18.4}$$

where Coulomb and exchange matrix elements are defined as

$$\left\langle vc\vec{k}\left|V_C\right|v'c'\vec{k}'\right\rangle = \iint \psi^*_{c\vec{k}}(\vec{r})\psi_{c'\vec{k}'}(\vec{r})V(\vec{r}-\vec{r}')\psi_{v\vec{k}}(\vec{r}')\psi^*_{v'\vec{k}'}(\vec{r}')d^3rd^3r'$$
$$\left\langle vc\vec{k}\left|V_x\right|v'c'\vec{k}'\right\rangle = \iint \psi^*_{c\vec{k}}(\vec{r})\psi_{c'\vec{k}'}(\vec{r}')V(\vec{r}-\vec{r}')\psi_{v\vec{k}}(\vec{r})\psi^*_{v'\vec{k}'}(\vec{r}')d^3rd^3r'.$$

In a more rigorous derivation [4], it turns out that the Coulomb interaction should be screened by surrounding charges, so that introducing the dielectric constant ε we find

$$\left\langle vc\vec{k}\left|V_C\right|v'c'\vec{k}'\right\rangle = \frac{1}{\varepsilon}\iint \psi^*_{c\vec{k}}(\vec{r})\psi_{c'\vec{k}'}(\vec{r})V(\vec{r}-\vec{r}')\psi_{v\vec{k}}(\vec{r}')\psi^*_{v'\vec{k}'}(\vec{r}')d^3rd^3r' .$$

This full matrix equation (using the screened Coulomb interaction) is known as the Bethe-Salpeter equation.

18.1 Wannier Model

The framework above is terribly complicated and extremely difficult to handle numerically. Fortunately, a much simplified version can be applied in many cases provided the Coulomb interaction is not too strong. To derive this "Wannier" model we first note that the eigenstates of a periodic solid can be written as

$$\psi_{v\vec{k}}(\vec{r}) = \frac{1}{\sqrt{\Omega}}u_{v\vec{k}}(\vec{r})e^{i\vec{k}\cdot\vec{r}},$$

where $u_{v\vec{k}}$ is the lattice-periodic part normalized so that

$$\frac{1}{\Omega_{UC}}\int_{UC}\left|u_{v\vec{k}}(\vec{r})\right|^2 d^3r = 1$$

with the integral taken over the unit cell volume Ω_{UC}. We first turn to the Coulomb matrix element. The product $\psi_{v\vec{k}}(\vec{r}')\psi^*_{v'\vec{k}'}(\vec{r}')V(\vec{r}-\vec{r}')$ has a rapidly varying periodic part and slow part. In analogy with Chapter 15, we will approximate the integral

$$\int \psi_{v\vec{k}}(\vec{r}')\psi^*_{v'\vec{k}'}(\vec{r}')V(\vec{r}-\vec{r}')d^3r' \approx \frac{1}{\Omega\Omega_{UC}}\int u_{v\vec{k}}(\vec{r}')u^*_{v'\vec{k}'}(\vec{r}')d^3r'\int e^{i(\vec{k}-\vec{k}')\cdot\vec{r}'}V(\vec{r}-\vec{r}')d^3r'$$
$$\equiv \frac{1}{\Omega}I_{v'\vec{k}',v\vec{k}}\int e^{i(\vec{k}-\vec{k}')\cdot\vec{r}'}V(\vec{r}-\vec{r}')d^3r', \quad I_{v'\vec{k}',v\vec{k}} = \frac{1}{\Omega_{UC}}\int u^*_{v'\vec{k}'}(\vec{r}')u_{v\vec{k}}(\vec{r}')d^3r'.$$

Making a similar approximation for the *r*-integration, we then find

$$\left\langle vc\vec{k}\left|V_C\right|v'c'\vec{k}'\right\rangle = \frac{1}{\varepsilon\Omega^2} I_{v'\vec{k}',v\vec{k}} I^*_{c'\vec{k}',c\vec{k}} \iint e^{i(\vec{k}-\vec{k}')\cdot(\vec{r}'-\vec{r})} V(\vec{r}-\vec{r}') d^3r d^3r'$$
$$= \frac{1}{\varepsilon\Omega} I_{v'\vec{k}',v\vec{k}} I^*_{c'\vec{k}',c\vec{k}} \int e^{i(\vec{k}'-\vec{k})\cdot\vec{r}} V(\vec{r})\, d^3r.$$

In a completely analogous manner, the exchange integral becomes

$$\left\langle vc\vec{k}\left|V_x\right|v'c'\vec{k}'\right\rangle = \frac{1}{\Omega} I_{c\vec{k},v\vec{k}} I^*_{c'\vec{k}',v'\vec{k}'} \int V(\vec{r})\, d^3r\,.$$

Now, at $\vec{k}=\vec{k}'$ we have $I_{\alpha\vec{k},\beta\vec{k}} = \delta_{\alpha\beta}$. Hence, if the *k*-dependence is not too severe we may assume that $I_{\alpha\vec{k},\beta\vec{k}'} = \delta_{\alpha\beta}$ hold approximately to a reasonable degree even when $\vec{k} \neq \vec{k}'$. In this case, we find the much simpler approximations

$$\left\langle vc\vec{k}\left|V_C\right|v'c'\vec{k}'\right\rangle \approx \frac{1}{\varepsilon\Omega} \int e^{i(\vec{k}'-\vec{k})\cdot\vec{r}} V(\vec{r})\, d^3r \delta_{vv'}\delta_{cc'}$$
$$\left\langle vc\vec{k}\left|V_x\right|v'c'\vec{k}'\right\rangle \approx 0.$$

Hence, the bands decouple and we can focus on a single pair *v* and *c*. The Hamiltonian matrix elements become

$$H_{\vec{k},\vec{k}'} = \left[\tilde{E}_{c\vec{k}} - \tilde{E}_{v\vec{k}}\right]\delta_{\vec{k}\vec{k}'} - \frac{1}{\varepsilon\Omega}\int e^{i(\vec{k}'-\vec{k})\cdot\vec{r}} V(\vec{r})\, d^3r\,,$$

where we have skipped the band indices on the matrix elements. Also, the unknown expansion coefficients can be re-labeled according to $\Psi_{ij} \to \Psi_{\vec{k}}$. It follows that the exciton eigenvalue problem is now

$$\left[\tilde{E}_{c\vec{k}} - \tilde{E}_{v\vec{k}}\right]\Psi_{\vec{k}} - \frac{1}{\varepsilon\Omega}\sum_{\vec{k}'}\int e^{i(\vec{k}'-\vec{k})\cdot\vec{r}} V(\vec{r})\, d^3r \Psi_{\vec{k}'} = E_{exc}\Psi_{\vec{k}}\,.$$

Here, the $\vec{k}'$ summation can be turned into an integral, i.e.

$$\left[\tilde{E}_{c\vec{k}} - \tilde{E}_{v\vec{k}}\right]\Psi_{\vec{k}} - \frac{1}{(2\pi)^3\varepsilon}\int\int e^{i(\vec{k}'-\vec{k})\cdot\vec{r}} V(\vec{r})\, d^3r \Psi_{\vec{k}'} d^3k' = E_{exc}\Psi_{\vec{k}}\,.$$

The final approximation of the Wannier model consists in applying the effective mass dispersion for both bands so that

$$\tilde{E}_{c\vec{k}} \approx E_g + \frac{\hbar^2k^2}{2m_e}, \quad \tilde{E}_{v\vec{k}} \approx -\frac{\hbar^2k^2}{2m_h} \quad \Rightarrow \quad \tilde{E}_{c\vec{k}} - \tilde{E}_{v\vec{k}} \approx E_g + \frac{\hbar^2k^2}{2m_{eh}}.$$

In this way, the *k*-space eigenproblem can be transformed into physical space by means of a simple inverse Fourier transform:

$$\frac{\sqrt{\Omega}}{(2\pi)^3}\int e^{i\vec{k}\cdot\vec{r}}\Psi_{\vec{k}}d^3k=\Psi_{exc}(\vec{r}),\quad \frac{\sqrt{\Omega}}{(2\pi)^3}\int e^{i\vec{k}\cdot\vec{r}}k^2\Psi_{\vec{k}}d^3k=-\nabla^2\Psi_{exc}(\vec{r})\,.$$

The $\sqrt{\Omega}$ factors are inserted to ensure that $\Psi_{exc}(\vec{r})$ is normalized. Moreover, the Coulomb term above is simply the convolution between the wave function and the Coulomb potential and, hence, we finally find

$$\left(E_g-\frac{\hbar^2\nabla^2}{2m_{eh}}\right)\Psi_{exc}(\vec{r})-\frac{1}{\varepsilon}V(\vec{r})\Psi_{exc}(\vec{r})=E_{exc}\Psi_{exc}(\vec{r})\,. \tag{18.5}$$

It is apparent that this so-called Wannier equation is mathematically similar to the Schrödinger equation for the hydrogen atom. The differences are that m_{eh} replaces the reduced electron-nucleus mass and that ε screens the Coulomb term. The physical interpretation is that the positive hole and negative electron interact via an attractive Coulomb potential. We note that only the relative motion of the electron-hole pair is present in the problem, so that the states have a vanishing centre-of-mass momentum. This is a consequence of our retaining only $v\vec{k}\to c\vec{k}$ excitations in the expansion, i.e. neglecting photon momentum. Hence, the centre-of-mass momentum must vanish both before and after the photon is emitted/absorbed.

To eventually calculate the optical properties, we need the momentum matrix element Eq.(18.3), which now reads as

$$\left\langle 0\left|\hat{P}_z\right|exc\right\rangle=\sqrt{2}\sum_{\vec{k}}\Psi_{\vec{k}}\left\langle v\vec{k}\left|\hat{p}_z\right|c\vec{k}\right\rangle=\frac{\sqrt{2\Omega}}{(2\pi)^3}\int\left\langle v\vec{k}\left|\hat{p}_z\right|c\vec{k}\right\rangle\Psi_{\vec{k}}d^3k\,.$$

As a simplification, we may take the single-electron momentum matrix element independent of $\vec{k}$ so that $\left\langle v\vec{k}\left|\hat{p}_z\right|c\vec{k}\right\rangle\approx p_{vc}$, which means that

$$\left\langle 0\left|\hat{P}_z\right|exc\right\rangle\approx\frac{\sqrt{2\Omega}}{(2\pi)^3}p_{vc}\int\Psi_{\vec{k}}d^3k=\sqrt{2\Omega}\,p_{vc}\Psi_{exc}(0)\,,$$

where $\Psi_{exc}(0)$ is the exciton wave function in physical space evaluated at the origin. This leads to a simple expression for the optical response

$$\chi(\omega)=\frac{4e^2\hbar^2\,|\,p_{vc}\,|^2}{\varepsilon_0 m^2}\sum_{exc}\frac{\left|\Psi_{exc}(0)\right|^2}{E_{exc}\left[E_{exc}^2-\hbar^2(\omega+i\Gamma)^2\right]}. \tag{18.6}$$

In this approximation, only exciton states that are finite at the origin ("*s*-type") contribute to the response. In the following, we evaluate the imaginary part in the limit of vanishing broadening

$$\mathrm{Im}\,\chi(\omega) = \frac{2\pi e^2 \,|\, p_{vc} \,|^2}{\varepsilon_0 m^2 \omega^2} \sum_{exc} \left|\Psi_{exc}(0)\right|^2 \delta(E_{exc} - \hbar\omega) \tag{18.7}$$

for bulk and low-dimensional cases.

Exercise: Natural exciton units

The Wannier equation Eq.(18.5) is formulated in SI units and it is advantageous to switch to more natural units.

a) Show that using $a_B^* = 4\pi\varepsilon\varepsilon_0\hbar^2 / m_{eh}e^2$ as the unit of length and $Ry^* = \hbar^2 / 2m_{eh}a_B^{*2}$ as the energy unit, the Wannier equation reduces to

$$\left(E_g - \nabla^2\right)\Psi_{exc}(\vec{r}) - \frac{2}{r}\Psi_{exc}(\vec{r}) = E_{exc}\Psi_{exc}(\vec{r})\,.$$

b) Show that $a_B^* = 0.529\text{Å}\cdot\varepsilon m / m_{eh}$ and $Ry^* = 13.6\text{eV}\cdot m_{eh} / \varepsilon^2 m$ (m is the free electron mass) and evaluate both for GaAs: $m_e = 0.066\,m$, $m_h = 0.5\,m$ (heavy hole), $\varepsilon = 12.9$ and for ZnO: $m_e = 0.28\,m$, $m_h = 0.59\,m$ (heavy hole), $\varepsilon = 6.7$ (including effective phonon contribution).

References

[1] G.E. Jellison and L.A. Boatner, Phys. Rev. B58, 3586 (1998).
[2] J.P. Dahl *Introduction to the Quantum World of Atoms and Molecules* (World Scientific, Singapore, 2001)
[3] L. Salem *The Molecular Orbital Theory of Conjugated Systems* (W.A. Benjamin, New York, 1966).
[4] M. Rohlfing and S.G. Louie, Phys. Rev. B62, 4927 (2000)

19. Excitons in Bulk and Two-dimensional Semiconductors

The Wannier model derived in the previous chapter provides a simple framework for the inclusion of excitons in the optical properties of semiconductors. In this chapter, we will evaluate the optical response for 3D and 2D semiconductors. As will become apparent, excitonic effects in low-dimensional semiconductors are hugely enhanced. The reason is that excitonic effects originate from the attractive interaction between electrons and holes. The stronger the attraction, the more pronounced the excitonic corrections to the response. But an additional confinement will also tend to localize electrons and holes in the same region of space and, hence, increase the overlap of their wave functions. This may increase the Coulomb attraction significantly.

We begin by investigating the 3D or bulk case. The starting point is the Wannier equation written in natural exciton units (using a_B^* and Ry^* as the units of length and energy, respectively)

$$\left(E_g - \nabla^2\right)\Psi_{exc}(\vec{r}) - \frac{2}{r}\Psi_{exc}(\vec{r}) = E_{exc}\Psi_{exc}(\vec{r})\,.$$

The expression for the susceptibility Eq.(18.6) shows that only *s*-type eigenstates are relevant. These are nothing but the usual eigenstates of the hydrogen atom. For the bound states with energy below E_g we have in terms of the principle quantum number n = 1, 2, 3,…,

$$\Psi_n(\vec{r}) = \frac{1}{\sqrt{\pi}n^{5/2}} L_{n-1}^1(2r/n)e^{-r/n}, \quad E_n = E_g - \frac{1}{n^2}\,.$$

Here, L_n^m is an associated Laguerre polynomial. As we use Ry^* as the unit of energy it follows that these states form a series of resonances located spectrally between E_g and $E_g - Ry^*$. Hence, bound excitons lead to discrete absorption peaks below the band gap. The continuum states with $E > E_g$ are harder, in part because they cannot be normalized in the usual sense. To circumvent this problem, we enclose the exciton in a large sphere of radius R [1]. We let the wave function vanish on the surface of the sphere and eventually let the radius go to infinity. We write $\alpha^2 = E - E_g$ so that the radial Schrödinger equation becomes

$$-\frac{1}{r}\frac{d^2}{dr^2}r\Psi(r) - \frac{2}{r}\Psi(r) = \alpha^2\Psi(\vec{r})\,.$$

A solution is

$$\Psi_\alpha(r) = Ne^{-i\alpha r} F(1+i/\alpha, 2; 2i\alpha r),$$

where F is called a confluent hypergeometric function (see e.g. [2] for details). We need to study the behavior as r approaches R to fulfill the boundary conditions. Since R is eventually taken very large, we can use the asymptotic expression for F. We can also apply the asymptotic expression to determine the normalization constant N because we mainly integrate over a region with r large. The asymptotic limit is

$$\begin{aligned}\left|F(1+i/\alpha, 2; 2i\alpha r)\right|^2 &\approx \left|\frac{(2i\alpha r)^{-1-i/\alpha}\, e^{\pi(i-1/\alpha)}}{\Gamma(1-i/\alpha)} + \frac{e^{2i\alpha r}(2i\alpha r)^{-1+i/\alpha}}{\Gamma(1+i/\alpha)}\right|^2 \\ &\approx \frac{e^{-\pi/\alpha}\sinh(\pi/\alpha)}{\pi\alpha r^2}\cos^2(\alpha r).\end{aligned}$$

Hence, to normalize we integrate within the sphere

$$1 = 4\pi\int_0^R \left|\Psi_\alpha(r)\right|^2 r^2 dr = N^2\,\frac{2\sinh(\pi/\alpha)}{\alpha e^{\pi/\alpha}}R\,.$$

It follows that the wave function at the origin is given by

$$\left|\Psi(0)\right|^2 = \begin{cases} \dfrac{1}{\pi n^3}, & E = E_g - \dfrac{1}{n^2} \qquad \text{bound states } E < E_g \\ \dfrac{\alpha e^{\pi/\alpha}}{2R\sinh(\pi/\alpha)}, & E = E_g + \alpha^2 \quad \text{continuum } E > E_g. \end{cases}$$

The main remaining problem lies in summing the continuum solutions while taking $R \to \infty$. For this purpose we introduce the weighted joint density of states

$$S(\omega) = \sum_{\text{cont.}} \left|\Psi(0)\right|^2 \delta(E - \hbar\omega),$$

where the sum is over all continuum states. From the asymptotic expression for the wave function above, it is clear that the allowed values of α fulfill $\alpha R = \pi(p+1/2)$, where p is an integer. Hence, the distance between allowed values of α is $\delta\alpha = \pi/R$. We can therefore write the sum as

$$S(\omega) = \sum_\alpha \frac{\alpha e^{\pi/\alpha}}{2\pi\sinh(\pi/\alpha)}\delta(E_g + \alpha^2 - \hbar\omega)\delta\alpha\,.$$

Taking $R \to \infty$ means $\delta\alpha \to 0$ and so, converting to an integral, we find the simple result

$$S(\omega) = \int_0^\infty \frac{\alpha e^{\pi/\alpha}}{2\pi \sinh(\pi/\alpha)} \delta(E_g + \alpha^2 - \hbar\omega) d\alpha$$
$$= \frac{\theta(\hbar\omega - E_g) e^{\pi/\sqrt{\hbar\omega - E_g}}}{4\pi \sinh(\pi/\sqrt{\hbar\omega - E_g})}, \tag{19.1}$$

where θ is the unit step function. Reverting to SI units means reintroducing a_B^* and Ry^*. Since S has units of $\text{length}^{-3} \cdot \text{energy}^{-1}$ we need to divide by a factor $a_B^{*3} Ry^* = \hbar^3/(2m_{eh})^{3/2}/\sqrt{Ry^*}$ using the relation $a_B^{*2} Ry^* = \hbar^2/2m_{eh}$ and we can then write the final result as

$$\text{Im}\,\chi(\omega) = \frac{e^2 |p_{vc}|^2 \sqrt{Ry^*}}{2\varepsilon_0 m^2 \omega^2} \left(\frac{2m_{eh}}{\hbar^2}\right)^{3/2} \left\{ \sum_n \frac{4}{n^3} \delta(\Delta + 1/n^2) + \frac{\theta(\Delta) e^{\pi/\sqrt{\Delta}}}{\sinh(\pi/\sqrt{\Delta})} \right\},$$

with $\Delta = (\hbar\omega - E_g)/Ry^*$. This formula is called the 3D Elliott formula. Based on this result, the real part can be obtained using the Kramers-Kronig relations and broadening can be introduced through convolution with a Lorentzian broadening function. This leads to the expression

$$\chi(\omega) = \frac{e^2\hbar^2 |p_{vc}|^2}{\pi^2 \varepsilon_0 m^2 E_g^{3/2}} \left(\frac{2m_{eh}}{\hbar^2}\right)^{3/2} X_3(w),$$

where $X_3(w)$ is now the excitonic susceptibility function given by [3]

$$X_3(w) = \frac{\pi Ry^{*1/2} E_g^{3/2}}{2(\hbar w)^2} \left\{ \sum_n \frac{8Ry^*(\hbar w)^2}{n^3 E_n \left[E_n^2 - (\hbar w)^2\right]} + g\left(\sqrt{\frac{Ry^*}{E_g - \hbar w}}\right) + g\left(\sqrt{\frac{Ry^*}{E_g + \hbar w}}\right) - 2g\left(\sqrt{\frac{Ry^*}{E_g}}\right) \right\}$$

$$g(z) \equiv 2\ln(z) + 2\psi(z) + \frac{1}{z}, \quad \psi(z): \text{ digamma function,}$$

where $w = \omega + i\Gamma$ and $E_n = E_g - \dfrac{Ry^*}{n^2}$ is the energy of the n'th bound exciton. In Fig. 19.1, the real and imaginary parts are plotted for a 3D material with E_g = 1.5 eV and varying values of the effective Rydberg and broadening.

To see that the Elliott formula has the correct limiting behavior if Coulomb effects become negligible (if e.g. ε becomes very large) we should take $Ry^* \to 0$ and accordingly $\Delta \to \infty$. We note that

$$\lim_{\Delta\to\infty}\frac{e^{\pi/\sqrt{\Delta}}}{\sinh(\pi/\sqrt{\Delta})}=\frac{\sqrt{\Delta}}{\pi}.$$

Also, the contribution from bound excitons vanishes. Hence, we find the approximate expression

$$\mathrm{Im}\,\chi(\omega)=\frac{e^2\,|\,p_{vc}\,|^2}{2\pi\varepsilon_0 m^2\omega^2}\left(\frac{2m_{eh}}{\hbar^2}\right)^{3/2}\sqrt{\hbar\omega-E_g}\,\theta(\hbar\omega-E_g).$$

This expression is identical to the imaginary part of the independent-electron result Eq.(16.2) if vanishing broadening is assumed.

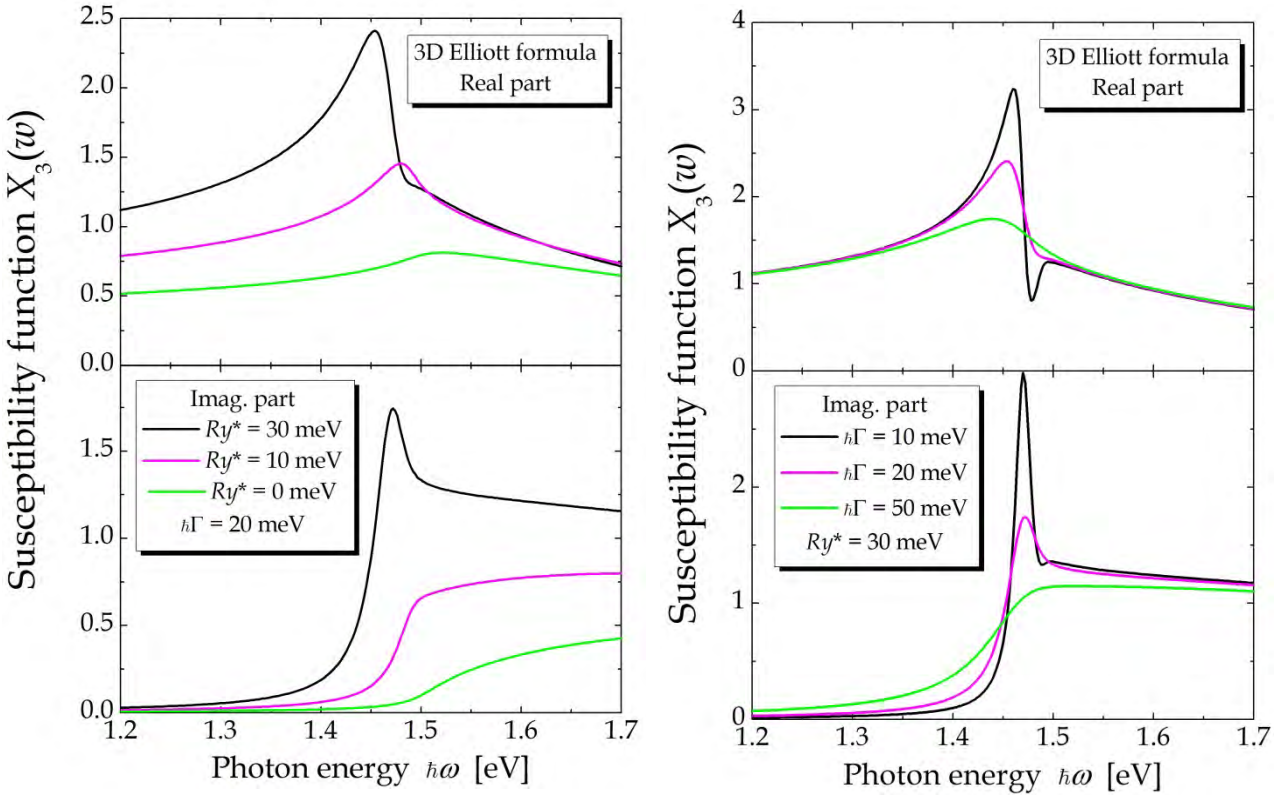


Figure 19.1. The excitonic susceptibility function for a bulk semiconductor. The plots illustrate the effect of varying the strength of the Coulomb interaction (left panels) and broadening (right panels).

19.1 Excitons in Quantum Wells

We will consider the case of an electron-hole pair confined to a thin quantum well. It is assumed that light is polarized along the quantum well plane so that the motion perpendicular to the plane (taken as the z direction) is not excited. As in the 3D case, we therefore require a vanishing centre-of-mass momentum in the plane. Hence, for the in-plane motion only the relative part is retained and the electron-hole pair is characterized by a Hamiltonian (in polar coordinates)

$$\hat{H}_{eh} = \hat{h}_e(z_e) + \hat{h}_h(z_h) + E_g - \frac{d^2}{dr^2} - \frac{1}{r}\frac{d}{dr} - \frac{1}{r^2}\frac{d^2}{d\theta^2} - \frac{2}{\sqrt{r^2 + (z_e - z_h)^2}}. \qquad (19.2)$$

Here, $\hat{h}_e(z_e) = -m_e^{-1} d^2 / dz_e^2 + V_e(z_e)$ is the Hamiltonian for the z motion of the electron confined by the quantum well potential V_e and $\hat{h}_h(z_h)$ is the analogous term for the hole. We will now specialize to the idealized case of an exciton confined to an extremely thin quantum well. The strong confinement means that $z_e \approx z_h$ and so we will use the simplifying approximation

$$\frac{2}{\sqrt{r^2 + (z_e - z_h)^2}} \approx \frac{2}{r}. \qquad (19.3)$$

This means that the Hamiltonian becomes a sum $\hat{H}_{eh} = \hat{h}_e(z_e) + \hat{h}_h(z_h) + \hat{H}(r)$, where $\hat{H}(r)$ describes the in-plane relative motion. As a consequence, the exciton wave function is a product $\Psi_{exc}(\vec{r}_e, \vec{r}_h) = \varphi_e(z_e)\varphi_h(z_h)\Psi(\vec{r})$, where φ_e is an eigenstate of $\hat{h}_e$ and similarly for the hole. We assume that the z-motion of both electrons and holes are frozen in the lowest eigenstates. This amounts to replacing the true band gap E_g by the effective one $\tilde{E}_g = E_g + E_1^e + E_1^h$, where $E_1^e + E_1^h$ is the sum of electron and hole quantization energies (in previous chapter we used the notation E_{g11} for $\tilde{E}_g$). The in-plane relative motion then leads to a purely two-dimensional Wannier equation given in polar coordinates by

$$\left(\tilde{E}_g - \frac{d^2}{dr^2} - \frac{1}{r}\frac{d}{dr} - \frac{1}{r^2}\frac{d^2}{d\theta^2}\right)\Psi(\vec{r}) - \frac{2}{r}\Psi(\vec{r}) = E\Psi(\vec{r}).$$

This problem is mathematically identical to a hydrogen atom in two dimensions. The *s*-type bound states (with no angular dependence) are quite similar to the 3D case and can be written

$$\Psi_n(\vec{r}) = \frac{1}{\sqrt{\pi}(n+\frac{1}{2})^{3/2}} L_n(2r/(n+\tfrac{1}{2}))e^{-r/(n+\frac{1}{2})}, \quad E_n = \tilde{E}_g - \frac{1}{(n+\frac{1}{2})^2}.$$

Here, the principle quantum number n is again an integer, but now the allowed values include zero, i.e. n = 0, 1, 2, … The most important feature of this result compared to the 3D case is that the binding energy of the lowest exciton is now $4Ry^*$ whereas excitons in bulk were bound by no more that $1Ry^*$. This is a direct manifestation of the increased electron-hole overlap in low-dimensional geometries. The continuum states are also highly similar to the bulk case and found to be

$$\Psi_\alpha(r) = Ne^{-i\alpha r} F(\tfrac{1}{2} + i/\alpha, 1; 2i\alpha r),$$

with a normalization condition that is now

$$1 = 2\pi \int_0^R |\Psi_\alpha(r)|^2 \, r dr = N^2 \frac{2\cosh(\pi/\alpha)}{\alpha e^{\pi/\alpha}} R.$$

This result corresponds to normalization within a circle of radius R. Eventually, it follows that the wave function at the origin is given by

$$|\Psi(0)|^2 = \begin{cases} \dfrac{1}{\pi(n+\frac{1}{2})^3}, & E = \tilde{E}_g - \dfrac{1}{(n+\frac{1}{2})^2} \text{ bound states } \; E < \tilde{E}_g \\ \dfrac{\alpha e^{\pi/\alpha}}{2R\cosh(\pi/\alpha)}, & E = \tilde{E}_g + \alpha^2 \;\; \text{continuum } \; E > \tilde{E}_g. \end{cases}$$

Summing over the continuum states leads to a result identical to Eq.(19.1) except that sinh is replaced by cosh. Finally, we arrive at a 2D Elliott formula describing the excitonic absorption in an ultrathin quantum well of width d:

$$\text{Im}\,\chi(\omega) = \frac{e^2 |\, p_{vc} \,|^2 \, m_{eh}}{\varepsilon_0 m^2 \hbar^2 \omega^2 d} \left\{ \sum_n \frac{4}{(n+\frac{1}{2})^3} \delta(\Delta + 1/(n+\tfrac{1}{2})^2) + \frac{\theta(\Delta) e^{\pi/\sqrt{\Delta}}}{\cosh(\pi/\sqrt{\Delta})} \right\}.$$

Here, $\Delta = (\hbar\omega - \tilde{E}_g)/Ry^*$ and the relation $a_B^{*2} Ry^* = \hbar^2/2m_{eh}$ has been utilized. Also, it is clear that in the limit of independent electrons we find

$$\text{Im}\,\chi(\omega) = \frac{e^2 |\, p_{vc} \,|^2 \, m_{eh}}{\varepsilon_0 m^2 \hbar^2 \omega^2 d} \theta(\hbar\omega - \tilde{E}_g),$$

in agreement with Chapter 16. Again, broadening can be introduced and the real part can be added so that the full 2D excitonic susceptibility becomes

$$\chi(\omega) = \frac{2e^2 |\, p_{vc} \,|^2 \, m_{eh}}{\pi \varepsilon_0 m^2 \tilde{E}_g^2 d} X_2(w),$$

with the susceptibility function [4]

$$X_2(w) = \frac{2\tilde{E}_g^2}{(\hbar w)^2} \left\{ \sum_n \frac{8Ry^* (\hbar w)^2}{(n+\frac{1}{2})^3 E_n \left[E_n^2 - (\hbar w)^2 \right]} + f\left(\sqrt{\frac{Ry^*}{\tilde{E}_g - \hbar w}} \right) + f\left(\sqrt{\frac{Ry^*}{\tilde{E}_g + \hbar w}} \right) - 2f\left(\sqrt{\frac{Ry^*}{\tilde{E}_g}} \right) \right\}$$

$$f(z) \equiv 2\ln(z) + 2\psi(z + \tfrac{1}{2}), \quad \psi(z): \text{ digamma function},$$

where $w=\omega+i\Gamma$ and $E_n=\tilde{E}_g-Ry^*/(n+\frac{1}{2})^2$ is the energy of the n'th bound exciton. In Fig. 19.2, we have plotted this result. It is clearly seen that excitons lead to a complete reorganization of the spectra. A huge excitonic resonance located $4Ry^*$ below the band gap emerges but also the continuum part of the spectrum is severely modified.

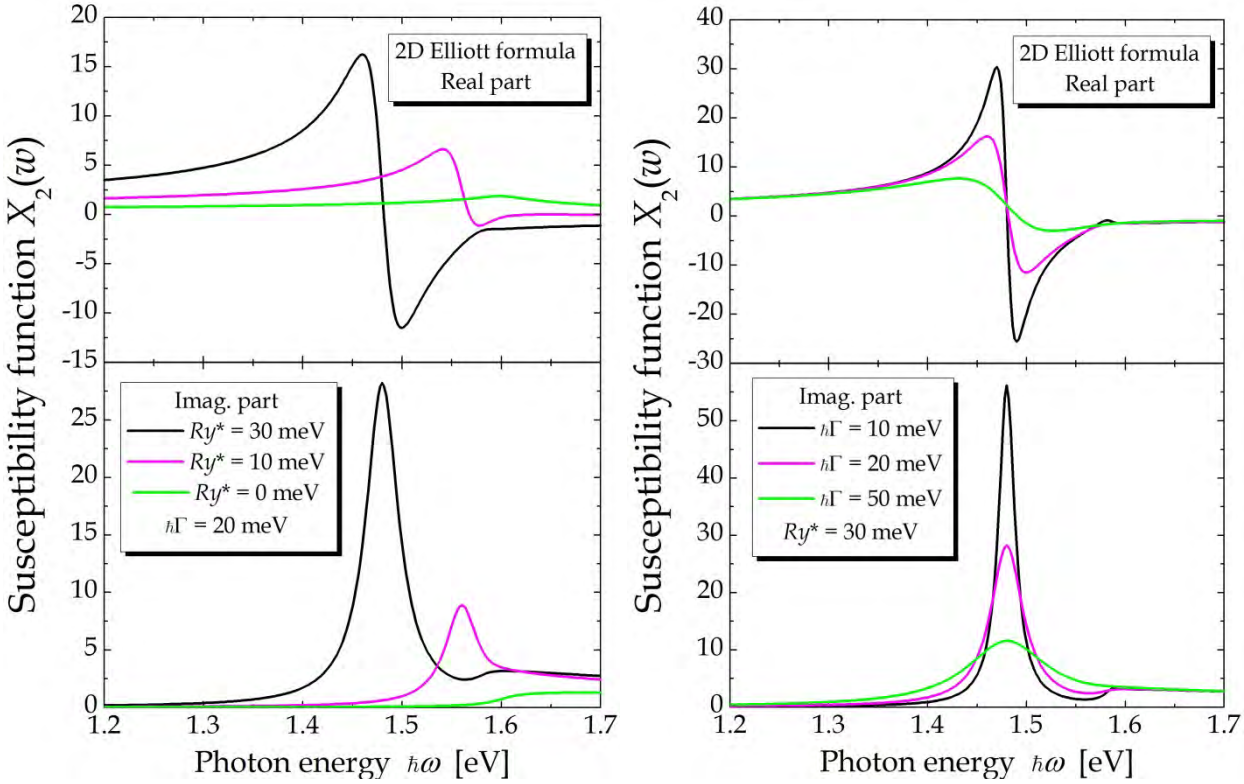


Figure 19.2. The excitonic susceptibility function for a semiconductor quantum well. In the plots, the effective band gap is taken as 1.6 eV.

Exercise: Variational treatment of quantum well excitons

In this exercise, we will return to the quantum well Hamiltonian Eq.(19.2) and look for more accurate solutions. Primarily, we will not apply the simple approximation Eq.(19.3). However, we will still look for solutions of the form $\Psi_{exc}(\vec{r}_e,\vec{r}_h)=\varphi_e(z_e)\varphi_h(z_h)\Psi(\vec{r})$ with φ_e an eigenstate of $\hat{h}_e$ and similarly for the hole. Hence, the improvement lies in finding a better estimate for $\Psi(\vec{r})$ and for this purpose we will use the variational method.

a) Show using Eq.(19.2) that the expectation value for the energy is

$$E=\tilde{E}_g+\left\langle\Psi(\vec{r})\right|-\frac{d^2}{dr^2}-\frac{1}{r}\frac{d}{dr}-\frac{1}{r^2}\frac{d^2}{d\theta^2}+V(r)\left|\Psi(\vec{r})\right\rangle,$$

with an effective potential

$$V(r) = -2\iint \frac{\left|\varphi_e(z_e)\varphi_h(z_h)\right|^2}{\sqrt{r^2 + (z_e - z_h)^2}} dz_e dz_h ,$$

As a simple example, we consider a quantum well of width d with infinite barriers. This means that the eigenstates for the z-motion are $\varphi(z) = \sqrt{\frac{2}{d}} \sin(\frac{\pi z}{d})$.

b) Using $u = (z_e - z_h)/d$ and $v = (z_e + z_h)/d$ show that

$$V(r) = -\frac{8}{d} \int_0^1 \int_u^{2-u} \frac{\sin^2[\frac{\pi}{2}(u+v)]\sin^2[\frac{\pi}{2}(u-v)]}{\sqrt{(r/d)^2 + u^2}} dv du .$$

Evaluating the v integral, we find

$$V(r) = -\frac{2}{d} \int_0^1 \frac{h(u)}{\sqrt{(r/d)^2 + u^2}} du .$$

with $h(u) = (1-u)[2 + \cos(2\pi u)] + \frac{3}{2\pi}\sin(2\pi u)$. As a particular variational ansatz, we will try the form $\Psi(\vec{r}) = \sqrt{\frac{2}{\pi}}\alpha e^{-\alpha r}$.

c) Show that the ansatz is normalized and that the kinetic energy is α^2. Use these results to demonstrate that

$$E = \tilde{E}_g + \alpha^2 + \int_0^1 h(u) W(u)\, du ,$$

where $W(u) = -\frac{8\alpha^2}{d} \int_0^\infty \frac{e^{-2\alpha r} r}{\sqrt{(r/d)^2 + u^2}} dr$.

In can be shown that $W(u) = 4\pi u \alpha^2 d \left\{ \frac{2}{\pi} + Y_1(2u\alpha d) - \mathrm{H}_1(2u\alpha d) \right\}$, where Y_1 and H_1 are Bessel and Struve functions, respectively. If d is sufficiently small, this rather complicated expression may be approximated by the 1. order expansion $W(u) \approx -4\alpha + 8u\alpha^2 d$.

d) Show that

$$E \approx \tilde{E}_g + \alpha^2 - 4\alpha + \left(\frac{8}{3} - \frac{10}{\pi^2}\right)\alpha^2 d,$$

and that by minimizing with respect to α that the optimal energy E_0 is given by

$$E_0 \approx \tilde{E}_g - \frac{4}{1 + \left(\frac{8}{3} - \frac{10}{\pi^2}\right)d}.$$

In the plot below, the exciton binding energy is plotted versus the width of the quantum well, both in natural exciton units.

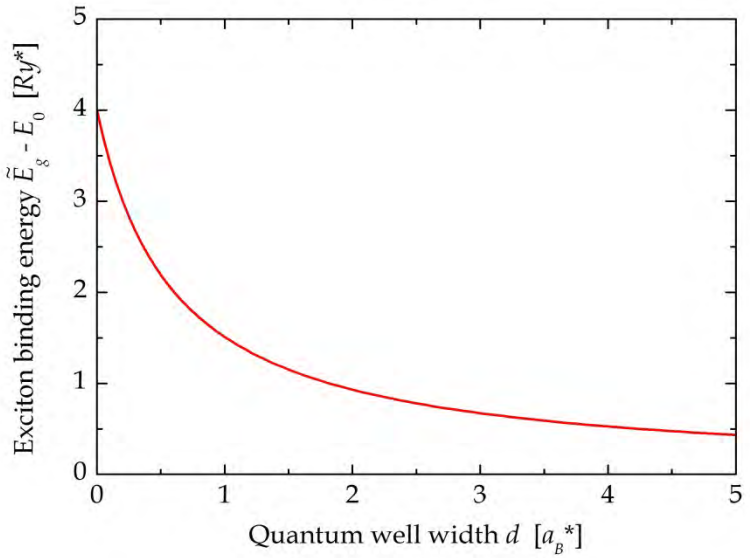


Figure 19.3. Variational calculation of the binding energy of excitons in a semiconductor quantum well.

References

[1] H. Haug and S.W. Koch *Quantum Theory of the Optical and Electronic Properties of Semiconductors* (World Scientific, Singapore, 1993).
[2] I.S. Gradsteyn and I.M. Ryzhik *Table of Integrals, Series and Products* (academic Press, San Diego, 1994).
[3] C. Tanguy, Phys. Rev. Lett. 75, 4090 (1995).
[4] C. Tanguy, Solid State Commun. 98, 65 (1996).

20. Excitons in Nanowires and Nanotubes

We have seen that confinement in quantum wells leads to enhanced excitonic effects in the optical response of semiconductors. The binding energy of the strongest bound excitons increase by a factor of 4 in the ideal 2D case. Consequently, one expects this trend to continue to 1D-structures with even stronger binding of excitons. As we will see, this is precisely what happens even to the point, where excitons completely dominate the response.

Following the previous chapter, we will limit ourselves to variational calculations of excitons in 1D-structures. These structures are assumed to be infinite along the long-axis direction and strongly confining in the two transverse dimensions. Also, the system is excited along the long-axis so that centre-of-mass momentum for this direction is to remain zero throughout. We take the z-axis as the long-axis and so the confined electron-hole pair is described by the Hamiltonian

$$\hat{H}_{eh} = \hat{h}_e(x_e, y_e) + \hat{h}_h(x_h, y_h) + E_g - \frac{d^2}{dz^2} - \frac{2}{\sqrt{(x_e - x_h)^2 + (y_e - y_h)^2 + z^2}}. \quad (20.1)$$

Here, $\hat{h}_e(x_e, y_e) = -m_e^{-1}(d^2/dx_e^2 + d^2/dy_e^2) + V_e(x_e, y_e)$ is the Hamiltonian for the transverse motion of the electron confined by the potential V_e and $\hat{h}_h$ is the analogous term for the hole. In this expression, the effective electron and hole masses should be taken in units of the reduced electron-hole pair mass, which is taken as the unit of mass. In a purely variational treatment, we attempt to describe the exciton state by the variational ansatz $\Psi_{exc}(\vec{r}_e, \vec{r}_h) = \varphi_e(x_e, y_e)\varphi_h(x_h, y_h)\Psi(z)$. The corresponding expectation value for the energy is then

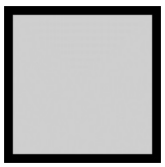
$$E = \tilde{E}_g + \left\langle \Psi(z) \middle| -\frac{d^2}{dz^2} + V(z) \middle| \Psi(z) \right\rangle, \quad (20.2)$$

with $\tilde{E}_g = E_g + E_1^e + E_1^h$ and an effective potential

$$V(z) = -2\iint \frac{|\varphi_e(x_e, y_e)\varphi_h(x_h, y_h)|^2}{\sqrt{(x_e - x_h)^2 + (y_e - y_h)^2 + z^2}} dx_e dy_e dx_h dy_h. \quad (20.3)$$

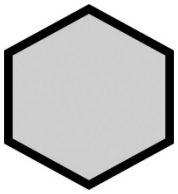
At this point, it is instructive to consider some specific examples of transverse confinement:

A. Rectangular nanowire with transverse dimensions $d\times d$ (Ref. [1]):



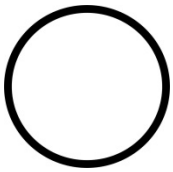
$$\varphi(x,y)=\frac{2}{d}\sin\left(\frac{\pi x}{d}\right)\sin\left(\frac{\pi y}{d}\right)$$

B. Hexagonal nanowire with diameter d (Ref. [2]):



$$\varphi(x,y)\approx\frac{8\pi}{3^{1/4}(16+5\pi^2)^{1/2}d}\cos\left(\frac{\pi x}{d}\right)\cos\left(\frac{\pi(x-\sqrt{3}y)}{2d}\right)\cos\left(\frac{\pi(x+\sqrt{3}y)}{2d}\right)$$

C. Circular nanotube with radius r (Ref. [3]):



$$|\varphi(x,y)|^2=\frac{1}{2\pi r}\delta(\sqrt{x^2+y^2}-r)$$

We note that in the nanotube model, the wave function is completely localized to the cylinder wall and rather than giving the wave function itself, we give the normalized square. In cases A and B, the effective potential can only be computed numerically. However, for the nanotube model an analytic result can be found. To this end, we introduce polar transverse coordinates $x=\rho\cos\theta$ and $y=\rho\sin\theta$. Due to the complete localization on the cylinder wall, we always have $\rho=r$. In this manner, $(x_e-x_h)^2+(y_e-y_h)^2=4r^2\sin^2((\theta_e-\theta_h)/2)$. Hence, in polar coordinates,

$$\begin{aligned}V(z)&=-\frac{2}{(2\pi r)^2}\int_0^{2\pi}\int_0^{\infty}\int_0^{2\pi}\int_0^{\infty}\frac{\delta(\rho_e-r)\delta(\rho_h-r)}{\sqrt{4r^2\sin^2((\theta_e-\theta_h)/2)+z^2}}\rho_e\rho_h d\rho_e d\theta_e d\rho_h d\theta_h\\&=-\frac{1}{\pi}\int_0^{2\pi}\frac{1}{\sqrt{4r^2\sin^2(\theta/2)+z^2}}d\theta\\&=-\frac{4}{\pi|z|}K\left(-\frac{4r^2}{z^2}\right).\end{aligned}\qquad(20.4)$$

In the last line, K is a so-called complete elliptic integral of the 1st kind defined by

$$K(x)\equiv\int_0^{\pi/2}\frac{1}{\sqrt{1-x\sin^2\theta}}d\theta.$$

From this definition, it is seen that $K(0)=\pi/2$ and the potential approaches the bare 1D Coulomb potential $V(z)=-2/|z|$ in the limit $r\to 0$. From Eq.(20.3) it is clear that this must always be the limit of a 1D effective potential whenever the confining potential becomes sufficiently narrow that $x_e\approx x_h$ and $y_e\approx y_h$ because of the confinement.

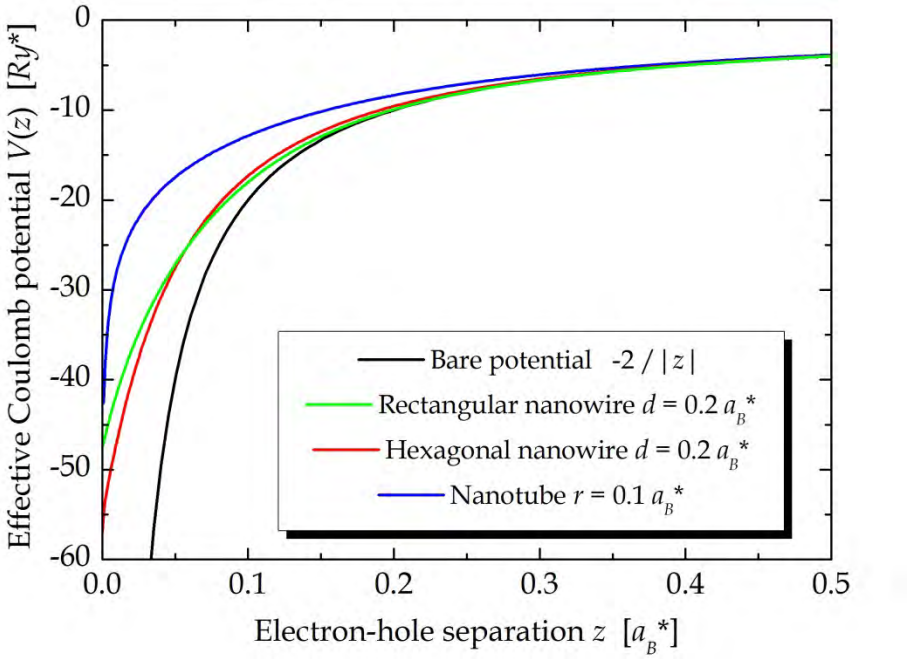


Figure 20.1. Effective Coulomb potential for three different 1D confinements compared to the bare potential.

The actual effective potentials for the three models listed above are illustrated in Fig. 20.1. When compared to the bare Coulomb potential, it is clear that the behaviour as $z \to 0$ is must less singular. Hence, for models A and B, the singularity is completely removed and for C, the singularity is now logarithmic instead of $\sim 1/|z|$. However, in all cases the bare potential $-2/|z|$ is found as a limit when the diameter of the nanowire or –tube becomes very small. It might then be thought that a viable and simple model for 1D excitons would result from using the pure 1D potential $V(z) = -2/|z|$ in Eq.(20.2) similarly to the 2D quantum well case. To look at the properties of this simple model we need to start with a "regularized" potential, however. For this purpose, we will take $V(z) = -2/(|z|+c)$ as our potential, where c is a positive constant, which should eventually be taken to zero. This "Loudon" model was originally analyzed by R. Loudon [4]. The form is mathematically simpler than any of the alternatives above. As our variational ansatz we will, as usual, try the exponential $\Psi(z) = \sqrt{\alpha} e^{-\alpha|z|}$. Differentiating twice leads to $d^2\Psi(z)/dz^2 = \alpha^2\Psi(z) - 2\alpha\Psi(0)\delta(z)$. Thus, the expectation value Eq.(20.2) is

$$E = \tilde{E}_g + \alpha^2 - 4\alpha \int_0^\infty \frac{e^{-2\alpha z}}{z+c} dz .$$

Evaluating this integral leads to yet another complicated function: the exponential integral Ei:

$$E = \tilde{E}_g + \alpha^2 + 4\alpha e^{2\alpha c} \mathrm{Ei}(-2\alpha c).$$

However, as c should go to zero we can use an expansion based on partial integration valid for small c

$$\begin{aligned}
E &= \tilde{E}_g + \alpha^2 - 4\alpha \int_0^\infty \frac{e^{-2\alpha z}}{z+c} dz \\
&= \tilde{E}_g + \alpha^2 - 4\alpha \left[e^{-2\alpha z} \ln(z+c) \right]_0^\infty - 8\alpha^2 \int_0^\infty e^{-2\alpha z} \ln(z+c) dz \\
&= \tilde{E}_g + \alpha^2 + 4\alpha \ln(c) - 8\alpha^2 \int_0^\infty e^{-2\alpha z} \ln(z+c) dz \qquad (20.5) \\
&\approx \tilde{E}_g + \alpha^2 + 4\alpha \ln(c) - 8\alpha^2 \int_0^\infty e^{-2\alpha z} \ln(z) dz \\
&= \tilde{E}_g + \alpha^2 + 4\alpha(\gamma + \ln(2\alpha c)),
\end{aligned}$$

where the integral $\int_0^\infty e^{-\beta z} \ln(z) dz = -(\gamma + \ln(\beta))/\beta$ with $\gamma = 0.577\ldots$ as Euler's constant has been applied. Differentiating, one finds $0 = 2\alpha + 4(\gamma + \ln(2\alpha c)) + 4$. With a few manipulations this condition can be reformulated as

$$\frac{\alpha}{2} \exp\left(\frac{\alpha}{2}\right) = \frac{1}{4ce^{1+\gamma}}.$$

The solution to this equation is given by the "product logarithm" (pl), c.f. Chapter 13. Thus, the solution for α is

$$\alpha = 2\,\mathrm{pl}\left(\frac{1}{4ce^{1+\gamma}}\right) \approx 2\ln\left(\frac{1}{4ce^{1+\gamma}}\right) - 2\ln\left(\ln\left(\frac{1}{4ce^{1+\gamma}}\right)\right),$$

where the second expression is the expansion for low c. Accordingly, the energy is

$$E = \tilde{E}_g - 8\,\mathrm{pl}\left(\frac{1}{4ce^{1+\gamma}}\right) - 4\,\mathrm{pl}^2\left(\frac{1}{4ce^{1+\gamma}}\right).$$

As the plot of the product logarithm in Chapter 13 shows, it is a monotonically increasing function that diverges logarithmically as the argument increases. The extremely important conclusion is this: as c goes to zero, we find $\alpha \to \infty$ and $E \to -\infty$. Hence, the wave function becomes completely localized to the point $z = 0$. This is *not* an artefact of the variational approach because the variational estimate for the energy is always higher than the true value. We therefore conclude that the 1D Coulomb model is "pathological" in that the ground state collapses and the ground state energy diverges. In Fig. 20.2, we illustrate the behaviour of the energy as c becomes smaller.

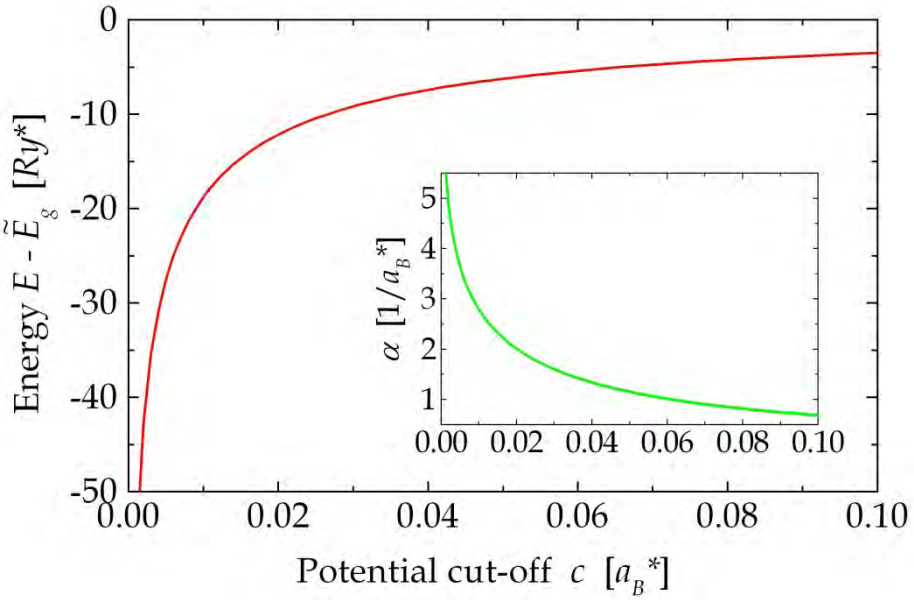


Figure 20.2. Binding energy and variational parameter (inset) in the Loudon model as a function of the potential cut-off.

20.1 1*s* Excitons in Carbon Nanotubes

Even though the pure 1D Coulomb model is clearly unphysical, it is still correct that the true potential for all realistic models approaches this strange situation as the confinement becomes stronger. Hence, even if actual exciton binding energies obviously do not diverge, they can still grow extremely large compared to bulk values. As an example of this, we now consider the true nanotube potential given by Eq.(20.4). With the same exponential ansatz as above, we find the energy expectation value

$$E = \tilde{E}_g + \alpha^2 + 2\pi\alpha J_0(2\alpha r) Y_0(2\alpha r) - \frac{32\alpha^2 r}{\pi} {}_2F_3\left[1,1;\tfrac{3}{2},\tfrac{3}{2},\tfrac{3}{2};-4\alpha^2 r^2\right],$$

where J_0 and Y_0 are Bessel functions of first and second kind, respectively, and ${}_2F_3$ is a generalized hypergeometric function [5]. The similarity with the Loudon model above becomes apparent if we again use partial integration to approximate to lowest order in r. Using the definition of the elliptic integral, we have

$$E = \tilde{E}_g + \alpha^2 - \frac{8\alpha}{\pi} \int_0^{\pi/2} \int_0^{\infty} \frac{e^{-2\alpha z}}{\sqrt{z^2 + 4r^2 \sin^2\theta}} dz d\theta \,.$$

The indefinite integral of $1/\sqrt{z^2+x^2}$ is $\ln(z+\sqrt{z^2+x^2})$ and it follows that

$$E = \tilde{E}_g + \alpha^2 - \frac{8\alpha}{\pi}\int_0^{\pi/2}\left\{\left[e^{-2\alpha z}\ln(z+\sqrt{z^2+4r^2\sin^2\theta})\right]_0^{\infty} + 2\alpha\int_0^{\infty} e^{-2\alpha z}\ln(z+\sqrt{z^2+4r^2\sin^2\theta})dz\right\}d\theta$$

$$= \tilde{E}_g + \alpha^2 - \frac{8\alpha}{\pi}\int_0^{\pi/2}\left\{-\ln(2r\sin\theta) + 2\alpha\int_0^{\infty} e^{-2\alpha z}\ln(z+\sqrt{z^2+4r^2\sin^2\theta})dz\right\}d\theta$$

$$\approx \tilde{E}_g + \alpha^2 + 4\alpha\ln(r) - \frac{16\alpha^2}{\pi}\int_0^{\pi/2}\int_0^{\infty} e^{-2\alpha z}\ln(2z)dzd\theta$$

$$= \tilde{E}_g + \alpha^2 + 4\alpha(\gamma + \ln(\alpha r)).$$

Comparing to Eq.(20.5), we see that r takes the place of $2c$. We therefore expect to find precisely the same behavior as above when r goes to zero. It should be noted that this similarity is obtained even though the nanotube potential is actually (logarithmically) divergent at the origin whereas the Loudon model potential is finite. The exciton binding energy for the nanotube model is illustrated in Fig. 20.3.

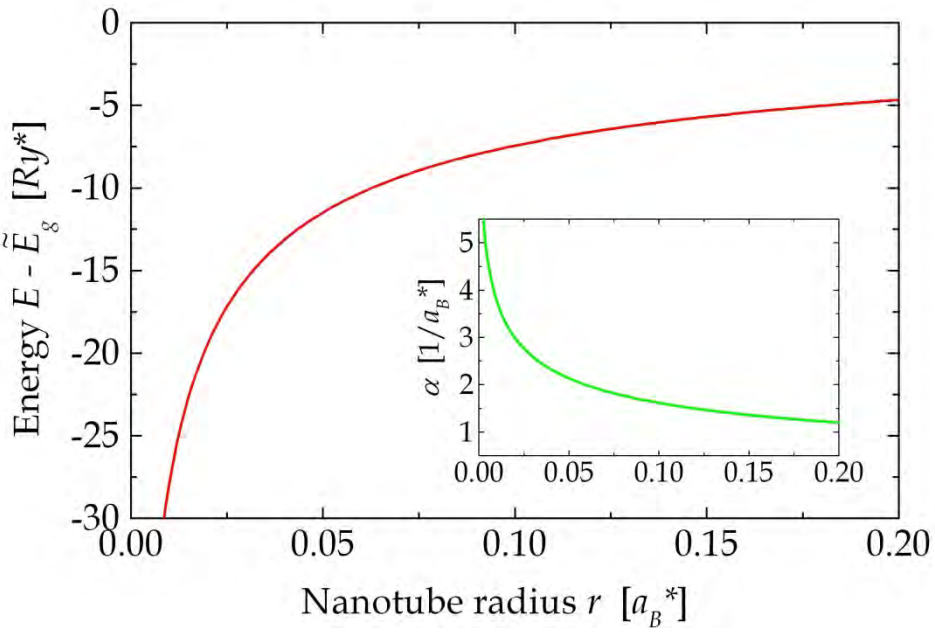


Figure 20.3. Binding energy and variational parameter (inset) of excitons in carbon nanotubes as a function of nanotube radius, all in natural exciton units.

From Fig. 20.3 it is apparent that the exciton binding energy may become very large if the nanotube radius is sufficiently small. The question is then: what are the actual values of a_B^* and Ry^*? We recall that $a_B^* = 0.529\text{Å}\cdot\varepsilon m/m_{eh}$ and $Ry^* = 13.6\text{eV}\cdot m_{eh}/m\varepsilon^2$. Hence, to answer this question we need the reduced effective electron-hole pair mass and a value of the dielectric constant. The latter is relatively straight-forward since most experiments are performed in liquid suspensions and a reasonable value describing the screening in this case is $\varepsilon = 3.5$ [3]. To compute m_{eh} we need to consider the band structure. For nanotubes excited along the long-axis the

allowed transitions are between bands symmetrically positioned above and below the Fermi level. In a simple nearest-neighbor tight-binding model of graphene, the transition energy is given by Eq.(17.2)

$$E_{cv}(\vec{k}) = 2\gamma\sqrt{1+4\cos^2\left(\frac{k_y a}{2}\right)+4\cos\left(\frac{k_y a}{2}\right)\cos\left(\frac{\sqrt{3}k_x a}{2}\right)}, \qquad (20.6)$$

where $\gamma \approx 3\,\text{eV}$ is the hopping integral and a = 2.46 Å is the lattice constant. Using zone-folding, this also applies to nanotubes as discussed in App. 1. The nanotube is characterized by the chiral indices (n,m) and in terms of these, the components of the k-vector are

$$k_x = k\frac{\sqrt{3}(n+m)}{2L} - q\frac{n-m}{2L}$$
$$k_y = k\frac{n-m}{2L} + q\frac{\sqrt{3}(n+m)}{2L}.$$

Here, k is the continuous long-axis component of the k-vector and q is the quantized short-axis component given by $q = p/r$, where p is an integer and r is the radius. Also, $L = \sqrt{n^2+m^2+nm}$ is the radius in units of $a/2\pi$, i.e. $2\pi r = aL$. An important point about the energy dispersion Eq.(20.6) is that $E_{cv}(\vec{K}) = 0$ with $\vec{K} = (\sqrt{3},1)\cdot 2\pi/3a$. Thus, the band gap is found at the allowed k-point closest to $\vec{K}$. To simplify the analysis, we expand the dispersion in the vicinity of $\vec{K}$ and find

$$E_{cv}(\vec{k}) \approx \sqrt{3}a\gamma\,|\vec{k}-\vec{K}\,|.$$

We introduce $\hat{l}$ and $\hat{s}$ as unit vectors for the long-axis and short-axis, respectively. If we express $\vec{K}$ and $\vec{k}$ in terms of the projections along these directions we find $\vec{k} = k\hat{l} + q\hat{s}$ and similarly $\vec{K} = K\hat{l} + Q\hat{s}$, where it can be shown that $Q = (2n+m)/3r$. Hence, setting q = Q, leads to the condition $p = (2n+m)/3$. This condition can *only* be fulfilled if $(2n+m)/3$ is an integer in which case the nanotube is a metal. If not, the minimum difference becomes $|q-Q|_{\min} = 1/3r$. In this case, the minimum transition energy, i.e. the band gap $\tilde{E}_g$, then becomes

$$\tilde{E}_g = \sqrt{3}a\gamma\,|q-Q|_{\min} = \frac{a\gamma}{\sqrt{3}r}.$$

To find the effective mass, we consider an approximately parabolic dispersion

$$E_{cv}(\vec{k}) \approx \tilde{E}_g + \frac{\hbar^2}{2m_{eh}}(k-K)^2$$

so that, to lowest order, the square becomes $E_{cv}^2(\vec{k}) \approx \tilde{E}_g^2 + \hbar^2(k-K)^2 \tilde{E}_g / m_{eh}$. On the other hand, the square of the energy dispersion for the nanotube near the minimum is

$$\begin{aligned} E_{cv}^2(\vec{k}) &\approx 3a^2\gamma^2 \left\{(k-K)^2 + |q-Q|_{\min}^2\right\} \\ &= \tilde{E}_g^2 + 3a^2\gamma^2(k-K)^2. \end{aligned}$$

Hence, a comparison demonstrates that $m_{eh} = \hbar^2 \tilde{E}_g / 3a^2\gamma^2$. Consequently the effective mass is proportional to the band gap. Plugging in the numbers it turns out that $\tilde{E}_g = 4.2 \text{ eVÅ}/r$ and $m_{eh}/m = 0.19\text{Å}/r$. An extremely important point is then that the effective Bohr radius $a_B^* = 0.529\text{Å} \cdot \varepsilon m / m_{eh}$ becomes a linear function of r given by $a_B^* = 9.7r$. As a consequence, the radius r measured in units of a_B^* is *always* roughly 0.1!. At this value, the exciton binding energy as computed above and illustrated in Fig. 20.3 is around -7.44 Ry^*. A more accurate calculation [3] finds a binding energy of approximately -8.1 Ry^*. It is noted that these values are substantially higher than the maximum value -4 Ry^* found for 2D structures. In analogy with the effective Bohr radius, the effective Rydberg also depends on r and inserting values we find $Ry^* = 0.21 \text{ eVÅ}/r$. It therefore follows that the ratio between exciton binding energy and band gap is a near constant of around -40%. This is obviously a huge value, which will completely rearrange the optical response.

As an example, we consider the (7,6) nanotube with a radius of r = 4.4 Å. For this structure, the exciton binding energy is then -0.39 eV. Unfortunately, there is no simple way to sum all the contributions to the optical response analytically. Instead, a numerical calculation of bound and unbound excitons can be made using a finite basis set [6]. Summing the different terms leads to the spectrum shown in Fig. 20.4, where the independent-particle result is included for comparison. The very large red-shift of the resonance corresponds to the value of the exciton binding energy. Also, it is noticed that the peak is now much more symmetric than the inverse square-root of the independent-particle spectrum.

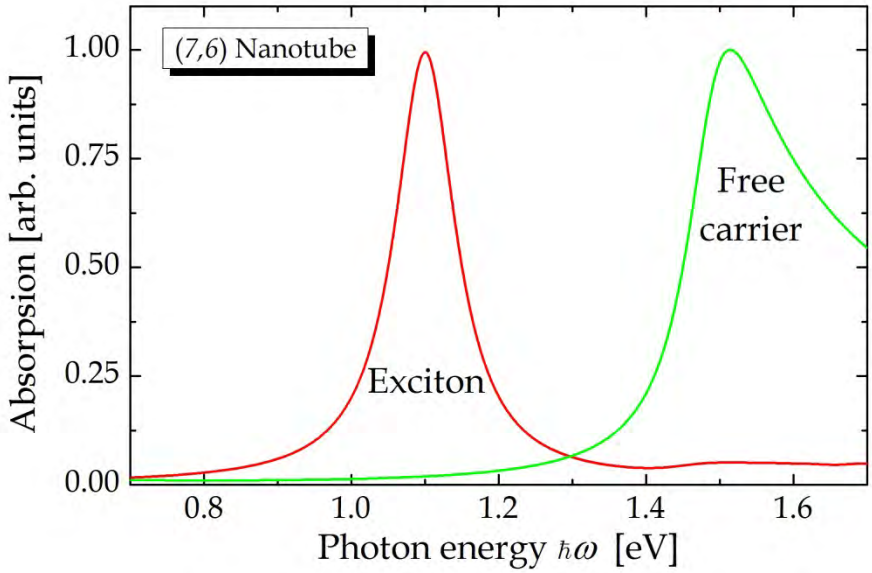


Figure 20.4. Normalized absorption spectra of a (7,6) carbon nanotube. The curves show the spectra including and neglecting excitonic effects, respectively.

Exercise: 2*p* nanotube excitons

In this exercise, we will attempt to compute the energy of the 2*p* exciton in a nanotube and for this purpose the ansatz $\Psi(z)=\sqrt{2\beta^3}ze^{-\beta|z|}$ will be used. Note that it is always orthogonal to the ground state.

a) Show that $\Psi(z)$ is normalized and that the kinetic energy is β^2.

The difficult part lies in determining an approximate expression for the Coulomb energy valid for small but finite *r*. It is given by

$$U=-\frac{8}{\pi}\int_0^{\pi/2}\int_0^{\infty}\frac{\Psi^2(z)}{\sqrt{z^2+4r^2\sin^2\theta}}dzd\theta\,.$$

As a start, we will consider the definite integral

$$W=\int_0^{\infty}\frac{f(z)}{\sqrt{z^2+x^2}}dz\,.$$

The first few indefinite integrals of the square-root are denoted $S_n(z,x)$ so that

$$S_0(z,x) \equiv \frac{1}{\sqrt{z^2+x^2}}$$
$$S_1(z,x) \equiv \int S_0(z,x)dz = \ln\left(z+\sqrt{z^2+x^2}\right)$$
$$S_2(z,x) \equiv \int S_1(z,x)dz = z\ln\left(z+\sqrt{z^2+x^2}\right)-\sqrt{z^2+x^2}$$
$$S_3(z,x) \equiv \int S_2(z,x)dz = \frac{1}{4}(2z^2-x^2)\ln\left(z+\sqrt{z^2+x^2}\right)-\frac{3}{4}z\sqrt{z^2+x^2}.$$

b) Under the assumption that $f(\infty)=0$, show by repeated use of partial integration that

$$\begin{aligned} W &= -f(0)\ln x - \int_0^\infty f'(z)S_1(z,x)dz \\ &= -f(0)\ln x - f'(0)x + \int_0^\infty f''(z)S_2(z,x)dz \\ &= -f(0)\ln x - f'(0)x + \frac{1}{4}f''(0)x^2\ln x - \int_0^\infty f'''(z)S_3(z,x)dz. \end{aligned}$$

At this stage, no approximations have been made. However, to actually calculate the integral, we will now use the small-x expansion

$$S_3(z,x) \approx \frac{1}{4}(2z^2-x^2)\ln(2z) - \frac{1}{4}(3z^2+x^2).$$

In the present example, f is the function $f(z)=\Psi^2(z)=2\beta^3z^2e^{-2\beta z}$ and so

$$f'(z)=4\beta^3z(1-\beta z)e^{-2\beta z},\ f''(z)=4\beta^3(1-4\beta z+2\beta^2z^2)e^{-2\beta z},\ f'''(z)=-8\beta^4(3-6\beta z+2\beta^2z^2)e^{-2\beta z}.$$

With this form and the approximate $S_3(z,x)$ we find

$$\int_0^\infty f'''(z)S_3(z,x)dz \approx -\frac{\beta}{2}-\frac{\beta^3x^2}{2}\{1+2\gamma+2\ln\beta\}.$$

In turn, W becomes $W \approx \frac{\beta}{2}+\frac{\beta^3x^2}{2}\{1+2\gamma+2\ln\beta x\}$ and we then have

$$\begin{aligned} U &= -\frac{8}{\pi}\int_0^{\pi/2}\left[\frac{\beta}{2}+2\beta^3r^2\sin^2\theta\{1+2\gamma+2\ln(2\beta r\sin\theta)\}\right]d\theta \\ &= -2\beta-8\beta^3r^2\{1+\gamma+\ln(\beta r)\}. \end{aligned}$$

The total energy is therefore $E_{2p} = \tilde{E}_g + \beta^2 - 2\beta - 8\beta^3 r^2 \{1 + \gamma + \ln(\beta r)\}$.

c) Show that the minimum energy is approximately

$$E_{2p} \approx \tilde{E}_g - 1 - 8r^2 \{1 + \gamma + \ln r\}.$$

This result is plotted below.

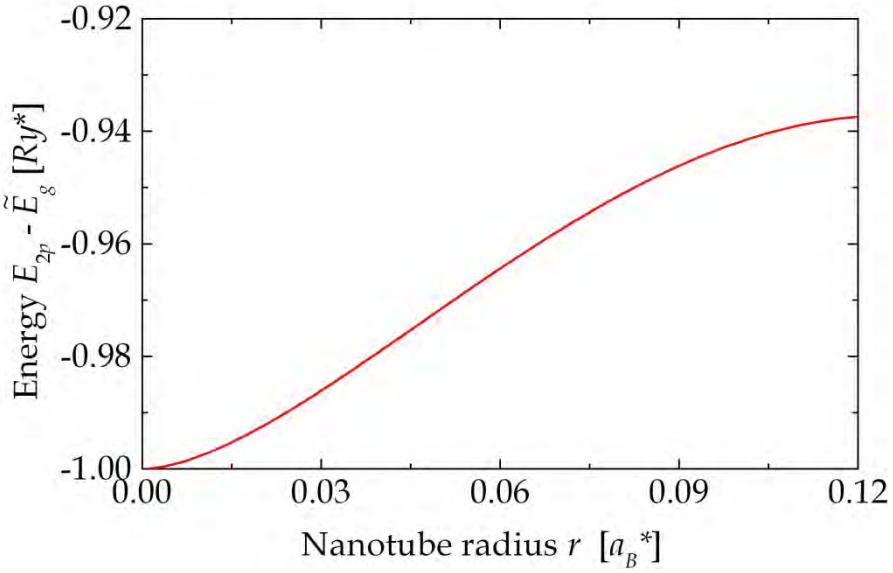


Figure 20.5. Variational 2*p* exciton binding energy.

References

[1] T.G. Pedersen, P.M. Johansen and H. C. Pedersen, Phys. Rev. B61, 10504 (2000).
[2] T.G. Pedersen, Phys. Stat. Sol. (c) 2, 4026 (2005).
[3] T.G. Pedersen, Phys. Rev. B67, 073401 (2003).
[4] R. Loudon, Am. J. Phys. 27, 649 (1959).
[5] I.S. Gradsteyn and I.M. Ryzhik *Table of Integrals, Series and Products* (Academic Press, San Diego, 1994).
[6] T.G. Pedersen, Carbon 42, 1007 (2004).

21. Electro-Optics

The term "electro-optics" refers to the fact that the optical properties of a material can be altered by the presence of an electrostatic field $\vec{\mathcal{E}}$. Basically, such a field displaces the electrons (and ions), thereby changing the eigenstates of the system. This, in turn, is reflected in the optical response. If the electric field is weak, the change can be treated as a perturbation. Hence, the dielectric constant is perturbed, i.e. $\varepsilon(\mathcal{E}) = \varepsilon(0) + \alpha\mathcal{E} + \beta\mathcal{E}^2 + \cdots$. Here, α is the linear (or Pockels) and β is the quadratic (or Kerr) electro-optic correction. In certain simple cases, however, it is possible to include the electrostatic field to all orders. This "non-perturbative" theory is applicable even when the field becomes comparable to the inter-atomic fields of the material itself. In this chapter, we will investigate this situation for semiconductors in the effective-mass approximation. If the *k*-dependence of momentum matrix elements can be ignored we will show that the model is sufficiently simple to allow including the field to all orders. This leads to the so-called Franz-Keldysh theory for the electro-optic response. The prerequisite for such a calculation is an exact solution for the stationary eigenstates of the semiconductor in the presence of $\vec{\mathcal{E}}$. We therefore need to compute these states.

Before turning to the states themselves, however, we will discuss the approach to the response calculation. The slightly subtle point here is that we will apply the framework developed for the excitonic response even if electron-hole interactions are ignored. Recall that for Wannier excitons the imaginary part of the optical response was given by Eq. (18.7)

$$\mathrm{Im}\,\chi(\omega) = \frac{2\pi e^2 \,|\, p_{vc} \,|^2}{\varepsilon_0 m^2 \omega^2} \sum_n \left|\Psi_n(0)\right|^2 \delta(E_n - \hbar\omega)\,. \tag{21.1}$$

Here, $\Psi_n(\vec{r})$ is the relative-motion part of electron-hole pair wave-function and E_n is the associated energy measured relative to the ground state. To make sure that Eq.(21.1), in fact, agrees with the expectations even if there is no Coulomb attraction, it is instructive to consider a free electron-hole pair. Hence, we start our analysis by revisiting the simple case of free carriers in a one-dimensional semiconductor.

21.1 One-Dimensional Materials

Specifically, we choose a one-dimensional material with a width *d* and length *L* and the long-axis along the *z*-direction. Hence, the relative-motion part of electron-hole pair wave-function is simply $\Psi(z) = e^{ikz} / (d\sqrt{L})$ and the energy is $E_g + \hbar^2 k^2 / 2m_{eh}$. It is then a simple matter to show that

$$\begin{aligned}\mathrm{Im}\,\chi(\omega) &= \frac{2\pi e^2 \,|\, p_{vc}\,|^2}{\varepsilon_0 m^2\omega^2}\frac{L}{2\pi}\int_{-\infty}^{\infty}\frac{1}{d^2L}\delta\left(E_g+\frac{\hbar^2k^2}{2m_{eh}}-\hbar\omega\right)dk \\ &= \frac{e^2\,|\,p_{vc}\,|^2\sqrt{2m_{eh}}}{\varepsilon_0 m^2\hbar\omega^2 d^2}\frac{\theta\left(\hbar\omega-E_g\right)}{\sqrt{\hbar\omega-E_g}},\end{aligned} \tag{21.2}$$

in full agreement with Eq. (16.4) in the limit of vanishing broadening.

Having established the validity for free carriers, we now turn to the problem in a field oriented along the $1D$ structure [1]. All the corrections result from the effect on the electron-hole given by $e\mathcal{E}(z_e - z_h) = e\mathcal{E}z$ as the electrons and hole have negative and positive charges, respectively. We therefore should solve the "Wannier equation"

$$\left\{-\frac{\hbar^2}{2m_{eh}}\frac{d^2}{dz^2}+e\mathcal{E}z\right\}\Psi(z) = (E-E_g)\Psi(z). \tag{21.3}$$

What are the natural distance and energy units for the problem? To answer this, we write $z=\lambda\zeta$ with λ a constant and equate the prefactors of the kinetic and potential energy terms: $\hbar^2/(2m_{eh}\lambda^2) = e\mathcal{E}\lambda$ leading to $\lambda = (\hbar^2/(2m_{eh}e\mathcal{E}))^{1/3}$. Defining, moreover, $e\mathcal{E}\lambda = (\hbar^2e^2\mathcal{E}^2/2m_{eh})^{1/3} \equiv \Theta$ and $\varepsilon \equiv (E-E_g)/\Theta$ we find

$$\left\{-\frac{d^2}{d\zeta^2}+\zeta\right\}\Psi(\zeta) = \varepsilon\Psi(\zeta).$$

From Airy's differential equation, which we encountered in Chapter 5, it is then clear that $\Psi(\zeta) = N\mathrm{Ai}(\zeta-\varepsilon)$.

In Fig. 21.1 below, we have plotted the Airy function together with its asymptotics as the argument goes to plus or minus infinity. As the argument increases, the function decays exponentially as expected for the range, in which the energy is less that the potential, i.e. when $\zeta > \varepsilon$. In contrast, the function oscillates as a damped sinusoidal at large negative arguments. This last fact allows us to provide a quantization condition for the energies and a normalization condition for the wave function. What we imagine is a "hard wall" erected at $\zeta = -L$ with L a large positive distance that is eventually taken to infinity. Using the sine-asymptotic that is valid near the wall it is then found that the allowed energies are

$$\varepsilon_n = \left[\frac{3\pi}{2}\left(n-\tfrac{1}{4}\right)\right]^{2/3} - L, \quad n=1,2,3,... \tag{21.4}$$

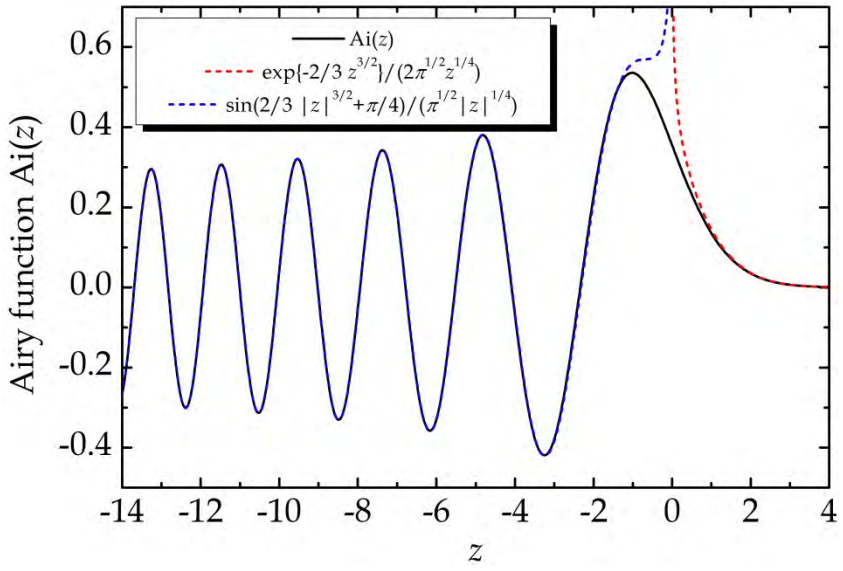


Figure 21.1. Airy function and its asymptotic expressions for large positive and negative arguments.

To normalize the wave function we should obviously integrate between $z=-L$ and $z=\infty$. To evaluate the integral, a useful property of the Airy function is that

$$\int \mathrm{Ai}^2(z)dz = z\mathrm{Ai}^2(z) - \mathrm{Ai}'^2(z), \tag{21.5}$$

where Ai' is the derivative of the Airy function. This indefinite integral vanishes at $z=\infty$ because of the exponential decay and at $z=-L$ we are allowed to use sine-asymptotic. In this manner, it follows that the normalized wave function is

$$\Psi_n(z) = \left(\left(\frac{2\pi^2}{3(n-\frac{1}{4})}\right)^{1/3} \frac{e\mathcal{E}}{\Theta d^2}\right)^{1/2} \mathrm{Ai}\left(\frac{e\mathcal{E}z}{\Theta} - \varepsilon_n\right). \tag{21.6}$$

At this stage we can then write the optical susceptibility as

$$\mathrm{Im}\,\chi(\omega) = \frac{2\pi e^2 |p_{vc}|^2 e\mathcal{E}}{\varepsilon_0 m^2\omega^2\Theta d^2} \sum_n \left(\frac{2\pi^2}{3(n-\frac{1}{4})}\right)^{1/3} \mathrm{Ai}^2(-\varepsilon_n)\delta(E_g + \Theta\varepsilon_n - \hbar\omega).$$

Because of the delta-function and using $e\mathcal{E}/\Theta^{3/2} = \sqrt{2m_{eh}}/\hbar$ we immediately see that

$$\mathrm{Im}\,\chi(\omega) = \frac{2\pi e^2 |p_{vc}|^2 \sqrt{2m_{eh}}}{\varepsilon_0 m^2\hbar\omega^2\Theta^{1/2} d^2} \mathrm{Ai}^2\left(\frac{E_g - \hbar\omega}{\Theta}\right) \sum_n \left(\frac{2\pi^2}{3(n-\frac{1}{4})}\right)^{1/3} \delta\left(\varepsilon_n - \frac{\hbar\omega - E_g}{\Theta}\right).$$

The sum can be evaluated by noting that as L becomes increasingly large, the important values of n (where the argument of the delta-function vanishes) also

increase. Now, suppose we fix our attention on some large n that contributes to the response for a particular frequency ω. Near n we find

$$\varepsilon_{n+\Delta n} \approx \varepsilon_n + \frac{d\varepsilon_n}{dn}\Delta n, \quad \frac{d\varepsilon_n}{dn} = \left(\frac{2\pi^2}{3}\right)^{1/3}\left(n-\tfrac{1}{4}\right)^{-1/3}$$

The point is that the derivative *decreases* with n and higher derivatives decrease even faster. Hence, the expansion can safely be terminated at first order. The allowed values of n are spaced by $\Delta n = 1$ but the allowed values of ε_n apparently lie closer and closer as n increases. This eventually means that we can convert the sum over n to an integral over ε_n as follows

$$\begin{aligned}\operatorname{Im}\chi_{1D}(\omega) &= \frac{2\pi e^2 \mid p_{vc} \mid^2 \sqrt{2m_{eh}}}{\varepsilon_0 m^2 \hbar\omega^2 \Theta^{1/2} d^2}\operatorname{Ai}^2\left(\frac{E_g-\hbar\omega}{\Theta}\right)\int\left(\frac{2\pi^2}{3(n-\frac{1}{4})}\right)^{1/3}\left(\frac{d\varepsilon}{dn}\right)^{-1}\delta\left(\varepsilon-\frac{\hbar\omega-E_g}{\Theta}\right)d\varepsilon \\ &= \frac{2\pi e^2 \mid p_{vc} \mid^2 \sqrt{2m_{eh}}}{\varepsilon_0 m^2 \hbar\omega^2 \Theta^{1/2} d^2}\operatorname{Ai}^2\left(\frac{E_g-\hbar\omega}{\Theta}\right).\end{aligned} \tag{21.7}$$

This important result is our analytical expression for the Franz-Keldysh effect in one-dimensional semiconductors. To verify that it conforms to the usual field-free result Eq.(21.2) we again use the asymptotics of the Airy function. As Θ goes to zero the Airy function dies exponentially if $\hbar\omega < E_g$. On the other hand, if $\hbar\omega > E_g$ the function approaches an increasing rapidly oscillating sine function. If we add a little broadening, the square of this wildly oscillating sine is approximately ½ and so we find

$$\begin{aligned}\operatorname{Im}\chi(\omega) &\approx \frac{2\pi e^2 \mid p_{vc} \mid^2 \sqrt{2m_{eh}}}{\varepsilon_0 m^2 \hbar\omega^2 \Theta^{1/2} d^2}\frac{1}{2\pi}\left(\frac{\hbar\omega-E_g}{\Theta}\right)^{-1/2}\theta(\hbar\omega-E_g) \\ &= \frac{e^2 \mid p_{vc} \mid^2 \sqrt{2m_{eh}}}{\varepsilon_0 m^2 \hbar\omega^2 d^2}\frac{\theta\left(\hbar\omega-E_g\right)}{\sqrt{\hbar\omega-E_g}}.\end{aligned}$$

This obviously agrees as it should.

21.2 Two- and Three-Dimensional Materials

Adding additional dimensions to the problem is, in fact, completely straight-forward. Since the dimensions are decoupled in the effective-mass approximation the wave function is a simple product and the energy is a sum of contributions from the different dimensions. For a two-dimensional semiconductor we then need to add the transverse kinetic energy to the band gap $E_g \to E_g + \hbar^2 k^2 / 2m_{eh}$, where k is the wave vector component perpendicular to the electrostatic field. If the added dimension is

in the x direction (and has an extent L_x) and we now take d to designate the thickness of the 2D slab, the Wannier wave function becomes

$$\Psi_n(x,z)=\left(\left(\frac{2\pi^2}{3(n-\frac{1}{4})}\right)^{1/3}\frac{e\mathcal{E}}{\Theta d}\right)^{1/2}\frac{e^{ikx}}{\sqrt{L_x}}\mathrm{Ai}\left(\frac{e\mathcal{E}z}{\Theta}-\varepsilon_n\right). \tag{21.8}$$

Thus, the upshot is a response given by

$$\mathrm{Im}\,\chi(\omega)=\frac{2\pi e^2\,|\,p_{vc}\,|^2\,e\mathcal{E}}{\varepsilon_0 m^2\omega^2\Theta d L_x}\sum_{n,k}\left(\frac{2\pi^2}{3(n-\frac{1}{4})}\right)^{1/3}\mathrm{Ai}^2(-\varepsilon_n)\delta\left(E_g+\frac{\hbar^2k^2}{2m_{eh}}+\Theta\varepsilon_n-\hbar\omega\right).$$

Converting both sums to integrals, we then find

$$\begin{aligned}\mathrm{Im}\,\chi(\omega)&=\frac{e^2\,|\,p_{vc}\,|^2\,e\mathcal{E}}{\varepsilon_0 m^2\omega^2\Theta d}\iint\mathrm{Ai}^2(-\varepsilon)\delta\left(E_g+\frac{\hbar^2k^2}{2m_{eh}}+\Theta\varepsilon-\hbar\omega\right)dkd\varepsilon\\&=\frac{e^2\,|\,p_{vc}\,|^2\,e\mathcal{E}}{\varepsilon_0 m^2\omega^2\Theta^2 d}\int_{-\infty}^{\infty}\mathrm{Ai}^2\left(\frac{E_g+\frac{\hbar^2k^2}{2m_{eh}}-\hbar\omega}{\Theta}\right)dk\\&=\frac{2e^2\,|\,p_{vc}\,|^2\,m_{eh}}{\varepsilon_0 m^2\hbar^2\omega^2\Theta^{1/2}d}\int_0^{\infty}\mathrm{Ai}^2\left(\frac{E_g+t-\hbar\omega}{\Theta}\right)\frac{dt}{\sqrt{t}}.\end{aligned}$$

This integral cannot be evaluated in closed form except in terms of hypergeometric functions. However, introducing the indefinite integral $\mathrm{Ai}_1(z)\equiv\int_z^{\infty}\mathrm{Ai}(t)dt$ it can be shown that [2]

$$\mathrm{Im}\,\chi_{2D}(\omega)=\frac{e^2\,|\,p_{vc}\,|^2\,m_{eh}}{\varepsilon_0 m^2\hbar^2\omega^2 d}\mathrm{Ai}_1\left(\frac{2^{2/3}(E_g-\hbar\omega)}{\Theta}\right). \tag{21.9}$$

The function $\mathrm{Ai}_1(z)$ approaches the step function $\theta(-z)$ at large negative arguments and so the entire expression is in agreement with our previous result Eq.(16.3). The three-dimensional case is even simpler [2]. Here, we take k to be the magnitude of the transverse wave vector and by simple generalization of the result above we see that

$$\mathrm{Im}\,\chi(\omega)=\frac{e^2\,|\,p_{vc}\,|^2\,e\mathcal{E}}{\varepsilon_0 m^2\omega^2\Theta^2}\int_0^{\infty}\mathrm{Ai}^2\left(\frac{E_g+\frac{\hbar^2k^2}{2m_{eh}}-\hbar\omega}{\Theta}\right)kdk\,.$$

Due to the additional k under the integral we immediately see that using Eq.(21.5)

$$\begin{aligned}\operatorname{Im}\chi_{3D}(\omega) &= \frac{e^2\,|\,p_{vc}\,|^2\,e\mathcal{E}m_{eh}}{\varepsilon_0 m^2\hbar^2\omega^2\Theta^2}\int_0^\infty \mathrm{Ai}^2\left(\frac{E_g+t-\hbar\omega}{\Theta}\right)dt \\ &= \frac{e^2\,|\,p_{vc}\,|^2\,(2m_{eh})^{3/2}\Theta^{1/2}}{2\varepsilon_0 m^2\hbar^3\omega^2}\left\{\mathrm{Ai}'^2\left(\frac{E_g-\hbar\omega}{\Theta}\right)-\left(\frac{E_g-\hbar\omega}{\Theta}\right)\mathrm{Ai}^2\left(\frac{E_g-\hbar\omega}{\Theta}\right)\right\}.\end{aligned} \tag{21.10}$$

The complicated Airy expression has the limit

$$\lim_{\Theta\to 0}\Theta^{1/2}\left\{\mathrm{Ai}'^2\left(\frac{E_g-\hbar\omega}{\Theta}\right)-\left(\frac{E_g-\hbar\omega}{\Theta}\right)\mathrm{Ai}^2\left(\frac{E_g-\hbar\omega}{\Theta}\right)\right\}=\frac{1}{\pi}\theta\left(\hbar\omega-E_g\right)\sqrt{\hbar\omega-E_g}.$$

This means that also Eq.(21.10) is in agreement with the field-free expression. To illustrate the effect of dimensionality on the response it is instructive to consider the electro-optic susceptibility functions with $z=(E_g-\hbar\omega)/\Theta$

$$X_n(\omega)=\begin{cases}\pi\Theta^{-1/2}\mathrm{Ai}^2(z) & n=1\\ \mathrm{Ai}_1\left(2^{2/3}z\right) & n=2\\ \pi\Theta^{1/2}\left\{\mathrm{Ai}'^2(z)-z\mathrm{Ai}^2(z)\right\} & n=3.\end{cases},$$

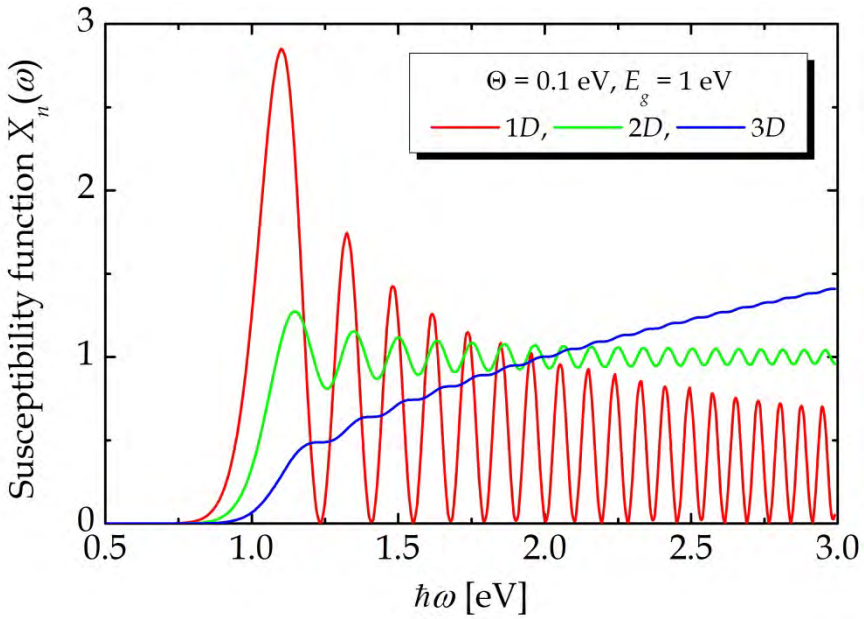


Figure 21.2. Electro-optic susceptibility functions for various dimensions.

As shown in Fig. 21.2, the modulation of the spectrum increases gradually as the dimension decreases. Hence, low-dimensional systems are most strongly affected by the field.

21.3 Beyond the Wannier Model

All of the results above relied on the Wannier approximations: effective-mass dispersion for the band structure and ignored *k*-dependence of the momentum matrix element. We now aim to improve this framework at the price of abandoning fully analytic results. First, we Fourier-transform the wave function, i.e. write

$$\Psi_{\vec{k}_\perp}(z) = \int \Psi_{\vec{k}} e^{-ik_z z} dk_z \,,$$

where $\vec{k}_\perp$ denotes the *k*-component perpendicular to the field. Next, if we Fourier-transform the Wannier equation, such as Eq.(21.3), we find

$$\left\{ E_{cv}(\vec{k}) - ie\mathcal{E}\frac{d}{dk_z} \right\} \Psi_{\vec{k}} = E\Psi_{\vec{k}}, \quad E_{cv}(\vec{k}) = E_g + \frac{\hbar^2 k^2}{2m_{eh}}.$$

This expression is really much more general than it appears. As it can be demonstrated using the approach in Chapter 18, it actually holds for *any* band structure, not only parabolic bands. Furthermore, this equation has a general solution

$$\Psi_{\vec{k}}^{(E)} = \frac{1}{2\pi\sqrt{e\mathcal{E}}} \exp\left\{ \frac{i}{e\mathcal{E}} \left(Ek_z - \int_0^{k_z} E_{cv}(\vec{k}') dk_z' \right) \right\}.$$

Here, the normalization is chosen such that $\int \Psi_{\vec{k}_\perp}^{(E)}(z) \Psi_{\vec{k}_\perp}^{(E')*}(z) dz = \delta(E - E')$. Thus, we are now able to go beyond the effective mass approximation. In fact, we can also easily include a *k*-dependent momentum matrix element if we apply the generalized response expression

$$\mathrm{Im}\,\chi(\omega) = \frac{2\pi e^2}{\varepsilon_0 m^2 \omega^2 A} \sum_{\vec{k}_\perp} \left| P_{\vec{k}_\perp}^{(\hbar\omega)} \right|^2, \quad P_{\vec{k}_\perp}^{(\hbar\omega)} = \int \Psi_{\vec{k}}^{(\hbar\omega)} p_{vc}(\vec{k}) dk_z \,. \tag{21.11}$$

Here, *A* is the area of the perpendicular dimensions. Applying this to a strictly 2*D* material such as graphene, it is convenient to introduce the sheet conductivity $\sigma = -id\varepsilon_0 \omega \chi$ so that

$$\sigma(\omega) = \frac{2\pi e^2}{m^2 \omega d} \sum_{k_\perp} \left| P_{k_\perp}^{(\hbar\omega)} \right|^2 = \frac{e^2}{m^2 \omega} \int \left| P_{k_\perp}^{(\hbar\omega)} \right|^2 dk_\perp \,.$$

The property that ensures that our results have the right behaviour in the field-free limit is that

$$\lim_{\mathcal{E}\to 0} P_{\vec{k}_\perp}^{(E)} \approx p_{vc}(\vec{k})\left|2\pi \tfrac{d}{dk_z} E_{cv}(\vec{k})\right|^{-1/2} \lim_{\mathcal{E}\to 0} \exp\left\{\frac{i}{e\mathcal{E}}\left(Ek_z - \int_0^{k_z} E_{cv}(\vec{k}')dk_z'\right)\right\},$$

where $\vec{k}$ is understood to satisfy the resonance condition $E_{cv}(\vec{k}) = E$. This result follows from the saddle-point approximation to the integral in Eq.(21.11).

Exercise: Franz-Keldysh effect in *k*-space

We consider the simplest possible case of a one-dimensional semiconductor in the effective mass approximation so that $E_{cv}(k) = E_g + \hbar^2 k^2 / 2m_{eh}$ and introduce the scaled wave vector $t \equiv k\Theta / e\mathcal{E}$

a) Show that

$$\Psi_k^{(E)} = \frac{1}{2\pi\sqrt{e\mathcal{E}}} \exp\left\{-i\left(\frac{E_g - E}{\Theta}\right)t - i\frac{t^3}{3}\right\}.$$

Using the integral representation $\mathrm{Ai}(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp\left\{-i(xt + \frac{t^3}{3})\right\} dt$ it then follows that

$$\Psi^{(E)}(z) = \frac{\sqrt{e\mathcal{E}}}{\Theta} \mathrm{Ai}\left(\frac{E_g - E + e\mathcal{E}z}{\Theta}\right).$$

b) Show that

$$\int \Psi^{(E)}(z)\Psi^{(E')*}(z)dz = \delta(E - E').$$

c) Using $P^{(\hbar\omega)} \approx p_{vc}\int \Psi_k^{(\hbar\omega)} dk = p_{vc}\Psi^{(\hbar\omega)}(z=0)$ demonstrate that

$$\mathrm{Im}\,\chi_{1D}(\omega) = \frac{2\pi e^2 |\, p_{vc}\,|^2 \sqrt{2m_{eh}}}{\varepsilon_0 m^2 \hbar\omega^2 \Theta^{1/2} d^2} \mathrm{Ai}^2\left(\frac{E_g - \hbar\omega}{\Theta}\right).$$

References

[1] T.G. Pedersen and T.B. Lynge, Phys. Rev. B65, 085201 (2002).
[2] D.E. Aspnes, Phys. Rev. 147, 554 (1966).

22. Semiconductor Lasers and LEDs

Lasers and light-emitting diodes (LEDs) are arguably among the most significant optoelectronic applications of semiconductors. These devices are central in optical communication, LED displays, laser printers and CD/DVD drives. In fact, lasers and LEDs are not fundamentally different and a laser run at a current below lasing threshold will operate as an LED. In both devices, *spontaneous emission* is present. This light is diffuse and covers a relatively broad frequency range. In a laser, however, population inversion produces *gain* inside the light emitting material. If the optical gain exceeds the optical losses at some frequency, light will be amplified as it propagates along the material. As light of this particular frequency is amplified, the amount of *stimulated emission* at the same frequency is greatly enhanced since the stimulated emission rate is proportional to the light intensity. This, in turn, stimulates emission even further and, eventually, nearly all available energy is channeled into light emitted with one specific frequency, direction and polarization mode. Thus, a prerequisite for lasing is gain and in an LED the injected current just isn't enough to produce amplification and, consequently, only spontaneous emission is obtained. In the present chapter, we will first investigate spontaneous emission and study the spectrum of the emission. Secondly, we will consider the condition for gain in bulk semiconductors and quantum wells.

The structure of lasers and LEDs is similar to pn junction diodes discussed in Chapters 9 and 10. Hence, we consider a structure comprised of two halves: to the right the *n*-doped half having an excess of electrons and to the left the *p*-doped half containing excess holes. An essential point in this respect is that lasers and LEDs are *not* in equilibrium. This means that the population of electrons in different states does not follow the usual statistical mechanical Fermi distribution. The reason is that we have placed the device in a circuit and keep injecting electrons from the right and extracting electrons (injecting holes) from the left end of the junction. This constant supply of carriers drives the device out of equilibrium. Consider now the *n*-doped half, in which additional electrons are injected. If all the electrons here thermalize, i.e. establish a local equilibrium among themselves, this amounts to pushing up the Fermi level locally. Similarly, adding holes to the *p*-side amount to a lowering of the Fermi level there. Hence, under the assumption of separate quasi-equilibria on both sides the electron concentration n in the *n*-side and the hole concentration p in the *p*-side are characterized by specific *electrochemical potentials* denoted μ_e and μ_h, respectively. Clearly, in true equilibrium we would have $\mu_e = \mu_h = E_F$. The actual non-equilibrium situation is illustrated in Fig. 22.1. This figure shows the band edges U_c and U_v in the junction and the position of the electrochemical potentials in cases with low (a) and high (b) injection. Panel (a) corresponds to the LED regime, for which injection is too small to achieve population inversion. Panel (b) is the laser situation, where the electrochemical potentials are pushed into the bands and a substantial density of excess electrons and holes is present.

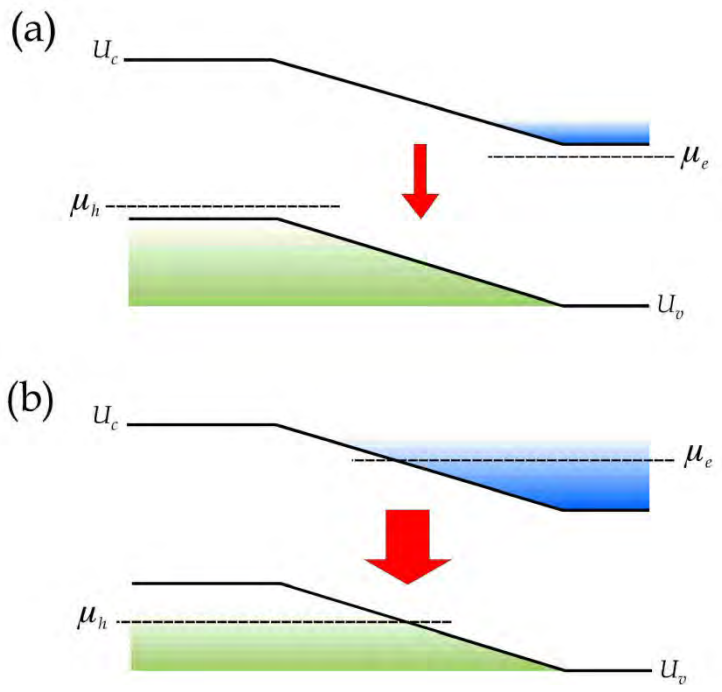


Figure 22.1. Biased pn junction in non-equilibrium. The red arrows indicate electron-hole recombination across the gap. In (a), the injected current is small and we are in the LED regime. In (b), we are in the laser regime for which the electrochemical potentials may be pushed into the bands.

A special situation arises in the interface where the two halves meet. Here, excess electrons in the conduction band have a chance to relax down into empty states in the valence band. This relaxation process is often referred to as electron-hole recombination. If the energy lost by the electron is given off in the form of radiation, the device is a light-emitting diode. The minimum energy loss is the band gap energy E_g and we therefore expect emission with a spectrum centered slightly above E_g. If the probability of radiative emission is assumed independent of the initial and final energy of the electron, a simple calculation of the spectrum can be done. To this end, consider an electron jumping from the conduction band into the valence band while emitting a photon with en energy equal to that lost by the electron. If the final energy of the electron in the valence band is E and the emitted photon energy is $\hbar\omega$, it follows that the initial electron energy of the transition must be $E+\hbar\omega$. The number of emitted photons will be proportional to the number of recombining electron-hole pairs. Thus, we can evaluate the intensity distribution as the integral

$$I(\omega) \propto \hbar\omega \int_{U_c-\hbar\omega}^{U_v} p(E)n(E+\hbar\omega)dE\,,$$

where the final electron energy E is restricted to the valence band $E \le U_v$. Note that this simplified model completely ignores "selection rules" in that all transitions are equally probable. The lower limit in the integral is derived from the requirement that the initial energy lies in the conduction band, i.e. $U_c \le E+\hbar\omega$. In this integral, $p(E)dE$

should be understood as the number of holes in the interval $[E, E+dE]$ and similarly for the electrons. With an electrochemical potential μ_e, the Fermi function for the electrons is given by

$$f_e(E) = \frac{1}{\exp\{(E-\mu_e)/kT\}+1}. \tag{22.1}$$

The electron density per unit energy is given by this function, evaluated at an energy $E+\hbar\omega$, times the density of states in the conduction band. Hence, in a bulk semiconductor the result is

$$n(E+\hbar\omega) = \frac{\sqrt{2}m_e^{3/2}}{\pi^2\hbar^3}\frac{\sqrt{E+\hbar\omega-U_c}}{\exp\{(E+\hbar\omega-\mu_e)/kT\}+1}$$

taking the electron density of states from Eq.(8.4). In the LED regime, the position of μ_e ensures that $(E+\hbar\omega-\mu_e)/kT \gg 1$ and so

$$n(E+\hbar\omega) \approx \frac{\sqrt{2}m_e^{3/2}}{\pi^2\hbar^3}\sqrt{E+\hbar\omega-U_c}\,\exp\{-(E+\hbar\omega-\mu_e)/kT\}.$$

The probability that a level is occupied by a hole must equal one minus the probability that the level is occupied by an electron. Hence, with an electrochemical potential μ_h the hole Fermi function is

$$f_h(E) = 1-\frac{1}{\exp\{(E-\mu_h)/kT\}+1} = \frac{1}{\exp\{(\mu_h-E)/kT\}+1}. \tag{22.2}$$

Using arguments similar to those used for electrons, the hole density per energy interval is

$$p(E) \approx \frac{\sqrt{2}m_h^{3/2}}{\pi^2\hbar^3}\sqrt{U_v-E}\,\exp\{(E-\mu_h)/kT\}.$$

Taking the product of electron and hole densities we find the following result for the emitted spectrum

$$\begin{aligned} I(\omega) &\propto \hbar\omega\exp\left\{-\frac{\hbar\omega}{kT}\right\}\int_{U_c-\hbar\omega}^{U_v}\sqrt{U_v-E}\sqrt{E+\hbar\omega-U_c}\,dE \\ &= \hbar\omega\exp\left\{-\frac{\hbar\omega}{kT}\right\}\frac{\pi}{8}\left(\hbar\omega-E_g\right)^2\theta(\hbar\omega-E_g). \end{aligned} \tag{22.3}$$

The step-function here indicates that the energy of the emitted light must be at least the band gap energy, as expected. The spectral shape based on this result is depicted in Fig. 22.2 for a GaAs based LED operating at room temperature.

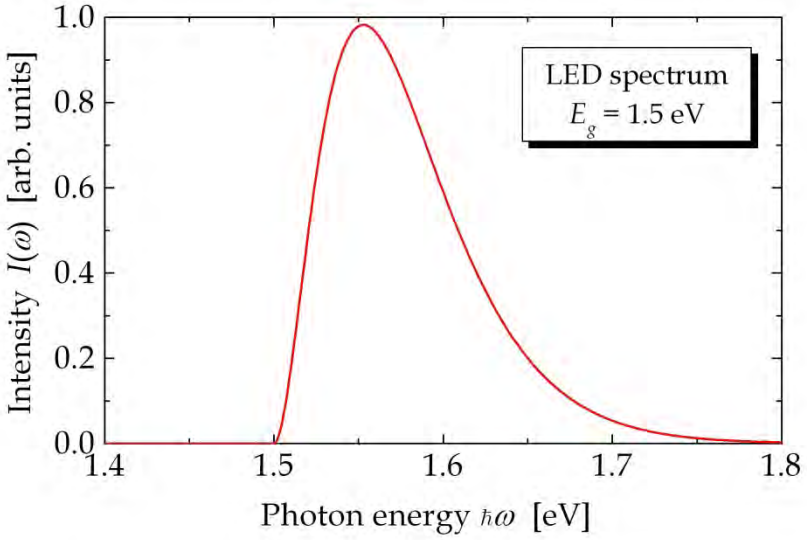


Figure 22.2. LED spectrum at room temperature.

22.1 Gain in Semiconductor Lasers

If the injected current is sufficiently high it may lead to gain inside the semiconductor. In this situation, light is amplified as it travels along rather than being partially absorbed as it normally would. To understand the conditions for this special situation, we need to describe the propagation of light in a material. If the complex refractive index is $n_r + in_i$ the complex amplitude of a wave propagating in the positive z direction evolves as

$$\mathcal{E}(z) = A\exp\left\{i\frac{2\pi}{\lambda}(n_r + in_i)z\right\},$$

where λ is the wavelength. The intensity of the wave is proportional to $|\mathcal{E}(z)|^2$ given by

$$|\mathcal{E}(z)|^2 = |A|^2\exp\left\{-\frac{4\pi}{\lambda}n_i z\right\}.$$

It is seen that the behavior of the intensity is determined by the imaginary part of the refractive index n_i: If $n_i > 0$ the field is damped and if $n_i < 0$ the field is amplified. As $n_i = \chi_i / 2n_r$ it follows that the sign of n_i is given by the sign of χ_i the imaginary part of the susceptibility (n_r is always positive in semiconductor laser materials). Hence, the condition for gain, which is a prerequisite for laser action, is that $\chi_i < 0$. To investigate this condition we return to the susceptibility expression. In the present

discussion, we wish to include a finite temperature and so the appropriate expression is

$$\chi(\omega)=\frac{4e^2\hbar^2\left|p_{cv}\right|^2}{\varepsilon_0 m^2\Omega}\sum_{\mu,\nu,\vec{k}} f_{\nu\mu}(\vec{k})\frac{\left|S_{\mu\nu}^{(D)}\right|^2}{E_{\mu\nu}(\vec{k})\left[E_{\mu\nu}^2(\vec{k})-\hbar^2(\omega+i\Gamma)^2\right]}, \tag{22.4}$$

in which $f_{\nu\mu}(\vec{k})$ is now the generalized occupation factor given by

$$f_{\nu\mu}(\vec{k})=1-f_h(E_{\nu\vec{k}})-f_e(E_{\mu\vec{k}}).$$

The fact that $1-f_h(E_{\nu\vec{k}})$ is the *electron* Fermi factor is used to obtain this result. It is seen that for the usual case of materials in thermal equilibrium at low temperatures with neither mobile electrons nor holes we'll have $f_{\nu\mu}(\vec{k})=1$ in agreement with the results of the last two chapters.

We wish to simplify the discussion by taking the broadening to zero, i.e. $\Gamma=0$. Furthermore, we will focus on the imaginary part of the susceptibility exclusively. The starting point is therefore to extract the imaginary part of Eq.(22.4) in the following manner:

$$\begin{aligned}\lim_{\Gamma\to 0}\mathrm{Im}\left\{\frac{1}{E_{\mu\nu}^2(\vec{k})-\hbar^2(\omega+i\Gamma)^2}\right\} &= \frac{1}{2E_{\mu\nu}(\vec{k})}\lim_{\Gamma\to 0}\mathrm{Im}\left\{\frac{1}{E_{\mu\nu}(\vec{k})-\hbar(\omega+i\Gamma)}+\frac{1}{E_{\mu\nu}(\vec{k})+\hbar(\omega+i\Gamma)}\right\}\\ &= \frac{1}{2E_{\mu\nu}(\vec{k})}\lim_{\Gamma\to 0}\left\{\frac{\hbar\Gamma}{(E_{\mu\nu}(\vec{k})-\hbar\omega)^2+\hbar^2\Gamma^2}-\frac{\hbar\Gamma}{(E_{\mu\nu}(\vec{k})+\hbar\omega)^2+\hbar^2\Gamma^2}\right\}\\ &= \frac{\pi}{2E_{\mu\nu}(\vec{k})}\delta\left(E_{\mu\nu}(\vec{k})-\hbar\omega\right).\end{aligned}$$

Used in Eq.(22.4) we find that

$$\chi_i(\omega)=\frac{2\pi e^2\left|p_{cv}\right|^2}{\varepsilon_0 m^2\Omega\omega^2}\sum_{\mu,\nu,\vec{k}} f_{\nu\mu}(\vec{k})\left|S_{\mu\nu}^{(D)}\right|^2\delta\left(E_{\mu\nu}(\vec{k})-\hbar\omega\right). \tag{22.5}$$

It is this general but simple result we now wish to evaluate for semiconductors of various dimensions just as we did for the ordinary susceptibility in the equilibrium situation.

We begin by looking at bulk semiconductors, for which size quantization is absent and states are labeled by $\vec{k}$ alone. The susceptibility above then becomes

$$\chi_i(\omega) = \frac{2\pi e^2 |p_{cv}|^2}{\varepsilon_0 m^2 \omega^2 (2\pi)^3} \int f_{\nu\mu}(\vec{k}) \delta\left(E_{\mu\nu}(\vec{k}) - \hbar\omega\right) d^3k .$$

To evaluate the integral, we rely on the technique used in the previous chapter and introduce $x = \hbar^2 k^2 / 2m_{eh}$. Now, in terms of x we have

$$E_{\mu\vec{k}} = E_c + \frac{\hbar^2 k^2}{2m_e} = E_c + \frac{m_{eh}}{m_e} x, \quad E_{\nu\vec{k}} = E_v - \frac{\hbar^2 k^2}{2m_h} = E_v - \frac{m_{eh}}{m_h} x, \quad E_{\mu\nu}(\vec{k}) = E_g + x .$$

Thus, upon changing the integration variable from $\vec{k}$ to x we find

$$\chi_i(\omega) = \frac{e^2 |p_{cv}|^2}{2\pi\varepsilon_0 m^2 \omega^2} \left(\frac{2m_{eh}}{\hbar^2}\right)^{3/2} \int_0^\infty \left\{1 - f_h\left(E_v - \tfrac{m_{eh}}{m_h} x\right) - f_e\left(E_c + \tfrac{m_{eh}}{m_e} x\right)\right\} \delta\left(E_g + x - \hbar\omega\right) \sqrt{x}\, dx$$

$$= \frac{e^2 |p_{cv}|^2}{2\pi\varepsilon_0 m^2 \omega^2} \left(\frac{2m_{eh}}{\hbar^2}\right)^{3/2} \left\{1 - f_h\left(E_v - \tfrac{m_{eh}}{m_h}(\hbar\omega - E_g)\right) - f_e\left(E_c + \tfrac{m_{eh}}{m_e}(\hbar\omega - E_g)\right)\right\} \sqrt{\hbar\omega - E_g}\, \theta\left(\hbar\omega - E_g\right).$$

The two electrochemical potentials are not completely independent because we expect partial neutrality. If we assume that the electron and hole densities are equal, then the results of Chapter 8 for a 3-dimensional semiconductor show that, as far as we are in the low-injection regime,

$$\mu_e + \mu_h = E_c + E_v + \frac{3}{2} kT \ln \frac{m_h}{m_e} .$$

We use this relation below even if we are not strictly in the low-injection regime to determine μ_h once μ_e has been specified. For GaAs at room temperature, the right-hand side of this relation is approximately 1.58 eV if $E_v = 0$. The spectra for four cases covering low, moderate and high injection are shown in Fig. 22.3. In the low-injection limit, μ_e is well below the conduction band edge E_c. In this case, the imaginary part of the susceptibility is always non-negative as shown in the figure. This corresponds to the usual material absorption, for which the spectrum follows a square root $\sqrt{\hbar\omega - E_g}$ behavior as derived in the previous chapter. When μ_e is pushed above E_c a region of negative absorption or, equivalently, positive gain develops. This range begins precisely at the band gap and the endpoint can be found by setting the gain expression to zero, which corresponds to the condition

$$1 - f_h\left(E_v - \tfrac{m_{eh}}{m_h}(\hbar\omega - E_g)\right) = f_e\left(E_c + \tfrac{m_{eh}}{m_e}(\hbar\omega - E_g)\right)$$

$$\Downarrow$$

$$\hbar\omega = \mu_e - \mu_h .$$

Hence, the spectral gain range is $E_g < \hbar\omega < \mu_e - \mu_h$. This clearly shows that at least one of the electrochemical potentials must be pushed into the band in order to achieve gain. The condition for gain $f_e > 1 - f_h$ is precisely that of population inversion: The population of electrons in the conduction band must exceed that of the valence band.

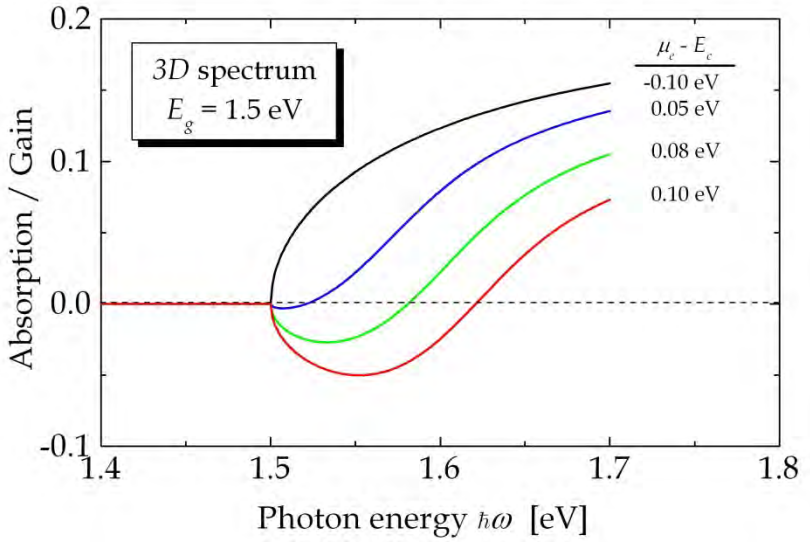


Figure 22.3. Spectra of the imaginary part of the susceptibility for different electron electrochemical potentials. Positive and negative values correspond to absorption and gain, respectively.

Next, we turn to a semiconductor quantum well for comparison. The calculation proceeds precisely as for the susceptibility in the previous chapter. Converting the sum over the two-dimensional $\vec{k}$ to an integral, Eq.(22.5) yields

$$\chi_i(\omega) = \frac{2\pi e^2 |p_{cv}|^2}{\varepsilon_0 m^2 d (2\pi)^2 \omega^2} \sum_{\mu,\nu} \left|S^{(2)}_{\mu\nu}\right|^2 \int f_{\nu\mu}(\vec{k}) \delta\left(E_{\mu\nu}(\vec{k}) - \hbar\omega\right) d^2k \,.$$

Writing again $E_{\mu\nu}(\vec{k}) = E_{g\mu\nu} + x$ we have

$$\chi_i(\omega) = \frac{2\pi e^2 |p_{cv}|^2}{\varepsilon_0 m^2 d (2\pi)^2 \omega^2} \frac{2\pi m_{eh}}{\hbar^2} \sum_{\mu,\nu} \left|S^{(2)}_{\mu\nu}\right|^2$$
$$\times \int_0^\infty \left\{1 - f_h\left(E_v + E^h_\nu - \tfrac{m_{eh}}{m_h} x\right) - f_e\left(E_c + E^e_\mu + \tfrac{m_{eh}}{m_e} x\right)\right\} \delta\left(E_{g\mu\nu} + x - \hbar\omega\right) dx.$$

The final result is therefore

$$\chi_i(\omega) = \frac{e^2 |p_{cv}|^2 m_{eh}}{\varepsilon_0 m^2 \hbar^2 \omega^2 d} \sum_{\mu,\nu} \left|S^{(2)}_{\mu\nu}\right|^2 \theta\left(\hbar\omega - E_{g\mu\nu}\right)$$
$$\times \left\{1 - f_h\left(E_v + E^h_\nu - \tfrac{m_{eh}}{m_h}(\hbar\omega - E_{g\mu\nu})\right) - f_e\left(E_c + E^e_\mu + \tfrac{m_{eh}}{m_e}(\hbar\omega - E_{g\mu\nu})\right)\right\}.$$

In Fig. 22.4, we plot the gain spectra for the quantum well similar to those of the bulk material. We have considered only the lowest subband transition $\nu = 1 \rightarrow \mu = 1$ in the plot and taken the band edge as $E_{g11} = 1.8\ \text{eV}$. To relate the electrochemical potentials we now use

$$\mu_e + \mu_h = E_c + E_1^e + E_v + E_1^h + kT \ln \frac{m_h}{m_e}.$$

The most striking difference between the 2D and 3D cases is the abrupt jump at the band edge in the 2D case. This difference simply reflects the difference in density of states: For 3D and 2D cases, the density of states varies as a square-root and a step function of the energy measured from the band edge, respectively. Hence, the steep profile in Fig. 22.4 is easily explained. If a quantum wire were considered, an even more abrupt inverse square-root behavior would have resulted.

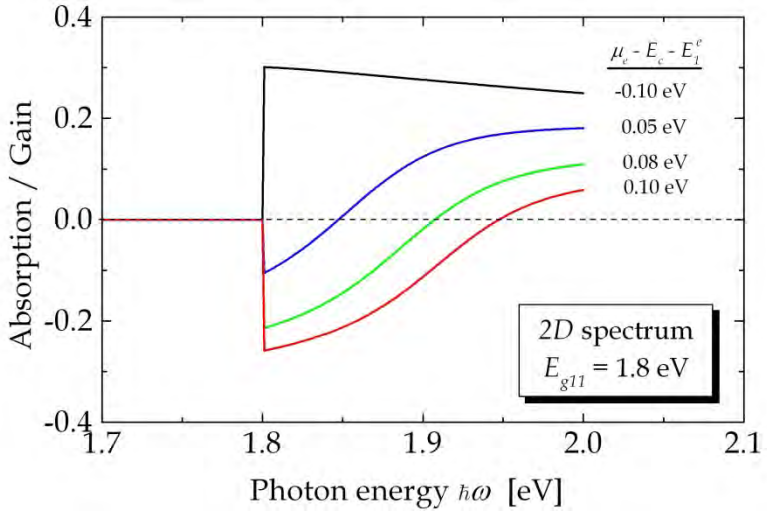


Figure 22.4. Absorption/gain spectra for a quantum well assuming different values of the electron electrochemical potential.

Thus, the laser gain directly reflects the density of states. In order to reduce the injection threshold for lasing in a given semiconductor structure it is obviously crucial that gain is maximized. As a consequence, lasing in low-dimensional structures is achieved at lower injection levels and more efficient lasers can be produced from these structures. At present, extremely efficient quantum well lasers are produced and serious efforts are put into developing lasers in quantum wire and even quantum dot structures in order to exploit their potentially large gain. Moreover, the low-dimensional structures may be used as waveguides for the optical field that will focus the intensity in the gain region. This will lower the lasing threshold even further.

22.2 Elementary Laser Model

In this section, we investigate a simple laser model. In fact, the model is not restricted to semiconductor devices and could apply to other types as well. The main point is to show how pumping, i.e. injected current, determines the laser output above and below threshold. To make things maximally simple, we consider a single electromagnetic mode, assuming that all emitted photons go into this mode. Einstein showed that a collection of excited electronic systems will emit into this mode through both spontaneous and stimulated emission. Hence, the number of photons N_p will change in time according to

$$\frac{dN_p}{dt} = \Gamma N_p + \Gamma ,$$

where the two terms represent stimulated and spontaneous processes, respectively, and Einstein used detailed balance to show that identical factors Γ enter these rates. The rate Γ is obviously related to gain. However, to obtain a realistic model, we should also include losses. These derive from two sources: absorption in the medium and outcoupling of photons through mirrors or end-facets of the laser. Since both losses are proportional to N_p, we may write

$$\frac{dN_p}{dt} = (\Gamma - \Gamma_{th})N_p + \Gamma ,$$

with Γ_{th} the threshold gain, i.e. the gain required to perfectly balance losses. Assuming $N_p(0) = 0$, this equation has the simple solution $N_p(t) = \Gamma / (\Gamma - \Gamma_{th})\{\exp(\Gamma - \Gamma_{th})t - 1\}$. Thus, if $\Gamma > \Gamma_{th}$ the photon number increases indefinitely and otherwise it approaches a constant. This behavior is clearly completely unphysical. The reason is that we have to include the exchange of energy between the photons and electronic system.

We previously found that the optical power varies as $P(z) = P_0 \exp(-2n_i k_0 z)$. If we follow a pulse travelling at speed c / n_r, this result can be converted into a time dependence $P(z = ct / n_r) = P_0 \exp(-2n_i \omega t / n_r)$ and, hence, $dP / dt = -2n_i \omega / n_r P$. Since $P \propto N_p$ it follows that Γ is related to the imaginary part of the susceptibility via $\Gamma = -2n_i \omega / n_r = -\chi_i \omega / n_r^2$. As we have seen, χ_i decreases with the density of injected electrons, eventually providing gain as $\chi_i < 0$. If we assume an approximately linear dependence on the number of injected electrons N_e, we can write $\Gamma = GN_e$. Then, the coupled dynamics of electrons and photons is given by

$$\frac{dN_p}{dt} = GN_e(N_p + 1) - \Gamma_{th} N_p,$$
$$\frac{dN_e}{dt} = -GN_e(N_p + 1) - \gamma N_e + I.$$

The first term in the electron rate is de-excitation: if a photon is emitted, an electron has dropped to the valence band. Hence, the electron and photon terms for this process are identical but opposite in sign. The second term γN_e is the rate, at which electrons relax to the valence band without emitting light. Finally, I is the pump flux of injected electrons, i.e. essentially the current. The coupled rate equations can only be solved numerically. However, at steady-state ($t \to \infty$), the time-derivatives vanish and we find

$$N_p(\infty) = \frac{GI - (\gamma + G)\Gamma_{th} + \sqrt{[GI + (\gamma + G)\Gamma_{th}]^2 - 4\gamma G\Gamma_{th} I}}{2G\Gamma_{th}},$$
$$N_e(\infty) = \frac{GI + (\gamma + G)\Gamma_{th} - \sqrt{[GI + (\gamma + G)\Gamma_{th}]^2 - 4\gamma G\Gamma_{th} I}}{2G\gamma}.$$

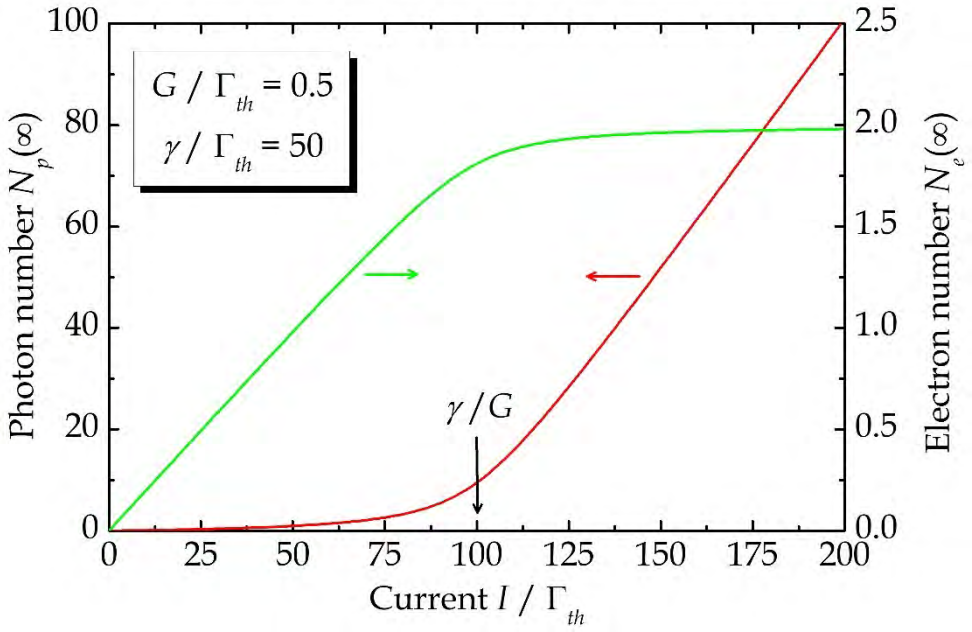


Figure 22.5. Illustration of steady-state photon and electron numbers in the laser model.

These quantities, of which $N_p(\infty)$ is essentially the laser output, are shown versus injected current in Fig. 22.5. The laser output rises dramatically above a threshold current $I_{th} = \gamma\Gamma_{th}/G$ that marks the onset of lasing. As expected, the threshold current increases with the non-radiative relaxation rate γ and decreases with gain coefficient G.

Exercise: LED peak wavelength.

It can be shown, that for a D-dimensional material, the LED spectrum is given by the following generalization of Eq.(22.3):

$$I(\omega) \propto \hbar\omega \exp\left\{-\frac{\hbar\omega}{kT}\right\}\left(\hbar\omega - E_g\right)^{D-1} \theta(\hbar\omega - E_g).$$

The different spectra are illustrated in Fig. 22.5. To simplify the analysis, we introduce a normalized photon energy $z = \hbar\omega / kT$ and the normalized band gap $g = E_g / kT$. Note that $g \gg 1$. In this manner, the spectrum can be written

$$I(z) \propto z \exp\{-z\}\left(z - g\right)^{D-1} \theta(z - g).$$

a) Show that for a three-dimensional LED the spectrum peaks at a normalized photon energy z given by

$$z = \frac{3 + g + \sqrt{9 + g^2 + 2g}}{2} \approx g + 2.$$

b) Repeat the question for a two-dimensional LED and show that here

$$z = \frac{2 + g + \sqrt{4 + g^2}}{2} \approx g + 1.$$

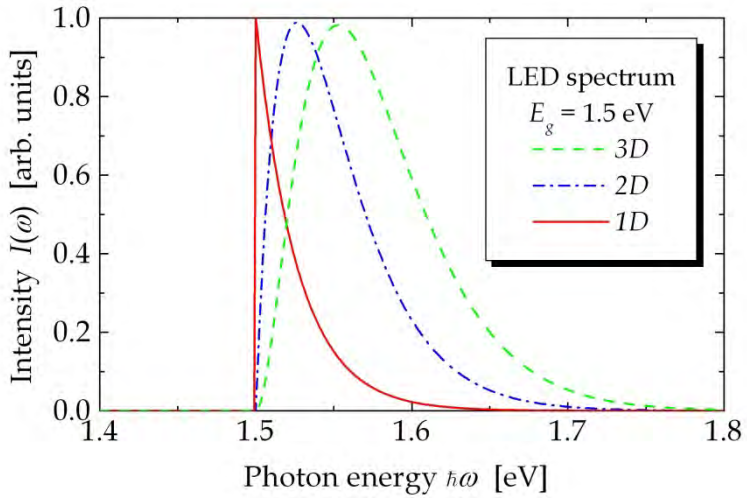


Figure 22.5. LED emission spectra for 3, 2 and 1 dimensional cases.

23. Solar Cells

The solar cell combines several aspects of both optical and electric properties of semiconductors. The optical aspects involve, obviously, absorption of solar radiation. But more subtle issues such as radiation produced by electron-hole recombination are important as well. The electric properties are equally significant as the bulk of the cell is a pn junction and the transport of carriers is clearly of great importance for the cell. In this chapter, we investigate the ideal solar cell. This means that all imperfections of real solar cells that could, in principle, be eliminated are assumed eliminated. Nevertheless, there are strict limits to the efficiency of the solar cell. The starting point is the flux of photons from the sun. The sun is approximately a black body radiator with a surface temperature of $T_{sun} \approx 5800\,\mathrm{K}$ corresponding to $kT_{sun} \approx 0.5\,\mathrm{eV}$. The photon flux (photons emitted per area×time×energy) is then given by

$$n_{sun}(E) = \frac{2\pi E^2}{h^3 c^2} \cdot \frac{1}{e^{E/kT_{sun}} - 1},$$

where E is the photon energy and $h = 2\pi\hbar$ is Planck's constant. The radiation spreads out as it travels from the sun, and on earth (outside the atmosphere) we receive a flux $n_E^{(0)}(E)$ given by

$$n_E^{(0)}(E) = n_{sun}(E)\frac{R_{sun}^2}{R_E^2} = \frac{R_{sun}^2}{R_E^2} \cdot \frac{2\pi E^2}{h^3 c^2} \cdot \frac{1}{e^{E/kT_{sun}} - 1},$$

where R_{sun} and R_E is the radius of the sun ($696 \cdot 10^6\,\mathrm{m}$) and the radius of earth's orbit around the sun ($150 \cdot 10^9\,\mathrm{m}$). We therefore find that the number of photons per area and time reaching us is

$$N_E^{(0)} = \int_0^\infty n_E^{(0)}(E)\,dE = \frac{R_{sun}^2}{R_E^2} \cdot \frac{4\pi (kT_{sun})^3 \zeta(3)}{h^3 c^2},$$

where $\zeta(x)$ is the Riemann zeta function. As each photon carries an energy E the intensity of the solar radiation becomes

$$I_E^{(0)} = \int_0^\infty n_E^{(0)}(E) E\,dE = \frac{R_{sun}^2}{R_E^2} \cdot \frac{2\pi^5 (kT_{sun})^4}{15 h^3 c^2}.$$

Plugging in numbers, we find

$$N_E^{(0)} = 6.4 \cdot 10^{21}\ \text{photons/m}^2\text{s}, \quad I_E^{(0)} = 1380\ \text{Watt/m}^2.$$

In reality, the sun is only approximately a black body radiator. Moreover, the influence of the atmosphere on earth means that a strongly modified spectrum is received. The flux corresponding to the actual spectrum is denoted $n_E(E)$ and looks as shown in Fig. 23.1. The spectrum received on earth when the sun is at an angle of 48 degrees relative to the normal is called the AM1.5 spectrum [1]. Numerical integration reveals that for this spectrum

$$N_E \approx 4.3 \cdot 10^{21} \text{ photons/m}^2\text{s}, \quad I_E \approx 1000 \text{ Watt/m}^2 . \tag{23.1}$$

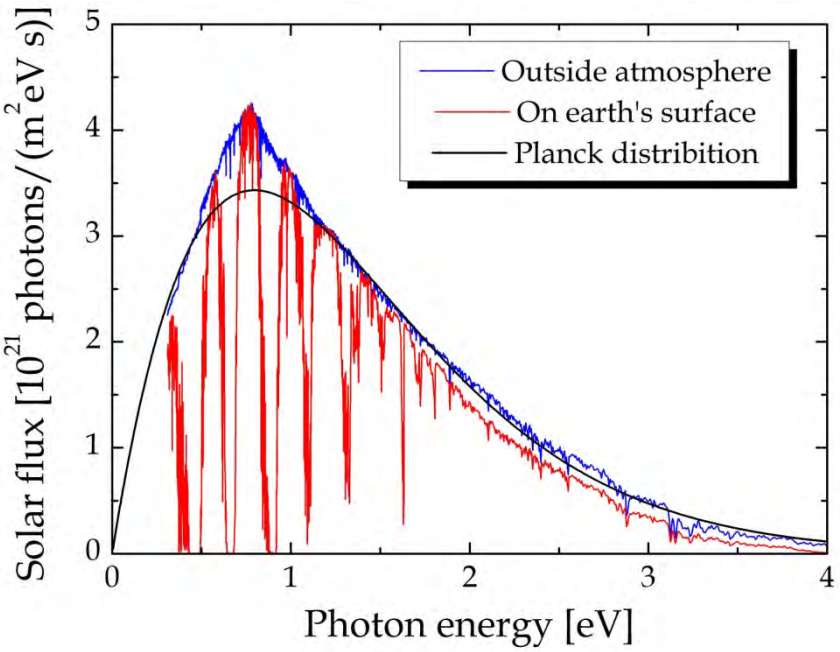


Figure 23.1. Experimental solar spectra above and below the atmosphere. The black line is the black body Planck distribution.

23.1 Ultimate Efficiency

A practical solar cell consists of semiconductor material characterized by a band gap. In ordinary cells, only a single layer is used but in more advanced "tandem solar cells", several different layers are combined with the highest band gap material on top, as illustrated in Fig. 23.2. The layers are doped to form pn junctions that produce built-in electric fields, which serve to dissociate the photo-excited electron hole pairs. If a solar cell is held at a temperature of 0 K, no energy is lost to radiative recombination. A calculation of the efficiency in this situation leads to the so-called "ultimate efficiency". Ideally, if every photon is absorbed, this might lead to an efficiency of 100%. However, immediately after absorption, an electron is a "hot carrier" with an energy higher than the conduction band edge. A certain time is needed before the electron is collected in a contact and during this time, the electron is practically certain to lose this excess energy to heat (phonons) because this process takes place on an extremely short time scale. As a consequence, the useable electron energy is reduced to that of the band edge. An analogous process applies to the holes

and, in turn, the useable energy of each electron-hole pair equals the band gap E_g. Hence, the band gap should be large to maximize this energy. However, with a large gap only a tiny fraction of the solar spectrum will be absorbed.

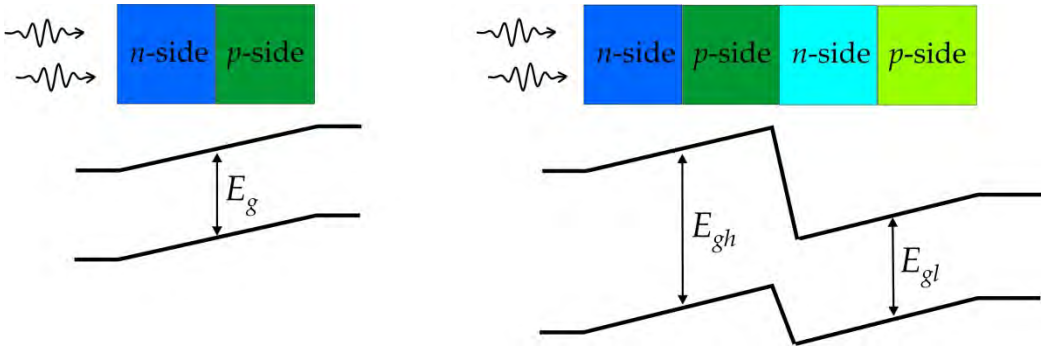


Figure 23.2. Schematic illustration of an ordinary single-layer solar cell (left) and a two-layer tandem solar cell (right).

Following this line of thought, the ultimate efficiency for a single-layer device is

$$\eta^{(0K)} = E_g \int_{E_g}^{\infty} n_E(E)dE \Big/ \int_0^{\infty} n_E(E)EdE\,. \tag{23.2}$$

Note that the lower limit of the integral in the numerator is E_g. The result is illustrated in Fig. 23.3.

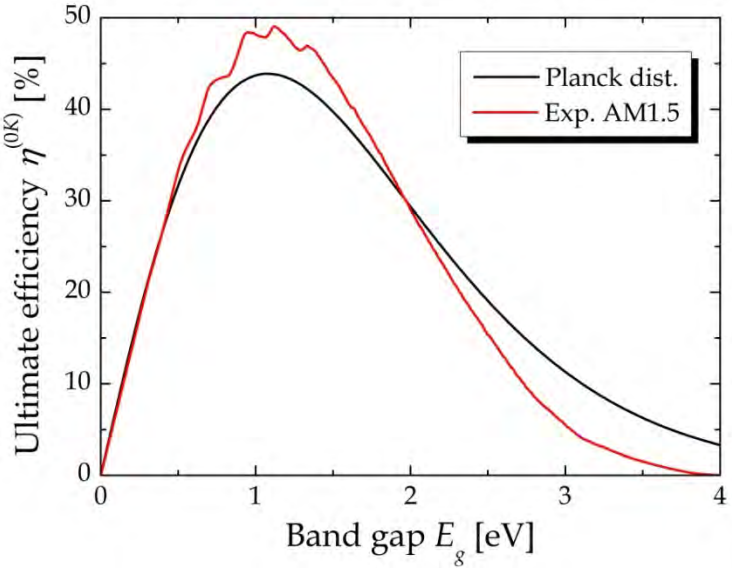


Figure 23.3. Ultimate efficiency of an ideal solar cell operated at zero temperature simulated using real (red) and black body (black) solar spectra.

Using the Planck distribution, a maximum of 44% is found at a band gap around 1.1 eV, which is very close to the value in silicon. To increase efficiency, multi-layer cells can be constructed. For example, a two-layer tandem device (Fig. 23.2) consists of two

different materials with the top layer having the highest band gap E_{gh} and the bottom layer having a lower value E_{gl}. In this structure, the ultimate efficiency becomes

$$\eta^{(0K)} = \frac{E_{gh}\int_{E_{gh}}^{\infty} n_E(E)dE + E_{gl}\int_{E_{gl}}^{E_{gh}} n_E(E)dE}{\int_0^{\infty} n_E(E)EdE}.$$

As shown in Fig. 23.4, a maximum efficiency of around 60% is obtained with E_{gl} = 0.76 eV and E_{gh} = 1.64 eV. With even more layers, the efficiency increases further and eventually reaches 100% for infinitely many layers.

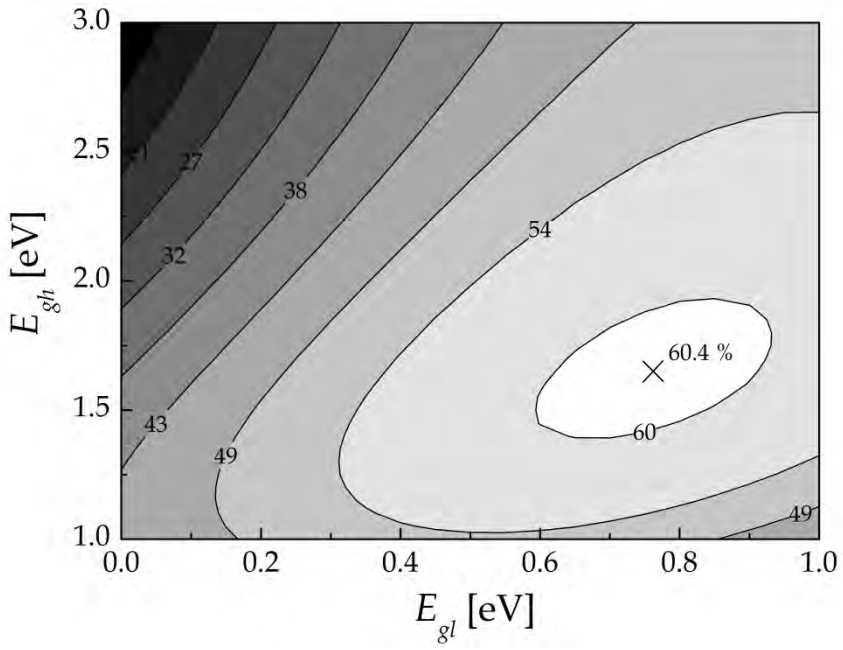


Figure 23.4. Ultimate efficiency of an ideal tandem solar cell operated at zero temperature. The maximum efficiency of 60.4 % is found for low and high band gaps of 0.76 and 1.64 eV, respectively.

It can be noted that band gaps of 0.76 and 1.64 eV match quite closely the values of the direct band gaps in Ge and GaAs, respectively.

23.2 Shockley-Queisser Limit

A solar cell at a temperature $T_{cell} \neq 0$ cannot attain the ultimate efficiency for a simple reason: the increased number of electrons and holes produced by absorption will increase the number of recombination events. The smallest possible rate is obtained when only radiative recombination occurs. Hence, energy is lost as thermal radiation emitted by the cell. This case, which was first analyzed by Shockley and Queisser [2],

leads to the maximum efficiency for a cell at a temperature $T_{cell} \neq 0$. In the literature, it is sometimes referred to as the "radiative" or "detailed balance" limit.

The useable power in the ultimate efficiency calculation is derived in terms of optical properties. It has, however, an equally simple physical interpretation in electrical terms. Hence, in the ideal limit each absorbed photon produces one electron-hole pair. This means that the cell produces an electric current density, which is just the absorbed photon flux times the elementary charge. This maximal current, which is called the short circuit current $J_{sc}^{(0K)}$, is therefore given by

$$J_{sc}^{(0K)} = e\int_{E_g}^{\infty} n_E(E)dE\,.$$

It follows from Eq.(23.1) that for a zero band gap material with perfect absorption a current density of $eN_E = 69$ mA/cm² can be produced. On the other hand, each electron-hole pair carries an energy E_g or, equivalently, a voltage of E_g/e. This voltage would be the open circuit voltage of an ideal cell at 0 K, i.e. $V_{oc}^{(0K)} = E_g/e$. Thus, the power produced by the ideal solar cell is simply the product $J_{sc}^{(0K)} \cdot V_{oc}^{(0K)}$. This product obviously equals the numerator in Eq.(23.2) as it should.

We now turn to calculating the energy lost by radiative recombination inside the solar cell. To analyze this effect, we first consider the situation, in which the cell is not illuminated. In this case, the product of electron and hole densities is simply n_i^2, i.e. the square of the intrinsic density, c.f. Chapter 9. The rate of radiative recombination events R_{rad} is proportional to the product of electron and hole densities so that $R_{rad} = Bn_i^2$. This energy is emitted as thermal radiation characterized by a Planck distribution with a temperature T_{cell} following the expression

$$n_{cell}(E) = \beta(E)\frac{4\pi E^2}{h^3c^2}\cdot\frac{1}{e^{E/kT_{cell}}-1}\,.$$

In this expression, $\beta(E)$ is the emissivity for photons of energy E. Moreover, the factor of 4 (instead of 2) accounts for the fact that the planar cell has two surfaces (front and back). In our model of a solar cell with perfect absorption above the band gap E_g and perfect transparency below, the emissivity is simply $\beta(E) = \theta(E - E_g)$. By balancing recombination events and thermal emission we therefore find that

$$R_{rad} = \int_{E_g}^{\infty} \frac{4\pi E^2}{h^3 c^2} \cdot \frac{1}{e^{E/kT_{cell}} - 1} dE$$
$$\approx \frac{1}{e} J_0, \quad J_0 \equiv e \frac{4\pi k T_{cell} E_g^2}{h^3 c^2} e^{-E_g/kT_{cell}}. \tag{23.3}$$

Here, the fact that $E_g \gg kT_{cell}$ has been used. Now, when the cell is illuminated, a bias V develops across the cell. In the case, the carrier product increases from n_i^2 to $n_i^2 e^{eV/kT_{cell}}$. Accordingly, the total radiative recombination current becomes

$$J_{rad} = J_0 e^{eV/kT_{cell}}.$$

The total current is then $J = J_{sc}^{(0K)} - J_{rad}$. It is convenient to write $J_{sc} = J_{sc}^{(0K)} - J_0$ so that

$$J(V) = J_{sc} - J_0 \left\{ e^{eV/kT_{cell}} - 1 \right\}.$$

It is clear that $J(0) = J_{sc}$ is the actual short circuit current at $T_{cell} \neq 0$. Similarly, we can find the open circuit voltage from the zero-crossing of the I/V characteristic:

$$J(V_{oc}) = 0 \quad \Rightarrow \quad eV_{oc} = kT_{cell} \ln\left\{ 1 + \frac{J_{sc}}{J_0} \right\}.$$

Using simple manipulations it follows that

$$eV_{oc} = E_g + kT_{cell} \ln\left\{ \frac{h^3 c^2}{4\pi k T_{cell} E_g^2} \int_{E_g}^{\infty} n_E(E) dE \right\}.$$

The quantity inside the curly brackets is much less than unity and so the correction is negative, i.e. $eV_{oc} \leq E_g$. The temperature dependence of the open circuit voltage is shown in Fig. 23.5. It is apparent, that it decreases roughly linearly with temperature and for E_g = 1.2 eV the open circuit voltage has dropped to 0.93 V at room temperature.

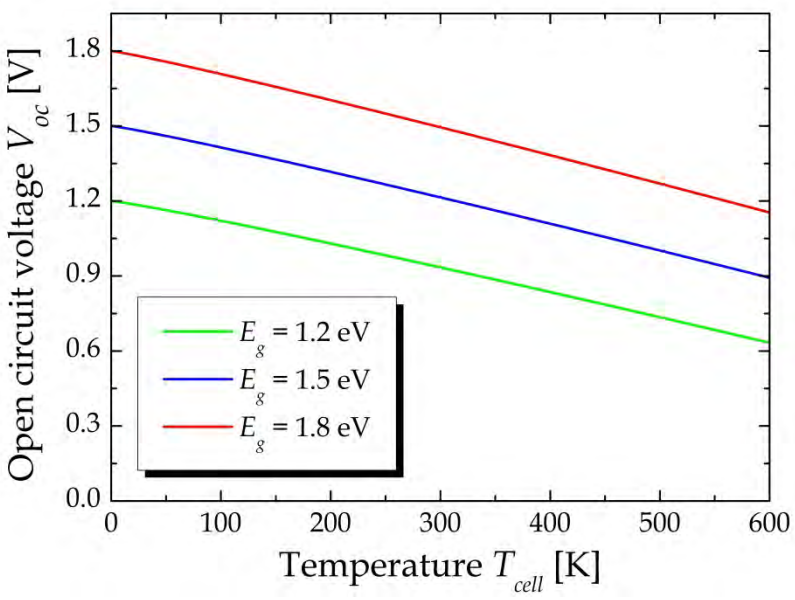


Figure 23.5. Open circuit voltage as a function of temperature and band gap.

The I/V characteristic can be written in the simple form

$$J(V) = J_0 \left\{ e^{eV_{oc}/kT_{cell}} - e^{eV/kT_{cell}} \right\}. \tag{23.4}$$

This I/V characteristic is illustrated in Fig. 23.6. The power per area that is extracted from a cell operating at a certain voltage V is given by $P = J(V) \cdot V$. The optimum working point (max power point) is found by differentiating $dP/dV = 0$ and the voltage at this point is the max point voltage V_{mp}.

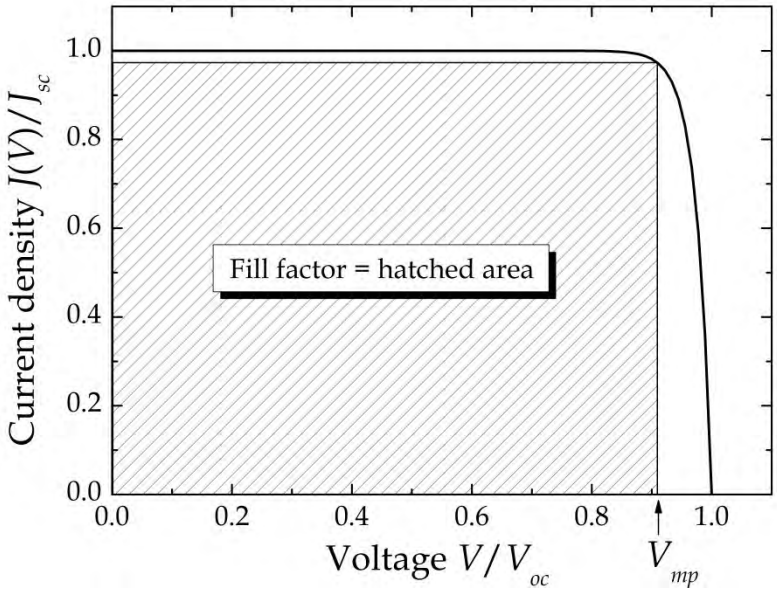


Figure 23.6. Illustration of the I/V characteristic. The maximum power is extracted when the cell voltage is at the max point V_{mp}.

Using Eq.(23.4) we find

$$J_0\left\{e^{eV_{oc}/kT_{cell}}-e^{eV_{mp}/kT_{cell}}\right\}-\frac{eV_{mp}}{kT_{cell}}J_0e^{eV_{mp}/kT_{cell}}=0 \quad \Rightarrow \quad \left\{1+\frac{eV_{mp}}{kT_{cell}}\right\}e^{eV_{mp}/kT_{cell}}=e^{eV_{oc}/kT_{cell}}.$$

The solution is given in terms of the product logarithm pl (see Chapter 13):

$$V_{mp}=\frac{kT_{cell}}{e}\left\{\mathrm{pl}\left(e^{eV_{oc}/kT_{cell}+1}\right)-1\right\}.$$

This result for the power density is usually written in terms of the fill factor *FF*

$$P_{mp}=J(V_{mp})\cdot V_{mp}=FF\cdot J_{sc}\cdot V_{oc}, \quad FF\equiv\frac{J(V_{mp})\cdot V_{mp}}{J_{sc}\cdot V_{oc}}. \tag{23.5}$$

In turn, the actual efficiency becomes

$$\eta=\eta^{(0K)}FF\cdot\frac{J_{sc}\cdot eV_{oc}}{J_{sc}^{(0K)}\cdot E_g}. \tag{23.6}$$

In Fig. 23.7, the temperature dependence of the efficiency is illustrated for three characteristic values of the band gap.

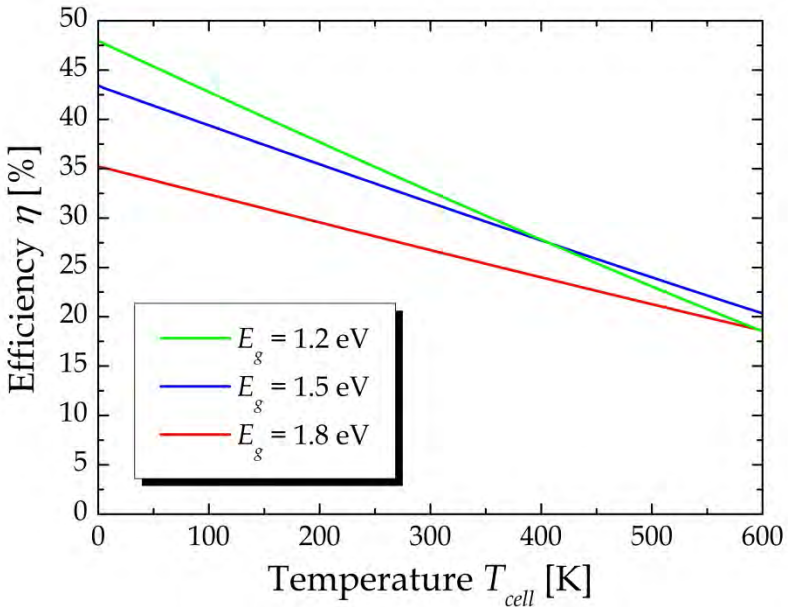


Figure 23.7. Solar cell efficiency as a function of temperature and band gap.

It is seen that the efficiency drops roughly linearly with temperature in the relevant range. Also, it is clear that at high enough temperatures, a higher band gap becomes favorable. For example, the efficiencies corresponding to band gaps of 1.2 and 1.5 eV cross around a temperature of 400 K. Fixing the temperature at room temperature

$T_{cell} =$ 300 K, the efficiency as a function of band gap is illustrated in Fig. 23.8. A maximum of 33.1% is found at a band gap of 1.35 eV.

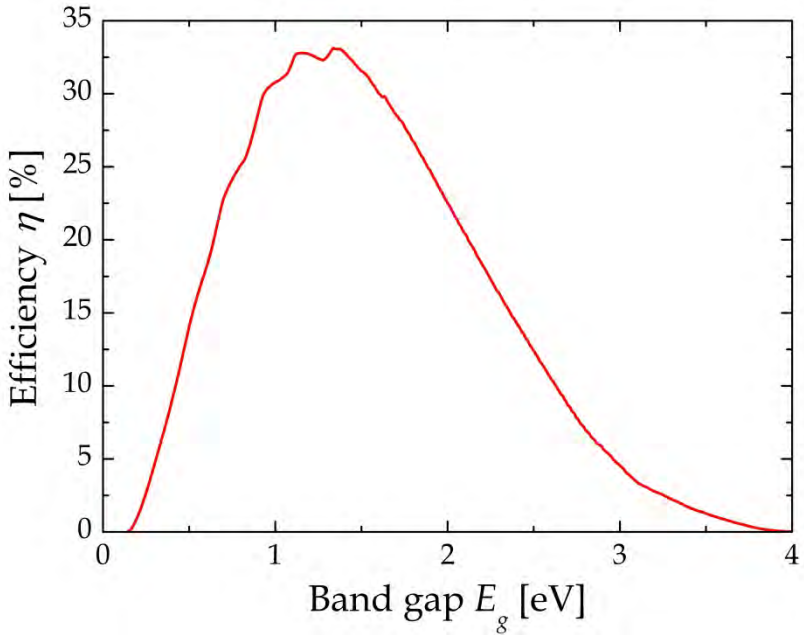


Figure 23.8. Room temperature efficiency as a function of band gap calculated in the Shockley-Queisser limit.

Exercise: Tandem solar cell in the Shockley-Queisser limit

In this exercise, we will study a tandem solar cell at finite temperature. The total voltage across the device is V and the voltage drops across the low and high gap regions are V_h and V_l, respectively, such that $V_h + V_l = V$.

a) Show that the radiative recombination currents generated in the higher gap front cell $J_{rad}^{(h)}$ and the lower gap back cell $J_{rad}^{(l)}$ are

$$J_{rad}^{(h)} = J_0^{(h)} e^{eV_h/kT_{cell}}, \quad J_0^{(h)} \equiv e\frac{4\pi kT_{cell}E_{gh}^2}{h^3c^2}e^{-E_{gh}/kT_{cell}}$$

$$J_{rad}^{(l)} = J_0^{(l)} e^{eV_l/kT_{cell}}, \quad J_0^{(l)} \equiv e\frac{4\pi kT_{cell}E_{gl}^2}{h^3c^2}e^{-E_{gl}/kT_{cell}}.$$

To compute the total currents, it is important to realize that each layer illuminates the other layer. Hence, half of the recombination radiation of the front cell is emitted in the direction of the back cell and vice versa. Hence, the total current in the high gap layer is

$$J_h = e\int_{E_{gh}}^{\infty} n_E(E)dE + e\int_{E_{gh}}^{\infty} \frac{2\pi E^2}{h^3c^2}\cdot\frac{1}{e^{E/kT_{cell}}-1}dE\cdot e^{eV_l/kT_{cell}} - J_0^{(h)}e^{eV_h/kT_{cell}}$$
$$\approx e\int_{E_{gh}}^{\infty} n_E(E)dE + \frac{1}{2}J_0^{(h)}e^{eV_l/kT_{cell}} - J_0^{(h)}e^{eV_h/kT_{cell}}.$$

b) Show that the current generated in lower gap back cell is

$$J_l = e\int_{E_{gl}}^{E_{gh}} n_E(E)dE + e\int_{E_{gh}}^{\infty} \frac{2\pi E^2}{h^3c^2}\cdot\frac{1}{e^{E/kT_{cell}}-1}dE\cdot e^{eV_h/kT_{cell}} - J_0^{(l)}e^{eV_l/kT_{cell}}$$
$$\approx e\int_{E_{gl}}^{E_{gh}} n_E(E)dE + \frac{1}{2}J_0^{(h)}e^{eV_h/kT_{cell}} - J_0^{(l)}e^{eV_l/kT_{cell}}.$$

The power generated by the tandem cell is $P = V_h J_h + V_l J_l$. We introduce $v_{l,h} \equiv eV_{l,h}/kT_{cell}$.

c) Show by differentiating that the conditions for obtaining maximum power are

$$J_h + v_h\frac{\partial J_h}{\partial v_h} + \frac{v_l}{2}J_0^{(h)}e^{v_h} = 0, \quad J_l + v_l\frac{\partial J_l}{\partial v_l} + \frac{v_h}{2}J_0^{(h)}e^{v_l} = 0.$$

These simultaneous equations reduce to

$$e\int_{E_{gh}}^{\infty} n_E(E)dE + \frac{1}{2}J_0^{(h)}\left\{e^{v_l} + v_l e^{v_h}\right\} - J_0^{(h)}\left\{1+v_h\right\}e^{v_h} = 0$$
$$e\int_{E_{gl}}^{E_{gh}} n_E(E)dE + \frac{1}{2}J_0^{(h)}\left\{e^{v_h} + v_h e^{v_l}\right\} - J_0^{(l)}\left\{1+v_l\right\}e^{v_l} = 0.$$

Their solution can only be obtained numerically. However, from these solutions the efficiency is immediately calculated. The result for room temperature conditions is illustrated in Fig. 23.9. The corresponding analysis for more than two layers can be found in Ref. [3].

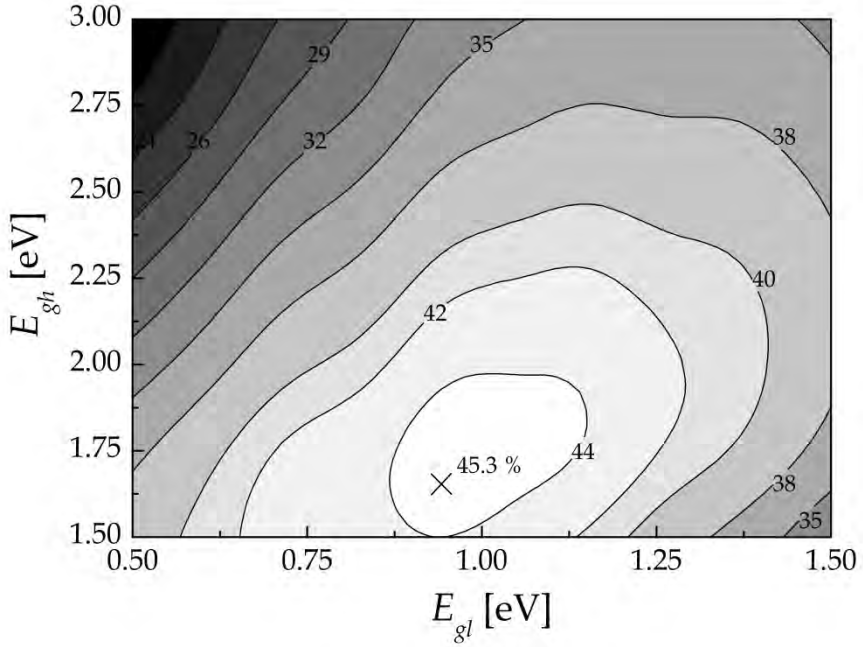


Figure 23.9. Room temperature efficiency of a tandem solar cell.

References

[1] http://rredc.nrel.gov/solar/spectra/am1.5/
[2] W. Shockley and H.J. Queisser, J. Appl. Phys. 32, 510 (1961).
[3] A. De Vos, J. Phys. D13, 839 (1980).

24. Photonic Band Gap Structures

The previous chapters have dealt with the microscopic background for the susceptibility of materials. The susceptibility χ determines the refractive index $n = \sqrt{1+\chi}$ and thereby the optical properties of the material. Since there is this link between the electronic structure and the optical properties we can to some extent design the optics by means of the atomic constituents of the material. In the discussion so far, however, we have focused on homogeneous materials. If we want even greater flexibility in the design we should go to *inhomogeneous* materials. That is, we should piece together regions of different materials in order to produce brand new composite materials with properties that are entirely different from those of the constituents. The simplest such example is a stack of alternating layers of two optically different materials. As we shall see, the properties of such a stack are radically different from a homogeneous material. To have a pronounced influence on the optical properties, the alternation of the materials should happen on a scale comparable to the optical wavelength. In a layered structure, for instance, the thickness of the different layers should be less than a wavelength and generally the typical length scale is around a few hundred nanometers.

An especially important application of such artificial optical materials is as reflectors. Typically, metals are used as reflectors (basically mirrors) but metals are not perfect reflectors, in particular not for small wavelengths. A different problem is that metals are highly absorbing materials that cannot be used with high power optical fields in e.g. lasers because they are destroyed by the heating. Finally, metallic reflectors cannot easily be designed to serve more general purposes. For instance, you might want a structure that reflects 95% at a particular wavelength but only 5% at a nearby wavelength. Such a "clever" mirror can only be made using artificial materials. To see how, we should think about band structures of ordinary materials such as GaAs shown in Fig. 8.1. The band structure is characterized by alternating regions of allowed energy bands and forbidden energy gaps. If an electron has an energy lying inside one of the gaps, it cannot propagate inside the material. This means that if such an electron is incident on a piece of the material, it will be reflected back. The idea behind *photonic band gap (PBG) structures* is to exploit this idea for optical fields. Hence, the periodic potential that is responsible for the band structure of electrons is mimicked by a periodically varying refractive index. By a proper design of the unit cell, optical fields of particular wavelength, propagation direction and polarization will be prevented from propagating in the structure. They are, therefore, necessarily reflected if the structure is composed from non-absorbing constituents. In the present chapter, we study the general principles behind the design and properties of PBG structures. For simplicity, the discussion is restricted to one and two dimensional PBG structures. For these cases, the refractive index varies along only one and two directions, respectively.

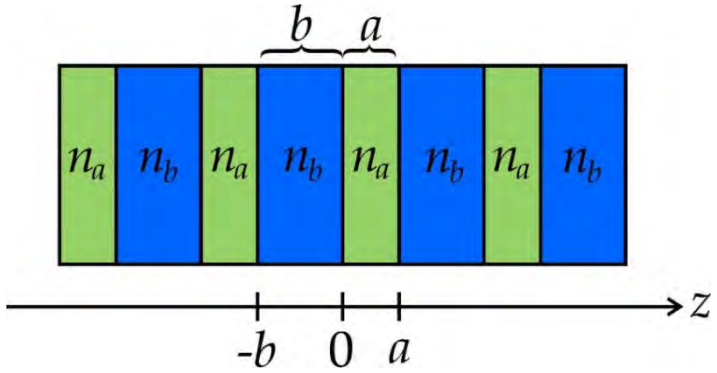


Figure 24.1. A simple one dimensional PBG structure with alternating layers of different thickness and refractive index.

24.1 One-Dimensional PBG Structures

A periodic arrangement of dielectric slabs can clearly be very complicated. However, the PBG effect appears even in the simple double-layer arrangement in Fig. 24.1. Here, the period consists of two slabs: one of thickness a and refractive index n_a and another of thickness b and refractive index n_b. The total period is $\Lambda = a + b$. The starting point for the analysis is the wave equation for the electric field. We take the z-axis as the direction, in which the refractive index varies. Furthermore, we make the simplifying assumption that the field propagates along z, i.e. perpendicular to the slabs. Hence, the polarization vector of the field is parallel to the slabs and, therefore, the field is entirely tangential. In this case, the wave equation for the amplitude of the field reads as

$$\frac{\partial^2 \mathcal{E}(z)}{\partial z^2} + n^2(z) k_0^2 \mathcal{E}(z) = 0,$$

where $k_0 = 2\pi/\lambda = \omega/c$ is the free-space wave number and $n(z)$ is the z-dependent refractive index. The key to solving this equation is the Bloch theorem, well known from electron wave functions in periodic solids. The theorem says that for a periodically varying potential the wave function $\varphi(z)$ satisfies the condition $\varphi(z+\Lambda) = \varphi(z) e^{ik\Lambda}$. In perfect analogy, the electric field satisfies $\mathcal{E}(z+\Lambda) = \mathcal{E}(z) e^{ik\Lambda}$, where k is a wave number labeling a particular solution and restricted to the Brillouin zone $-\pi/\Lambda \le k < \pi/\Lambda$. Now, inside slabs of type a and b having refractive indices n_a and n_b, respectively, the complete solution is

$$\mathcal{E}(z) = \begin{cases} A\exp(ik_a z) + B\exp(-ik_a z), & 0 < z < a \\ C\exp(ik_b z) + D\exp(-ik_b z), & -b < z < 0, \end{cases}$$

where $k_{a,b} = n_{a,b} k_0$. We need to determine the coefficients A, B, C and D and for this purpose we require continuity of the field and its derivative at $z = 0$ and $z = a$.

When considering the latter point $z = a$, we should know the field to the right of this point. This field is found directly from the field specified above together with the condition $\mathcal{E}(z+\Lambda) = \mathcal{E}(z)e^{ik\Lambda}$. Thus, the four boundary conditions are

$$\begin{gathered} A + B = C + D \\ A\exp(ik_a a) + B\exp(-ik_a a) = \{C\exp(-ik_b b) + D\exp(ik_b b)\}\exp(ik\Lambda) \\ k_a(A - B) = k_b(C - D) \\ k_a\{A\exp(ik_a a) - B\exp(-ik_a a)\} = k_b\{C\exp(-ik_b b) - D\exp(ik_b b)\}\exp(ik\Lambda). \end{gathered}$$

The simplest way of dealing with this set of equations is to rewrite them in the form $\overleftrightarrow{M}\cdot\{A,B,C,D\} = 0$, where $\overleftrightarrow{M}$ is a 4 by 4 matrix containing the coefficients of A, B, C and D in the equations above. For non-trivial solutions we require $\det\overleftrightarrow{M} = 0$, which readily leads to the condition

$$\begin{aligned} \cos k\Lambda &= \cos k_a a\cos k_b b - \frac{k_a^2 + k_b^2}{2k_a k_b}\sin k_a a\sin k_b b \\ &= \cos(n_a k_0 a)\cos(n_b k_0 b) - \frac{n_a^2 + n_b^2}{2n_a n_b}\sin(n_a k_0 a)\sin(n_b k_0 b). \end{aligned} \tag{24.1}$$

What does this equation tell us? Well, the left-hand side clearly only involves the wave number k. The right-hand side, on the other hand, does not involve k but rather $k_0 = \omega/c$. Hence, we may view this equation as a relation between the wave number k of a particular field mode and the frequency ω of this mode. A set of solutions to the equation is easily obtained: We simply find k from $k\Lambda = \cos^{-1}(\text{right-hand side})$! This provides us with k as a function of frequency. If we then invert this set of data we get frequency as a function wave number, just as a regular band structure. An example of such a calculation is shown in Fig. 24.2.

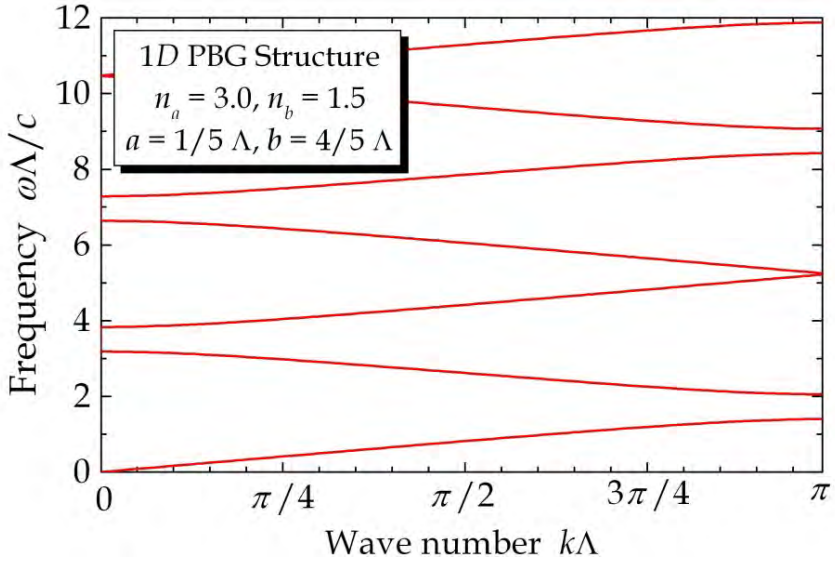


Figure 24.2. Photonic bands for a 1D PBG structure with parameters as listed in the legend.

Actually, if we simply evaluate *k* from Eq.(24.1) we find that it is sometimes complex. Those solutions clearly cannot be used because they don't correspond to propagating waves but, rather, exponentially damped waves. Hence, what is plotted in Fig. 24.2 is solely the purely real-valued solutions. What is crucially important about this result is the appearance of gaps in the band structure: *For some values of the frequency there are no propagating solutions*! The physical consequence of this fact is clear: If light of this frequency is incident on the structure it is certain to be reflected since it is not allowed to propagate into the material. From the plot, we see that the first gap is encountered if $\omega\Lambda/c$ lies in the range from $\omega\Lambda/c \approx 1.4$ to $\omega\Lambda/c \approx 2.0$. In fact, those limits can be found analytically since the parameters chosen obey the condition $n_b b/n_a a = 2$. Hence, using some trigonometric equalities [2] it can be shown that the band gap covers the range

$$\frac{2\Lambda}{n_a a}\tan^{-1}\sqrt{\frac{n_b}{2n_a+n_b}} < \frac{\omega\Lambda}{c} < \frac{2\Lambda}{n_a a}\tan^{-1}\sqrt{\frac{n_a}{2n_b+n_a}}$$

$$\Downarrow$$

$$1.402 < \frac{\omega\Lambda}{c} < 2.052,$$

where the specific parameters of the example are inserted in the last line. This is a rather substantial gap amounting to roughly 38% of the frequency at the midpoint of the gap. It is seen from the limits of the interval that such a large gap requires a large *dielectric contrast*, i.e. the difference between n_a and n_b must be large. If the two indices are nearly equal the size of the gap is $\Lambda|n_b - n_a|/(\sqrt{3}n_a^2 a)$ if the condition $n_b b/n_a a = 2$ still holds. Hence, if we want a perfect mirror over a large range of frequencies we should choose layers with very dissimilar refractive indices. It is notable, however, that a band gap can always be obtained even if the indices are almost identical. This is not the case in higher dimensions, for which a minimum index difference is required in order to produce a full band gap. Similarly, there are strict requirements for the geometries capable of producing a full band gap in 2 and 3 dimensional structures. In 1D, slabs of all thicknesses can produce gaps, albeit not necessarily large ones.

24.2 Two-Dimensional PBG Structures

A potential problem with the one dimensional PBG structure is that the band gap is highly directional. If light is incident at an angle rather than perpendicular to the stack of slabs the size of the gap shrinks and eventually disappears at sufficiently large angles of incidence. A solution to the problem is to make the periodic arrangement two dimensional as illustrated in Fig. 24.3. Hence, we imagine columns of one material embedded in another material in a periodic fashion.

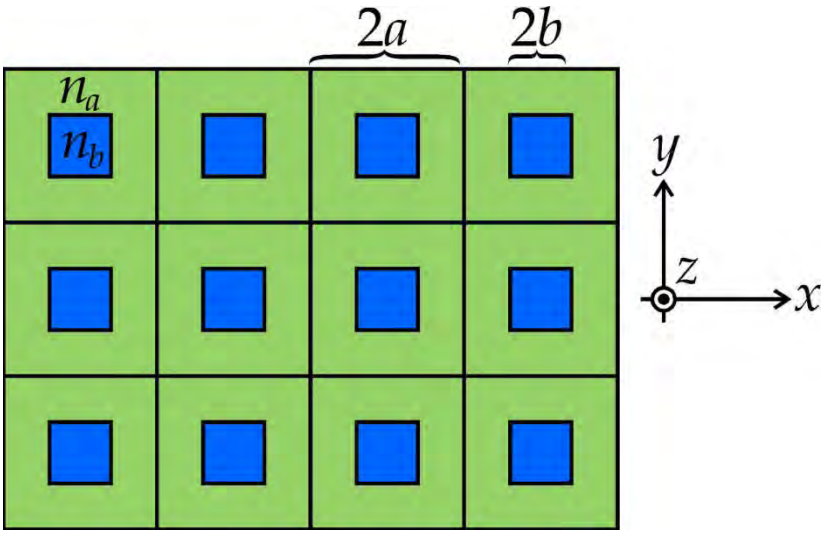


Figure 24.3. Two dimensional PBG structure with a square geometry.

The analysis of this situation is much more involved and we begin by considering the Maxwell equations

$$\nabla\times\vec{\mathcal{E}} = iZ_0k_0\vec{\mathcal{H}},\quad \nabla\times\vec{\mathcal{H}} = -\frac{i}{Z_0}n^2k_0\vec{\mathcal{E}}\,, \tag{24.2}$$

where $Z_0=\sqrt{\mu_0/\varepsilon_0}$ is the impedance of vacuum. We can isolate $\vec{\mathcal{H}}$ in these equations if we divide the latter by n^2, take the curl and use the former to substitute:

$$\nabla\times\left\{\frac{1}{n^2}\nabla\times\vec{\mathcal{H}}\right\} = -\frac{i}{Z_0}k_0\nabla\times\vec{\mathcal{E}} = k_0^2\vec{\mathcal{H}}\,. \tag{24.3}$$

This wave equation is simpler than the corresponding one for the electric field. For simplicity we consider a case for which the electric field vector $\vec{\mathcal{E}}$ is perpendicular to the columns, i.e. lies in the *xy* plane of Fig. 24.3. Accordingly, $\vec{\mathcal{H}}$ will be directed along the z axis and, therefore, perpendicular to the plane in Fig. 24.3. This configuration of the field is sometimes known as TE polarization and other times as *p*-polarization. A straightforward way of solving is by Fourier transforming $\vec{\mathcal{H}}$ as well as $\eta\equiv 1/n^2$. In our particular case, $\vec{\mathcal{H}}$ is of the form $\vec{\mathcal{H}}=\vec{e}_z\mathcal{H}(x,y)e^{i\vec{k}\cdot\vec{r}}$, where $\vec{k}=(k_x,k_y)$ is a two dimensional Bloch wave vector. The amplitude $\mathcal{H}(x,y)$ is a lattice-periodic function that can be Fourier transformed according to

$$\mathcal{H}(x,y)=\sum_{\vec{G}}\mathcal{H}_{\vec{G}}e^{i\vec{G}\cdot\vec{r}}\,,$$

where $\{\vec{G}\}$ is a complete set of reciprocal lattice vectors. In the simple square lattice, $\vec{G}=(p,q)\pi/a$ with p and q integers. Similarly, we write the inverse square of the refractive index as

$$\eta(x,y)=\sum_{\vec{G}}\eta_{\vec{G}}e^{i\vec{G}\cdot\vec{r}}. \tag{24.4}$$

The coefficients in this expansion follow from the Fourier transform of $\eta(x,y)$ as

$$\eta_{\vec{G}}=\frac{1}{(2a)^2}\int_{-a}^{a}\int_{-a}^{a}\eta(x,y)\exp\{-i(G_x x+G_y y)\}dxdy\,. \tag{24.5}$$

To see the use of these expansions we go through their use in the wave equation Eq.(24.3) step by step. First, we apply the curl operation to $\vec{\mathcal{H}}$

$$\nabla\times\vec{\mathcal{H}}=\sum_{\vec{G}}i(\vec{G}+\vec{k})\times\vec{e}_z\mathcal{H}_{\vec{G}}e^{i(\vec{G}+\vec{k})\cdot\vec{r}}.$$

Second, we multiply this result by $\eta(x,y)$ using the expansion Eq.(24.4) but relabeling from $\vec{G}$ to $\vec{G}'$ to avoid confusion:

$$\eta(x,y)\nabla\times\vec{\mathcal{H}}=\sum_{\vec{G},\vec{G}'}i\eta_{\vec{G}'}(\vec{G}+\vec{k})\times\vec{e}_z\mathcal{H}_{\vec{G}}e^{i(\vec{G}'+\vec{G}+\vec{k})\cdot\vec{r}}.$$

Third, we apply the curl to this expression

$$\nabla\times\left\{\eta(x,y)\nabla\times\vec{\mathcal{H}}\right\}=-\sum_{\vec{G},\vec{G}'}\eta_{\vec{G}'}(\vec{G}'+\vec{G}+\vec{k})\times\left\{(\vec{G}+\vec{k})\times\vec{e}_z\right\}\mathcal{H}_{\vec{G}}e^{i(\vec{G}'+\vec{G}+\vec{k})\cdot\vec{r}}.$$

Because all vectors $\vec{G}$, $\vec{G}'$ and $\vec{k}$ are perpendicular to $\vec{e}_z$ it is easily demonstrated that

$$(\vec{G}'+\vec{G}+\vec{k})\times\left\{(\vec{G}+\vec{k})\times\vec{e}_z\right\}=-\vec{e}_z(\vec{G}'+\vec{G}+\vec{k})\cdot(\vec{G}+\vec{k}),$$

and so

$$\nabla\times\left\{\eta(x,y)\nabla\times\vec{\mathcal{H}}\right\}=\vec{e}_z\sum_{\vec{G},\vec{G}'}\eta_{\vec{G}'}(\vec{G}'+\vec{G}+\vec{k})\cdot(\vec{G}+\vec{k})\mathcal{H}_{\vec{G}}e^{i(\vec{G}'+\vec{G}+\vec{k})\cdot\vec{r}}.$$

Fourth, we insert this expression as well as the expansion of $\vec{\mathcal{H}}$ in the wave equation Eq.(24.3)

$$\vec{e}_z \sum_{\vec{G},\vec{G}'} \eta_{\vec{G}'} (\vec{G}' + \vec{G} + \vec{k}) \cdot (\vec{G} + \vec{k}) \mathcal{H}_{\vec{G}} e^{i(\vec{G}' + \vec{G} + \vec{k}) \cdot \vec{r}} = k_0^2 \vec{e}_z \sum_{\vec{G}''} \mathcal{H}_{\vec{G}''} e^{i(\vec{G}'' + \vec{k}) \cdot \vec{r}}.$$

Fifth, we see by comparison that term-wise $\vec{G}'' = \vec{G} + \vec{G}'$ or, equivalently, $\vec{G}' = \vec{G}'' - \vec{G}$ and therefore equating identical terms yields

$$\sum_{\vec{G}} \eta_{\vec{G}'' - \vec{G}} (\vec{G}'' + \vec{k}) \cdot (\vec{G} + \vec{k}) \mathcal{H}_{\vec{G}} = k_0^2 \mathcal{H}_{\vec{G}''}.$$

Relabeling yet another time provides the final result

$$\sum_{\vec{G}'} \eta_{\vec{G} - \vec{G}'} (\vec{G} + \vec{k}) \cdot (\vec{G}' + \vec{k}) \mathcal{H}_{\vec{G}'} = k_0^2 \mathcal{H}_{\vec{G}}. \tag{24.6}$$

In this form, it is apparent that k_0^2 is an eigenvalue of a matrix having $\vec{G}$ and $\vec{G}'$ as indices and elements given by $\eta_{\vec{G} - \vec{G}'} (\vec{G} + \vec{k}) \cdot (\vec{G}' + \vec{k})$. Thus, the relation between wave vector and frequency is now formulated as an eigenvalue problem that can be solved using standard numerical routines if a finite set of reciprocal lattice vectors is selected. Typically, a few hundred is sufficient for convergence in such two-dimensional problems.

Generally, Eq.(24.5) is difficult to evaluate but in the present simple example illustrated in Fig. 24.3 the Fourier integral can be carried out analytically and the result is

$$\eta_{\vec{G}} = \frac{1}{a^2 G_x G_y} \left\{ \frac{1}{n_a^2} \sin G_x a \sin G_y a + \left(\frac{1}{n_b^2} - \frac{1}{n_a^2} \right) \sin G_x b \sin G_y b \right\}.$$

The band structure calculated from this expression is illustrated in Fig. 24.4. To produce the plot all $\vec{G}$ vectors of the form $\vec{G} = (p, q)\pi / a$ with $-7 \le p, q \le 7$ have been included giving a total of 225 vectors. The matrix problem is therefore of dimension 225 times 225. The parameters chosen are $n_a = 3.5$ and $n_b = 1.0$ so that, actually, the structure consists of an array of air holes in a high-index material such as GaAs. The ratio between hole size and lattice constant is taken to be $b / a = 0.8$. Again, the appearance of a band gap is noted, just as for the 1D case. In the present case, the magnitude of the gap corresponds to roughly 30% of the midpoint frequency. Due to the 2D nature of the structure, however, the gap will exist for light propagating in all directions within the *xy* plane. If light is incident at an angle to this plane, the gap may disappear and others may appear. Also, our results are valid only for light polarized with the electric field lying in the *xy* plane. For light of the perpendicular polarization, no gap is observed.

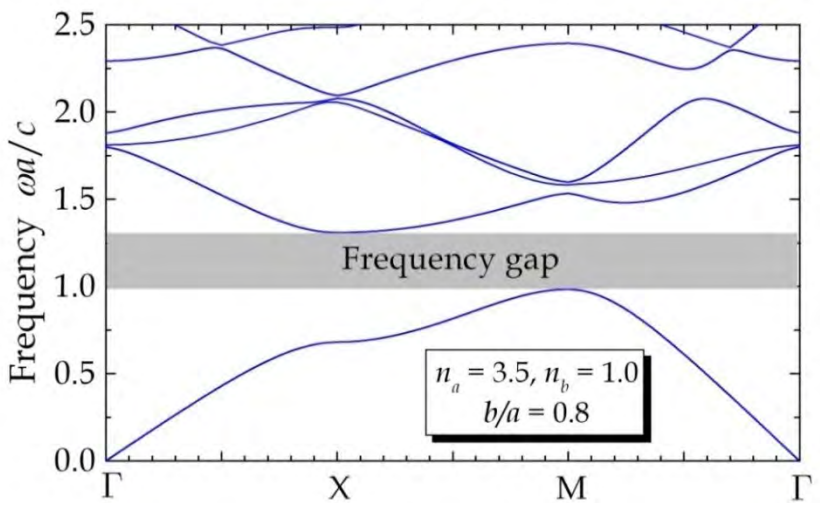


Figure 24.4. Band structure of a 2D PBG structure with a square array of holes in a high-index material. Light is TE polarized and the notation on the horizontal axis represents high symmetry points of the 2D Brillouin zone.

Exercise: TM polarized solutions

By combining the Maxwell equations in Eq.(24.2) so that $\vec{\mathcal{E}}$ is isolated we get the wave equation for the electric field

$$\frac{1}{n^2}\nabla\times\nabla\times\vec{\mathcal{E}} = k_0^2\vec{\mathcal{E}}$$

Because $\nabla\times\nabla\times\vec{\mathcal{E}} = \nabla(\nabla\cdot\vec{\mathcal{E}}) - \nabla^2\vec{\mathcal{E}} = -\nabla^2\vec{\mathcal{E}}$ using $\nabla\cdot\vec{\mathcal{E}} = 0$, this reduces to $-\eta\nabla^2\vec{\mathcal{E}} = k_0^2\vec{\mathcal{E}}$. In this exercise, we look at TM polarized solutions, for which $\vec{\mathcal{E}}$ can be written in the general form $\vec{\mathcal{E}} = \vec{e}_z\mathcal{E}(x,y)e^{i\vec{k}\cdot\vec{r}}$.

a) Show by applying Eq.(24.4) and the expansion $\mathcal{E}(x,y) = \sum_{\vec{G}}\mathcal{E}_{\vec{G}}e^{i\vec{G}\cdot\vec{r}}$ that the following eigenvalue problem is obtained:

$$\sum_{\vec{G}'}\eta_{\vec{G}-\vec{G}'}(\vec{G}'+\vec{k})^2\mathcal{E}_{\vec{G}'} = k_0^2\mathcal{E}_{\vec{G}}.$$

b) Show using $\mathcal{F}_{\vec{G}} = |\vec{G}+\vec{k}|\,\mathcal{E}_{\vec{G}}$ that this equation can be recast in the Hermitian form

$$\sum_{\vec{G}'}\eta_{\vec{G}-\vec{G}'}|\vec{G}+\vec{k}||\vec{G}'+\vec{k}|\,\mathcal{F}_{\vec{G}'} = k_0^2\mathcal{F}_{\vec{G}}.$$

References

[1] F. Szmulowicz, Am. J. Phys. 72, 1392 (2004).
[2] M.R. Spiegel *Mathematical Handbook* (McGraw-Hill, New York, 1990).

25. Optical Processes

The optical properties of nanostructures (as well as atoms and molecules) manifest themselves via optical processes. Most prominent among these is absorption, which is associated with the imaginary part of the susceptibility as discussed in the previous chapters. But other processes such as scattering and fluorescence are of importance. The processes are not independent. The energy emitted in as scattering and fluorescence processes must originate from energy transferred from light to matter, ultimately absorption in a broad sense. In this chapter, we will study the balance between these processes and study their connection.

Quite generally, the exchange of energy between light and matter is governed by the balance of electromagnetic power density (the Poynting vector $\vec{S} = \vec{\mathcal{E}} \times \vec{\mathcal{H}}$) reaching matter, on the one hand, and energy stored in the fields (electromagnetic energy density u) plus absorbed power, on the other. The energy balance is expressed as

$$-\nabla \cdot \vec{S} = \frac{\partial u}{\partial t} + \vec{j} \cdot \vec{\mathcal{E}} \, .$$

The electromagnetic energy density u is given by $u = \frac{1}{2}(\vec{\mathcal{E}} \cdot \vec{\mathcal{D}} + \vec{\mathcal{B}} \cdot \vec{\mathcal{H}})$. We will make the simplifying assumption that all processes are elastic. Hence, if we restrict ourselves to monochromatic incident fields with a frequency ω all fields vary with this frequency and we find that u contains terms varying at twice the frequency 2ω as well as temporally constant terms. Taking the time derivative, the constant terms vanish. In addition, we will average the energy balance over one period of the field. This kills off the 2ω terms as well. It should be noted that the elastic assumption means that fluorescence is ignored. Also, conversion of electromagnetic energy into heat is neglected. We write the time-averaged quantities using pointed brackets such as $<\vec{S}>$ and find

$$-\nabla \cdot <\vec{S}> = <\vec{j} \cdot \vec{\mathcal{E}}> \, .$$

Next, we integrate this relation over a finite volume V and use Gauss' theorem to transform into an integral over the bounding surface S

$$-\int_V \nabla \cdot <\vec{S}> d^3r = \int_V <\vec{j} \cdot \vec{\mathcal{E}}> d^3r$$

$$\Downarrow$$

$$-\int_S \vec{e}_n \cdot <\vec{S}> dS = \int_V <\vec{j} \cdot \vec{\mathcal{E}}> d^3r,$$

where $\vec{e}_n$ is the outward pointing normal. The left-hand side has a simple interpretation as the net electromagnetic intensity radiated into the volume. Also, the $\vec{j}\cdot\vec{\mathcal{E}}$ term is the absorbed optical power inside the volume Both electric and magnetic fields contain an incident (subscript "0") and a scattered (subscript "scat") part and so the Poynting vector has three contributions

$$\vec{S} = \vec{S}_0 + \vec{S}_{scat} + \vec{S}_{ext}$$
$$\vec{S}_0 = \vec{\mathcal{E}}_0 \times \vec{\mathcal{H}}_0, \quad \vec{S}_{scat} = \vec{\mathcal{E}}_{scat} \times \vec{\mathcal{H}}_{scat}, \quad \vec{S}_{ext} = \vec{\mathcal{E}}_0 \times \vec{\mathcal{H}}_{scat} + \vec{\mathcal{E}}_{scat} \times \vec{\mathcal{H}}_0.$$

The integral over $\vec{e}_n \cdot < \vec{S}_0 >$ vanishes because of conservation of electromagnetic power in the absence of matter. The integral of $\vec{e}_n \cdot < \vec{S}_{scat} >$ is the power radiated by scattering. Hence, the energy balance yields

$$P_{ext} = P_{scat} + P_{abs},$$

where the three terms are, respectively, the extinction, scattered, and absorbed power given by

$$P_{ext} = -\int_S \vec{e}_n \cdot < \vec{S}_{ext} > dS, \quad P_{scat} = \int_S \vec{e}_n \cdot < \vec{S}_{scat} > dS, \quad P_{abs} = \int_V < \vec{j} \cdot \vec{\mathcal{E}} > d^3r .$$

25.1 Single Dipole

We now specialize to a single point dipole, that is, a single nanostructure, atom or molecule that is sufficiently small compared to the optical wavelength that it can be regarded as a point source. The dipole moment varies with time as $\vec{p}(t) = \frac{1}{2}\left(\vec{p}(\omega)e^{-i\omega t} + c.c.\right)$ and taking the dipole position as the origin, the associated density is $\vec{p}(t)\delta(\vec{r})$. The accompanying current density is

$$\vec{j}(\vec{r},t) = \frac{\partial \vec{p}(t)}{\partial t}\delta(\vec{r}) = \frac{1}{2}\left(-i\omega\vec{p}(\omega)e^{-i\omega t} + c.c.\right)\delta(\vec{r}) . \tag{25.1}$$

With an electric field $\vec{\mathcal{E}}(\vec{r},t) = \frac{1}{2}\left(\vec{\mathcal{E}}(\vec{r},\omega)e^{-i\omega t} + c.c.\right)$ it then readily follows that

$$\begin{aligned} P_{abs} &= \int_V \frac{1}{4}\left(-i\omega\vec{p}(\omega)\cdot\vec{\mathcal{E}}^*(\vec{r},\omega) + i\omega\vec{p}^*(\omega)\cdot\vec{\mathcal{E}}(\vec{r},\omega)\right)\delta(\vec{r})\, d^3r \\ &= \frac{1}{2}\omega\,\mathrm{Im}\left\{\vec{p}(\omega)\cdot\vec{\mathcal{E}}^*(0,\omega)\right\}. \end{aligned}$$

In our case, the dipole is induced by the electric field and we therefore write $\vec{p}(\omega) = \overleftrightarrow{\alpha}_0(\omega)\cdot\vec{\mathcal{E}}(0,\omega)$, where $\overleftrightarrow{\alpha}_0$ is the polarizability (tensor). The subscript "0"

signifies that it is the "bare" polarizability, which relates the dipole moment to the total driving field. Writing the field vector as $\vec{\mathcal{E}}(\vec{r},\omega)=\mathcal{E}(\vec{r},\omega)\vec{e}$ we subsequently have

$$P_{abs}=\tfrac{1}{2}\omega\,\mathrm{Im}\left\{\vec{e}\cdot\overleftrightarrow{\alpha}_0(\omega)\cdot\vec{e}\right\}|\,\mathcal{E}(0,\omega)|^2\,. \tag{25.2}$$

This result demonstrates that the absorbed power is proportional to the imaginary part of the polarizability.

Our next step is to compute the fields radiated by the dipole, i.e. the scattered fields. The simplest strategy is to obtain the vector potential $\vec{\mathcal{A}}$ (choosing Lorentz gauge) and then find the magnetic field via $\vec{\mathcal{B}}=\nabla\times\vec{\mathcal{A}}$ and finally the electric field from the magnetic field [1]. We write $\vec{\mathcal{A}}_{scat}(\vec{r},t)=\tfrac{1}{2}\left(\vec{\mathcal{A}}_{scat}(\vec{r},\omega)e^{-i\omega t}+c.c.\right)$. For a dipole source embedded in a homogeneous medium with refractive index $n(\omega)$ (assumed real i.e. without absorption), $\vec{\mathcal{A}}_{scat}(\vec{r},\omega)$ is governed by the wave equation

$$\nabla^2\vec{\mathcal{A}}_{scat}(\vec{r},\omega)+n^2\frac{\omega^2}{c^2}\vec{\mathcal{A}}_{scat}(\vec{r},\omega)=-\frac{i\omega}{\varepsilon_0 c^2}\vec{j}(\vec{r},\omega)\,,$$

where the current amplitude is given by $\vec{j}(\vec{r},\omega)=-i\omega\vec{p}(\omega)\delta(\vec{r})$ c.f. Eq.(25.1). The solution is then

$$\vec{\mathcal{A}}_{scat}(\vec{r},\omega)=-\frac{ik_0}{4\pi\varepsilon_0 c}\vec{p}(\omega)\frac{e^{ikr}}{r},\quad k_0\equiv\frac{\omega}{c},\quad k\equiv n(\omega)k_0\,. \tag{25.3}$$

In turn, the magnetic field becomes $\vec{\mathcal{B}}_{scat}(\vec{r},t)=\tfrac{1}{2}\left(\vec{\mathcal{B}}_{scat}(\vec{r},\omega)e^{-i\omega t}+c.c.\right)$ with

$$\begin{aligned}\vec{\mathcal{B}}_{scat}(\vec{r},\omega)=\nabla\times\vec{\mathcal{A}}_{scat}(\vec{r},\omega)&=-\frac{ik_0}{4\pi\varepsilon_0 c}\nabla\left(\frac{e^{ikr}}{r}\right)\times\vec{p}(\omega)\\&=-\frac{ik_0}{4\pi\varepsilon_0 c}\frac{\partial}{\partial r}\left(\frac{e^{ikr}}{r}\right)\vec{e}_r\times\vec{p}(\omega)\\&\approx\frac{kk_0e^{ikr}}{4\pi\varepsilon_0 cr}\vec{e}_r\times\vec{p}(\omega),\end{aligned}$$

where we have applied the far-field assumption that e^{ikr} is much faster varying than r^{-1}, which is clearly the case whenever $kr\gg1$. Similarly, the electric field in the far-field limit becomes

$$\vec{\mathcal{E}}_{scat}(\vec{r},\omega)=ic^2/(n^2\omega)\nabla\times\vec{\mathcal{B}}_{scat}(\vec{r},\omega)\approx-\frac{k_0^2e^{ikr}}{4\pi\varepsilon_0 r}\vec{e}_r\times[\vec{e}_r\times\vec{p}(\omega)]=\frac{k_0^2e^{ikr}}{4\pi\varepsilon_0 r}[\vec{U}-\vec{e}_r\vec{e}_r]\cdot\vec{p}(\omega)\,.$$

The last form is obtained by considering the direction of $\vec{e}_r \times [\vec{e}_r \times \vec{p}(\omega)]$. From these fields we now compute the time-averaged Poynting vector of the dipole radiation

$$< \vec{S}_{scat} >= \frac{\varepsilon_0 c^2}{2} \mathrm{Re}\left\{ \vec{\mathcal{E}}_{scat}(\vec{r},\omega) \times \vec{\mathcal{B}}^*_{scat}(\vec{r},\omega) \right\} = -\frac{\varepsilon_0 c^2}{2} \mathrm{Re}\left\{ \mathcal{E}_{scat}(\vec{r},\omega) \mathcal{B}^*_{scat}(\vec{r},\omega) \right\} \vec{e}_r .$$

The last equality follows from the direction of the cross product between $\vec{e}_r \times \vec{p}(\omega)$ and $\vec{e}_r \times [\vec{e}_r \times \vec{p}(\omega)]$. Introducing θ as the angle between $\vec{e}_r$ and $\vec{p}(\omega)$ as illustrated in Fig. 25.1 we realize that $|\vec{e}_r \times \vec{p}(\omega)| = |\vec{e}_r \times [\vec{e}_r \times \vec{p}(\omega)]| = p(\omega)\sin\theta$ and so

$$< \vec{S}_{scat} >= \frac{cnk_0^4}{32\pi^2 \varepsilon_0} |p(\omega)|^2 \frac{\sin^2\theta}{r^2} \vec{e}_r .$$

This result can be integrated over a large sphere to provide the scattered power

$$\begin{aligned} P_{scat} &= \int_S \frac{cnk_0^4}{32\pi^2 \varepsilon_0} |p(\omega)|^2 \frac{\sin^2\theta}{r^2} dS = \frac{cnk_0^4}{12\pi\varepsilon_0} |p(\omega)|^2 \\ &= \frac{cnk_0^4}{12\pi\varepsilon_0} |\vec{e} \cdot \vec{\vec{\alpha}}_0(\omega) \cdot \vec{e}\,|^2 |\mathcal{E}(0,\omega)|^2 \end{aligned} \tag{25.4}$$

using the fact that $\int_S \frac{\sin^2\theta}{r^2} dS = 8\pi/3$.

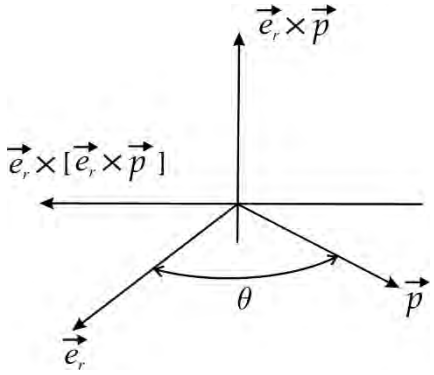


Figure 25.1 Vector diagram for the directions of scattered vector potential ($\vec{p}$), magnetic field ($\vec{e}_r \times \vec{p}$) and electric field ($\vec{e}_r \times [\vec{e}_r \times \vec{p}]$).

25.2 Bare and Dressed Polarizabilities

Will a particle be influenced by the scattered field that it emits? This may sound as a strange question but, in fact, it's important for a full understanding of optical processes. The answer is yes and the basic reason is simple. In the process of emitting light, momentum is lost and so a force acts on the particle. This effect, however, can

be incorporated into the response of the particle. Hence, instead of a "bare" particle interacting with both incident and scattered fields we find an equivalent picture of a "dressed" particle interacting with the incident filed only. To set up this equivalent picture, a careful analysis of external and local fields is needed. Primarily, we need to consider the fields without using the far-field approximation. To this end, we use the identity

$$\frac{e^{ikr}}{r} = \frac{1}{2\pi^2}\int \frac{e^{i\vec{q}\cdot\vec{r}}}{q^2-k^2}d^3q$$

and write the vector potential as

$$\vec{A}_{scat}(\vec{r},\omega) = -\frac{ik_0}{8\pi^3\varepsilon_0 c}\int \frac{e^{i\vec{q}\cdot\vec{r}}}{q^2-k^2}d^3q\,\vec{p}(\omega).$$

Taking various curls, the electric field is then readily found as

$$\begin{aligned}\vec{\mathcal{E}}_{scat}(\vec{r},\omega) &= -\frac{1}{8\pi^3 n^2\varepsilon_0}\int \frac{e^{i\vec{q}\cdot\vec{r}}}{q^2-k^2}\vec{q}\times\vec{q}\times\vec{p}(\omega)d^3q \\ &= \frac{1}{8\pi^3 n^2\varepsilon_0}\int \frac{e^{i\vec{q}\cdot\vec{r}}}{q^2-k^2}\left[q^2\overleftrightarrow{U}-\vec{q}\,\vec{q}\right]d^3q\cdot\vec{p}(\omega).\end{aligned}$$

This field, which is radiated by the dipole, also acts on the dipole! In fact, the total local field is the sum of the incident field and this radiated field evaluated at $\vec{r}=0$. The imaginary part is particularly important because it acts as a damping force being 90 degrees out of phase with the dipole. The real part (which, incidentally, cannot be correctly handled in a classical scheme but requires a fully quantum-electrodynamical theory) leads to a frequency shift of the resonance, the so-called Lamb shift. The imaginary part of the field evaluated at $\vec{r}=0$ is called the Radiation-Reaction field $\vec{\mathcal{E}}_{RR}(\omega)$. It is obtained from the imaginary part of the pole, i.e.

$$\begin{aligned}\vec{\mathcal{E}}_{RR}(\omega) &= \frac{i}{8\pi^2 n^2\varepsilon_0}\int \delta(q^2-k^2)\left[q^2\overleftrightarrow{U}-\vec{q}\vec{q}\,\right]d^3q\cdot\vec{p}(\omega) \\ &= \frac{ikk_0^2}{6\pi\varepsilon_0}\vec{p}(\omega).\end{aligned} \tag{25.5}$$

The effects of radiation reaction are clear if we note that the total field driving the dipole becomes $\vec{\mathcal{E}}(0,\omega)=\vec{\mathcal{E}}_0(0,\omega)+\vec{\mathcal{E}}_{RR}(\omega)$. Hence, the induced dipole moment is

$$\vec{p}(\omega) = \overleftrightarrow{\alpha}_0(\omega)\cdot\left\{\vec{\mathcal{E}}_0(0,\omega)+\vec{\mathcal{E}}_{RR}(\omega)\right\}.$$

Applying the expression for the radiation-reaction field, this result can be rearranged as

$$\vec{p}(\omega) = \left\{ \overleftrightarrow{U} - \frac{ikk_0^2}{6\pi\varepsilon_0}\overleftrightarrow{\alpha}_0(\omega) \right\}^{-1} \cdot \overleftrightarrow{\alpha}_0(\omega) \cdot \vec{\mathcal{E}}_0(0,\omega)$$
$$\equiv \overleftrightarrow{\alpha}(\omega) \cdot \vec{\mathcal{E}}_0(0,\omega).$$

Here, $\overleftrightarrow{\alpha}$ is the so-called dressed polarizability relating the dipole moment to the incident field alone. Hence, the effects of radiation reaction have been absorbed into the dressed polarizability. The bare polarizability was calculated previously in Chapter 2

$$\overleftrightarrow{\alpha}_0(\omega) = 2e^2 \sum_n \frac{E_n}{E_n^2 - \hbar^2\omega^2} \vec{D}_{0n}\vec{D}_{n0}\,.$$

Here, E_n is the many body energy eigenvalue of the n'the state measured relative to the ground state. Also, $\vec{D}_{0n} = \langle 0 | \sum_e \vec{r}_e | n \rangle$ is the many-electron transition dipole moment. We now make the simplifying assumption of isotropy in the polarizability as appropriate for atoms or spherical nanostructures. Also, the approximations applied above are only expected to hold when we are close to a resonance $\omega \approx E_n / \hbar \equiv \omega_n$. Hence,

$$\overleftrightarrow{\alpha}_0(\omega) \approx \frac{e^2}{\hbar} \frac{|D_{0n}|^2}{\omega_n - \omega} \overleftrightarrow{U}$$

and, thereby, with $k_n \equiv \omega_n / c$

$$\overleftrightarrow{\alpha}(\omega) = \left\{ 1 - \frac{ink_n^3}{6\pi\varepsilon_0} \frac{e^2}{\hbar} \frac{|D_{0n}|^2}{\omega_n - \omega} \right\}^{-1} \cdot \frac{e^2}{\hbar} \frac{|D_{0n}|^2}{\omega_n - \omega} \overleftrightarrow{U}\,.$$

Using simple manipulations, this result can be rewritten as

$$\overleftrightarrow{\alpha}(\omega) = \frac{e^2}{\hbar} \frac{|D_{0n}|^2}{\omega_n - \omega - i\frac{1}{2}\Gamma_n} \overleftrightarrow{U}\,, \tag{25.6}$$

where

$$\Gamma_n = \frac{e^2 n k_n^3 |D_{0n}|^2}{3\pi\varepsilon_0\hbar} \tag{25.7}$$

This quantity, which is actually the spontaneous decay rate of the *n*'th excited state [2], then results in a finite line width equal to Γ_n for the resonance. Without Γ_n, the resonance would diverge precisely at $\omega = \omega_n$. This modification should be done for every transition and so

$$\overleftrightarrow{\alpha}(\omega) = \frac{e^2}{\hbar}\sum_n \frac{|D_{0n}|^2}{\omega_n - \omega - i\frac{1}{2}\Gamma_n}\overleftrightarrow{U}.$$

We have defined the dressed polarizability so that $\overleftrightarrow{\alpha}_0(\omega)\cdot\vec{\mathcal{E}}(0,\omega) = \overleftrightarrow{\alpha}(\omega)\cdot\vec{\mathcal{E}}_0(0,\omega)$. This means that the correct versions of Eq.(25.2) and (25.4) can be written as

$$\begin{aligned} P_{abs} &= \tfrac{1}{2}\omega\,\mathrm{Im}\left\{\vec{e}\cdot\overleftrightarrow{\alpha}(\omega)\cdot\vec{e}\right\}|\mathcal{E}_0(0,\omega)|^2 \\ P_{scat} &= \frac{cnk_0^4}{12\pi\varepsilon_0}|\vec{e}\cdot\overleftrightarrow{\alpha}(\omega)\cdot\vec{e}\,|^2|\mathcal{E}_0(0,\omega)|^2 \end{aligned}.$$

We wish to introduce cross sections for the two processes and, to this end, need the intensity of the incident field $I_{inc} = |\vec{S}_0| = \frac{1}{2}n\varepsilon_0 c|\mathcal{E}_0|^2$. Dividing the power expressions above by this intensity provides absorption and scattering cross sections

$$\begin{aligned} \sigma_{abs} &= \frac{k_0}{n\varepsilon_0}\mathrm{Im}\left\{\vec{e}\cdot\overleftrightarrow{\alpha}(\omega)\cdot\vec{e}\right\} \\ \sigma_{scat} &= \frac{k_0^4}{6\pi\varepsilon_0^2}|\vec{e}\cdot\overleftrightarrow{\alpha}(\omega)\cdot\vec{e}\,|^2 \end{aligned}.$$

Now, if energy is not accumulated in the particle and it is not dissipated by other means, the absorbed power must equal the scattered power. Thus, equating the above cross sections yields

$$\mathrm{Im}\left\{\vec{e}\cdot\overleftrightarrow{\alpha}(\omega)\cdot\vec{e}\right\} \approx \frac{nk_0^3}{6\pi\varepsilon_0}|\vec{e}\cdot\overleftrightarrow{\alpha}(\omega)\cdot\vec{e}\,|^2. \tag{25.8}$$

This condition is only expected hold exactly at a resonance $\omega = \omega_n$ so that Eq.(25.6) yields $\overleftrightarrow{\alpha}(\omega_n) = 2ie^2|D_{0n}|^2/(\hbar\Gamma_n)\overleftrightarrow{U}$. Plugging this into Eq.(25.8) shows that the requirement is precisely obeyed if Γ_n is given by the expression Eq.(25.7). Hence, our expression for the spontaneous decay rate is consistent with all absorbed power eventually being re-emitted as scattering.

Exercise: Near-field relations

The calculation of the scattered field above relied on the far-field approximation $kr \gg 1$. If all terms are retained, a somewhat tedious computation shows that

$$\vec{\mathcal{E}}_{scat}(\vec{r},\omega) = \frac{k_0^2 e^{ikr}}{4\pi\varepsilon_0 r}\left\{\left[1+\frac{i}{kr}-\frac{1}{k^2r^2}\right]\overleftrightarrow{U} - \left[1+\frac{3i}{kr}-\frac{3}{k^2r^2}\right]\vec{e}_r\vec{e}_r\right\}\cdot\vec{p}(\omega).$$

a) Show by expansion around x = 0 that

$$\begin{aligned} e^{ix}\left[\frac{1}{x}+\frac{i}{x^2}-\frac{1}{x^3}\right] &\approx -\frac{1}{x^3}+\frac{1}{2x}+\frac{2i}{3} \\ e^{ix}\left[\frac{1}{x}+\frac{3i}{x^2}-\frac{3}{x^3}\right] &\approx -\frac{3}{x^3}-\frac{1}{2x} \end{aligned}.$$

b) Use this result to show that the radiation-reaction field is precisely the imaginary part of the scattered field in the limit $r \to 0$.

c) Show that the near-field, i.e. the electric field very close to the dipole, is approximately given by

$$\vec{\mathcal{E}}_{near-field}(\vec{r},\omega) \approx \frac{1}{4\pi\varepsilon_0 n^2 r^3}\left\{3\vec{e}_r\vec{e}_r - \overleftrightarrow{U}\right\}\cdot\vec{p}(\omega).$$

The same expression is obtained for the electrostatic field produced by a dipole. In a static calculation, we put $\vec{\mathcal{E}} = -\nabla\Phi$, where Φ is the electrostatic potential.

d) Show that the accompanying near-field potential is

$$\Phi_{near-field}(\vec{r},\omega) \approx \frac{1}{4\pi\varepsilon_0 n^2 r^2}\vec{e}_r\cdot\vec{p}(\omega).$$

We now consider a small nanosphere of radius *a* and refractive index n_2 embedded in a medium with refractive index n_1 subjected to a constant incident field $\vec{\mathcal{E}}_0 = \mathcal{E}_0\vec{e}_z$. Hence, the incident potential must be $\Phi_0 = -\mathcal{E}_0 z$. In polar coordinates, $\Phi_0 = -\mathcal{E}_0 r\cos\theta$ and we therefore write the full solution as $\Phi(r,\theta) = f(r)\cos\theta$. Laplace's equation for the potential $\nabla^2\Phi(r,\theta) = 0$ consequently simplifies to $r\partial^2(rf)/\partial r^2 = 2f$.

d) Show that $f(r) = r$ and $f(r) = 1/r^2$ are solutions to Laplace's equation.

e) Of the above, only the first type is allowed inside the sphere while both forms are applicable outside. Apply the boundary conditions $f(a_-) = f(a_+)$ and $n_2^2 f'(a_-) = n_1^2 f'(a_+)$ to demonstrate that the full solution outside the sphere is $\Phi = \Phi_0 + \Phi_{scat}$ with

$$\Phi_{scat}(r,\theta) \approx \frac{a^3}{r^2} \frac{n_2^2 - n_1^2}{n_2^2 + 2n_1^2} \mathcal{E}_0 \cos\theta .$$

Comparing to the general expression for the near-field potential, this demonstrates that the polarizability is $\alpha(\omega) \approx 4\pi\varepsilon_0 a^3 n_1^2 \frac{n_2^2 - n_1^2}{n_2^2 + 2n_1^2}$.

References

[1] J.R. Reitz, F.J. Milford, and R.W. Christy *Foundations of Electromagnetic Theory* (Addison-Wesley, Massachusetts, 1979).
[2] R. Loudon *The Quantum Theory of Light* (Oxford University Press, Oxford, 1991)

26. Optical Properties of Nanospheres

There are two limits, in which the optical properties of an object are relatively simply described: Either the object should be much smaller than the optical wavelength or the object should be much larger. In the former case, the electric field practically doesn't vary across the object and the field can be approximated by a constant in space. In the latter case, we enter the macroscopic regime and so-called geometric (or ray) optics can be applied. This is the regime of ordinary lenses and similar optical components. The scientific field of *nanooptics* deals with the complicated regime in between these extremes. In nanooptics, objects are typically comparable to the wavelength and so the optical fields vary inside and around the object in a very complicated fashion. This leads to quite novel phenomena such as the extraordinarily large transmission through sub-wavelength holes in metallic films [1].

To enter the field of nanooptics, we will restrict ourselves to a simple but important example: A nanosphere illuminated by a plane wave. This case has a sufficiently simple geometry that it is possible to calculate the optical field exactly. At the same time, however, it is a case of great practical relevance. For instance, it describes the optical properties of colloidal metal particles. Also, natural phenomena such as the rainbow and light scattered by fog is explained by this example. In the present chapter, we formulate a theory for such phenomena. Basically, light interacting with a nanosphere can be either scattered or absorbed by the particle. If a detector is placed behind a sample containing such spheres it will record the amount of light lost in both processes. The combined effects (scattering and absorption) are known as extinction and to compare with measurements we should discuss how scattering, absorption and extinction depend on particle size, optical wavelength and refractive indices of spheres and the surrounding medium.

Our strategy is as follows: Given an incident field, we need to find the scattered field as well as the field inside the sphere. We do this by decomposing the incident field in a sum of spherical waves. Then, we write the unknown fields as similar expansions but containing unknown expansion coefficients. Finally, the usual boundary conditions are invoked in order to determine the expansion coefficients. Once the fields are determined, the scattered and absorbed power can be calculated. This approach was established by G. Mie in 1908 and the scattering of light from particles comparable to the wavelength in size is known as Mie scattering. We start by introducing general solutions to the Helmholtz equation

$$(\nabla^2 + k^2) g(\vec{r}) = 0,$$

where $k = nk_0$ is the wave number of a field in a homogeneous medium with a refractive index n. Later, we will use these solutions to build the different fields in the problem. For a very small particle, the relevant solution is the simple outgoing

spherical wave $g(\vec{r}) = \exp(ikr)/r$. In a more general calculation, though, we should consider all possible solutions. We obviously should take advantage of the simple spherical geometry and consequently find solutions with a simple spherical symmetry. The appropriate spherical coordinates are $\{r, \theta, \phi\}$ and we seek solutions that are separable in these variables. By direct calculation it can be shown that these solutions are

$$\begin{aligned} g^{(1)}_{emn}(\vec{r}) &= \cos m\phi P^m_n(\cos\theta) j_n(kr), \quad g^{(1)}_{omn}(\vec{r}) = \sin m\phi P^m_n(\cos\theta) j_n(kr) \\ g^{(2)}_{emn}(\vec{r}) &= \cos m\phi P^m_n(\cos\theta) h_n(kr), \quad g^{(2)}_{omn}(\vec{r}) = \sin m\phi P^m_n(\cos\theta) h_n(kr), \end{aligned} \tag{26.1}$$

where $P^m_n(\cos\theta)$ is an associated Legendre polynomial, $j_n(kr)$ is a spherical Bessel function and $h_n(kr)$ is a spherical Hankel function. The superscripts are used to distinguish between these two kinds of radial behavior. Also, subscripts *e* and *o* are used for even and odd functions of ϕ, respectively. The radial functions are defined by

$$j_n(x) = \sqrt{\frac{\pi}{2x}} J_{n+1/2}(x), \quad h_n(x) = \sqrt{\frac{\pi}{2x}} \left(J_{n+1/2}(x) + iY_{n+1/2}(x) \right),$$

where $J_{n+1/2}(x)$ and $Y_{n+1/2}(x)$ are ordinary Bessel functions of first and second kind, respectively. At large values of x these functions behave as

$$j_n(x) \approx \frac{1}{x} \cos\left(x - \frac{(n+1)\pi}{2} \right), \quad h_n(x) \approx \frac{1}{i^{n+1} x} e^{ix}, \quad x \text{ large.} \tag{26.2}$$

On the other hand, when x is small $j_n(x) \approx 0$ while $\lim_{x\to 0} h_n(x) = -i\infty$. This behavior means that only the $j_n(kr)$ - type solution can be used inside the sphere. For the scattered field, only the $h_n(kr)$ - type solution can be used because it approaches an outgoing spherical wave at large *r*.

The next step is to construct solutions for the actual vectorial fields in a spherical geometry. In regions of space with a constant, isotropic refractive index, the vectorial $\mathcal{E}$ and $\mathcal{H}$ fields satisfy the wave equations

$$(\nabla^2 + k^2)\vec{\mathcal{E}}(\vec{r}) = 0, \quad (\nabla^2 + k^2)\vec{\mathcal{H}}(\vec{r}) = 0$$

as well as the transversality conditions

$$\nabla \cdot \vec{\mathcal{E}}(\vec{r}) = 0, \quad \nabla \cdot \vec{\mathcal{H}}(\vec{r}) = 0\,.$$

Also, we should make sure that $\mathcal{E}$ and $\mathcal{H}$ are related by the Maxwell equations

$$\nabla\times\vec{\mathcal{E}}(\vec{r})=ikZ\,\vec{\mathcal{H}}(\vec{r}),\quad \nabla\times\vec{\mathcal{H}}(\vec{r})=-\frac{ik}{Z}\vec{\mathcal{E}}(\vec{r}), \tag{26.3}$$

where $Z=\frac{1}{n}\sqrt{\mu_0/\varepsilon_0}$ is the impedance of the medium whose refractive index is n. We may now apply the set of g's to construct solutions to the field equations in spherical coordinates. To get from the scalar g's to the vectorial fields we introduce the operator

$$\hat{\vec{A}}\equiv-\vec{r}\times\nabla=\vec{e}_\theta\frac{1}{\sin\theta}\frac{\partial}{\partial\phi}-\vec{e}_\phi\frac{\partial}{\partial\theta}. \tag{26.4}$$

The expression of this operator in spherical coordinates is demonstrated using the relation

$$\nabla=\vec{e}_r\frac{\partial}{\partial r}+\vec{e}_\theta\frac{1}{r}\frac{\partial}{\partial\theta}+\vec{e}_\phi\frac{1}{r\sin\theta}\frac{\partial}{\partial\phi}.$$

In these formulas, $\vec{e}_r$, $\vec{e}_\theta$ and $\vec{e}_\phi$ are the usual unit vectors of spherical geometry. The operator commutes with the Laplacian, i.e. $\hat{\vec{A}}\nabla^2=\nabla^2\hat{\vec{A}}$. Hence, it follows directly that if $(\nabla^2+k^2)g(\vec{r})=0$ then $\vec{M}\equiv\hat{\vec{A}}g(\vec{r})$ fulfils the equation

$$(\nabla^2+k^2)\vec{M}=0.$$

However, if for instance $\mathcal{E}$ contains M functions then Eq.(26.3) shows that $\mathcal{H}$ will contain N functions given by

$$\vec{N}\equiv\frac{1}{k}\nabla\times\vec{M}.$$

These functions also satisfy the wave equation and can be written in terms of the g's according to $\vec{N}\equiv\hat{\vec{B}}g(\vec{r})$, where the new operator is given by

$$\hat{\vec{B}}\equiv\frac{1}{k}\nabla\times\hat{\vec{A}}=\vec{e}_\theta\frac{1}{kr}\frac{\partial^2}{\partial\theta\partial r}r+\vec{e}_\phi\frac{1}{kr\sin\theta}\frac{\partial^2}{\partial\phi\partial r}r-\vec{e}_r\frac{1}{kr}\left(\frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial}{\partial\theta}+\frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2}\right).$$

Using the Laplacian in spherical coordinates

$$\nabla^2=\frac{1}{r}\frac{\partial^2}{\partial r^2}r+\frac{1}{r^2\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial}{\partial\theta}+\frac{1}{r^2\sin^2\theta}\frac{\partial^2}{\partial\phi^2}.$$

and the fact that $(\nabla^2 + k^2)g(\vec{r}) = 0$, the B operator can be written as

$$\hat{\vec{B}} = \vec{e}_\theta \frac{1}{kr}\frac{\partial^2}{\partial\theta\partial r}r + \vec{e}_\phi \frac{1}{kr\sin\theta}\frac{\partial^2}{\partial\phi\partial r}r + \vec{e}_r \frac{1}{k}\left(\frac{\partial^2}{\partial r^2}r + rk^2\right). \tag{26.5}$$

It might be thought that, in turn, this would introduce entirely new functions $\nabla\times\vec{N}/k$ and so on without end. However,

$$\begin{aligned}\frac{1}{k}\nabla\times\vec{N} &= \frac{1}{k^2}\nabla\times\nabla\times\vec{M}\\ &= \frac{1}{k^2}\left\{\nabla(\nabla\cdot\vec{M}) - \nabla^2\vec{M}\right\}\\ &= -\frac{1}{k^2}\nabla^2\vec{M}\\ &= \vec{M},\end{aligned}$$

using the fact that the divergence of a curl vanishes.

26.1 Finding the Fields

As explained above, the analysis proceeds by (i) expressing the incident field in terms of spherical waves, i.e. the $\vec{M}$ and $\vec{N}$ functions introduced above, (ii) expressing the unknown fields in a similar form, and (iii) determine the unknowns by means of the boundary conditions of the problem. We choose the geometry (angles) as indicated in Fig. 26.1 below.

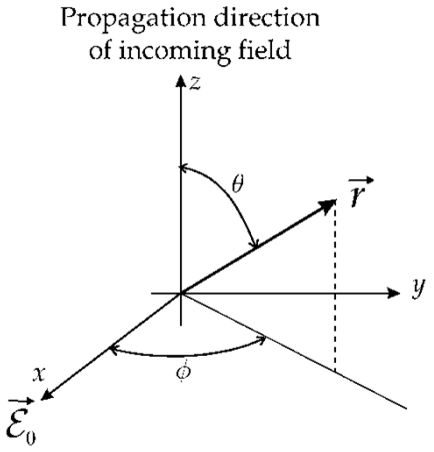


Figure 26.1. Illustration of the spherical coordinates and their relation to the electric field.

The incoming field is represented by a standard plane wave propagating along z and polarized along x as shown in Fig. 26.1 and, consequently, given by

$$\vec{\mathcal{E}}_{inc} = \mathcal{E}_0 \vec{e}_x \exp(i\,kz) = \mathcal{E}_0 \vec{e}_x \exp(i\,kr\cos\theta)$$

The phase $\exp(i\,kr\cos\theta)$ can be rewritten using the identity (using the positive sign convention for P_n^1, i.e. $P_n^1(x) = \sqrt{1-x^2}\, dP_n(x)/dx$)

$$\exp(i\,kr\cos\theta) = \frac{1}{ikr\sin\theta}\sum_{n=1}^{\infty} i^n (2n+1) P_n^1(\cos\theta) j_n(kr)\,.$$

In addition, we need to re-express the polarization vector $\vec{e}_x$ using the spherical unit vectors. The transformation between Cartesian and spherical coordinates is given by $\vec{e}_x = \sin\theta\cos\phi\vec{e}_r + \cos\theta\cos\phi\vec{e}_\theta - \sin\phi\vec{e}_\phi$ and consequently

$$\vec{\mathcal{E}}_{inc} = \left\{\sin\theta\cos\phi\vec{e}_r + \cos\theta\cos\phi\vec{e}_\theta - \sin\phi\vec{e}_\phi\right\}\frac{\mathcal{E}_0}{ikr\sin\theta}\sum_{n=1}^{\infty} i^n (2n+1) P_n^1(\cos\theta) j_n(kr). \quad (26.6)$$

The crucial point is now to express $\vec{\mathcal{E}}_{inc}$ in terms of *M* and *N* functions, i.e. as a sum of terms given by $\hat{\vec{A}} g(\vec{r})$ and $\hat{\vec{B}} g(\vec{r})$. Looking at the expression above, it is obvious that we don't need all the *g*'s defined in Eq.(26.1) to construct the field. The simplifying point is that among all the associated Legendre polynomials only P_n^1 appears in Eq.(26.6). As a consequence, we should put $m=1$ in Eq.(26.1) and focus on the corresponding *M* and *N* functions given by

$$\vec{M}_{e1n}^{(1)} \equiv \hat{\vec{A}} g_{e1n}^{(1)}(\vec{r}) = -\frac{\sin\phi}{\sin\theta} P_n^1(\cos\theta) j_n(kr)\vec{e}_\theta - \cos\phi\frac{\partial P_n^1(\cos\theta)}{\partial\theta} j_n(kr)\vec{e}_\phi$$

$$\vec{M}_{o1n}^{(1)} \equiv \hat{\vec{A}} g_{o1n}^{(1)}(\vec{r}) = \frac{\cos\phi}{\sin\theta} P_n^1(\cos\theta) j_n(kr)\vec{e}_\theta - \sin\phi\frac{\partial P_n^1(\cos\theta)}{\partial\theta} j_n(kr)\vec{e}_\phi,$$

and

$$\vec{N}_{e1n}^{(1)} \equiv \frac{1}{k}\nabla\times\vec{M}_{e1n}^{(1)}(\vec{r})$$

$$= \cos\phi P_n^1(\cos\theta) n(n+1)\frac{j_n(kr)}{kr}\vec{e}_r + \frac{1}{kr}\frac{\partial[rj_n(kr)]}{\partial r}\left\{\cos\phi\frac{\partial P_n^1(\cos\theta)}{\partial\theta}\vec{e}_\theta - \sin\phi\frac{P_n^1(\cos\theta)}{\sin\theta}\vec{e}_\phi\right\}$$

$$\vec{N}_{o1n}^{(1)} \equiv \frac{1}{k}\nabla\times\vec{M}_{o1n}^{(1)}(\vec{r})$$

$$= \sin\phi P_n^1(\cos\theta) n(n+1)\frac{j_n(kr)}{kr}\vec{e}_r + \frac{1}{kr}\frac{\partial[rj_n(kr)]}{\partial r}\left\{\sin\phi\frac{\partial P_n^1(\cos\theta)}{\partial\theta}\vec{e}_\theta + \cos\phi\frac{P_n^1(\cos\theta)}{\sin\theta}\vec{e}_\phi\right\}.$$

In addition, type-2 *M* and *N* functions (with superscript 2) are given by identical expressions except that $j_n(kr)$ is replaced by $h_n(kr)$. Now, a direct comparison between these results and Eq.(26.6) demonstrates that

$$\vec{\mathcal{E}}_{inc}(\vec{r}) = \mathcal{E}_0 \sum_{n=1}^{\infty} i^n \frac{(2n+1)}{n(n+1)} \left\{ \vec{M}_{o1n}^{(1)} - i\vec{N}_{e1n}^{(1)} \right\}.$$

We now write the remaining fields in analogous ways. As argued above, only type-2 solutions are applicable for the scattered field $\vec{\mathcal{E}}_{scat}$ while the field inside the sphere $\vec{\mathcal{E}}_{sph}$ must be constructed from type-1 solutions in order to remain finite at the origin. Thus,

$$\vec{\mathcal{E}}_{scat}(\vec{r}) = \mathcal{E}_0 \sum_{n=1}^{\infty} i^n \frac{(2n+1)}{n(n+1)} \left\{ a_n \vec{M}_{o1n}^{(2)} - ib_n \vec{N}_{e1n}^{(2)} \right\}$$

and

$$\vec{\mathcal{E}}_{sph}(\vec{r}) = \mathcal{E}_0 \sum_{n=1}^{\infty} i^n \frac{(2n+1)}{n(n+1)} \left\{ c_n \vec{M}_{o1n}^{(1)} - id_n \vec{N}_{e1n}^{(1)} \right\}.$$

The unknown coefficients a_n, b_n, c_n and d_n must now be determined by matching boundary conditions for the radial and tangential components of the fields:

$$n_1^2 \left(\vec{\mathcal{E}}_{inc} + \vec{\mathcal{E}}_{scat} \right) \cdot \vec{e}_r = n_2^2 \vec{\mathcal{E}}_{sph} \cdot \vec{e}_r \quad \text{and} \quad \left(\vec{\mathcal{E}}_{inc} + \vec{\mathcal{E}}_{scat} \right) \cdot \vec{e}_\theta = \vec{\mathcal{E}}_{sph} \cdot \vec{e}_\theta,$$

where n_1 and n_2 are refractive indices outside and inside the sphere, respectively. In this manner, for a sphere with radius a the boundary conditions require that

$$\begin{aligned}
c_n j_n(k_2 a) &= j_n(k_1 a) + a_n h_n(k_1 a) \\
c_n [k_2 a\, j_n(k_2 a)]' &= [k_1 a\, j_n(k_1 a)]' + a_n [k_1 a\, h_n(k_1 a)]' \\
d_n k_2 j_n(k_2 a) &= k_1 j_n(k_1 a) + b_n k_1 h_n(k_1 a) \\
d_n \frac{1}{k_2 a} [k_2 a\, j_n(k_2 a)]' &= \frac{1}{k_1 a} [k_1 a\, j_n(k_1 a)]' + b_n \frac{1}{k_1 a} [k_1 a\, h_n(k_1 a)]'.
\end{aligned}$$

Here, the notation $k_i = n_i \omega / c$ and $[xf(x)]' \equiv \dfrac{\partial (xf(x))}{\partial x}$ is used. Solving the four coupled equations yields

$$\begin{aligned}
a_n &= -\frac{j_n(k_2 a)[k_1 a\, j_n(k_1 a)]' - j_n(k_1 a)[k_2 a\, j_n(k_2 a)]'}{j_n(k_2 a)[k_1 a\, h_n(k_1 a)]' - h_n(k_1 a)[k_2 a\, j_n(k_2 a)]'} \\
b_n &= -\frac{n_1^2 j_n(k_1 a)[k_2 a\, j_n(k_2 a)]' - n_2^2 j_n(k_2 a)[k_1 a\, j_n(k_1 a)]'}{n_1^2 h_n(k_1 a)[k_2 a\, j_n(k_2 a)]' - n_2^2 j_n(k_2 a)[k_1 a\, h_n(k_1 a)]'}.
\end{aligned} \tag{26.7}$$

26.2 Scattering, Absorption and Extinction

To calculate the scattered intensity, we need the scattered magnetic field as determined by Eq.(26.3)

$$\vec{\mathcal{H}}_{scat}(\vec{r}) = \frac{1}{ikZ}\nabla\times\vec{\mathcal{E}}_{scat}(\vec{r}) = \frac{\mathcal{E}_0}{iZ}\sum_{n=1}^{\infty} i^n \frac{(2n+1)}{n(n+1)}\left\{a_n \vec{N}^{(2)}_{o1n} - ib_n \vec{M}^{(2)}_{e1n}\right\}.$$

The corresponding intensity is the radial component of the Poynting vector

$$I_{scat} = \frac{1}{2}\mathrm{Re}\left\{\vec{\mathcal{E}}_{scat}(\vec{r})\times\vec{\mathcal{H}}^{*}_{scat}(\vec{r})\right\}\cdot\vec{e}_r\,.$$

The power scattered into a solid angle $d\Omega$ at a distance *r* from the scatterer is given by $I_{scat}r^2 d\Omega$. Also, the intensity of the incoming field is $I_{inc} = |\mathcal{E}_0|^2/2Z$. The scattered power per solid angle divided by the intensity of the incoming beam is the so-called differential scattering cross section, i.e. the cross section per solid angle, and it can be obtained via the relation

$$\frac{\partial\sigma_{scat}(\theta,\phi)}{\partial\Omega} = \frac{2I_{scat}Zr^2}{|\mathcal{E}_0|^2}.$$

The cross section should be independent of distance from the scatterer and for that reason we may look at the fields far away from the sphere where the asymptotic expansions Eq.(26.2) apply. Hence, we find

$$\vec{\mathcal{E}}_{scat}(\vec{r}) \approx \frac{\mathcal{E}_0 e^{ikr}}{ikr}\sum_{n=1}^{\infty}\frac{(2n+1)}{n(n+1)}\left\{\vec{e}_\theta(a_n\pi_n + b_n\tau_n)\cos\phi - \vec{e}_\phi(a_n\tau_n + b_n\pi_n)\sin\phi\right\}$$
$$\vec{\mathcal{H}}_{scat}(\vec{r}) \approx \frac{\mathcal{E}_0 e^{ikr}}{Zikr}\sum_{n=1}^{\infty}\frac{(2n+1)}{n(n+1)}\left\{\vec{e}_\phi(a_n\pi_n + b_n\tau_n)\cos\phi + \vec{e}_\theta(a_n\tau_n + b_n\pi_n)\sin\phi\right\},$$

where

$$\pi_n = \frac{P^1_n(\cos\theta)}{\sin\theta}, \quad \tau_n = \frac{\partial P^1_n(\cos\theta)}{\partial\theta}.$$

We subsequently find

$$\frac{\partial\sigma_{scat}(\theta,\phi)}{\partial\Omega} = \gamma_p(\theta)\cos^2\phi + \gamma_s(\theta)\sin^2\phi\,, \tag{26.8}$$

where γ_p and γ_s are the cross sections for two special cases: scattering observed in the plane containing the x-axis (γ_p) and scattering observed in the plane perpendicular to $\vec{e}_x$ (γ_s). These cases correspond to $\phi = 0$ and $\phi = \pi/2$, respectively. Their expressions are

$$\gamma_p(\theta) = \frac{1}{k_1^2}\left|\sum_{n=1}^{\infty}\frac{(2n+1)}{n(n+1)}(a_n\pi_n + b_n\tau_n)\right|^2, \quad \gamma_s(\theta) = \frac{1}{k_1^2}\left|\sum_{n=1}^{\infty}\frac{(2n+1)}{n(n+1)}(a_n\tau_n + b_n\pi_n)\right|^2. \quad (26.9)$$

Figure 26.2 below illustrates $\gamma_p(\theta)$ for two cases. Here, a wavelength of $0.5\ \mu$m is assumed and the refractive indices are taken as those of latex spheres in water. This is a common test system for the angular dependence of optical scattering.

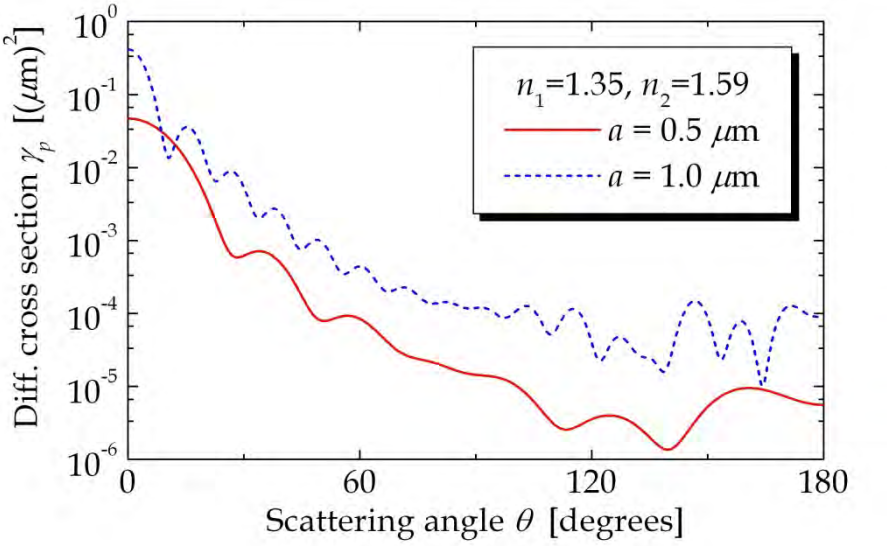


Figure 26.2. Differential scattering cross section for spheres of two different radii and wavelength $0.5\ \mu$m. The refractive indices correspond to latex spheres in water.

In addition, we wish to calculate the total (integrated over solid angle) scattering cross section. To this end, we need the result

$$\int_0^{\pi}(\pi_n\pi_m + \tau_n\tau_m)\sin\theta d\theta = \frac{2n^2(n+1)^2}{2n+1}\delta_{nm}.$$

We therefore find

$$\sigma_{scat} = \frac{2\pi}{k_1^2}\sum_{n=1}^{\infty}(2n+1)\left\{|a_n|^2 + |b_n|^2\right\}. \quad (26.10)$$

In a similar manner, the total extinction cross section is given by [2]

$$\sigma_{ext} = -\frac{2\pi}{k_1^2}\sum_{n=1}^{\infty}(2n+1)\,\mathrm{Re}(a_n + b_n). \tag{26.11}$$

The absorption cross section is given by the difference between these results. If a series expansion with respect to a is made the first non-vanishing term comes from b_1:

.

$$b_1 \approx \frac{2i}{3}\cdot\frac{n_2^2 - n_1^2}{2n_1^2 + n_2^2}(k_1 a)^3 \tag{26.12}$$

The real part is then

$$\mathrm{Re}(b_1) \approx -\frac{2n_1^2\,\mathrm{Im}(n_2^2)}{|\,2n_1^2 + n_2^2\,|^2}(k_1 a)^3.$$

and the extinction (actually absorption) cross section becomes

$$\sigma_{ext} \approx \frac{12\pi n_1^2\,\mathrm{Im}(n_2^2)}{|\,2n_1^2 + n_2^2\,|^2}k_1 a^3. \tag{26.13}$$

This is the celebrated Rayleigh cross section, which is valid for small particles. We see that it emerges as a special case of the general theory. Also, since $\pi_1 = -1$ and $\tau_1 = -\cos\theta$ we find in this limit

$$\begin{aligned}\frac{\partial\sigma_{scat}(\theta,\phi)}{\partial\Omega} &\approx \frac{9\,|\,b_1\,|^2}{4k_1^2}\left\{\cos^2\theta\cos^2\phi + \sin^2\phi\right\}\\ &= \frac{9\,|\,b_1\,|^2}{4k_1^2}\left\{1 - (\vec{e}_r\cdot\vec{e}_x)^2\right\},\end{aligned}$$

which shows that scattering from very small particles predominantly is in the plane perpendicular to the polarization of the incident light. The plots below show numerical spectra (extinction cross section vs. photon energy) calculated for diameters of 40 nm, 80 nm and 160 nm assuming $n_1 = 1.35$ and using a free electron model for n_2, c.f. Chapter 3

$$n_2^2(\omega) = \varepsilon_\infty - \frac{\omega_p^2}{\omega(\omega + i\Gamma)}. \tag{26.14}$$

with $\hbar\omega_p = 9.3$ eV, $\hbar\Gamma = 0.15$ eV and $\varepsilon_\infty = 6$. These parameters are appropriate for silver nanospheres. Looking at Eq.(26.12) it is clear that in the Rayleigh limit, a resonance appears when $|2n_1^2 + n_2^2|$ is small. The approximate minimum value is found at the frequency, for which the real part vanishes. In the free electron model, the happens when $2n_1^2 + \varepsilon_\infty - \omega_p^2/\omega^2 = 0$ so that $\omega = \omega_p/\sqrt{2n_1^2 + \varepsilon_\infty}$. For silver the resonance occurs at $\hbar\omega \approx 3.0$ eV corresponding to a wavelength of 415 nm. This value is only correct for very small spheres, however. From Fig. 26.3 it is seen that even for 40 nm spheres the resonance is slightly shifted (by roughly 15 nm) compared to the Rayleigh limit.

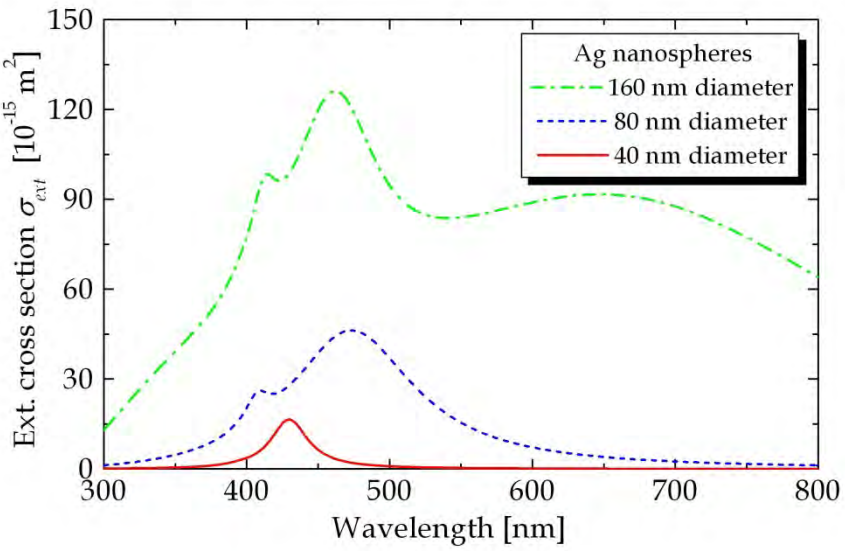


Figure 26.3. Total extinction cross section for silver nanospheres of different radii.

Exercise: Intensity enhancement

At the surface of a metallic sphere, the optical intensity can be significantly larger than the incident intensity.

a) Show from the definition of the differential scattering cross section that at the surface

$$\frac{I_{scat}}{I_{inc}} = \frac{1}{a^2}\frac{\partial\sigma_{scat}(\theta,\phi)}{\partial\Omega}.$$

It follows that the angular average of this ratio is

$$\left\langle \frac{I_{scat}}{I_{inc}} \right\rangle = \frac{1}{4\pi a^2}\int \frac{\partial\sigma_{scat}(\theta,\phi)}{\partial\Omega} d\Omega = \frac{\sigma_{scat}}{4\pi a^2}.$$

b) Show that in the Rayleigh limit, where Eq.(26.12) applies

$$\left\langle \frac{I_{scat}}{I_{inc}} \right\rangle = \frac{3|b_1|^2}{2k_1^2 a^2} = \frac{2}{3}\left|\frac{n_2^2 - n_1^2}{2n_1^2 + n_2^2}\right|^2 (k_1 a)^4 .$$

The maximum enhancement is found at the resonance frequency $\omega = \omega_p / \sqrt{2n_1^2 + \varepsilon_\infty}$ at which $2n_1^2 + n_2^2 \approx i\omega_p^2 \Gamma / \omega^3$.

c) Show that provided $\omega_p^2 \Gamma / \omega^3 \ll 1$

$$\left\langle \frac{I_{scat}}{I_{inc}} \right\rangle \approx \frac{6n_1^4}{(2n_1^2 + \varepsilon_\infty)^3}\left(\frac{\omega_p}{\Gamma}\right)^2 (k_1 a)^4 .$$

d) Evaluate this ratio for silver spheres of radius 10, 20 and 30 nm in a medium whose refractive index is 1.35.

References

[1] T.W. Ebbesen *et al.* Nature 391, 667 (1998).
[2] J.D. Jackson *Classical Electrodynamics* (Wiley, New York, 1999).

27. Nanoparticle Optics in the Electrostatic Limit

In the previous chapter, we were restricted to nanospheres because a full solution of Maxwell's equation was needed. Such an analytic calculation is only possible for very simple geometries. However, if the particles become sufficiently small so that the spatial variation of the electromagnetic field can be simplified, more complicated shapes can be studied. Hence, in this chapter, we adopt the "electrostatic limit" that is equivalent to ignoring the spatial variation of the incident electric field inside the particle.

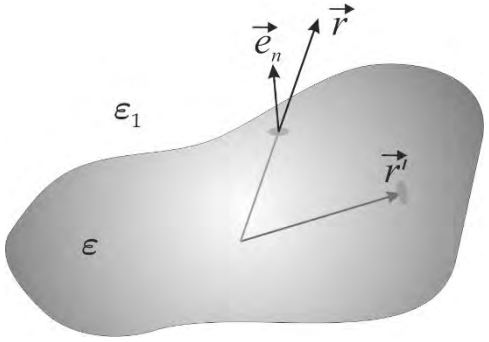


Figure 27.1. Geometry of metal nanoparticle having a dielectric constant ε embedded in a homogeneous medium with a dielectric constant ε_1.

We consider a nanoparticle such as the one depicted in Fig. 27.1 and apply the electrostatic method presented in Ref. [1]. In the electrostatic picture, an electric field only induces polarization charges on the surface of the particle. Hence, the total electrostatic potential $\Phi(\vec{r})$ is the sum of an incident part $\Phi_0(\vec{r})$ and the contribution generated by the surface charges. If the surface charge density is $\sigma(\vec{r})$, the potential is given by

$$\Phi(\vec{r}) = \Phi_0(\vec{r}) + \frac{1}{4\pi\varepsilon_0} \oint \frac{\sigma(\vec{r}')}{|\vec{r} - \vec{r}'|} dS',$$

where the integral is over the entire surface of the particle. We can now use this to compute the electric field via the relation $\vec{\mathcal{E}}(\vec{r}) = -\nabla\Phi(\vec{r})$. In particular, we wish to compute the electric field on the surface. In this situation, care has to be taken because the normal component is discontinuous. The normal component is given by $\mathcal{E}_n(\vec{r}) = -\vec{e}_n \cdot \nabla\Phi(\vec{r})$, where $\vec{e}_n$ is the outward unit normal vector at position $\vec{r}$. The discontinuity means that the normal component just outside the particle $\mathcal{E}_n^{out}$ is related to the one just inside $\mathcal{E}_n^{in}$ via the relation

$$\mathcal{E}_n^{out}(\vec{r}) - \frac{\sigma(\vec{r})}{2\varepsilon_0} = \mathcal{E}_n^{in}(\vec{r}) + \frac{\sigma(\vec{r})}{2\varepsilon_0}.$$

Using these relations, we find that

$$\mathcal{E}_n^{out/in}(\vec{r}) = \vec{e}_n \cdot \vec{\mathcal{E}}_0 + \frac{1}{4\pi\varepsilon_0} \mathrm{PV} \oint \sigma(\vec{r}\,') g(\vec{r}, \vec{r}\,') dS' \pm \frac{\sigma(\vec{r})}{2\varepsilon_0}, \tag{27.1}$$

where "PV" indicates principle value, and + and – go with the fields outside and inside the particle, respectively. Also, $\vec{\mathcal{E}}_0$ is the (constant) incident field and g denotes the so-called Green's function

$$g(\vec{r}, \vec{r}\,') = -\vec{e}_n \cdot \nabla \frac{1}{|\vec{r} - \vec{r}\,'|} = \vec{e}_n \cdot \frac{\vec{r} - \vec{r}\,'}{|\vec{r} - \vec{r}\,'|^3}.$$

However, we also know that the normal components are related via the dielectric constants, i.e. $\varepsilon_1 \mathcal{E}_n^{out}(\vec{r}) = \varepsilon \mathcal{E}_n^{in}(\vec{r})$. This means that with a bit of rearrangement, Eq. (27.1) can be reformulated as

$$\sigma(\vec{r}) = 2\varepsilon_0 \lambda \vec{e}_n \cdot \vec{\mathcal{E}}_0 + \frac{\lambda}{2\pi} \oint \sigma(\vec{r}\,') g(\vec{r}, \vec{r}\,') dS', \tag{27.2}$$

where $\lambda = (\varepsilon - \varepsilon_1)/(\varepsilon + \varepsilon_1)$. This formulation is very convenient because it allows us to compute the distribution of surface charges from a single equation.

27.1 Cylindrical Nanoparticles

The framework above applies to nanoparticles of completely general shape. It even works for collections of nanoparticles if the surface is taken as the sum of surfaces. In practice, however, Eq. (27.2) is difficult to solve in the general case. Fortunately, many important cases are much simpler. In particular, many relevant nanoparticle geometries have cylindrical symmetry, i.e. they have a rotational symmetry axis as illustrated in Fig. 27.2. In this case, the general problem can be reduced significantly.

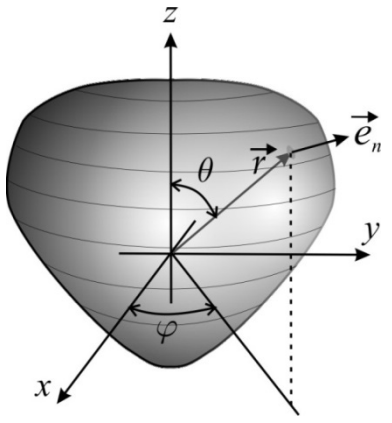


Figure 27.2. Cylindrically symmetric nanoparticle.

Using the geometry of Fig. 27.2, we can without loss of generality choose to keep the incident field $\vec{\mathcal{E}}_0$ in the (x,z)-plane. The x- and z-axes are the horizontal (h) and vertical (v) directions, respectively, and so we decompose $\vec{\mathcal{E}}_0 = \mathcal{E}_0^{(h)}\vec{e}_x + \mathcal{E}_0^{(v)}\vec{e}_z$. Due to the superposition principle we can, in fact, treat the horizontal and vertical cases separately. We then benefit significantly from the simple angular dependence of these cases.

In a cylindrical geometry, we may express the geometrical vectors using polar angles in a simple manner. Hence,

$$\begin{aligned} \vec{r} &= r(\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta) \\ \vec{e}_n &= (n_x\cos\varphi, n_x\sin\varphi, n_z) \\ \vec{r}' &= r'(\sin\theta'\cos\varphi', \sin\theta'\sin\varphi', \cos\theta') \end{aligned} \quad ,$$

where r, n_x, n_z are all functions of θ and r' is a function of θ'. It follows that $\vec{e}_n \cdot \vec{\mathcal{E}}_0$ is $\mathcal{E}_0 n_x \cos\varphi$ and $\mathcal{E}_0 n_z$ in the horizontal and vertical cases, respectively. It is then readily shown that the surface charge follows exactly the same dependence on the angle φ. Due to the symmetry, the surface area element dS must be independent of φ and we may write $dS' = S(\theta')d\theta' d\varphi'$. We will return to the θ dependence later. We also need the following relations to reduce the Green's function:

$$\begin{aligned} \vec{e}_n \cdot (\vec{r} - \vec{r}') &= n_z(r\cos\theta - r'\cos\theta') + n_x(r\sin\theta - r'\sin\theta'\cos(\varphi - \varphi')) \\ |\vec{r} - \vec{r}'|^2 &= r^2 + r'^2 - 2rr'\cos\theta\cos\theta' - 2rr'\sin\theta\sin\theta'\cos(\varphi - \varphi') \end{aligned} .$$

The simplest case is that of vertical polarization, for which $\vec{e}_n \cdot \vec{\mathcal{E}}_0$ is independent of φ. We wish to prove that the surface charge is a function of θ only and so we write $\sigma(\vec{r}) = \sigma(\theta)$. The charge balance Eq.(27.2) therefore reduces to

$$\text{Vertical:} \quad \sigma(\theta) = 2\varepsilon_0\lambda\mathcal{E}_0 n_z(\theta) + \lambda\int_0^{\pi}\sigma(\theta')G^{(v)}(\theta,\theta')S(\theta')d\theta' , \tag{27.3}$$

where

$$G^{(v)}(\theta,\theta') = \int_0^{2\pi}\frac{\vec{e}_n\cdot(\vec{r}-\vec{r}')}{2\pi|\vec{r}-\vec{r}'|^3}d\varphi' .$$

Performing the integral using the geometrical relations we then find

$$G^{(v)}(\theta,\theta') = \frac{1}{2\pi}\left\{\left[n_z(r\cos\theta - r'\cos\theta') + n_x r\sin\theta\right]F_{0,1}(x,y) - n_x r'\sin\theta' F_{1,1}(x,y)\right\}, \tag{27.4}$$

with $x = r^2 + r'^2 - 2rr'\cos\theta\cos\theta'$, $y = -2rr'\sin\theta\sin\theta'$ and introducing the functions

$$F_{m,n}(x,y) \equiv \int_0^{2\pi} \frac{\cos^m\varphi \, d\varphi}{(x + y\cos\varphi)^{(2n+1)/2}}.$$

The first few of these functions are

$$\begin{aligned}
F_{0,0}(x,y) &= \frac{4}{\sqrt{x+y}} K\left(\frac{2y}{x+y}\right) \\
F_{0,1}(x,y) &= \frac{4}{(x-y)\sqrt{x+y}} E\left(\frac{2y}{x+y}\right) \\
F_{1,0}(x,y) &= \frac{4\sqrt{x+y}}{y} E\left(\frac{2y}{x+y}\right) - \frac{4x}{y\sqrt{x+y}} K\left(\frac{2y}{x+y}\right) \\
F_{1,1}(x,y) &= \frac{4x}{y(y-x)\sqrt{x+y}} E\left(\frac{2y}{x+y}\right) + \frac{4}{y\sqrt{x+y}} K\left(\frac{2y}{x+y}\right) \\
F_{2,1}(x,y) &= \frac{4y^2 - 8x^2}{y^2(y-x)\sqrt{x+y}} E\left(\frac{2y}{x+y}\right) - \frac{8x}{y^2\sqrt{x+y}} K\left(\frac{2y}{x+y}\right)
\end{aligned} \quad ,$$

where K and E are complete elliptic integrals. Higher terms can be generated using $F_{m,n} = (F_{m-1,n-1} - xF_{m-1,n})/y$. The fact that these functions depend only on θ and θ' completes the proof.

For the horizontal case, we proceed in almost complete analogy but now the surface charge is found to follow the $\cos\varphi$ behaviour of $\vec{e}_n \cdot \vec{\mathcal{E}}_0$. Hence, we write $\sigma(\vec{r}) = \sigma(\theta)\cos\varphi$ in this case and using elementary mathematical manipulations (rewriting $\cos\varphi'$ as $\mathrm{Re}\left\{e^{i(\varphi-\varphi')}e^{-i\varphi}\right\}$ and doing the integral before taking the real part) find that

$$\text{Horizontal:} \quad \sigma(\theta) = 2\varepsilon_0\lambda\mathcal{E}_0 n_x(\theta) + \lambda\int_0^{\pi} \sigma(\theta')G^{(h)}(\theta,\theta')S(\theta')d\theta', \tag{27.5}$$

where

$$G^{(h)}(\theta,\theta') = \int_0^{2\pi} \frac{\vec{e}_n \cdot (\vec{r} - \vec{r}')\cos(\varphi - \varphi')}{2\pi\,|\,\vec{r} - \vec{r}'\,|^3} d\varphi'$$

can be expressed as

$$G^{(h)}(\theta,\theta') = \frac{1}{2\pi}\left\{\left[n_z(r\cos\theta - r'\cos\theta') + n_x r\sin\theta\right]F_{1,1}(x,y) - n_x r'\sin\theta' F_{2,1}(x,y)\right\}. \quad (27.6)$$

27.2 Oblate Spheroids

As a relatively simple but still technologically important example we will consider oblate spheroids: "pancake" shaped particles obtained by flattening spheres along one direction. Such a particle is illustrated in Fig. 27.3.

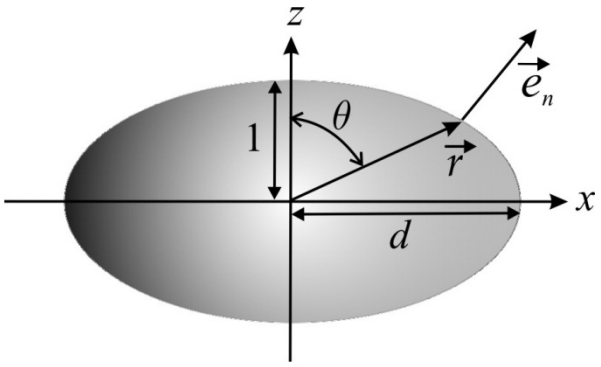


Figure 27.3. Cross section of an oblate spheroid.

The geometry of the spheroid is taken such that the thickness of the "pancake" is unity and the radius is d. Hence, all distances are actually measured in units of half the particle height. A particular point on the surface obeys the ellipse parameterization

$$1 = z^2 + \frac{x^2}{d^2} \quad \Rightarrow \quad 1 = r^2\cos^2\theta + \frac{r^2}{d^2}\sin^2\theta \quad \Rightarrow \quad r = \frac{d}{\sqrt{d^2\cos^2\theta + \sin^2\theta}}.$$

Also, the surface normal is calculated from the requirement that $\vec{e}_n \cdot (d\vec{r}/d\theta) = 0$. Hence, differentiating and ensuring normalization it is found that

$$n_x = \frac{\sin\theta}{\sqrt{d^4\cos^2\theta + \sin^2\theta}}, \quad n_z = \frac{d^2\cos\theta}{\sqrt{d^4\cos^2\theta + \sin^2\theta}}.$$

In general, the surface areal function $S(\theta)$ is to be calculated as $S(\theta) = r^3\sin\theta / \vec{e}_n \cdot \vec{r}$ and in the spheroid case $S(\theta) = d^2\sin\theta(d^4\cos^2\theta + \sin^2\theta)^{1/2} / (d^2\cos^2\theta + \sin^2\theta)^2$.

To solve equations like Eq.(27.3) and Eq.(27.5) numerically we need to discretize the angle θ. On the interval $\theta \in [0,\pi[$ we therefore select N discrete values θ_i with separations $\Delta_i = \theta_{i+1} - \theta_i$ (for i = N we take $\Delta_N = \pi - \theta_i$). Hence, the equations are reformulated as

$$\sigma(\theta_i) \approx 2\varepsilon_0 \lambda \mathcal{E}_0 n(\theta_i) + \lambda \sum_{j=1}^{N} \sigma(\theta_j) G(\theta_i, \theta_j) S(\theta_j) \Delta_j \,, \tag{27.7}$$

where the appropriate Green's function and normal vector component should be chosen for the two polarizations. Equations of this sort are easily converted into a tractable form by introducing vectors

$$\vec{\sigma} = \begin{pmatrix} \sigma(\theta_1) \\ \sigma(\theta_2) \\ \vdots \\ \sigma(\theta_N) \end{pmatrix}, \quad \vec{e} = 2\varepsilon_0 \mathcal{E}_0 \begin{pmatrix} n(\theta_1) \\ n(\theta_2) \\ \vdots \\ n(\theta_N) \end{pmatrix}$$

as well as a matrix

$$\overleftrightarrow{G} = \begin{pmatrix} G(\theta_1,\theta_1)S(\theta_1)\Delta_1 & G(\theta_1,\theta_2)S(\theta_2)\Delta_2 & \cdots & G(\theta_1,\theta_N)S(\theta_N)\Delta_N \\ G(\theta_2,\theta_1)S(\theta_1)\Delta_1 & G(\theta_2,\theta_2)S(\theta_2)\Delta_2 & \cdots & G(\theta_2,\theta_N)S(\theta_N)\Delta_N \\ \vdots & \vdots & \ddots & \vdots \\ G(\theta_N,\theta_1)S(\theta_1)\Delta_1 & G(\theta_N,\theta_2)S(\theta_2)\Delta_2 & \cdots & G(\theta_N,\theta_N)S(\theta_N)\Delta_N \end{pmatrix}.$$

In terms of these quantities, the discretized equation reads as

$$\left\{\lambda^{-1}\overleftrightarrow{U} - \overleftrightarrow{G}\right\} \cdot \vec{\sigma} = \vec{e}, \quad \Rightarrow \quad \vec{\sigma} = \left\{\lambda^{-1}\overleftrightarrow{U} - \overleftrightarrow{G}\right\}^{-1} \cdot \vec{e} \,. \tag{27.8}$$

Thus, the unknown surface charges in vector $\vec{\sigma}$ are found by inverting a matrix and multiplying onto a known vector. Furthermore, it is realized that certain *eigenmodes* of the surface charge can be found whenever the determinant $\left|\lambda^{-1}\overleftrightarrow{U} - \overleftrightarrow{G}\right|$ vanishes. This is because this condition corresponds to a situation, in which a surface charge exists even with a vanishingly small incident field. This is clearly a mathematical abstraction but the significance is that, in actual calculations, resonances in absorption or scattering cross sections may appear near these eigenmodes. From the form of the matrix it is also evident that eigenmodes are found whenever λ^{-1} is an eigenvalue of $\overleftrightarrow{G}$. In practice, the matrix $\overleftrightarrow{G}$ is slightly problematic because the diagonal elements diverge! By clever usage of the general properties of the Green's function, however, appropriate values of the diagonal elements can be found (see the exercise). If the dielectric constant of the metal nanoparticle is assumed to be of the lossless Drude form $\varepsilon(\omega) = \varepsilon_\infty - \omega_p^2 / \omega^2$ the relation between eigenvalue and resonance frequency is given by

$$\omega = \frac{\omega_p}{\sqrt{\varepsilon_\infty + \varepsilon_1 \dfrac{\lambda+1}{\lambda-1}}}.$$

As an example, we now take Ag spheroids ($\varepsilon_\infty = 6$, $\hbar\omega_p = 9.3\,\text{eV}$) embedded in Si ($\varepsilon_1 = 12$). For a particular geometry, the spheroid is characterized by its ellipticity $e = \sqrt{1-1/d^2}$, which ranges between 0 for a sphere and 1 for a plane. In Fig. 27.4, results for the resonance wavelengths in this case are depicted. It is noted that the fundamental horizontal and vertical modes shift towards longer and shorter wavelengths, respectively, as the particle is flattened. In fact, for the dipole mode, simple analytical expressions for vertical and horizontal eigenvalues can be found

$$\lambda_v = \frac{1}{4L-1}, \quad \lambda_h = \frac{1}{1-2L}, \quad L = \frac{1}{2}\left\{1 - \frac{1}{e^2} + \frac{\sqrt{1-e^2}\,\sin^{-1} e}{e^3}\right\}.$$

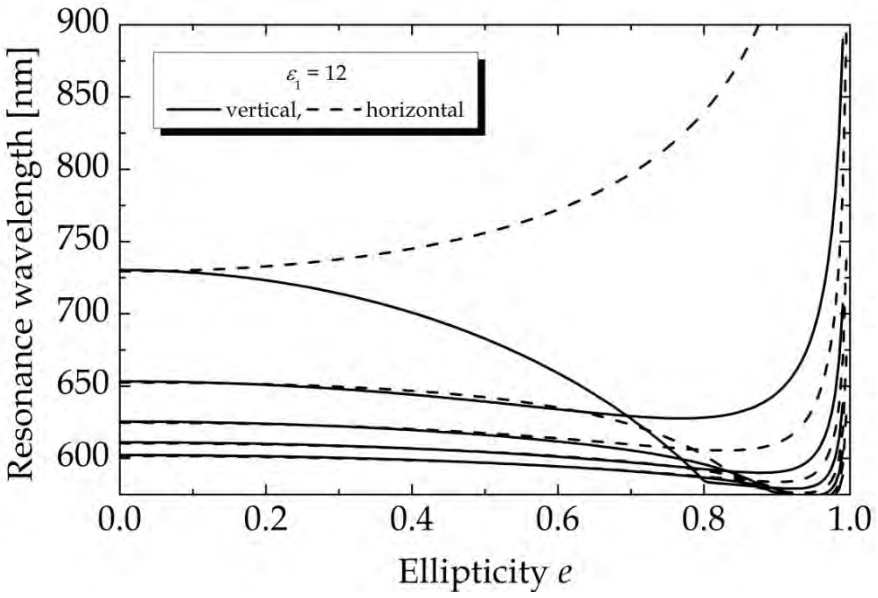


Figure 27.4. Resonance wavelengths of Ag spheroids embedded in Si. Solid and dashed curves illustrate vertical and horizontal resonances versus nanoparticle ellipticity.

The present technique can be extended in several directions. Primarily, based on solution of the inhomogeneous equation Eq.(27.8), the absorption cross section is calculated via

$$\sigma_{abs}(\omega) = \frac{\omega}{c\varepsilon_0\sqrt{\varepsilon_1}}\,\text{Im}\left\{\frac{p}{\mathcal{E}_0}\right\}, \tag{27.9}$$

where p is the induced dipole moment that is easily calculated using the formulas

$$\text{Vertical:}\quad \vec{p} = 2\pi\varepsilon_1 \vec{e}_z \int_0^{\pi} \sigma(\theta) \frac{r^4(\theta)}{\vec{e}_n \cdot \vec{r}} \cos\theta \sin\theta d\theta$$

$$\text{Horizontal:}\quad \vec{p} = \pi\varepsilon_1 \vec{e}_x \int_0^{\pi} \sigma(\theta) \frac{r^4(\theta)}{\vec{e}_n \cdot \vec{r}} \sin^2\theta d\theta$$

.

In addition, nanoparticles positioned on a surface or embedded in thin layers can be handled by proper modifications to the Green's functions [2].

Exercise: Properties of the Green's function

a) Using Gauss' theorem, show that (remembering that the singularity of *g* lies on the boundary and therefore only contributes half)

$$\oint g(\vec{r}, \vec{r}') dS = 2\pi .$$

b) Based on this result, show that the reduced Green's function for vertical polarization satisfies the condition (note that integration over φ' and φ is equivalent)

$$\int_0^{\pi} G^{(v)}(\theta, \theta') S(\theta) d\theta = 1 .$$

This result can be used to handle the singularity of the Green's function. In a discretized version, it reads as

$$\sum_{i=1}^{N} G^{(v)}(\theta_i, \theta_j) S(\theta_i) \Delta_i = 1 .$$

It follows that the diagonal element must be

$$G^{(v)}(\theta_j, \theta_j) = \frac{1}{S(\theta_j)\Delta_j} \left\{ 1 - \sum_{i=1, i\neq j}^{N} G^{(v)}(\theta_i, \theta_j) S(\theta_i) \Delta_i \right\} .$$

For the horizontal case, unfortunately, no such simple result applies. Hence, special numerical handling of the diagonal terms is needed.

References

[1] I.D. Mayergoyz, D.R. Fredkin, and Z. Zhang, Phys. Rev. B72, 155412 (2005).
[2] J. Jung, T.G. Pedersen, T. Søndergaard, K. Pedersen, A. Nylandsted Larsen, and B. Bech Nielsen, Phys. Rev. 81, 125413 (2010).

28. Cylindrical Waveguides

In this chapter, we look at cylindrical waveguides as transmission lines using both classical impedance analysis and a full wave equation approach. A transmission line is a cable transmitting electrical signals. An example would be a coaxial cable such as the one shown in the figure below. The outer conductor is grounded and the signal propagates in the core. In communication networks, electrical pulses carry the information. The propagation of such pulses, however, can be understood by studying signals with a definite frequency using Fourier analysis. We will investigate signals propagating along the cable assuming one end to be driven by a sinusoidal (i.e. single-frequency or monochromatic) signal generator. In this case, the voltage along the cable will also vary sinusoidally in time. The amplitude, however, will depend on the properties of the cable. First, a classical impedance or circuit analysis will be applied to tackle the problem.

28.1 Impedance Analysis

Generally, a homogeneous cable is described by a characteristic frequency-dependent impedance, which is determined by (1) ohmic losses, (2) capacitance between inner and outer conductor, and (3) inductance due to magnetic effects. We will consider a small piece of wire with a length Δx having a resistance ΔR, capacitance ΔC and inductance ΔL. All of these are proportional to Δx, i.e. $\Delta R = R\Delta x$, $\Delta C = C\Delta x$, and $\Delta L = L\Delta x$ with R, C and L the respective quantities per unit length. The insulator between the two conductors is assumed lossless. The model for the wire piece is shown in the figure. You might wonder why the capacitance is added to the right of resistance and inductance instead of left. The answer is that it doesn't matter as long as we consider an infinitesimal piece, i.e. we take the limit $\Delta x \to 0$.

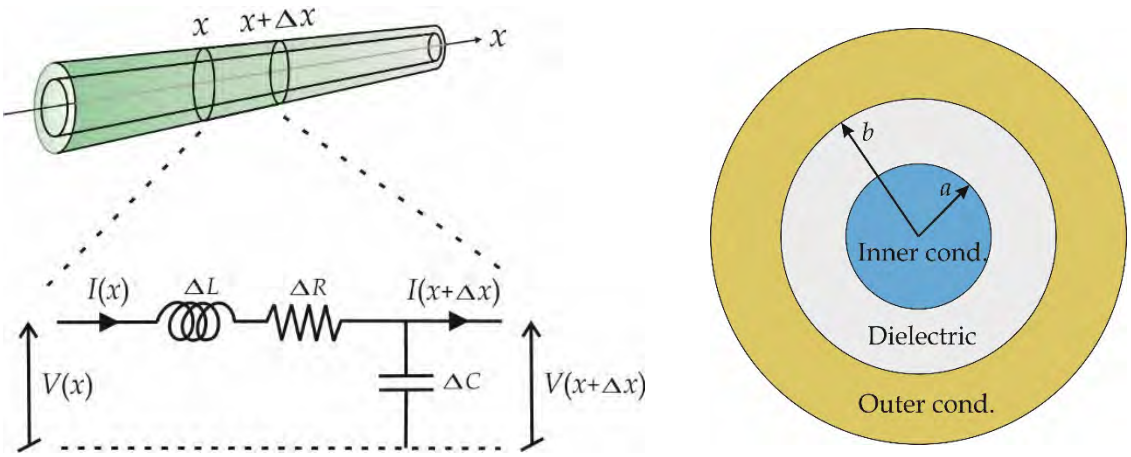


Figure 28.1. Coax cable with circuit representing a tiny piece of length Δx. Symbols ΔL, ΔR, and ΔC indicate inductance, resistance, and capacitance of the piece, respectively. The sketch to the right is a cross section of the cable.

We want to construct a model describing the variation of the voltage $V(x,t)$ along the wire. Both voltage and current naturally also depend on time *t*. To construct a model, we first note that the current running through the capacitor to ground is given by $I_C = \Delta C \times \partial V(x+\Delta x,t)/\partial t$. Here, however, ΔC is already an infinitesimal and, so, if we Taylor expand the voltage, we find $I_C \approx \Delta C \times \partial V(x,t)/\partial t$ to first order in Δx. This means that the current balance (Kirchhoff's law) is

$$I(x+\Delta x,t) = I(x,t) - \Delta C \frac{\partial V(x,t)}{\partial t}.$$

Similarly, the voltage drop along the wire is found from

$$V(x+\Delta x,t) = V(x,t) - \Delta R \times I(x,t) - \Delta L \frac{\partial I(x,t)}{\partial t}.$$

Now, dividing through with Δx and taking the limit $\Delta x \to 0$, we find

$$\frac{\partial I(x,t)}{\partial x} = -C \frac{\partial V(x,t)}{\partial t}, \tag{28.1}$$

and

$$\frac{\partial V(x,t)}{\partial x} = -RI(x,t) - L \frac{\partial I(x,t)}{\partial t}. \tag{28.2}$$

Differentiating Eq.(28.2) once more and applying Eq.(28.1), it is readily shown that the *telegrapher's equation* is

$$\frac{\partial^2 V(x,t)}{\partial x^2} = RC \frac{\partial V(x,t)}{\partial t} + LC \frac{\partial^2 V(x,t)}{\partial t^2}. \tag{28.3}$$

This is the equation we wish to solve subject to the boundary condition $V(0,t) = V_0 \cos \omega t$. Physically, this means that the monochromatic voltage source is placed at $x = 0$ and we take the wire to lie along the positive *x*-axis. In the lossless case $R = 0$, finding the solution is a simple matter. With losses, things are greatly simplified if we use complex analysis. Hence, we writing $V(x,t) = V_0 e^{i(kx-\omega t)}$, with the understanding that only the real part is retained in the end, it is easily shown that

$$-k^2 V_0 e^{i(kx-\omega t)} = -i\omega RC V_0 e^{i(kx-\omega t)} - \omega^2 LC V_0 e^{i(kx-\omega t)}.$$

This obviously means that

$$k = \sqrt{\omega^2 LC + i\omega RC}. \tag{28.4}$$

Here, it is understood that if the signal propagates in the positive x-direction, the sign of the square-root should be chosen so that $\mathrm{Im}\,k > 0$ to ensure an exponentially damped (rather than growing) wave. Splitting according to $k = k_R + ik_I$, we have

$$V(x,t) = \mathrm{Re}\{V_0 e^{i(kx-\omega t)}\} = V_0 e^{-k_I x}\cos(k_R x - \omega t)\,.$$

This describes a monochromatic wave that is damped during propagation along the wire. Expanding the complex square-root, the damping constant k_I is given by

$$k_I = \frac{1}{\sqrt{2}}\left\{\omega\sqrt{\omega^2(LC)^2+(RC)^2}-\omega^2 LC\right\}^{1/2} = \omega\sqrt{\frac{LC}{2}}\left\{\sqrt{1+\left(\frac{R}{\omega L}\right)^2}-1\right\}^{1/2} \approx \frac{R}{2}\sqrt{\frac{C}{L}}\,,$$

where the last approximate result is valid at very high frequencies $\omega \gg R/L$. Not surprisingly, the damping is proportional to the resistance. In the same limit, $k_R \approx \omega\sqrt{LC}$ and the phase velocity of the signal is $v_{phase} = \omega/k_R \approx 1/\sqrt{LC}$.

The current follows from Eq.(28.1) and, similarly to the voltage, we write $I(x,t) = \mathrm{Re}\{I_0 e^{i(kx-\omega t)}\}$ and find $I_0 = (C\omega/k)V_0$. The ratio between the *complex* voltage and current is the impedance

$$Z(\omega) \equiv \frac{V_0}{I_0} = \sqrt{\frac{L}{C}+\frac{iR}{\omega C}}\,.$$

Now, a coax cable consists of an inner conductor with radius a and an outer cylindrical conductor with inner radius b. If the relative dielectric constant of the (non-magnetic) insulator is ε, the capacitance and inductance are

$$C = \frac{2\pi\varepsilon\varepsilon_0}{\ln(b/a)}, \quad L = \frac{\mu_0}{2\pi}\ln(b/a)\,.$$

Hence, phase velocity is $v_{phase} = c/\sqrt{\varepsilon}$ with $c = 1/\sqrt{\varepsilon_0\mu_0}$ the speed of light. Also, the infinite-frequency (or "characteristic") impedance Z_0 is

$$Z_0 = \sqrt{\frac{L}{C}} = \sqrt{\frac{\mu_0}{\varepsilon\varepsilon_0}}\,\frac{\ln(b/a)}{2\pi} \approx 60\ \Omega \times \frac{\ln(b/a)}{\sqrt{\varepsilon}}\,.$$

A standard "50 Ω" coax cable has $b/a \approx 3.5$ and insulation made of polyethylene ($\varepsilon = 2.25$) leading to $Z_0 \approx 50\ \Omega$, in case you wondered. A more realistic calculation requires the resistance R that is the series connection between inner and outer

conductors $R = R_{inner} + R_{outer}$. This, in fact, is a tricky point because AC currents are not uniformly distributed across the wire at high frequency. The current actually tends to run on the surface of the conductor and, hence, the effective area of, e.g., the inner conductor is much less than the geometric area πa^2. At low frequency, however, we may approximate $R_{inner} \approx \rho/(\pi a^2)$, where $\rho \approx 1.7\times10^{-8}$ Ωm is the resistivity of Cu. Hence, for a $a = 1/2$ mm cable, $R_{inner} \approx 22$ mΩ/m. For simplicity, we will simply take $R = 2R_{inner}$. In the presence of losses, the impedance becomes complex and we may split into modulus and phase according to $Z(\omega) = |Z(\omega)|e^{i\varphi(\omega)}$. In the plots below, we consider a 50 Ω coax cable with $a = 1/2$ mm.

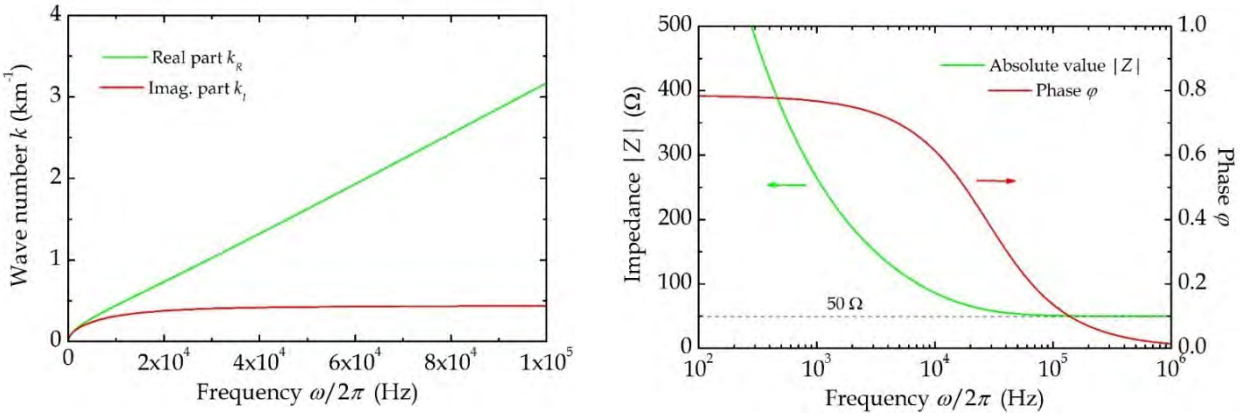


Figure 28.2. Real and imaginary parts of the wave number (left) as well as modulus and phase of the complex impedance (right) for a 50 Ω coax cable.

If two cables having different impedances are joined at $x = 0$, it turns out that a certain part of the electric pulse is reflected at the junction between the two cables. The voltage at the incoming side is $V_0(e^{ikx} + re^{-ikx})e^{-i\omega t}$ and after passing the junction $V_0(1+r)e^{iqx}e^{-i\omega t}$, where q is the wave number on the receiving side and the amplitude has been found by requiring continuity. If the impedances are Z_1 and Z_2, the currents on the two sides, given by Eq.(28.1), are $V_0/Z_1(e^{ikx} - re^{-ikx})e^{-i\omega t}$ and $V_0/Z_2(1+r)e^{iqx}e^{-i\omega t}$, respectively. Demanding continuity of the current as well shows that the reflection coefficient is $r = (Z_2 - Z_1)/(Z_2 + Z_1)$. Writing, again, $r = |r|e^{i\varphi}$ the modulus $|r|$ is the ratio between reflected and incident electric field amplitudes and φ is the reflection phase. As an example, we assume the cable carrying the incident pulse to be the same as above with resistance $R_1 \approx 44$ mΩ/m and that the other cable is similar but has twice the resistance i.e. $R_2 \approx 88$ mΩ/m. This leads to the amplitude reflection and phase shown in Fig. 28.3. At low frequency $r = (\sqrt{R_2} - \sqrt{R_1})/(\sqrt{R_2} + \sqrt{R_1})$ and at high frequency $\varphi = \pi/2$ but $|r| \approx 0$.

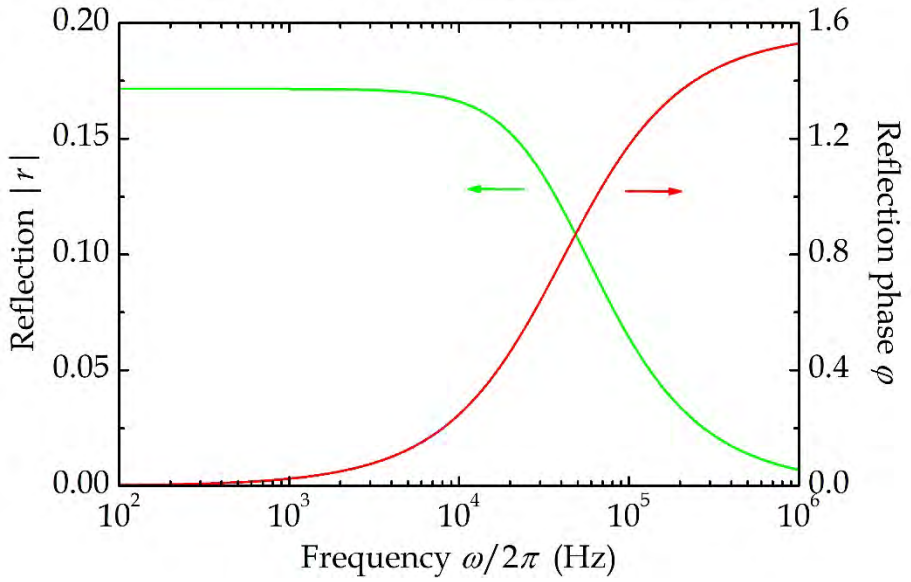


Figure 28.3. Modulus and phase of the complex reflection for a junction between 50 Ω coax cables with resistances $44\ \mathrm{m\Omega/m}$ and $88\ \mathrm{m\Omega/m}$.

28.2 Wave Equation Analysis

We now want to study the full solution for the electric field based on the wave equation. In cylindrical coordinates, the equation for a radially symmetric field of the form $\vec{E} = \{E_r(r)\vec{e}_r + E_x(r)\vec{e}_x\}e^{ikx}$ is

$$\frac{1}{r}\frac{d}{dr}r\frac{dE_r(r)}{dr} + \{\varepsilon(r)k_0^2 - k^2\}E_r(r) = 0, \quad \frac{1}{r}\frac{d}{dr}r\frac{dE_x(r)}{dr} + \{\varepsilon(r)k_0^2 - k^2\}E_x(r) = 0$$

This is supplemented by the divergence condition $d\{rE_r(r)\}/dr + ikrE_x(r) = 0$. Across an interface, $E_x(r)$ and $\varepsilon(r)E_r(r)$ are continuous. If follows that $d\{rE_r(r)\}/dr$ is also continuous. For simplicity, we take the outer conductor to be perfectly conducting. Hence, the field vanishes in this material and we have with $k_i = (\varepsilon_i k_0^2 - k^2)^{1/2}$

$$E_x(r) = \begin{cases} AJ_0(k_1 r) & r < a \\ B\left[J_0(k_2 r)Y_0(k_2 b) - Y_0(k_2 r)J_0(k_2 b)\right]' & a < r < b \end{cases}$$

Writing the field in the dielectric in this particular form ensures a vanishing value on the outer conductor. We require continuity of $E_x(r)$ and $\varepsilon(r)/[\varepsilon(r)k_0^2 - k^2]dE_x(r)/dr$. In turn, this means

$$\frac{\varepsilon_1 k_2}{\varepsilon_2 k_1}\frac{J_1(k_1 a)}{J_0(k_1 a)} = \frac{J_1(k_2 a)Y_0(k_2 b) - Y_1(k_2 a)J_0(k_2 b)}{J_0(k_2 a)Y_0(k_2 b) - Y_0(k_2 a)J_0(k_2 b)}. \tag{28.5}$$

Expanding to first order in k_i^2, as is allowed at low frequency, we find

$$k^2 \approx \varepsilon_2 k_0^2 - \frac{4\varepsilon_2}{\varepsilon_2(b^2 - a^2) + 2a^2(\varepsilon_1 - \varepsilon_2)\ln b/a}.$$

Because $|\varepsilon_1| \gg |\varepsilon_2|$ we may approximate

$$k^2 \approx \varepsilon_2 k_0^2 - \frac{2\varepsilon_2}{a^2 \varepsilon_1 \ln b/a}.$$

Now, for a metal $\varepsilon(\omega) = 1 + i\sigma/(\varepsilon_0\omega) \approx i/(R\pi a^2 \varepsilon_0 \omega)$. The impedance analysis Eq.(28.4) above showed that

$$k^2 = \varepsilon k_0^2 + i\omega R \frac{2\pi\varepsilon\varepsilon_0}{\ln b/a}.$$

Hence, the two models agree in this limit. In the plot below, you find a comparison between losses, i.e. k_I, computed from Eq.(28.4) and the numerical solution of Eq.(28.5). One notes that an increased loss at high frequency is seen in the full solution but excellent agreement is found at low ω. In Fig. 28.5, we plot the radial behavior of the field. It is seen that at high frequency, the field is strongly varying inside the conductor. This, in fact, is the main reason for the failure of the impedance analysis in this case.

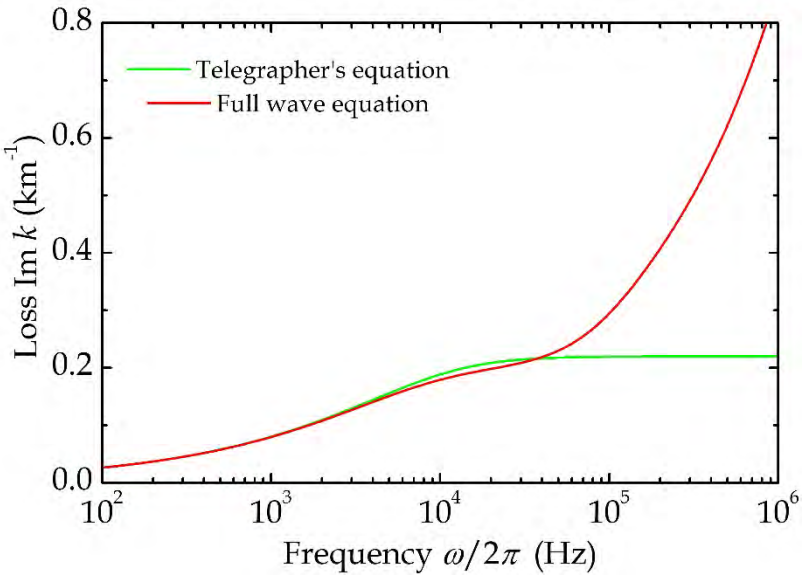


Figure 28.4. Comparison between losses based on telegrapher's equation and the full wave equation.

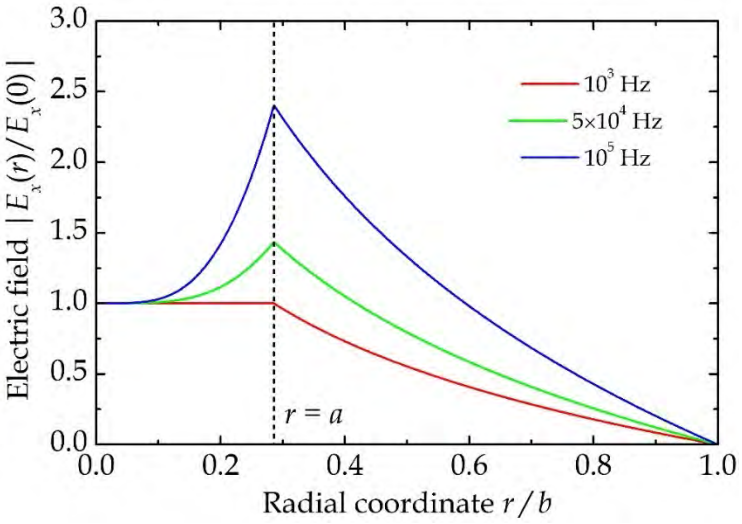


Figure 28.5. Radial dependence of the longitudinal electric field normalized by the center value.

Exercise: High-frequency behavior

As apparent from Fig. 28.5, the field behaves as if it is pushed to the surface. Hence, the inner parts are poorly "utilized" in terms of spreading out the current. One may say that a radial segment is "filled" by a fraction $\mathrm{Re}\, J_0(k_1 r)/\mathrm{Re}\, J_0(k_1 a) \approx \exp(k_I(r-a))$, where the approximation is valid whenever $|k_1 r| \gg 1$. Averaging this quantity over the entire area produces a measure of the effective area. Hence,

$$A_{eff} = 2\pi \int_0^a e^{k_I(r-a)} r dr \approx \frac{2\pi a}{k_I}.$$

Comparing to the geometric area πa^2, we see that the current is effectively confined to a thin annulus of area $2\pi a\delta$, where $\delta = 1/k_I$ is the "skin depth". If we use $\varepsilon(\omega) \approx i/(\varepsilon_0 \rho\omega)$ and $k_1 \approx \sqrt{\varepsilon(\omega)} k_0$ it follows that $A_{eff} \approx 2\pi a (2\varepsilon_0 \rho\omega)^{1/2}/k_0$. In turn, the effective resistance becomes

$$R_{eff} = \rho / A_{eff} \approx \left(\frac{\rho}{2\varepsilon_0 \omega}\right)^{1/2} \frac{k_0}{2\pi a} = \frac{\left(\frac{1}{2}\mu_0 \rho\omega\right)^{1/2}}{2\pi a}.$$

a) Use this expression instead of R in Eq.(28.4) and compare with the full numerical solution for the loss versus frequency. Also, try an improved interpolation $R_{int} = (R^2 + R_{eff}^2)^{1/2}$. The results are shown in Fig. 28.6 below.

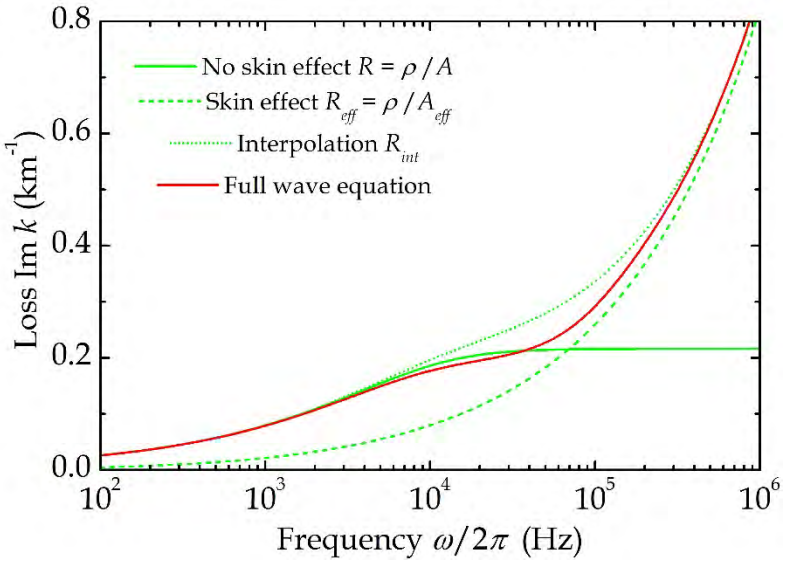


Figure 28.6. Losses based on telegrapher's equation with geometric area (green solid), effective area (green dashed), and interpolated area (green dotted) in comparison to the full wave equation solution (red).

29. Graphene Plasmonics

In this chapter, we will apply the optical characteristics of graphene, derived in Chapter 17, to the specific goal of understanding plasmons in graphene. Graphene is special in the sense that the extremely low thickness means that charges and currents are essentially confined to an infinitely thin layer. We will begin by investigating appropriate boundary conditions for a dielectric interface containing the graphene sheet.

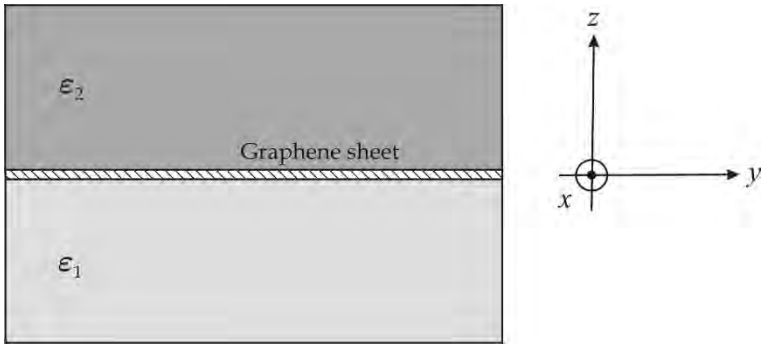


Figure 29.1. A graphene sheet sandwiched between two dielectric media.

We treat a planar interface between media with dielectric constants ε_1 and ε_2, respectively, as illustrated in Fig. 29.1. The presence of the graphene sheet is then modelled as the addition of a sheet current $\vec{j}$ and charges ρ located at the interface. In such a case, the standard Maxwell boundary conditions are modified and read

$$\begin{aligned} &\mathcal{E}_t^{(2)} - \mathcal{E}_t^{(1)} = 0, \quad \varepsilon_2 \mathcal{E}_n^{(2)} - \varepsilon_1 \mathcal{E}_n^{(1)} = \rho / \varepsilon_0 \\ &\vec{e}_z \times (\vec{\mathcal{B}}^{(2)} - \vec{\mathcal{B}}^{(1)}) = \mu_0 \vec{j}, \quad \mathcal{B}_n^{(2)} - \mathcal{B}_n^{(1)} = 0. \end{aligned} \tag{29.1}$$

Here, subscripts *t* and *n* indicate tangential and normal components. Note that the sheet quantities are themselves related via the continuity equation $\nabla \cdot \vec{j} = i\omega\rho$. We first assume that only a single plane wave exists on either side of the interface. The *z*-direction is perpendicular to the interface and the (*x*,*z*) plane is taken as the plane of incidence. The parallel component of the wave vector k_x must be identical for all waves in order to fulfil boundary conditions. The *z*-components are different, however. We denote these wave vector *z*-components by q_i for the *i*'th side. They are related to k_x via $k_x^2 + q_i^2 = \varepsilon_i k_0^2$ with $k_0 = \omega / c$. The surrounding dielectric constants are assumed real and positive. The sheet current itself is $j = \sigma \mathcal{E}_t$ and we now analyze the cases of *s*- and *p*-polarization, for which the tangential electric field is along *y* and *x*, respectively.

29.1 S-polarization

In the *s*-polarized case, we have

$$\vec{\mathcal{E}}^{(i)} = \mathcal{E}^{(i)}\,\vec{e}_y e^{i(k_x x+q_i z)},\quad \vec{\mathcal{B}}^{(i)} = \frac{1}{i\omega}\nabla\times\vec{\mathcal{E}}^{(i)} = \frac{\mathcal{E}^{(i)}}{\omega}\left(-q_i\vec{e}_x + k_x\vec{e}_z\right)e^{i(k_x x+q_i z)}.$$

Here, the boundary conditions imply

$$\mathcal{E}^{(2)} - \mathcal{E}^{(1)} = 0,\quad -q_2\mathcal{E}^{(2)} + q_1\mathcal{E}^{(1)} = \mu_0\omega j.$$

The sheet current is $j = \sigma\mathcal{E}^{(1)} = \sigma\mathcal{E}^{(2)}$ and so the dispersion relation is simply

$$q_1 - q_2 - \mu_0\omega\sigma = 0. \tag{29.2}$$

A similar analysis applies to the computation of reflectances. In this case, the field on one side consists of incident $\mathcal{E}_{inc}$ and reflected $\mathcal{E}_{ref}$ components whereas a transmitted part $\mathcal{E}_{trans}$ is found on the other side. Hence, taking the field to be incident from the $i = 1$ side

$$\vec{\mathcal{E}}^{(1)} = \vec{e}_y\left(\mathcal{E}_{inc}e^{iq_1 z} + \mathcal{E}_{ref}e^{-iq_1 z}\right)e^{ik_x x},\quad \vec{\mathcal{E}}^{(2)} = \vec{e}_y\mathcal{E}_{trans}e^{iq_2 z}e^{ik_x x},$$
$$\vec{\mathcal{B}}^{(1)} = \frac{1}{\omega}\left(-q_1\left[\mathcal{E}_{inc}e^{iq_1 z} - \mathcal{E}_{ref}e^{-iq_1 z}\right]\vec{e}_x + k_x\left[\mathcal{E}_{inc}e^{iq_1 z} + \mathcal{E}_{ref}e^{-iq_1 z}\right]\vec{e}_z\right)e^{ik_x x},$$
$$\vec{\mathcal{B}}^{(2)} = \frac{\mathcal{E}_{trans}}{\omega}\left(-q_2\vec{e}_x + k_x\vec{e}_z\right)e^{iq_2 z}e^{ik_x x}.$$

The boundary conditions then lead to the following field relations

$$\mathcal{E}_{trans} - \mathcal{E}_{inc} - \mathcal{E}_{ref} = 0,\quad -q_2\mathcal{E}_{trans} + q_1\left(\mathcal{E}_{inc} - \mathcal{E}_{ref}\right) = \mu_0\omega\sigma\mathcal{E}_{trans}.$$

The reflection and transmission coefficients are, respectively, $r = \mathcal{E}_{ref}/\mathcal{E}_{inc}$ and $t = \mathcal{E}_{trans}/\mathcal{E}_{inc}$ and solving, we find

$$r_s = \frac{q_1 - q_2 - \mu_0\omega\sigma}{q_1 + q_2 + \mu_0\omega\sigma}. \tag{29.3}$$

Note that the plasmon dispersion relation Eq.(29.2) follows by taking $r_s = 0$. This is no coincidence, as a localized mode is precisely one that propagates without radiating into the incident direction. In the ideal case of lossless graphene sheets sufficiently doped that the intraband response dominates, we have $\sigma = i\varepsilon_0\omega_p^2/\omega$. To fulfil Eq.(29.2), it is then clear that both *z*-components of the wave vector must be

purely imaginary, i.e. for both sides we need $q_i > \sqrt{\varepsilon_i} k_0$. However, we are looking for field solutions that are localized to the interface. Hence, we require $q_1 = -i\kappa_1$ and $q_2 = i\kappa_2$, where $\kappa_i = (k_x^2 - \varepsilon_i k_0^2)^{1/2}$ must be positive on both sides. We then find

$$\kappa_1 + \kappa_2 - i\mu_0\omega\sigma = 0 \quad \Rightarrow \quad (k_x^2 - \varepsilon_1 k_0^2)^{1/2} + (k_x^2 - \varepsilon_2 k_0^2)^{1/2} = -\omega_p^2 / c^2 . \quad (29.4)$$

This equation clearly has no solutions as the signs on either side differ.

29.2 P-polarization

For *p*-polarization, we have instead

$$\vec{\mathcal{B}}^{(i)} = \mathcal{B}^{(i)}\, \vec{e}_y e^{i(k_x x + q_i z)}, \quad \vec{\mathcal{E}}^{(i)} = \frac{ic^2}{\varepsilon_i \omega} \nabla \times \vec{\mathcal{B}}^{(i)} = \frac{c^2}{\varepsilon_i \omega} \mathcal{B}^{(i)} \left(q_i \vec{e}_x - k_x \vec{e}_z \right) e^{i(k_x x + q_i z)} .$$

In this case, we find

$$\frac{q_2}{\varepsilon_2} \mathcal{B}^{(2)} - \frac{q_1}{\varepsilon_1} \mathcal{B}^{(1)} = 0, \quad -\mathcal{B}^{(2)} + \mathcal{B}^{(1)} = \mu_0 j .$$

We need the current $j = \sigma \mathcal{E}_t = \sigma c^2 q_i \mathcal{B}^{(i)} / (\varepsilon_i \omega)$ that can be evaluated for i = 1 or 2 with identical results. Hence, a non-trivial solution implies

$$\frac{\varepsilon_1}{\kappa_1} + \frac{\varepsilon_2}{\kappa_2} + \frac{i\sigma}{\varepsilon_0 \omega} = 0 \quad \Rightarrow \quad \frac{\varepsilon_1}{(k_x^2 - \varepsilon_1 k_0^2)^{1/2}} + \frac{\varepsilon_2}{(k_x^2 - \varepsilon_2 k_0^2)^{1/2}} = \frac{\omega_p^2}{\omega^2} . \quad (29.5)$$

Similarly, solving for the reflection coefficient we find

$$r_p = \frac{\varepsilon_2 q_1 - \varepsilon_1 q_2 + q_1 q_2 \sigma / (\varepsilon_0 \omega)}{\varepsilon_2 q_1 + \varepsilon_1 q_2 + q_1 q_2 \sigma / (\varepsilon_0 \omega)} .$$

Again, the plasmon dispersion follows from setting $r_p = 0$. The dispersion relation Eq.(29.5) for *p*-polarized plasmons can be converted into a polynomial equation. However, the generally complicated solution is greatly simplified in the quasi-static limit $k_x \gg k_0$. In this limit, it is readily seen that $k_x \approx (\varepsilon_1 + \varepsilon_2)\omega^2 / \omega_p^2$. In highly doped graphene, the plasma frequency is approximately $\omega_p^2 = e^2 E_F / (\pi \varepsilon_0 \hbar^2) = 4\alpha c \omega_F$, where $\omega_F = E_F / \hbar$ and α is the fine-structure constant, c.f. chapter 17. It then follows that

$$k_x \approx k_0 \left(\frac{\varepsilon_1 + \varepsilon_2}{4\alpha} \right) \frac{\omega}{\omega_F} .$$

Thus, writing $k_0 = 2\pi/\lambda_0$ and introducing the surface plasmon wavelength via $k_x = 2\pi/\lambda_{sp}$ we have

$$\lambda_{sp} \approx \lambda_0 \left(\frac{4\alpha}{\varepsilon_1 + \varepsilon_2} \right) \frac{\omega_F}{\omega}.$$

In the experiments of Ref. [1], parameters were $\hbar\omega \approx 0.13$ eV, $E_F \approx 0.4$ eV, $\varepsilon_1 = 1$ and $\varepsilon_2 \approx 1.9$. Thus, we expect $\lambda_{sp}/\lambda_0 \approx 1/32$ in agreement with the experiments. The full numerical solution is shown in the inset below. Another way of illustrating the dispersion relation is by plotting ω/ω_F versus k_x/k_F, with $k_F = \omega_F/v_F$ the Fermi wave vector and $v_F \approx c/300$ the Fermi velocity. According to the analysis above, we expect a square-root in the electro-static limit

$$\frac{\omega}{\omega_F} \approx \left(\frac{4\alpha c}{(\varepsilon_1 + \varepsilon_2) v_F} \cdot \frac{k_x}{k_F} \right)^{1/2}. \tag{29.6}$$

In the main plot below, obtained from the full numerical solution, we see that this expectation is borne out.

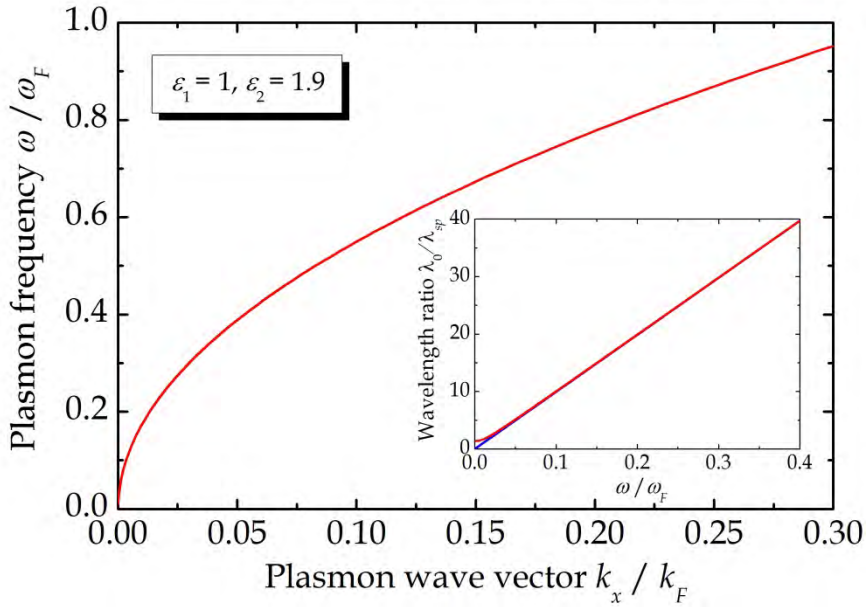


Figure 29.2. Plasmon dispersion relation. The inset shows a comparison of surface plasmon wave lengths for the full (red) and quasi-static (blue) solution.

29.3 Graphene Disk Plasmons

We now specialize to an area of increasing interest in graphene physics: plasmons in nanostructured graphene. In particular, we focus on plasmons in circular graphene disks such as the one shown in Fig. 29.3.

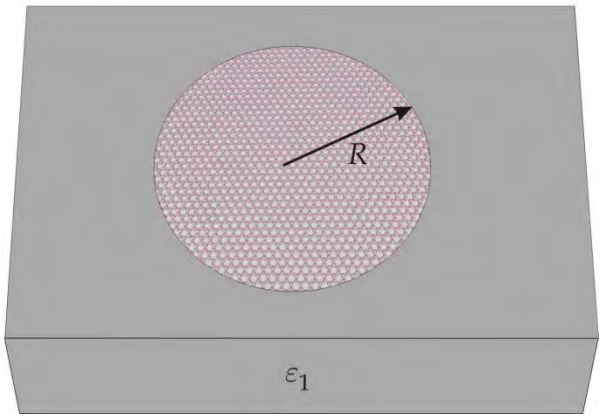


Figure 29.3. Circular graphene disk on a dielectric substrate.

We will apply the electro-static approximation similarly to the case of small metallic particles in chapter 27. The disk is characterized by a sheet conductivity $\sigma\chi(\vec{r})$, where χ is the characteristic function of the disk, i.e. χ equals unity inside the disk and zero outside. As a starting point, the electric potential produced by a sheet charge density $\rho(\vec{r})$ is

$$\phi(\vec{r}) = \phi_0(\vec{r}) + \frac{1}{4\pi\varepsilon\varepsilon_0}\int \frac{\rho(\vec{r}')}{|\vec{r}-\vec{r}'|} d^2r',$$

where ϕ_0 represents the incident field and $\varepsilon = (\varepsilon_1+\varepsilon_2)/2$. Now, from the continuity equation we have $\rho(\vec{r}) = (i\omega)^{-1}\nabla\cdot\vec{j}(\vec{r}) = (i\omega)^{-1}\sigma\nabla\cdot\left\{\chi(\vec{r})\vec{\mathcal{E}}(\vec{r})\right\}$. Moreover, $\vec{\mathcal{E}}(\vec{r}) = -\nabla\phi(\vec{r})$ and so

$$\rho(\vec{r}) = \frac{i\sigma}{\omega}\nabla\cdot\left\{\chi(\vec{r})\nabla\phi(\vec{r})\right\}. \tag{29.7}$$

Next, we take advantage of the circular geometry. This implies that all eigenmodes will be of the form $\phi(\vec{r}) = \phi_l(r)e^{il\varphi}$ and $\rho(\vec{r}) = \rho_l(r)e^{il\varphi}$ for potential and charge, respectively, where l is an integer. This means that we need the integral

$$K_l(r,r') = \frac{1}{2\pi}\int_0^{2\pi} \frac{\cos l\varphi}{\sqrt{r^2+r'^2-2rr'\cos\varphi}} d\varphi.$$

Unfortunately, no simple expression exists for general l. For l = 1, the result is given by $(2\pi)^{-1}$ times $F_{1,0}$ defined in chapter 27. Via this function, we find

$$\phi_l(r) = \phi_{0,l}(r) + \frac{1}{2\varepsilon\varepsilon_0}\int_0^R K_l(r,r')\rho_l(r')r'dr' .$$

The continuity equation Eq.(29.7) can be rewritten

$$\rho_l(r) = \frac{i\sigma}{\omega}\left\{\frac{1}{r}\frac{d}{dr}r\frac{d}{dr} - \frac{l^2}{r^2}\right\}\phi_l(r) .$$

To invert this equation, we need the corresponding Green's function. The appropriate function should satisfy the condition

$$\left\{\frac{1}{r}\frac{d}{dr}r\frac{d}{dr} - \frac{l^2}{r^2}\right\}G_l(r,r') = -\frac{1}{r}\delta(r-r') .$$

In this manner, we can rewrite the equation as

$$\phi_l(r) = \frac{i\omega}{\sigma}\int_0^R G_l(r,r')\rho_l(r')r'dr' .$$

The boundary condition for the Green's function follows from behavior of the field on the edge. Hence, we require the current density to be tangential at the edge of the disk. The normal component of the current j_n is given by $j_n = \sigma\mathcal{E}_n = -\sigma(d\phi/dr)_{r=R}$. Thus, we require $dG_l(r,r')/dr|_{r=R} = 0$. It then follows that with $r_< = \min(r,r')$ and $r_> = \max(r,r')$

$$G_l(r,r') = \frac{1}{2l}\left\{\left(\frac{r_<}{r_>}\right)^l + \left(\frac{rr'}{R^2}\right)^l\right\}.$$

This may be verified by noting that

$$\frac{d}{dr}\left(\frac{r_<}{r_>}\right)^l = l\frac{\text{sign}(r'-r)}{r}\left(\frac{r_<}{r_>}\right)^l, \quad \frac{1}{r}\frac{d}{dr}r\frac{d}{dr}\left(\frac{r_<}{r_>}\right)^l = \frac{l^2}{r^2}\left(\frac{r_<}{r_>}\right)^l - 2l\frac{\delta(r-r')}{r}.$$

The second term in the Green's function is the homogeneous solution added to satisfy the boundary condition $dG_l(r,r')/dr|_{r=R} = 0$, as is readily demonstrated. We finally see that, in the absence of an external field, the condition for a plasmon is

$$\int_0^R G_l(r,r')\rho_l(r')r'dr' = \frac{\sigma}{2i\omega\varepsilon\varepsilon_0}\int_0^R K_l(r,r')\rho_l(r')r'dr' .$$

It is advantageous to introduce normalized radial coordinates $x = r/R$ and $y = r'/R$ so that because $K \propto 1/r$

$$\int_0^1 G_l(x,y)\rho_l(y)ydy = \frac{\sigma}{2i\omega\varepsilon\varepsilon_0 R}\int_0^1 K_l(x,y)\rho_l(y)ydy . \tag{29.8}$$

An elegant way of solving this eigenvalue problem is by expanding in an appropriate basis. A.L. Fetter [2] has shown that by using a Jacobi polynomial basis

$$\rho_l(y) = \sum_{n=0}^{\infty} c_n f_n(y), \quad f_n(y) = y^l P_n^{l,0}(1-2y^2)$$

the required (non zero) matrix elements for $l > 0$ are

$$G_{nm}^l = \int_0^1\int_0^1 G_l(x,y)f_n(x)f_m(y)xydxdy = \begin{cases} (3l+4)/[8l(l+1)^2(l+2)] & n=m=0 \\ 1/[4(l+2n)(l+2n+1)(l+2n+2)] & n=m\neq 0 \\ 1/[8(l+2n+1)(l+2n+2)(l+2n+3)] & n=m-1 \\ 1/[8(l+2m+1)(l+2m+2)(l+2m+3)] & m=n-1 \end{cases}$$

$$K_{nm}^l = \int_0^1\int_0^1 K_l(x,y)f_n(x)f_m(y)xydxdy = (-1)^{m-n+1}/[\pi(4(n-m)^2-1)(l+n+m+\tfrac{1}{2})(l+n+m+\tfrac{3}{2})].$$

The polynomials themselves can be generated using the recursion relation $P_0^{l,0}(z)$ = 1, $P_1^{l,0}(z) = ((l+2)z+l)/2$ and

$$P_n^{l,0}(z) = \frac{(-1+l+2n)(l^2+z(-2+l+2n)(l+2n))}{2n(l+n)(-2+l+2n)}P_{n-1}^{l,0}(z) - \frac{(-1+l+n)(-1+n)(l+2n)}{n(l+n)(-2+l+2n)}P_{n-2}^{l,0}(z).$$

In this manner, Eq.(29.8) is transformed into a generalized eigenvalue problem

$$\sum_m K_{nm}^l c_m = \lambda \sum_m G_{nm}^l c_m$$

with the eigenvalue $\lambda = 2i\omega\varepsilon\varepsilon_0 R/\sigma = 2\varepsilon R\omega^2/\omega_p^2$. Using a basis of 300 polynomials, the eigenvalues are plotted in Fig. 29.4. Of particular importance is the fundamental dipole $(l=1)$ mode, for which the eigenvalue is $\lambda = 1.098$. The associated resonance frequency is given by Eq.(29.6) with wave vector $k = \lambda/R$

$$\frac{\omega}{\omega_F} \approx \left(\frac{4\alpha c}{(\varepsilon_1 + \varepsilon_2) v_F} \cdot \frac{\lambda}{k_F R} \right)^{1/2} . \tag{29.9}$$

For an $R = 100$ nm disk on glass with $E_F \approx 0.4$ eV, $\varepsilon_1 = 1$ and $\varepsilon_2 \approx 1.9$ this corresponds to $\hbar\omega \approx 0.093$ eV.

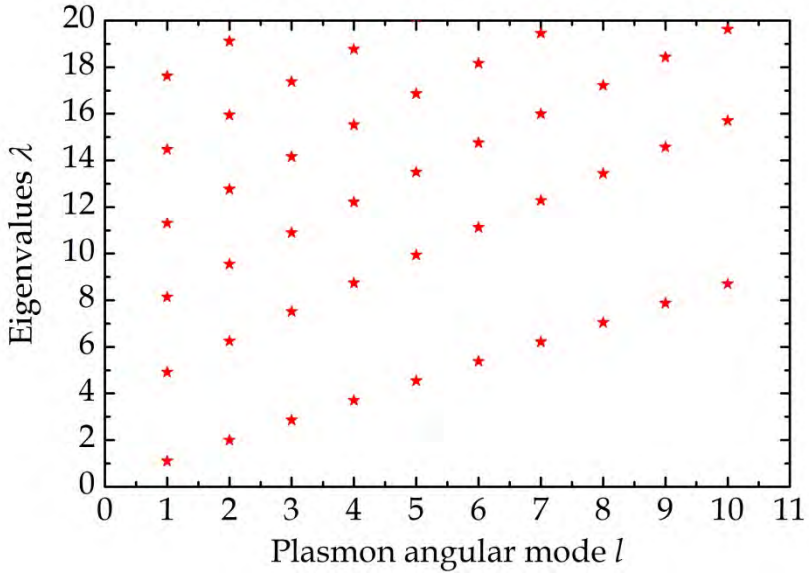


Figure 29.4. Eigenvalues of the graphene disk plasmon problem for a range of angular modes.

29.4 Graphene Nanoribbon Plasmons

Next, a similar treatment will be given to graphene nanoribbons, i.e. infinitely long strips. we take the y-axis along the ribbon and so, placing ourselves at $y = 0$, find

$$\phi(x) = \phi_0(x) + \frac{1}{4\pi\varepsilon\varepsilon_0} \int \frac{\rho(x')}{\sqrt{(x-x')^2 + y'^2}} d^2 r' .$$

To integrate over y', we first consider a finite ribbon of length L so that

$$\int_{-L/2}^{L/2} \frac{1}{\sqrt{(x-x')^2 + y'^2}} dy' = \ln \left(\frac{\sqrt{L^2 + 4(x-x')^2} + L}{\sqrt{L^2 + 4(x-x')^2} - L} \right) \approx 2\ln L - 2\ln|x - x'| .$$

In the last approximation, we used the fact that L is large. Now, since the induced charge is overall neutral, the constant piece $2\ln L$ has a vanishing contribution to the integral. Thus, for a ribbon of width W we find (taking the $L \to \infty$ limit)

$$\phi(x)=\phi_0(x)-\frac{1}{2\pi\varepsilon\varepsilon_0}\int_0^W \rho(x')\ln|x-x'|\,dx'\,.$$

Finally, using again Eq.(29.7) we have

$$\phi(x)=\phi_0(x)+\frac{\sigma}{2i\pi\varepsilon\varepsilon_0\omega}\int_0^W \ln|x-x'|\frac{d}{dx'}\left\{\chi(x')\frac{d}{dx'}\phi(x')\right\}dx'\,.$$

This is the integro-differential equation to solve. Again, in the absence of an incident potential and introducing $\lambda=2i\omega\varepsilon\varepsilon_0 W/\sigma$, we find an eigenvalue problem in normalized coordinates

$$\lambda\phi(x)=\frac{1}{\pi}\int_0^1 \ln|x-x'|\frac{d}{dx'}\left\{\chi(x')\frac{d}{dx'}\phi(x')\right\}dx'\,.$$

Numerically, we solve by discretizing and considering a ribbon that is slightly narrower than unity such that the characteristic function χ vanishes in the outermost four points (two in each side: one outside and one inside the ribbon). Applying a grid of 2502 points and midpoint sampling [3] then leads to eigenvalues $\lambda=\{0.000, 2.316, 5.510, 8.635,...\}$. The associated eigenmodes are shown in Fig. 29.5.

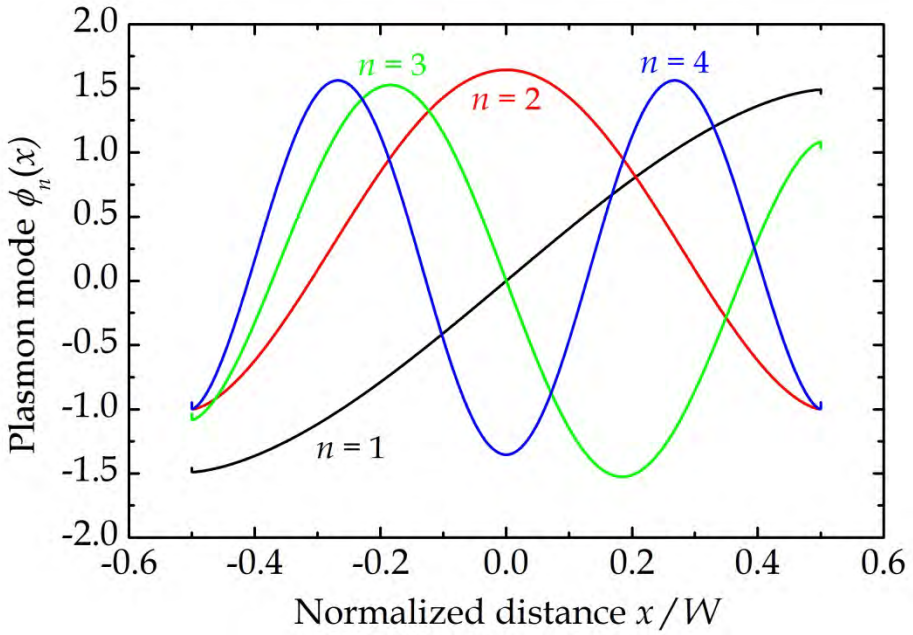


Figure 29.5. Four lowest graphene nanoribbon plasmon eigenmodes.

Exercise: Graphene magnetoplasmons

In this exercise, we return to the plasmon mode problem for the infinite sheet. Hence, we consider the geometry in Fig. 29.1 but in addition to the optical fields we imagine that a static magnetic field is applied perpendicular to the graphene sheet. In the presence of a $\mathcal{B}$- field, the optical conductivity tensor is no longer diagonal but becomes [4]

$$\vec{\vec{\sigma}} = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} \\ -\sigma_{xy} & \sigma_{xx} \end{pmatrix}.$$

In the case of strongly doped graphene, the conductivities in the collisionless limit $\tau \to \infty$ are (see Chapter 37)

$$\sigma_{xx} = \frac{i\varepsilon_0 \omega_p^2 \omega}{\omega^2 - \omega_D^2}, \quad \sigma_{xy} = \frac{\varepsilon_0 \omega_p^2 \omega_D}{\omega^2 - \omega_D^2}.$$

Here, $\omega_D = v_F \sqrt{2e\mathcal{B}/\hbar}$ is the Dirac cyclotron frequency. In this case, the plasmon eigenmodes are no longer purely *s*- or *p*-polarized, but rather a mixture. Hence, a general field is of the form $\vec{\mathcal{E}}^{(i)} = \left\{ \mathcal{E}_x \vec{e}_x + \mathcal{E}_y \vec{e}_y + \mathcal{E}_z^{(i)} \vec{e}_z \right\} e^{i(k_x x + q_i z)}$. Here, we utilized the fact that the tangential field components are the same in the two media. Since the fields have both *x*- and *y*-components we find a sheet current

$$\vec{j} = \left\{ (\sigma_{xx}\mathcal{E}_x + \sigma_{xy}\mathcal{E}_y)\vec{e}_x + (-\sigma_{xy}\mathcal{E}_x + \sigma_{xx}\mathcal{E}_y)\vec{e}_y \right\} e^{ik_x x}.$$

a) Using $\vec{\mathcal{B}}^{(i)} = (i\omega)^{-1} \nabla \times \vec{\mathcal{E}}^{(i)}$ and the fact that $\nabla \cdot \vec{\mathcal{E}}^{(i)} = 0$, show that for $z \neq 0$

$$\vec{e}_z \times \vec{\mathcal{B}}^{(i)} = -\frac{1}{\omega} \left\{ \left(\frac{k_x^2}{q_i} + q_i \right) \mathcal{E}_x \vec{e}_x + q_i \mathcal{E}_y \vec{e}_y \right\} e^{i(k_x x + q_i z)}.$$

b) By combining with the boundary conditions Eq.(29.1), show that

$$\begin{pmatrix} k_x^2/q_1 + q_1 - k_x^2/q_2 - q_2 - \mu_0 \omega \sigma_{xx} & -\mu_0 \omega \sigma_{xy} \\ \mu_0 \omega \sigma_{xy} & q_1 - q_2 - \mu_0 \omega \sigma_{xx} \end{pmatrix} \cdot \begin{pmatrix} \mathcal{E}_x \\ \mathcal{E}_y \end{pmatrix} = 0.$$

As usual, the mode condition is that the determinant vanishes.

c) Show that this eventually implies

$$\left(\frac{\varepsilon_1}{\kappa_1} + \frac{\varepsilon_2}{\kappa_2} + \frac{i\sigma_{xx}}{\varepsilon_0 \omega} \right) \left(\kappa_1 + \kappa_2 - i\mu_0 \omega \sigma_{xx} \right) + \frac{\mu_0}{\varepsilon_0} \sigma_{xy}^2 = 0.$$

Hence, it is obvious that the mode condition is a mixture of the purely *s*-polarized case Eq.(29.4) and the *p*-polarized case Eq.(29.5) and, in the absence of a magnetic field, the two original modes emerge. As before, we apply the quasi-static approximation $\kappa_i \approx k_x$ to solve. In this manner, using the conductivities above, one finds

$$\frac{\omega}{\omega_F} = \left(\frac{4\alpha c}{(\varepsilon_1+\varepsilon_2)v_F} \cdot \frac{k_x}{k_F} + \frac{2\frac{\omega_D^2}{\omega_F^2}\cdot\frac{k_x}{k_F}}{2\frac{k_x}{k_F}+\frac{4\alpha v_F}{c}} \right)^{1/2} \approx \left(\frac{4\alpha c}{(\varepsilon_1+\varepsilon_2)v_F} \cdot \frac{k_x}{k_F} + \frac{\omega_D^2}{\omega_F^2} \right)^{1/2}.$$

The second approximate expression holds because $4\alpha v_F / c \ll 1$. It is clear that the correct limit Eq.(29.6) is found for $\omega_D = 0$. In Fig. 29.6 below, we plot the dispersion relation for three values of the cyclotron frequency. A cyclotron energy of 0.3 eV corresponds to a field of about 45 T, so this is a huge energy. Note that the red curve is identical to the result in Fig. 29.2.

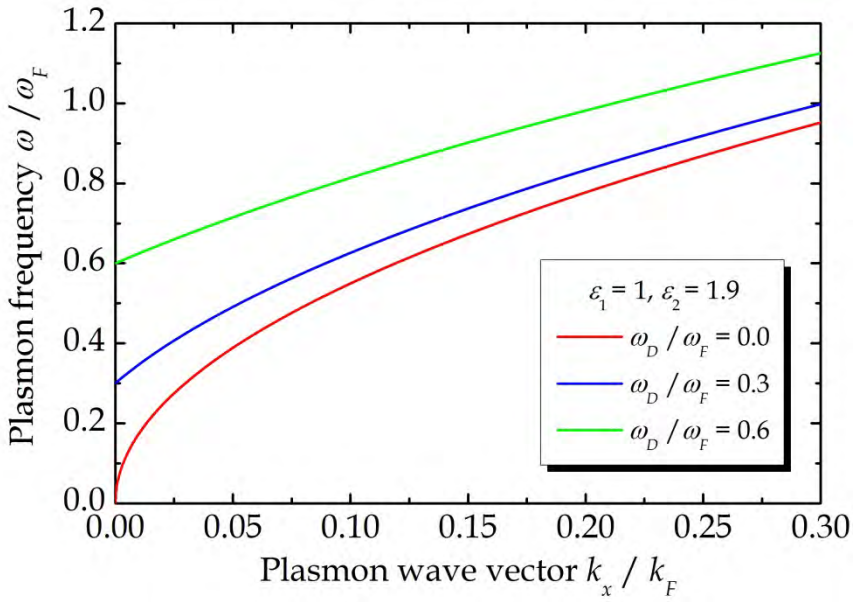


Figure 29.6. Solution of the magnetoplasmon dispersion relation for different values of the cyclotron energy.

References

[1] J. Chen *et al.*, Nature 487, 77 (2012).
[2] A.L. Fetter, Phys. Rev. B33, 5221 (1986).
[3] T. Christensen, W. Yan, A.-P. Jauho, M. Wubs, and N.A. Mortensen, Phys. Rev. B92, 121407(R) (2015).
[4] T.G. Pedersen, Phys. Rev. B68, 245104 (2003).

30. Optical Properties of Ultrathin Metal Films

In previous chapters, we looked at the optical properties of bulk metals (Chapter 3) and metallic nanoparticles (Chapter 27). In addition, we have studied the optical response of two-dimensional semiconductors and graphene. The properties of two-dimensional metals, however, have not been discussed. Such thin films are of importance as coatings but, equally, as wave guides for plasmons. Free-electron metal films support quantum well states similarly to semiconductors. The large density of free carriers leads to quite different material properties in metals, however. In the present chapter, we will describe the optical properties of both ideal and realistic metal films.

We consider an ideal one-dimensional quantum well such as the one in Fig. A1.2. By ideal, we mean a structure, in which the potential only varies along z, while we have perfect translational invariance along x and y. In this case, the k-vector $\vec{k}$ is two-dimensional and the eigenstates and energy eigenvalues are always of the form

$$\psi_{n\vec{k}}(\vec{r}) = A^{-1/2}\varphi_n(z)e^{i\vec{k}\cdot\vec{r}}, \quad E_{n\vec{k}} = E_n + \frac{\hbar^2 k^2}{2m_e} \tag{30.1}$$

when normalized to a surface area A and labeled by band index n. The thickness of the film is d and so the volume is $\Omega = Ad$. As usual, the susceptibility in the direction perpendicular to the film is given by the standard expression

$$\chi_{zz}(\omega) = \frac{4e^2}{\varepsilon_0\Omega}\sum_{m,n,\vec{k}} f_{n\vec{k}} \frac{E_{mn\vec{k}}\left|\left\langle\psi_{m\vec{k}}\left|z\right|\psi_{n\vec{k}}\right\rangle\right|^2}{E_{mn\vec{k}}^2 - \hbar^2(\omega+i\Gamma)^2}.$$

In the semiconducting case, we always assumed that a dominant Bloch part of the wave function existed. This was subsequently used to approximate the momentum matrix elements in Chapters 15-19 and several other places. Here, however, we will assume that our free-electron states are really of the simple form Eq.(30.1) and, hence, all bands follow exactly identical dispersions in k-space. Hence, $E_{mn\vec{k}} = E_{mn}$ and $\left\langle\psi_{m\vec{k}}\left|z\right|\psi_{n\vec{k}}\right\rangle = \left\langle\varphi_m\left|z\right|\varphi_n\right\rangle$ are actually independent of k-vector. In a two-dimensional material at zero temperature, $\sum_{\vec{k}} f_{n\vec{k}} = m_e A/(2\pi\hbar^2)(E_F - E_n)\theta(E_F - E_n)$ and so

$$\chi_{zz}(\omega) = \frac{2e^2 m_e}{\pi\varepsilon_0\hbar^2 d}\sum_{m,n}\left|\left\langle\varphi_m\left|z\right|\varphi_n\right\rangle\right|^2 \frac{E_{mn}(E_F - E_n)\theta(E_F - E_n)}{E_{mn}^2 - \hbar^2(\omega+i\Gamma)^2}.$$

This susceptibility can be converted to a sheet conductivity via the relation $\sigma = -i\omega d\varepsilon_0\chi$ such that

$$\sigma_{zz}(\omega) = \sigma_0 \frac{8m_e\omega}{i\pi\hbar} \sum_{m,n} \left|\left\langle \varphi_m \middle| z \middle| \varphi_n \right\rangle\right|^2 \frac{E_{mn}(E_F - E_n)\theta(E_F - E_n)}{E_{mn}^2 - \hbar^2(\omega + i\Gamma)^2}$$

with $\sigma_0 = e^2/4\hbar$ the usual graphene DC value. Finally, in terms of the oscillator strength $g_{mn} = 2m_e \left|\left\langle \varphi_m \middle| z \middle| \varphi_n \right\rangle\right|^2 E_{mn}/\hbar^2$ of Chapter 2, we can write

$$\sigma_{zz}(\omega) = \sigma_0 \frac{4\hbar\omega}{i\pi} \sum_{m,n} g_{mn} \frac{(E_F - E_n)\theta(E_F - E_n)}{E_{mn}^2 - \hbar^2(\omega + i\Gamma)^2}. \tag{30.2}$$

The *xx* component of the conductivity is given by the Drude form since the *x*-direction is translationally invariant. The density is given by

$$n = \frac{2}{\Omega} \sum_{n\vec{k}} f_{n\vec{k}} = \frac{m_e}{\pi d\hbar^2} \sum_n (E_F - E_n)\theta(E_F - E_n).$$

Hence, from the free-electron plasma frequency $\omega_p^2 = e^2 n/(\varepsilon_0 m_e)$ we find the sheet conductivity of the Drude form $\sigma_{xx}(\omega) = \sigma_D(\omega)$ with

$$\sigma_D(\omega) = \sigma_0 \frac{4i\omega}{\pi\hbar(\omega + i\Gamma)^2} \sum_n (E_F - E_n)\theta(E_F - E_n). \tag{30.3}$$

This has the interesting implication that at high frequencies, for which $\hbar\omega \gg E_{mn}$ for all important transitions, and using the sum rule $\sum_m g_{mn} = 1$ we find

$$\sigma_{zz}(\omega) \approx \sigma_0 \frac{4i\omega}{\pi\hbar(\omega + i\Gamma)^2} \sum_{m,n} g_{mn}(E_F - E_n)\theta(E_F - E_n) = \sigma_{xx}(\omega) \quad \text{(high frequency limit)}.$$

Thus, at high frequencies the material becomes approximately isotropic. What "high" means precisely depends on thickness *d*, but generally we need $\hbar\omega$ much greater than the spacing between subsequent energy levels. In a simple square-well model, this means $\hbar\omega \gg \hbar^2\pi^2/(m_e d^2)$. In practice, it is impossible to sum all the infinitely many *m*-levels and, hence, it can be advantageous to write the *zz*-response in the form (using the identity $(E^2 - \omega^2)^{-1} = E^2\omega^{-2}(E^2 - \omega^2)^{-1} - \omega^{-2}$)

$$\sigma_{zz}(\omega) = \sigma_D(\omega) + \sigma_0 \frac{4\omega}{i\pi\hbar(\omega + i\Gamma)^2} \sum_{m,n} g_{mn} \frac{E_{mn}^2(E_F - E_n)\theta(E_F - E_n)}{E_{mn}^2 - \hbar^2(\omega + i\Gamma)^2}, \tag{30.4}$$

which ensures an isotropic response at high frequencies because the second term will be small.

The expressions Eq.(30.2) and (30.3) can be evaluated in various models. The jellium model discussed in Appendix 6 provides a rather accurate description of the quantum well including electron-electron interactions at the mean-field level. We will return to this model below. For now, however, we will apply the exceedingly simple infinite-barrier square well (free-electron) model (as in the exercise in Chapter 2). Hence, the states are

$$\varphi_n(z) = \sqrt{\frac{2}{d}} \cdot \begin{cases} \cos(nz\pi/d) & n \text{ odd} \\ \sin(nz\pi/d) & n \text{ even} \end{cases},$$

and the energies are $E_n = \hbar^2 n^2 \pi^2 / 2m_e d^2$. An elementary calculation shows that for $n+m$ odd $\left|\left\langle \varphi_m \middle| z \middle| \varphi_n \right\rangle\right| = 8dnm/[\pi^2(m^2-n^2)^2]$ while the matrix element vanishes for $n+m$ even. Consequently, $g_{mn} = 64n^2m^2/[\pi^2(m^2-n^2)^3]$. In Fig. 30.1 below, we plot the response for a free-electron slab with Fermi level E_F = 5 eV and $\hbar\Gamma = 25$ meV. The thicknesses 5 and 25 nm correspond to plasma frequencies of 8.21 and 8.35 eV, respectively.

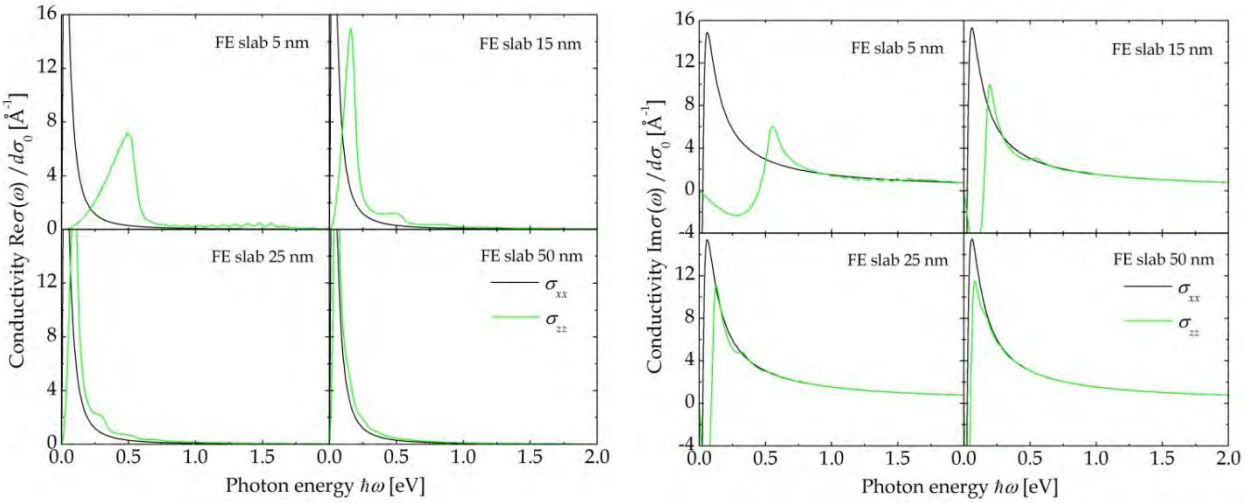


Figure 30.1. Real and imaginary parts of the anisotropic conductivity for free-electron slabs of various thicknesses.

30.1 Plasmons in Ultrathin Metal Film

We will now proceed by investigating plasmons in the thin films studied above. The surrounding dielectrics are assumed isotropic, as usual. Hence, we will look for bound electromagnetic modes in the geometry shown in Fig. 30.2.

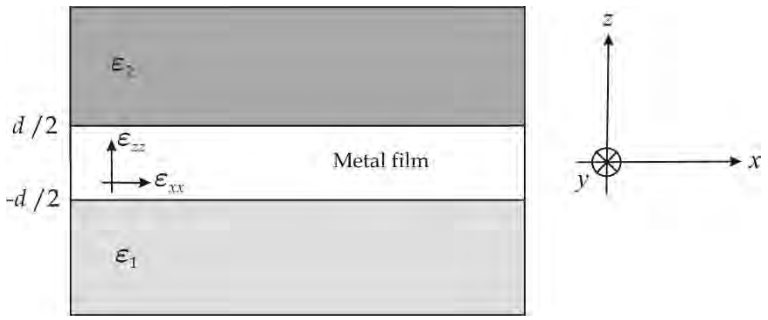


Figure 30.2. Schematic geometry of an anisotropic metal film sandwiched between isotropic dielectrics.

This geometry is obviously very similar to the graphene case studied in Chapter 29. We will keep a finite thickness d, however. The interesting case is that of p-polarization, in which the magnetic field is of the form $\vec{\mathcal{B}}(x,z) = \mathcal{B}(z)\,\vec{e}_y e^{ik_x x}$. In an anisotropic medium with an inverse dielectric tensor $\overleftrightarrow{\eta} = \overleftrightarrow{\varepsilon}^{-1}$, the wave equation is $\nabla\times(\overleftrightarrow{\eta}\cdot\nabla\times\vec{\mathcal{B}}) = k_0^2\vec{\mathcal{B}}$. In the present case, the tensor is diagonal $\overleftrightarrow{\eta} = \mathrm{diag}(\varepsilon_{xx}^{-1}, \varepsilon_{xx}^{-1}, \varepsilon_{zz}^{-1})$. Hence, if $\mathcal{B}(z) = \mathcal{B}e^{iqz}$, we find $q^2 = \varepsilon_{xx}k_0^2 - (\varepsilon_{xx}/\varepsilon_{zz})k_x^2$. Across an interface, $\mathcal{B}(z)$ as well as $\mathcal{B}'(z)/\varepsilon_{xx}$ are continuous. To look for bound modes, we write $q_1 = -i\kappa_1$ and $q_2 = i\kappa_2$ with $\kappa_i = (k_x^2 - \varepsilon_i k_0^2)^{1/2}$ for the isotropic dielectrics and assume

$$\mathcal{B}(z) = \begin{cases} A_1 e^{\kappa_1 z} & z < -d/2 \\ B_1\cos qz + B_2 \sin qz, & |z| < d/2 \\ A_2 e^{-\kappa_2 z} & z > d/2. \end{cases}$$

Matching across the interfaces leads to a dispersion relation

$$q\varepsilon_{xx}(\varepsilon_1\kappa_2 + \varepsilon_2\kappa_1) = (\varepsilon_1\varepsilon_2 q^2 - \varepsilon_{xx}^2\kappa_1\kappa_2)\tan(qd). \tag{30.5}$$

In a symmetric geometry with $\varepsilon_1 = \varepsilon_2 \equiv \varepsilon$, we find even and odd modes (of the magnetic field) with simplified dispersion relations

$$\text{even:}\quad \varepsilon_{xx}\kappa = \varepsilon q\tan(\tfrac{1}{2}qd),$$
$$\text{odd:}\quad \varepsilon_{xx}\kappa = -\varepsilon q\cot(\tfrac{1}{2}qd).$$

Notice the close resemblance with the corresponding equations for the electronic quantum well modes themselves in Appendix 1. For numerical purposes, the general expression Eq.(30.5) can conveniently be written using sheet conductivities via $\varepsilon_{ii} = 1 + i\sigma_{ii}/(\varepsilon_0\omega d) = 1 + i\pi\alpha\sigma_{ii}/(\sigma_0 k_0 d)$. In turn, if we take the limit $d\to 0$, we find the approximation

$$\frac{\varepsilon_1}{\kappa_1}+\frac{\varepsilon_2}{\kappa_2}+\frac{i\sigma_{xx}}{\varepsilon_0\omega}=0\,.$$

Thus, in this limit, we recover the graphene result Eq.(29.5) and anisotropy plays no role. For $\varepsilon_1=\varepsilon_2$, this mode is the odd one. In the quasi-static limit, the solution is $k_x/k_0=i(\varepsilon_1+\varepsilon_2)\sigma_0/(\pi\alpha\sigma_{xx})$. More accurately, the solution for $\varepsilon_1=\varepsilon_2$ is $k_x/k_0=\{\varepsilon_1-[2\varepsilon_1\sigma_0/(\pi\alpha\sigma_{xx})]^2\}^{1/2}$. The numerical solution for the plasmon dispersion relation for a 5 nm slab embedded in glass $(\varepsilon=1.5^2)$ taking $\hbar\Gamma=25$ meV is shown in Figs. 30.3 and 30.4 for odd and even modes, respectively. One may note that the real part of k_x is only slightly influenced be the anisotropy. On the other hand, the imaginary part related to absorption is increased whenever an allowed transition between quantum well levels is encountered. At low photon energies, absorption via the isotropic Drude term dominates and the effect of anisotropy is negligible.

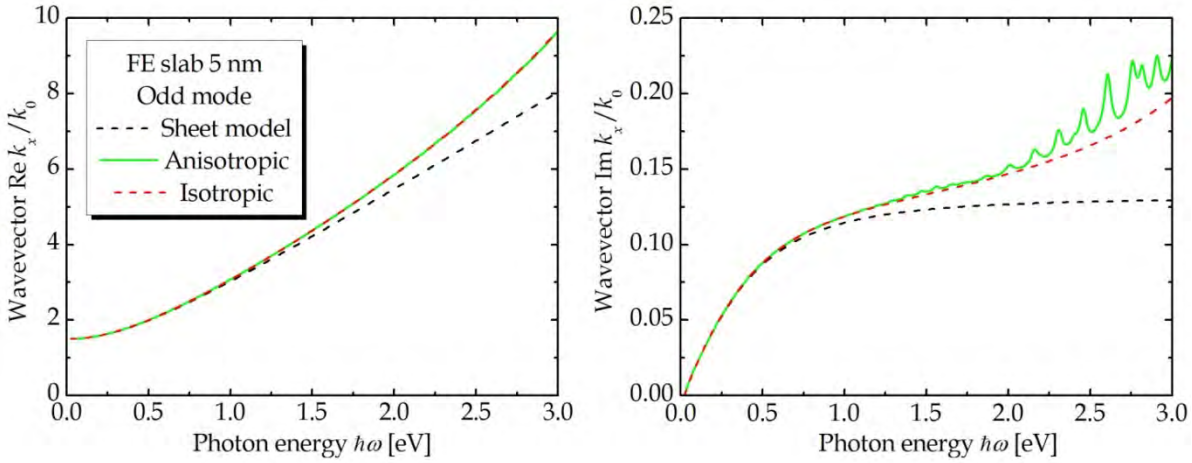


Figure 30.3. Plasmon dispersion relation for the odd mode with and without anisotropy.

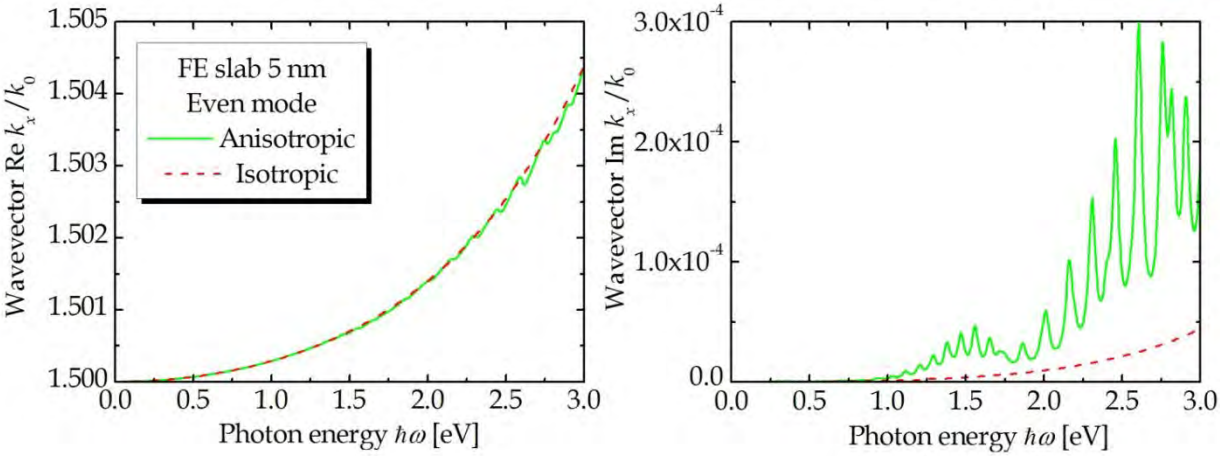


Figure 30.4. Plasmon dispersion relation for the even mode with and without anisotropy.

30.2 Jellium Model

A significantly more realistic description of the quantum well can be found in the jellium model presented in Appendix 6. Here, electron-electron interactions are accounted for by a self-consistent solution of the coupled Schrödinger and Poisson equations. Moreover, exchange and correlation effects are taken into account. In Fig. 30.5, the self-consistent potential and associated energy levels of 1 and 5 nm quantum wells are shown. The positive jellium charge corresponds to that of gold ($r_s = 3.01$ Bohr). It is clear that, for the wider structure, the potential is close to the simple square well.

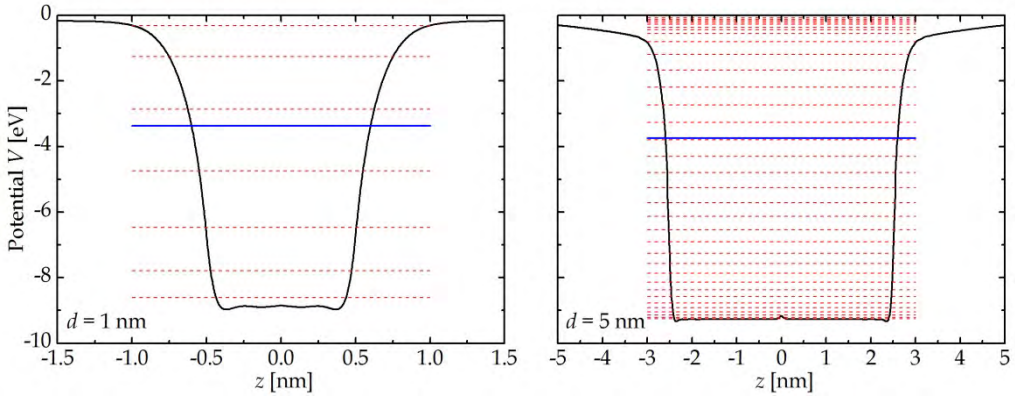


Figure 30.5. Potentials and energies of 1 and 5 nm quantum wells.
The Fermi level is indicated by the blue line.

In Fig. 30.6, we compare the response of jellium and square-well models with densities corresponding to bulk gold in all cases. In the 5 nm case, the two agree rather well with the only noticeable difference being that jellium energy levels are slightly closer due to spill-out.

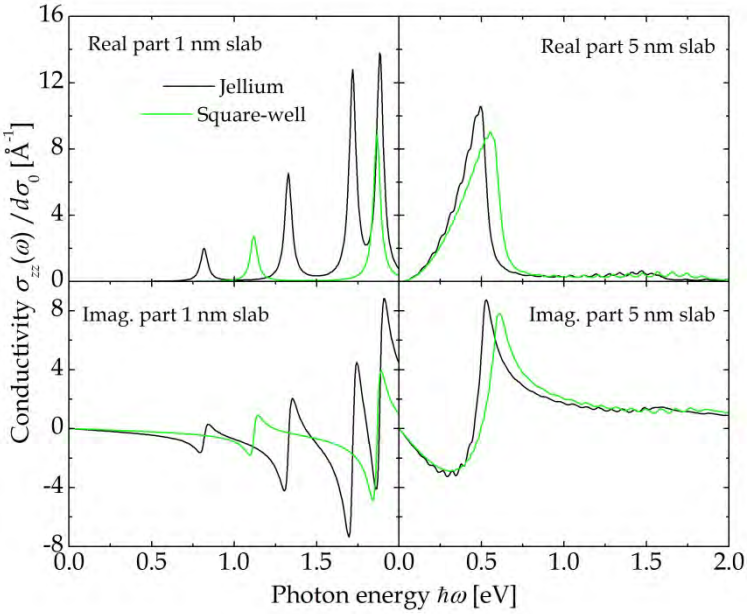


Figure 30.6. Comparison of jellium and square-well models for two quantum well widths.

30.3 Nonlocal Response and Spill-Out

In all calculations so far, the response has been assumed local, i.e. so that the local current is proportional to the local field. In fact, this is an approximation, and the current $\vec{j}(z)$ may depend on the field at another position $\vec{\mathcal{E}}(z')$, i.e. the response is nonlocal. This is a slightly subtle issue and, if the spatial variation of the field is taken into account, one cannot apply the dipole approximation for the interaction. Generally, a nonlocal response in the thin-film case is of the form $\vec{j}(z) = \int \overleftrightarrow{\sigma}(z,z') \cdot \vec{\mathcal{E}}(z')\, dz'$. The in-plane response is given by [1]

$$\sigma_{xx}(z,z',\omega) = \sigma_D(z,\omega)\delta(z-z') + \sigma_0 \frac{16\omega}{i\pi\hbar^3(\omega+i\Gamma)^2} \sum_{m,n} \frac{E_{mn}(E_F-E_n)^2\theta(E_F-E_n)}{E_{mn}^2-\hbar^2(\omega+i\Gamma)^2} \rho_{nm}(z)\rho_{mn}(z')$$

with $\rho_{nm}(z) = \varphi_n^*(z)\varphi_m(z)$ and the local Drude term $\sigma_D(z,\omega)$ given by

$$\sigma_D(z,\omega) = \sigma_0 \frac{4i\omega}{\pi\hbar(\omega+i\Gamma)^2} \sum_n (E_F-E_n)\theta(E_F-E_n)\,|\,\varphi_n(z)\,|^2 . \tag{30.6}$$

Hence, the Drude part remains local, but spatially dependent. The average of the in-plane nonlocal term vanishes due to orthogonality of the wave functions. The out-of-plane component is significantly modified, however. Hence, one finds [1]

$$\sigma_{zz}(z,z',\omega) = \sigma_D(z,\omega)\delta(z-z') + \sigma_{NL}(z,z',\omega)$$

with the nonlocal contribution

$$\sigma_{NL}(z,z',\omega) = \sigma_0 \frac{8\omega}{i\pi m_e\hbar(\omega+i\Gamma)^2} \sum_{m,n} \frac{E_{mn}(E_F-E_n)\theta(E_F-E_n)}{E_{mn}^2-\hbar^2(\omega+i\Gamma)^2} j_{nm}(z)j_{mn}(z') . \tag{30.7}$$

Here, the transition momentum density is defined by

$$j_{nm}(z) = \tfrac{1}{2}\left\{\varphi_n^*(z)\hat{p}_z\varphi_m(z) - \varphi_m(z)\hat{p}_z\varphi_n^*(z)\right\}.$$

It is readily shown that $\iint \sigma_{xx}(z,z',\omega)\,dz\,dz' = \sigma_{xx}(\omega)$ and similarly for the zz-response.

The local Drude term Eq.(30.6) leads to some significant new effects itself. This is because the response is now spatially dependent and, hence, we cannot simply match the fields across a few boundaries to solve the electromagnetic problem. In fact, in a

general anistropic local medium, the wave equation $\nabla\times(\vec{\vec{\eta}}(z)\cdot\nabla\times\vec{\mathcal{B}})=k_0^2\vec{\mathcal{B}}$ leads to the differential equation

$$\varepsilon_{zz}(z)\left\{\frac{d}{dz}\left(\frac{1}{\varepsilon_{xx}(z)}\frac{d\mathcal{B}(z)}{dz}\right)+k_0^2\mathcal{B}(z)\right\}=k_x^2\mathcal{B}(z). \tag{30.8}$$

If only the spatially dependent Drude term is retained, we have $\varepsilon_{xx}(z)=\varepsilon_{zz}(z)=1+i\sigma_D(z,\omega)/(\varepsilon_0\omega)\equiv\varepsilon(z)$. Thus, in this case

$$\varepsilon(z)\left\{\frac{d}{dz}\left(\frac{1}{\varepsilon(z)}\frac{d\mathcal{B}(z)}{dz}\right)+k_0^2\mathcal{B}(z)\right\}=k_x^2\mathcal{B}(z).$$

This equation is a (complex) eigenvalue condition for k_x^2. Numerically, it is problematic that all the variations in $\varepsilon(z)$ occur in a nanometer range, while $\mathcal{B}(z)$ itself varies on a hundreds-of-nanometers scale. The eigenvalue problem can be formulated in the Schrödinger form by writing $\mathcal{B}(z)=n(z)\mathcal{F}(z)$ with $n(z)=\sqrt{\varepsilon(z)}$ that leads to

$$\mathcal{F}''(z)+\left\{n^2(z)k_0^2-n(z)\left(\frac{1}{n(z)}\right)''\right\}\mathcal{F}(z)=k_x^2\mathcal{F}(z).$$

The classical case corresponds to taking $n(z)=n_1+(n-n_1)\theta(d/2-|z|)$ with $n_1=\sqrt{\varepsilon_1}$ in the symmetric case.

An alternative to a complicated numerical solution is to rely on perturbation theory. Suppose that the dielectric constant profile in Eq.(30.8) is close to the classical one, i.e. that spill-out is a minor correction. We will denote the classical quantities with a superscript (0) so that we have $\varepsilon_{xx}^{(0)}(z)$, $\varepsilon_{zz}^{(0)}(z)$, $\mathcal{B}^{(0)}(z)$ and $k_x^{(0)}$ in this case. The corresponding quantum parameters are without the superscript. The essence of perturbation theory is that $\mathcal{B}(z)\approx\mathcal{B}^{(0)}(z)$ and $\mathcal{B}'(z)/\varepsilon_{xx}(z)\approx\mathcal{B}^{(0)\prime}(z)/\varepsilon_{xx}^{(0)}(z)$ because these are the continuous quantities. If we now multiply Eq.(30.8) by $\mathcal{B}^{(0)*}(z)$ after dividing by $\varepsilon_{zz}(z)$ and integrate, we find

$$\begin{aligned}k_x^2\int\mathcal{B}^{(0)*}(z)\frac{1}{\varepsilon_{zz}(z)}\mathcal{B}(z)dz&=\int\mathcal{B}^{(0)*}(z)\left\{\frac{d}{dz}\left(\frac{1}{\varepsilon_{xx}(z)}\frac{d\mathcal{B}(z)}{dz}\right)+k_0^2\mathcal{B}(z)\right\}dz\\&\approx\int\mathcal{B}^{(0)*}(z)\left\{\frac{d}{dz}\left(\frac{1}{\varepsilon_{xx}^{(0)}(z)}\frac{d\mathcal{B}^{(0)}(z)}{dz}\right)+k_0^2\mathcal{B}^{(0)}(z)\right\}dz.\end{aligned}$$

On the other hand, for the classical case we have without approximation

$$(k_x^{(0)})^2 \int \mathcal{B}^{(0)*}(z) \frac{1}{\varepsilon_{zz}^{(0)}(z)} \mathcal{B}^{(0)}(z) dz = \int \mathcal{B}^{(0)*}(z) \left\{ \frac{d}{dz} \left(\frac{1}{\varepsilon_{xx}^{(0)}(z)} \frac{d\mathcal{B}^{(0)}(z)}{dz} \right) + k_0^2 \mathcal{B}^{(0)}(z) \right\} dz.$$

Hence, the two right-hand sides are equal and if we use $\mathcal{B}(z) \approx \mathcal{B}^{(0)}(z)$ in the left-hand side of the quantum case, we find

$$\left(\frac{k_x}{k_x^{(0)}} \right)^2 \approx \frac{\int \left| \mathcal{B}^{(0)}(z) \right|^2 \frac{1}{\varepsilon_{zz}^{(0)}(z)} dz}{\int \left| \mathcal{B}^{(0)}(z) \right|^2 \frac{1}{\varepsilon_{zz}(z)} dz}.$$

This means that the correction to the mode index can be found approximately without computing the actual field.

In the presence on nonlocality in the z-direction, we find a modified wave equation

$$k_0^2 \mathcal{B}(z) = -\frac{d}{dz} \int \varepsilon_{xx}^{-1}(z,z') \frac{d\mathcal{B}(z')}{dz'} dz' + k_x^2 \int \varepsilon_{zz}^{-1}(z,z') \mathcal{B}(z') dz'.$$

Here, the inverse response tensor must be understood in the sense

$$\int \vec{\vec{\varepsilon}}^{-1}(z,z'') \cdot \vec{\vec{\varepsilon}}(z'',z') dz'' = \vec{\vec{I}} \delta(z-z').$$

In the thin-film geometry, a sensible approximation is

$$\varepsilon_{xx}^{-1}(z,z') \approx \frac{1}{\varepsilon_{xx}(z)} \delta(z-z'), \quad \varepsilon_{zz}^{-1}(z,z') \approx \frac{1}{\varepsilon_{zz}(z)} \delta(z-z') - \frac{\chi_{NL}(z,z')}{\varepsilon_{zz}(z)\varepsilon_{zz}(z')},$$

where $\chi_{NL}(z,z')$ denotes the nonlocal contribution to the response, i.e. $\chi_{NL}(z,z') = i/(\varepsilon_0 \omega) \sigma_{NL}(z,z')$. Perturbatively, we then have

$$\left(\frac{k_x}{k_x^{(0)}} \right)^2 \approx \frac{\int \left| \mathcal{B}^{(0)}(z) \right|^2 \frac{1}{\varepsilon_{zz}^{(0)}(z)} dz}{\int \left| \mathcal{B}^{(0)}(z) \right|^2 \frac{1}{\varepsilon_{zz}(z)} dz - \iint \mathcal{B}^{(0)*}(z) \frac{\chi_{NL}(z,z')}{\varepsilon_{zz}(z)\varepsilon_{zz}(z')} \mathcal{B}^{(0)}(z') dz dz'}.$$

Exercise: Reflection from thin films

In this exercise, we consider a *p*-polarized optical beam incident on the film. First, we look at an interface between two anisotropic media, with the beam incident from medium 1 and going into medium 2. Hence, in the two media, we write the magnetic field as $\mathcal{B}_1(z) = e^{iq_1 z} + r_{12} e^{-iq_1 z}$ and $\mathcal{B}_2(z) = t_{12} e^{iq_1 z}$, respectively, with r_{12} and t_{12} the coefficients of reflection and transmission.

a) Show that the boundary conditions imply

$$r_{12} = \frac{\varepsilon_{2,xx} q_1 - \varepsilon_{1,xx} q_2}{\varepsilon_{2,xx} q_1 + \varepsilon_{1,xx} q_2}, \quad t_{12} = 1 + r_{12}\,.$$

Next, we look at the thin film system in Fig. 30.2 with a beam incident from below. In this case, the reflection coefficient of the stack becomes $r = (r_{1F} e^{-iqd} + r_{F2} e^{iqd}) / (e^{-iqd} + r_{1F} r_{F2} e^{iqd})$.

b) Show that setting $r = 0$ leads to the plasmon dispersion relation Eq.(30.5).

References

[1] P. J. Feibelman, Phys. Rev. B12, 1319 (1975).

31. Electron Energy Loss Spectroscopy

In this chapter, we will investigate a particular kind of interaction between electrons and light: light emission due to a moving charge. In turn, this will lead to absorption losses in any material in the vicinity of the electron. This energy loss must come from the electron, which will therefore lose speed. Measurements of this loss are known as electron energy loss spectroscopy (EELS) and constitute an important spectroscopic tool for both bulk and nanostructures. We will consider an electron travelling with speed v along the z-direction. The x and y coordinates will be denoted $\vec{r}_\parallel$ collectively. Below, z will be the direction perpendicular to the material surface. The charge and current densities are

$$\rho(\vec{r},t) = -e\delta(z-vt)\delta(\vec{r}_\parallel), \quad j_z(\vec{r},t) = -ev\delta(z-vt)\delta(\vec{r}_\parallel).$$

When Fourier transformed in time as well as the parallel coordinates, we have

$$\rho(z) = -e/v e^{iqz}, \quad j_z(z) = -ee^{iqz}$$

with $q = \omega / v$. In three-dimensional space with $\vec{k}_\parallel$ as the parallel wave vector, the full quantities are $\rho(\vec{r}) = (2\pi)^{-2}\int \rho(z)e^{i\vec{k}_\parallel\cdot\vec{r}_\parallel}d^2k_\parallel$ and similarly for all other quantities. This looks a little strange because $\rho(z)$ does not depend on $\vec{k}_\parallel$ but other quantities might. We now consider the electric field $\vec{\mathcal{E}}(z) = \vec{\mathcal{E}}_\parallel(z) + \mathcal{E}_z(z)\vec{e}_z$. Also, we will allow for anisotropic media, however restricting ourselves to a uniaxial dielectric response given by $\overleftrightarrow{\varepsilon} = \mathrm{diag}(\varepsilon_\parallel, \varepsilon_\parallel, \varepsilon_z)$. Below, we will introduce material interfaces located at certain z-values. However, everything will remain invariant in the parallel coordinates and, hence, we can solve for each $\vec{k}_\parallel$ separately. Now, the Maxwell divergence equation $\nabla\cdot\vec{\mathcal{D}} = \rho$ with $\vec{\mathcal{D}} = \varepsilon_0\overleftrightarrow{\varepsilon}\cdot\vec{\mathcal{E}}$ tells us that (with primes denoting d/dz)

$$i\varepsilon_\parallel\vec{k}_\parallel\cdot\vec{\mathcal{E}}_\parallel(z) + \varepsilon_z\mathcal{E}_z'(z) = -e/(v\varepsilon_0)e^{iqz}.$$

It follows that $\nabla\cdot\vec{\mathcal{E}} = (1-\varepsilon_z/\varepsilon_\parallel)\mathcal{E}_z'(z) - e/(v\varepsilon_0\varepsilon_\parallel)e^{iqz}$. The wave equation is, as usual,

$$\nabla^2\vec{\mathcal{E}}(\vec{r}) - \nabla\nabla\cdot\vec{\mathcal{E}}(\vec{r}) + k_0^2\overleftrightarrow{\varepsilon}\cdot\vec{\mathcal{E}}(\vec{r}) = -i\mu_0\omega\vec{j}(\vec{r}).$$

We use the divergence equation to isolate the z-component

$$\mathcal{E}_z''(z) - k_\parallel^2\mathcal{E}_z(z) - \frac{d}{dz}\left\{(1-\varepsilon_z/\varepsilon_\parallel)\mathcal{E}_z'(z) - e/(v\varepsilon_0\varepsilon_\parallel)e^{iqz}\right\} + \varepsilon_z k_0^2\mathcal{E}_z(z) = -i\mu_0\omega j_z(z).$$

This finally means that

$$\frac{\varepsilon_z}{\varepsilon_\parallel}\mathcal{E}_z''(z)+(\varepsilon_z k_0^2-k_\parallel^2)\mathcal{E}_z(z)=\frac{e}{i\varepsilon_0}\left(\frac{q}{v\varepsilon_\parallel}-\frac{k_0}{c}\right)e^{iqz}.$$

This is an important result, which we will now analyze. First, if we simply consider a homogeneous medium without boundaries, the full solution inside the medium is simply the partial one driven by the right-hand side, i.e. $\mathcal{E}_z(z)=\mathcal{E}_{in}e^{iqz}$ with

$$\mathcal{E}_{in}=\frac{e}{i\varepsilon_0 K_{in}^2}\left(\frac{q}{v\varepsilon_\parallel}-\frac{k_0}{c}\right),\quad K_{in}^2\equiv\varepsilon_z k_0^2-k_\parallel^2-(\varepsilon_z/\varepsilon_\parallel)q^2.$$

Clearly, $k_0=qv/c\equiv q\beta$ and

$$\mathcal{E}_{in}=\frac{eq}{iv\varepsilon_0\varepsilon_\parallel K_{in}^2}(1-\varepsilon_\parallel\beta^2),\quad K_{in}^2=-(\varepsilon_z/\varepsilon_\parallel)(1-\varepsilon_\parallel\beta^2)q^2-k_\parallel^2.$$

If we introduce $k_z^2\equiv(\varepsilon_z/\varepsilon_\parallel)(1-\varepsilon_\parallel\beta^2)q^2$ it can be seen that

$$\mathcal{E}_{in}=\frac{iek_z}{v\varepsilon_0 k^2}\left(\frac{1-\varepsilon_\parallel\beta^2}{\varepsilon_z\varepsilon_\parallel}\right)^{1/2}.$$

Now, since $dk_z=[(\varepsilon_z/\varepsilon_\parallel)(1-\varepsilon_\parallel\beta^2)]^{1/2}d\omega/v$ it follows that with $k^2=k_z^2+k_\parallel^2$

$$\mathcal{E}_{in}=\frac{iek_z}{\varepsilon_z\varepsilon_0 k^2}\frac{dk_z}{d\omega}.$$

In space-time, the field must be calculated from $\mathcal{E}_{in}(\vec{r},t)=(2\pi)^{-3}\int\mathcal{E}_{in}e^{i\vec{k}_\parallel\cdot\vec{r}_\parallel}e^{iqz}e^{-i\omega t}d^2k_\parallel d\omega$. If we introduce $\vec{R}=\vec{r}_\parallel+Z\vec{e}_z$ with $Z=[(\varepsilon_\parallel/\varepsilon_z)/(1-\varepsilon_\parallel\beta^2)]^{1/2}(z-vt)$ we then find

$$\mathcal{E}_{in}(\vec{r},t)=\frac{ie}{(2\pi)^3\varepsilon_0}\int\frac{k_z}{\varepsilon_z k^2}e^{i\vec{k}\cdot\vec{R}}d^3k=\frac{e}{(2\pi)^3\varepsilon_0}\frac{d}{dZ}\int\frac{1}{\varepsilon_z k^2}e^{i\vec{k}\cdot\vec{R}}d^3k.$$

The integral here is nothing but the Coulomb potential (in an anisotropic medium) and so the field agrees with standard expressions apart from relativistic and anisotropy corrections. In vacuum, we simply have

$$\mathcal{E}_{out} = \frac{e}{i\varepsilon_0 K_{out}^2}\left(\frac{q}{v} - \frac{k_0}{c}\right), \quad K_{out}^2 \equiv k_0^2 - k_\parallel^2 - q^2 .$$

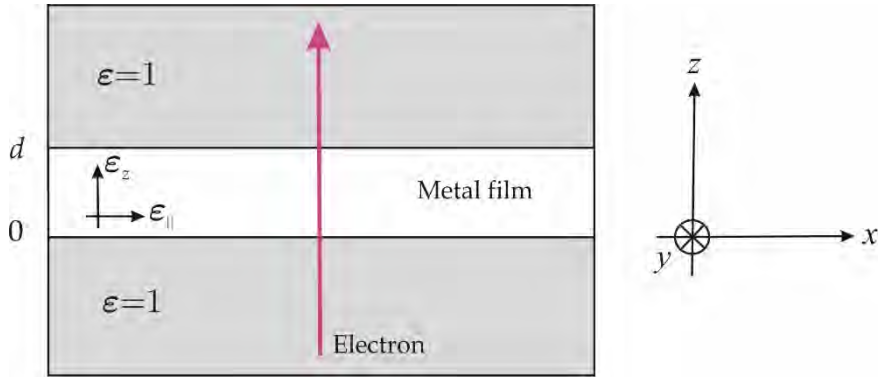


Figure 31.1. Thin-film geometry with incident electron from below.

31.1 Thin Films and Surfaces

Now, we will consider the more realistic case of a thin film, as illustrated in Fig. 31.1. This is more realistic, (1) because the electron is incident from some source and (2) for a thin film it is permissible to ignore that the electron velocity decreases due to losses during propagation. For a film in the region $0 < z < d$, we have generally

$$\mathcal{E}_z(z) = \begin{cases} re^{-ik_{out}z} + \mathcal{E}_{out}e^{iqz} & z < 0 \\ Ae^{ik_{in}z} + Be^{-ik_{in}z} + \mathcal{E}_{in}e^{iqz}, & 0 < z < d \\ te^{ik_{out}z} + \mathcal{E}_{out}e^{iqz} & z > d. \end{cases}$$

Here, $k_{in}^2 = \varepsilon_\parallel k_0^2 - (\varepsilon_\parallel / \varepsilon_z)k_\parallel^2$ and $k_{out}^2 = k_0^2 - k_\parallel^2$. The unknowns must be found from boundary conditions requiring $\varepsilon_z\mathcal{E}_z(z)$ and the tangential field continuous. Hence,

$$r + \mathcal{E}_{out} = \varepsilon_z(A + B + \mathcal{E}_{in}), \quad te^{ik_{out}d} + \mathcal{E}_{out}e^{iqd} = \varepsilon_z(Ae^{ik_{in}d} + Be^{-ik_{in}d} + \mathcal{E}_{in}e^{iqd}) .$$

Isolating from the divergence equation, we see that $(\varepsilon_z / \varepsilon_\parallel)\mathcal{E}_z'(z) + e/(v\varepsilon_0\varepsilon_\parallel)e^{iqz}$ must be conserved across all interfaces. Thus,

$$-k_{out}r + q\mathcal{E}_{out} + \frac{e}{iv\varepsilon_0} = \frac{\varepsilon_z}{\varepsilon_\parallel}(k_{in}A - k_{in}B + q\mathcal{E}_{in}) + \frac{e}{iv\varepsilon_0\varepsilon_\parallel}$$

$$k_{out}te^{ik_{out}d} + q\mathcal{E}_{out}e^{iqd} + \frac{e}{iv\varepsilon_0}e^{iqd} = \frac{\varepsilon_z}{\varepsilon_\parallel}(k_{in}Ae^{ik_{in}d} - k_{in}Be^{-ik_{in}d} + q\mathcal{E}_{in}e^{iqd}) + \frac{e}{iv\varepsilon_0\varepsilon_\parallel}e^{iqd}.$$

The work W done by the electron is the integral of the force $-e\,\mathrm{Re}\{\mathcal{E}_z(z)e^{-iqz}\}$, i.e. $W=(2\pi)^{-3}\int \Gamma d^2k_\parallel d\omega$ with

$$\Gamma=-e\int \mathrm{Re}\{\mathcal{E}_z(z)e^{-iqz}\}dz\,.$$

This becomes

$$\Gamma=\Gamma_{bulk}+e\,\mathrm{Im}\left\{\frac{r}{k_{out}+q}+\frac{te^{i(k_{out}-q)d}}{k_{out}-q}+A\frac{1-e^{i(k_{in}-q)d}}{k_{in}-q}-B\frac{1-e^{-i(k_{in}+q)d}}{k_{in}+q}\right\},$$

where the bulk loss is $\Gamma_{bulk}=-ed\,\mathrm{Re}\,\mathcal{E}_{in}$. In the quasi-static (non-retarded) limit $k_0=0$, the bulk loss is

$$\Gamma_{bulk}=-\frac{e^2qd}{\varepsilon_0 v}\mathrm{Im}\frac{1}{\varepsilon_\parallel k_\parallel^2+\varepsilon_z q^2}\,.$$

Things simplify considerably for an isotropic film, in which

$$\Gamma_{bulk}=-\frac{e^2qd}{\varepsilon_0 v(k_\parallel^2+q^2)}\mathrm{Im}\frac{1}{\varepsilon} \tag{31.1}$$

and the full result becomes

$$\Gamma=\Gamma_{bulk}+\frac{2e^2qk_\parallel}{\varepsilon_0 v(k_\parallel^2+q^2)^2}\mathrm{Im}\left\{\frac{(\varepsilon-1)^2}{\varepsilon}\cdot\frac{\cosh(k_\parallel d)+\varepsilon\sinh(k_\parallel d)-\cos(qd)}{2\varepsilon\cosh(k_\parallel d)+(\varepsilon^2+1)\sinh(k_\parallel d)}\right\}. \tag{31.2}$$

In the non-retarded limit, the reflection coefficient for the film is

$$r_p=\frac{(\varepsilon^2-1)\sinh(k_\parallel d)}{2\varepsilon\cosh(k_\parallel d)+(\varepsilon^2+1)\sinh(k_\parallel d)}\,.$$

One can therefore write the loss function as

$$\Gamma=\Gamma_{bulk}+\frac{2e^2qk_\parallel}{\varepsilon_0 v(k_\parallel^2+q^2)^2}\mathrm{Im}\left\{\frac{(\varepsilon-1)}{\varepsilon(\varepsilon+1)}\cdot\frac{\cosh(k_\parallel d)+\varepsilon\sinh(k_\parallel d)-\cos(qd)}{\sinh(k_\parallel d)}r_p\right\}.$$

We now wish to study the limit of a thin film, for which $\varepsilon=1+i\sigma/(d\varepsilon_0\omega)$. Plugging this into the general result and expanding in d, we find that

$$\Gamma \approx \frac{2e^2 q k_\parallel}{\varepsilon_0 v (k_\parallel^2 + q^2)^2} \operatorname{Im} r_p, \quad r_p \approx k_\parallel \sigma / (k_\parallel \sigma - 2i\varepsilon_0 \omega) \quad \text{(thin film)}. \qquad (31.3)$$

Similarly, expanding Eq.(31.2) in the limit of large d we see that

$$\Gamma \approx \Gamma_{bulk} + \frac{2e^2 q k_\parallel}{\varepsilon_0 v (k_\parallel^2 + q^2)^2} \operatorname{Im}\left\{\frac{(\varepsilon-1)^2}{\varepsilon(\varepsilon+1)}\right\}, \quad \text{(thick film)}.$$

The second, thickness-independent term is the contribution from a single surface. This form clearly shows that resonances are expected whenever $\varepsilon = 0$ or $\varepsilon + 1 = 0$. These conditions correspond to bulk and surface plasmons, respectively, and for a Drude metal they are located at $\omega_p / \sqrt{\varepsilon_\infty}$ and $\omega_p / \sqrt{\varepsilon_\infty + 1}$.

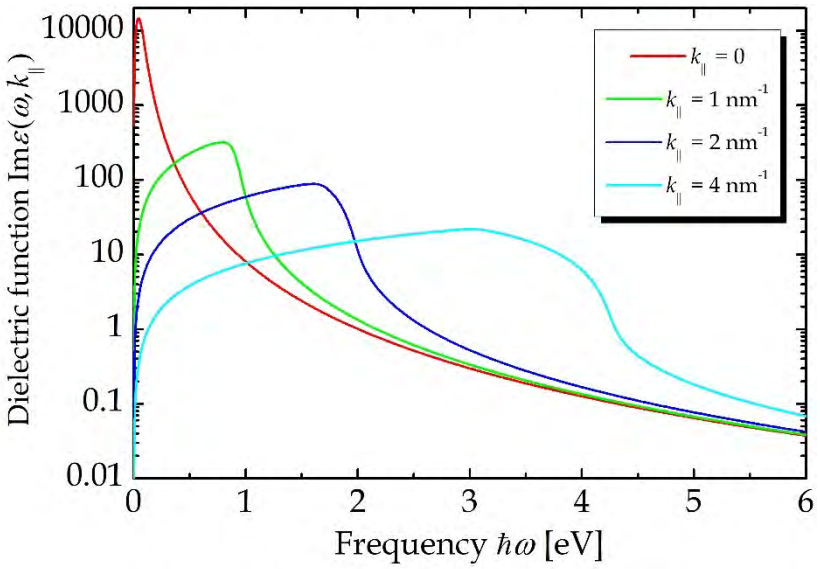


Figure 31.2. 3D Lindhard dielectric function for different wave vectors.

As a preliminary estimate, we will now use the 3D Lindhard model from Chapter 32 to compute the loss function including its wave vector dependence, adding a broadening of 0.05 eV. Hence, we take

$$\varepsilon(\omega, k_\parallel) = \varepsilon_\infty + \frac{e^2 m_* k_F}{\varepsilon_0 \pi^2 \hbar^2 k_\parallel^2} F_{3D}\left(\frac{k_\parallel}{k_F}, \frac{\hbar(\omega + i\Gamma)}{E_F}\right),$$

with

$$F_{3D}(x,y) = \int_0^1 \int_0^\pi \int_0^{2\pi} \left\{\frac{1}{y - x^2 - 2xz\cos\theta} - \frac{1}{y + x^2 + 2xz\cos\theta}\right\} \sin\theta z^2 d\phi d\theta dz$$
$$= \frac{1}{2} - \frac{x^4 + y^2 - 2x^2(y+2)}{16x^3} \ln\frac{x^2 + 2x - y}{x^2 - 2x - y} - \frac{x^4 + y^2 + 2x^2(y-2)}{16x^3} \ln\frac{x^2 + 2x + y}{x^2 - 2x + y}.$$

In Fig. 31.2, we have assumed $\varepsilon_\infty = 5$ and $E_F = 5.5$ eV corresponding roughly to Silver. The four curves are for $k_\parallel$ of 0, 1, 2 and 4 nm^{-1}. It is observed that a significant portion of the absorption shifts to higher momenta compared to the local response. We define the integrated loss function such that $W = \int \hbar\omega\gamma(\omega)d\hbar\omega$ i.e.

$$\gamma(\omega) = \frac{1}{(2\pi)^3\hbar^2\omega}\int \Gamma d^2k_\parallel .$$

In the case of a thin film, for which the sheet conductivity can be considered independent of $k_\parallel$, one finds

$$\gamma(\omega) = \frac{e^2}{8\pi^2\varepsilon_0\hbar^2v^2}\,\mathrm{Im}\left\{\frac{\pi q(q^2-3w^2)+2iw(q^2-w^2)-4iw^3\ln(q/iw)}{(q^2-w^2)^2}\right\}, \quad w = \frac{2\varepsilon_0\omega}{\sigma} \quad \text{(thin film)}.$$

In Fig. 31.3 below, we show the total (bulk + surface) loss for 1 and 10 nm films at $v = 0.1c$. The distinct bulk and surface responses are clearly resolved for the thicker film.

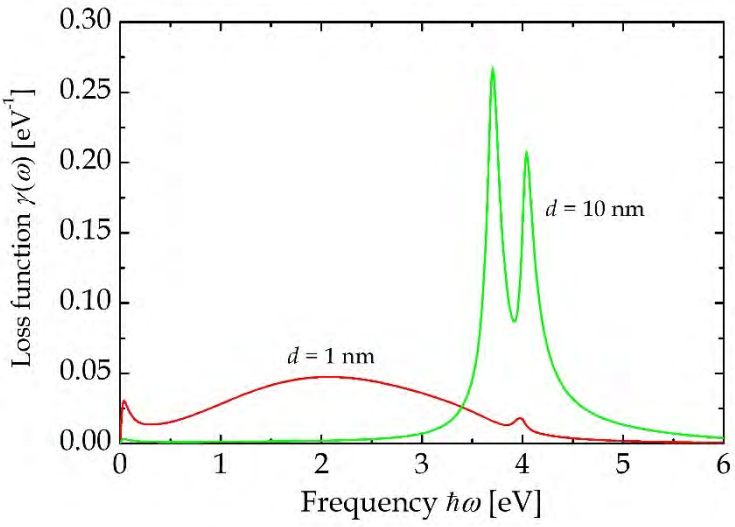


Figure 31.3. Loss function for thin and thick Ag films modelled using 3D Lindhard functions.

31.2 Quantum Size Effects

At this point it is reasonable to ask if a 3D Lindhard function really captures the response of a 1 nm film. To include size effects in the thin-film response, we return to the full space-dependent response function Eq.(31.2) but assume a confinement that depends on z alone. This means free-electron behavior in the parallel coordinates $\varphi_{n\vec{k}}(\vec{r}) = A^{-1/2}\psi_n(z)\exp(i\vec{k}\cdot\vec{r})$ and $\varphi_{m,\vec{k}+\vec{k}_\parallel}(\vec{r}) = A^{-1/2}\psi_m(z)\exp(i(\vec{k}+\vec{k}_\parallel)\cdot\vec{r})$. In turn, the single particle susceptibility is

$$\chi^S(\vec{r},\vec{r}') = \frac{1}{A}\sum_{\vec{k}_\parallel} \chi^S(z,z';\vec{k}_\parallel) e^{i\vec{k}_\parallel\cdot(\vec{r}-\vec{r}')},$$

with $\chi^S(z,z';\vec{k}_\parallel) = \sum_{nm} \alpha_{nm}(\omega,k_\parallel)\rho_{nm}(z)\rho_{mn}(z')$, where

$$\alpha_{nm}(\omega,k_\parallel) = \frac{2}{A}\sum_{\vec{k}} \frac{f(E_{n,\vec{k}}) - f(E_{m,\vec{k}+\vec{k}_\parallel})}{\hbar\omega + E_{n,\vec{k}} - E_{m,\vec{k}+\vec{k}_\parallel}}, \quad \rho_{nm}(z) = \psi_n^*(z)\psi_m(z).$$

In the 2D geometry, we can use the following representation of the Coulomb interaction

$$\frac{1}{4\pi\sqrt{r_\parallel^2 + z^2}} = \frac{1}{2A}\sum_{\vec{p}} \frac{e^{-i\vec{p}\cdot\vec{r}} e^{-p|z|}}{p},$$

where $\vec{p}$ is a two-dimensional vector. Hence, the dielectric function becomes

$$\varepsilon(z,z';\vec{k}_\parallel) = \varepsilon_\infty \delta(z-z') - \frac{e^2}{2\varepsilon_0 k_\parallel}\int e^{-k_\parallel|z-z''|}\chi^S(z'',z';\vec{k}_\parallel)dz''.$$

Hence, we introduce

$$f_{nm}(k_\parallel,z) = \int_0^d e^{-k_\parallel|z-z'|}\rho_{nm}(z')dz'$$

so that

$$\varepsilon(z,z';\vec{k}_\parallel) = \varepsilon_\infty \delta(z-z') - \frac{e^2}{2\varepsilon_0 k_\parallel}\sum_{n,m} \alpha_{nm}(\omega,k_\parallel) f_{nm}(z,k_\parallel)\rho_{mn}(z').$$

The spatial average will be

$$\varepsilon_S(\omega,k_\parallel) = \varepsilon_\infty + \frac{e^2}{2\varepsilon_0 k_\parallel^2 d}\sum_n \alpha_{nn}(\omega,k_\parallel) g_n(k_\parallel)$$

with

$$g_n(k_\parallel) = \int_0^d |\psi_n(z)|^2 \left(2 - e^{-k_\parallel z} - e^{-k_\parallel(d-z)}\right) dz.$$

Unfortunately, the macroscopic response is really the spatial average of the inverse dielectric function so that $\varepsilon_M^{-1}(\omega,k_\parallel)=d^{-1}\iint\varepsilon^{-1}(z,z';\vec{k}_\parallel)dzdz'$. This inverse is given by

$$\varepsilon^{-1}(z,z';\vec{k}_\parallel)=\frac{\delta(z-z')}{\varepsilon_\infty}+\frac{e^2}{2\varepsilon_0\varepsilon_\infty k_\parallel}\int e^{-k_\parallel|z-z''|}\chi(z'',z';\vec{k}_\parallel)dz'',$$

where χ is the many-body susceptibility related to χ^S via the Lipmann-Schwinger equation

$$\chi(z,z';\vec{k}_\parallel)=\frac{1}{\varepsilon_\infty}\chi^S(z,z';\vec{k}_\parallel)-\frac{e^2}{2\varepsilon_0\varepsilon_\infty k_\parallel}\iint\chi^S(z,z'';\vec{k}_\parallel)e^{-k_\parallel|z''-z'''|}\chi(z''',z';\vec{k}_\parallel)dz''dz''',$$

Hence, we define a generalized 2D Lindhard function

$$L_{nm}(\vec{q},\omega)=\frac{2}{A}\sum_{\vec{k}}f(E_n+E_{\vec{k}})\left\{\frac{1}{\hbar\omega-E_{mn}+E_{\vec{k}}-E_{\vec{k}+\vec{q}}}-\frac{1}{\hbar\omega+E_{mn}+E_{\vec{k}-\vec{q}}-E_{\vec{k}}}\right\}.$$

Introducing $E_{Fn}\equiv E_F-E_n=\hbar^2k_{Fn}^2/2m_*$ and $w_\pm=\frac{2m_*}{\hbar^2}(\hbar\omega+i\hbar\Gamma\pm E_{mn})$, this becomes

$$\begin{aligned}L_{nm}(\vec{k}_\parallel,\omega)&=\frac{m_*}{\pi^2\hbar^2}\int\theta(k_{Fn}-k)\left\{\frac{1}{k^2-|\vec{k}+\vec{k}_\parallel|^2+w_-}+\frac{1}{k^2-|\vec{k}-\vec{k}_\parallel|^2-w_+}\right\}d^2k\\&=-\frac{m_*}{2\pi\hbar^2}F_{2D}\left(\frac{k_\parallel}{k_{Fn}},\frac{\hbar(\omega+i\Gamma)}{E_{Fn}},\frac{E_{mn}}{E_{Fn}}\right),\end{aligned}$$

with

$$F_{2D}(x,y,z)=\frac{x^2-y+z}{x^2}\left\{1-\sqrt{1-\frac{4x^2}{(x^2-y+z)^2}}\right\}+\frac{x^2+y+z}{x^2}\left\{1-\sqrt{1-\frac{4x^2}{(x^2+y+z)^2}}\right\}.$$

Eventually,

$$\alpha_{nm}(\omega,k_\parallel)=-\frac{m_*}{2\pi\hbar^2}F_{2D}\left(\frac{k_\parallel}{k_{Fn}},\frac{\hbar(\omega+i\Gamma)}{E_{Fn}},\frac{E_{mn}}{E_{Fn}}\right).\qquad(31.4)$$

We will apply the infinite-barrier model, for which the energies are $E_n=\hbar^2k_n^2/2m_*$ and $\psi_n(z)=(2/d)^{1/2}\sin(k_nz)$ with $k_n=n\pi/d$. In this case,

$$g_n(k_\parallel) = 2 + \frac{8(e^{-k_\parallel d} - 1)k_n^2}{k_\parallel d(k_\parallel^2 + 4k_n^2)}$$

and

$$f_{nm}(k_\parallel, z) = 4k_\parallel \frac{2\cos(k_n z)\cos(k_m z)k_n k_m + \sin(k_n z)\sin(k_m z)(k_\parallel^2 + k_n^2 + k_m^2)}{d[k_\parallel^2 + (k_n - k_m)^2][k_\parallel^2 + (k_n + k_m)^2]}$$
$$-4k_\parallel \frac{(e^{-k_\parallel z} + (-1)^{n+m} e^{-k_\parallel (d-z)})k_n k_m}{d[k_\parallel^2 + (k_n - k_m)^2][k_\parallel^2 + (k_n + k_m)^2]}.$$

A convenient way of finding the many-body susceptibility is via Fourier-decomposition. Thus, we write $\chi^S(z, z'; \vec{k}_\parallel) = \sum_{nm} \chi^S_{nm} \sin(k_n z)\sin(k_m z')$ and similarly for $\chi(z, z'; \vec{k}_\parallel)$. It follows that

$$\varepsilon_\infty \chi_{nm} = \chi^S_{nm} - \sum_{pq} \chi^S_{np} W_{pq} \chi_{qm}$$

We have $\chi^S_{pq} = \frac{4}{d^2} \sum_{nm} \alpha_{nm}(\omega, k_\parallel) \rho^{(p)}_{nm} \rho^{(q)}_{mn}$ with

$$\rho^{(n)}_{pq} = \int_0^d \rho_{pq}(z)\sin(k_n z)dz = \frac{4[(-1)^{n+p+q} - 1]npq}{\pi[n^2 - (p+q)^2][n^2 - (p-q)^2]}$$

And

$$W_{pq} = \frac{e^2}{2\varepsilon_0 k_\parallel} \int_0^d \int_0^d e^{-k_\parallel |z - z'|} \sin(k_p z)\sin(k_q z')dzdz'$$
$$= \frac{e^2}{2\varepsilon_0 k_\parallel} \begin{cases} \dfrac{k_p k_q \{1 + (-1)^{p+q} - [(-1)^p + (-1)^q]e^{-k_\parallel d}\}}{(k_\parallel^2 + k_p^2)(k_\parallel^2 + k_q^2)} & p \neq q \\ \dfrac{(k_\parallel^2 + k_p^2)k_\parallel d + 2\{1 - (-1)^p e^{-k_\parallel d}\}k_p^2}{(k_\parallel^2 + k_p^2)^2} & p = q. \end{cases}$$

We also need $S_n = \int_0^d \sin(k_n z)dz = [1 - (-1)^n]/k_n$ and

$$\tilde{W}_{nm} = \frac{e^2}{2\varepsilon_0 k_\parallel} \int_0^d \int_0^d \int_0^d e^{-k_\parallel |z - z''|} \sin(k_n z'')\sin(k_m z')dzdz'dz''$$
$$= \frac{e^2}{2\varepsilon_0} \frac{\left(2k_\parallel^2 + (1 - e^{-k_\parallel d})k_n^2\right)}{k_\parallel^2\left(k_\parallel^2 + k_n^2\right)} S_n S_m.$$

Hence,

$$\varepsilon_S(\omega,k_\parallel)=\varepsilon_\infty-\frac{1}{d}\sum_{nm}\chi^S_{nm}\tilde{W}_{nm},\quad \varepsilon^{-1}_M(\omega,k_\parallel)=\frac{1}{\varepsilon_\infty}+\frac{1}{\varepsilon_\infty d}\sum_{nm}\chi_{nm}\tilde{W}_{nm}$$

The comparison between 2D and 3D models is shown in Fig. 31.4. The 2D response is based on a corrected Fermi level in order to match the bulk electron density. Also, for the many body response, the calculation is based on 1000 wave vectors in the range $k_\parallel\in[0,\frac{1}{2}k_F]$ and the z-coordinate is sampled with at least 20 points per nm. In the m-sum over quantum states, we include twice the number of occupied states. Note that this k-sampling is insufficient and produced unphysical low-energy oscillations in the spectra. It is seen that quantum size effects are huge for thin films and, even for thick ones, an increased broadening is observed. Also, a substantial difference between single-particle and many-body responses is seen for the thicker films.

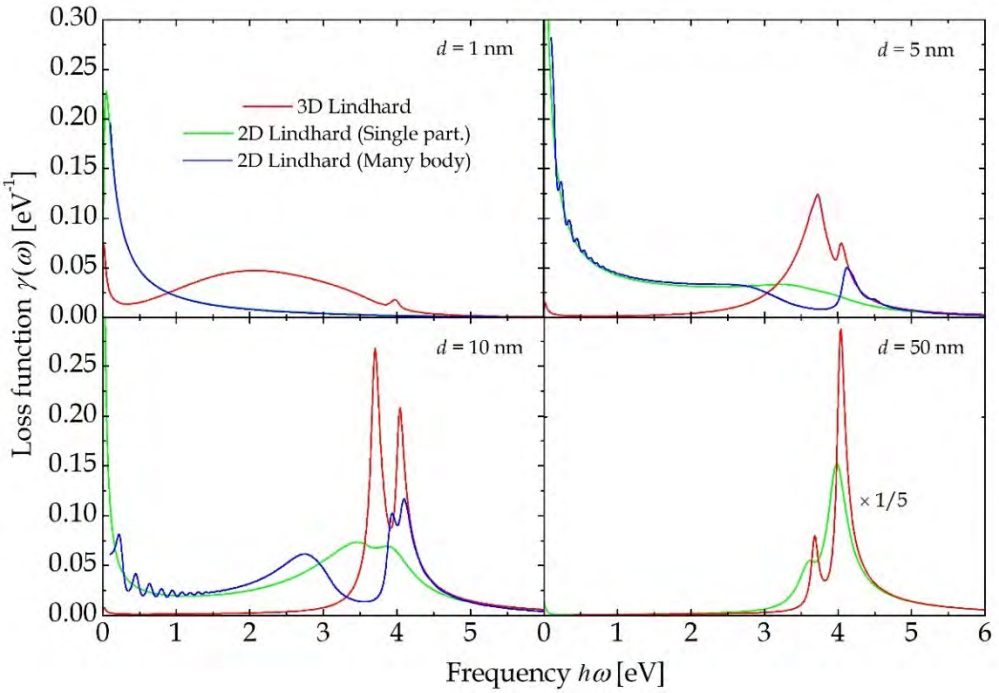


Figure 31.4. Comparison of loss functions for Ag films described using either 2D or 3D Lindhard functions.

Exercise: Limiting behavior of the Lindhard function

In the limit of small $k_\parallel$, the $F_{2D}(x,y,0)$ function becomes $F_{2D}(x,y,0) \approx -4x^2 / y^2$.

a) Show that this leads to a generalized Drude model

$$\varepsilon_S(\omega, k_\parallel) \approx \varepsilon_\infty - \frac{\omega_p^2(k_\parallel)}{(\omega + i\Gamma)^2}, \quad \omega_p^2(k_\parallel) = \frac{e^2}{2\pi\varepsilon_0 \hbar^2 d} \sum_n g_n(k_\parallel) E_{Fn} .$$

b) Show that the small and large d limits of $g_n(k_\parallel)$ are $k_\parallel d$ and 2, respectively. Since the electron density in the film is $n = m_* / (\pi\hbar^2 d) \sum_n E_{Fn}$, we then find

$$\text{thin film}: \omega_p^2(k_\parallel) \approx \frac{e^2 n}{\varepsilon_0 m_*} \cdot \frac{k_\parallel d}{2}, \quad \text{thick film}: \omega_p^2(k_\parallel) \approx \frac{e^2 n}{\varepsilon_0 m_*}.$$

32. Many-Body Polarizability

We have previously discussed the polarizability concept at length, c.f. Chapter 2 and 25. However, we restricted the actual computations in Chapter 2 to simple one-electron examples. We now wish to examine the many-body case and try to establish useful approximations. First, we will focus on the static case and later expand to include finite excitation frequencies. We will start by presenting one particular approach based on Dalgardo-Lewis perturbation theory [1], see also Appendix 11. This approach is easily applied to hydrogen.

The total problem in the static case can be formulated as $(\hat{H}_0 + \hat{H}_1)\psi = E\psi$, where $\hat{H}_0$ is the usual field-free Hamiltonian and $\hat{H}_1 = -eF\sum_i z_i$ is the perturbation. We restrict the discussion to systems that are symmetric in z. In such a system, the energy must be an even function of F. This means that by symmetry there cannot be any energy correction at first order of the perturbation. Including second order we therefore write

$$E = E_0 + E_2 + \ldots, \quad \psi = \psi_0 + \psi_1 + \psi_2 + \ldots$$

where the subscript indicates the order of the electric field. The zeroth order term is the usual unperturbed equation, whereas first order terms yield

$$(\hat{H}_0 - E_0)\psi_1 = -\hat{H}_1\psi_0 \,. \tag{32.1}$$

This constitutes a normal differential equation for ψ_1. Similarly, we may proceed by setting up an equation for the second order terms. If the inner product with the unperturbed state is taken it is then found that $E_2 = \langle \psi_0 | \hat{H}_1 | \psi_1 \rangle$.

We start by looking at hydrogen-like atoms for which the nuclear charge is ζe. In this case, the ground state is $\psi_0 = (\zeta^3/\pi)^{1/2} e^{-\zeta r}$. In polar coordinates, the perturbation is $-eFr\cos\theta$. From hereon, we work in atomic units such that distances are in Bohr a_0 and energies in Hartrees. If we write the wave function correction as $\psi_1 = g(r)\cos\theta\psi_0$ we find after insertion

$$-\frac{1}{2}g'' + \left(\xi - \frac{1}{r}\right)g' + \frac{1}{r^2}g = Fr \,.$$

This equation has the solution $g(r) = Fr(\frac{1}{\xi^2} + \frac{r}{2\xi})$. The correction to the energy thus becomes

$$E_2 = -F\frac{\zeta^3}{\pi}\int e^{-2\zeta r} g(r) r\cos^2\theta d^3r = -\frac{9F^2}{4\zeta^4}.$$

By definition, we introduce the static polarizability α by writing $E_2 = -\frac{1}{2}\alpha F^2$ and so $\alpha = 9/(2\zeta^4)$. This is an exact result for the polarizability. Reinstating units leads to the result $\alpha = 9a_0^4/(2\zeta^4)(e^2m/\hbar^2)$. For actual hydrogen, the exponent is $\zeta = 1$ and so the polarizability is $\alpha_H = 9/2$.

We now will try to develop a similar theory for two-electron atoms, i.e. helium and helium-like ions. A problem arises, however, because we don't know the actual ground state wave function. However, an excellent approximation is given by the Hylleraas form $\psi_0 \approx Ne^{-\zeta(r_1+r_2)}(1+cr_{12})$, where ζ and c are variational parameters. To perform the required integrations it is convenient to introduce so-called Hylleraas coordinates [2] given by

$$s = r_1 + r_2, \quad t = r_1 - r_2, \quad u = r_{12}.$$

If the integrand has no angular dependence, the Jacobian of the transformation is uJ with $J = s^2 - t^2$ and in terms of these parameters, integrals become

$$\int \ldots d^3r_1 d^3r_2 = \pi^2 \int_0^\infty \int_0^s \int_{-u}^u \ldots uJ dt du ds.$$

Thus, we readily find that

$$N = \frac{2\sqrt{2}\zeta^4}{\pi\left(8\zeta^2 + 35c\zeta + 48c^2\right)^{1/2}}.$$

Calculating the expectation values of the energy is a little more cumbersome as we need the Laplacian in the new coordinates. After some manipulations one finds [2]

$$\hat{H}_0\psi = -\psi_{ss} - \psi_{uu} - \psi_{tt} - \frac{2s(u^2-t^2)}{uJ}\psi_{su} - \frac{2t(s^2-u^2)}{uJ}\psi_{tu} - \frac{4s}{J}\psi_s - \frac{2}{u}\psi_u + \frac{4t}{J}\psi_t - \frac{4Zs}{J}\psi + \frac{1}{u}\psi.$$

Thus, after the integration the energy becomes

$$E_0 = \zeta\frac{2\zeta^2(5+8\zeta-16Z) + 2c\zeta(16+25\zeta-60Z) + c^2(35+64\zeta-144Z)}{16\zeta^2 + 70c\zeta + 96c^2}.$$

If this expression is minimized with respect to ζ and *c*, a minimum of $E_0 = -2.8911$ is found at $\zeta = 1.8497$ and $c = 0.3658$. This variational result is obviously quite close to the exact value of $E_0 = -2.903$. One can do even better by adding more variational parameters. A list of increasingly advanced functions is shown in Table 32.1 along with optimized parameters and energies. For the most advanced 6-parameter Hylleraas function, the agreement with experiments is practically perfect.

Wave function	Parameters	Energy	Pol.
$\psi_0 \approx Ne^{-\zeta(r_1+r_2)}$	$\zeta = 1.6875$	-2.8477	1.110
$\psi_0 \approx Ne^{-\zeta(r_1+r_2)}(1+cr_{12})$	$\zeta = 1.8497, c = 0.3658$	-2.8911	0.999
$\psi_0 \approx Ne^{-\zeta(r_1+r_2)}(1+cr_{12}+d(r_1-r_2)^2)$	$\zeta = 1.8161, c = 0.2918, d = 0.1308$	-2.9024	1.311
$\psi_0 \approx Ne^{-\zeta(r_1+r_2)}[1+cr_{12}+d(r_1-r_2)^2$ $+e(r_1+r_2)+f(r_1+r_2)^2+hr_{12}^2]$	$\zeta = 1.7557, c = 0.3373, d = 0.1125$ $e = -0.1459, f = 0.0236,$ $h = -0.0370$	-2.9033	1.386

Table 32.1. Various variational ansätze and the corresponding optimized parameters, energies and static polarizabilities.

We can now return to the problem of the polarizability. Because we don't know the exact ground state wave function, we need to modify the approach slightly. Taking inspiration from the hydrogen case, we approximate the full wave function in the presence of an electric field by $\psi \approx \psi_0 + \psi_1$, where ψ_1 is a first order correction. Now, exploiting the fact that the first order correction to the energy vanishes we find that

$$E = \frac{\langle\psi|\hat{H}_0 + \hat{H}_1|\psi\rangle}{\langle\psi|\psi\rangle} \approx \frac{E_0 + 2\langle\psi_0|\hat{H}_1|\psi_1\rangle + \langle\psi_1|\hat{H}_0|\psi_1\rangle}{\langle\psi_0|\psi_0\rangle + \langle\psi_1|\psi_1\rangle}.$$

Since $\langle\psi_0|\psi_0\rangle = 1$ and $\langle\psi_1|\psi_1\rangle$ is small, we expand and thereby find the second order energy

$$E_2 \approx 2\langle\psi_0|\hat{H}_1|\psi_1\rangle + \langle\psi_1|\hat{H}_0 - E_0|\psi_1\rangle.$$

We see that if Eq.(32.1) still holds, we obtain $E_2 = \langle\psi_0|\hat{H}_1|\psi_1\rangle$ as before. The above expression, however, is still valid as a variational estimate even if Eq.(32.1) is not fulfilled and only an approximate ground state is used. We now take additional inspiration from the hydrogen case and write the modification as $\psi_1 = [g(\vec{r}_1) + g(\vec{r}_2)]\psi_0$, with more generally $g(\vec{r}) = Fz(a+br)$. To simplify the computation we split the terms arising by having $\hat{H}_0$ act on a product. In this manner, we find

$$E_2 \approx -4F\langle\psi_0|g(\vec{r}_1)(z_1+z_2)|\psi_0\rangle + 2\langle\psi_0|g(\vec{r}_1)[g(\vec{r}_1)+g(\vec{r}_2)](\hat{H}_0-E_0)|\psi_0\rangle + \langle\psi_0|[\nabla_1 g(\vec{r}_1)]^2|\psi_0\rangle.$$

At this point, we again make the approximation that we're dealing with the exact ground state. In that case, the middle term above vanishes identically and we have

$$\frac{E_2}{F^2} \approx -4\langle\psi_0|(z_1+z_2)z_1(a+br_1)|\psi_0\rangle + \langle\psi_0|\left(a^2+2ab(r_1+z_1^2/r_1)+b^2(r_1^2+3z_1^2)\right)|\psi_0\rangle.$$

We now face the task of minimizing the second order energy with respect to *a* and *b*. Differentiating and solving, we immediately find

$$\alpha = 2\frac{18\left[6\langle z_1^2\rangle\left(\langle z_1^2\rangle+\langle z_1z_2\rangle\right)^2+\left(\langle z_1^2r_1\rangle+\langle z_1z_2r_1\rangle\right)^2\right]-48\left(\langle z_1^2\rangle+\langle z_1z_2\rangle\right)\left(\langle z_1^2r_1\rangle+\langle z_1z_2r_1\rangle\right)\langle r_1\rangle}{27\langle z_1^2\rangle-8\langle r_1\rangle^2}.$$

Here, the averages are to be made using ψ_0^2 as weight. To simplify, we may next symmetrize over *x*, *y*, and *z* and replace $z_1^2 \to r_1^2/3$ and $z_1z_2 \to (r_1^2+r_2^2-u^2)/6$. Writing $r_1=(s+t)/2$ and $r_2=(s-t)/2$ makes it straightforward to evaluate the remaining integrals in Hylleraas coordinates.

The results using various variational guesses for the ground state are listed in the Table. For the simplest form $\psi_0 \approx Ne^{-\zeta(r_1+r_2)}$, the He polarizability is simply doubled compared to hydrogen, i.e. $\alpha = 9/\zeta^4$. For helium, the optimized exponent is $\zeta = 27/16$ and so the polarizability becomes $\alpha \approx 1.11$. This is rather far from the exact value of $\alpha = 1.380$. In contrast, if a highly accurate (and complicated) variational form is used, both the ground state energy and polarizability are very close to exact results. Results for various variational ansätze are given in Table 32.1.

32.1 Frequency Dependence

Having tackled the static polarizability, we now return to the frequency dependent one given in Chapter 2 as (using atomic units)

$$\alpha(\omega) = \sum_n \frac{g_{n0}}{E_{n0}^2-(\omega+i\Gamma)^2}, \quad g_{n0} = 2\left|\langle\varphi_n|\sum_i z_i|\varphi_0\rangle\right|^2 E_{n0},$$

where g_{n0} is the oscillator strength. In these expressions, *n* denotes the *n*'th excited state and 0 is the ground state. For the hydrogen atom, the ground state is the 1s state and, as discussed in Chapter 2, the only excited states that couple to this ground state are *p*-type with a matrix element given by

$$\left\langle \varphi_{np} \left| z \right| \varphi_{1s} \right\rangle = \frac{1}{\sqrt{3}} \int_0^\infty R_{np}(r) R_{1s}(r) r^3 dr \ . \tag{32.2}$$

The radial eigenfunctions needed here are $R_{1s}(r) = 2e^{-r}$ and

$$R_{np}(r) = \frac{4}{n^3} \sqrt{\frac{(n-2)!}{(n+1)!}} r L_{n-2}^3 \left(\tfrac{2r}{n}\right) e^{-r/n} \ . \tag{32.3}$$

Here, L_n^m is an associated Laguerre polynomial. The integral can be carried out analytically with the result [3]

$$g_{n0} = \frac{2^8 n^5 (n-1)^{2n-4}}{3(n+1)^{2n+4}}.$$

Using transition energies given by $E_{n0} = \frac{1}{2} - \frac{1}{2n^2}$, the first few terms are $g_{20} = 0.4162$, $g_{30} = 0.0791$, $g_{40} = 0.0290$, $g_{50} = 0.0139$ etc. Recalling the Thomas-Reiche-Kuhn sum rule that the g's should sum to 1 (for a one-electron system) we see that this appears to be violated. In fact, summing the first 200 terms yields a total of 0.5650. The reason is, of course, that transitions to the continuum of ionized states have been ignored. These states form a continuum at energies above the ionization threshold that in atomic units lies at $\omega = \frac{1}{2}$. To account for these transitions, we must therefore write

$$\alpha(\omega) = \sum_n \frac{g_{n0}}{E_{n0}^2 - (\omega + i\Gamma)^2} + \int_{1/2}^\infty \frac{1}{E^2 - (\omega + i\Gamma)^2} \frac{dg_{E0}}{dE} dE \, .$$

where g_{E0} is the "oscillator strength" of the continuum transition and the discrete sum only covers bound states. It turns out that, writing $E = \frac{1}{2} + \frac{k^2}{2}$ i.e. separating out the kinetic energy, we have $dg_{E0} / dE \equiv g'(k)$ with [3]

$$g'(k) = \frac{2^8}{3(1+k^2)^4} \exp\left\{-4 \frac{\tan^{-1} k}{k}\right\} / \left(1 - \exp\left\{-2\pi / k\right\}\right). \tag{32.4}$$

And so

$$\alpha(\omega) = \sum_n \frac{g_{n0}}{E_{n0}^2 - (\omega + i\Gamma)^2} + \int_0^\infty g'(k) \frac{k}{\frac{1}{4}(1+k^2)^2 - (\omega + i\Gamma)^2} dk \, .$$

In this manner, the high-frequency limit becomes

$$\alpha(\omega) \approx -\frac{1}{\omega^2}\left\{\sum_n g_{n0} + \int_0^\infty g'(k)kdk\right\}.$$

Integration readily shows that the combination inside the curly brackets is, indeed, unity. Similarly, in the static limit we find

$$\alpha(0) = \sum_n \frac{g_{n0}}{E_{n0}^2} + 4\int_0^\infty g'(k)\frac{k}{(1+k^2)^2}dk\,.$$

This expression evaluates to 9/2 as it must. Quite remarkably, Gavrila [4] has managed to evaluate the full polarizability is closed form

$$\alpha(\omega) = -\frac{1}{2\omega^2} + \frac{128(1-2\omega)^2}{\omega^2(\sqrt{1-2\omega}+1)^8(2\sqrt{1-2\omega}-1)}\times$$
$${}_2F_1\left[4, 2-\tfrac{1}{\sqrt{1-2\omega}}, 3-\tfrac{1}{\sqrt{1-2\omega}}, \left(\tfrac{\sqrt{1-2\omega}-1}{\sqrt{1-2\omega}+1}\right)^2\right] + (\omega \to -\omega).$$

In Fig. 32.1, real and imaginary parts of the polarizability are plotted using $\Gamma = 0.005$ The static limit is clearly seen to be correct. Note, also, the discrete resonances below ionization $\omega < \frac{1}{2}$ and the continuum above.

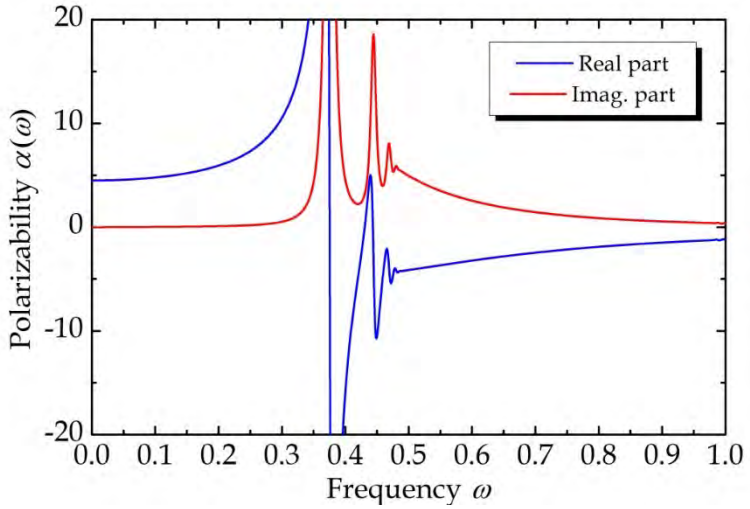


Figure 32.1. Complex polarizability of hydrogen atoms.

To make a similar calculation for He, we need appropriate excited states. A simple variational estimate can be constructed using the singlet ansatz

$$\psi_{nP} = \frac{1}{\sqrt{2}}\left\{\varphi_{1s}(\vec{r}_1)\varphi_{np}(\vec{r}_2) + \varphi_{1s}(\vec{r}_2)\varphi_{np}(\vec{r}_1)\right\}.$$

Here, we use hydrogenic eigenstates of the form $\varphi_{1s}(\vec{r}) = Y_{00}(\theta,\phi)R_s(r)$ and $\varphi_{np}(\vec{r}) = Y_{10}(\theta,\phi)R_{np}(r)$ with $R_{1s}(r) = 2\zeta_s^{3/2}e^{-\zeta_s r}$ and $R_{np}(r)$ given by Eq.(32.3) but with a screened exponent, i.e.

$$R_{np}(r) = \frac{4\zeta_p^{5/2}}{n^3}\sqrt{\frac{(n-2)!}{(n+1)!}} r L_{n-2}^3\left(\tfrac{2\zeta_p r}{n}\right) e^{-\zeta_p r/n} \quad .$$

It should be noted that these states are automatically orthogonal to the ground state. However, *different* excited states corresponding to different values of n are only orthogonal if the same exponent ζ_p is used for all states. We determine the best choice of ζ_p and ζ_s by optimizing the energy of the lowest excited state, i.e. n = 2. Similarly to a standard Hartree-Fock calculation (see App. 4) we find

$$\langle \psi_{nP} | \hat{H}_0 | \psi_{nP} \rangle = h_{1s} + h_{np} + J_{1snp} + K_{1snp} \, .$$

Evaluating all terms we find for n = 2

$$\langle \psi_{2P} | \hat{H}_0 | \psi_{2P} \rangle = \frac{\zeta_s^2}{2} - Z\zeta_s + \frac{\zeta_p^2}{8} - \frac{Z\zeta_p}{4} + \frac{\zeta_p^5}{4}\left\{\frac{1}{\zeta_p^4} - \frac{\zeta_p + 6\zeta_s}{(\zeta_p + 2\zeta_s)^5}\right\} + \frac{112\zeta_s^3\zeta_p^5}{3(\zeta_p + 2\zeta_s)^7} .$$

Minimizing, we find $E_{2P} \approx -2.1224$ at $\zeta_s \approx Z = 2$ and $\zeta_p = 0.9646$. For $n > 2$, the energies follow the Rydberg series $E_{nP} \approx -2 - \frac{1}{2n^2}$ to very high accuracy. The values obtained using these exponents for some low values of n are shown in Table 32.2.

Excited state	Energy E_{nP}	Osc. strength g_{nP}
ψ_{2P}	-2.1224	0.2633
ψ_{3P}	-2.0556	0.0717
ψ_{4P}	-2.0313	0.0275
ψ_{5P}	-2.0200	0.0145

Table 32.2. Energies and oscillator strengths of some excited *p* states.

To compute the oscillator strengths of the $0 \to nP$ transitions we need the dipole matrix elements given by

$$\langle \psi_0 | z_1 + z_2 | \psi_{nP} \rangle = \frac{\sqrt{6}}{4\pi} \int\int \psi_0 \frac{z_1}{r_1}(z_1 + z_2) R_{np}(r_1) R_{1s}(r_2) d^3r_1 d^3r_2 \, .$$

These integrals may again be evaluated by symmetrizing over x, y, and z and using Hylleraas coordinates. The corresponding oscillator strengths are given by

$$g_{nP} = 2\left|\langle \psi_0 | z_1 + z_2 | \psi_{nP} \rangle\right|^2 (E_{nP} - E_0) \, .$$

The table above lists a few results for the lowest transitions. Here, it is very clearly seen that the sum rule $\sum g_{nP} = 2$ is badly broken if only bound states are counted. Adding continuum contributions is tricky. A simple approximation can be obtained, however, by appealing to the hydrogen result with a few modifications. First, for this part, we approximate the ground state by the simple top-most product in Table 32.1 $\psi_0 \approx (\zeta_s^3/\pi)e^{-\zeta_s(r_1+r_2)}$. In this manner,

$$\langle \psi_0 | z_1 + z_2 | \psi_{nP} \rangle \approx 4\sqrt{\frac{2}{3}}\zeta_s^3 \int_0^\infty R_{np}(r)e^{-\zeta_s r} r^3 dr \int_0^\infty R_{1s}(r)e^{-\zeta_s r} r^2 dr \ .$$

We further approximate $\zeta_p \approx 1$. Thus, based on the results of the exercise below, we will use the modified oscillator strength distribution

$$g'(k) = C\frac{2^8\zeta_s^3(2\zeta_s - 1)^2(2E_{IP} + k^2)(1 + k^2)}{3(\zeta_s^2 + k^2)^6}\exp\left\{-4\frac{\tan^{-1} k/\zeta_s}{k}\right\}/\left(1 - \exp\{-2\pi/k\}\right).$$

It is apparent that we have modified the hydrogen result in three ways: (1) the H ionization potential of $1/2$ is replaced by the He value $E_{IP} = -2.0 + 2.9033 = 0.9033$, (2) the stretching of the *s*-state for the ground state approximation is accounted for via ζ_s, and (3) the appearance of a prefactor $C = 2|2\zeta_s^{3/2}\int_0^\infty R_{1s}(r)e^{-\zeta_s r} r^2 dr|^2$ that is given by the overlap between the two 1s states. Using $\zeta_s = 27/16$ (and $\zeta_s = 2$ in R_{1s}) leads to $C = 1.96$. With these modifications, we calculate the polarizability from

$$\alpha(\omega) = \sum_n \frac{g_{nP}}{(E_{nP} - E_0)^2 - (\omega + i\Gamma)^2} + \int_0^\infty g'(k)\frac{k}{\frac{1}{4}(2E_{IP} + k^2)^2 - (\omega + i\Gamma)^2}dk \ .$$

The result is illustrated in Fig. 32.2.

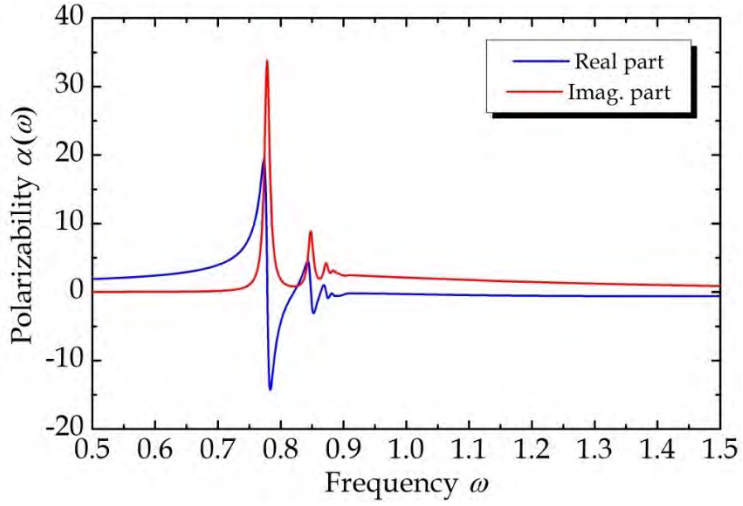


Figure 32.2. Complex polarizability of helium atoms.

As a consequence of various approximations, we clearly expect sum rules etc. to be violated. However, in fact, it turns out that the sum of oscillator strengths is 2.14, only slightly above the exact Thomas-Reiche-Kuhn value of 2. Moreover, it is observed that a static value of $\alpha(0) \approx 1.33$ is found in reasonable agreement with the exact value of 1.38.

32.2 Alternative Green's Function Approach

A very different approach to the frequency dependent polarizability problem can be formulated using Green's functions. We concentrate here exclusively on the single-electron case. First, we notice that the polarizability can be written as

$$\alpha(\omega) = \sum_n \left\{ \frac{\left|\left\langle \varphi_n \middle| z \middle| \varphi_0 \right\rangle\right|^2}{E_{n0} - \omega} + \frac{\left|\left\langle \varphi_n \middle| z \middle| \varphi_0 \right\rangle\right|^2}{E_{n0} + \omega} \right\},$$

where the imaginary part of the frequency is absorbed into ω. Reorganizing, this becomes

$$\alpha(\omega) = \left\langle \varphi_0 \middle| z \left\{ \sum_n \frac{\left|\varphi_n\right\rangle\left\langle\varphi_n\right|}{E_n - E_0 - \omega} + \sum_n \frac{\left|\varphi_n\right\rangle\left\langle\varphi_n\right|}{E_n - E_0 + \omega} \right\} z \middle| \varphi_0 \right\rangle.$$

The operator sum $\sum_n (E - E_n)^{-1} \left|\varphi_n\right\rangle\left\langle\varphi_n\right|$ is the Green's function for the unperturbed problem $G(\vec{r}, \vec{r}'; E) = (E - \hat{H}_0)^{-1}$ satisfying the relation $(E - \hat{H}_0) G(\vec{r}, \vec{r}'; E) = \delta(\vec{r} - \vec{r}')$. Hence,

$$\alpha(\omega) = -\left\langle \varphi_0 \middle| z \left\{ G(\vec{r}, \vec{r}'; E_0 + \omega) + G(\vec{r}, \vec{r}'; E_0 - \omega) \right\} z \middle| \varphi_0 \right\rangle.$$

In spherical coordinates, the delta function becomes $\delta(r - r')\delta(\Omega - \Omega') / (r')^2$, where Ω denotes the polar angles. For problems with spherical symmetry, the states are of the form $\varphi_{nlm}(\vec{r}) = Y_{lm}(\Omega) R_{nl}(r)$ and so the Green's function becomes

$$G(\vec{r}, \vec{r}'; E) = \sum_{nlm} (E - E_{nl})^{-1} \left|\varphi_{nlm}\right\rangle\left\langle\varphi_{nlm}\right| = \sum_{lm} Y^*_{lm}(\Omega') Y_{lm}(\Omega) g_l(r, r'; E),$$

where g_l is the reduced Green's function for the channel with angular momentum l. The general spherically symmetric Hamiltonian is

$$\hat{H}_0^{(l)} = -\frac{1}{2}\frac{d^2}{dr^2} - \frac{1}{r}\frac{d}{dr} + \frac{l(l+1)}{2r^2} + V(r).$$

Hence, the reduced Green's function satisfies

$$\left\{E+\frac{1}{2}\frac{d^2}{dr^2}+\frac{1}{r}\frac{d}{dr}-\frac{l(l+1)}{2r^2}-V(r)\right\}g_l(r,r';E)=\delta(r-r')/(r')^2. \qquad (32.5)$$

The homogeneous version of this second order equation has two solutions: $j_E(r)$, which is finite at the origin, and $h_E(r)$ that is irregular at the origin but vanishes as $r\to\infty$. The Green's function is given in terms of these as $g_l(r,r';E)=C(E)j_E(r_<)h_E(r_>)$, where *C* is a (possibly r'- dependent) constant. If we integrate the above equation with respect to *r* from r'_- to r'_+ we find that $C^{-1}(E)=\frac{1}{2}(r')^2[j_E(r')h'_E(r')-j'_E(r')h_E(r')]$. This quantity, called the Wronskian, turns out to be independent of r' as can be demonstrated by differentiating and using $j\hat{H}_0h-h\hat{H}_0j=0$. Finally, the polarizability can be calculated as

$$\alpha(\omega)=-\frac{1}{3}\int_0^\infty\int_0^\infty\left\{g_1(r,r';E_0+\omega)+g_1(r,r';E_0-\omega)\right\}R_{1s}(r)R_{1s}(r')r^3(r')^3\,drdr'.$$

In this manner, polarizabilities can be obtained for arbitrary spherical one-electron systems. With a few manipulations

$$\alpha(\omega)=-\frac{1}{3}C(E_0+\omega)\int_0^\infty R_{1s}(x)x^3h_{E_0+\omega}(x)\int_0^x j_{E_0+\omega}(y)R_{1s}(y)y^3\,dydx$$
$$-\frac{1}{3}C(E_0+\omega)\int_0^\infty R_{1s}(x)x^3j_{E_0+\omega}(x)\int_x^\infty h_{E_0+\omega}(y)R_{1s}(y)y^3\,dydx+(\omega\to-\omega).$$

For a hydrogen atom, the regular and irregular solutions to the homogeneous version of Eq.(32.5) are

$$\begin{aligned}j_E(r)&=Ne^{-qr}r^lF(1+l-\tfrac{1}{q},2(1+l);2qr)\\ h_E(r)&=N'e^{-qr}r^lU(1+l-\tfrac{1}{q},2(1+l);2qr),\end{aligned} \qquad (32.6)$$

where *F* and *U* are the Kummer and confluent hypergeometric functions, respectively, and $E=-q^2/2$. The normalization constants *N* and N' always cancel and we may take $N=N'=1$. For $l=1$, the inverse of the Wronskian turns out to be

$$C(E)=2\left\{(r')^2[j_E(r')h'_E(r')-j'_E(r')h_E(r')]\right\}^{-1}=\frac{8q^2}{3}(1-\tfrac{1}{q})\Gamma(-\tfrac{1}{q}).$$

The Gamma function has singularities when the argument is a negative integer, i.e. for $E=-1/(2n^2), n=1,2,3,...$ corresponding to the usual hydrogen energy eigenvalues.

Exercise: Continuum transitions

In a hydrogen atom, the ionized *p*-type states given by Eq.(32.6) with $q = ik$ are

$$R_p(r) = Ne^{-ikr} rF(2 + \tfrac{i}{k}, 4; 2ikr).$$

These states cannot be normalized in the conventional sense. However, if we "enclose" the atom in a large sphere of radius R with infinite barrier walls, we find $k = n\pi / R$ and

$$N^2 = \frac{4\pi k(k^2+1)}{9R(1-e^{-2\pi/k})}.$$

Now, for a 1s state of the form $R_{1s}(r) = 2\zeta_s^{3/2} e^{-\zeta_s r}$ the matrix element Eq.(32.2) becomes

$$\left\langle \varphi_{np} \middle| z \middle| \varphi_{1s} \right\rangle = \frac{2\zeta_s^{3/2} N}{\sqrt{3}} \int_0^R e^{-\zeta_s r} e^{-ikr} F(2 + \tfrac{i}{k}, 4; 2ikr) r^4 dr \ . \tag{32.7}$$

Because R is large we can extend the upper integration limit to infinity. The integral is still complicated but (using an integral representation for F) turns out to be

$$\left\langle \varphi_{np} \middle| z \middle| \varphi_{1s} \right\rangle = \frac{\zeta_s^{3/2} N}{\sqrt{3}} \frac{24(2\zeta_s - 1)}{(k^2 + \zeta_s^2)^3} \left(\frac{\zeta_s - ik}{\zeta_s + ik} \right)^{-i/k}.$$

a) Show that

$$\left(\frac{\zeta_s - ik}{\zeta_s + ik} \right)^{-i/k} = \exp\left\{ -\frac{2}{k} \tan^{-1}(k / \zeta_s) \right\}.$$

The transition energy is $k^2 / 2 - E_{1s}$. Thus, the oscillator strength becomes

$$g_{n0} = \frac{2^8 \pi \zeta_s^3 (2\zeta_s - 1)^2 k(k^2+1)(k^2 - 2E_{1s})}{3(k^2 + \zeta_s^2)^6 R(1 - e^{-2\pi/k})} \exp\left\{ -\frac{4}{k} \tan^{-1}(k / \zeta_s) \right\}.$$

b) Show that in the limit $R \to \infty$

$$\sum_{n=0}^{\infty} f(k) = \sum_{n=0}^{\infty} \left[\frac{R}{\pi k} f(k) \right] k \frac{\pi}{R} \to \int_0^{\infty} f'(k) k dk, \quad f'(k) \equiv \frac{R}{\pi k} f(k).$$

c) Show that this implies Eq.(32.4) if $\zeta_s = 1$.

References

[1] A. Dalgarno and J.T. Lewis, Proc. R. Soc. Lond. A233, 70 (1955).
[2] T. Kinoshita, Phys. Rev. 105, 1490 (1957).
[3] H.A. Bethe and E.E. Salpeter *Quantum mechanics of one- and two-electron atoms* (Springer, Berlin, 1957).
[4] M. Gavrila, Phys. Rev. 163, 147 (1967).

33. Density Response Theory

In this chapter, we explore the fundamental relations between many-body and single-particle response functions. Such relations may be set up using heuristic arguments but can, in fact, be made rigorous within the realm of density-functional theory, more precisely the time-dependent version TDDFT. We will begin the analysis using general response theory based on many-body states and then compare to approximations and establish relations between the two approaches.

We limit the discussion to perturbations given as time-dependent scalar potentials. In a true many-body approach, an electronic system is perturbed only by the external potential $V_{ext}(\vec{r})$. The full quantum mechanical perturbation Hamiltonian is then $-e\sum_i V_{ext}(\vec{r}_i)$ with a sum over all N electrons. The perturbation induces an electron density $n_{ind}(\vec{r})$ at position $\vec{r}$ given by standard response theory. The associated operator is $\sum_i \delta(\vec{r}-\vec{r}_i)$ and so

$$n_{ind}(\vec{r}) = -\sum_{m,n} f_{nm} \frac{\langle n|\sum_i \delta(\vec{r}-\vec{r}_i)|m\rangle\langle m|-e\sum_i V_{ext}(\vec{r}_i)|n\rangle}{E_m - E_n - \hbar\omega}.$$

Here, the states are true many-body eigenstates. The matrix elements of a symmetric function of the form $W = \sum_i w(\vec{r}_i)$ can be simplified as

$$\begin{aligned}\langle n|W|m\rangle &= \int \psi_n^*(\vec{r}_1,\ldots,\vec{r}_N)\sum_i w(\vec{r}_i)\psi_m(\vec{r}_1,\ldots,\vec{r}_N)d^3r_1\ldots d^3r_N \\ &= N\int \psi_n^*(\vec{r}_1,\ldots,\vec{r}_N)w(\vec{r}_1)\psi_m(\vec{r}_1,\ldots,\vec{r}_N)d^3r_1\ldots d^3r_N \\ &= \int \rho_{nm}(\vec{r})w(\vec{r})d^3r, \quad \rho_{nm}(\vec{r}) \equiv N\int \psi_n^*(\vec{r},\vec{r}_2,\ldots,\vec{r}_N)\psi_m(\vec{r},\vec{r}_2,\ldots,\vec{r}_N)d^3r_2\ldots d^3r_N.\end{aligned}$$

Here, anti-symmetry under exchange of electron-arguments has been applied in the second line. Eventually, we see that we can write

$$n_{ind}(\vec{r}) = -e\int \chi(\vec{r},\vec{r}')V_{ext}(\vec{r}')d^3r',$$

where $\chi(\vec{r},\vec{r}')$ is the density response function given by

$$\chi(\vec{r},\vec{r}') = \sum_{m,n} f_{nm} \frac{\rho_{nm}(\vec{r})\rho_{mn}(\vec{r}')}{\hbar\omega - E_m + E_n}. \tag{33.1}$$

It is easily demonstrated that this expression agrees with our previous approach to the optical response calculation. Hence, if the perturbation is due to a constant electric

field $V_{ext}(\vec{r}) = -\vec{\mathcal{E}} \cdot \vec{r}$ we can compute the induced dipole moment as $\vec{d} = -e\int \vec{r} n_{ind}(\vec{r}) d^3 r$ and so

$$\vec{d} = -e^2 \sum_{m,n} f_{nm} \frac{\int \vec{r} \rho_{nm}(\vec{r}) d^3 r \int \vec{r}' \rho_{mn}(\vec{r}') d^3 r'}{\hbar\omega - E_m + E_n} \cdot \vec{\mathcal{E}} .$$

The proportionality factor between d and $\mathcal{E}$ is the polarizability, which clearly agrees with the results of Chapter 2.

Now, as indicated above, we will proceed using heuristic arguments and then exactify the results later. The argument goes as follows. If single-particle rather than many-body states are applied we should try to remedy the error. By ignoring electron-electron repulsion in the Schrödinger equation we miss an important part of the physics. This can be partly fixed, however, by assuming that electrons are perturbed not only by the external potential but also by the induced Coulomb potential due to the other electrons. The basic relation is given by $V_{tot}(\vec{r}) = V_{ext}(\vec{r}) + V_{ind}(\vec{r})$. The density operator is now simply $\delta(\vec{r} - \vec{r}_1)$ and we can write the expression for the induced density in the form

$$n_{ind}(\vec{r}) = -e\int \chi^S(\vec{r}, \vec{r}') V_{tot}(\vec{r}') d^3 r' ,$$

where $\chi^S(\vec{r}, \vec{r}')$ is the single-particle density response function given by

$$\chi^S(\vec{r}, \vec{r}') = \sum_{m,n} f_{nm} \frac{\varphi_n^*(\vec{r}) \varphi_n(\vec{r}') \varphi_m^*(\vec{r}') \varphi_m(\vec{r})}{\hbar\omega - E_m + E_n} . \qquad (33.2)$$

Of course, the energy-eigenvalues in Eqs.(33.1) and (33.2) are many-body and single-particle ones, respectively. At the moment, we will ignore exchange and correlation effects (see App. 3) and only retain the classical part of the induced potential. This is known as the random phase approximation. Classically, potential and density are related via the Coulomb expression

$$V_{ind}(\vec{r}) = \frac{-e}{4\pi\varepsilon_0} \int \frac{n_{ind}(\vec{r}')}{|\vec{r} - \vec{r}'|} d^3 r' .$$

Hence, the induced potential is identified with the time-dependent part of the Hartree potential. In turn,

$$V_{tot}(\vec{r}) = V_{ext}(\vec{r}) + \frac{e^2}{4\pi\varepsilon_0} \int\int \frac{1}{|\vec{r} - \vec{r}'|} \chi^S(\vec{r}', \vec{r}'') V_{tot}(\vec{r}'') d^3 r' d^3 r'' . \qquad (33.3)$$

This is the central integral equation for the total perturbation. For general geometries it is obviously hard to solve.

By definition, the total and external potentials are related via the dielectric matrix

$$V_{tot}(\vec{r}) \equiv \int \varepsilon^{-1}(\vec{r},\vec{r}')V_{ext}(\vec{r}')d^3r' .$$

Hence, Eq.(33.3) shows that

$$\varepsilon(\vec{r},\vec{r}') = \delta(\vec{r}-\vec{r}') - \frac{e^2}{4\pi\varepsilon_0}\int \frac{1}{|\vec{r}-\vec{r}''|}\chi^S(\vec{r}',\vec{r}'')d^3r'' .$$

On the other hand, in the many-body analysis we can write

$$V_{tot}(\vec{r}) = V_{ext}(\vec{r}) + \frac{e^2}{4\pi\varepsilon_0}\int\int \frac{1}{|\vec{r}-\vec{r}'|}\chi(\vec{r}',\vec{r}'')V_{ext}(\vec{r}'')d^3r'd^3r'' . \qquad (33.4)$$

Thus, here

$$\varepsilon^{-1}(\vec{r},\vec{r}') = \delta(\vec{r}-\vec{r}') + \frac{e^2}{4\pi\varepsilon_0}\int \frac{1}{|\vec{r}-\vec{r}''|}\chi(\vec{r}',\vec{r}'')d^3r'' .$$

The two expressions for the dielectric constant establish a link between the two approaches. We can put them together and obtain the following link between the density responses

$$\chi(\vec{r},\vec{r}') = \chi^S(\vec{r},\vec{r}') + \frac{e^2}{4\pi\varepsilon_0}\int \chi^S(\vec{r},\vec{r}'')\frac{1}{|\vec{r}'''-\vec{r}''|}\chi(\vec{r}',\vec{r}''')d^3r''d^3r''' .$$

Alternatively,

$$\chi^{-1}(\vec{r},\vec{r}') = (\chi^S)^{-1}(\vec{r},\vec{r}') - \frac{e^2}{4\pi\varepsilon_0}\frac{1}{|\vec{r}-\vec{r}'|}$$

as can be verified by insertion.

33.1 High-Frequency Limit

The complicated set of equations above can be reduced significantly in the high-frequency limit. In this case, we assume that $\hbar\omega \gg |E_m - E_n|$ for all important transition pairs n, m. Hence, we try to expand

$$\chi^S(\vec{r},\vec{r}') \approx \sum_{m,n} f_{nm}\, \varphi_n^*(\vec{r})\varphi_n(\vec{r}')\varphi_m^*(\vec{r}')\varphi_m(\vec{r})\left(\frac{1}{\hbar\omega}+\frac{E_m-E_n}{(\hbar\omega)^2}\right).$$

The first term is easily seen to vanish as the $f_{nm}=f(E_n)-f(E_m)$ factor averages to zero. The second term can be written

$$\chi^S(\vec{r},\vec{r}') \approx \frac{2}{(\hbar\omega)^2}\sum_{m,n} f(E_n)\varphi_n^*(\vec{r})\varphi_n(\vec{r}')\varphi_m^*(\vec{r}')\varphi_m(\vec{r})\left(E_m-E_n\right).$$

We now exploit the fundamental properties

$$\sum_n \varphi_n^*(\vec{r})\varphi_n(\vec{r}') = \delta(\vec{r}-\vec{r}')$$
$$\sum_n E_n\varphi_n^*(\vec{r})\varphi_n(\vec{r}') = \hat{H}(\vec{r})\delta(\vec{r}-\vec{r}')$$
$$\sum_n f(E_n)\varphi_n^*(\vec{r})\varphi_n(\vec{r}') = \rho(\vec{r},\vec{r}')$$
$$\rho(\vec{r},\vec{r}) = \sum_n f(E_n)\left|\varphi_n(\vec{r})\right|^2 = n(\vec{r})$$

and find

$$\chi^S(\vec{r},\vec{r}') \approx \frac{2}{(\hbar\omega)^2}\left\{\rho(\vec{r},\vec{r}')\hat{H}(\vec{r})\delta(\vec{r}-\vec{r}')-\delta(\vec{r}-\vec{r}')\hat{H}(\vec{r})\rho(\vec{r},\vec{r}')\right\}.$$

However, the only part of $\hat{H}$ that contributes here is the kinetic energy and so

$$\chi^S(\vec{r},\vec{r}') \approx -\frac{1}{m\omega^2}\left\{\rho(\vec{r},\vec{r}')\nabla^2\delta(\vec{r}-\vec{r}')-\delta(\vec{r}-\vec{r}')\nabla^2\rho(\vec{r},\vec{r}')\right\}.$$

If we reinsert this expression into Eq. (33.3) and integrate by parts we find

$$V_{tot}(\vec{r}) = V_{ext}(\vec{r}) - \frac{e^2}{4\pi\varepsilon_0 m\omega^2}\times$$
$$\int\int\left\{\rho(\vec{r}',\vec{r}'')\nabla'^2\frac{1}{|\vec{r}-\vec{r}'|}+2\nabla'\frac{1}{|\vec{r}-\vec{r}'|}\cdot\nabla'\rho(\vec{r}',\vec{r}'')\right\}V_{tot}(\vec{r}'')\delta(\vec{r}'-\vec{r}'')d^3r'd^3r''.$$

The integral over $\vec{r}''$ is easily made due to the delta-function

$$V_{tot}(\vec{r}) = V_{ext}(\vec{r}) - \frac{e^2}{4\pi\varepsilon_0 m\omega^2}\int\left\{n(\vec{r}')\nabla'^2\frac{1}{|\vec{r}-\vec{r}'|}+\nabla'\frac{1}{|\vec{r}-\vec{r}'|}\cdot\nabla' n(\vec{r}')\right\}V_{tot}(\vec{r}')d^3r'.$$

Note that the factor of 2 disappears because we differentiate ρ *before* integrating over $\vec{r}''$. The first term is simplified by recalling the Poisson equation for a point charge and we find

$$V_{tot}(\vec{r}) = V_{ext}(\vec{r}) + \frac{e^2 n(\vec{r})}{\varepsilon_0 m\omega^2} V_{tot}(\vec{r}) - \frac{e^2}{4\pi\varepsilon_0 m\omega^2} \int \left\{ \nabla' \frac{1}{|\vec{r}-\vec{r}'|} \cdot \nabla' n(\vec{r}') \right\} V_{tot}(\vec{r}') d^3 r' . \quad (33.5)$$

The term $e^2 n(\vec{r})/\varepsilon_0 m$ is seen to act as an effective, position dependent squared plasma frequency. In fact, if we introduce the usual local dielectric constant $\varepsilon(\vec{r}) = 1 - e^2 n(\vec{r})/(\varepsilon_0 m\omega^2)$ we can write

$$\varepsilon(\vec{r}) V_{tot}(\vec{r}) = V_{ext}(\vec{r}) + \frac{1}{4\pi} \int \left\{ \nabla' \frac{1}{|\vec{r}-\vec{r}'|} \cdot \nabla' \varepsilon(\vec{r}') \right\} V_{tot}(\vec{r}') d^3 r' . \quad (33.6)$$

It can be shown that this form follows quite generally from usual electrostatics if one starts from

$$V_{tot}(\vec{r}) = V_{ext}(\vec{r}) + \frac{1}{4\pi\varepsilon_0} \int \frac{1}{|\vec{r}-\vec{r}'|} \rho_{ind}(\vec{r}') d^3 r'$$

and writes the induced charge density as $\rho_{ind}(\vec{r}) = \varepsilon_0 \nabla \cdot \left[\varepsilon(\vec{r}) - 1 \right] \nabla V_{tot}(\vec{r})$. Note, however, that really this approach is only valid in the high-frequency limit. In addition, we see that the connection between this local $\varepsilon(\vec{r})$ and the non-local $\varepsilon(\vec{r}, \vec{r}')$ must be

$$\varepsilon(\vec{r}, \vec{r}') = \varepsilon(\vec{r}) \delta(\vec{r}-\vec{r}') - \frac{1}{4\pi} \nabla' \frac{1}{|\vec{r}-\vec{r}'|} \cdot \nabla' \varepsilon(\vec{r}') .$$

33.2 Spherical Symmetry

In cases of spherical symmetry, we have $\varepsilon(\vec{r}) = \varepsilon(r)$, $V_{ext}(\vec{r}) = -r\mathcal{E}_0 \cos\theta$, and $V_{tot}(\vec{r}) = f(r)\cos\theta$. We now use the identity

$$\frac{1}{|\vec{r}-\vec{r}'|} = \sum_{l,m} \frac{4\pi}{2l+1} Y_{lm}^*(\theta,\phi) Y_{lm}(\theta',\phi') \frac{r_<^l}{r_>^{l+1}} ,$$

where $r_< = \min\{r, r'\}$ and $r_> = \max\{r, r'\}$. We then find

$$\varepsilon(r)f(r) = -\mathcal{E}_0 r + \frac{1}{3}\int_0^\infty \frac{d\varepsilon(r')}{dr'}\frac{d}{dr'}\left(\frac{r_<}{r_>^2}\right) f(r')r'^2 dr'$$
$$= -\mathcal{E}_0 r + \frac{1}{3r^2}\int_0^r \frac{d\varepsilon(r')}{dr'} f(r')r'^2 dr' - \frac{2r}{3}\int_r^\infty \frac{d\varepsilon(r')}{dr'} f(r')\frac{1}{r'}dr'.$$

For an ordinary sphere, with $\varepsilon(r) = \varepsilon$ for $r < R$ and $\varepsilon(r) = 1$ for $r > R$ it then follows that

$$f(r) = \begin{cases} -\dfrac{1}{\varepsilon}\mathcal{E}_0 r - \dfrac{2(1-\varepsilon)r}{3\varepsilon R} f(R) & r < R \\ -\mathcal{E}_0 r + \dfrac{(1-\varepsilon)R^2}{3r^2} f(R) & r > R \end{cases}$$

These equations agree at $r = R$, both of them yielding $f(R) = -3\mathcal{E}_0 R/(\varepsilon+2)$. The plasmon resonance condition is therefore $\varepsilon + 2 = 0$. Recalling the polarizability $\alpha = 4\pi\varepsilon_0 R^3(\varepsilon-1)/(\varepsilon+2)$, the solution can be written

$$f(r) = \begin{cases} -\dfrac{3\mathcal{E}_0 r}{\varepsilon+2} & r < R \\ -\mathcal{E}_0 r + \dfrac{\alpha\mathcal{E}_0}{4\pi\varepsilon_0 r^2} & r > R. \end{cases}$$

Hence, the scattered field is proportional to the polarizability, as expected. The induced dipole moment d is generally given by $d = \int z\rho_{ind}(\vec{r})d^3r$. Utilizing the relation between induced charge and potential, it can be shown that this becomes

$$d = \frac{4\pi\varepsilon_0}{3}\int_0^\infty [1-\varepsilon(r)]\frac{d}{dr}\left[f(r)r^2\right]dr.$$

For the ordinary, homogeneous sphere, this is readily shown to be $d = \alpha\mathcal{E}_0$.

33.3 Cylindrical Symmetry

In cases of cylindrical symmetry, we have $n(\vec{r}) = n(r)$, $V_{ext}(\vec{r}) = -r\mathcal{E}_0\cos\phi$, and $V_{tot}(\vec{r}) = f(r)\cos\phi$. We consider formally a piece of a cylindrical wire extending from $z = -L$ to $z = L$. Eventually, the limit $L \to \infty$ should be taken. However, as nothing depends on z we find by integrating Eq. (33.6) and assuming $L \gg r$ that

$$\varepsilon(r)f(r) = -\mathcal{E}_0 r + \frac{1}{4\pi\cos\phi}\int_0^{2\pi}\int_0^\infty \frac{d\varepsilon(r')}{dr'}\frac{d}{dr'}\left(2\ln\frac{2L}{|\vec{r}-\vec{r}'|}\right) f(r')r'dr'\cos\phi' d\phi'.$$

Here, the vectors should be understood as two-dimensional ones. We now use the identity

$$\ln\frac{2L}{|\vec{r}-\vec{r}'|}=\ln\frac{2L}{r_>}+\sum_{m=1}^{\infty}\frac{1}{m}\cos[m(\phi-\phi')]\frac{r_<^m}{r_>^m}.$$

The first terms doesn't contribute due to the angular integral and we are left with

$$\begin{aligned}\varepsilon(r)f(r)&=-\mathcal{E}_0 r+\frac{1}{2}\int_0^{\infty}\frac{d\varepsilon(r')}{dr'}\frac{d}{dr'}\left(\frac{r_<}{r_>}\right)f(r')r'dr'\\&=-\mathcal{E}_0 r+\frac{1}{2r}\int_0^{r}\frac{d\varepsilon(r')}{dr'}f(r')r'dr'-\frac{r}{2}\int_r^{\infty}\frac{d\varepsilon(r')}{dr'}f(r')\frac{1}{r'}dr'.\end{aligned}$$

For an ordinary cylinder, with $\varepsilon(r)=\varepsilon$ for $r<R$ and $\varepsilon(r)=1$ for $r>R$ it then follows that

$$f(r)=\begin{cases}-\dfrac{1}{\varepsilon}\mathcal{E}_0 r-\dfrac{(1-\varepsilon)r}{2\varepsilon R}f(R) & r<R\\ -\mathcal{E}_0 r+\dfrac{(1-\varepsilon)R}{2r}f(R) & r>R.\end{cases}$$

These equations agree at $r=R$, both of them yielding $f(R)=-2\mathcal{E}_0 R/(\varepsilon+1)$. In this case, the plasmon resonance condition becomes $\varepsilon+1=0$.

33.4 Planar Symmetry

In the simple case of planar symmetry so that $\varepsilon(\vec{r})=\varepsilon(x)$ and $V_{ext}(\vec{r})=-x\mathcal{E}_0$, we find the condition for the self-consistent potential $V(x)$ after integrating over y and z

$$\varepsilon(x)V(x)=-\mathcal{E}_0 x+\frac{1}{2}\int_{-\infty}^{\infty}\mathrm{sgn}(x-x')\frac{d\varepsilon(x')}{dx'}V(x')dx'.$$

Again, for the classical case $\varepsilon(x)=\varepsilon$ for $|x|<a$ and $\varepsilon(x)=1$ for $|x|>a$, the solution is

$$V(x)=-\mathcal{E}_0\begin{cases}\dfrac{x}{\varepsilon} & |x|<a\\ x-a\,\mathrm{sgn}(x)\dfrac{\varepsilon-1}{\varepsilon} & |x|>a.\end{cases}$$

The plasmon resonance condition therefore becomes $\varepsilon=0$.

33.5 Time-Dependent Density-Functional Theory

If we wish to go beyond the random phase approximation, we need to go back to the relation between density and potential

$$n_{ind}(\vec{r}) = -e\int \chi^S(\vec{r},\vec{r}')V_{tot}(\vec{r}')d^3r' .$$

To linear order, this relation still holds but we now add the exchange-correlation V_{xc} part to V_{tot}. However, it is important to note that it is only the small time-dependent portion of V_{xc} that should be considered. We know that V_{xc} is a functional of density n and this also holds in the time-dependent case. Hence, to linear order we can write

$$V_{xc} = V_{xc}[n_0 + n_{ind},\omega] \approx V_{xc}[n_0] + \int f_{xc}(\vec{r},\vec{r}',\omega)n_{ind}(\vec{r}')d^3r', \quad f_{xc}(\vec{r},\vec{r}',\omega) = \left.\frac{\delta V_{xc}(\vec{r})}{\delta n(\vec{r}')}\right|_{n=n_0} .$$

Here, $f_{xc}(\vec{r},\vec{r}',\omega)$ is a frequency-dependent xc kernel that depends on the ground state density n_0 only. Adding the induced part of V_{xc} we now find that

$$V_{tot}(\vec{r}) = V_{ext}(\vec{r}) + \int\int f_{Hxc}(\vec{r},\vec{r}',\omega)\chi^S(\vec{r}',\vec{r}'')V_{tot}(\vec{r}'')d^3r'd^3r'', \qquad (33.7)$$

where

$$f_{Hxc}(\vec{r},\vec{r}',\omega) \equiv \frac{e^2}{4\pi\varepsilon_0}\frac{1}{|\vec{r}-\vec{r}'|} + f_{xc}(\vec{r},\vec{r}',\omega).$$

This integral equation is the foundation of time-dependent density-functional theory (TDDFT). Since we don't even know V_{xc} in the static case, we obviously don't know $f_{xc}(\vec{r},\vec{r}',\omega)$ either. However, useful approximations can be constructed. Among these, the adiabatic approximation, in which the frequency dependence of f_{xc} is ignored, is particularly important as it allows us to construct the kernel from excellent, well-tested approximations to the static V_{xc}. The equation determining the non-local dielectric function now generalizes to

$$\varepsilon(\vec{r},\vec{r}') = \delta(\vec{r}-\vec{r}') - \int f_{Hxc}(\vec{r},\vec{r}'',\omega)\chi^S(\vec{r}',\vec{r}'')d^3r'' \qquad (33.8)$$

and the relation between χ's becomes

$$\chi^{-1}(\vec{r},\vec{r}') = (\chi^S)^{-1}(\vec{r},\vec{r}') - f_{Hxc}(\vec{r},\vec{r}',\omega).$$

At this point, we wish to establish the resonance condition given by the so-called Casida equation [1,2]. The resonances occur at certain discrete frequencies $\hbar\omega = \Omega$. To find them, we combine Eq.(33.2) with Eq.(33.7) in the case of no external potential so that

$$V_{tot}(\vec{r}) = \sum_{m,n} f_{nm} \frac{\langle \varphi_m | V_{tot} | \varphi_n \rangle}{\Omega - E_{mn}} \int f_{Hxc}(\vec{r}, \vec{r}', \omega) \varphi_n^*(\vec{r}') \varphi_m(\vec{r}') d^3 r' .$$

Taking matrix elements means that

$$\langle \varphi_p | V_{tot} | \varphi_q \rangle = \sum_{m,n} f_{nm} K_{pq,mn} \frac{\langle \varphi_m | V_{tot} | \varphi_n \rangle}{\Omega - E_{mn}}, K_{pq,mn} = \int \int \varphi_p^*(\vec{r}) \varphi_q(\vec{r}) f_{Hxc}(\vec{r}, \vec{r}', \omega) \varphi_n^*(\vec{r}') \varphi_m(\vec{r}') d^3 r d^3 r'.$$

If we now define $\beta_{mn} \equiv \langle \varphi_m | V_{tot} | \varphi_n \rangle / (\Omega - E_{mn})$ we see that

$$\left[\Omega - E_{pq}\right] \beta_{pq} = \sum_{m,n} f_{nm} K_{pq,mn} \beta_{mn} .$$

Hence, the resonances $\hbar\omega$ are eigenvalues of a matrix with entries $f_{nm} K_{pq,mn} + E_{pq} \delta_{pq,mn}$. The problem can be further reformulated in the low-temperature limit, where the Fermi factors are either 0 or 1 so that

$$\left[\Omega - E_{pq}\right] \beta_{pq} = \sum_{m \in c, n \in v} \left\{ K_{pq,mn} \beta_{mn} - K_{pq,nm} \beta_{nm} \right\}.$$

Here, we relabeled in the last term and so the summations are restricted to empty m states and occupied n states. An additional equation is obtained by interchanging p and q

$$\left[\Omega + E_{pq}\right] \beta_{qp} = \sum_{m \in c, n \in v} \left\{ K_{qp,mn} \beta_{mn} - K_{qp,nm} \beta_{nm} \right\}.$$

These equations can be combined into a single matrix eigenvalue problem. However, if we restrict the analysis to real-valued wave functions we have $K_{pq,mn} = K_{pq,nm} = K_{qp,mn} = K_{qp,nm}$. Hence, if we multiply the first by $\Omega + E_{pq}$ and the last by $\Omega - E_{pq}$ and, subsequently, subtract the two we find

$$\left[\Omega^2 - E_{pq}^2\right] \left\{ \beta_{pq} - \beta_{qp} \right\} = 2 E_{pq} \sum_{m \in c, n \in v} K_{pq,mn} \left\{ \beta_{mn} - \beta_{nm} \right\}.$$

This expression can easily be converted into a Hermetian eigenvalue problem by means of the new set of unknowns $\overline{\beta}_{mn} \equiv \left\{ \beta_{mn} - \beta_{nm} \right\} / \sqrt{E_{mn}}$ so that

$$\Omega^2 \overline{\beta}_{pq} = \sum_{m\in c, n\in v} \left\{ E_{pq}^2 \delta_{pq,mn} + F_{pq,mn} \right\} \overline{\beta}_{mn}, \quad F_{pq,mn} = 2\sqrt{E_{pq}} K_{pq,mn} \sqrt{E_{mn}} \, .$$

Hence, Ω^2 is an eigenvalue of the Casida matrix $\overleftrightarrow{C}$ with elements $c_{pq,mn} = E_{pq}^2 \delta_{pq,mn} + F_{pq,mn}$. We will now specialize to the important case of spin-balanced systems. In this case, similarly to Chapter 18, the ground state is of the form

$$|0\rangle = |(v_1\uparrow),(v_1\downarrow),.....(v_N\uparrow),(v_N\downarrow)|$$

with all single-electron valence states occupied by both spin-up and –down electrons. Similarly, we consider two types of singly-excited states

$$|(v_n\uparrow)\to(c_m\uparrow)\rangle = |.....,(c_m\uparrow),(v_n\downarrow),.....|$$
$$|(v_n\downarrow)\to(c_m\downarrow)\rangle = |.....,(v_n\uparrow),(c_m\downarrow),.....|.$$

From these, we form the singlet $(+)$ and triplet $(-)$ combinations

$$|v_n \to c_m\rangle_{\pm} \equiv \left\{ |(v_n\uparrow)\to(c_m\uparrow)\rangle \pm |(v_n\downarrow)\to(c_m\downarrow)\rangle \right\} / \sqrt{2} \, .$$

The corresponding expansion coefficients are denoted $\overline{\beta}_{mn}^{\pm}$ and the matrix problem becomes

$$\left[\Omega^2 - E_{pq}^2\right] \overline{\beta}_{pq}^{\pm} = \sum_{m\in c, n\in v} F_{pq,mn}^{\pm} \overline{\beta}_{mn}^{\pm}, \quad F_{pq,mn}^{\pm} = F_{pq,mn}^{\uparrow\uparrow} \pm F_{pq,mn}^{\uparrow\downarrow} \, .$$

Here, the kernel is determined from

$$K_{pq,mn}^{\sigma\sigma'} = \int\int \varphi_{c_p}^*(\vec{r}) \varphi_{v_q}(\vec{r}) \left\{ \frac{e^2}{4\pi\varepsilon_0} \frac{1}{|\vec{r}-\vec{r}'|} + f_{xc,\sigma\sigma'}(\vec{r},\vec{r}',\Omega) \right\} \varphi_{v_n}^*(\vec{r}') \varphi_{c_m}(\vec{r}') d^3r d^3r' \, .$$

In the adiabatic approximation, the frequency-dependence of f_{xc} is ignored. Hence, the ground state expression for the density-dependence of E_{xc} is assumed to be valid even for time-dependent density perturbations. If, moreover, the local spin-density approximation (LSDA) is adopted $E_{xc} = \int e_{xc}[n_\uparrow(\vec{r}), n_\downarrow(\vec{r})] \, d^3r$ it follows that

$$f_{xc,\sigma\sigma'}(\vec{r},\vec{r}',\Omega) \approx \frac{\partial^2 e_{xc}[n_\uparrow(\vec{r}), n_\downarrow(\vec{r})]}{\partial n_\sigma(\vec{r}) \partial n_{\sigma'}(\vec{r})} \delta(\vec{r}-\vec{r}') \, .$$

Eventually, the kernels become $F_{pq,mn}^{\pm} = 2\sqrt{E_{pq}} K_{pq,mn}^{\pm} \sqrt{E_{mn}}$ with

$$K^{+}_{pq,mn} = 2\int\int \varphi^{*}_{c_p}(\vec{r})\varphi_{v_q}(\vec{r})\frac{e^2}{4\pi\varepsilon_0}\frac{1}{|\vec{r}-\vec{r}'|}\varphi^{*}_{v_n}(\vec{r}')\varphi_{c_m}(\vec{r}')d^3rd^3r'$$
$$+\int \varphi^{*}_{c_p}(\vec{r})\varphi_{v_q}(\vec{r})\left\{\frac{\partial^2 e_{xc}[n_\uparrow(\vec{r}),n_\downarrow(\vec{r})]}{\partial n^2_\uparrow(\vec{r})}+\frac{\partial^2 e_{xc}[n_\uparrow(\vec{r}),n_\downarrow(\vec{r})]}{\partial n_\uparrow(\vec{r})\partial n_\downarrow(\vec{r})}\right\}\varphi^{*}_{v_n}(\vec{r})\varphi_{c_m}(\vec{r})d^3r,$$

and

$$K^{-}_{pq,mn} = \int \varphi^{*}_{c_p}(\vec{r})\varphi_{v_q}(\vec{r})\left\{\frac{\partial^2 e_{xc}[n_\uparrow(\vec{r}),n_\downarrow(\vec{r})]}{\partial n^2_\uparrow(\vec{r})}-\frac{\partial^2 e_{xc}[n_\uparrow(\vec{r}),n_\downarrow(\vec{r})]}{\partial n_\uparrow(\vec{r})\partial n_\downarrow(\vec{r})}\right\}\varphi^{*}_{v_n}(\vec{r})\varphi_{c_m}(\vec{r})d^3r.$$

It is noted that if coupling between different singly-excited states is neglected, the resonances are found analytically as

$$\Omega \approx \sqrt{E^2_{pq}+F^{\pm}_{pq,pq}}\,. \tag{33.9}$$

If, moreover, $|K^{\pm}_{pq,pq}| \ll E_{pq}$ we find $\Omega \approx E_{pq}+K^{\pm}_{pq,pq}$.

Now, we may go back to the case of a non-vanishing external potential at an arbitrary frequency ω. In this case, it can be shown that

$$\overline{\beta}_{mn} = 2\sum_{pq} R_{mn,pq}\sqrt{E_{pq}}\left\langle\varphi_p\middle|V_{ext}\middle|\varphi_q\right\rangle,$$

where the *R* matrix is defined as the inverse $\overleftrightarrow{R} = [(\hbar\omega)^2\overleftrightarrow{I}-\overleftrightarrow{C}]^{-1}$. Above, we discussed eigenvalues Ω_r and eigenvectors $|\lambda_r\rangle$ of the Casida matrix $\overleftrightarrow{C}$. Using the so-called spectral representation it follows that

$$\overleftrightarrow{R} = \sum_r \frac{|\lambda_r\rangle\langle\lambda_r|}{(\hbar\omega)^2-\Omega^2_r}\,.$$

If the external potential is taken as simply $V_{ext}(\vec{r}) = z$ the dipole moment is identical to the polarizability $\alpha(\omega)$. It is then easily seen that $\alpha(\omega) = -2\sum_{mn}\sqrt{E_{mn}}\left\langle\varphi_m\middle|z\middle|\varphi_n\right\rangle\overline{\beta}_{mn}$. Hence, we can finally write

$$\alpha(\omega) = \sum_r \frac{g_r}{\Omega^2_r-(\hbar\omega)^2},\quad g_r = 2\left|\sum_{mn}\lambda^r_{mn}\sqrt{E_{mn}}\left\langle\varphi_m\middle|z\middle|\varphi_n\right\rangle\right|^2,$$

where g_r is the oscillator strength.

Exercise: Application of LSDA to atoms

In the LSDA, the spin-dependence is given by (c.f. Eqs.(A4.4) and (A4.5))

$$e_{xc}[n_\uparrow(\vec{r}),n_\downarrow(\vec{r})]=\varepsilon_{xc}[n_\uparrow(\vec{r}),n_\downarrow(\vec{r})]n(\vec{r}),\quad \varepsilon_{xc}[n_\uparrow(\vec{r}),n_\downarrow(\vec{r})]=\varepsilon_{xc}^{U}+f\cdot\left(\varepsilon_{xc}^{P}-\varepsilon_{xc}^{U}\right),$$

where

$$\varepsilon_{xc}^{U}=\varepsilon_{xc}[\tfrac{1}{2}n(\vec{r}),\tfrac{1}{2}n(\vec{r})],\quad \varepsilon_{xc}^{P}=\varepsilon_{xc}[n(\vec{r}),0],\quad f[n_\uparrow(\vec{r}),n_\downarrow(\vec{r})]=\frac{\left(\frac{n_\uparrow(\vec{r})}{n(\vec{r})}\right)^{4/3}+\left(\frac{n_\downarrow(\vec{r})}{n(\vec{r})}\right)^{4/3}-2^{-1/3}}{1-2^{-1/3}}.$$

a) Show by differentiation that

$$\left.\frac{\partial f}{\partial n_\uparrow}\right|_{n_\uparrow=n_\downarrow=n/2}=\left.\frac{\partial f}{\partial n_\downarrow}\right|_{n_\uparrow=n_\downarrow=n/2}=0\quad\text{and}\quad \left.\frac{\partial^2 f}{\partial n_\uparrow^2}\right|_{n_\uparrow=n_\downarrow=n/2}=-\left.\frac{\partial^2 f}{\partial n_\uparrow\partial n_\downarrow}\right|_{n_\uparrow=n_\downarrow=n/2}=\frac{4}{9n^2(2^{1/3}-1)}.$$

b) Use these results to show that

$$\left.\frac{\partial^2 e_{xc}}{\partial n_\uparrow^2}\right|_{n_\uparrow=n_\downarrow=n/2}=\frac{\partial^2(n\varepsilon_{xc}^{U})}{\partial n^2}+\frac{4(\varepsilon_{xc}^{P}-\varepsilon_{xc}^{U})}{9n(2^{1/3}-1)},\quad \left.\frac{\partial^2 e_{xc}}{\partial n_\uparrow\partial n_\downarrow}\right|_{n_\uparrow=n_\downarrow=n/2}=\frac{\partial^2(n\varepsilon_{xc}^{U})}{\partial n^2}-\frac{4(\varepsilon_{xc}^{P}-\varepsilon_{xc}^{U})}{9n(2^{1/3}-1)}.$$

For the exchange part, $\varepsilon_x^U=-\frac{3}{4}\left(\frac{3n}{\pi}\right)^{1/3}$ and $\varepsilon_x^P=2^{1/3}\varepsilon_x^U$ in atomic units.

c) Show that

$$\frac{\partial^2(n\varepsilon_x^U)}{\partial n^2}=\frac{4(\varepsilon_x^P-\varepsilon_x^U)}{9n(2^{1/3}-1)}=-\frac{1}{(9\pi)^{1/3}n^{2/3}}.$$

Explicit formulas for the correlation part in the Perdew-Zunger parameterization (slightly modified to ensure continuous second derivatives) are given in Ref. [2].

Implementation of the above framework is relatively straight-forward for closed-shell atoms with a spherically symmetric ground state density. Here, we investigate excitations of an electron from the outermost occupied *s*-orbital to the lowest empty *p*-orbital. We adopt the restricted model of App. 3 and require spherical symmetry for the excited states as well. Moreover, we first use Eq.(33.9), i.e. ignore coupling between excited configurations. In the table below, some results for closed-shell atoms are shown.

Atom	$E_p - E_s$ [eV]	Singlet E_S [eV]	Triplet E_T [eV]	Exp. [eV]
Be	3.50	5.08	2.58	5.28/2.72
Mg	3.40	4.57	2.87	4.34/2.72
Ca	2.40	3.39	2.00	2.94/1.89

It is seen that the corrections improve agreement with experiment. If, next, coupling between the 10 lowest excitations is neglected, we find the results of the following table.

Atom	$E_p - E_s$ [eV]	Singlet E_S [eV]	Triplet E_T [eV]	Exp. [eV]
Be	3.50	4.84	2.51	5.28/2.72
Mg	3.40	4.25	2.82	4.34/2.72
Ca	2.40	3.16	1.97	2.94/1.89

It is noted that results for Be actually get worse but the other cases improve.

References

[1] M.E. Casida, in *Recent Advances in Density-Functional Methods, Part 1*, edited by D.P. Chong (World Scientific, Singapore, 1995).
[2] I. Vasiliev, S. Ögüt, and J.R. Chelikowsky, Phys. Rev. B65, 115416 (2002).

34. Screening

In this chapter, we look at screening in periodic systems. Screening is essential for a correct understanding of many-body effects in crystals. In particular, excitons in semiconductors require a careful treatment of the screened Coulomb interaction. Moreover, doping leads to the presence of free carriers that efficiently screen interactions between localized and mobile charges. We apply the density response formalism developed in Chapter 33 for our purpose. We wish to apply the formalism to the periodic situation, i.e. to the response of crystals. In that case, we write the wave functions on the form

$$\varphi_n(\vec{r}) = \frac{1}{\sqrt{\Omega}} u_{n,\vec{k}}(\vec{r}) e^{i\vec{k}\cdot\vec{r}}, \quad \varphi_m(\vec{r}) = \frac{1}{\sqrt{\Omega}} u_{m,\vec{k}+\vec{q}}(\vec{r}) e^{i(\vec{k}+\vec{q})\cdot\vec{r}},$$

where the u's are lattice periodic and n is a band index. Noting that the product of u's is also lattice periodic, we introduce $u^*_{n,\vec{k}}(\vec{r}) u_{m,\vec{k}+\vec{q}}(\vec{r}) = \sum_{\vec{G}} \Phi_{nm\vec{k}\vec{q}}(\vec{G}) e^{i\vec{G}\cdot\vec{r}}$. It follows that

$$\chi^S(\vec{r},\vec{r}') = \frac{1}{\Omega} \sum_{\vec{G},\vec{G}',\vec{q}} \chi^S_{\vec{G},\vec{G}'}(\vec{q},\omega) e^{i(\vec{G}\cdot\vec{r}-\vec{G}'\cdot\vec{r}')} e^{i\vec{q}\cdot(\vec{r}-\vec{r}')},$$

where (after spin summation)

$$\chi^S_{\vec{G},\vec{G}'}(\vec{q},\omega) = \frac{2}{\Omega} \sum_{n,m,\vec{k}} \left[f(E_{n,\vec{k}}) - f(E_{m,\vec{k}+\vec{q}}) \right] \frac{\Phi_{nm\vec{k}\vec{q}}(\vec{G}) \Phi^*_{nm\vec{k}\vec{q}}(\vec{G}')}{\hbar\omega + E_{n,\vec{k}} - E_{m,\vec{k}+\vec{q}}}. \tag{34.1}$$

In should be noted that $\vec{k}$ and $\vec{q}$ both lie within the Brillouin zone. We would like to show that, upon averaging over a macroscopic volume, the expression Eq.(34.1) agrees with the usual dielectric constant Eq. (15.3) if the many-body correction f_{xc} is ignored. In fact, it can be shown that one should average the *inverse* of the dielectric function. This complication, however, will be ignored here. We use the Fourier expansion of the Coulomb interaction

$$\frac{1}{4\pi r} = \frac{1}{\Omega} \sum_{\vec{G},\vec{p}} \frac{e^{-i(\vec{p}+\vec{G})\cdot\vec{r}}}{|\vec{p}+\vec{G}|^2}.$$

Here, again, $\vec{p}$ is inside the Brillouin zone. Integrating, we find that only $\vec{p} = \vec{q}$ survives and the RPA response becomes

$$\varepsilon(\vec{r},\vec{r}') = \delta(\vec{r}-\vec{r}') - \frac{e^2}{\varepsilon_0 \Omega} \sum_{\vec{G},\vec{G}',\vec{q}} \frac{\chi^S_{\vec{G},\vec{G}'}(\vec{q},\omega)}{|\vec{q}+\vec{G}|^2} e^{i(\vec{G}\cdot\vec{r}-\vec{G}'\cdot\vec{r}')} e^{i\vec{q}\cdot(\vec{r}-\vec{r}')} \tag{34.2}$$

We now calculate the Fourier transform from the integral

$$\varepsilon(\vec{q},\omega)=\frac{1}{\Omega}\iint \varepsilon(\vec{r},\vec{r}')e^{-i\vec{q}\cdot(\vec{r}-\vec{r}')}d^3rd^3r' .$$

Assuming that $\vec{q}$ is inside the Brillouin zone, we find

$$\varepsilon(\vec{q},\omega)=1-\frac{e^2}{\varepsilon_0 q^2}\chi^S_{00}(\vec{q},\omega).$$

The long-range spatial average in then $\overline{\varepsilon}(\omega)=\varepsilon(0,\omega)=1-\frac{e^2}{\varepsilon_0}\lim_{\vec{q}\to 0}\chi^S_{00}(\vec{q},\omega)/q^2$. Now, for simplicity we will consider an intrinsic semiconductor material with a finite gap separating filled and empty bands. Thus,

$$\chi^S_{\vec{G},\vec{G}'}(\vec{q},\omega)=\frac{4}{\Omega}\sum_{n\in v,m\in c,\vec{k}}\frac{(E_{m,\vec{k}+\vec{q}}-E_{n,\vec{k}})\Phi_{nm\vec{k}\vec{q}}(\vec{G})\Phi^*_{nm\vec{k}\vec{q}}(\vec{G}')}{\hbar^2\omega^2-(E_{m,\vec{k}+\vec{q}}-E_{n,\vec{k}})^2}.$$

Next, $\Phi_{nm\vec{k}\vec{q}}(\vec{G})=\Omega^{-1}_{UC}\int_{UC}u^*_{n,\vec{k}}(\vec{r})u_{m,\vec{k}+\vec{q}}(\vec{r})e^{-i\vec{G}\cdot\vec{r}}d^3r$ and using Eq.(3.7) leads to

$$\begin{aligned}(E_{m,\vec{k}+\vec{q}}-E_{n,\vec{k}})\Phi_{nm\vec{k}\vec{q}}(0)&=\Omega^{-1}_{UC}\int_{UC}u^*_{n,\vec{k}}(\vec{r})\left[\hat{h}_0(\vec{k}+\vec{q})-\hat{h}_0(\vec{k})\right]u_{m,\vec{k}+\vec{q}}(\vec{r})d^3r\\&\simeq\frac{\hbar}{m}\vec{q}\cdot\left\langle u_{n,\vec{k}}\middle|\hat{\vec{p}}\middle|u_{m,\vec{k}}\right\rangle.\end{aligned}$$

Here, an expansion in $\vec{q}$ was applied. It finally follows that

$$\overline{\varepsilon}(\omega)=1-\frac{4e^2\hbar^2}{\varepsilon_0 m^2\Omega}\sum_{n\in v,m\in c,\vec{k}}\frac{\left|\left\langle u_{n,\vec{k}}\middle|\hat{p}_z\middle|u_{m,\vec{k}}\right\rangle\right|^2}{(E_{m,\vec{k}}-E_{n,\vec{k}})\left[\hbar^2\omega^2-(E_{m,\vec{k}}-E_{n,\vec{k}})^2\right]}.$$

This expression agrees with our previous result in Eq. (15.3).

34.1 Lindhard Functions

Having established this agreement, we will now turn to simplified treatments. To this end, we note that for small $\vec{q}$ we have $\Phi_{nm\vec{k}\vec{q}}(\vec{G})\approx\delta_{nm}\delta_{\vec{G},0}$. If only a single band contributes significantly, this means that $\chi^S_{00}(\vec{q},\omega)\approx L(\vec{q},\omega)$, where L is the Lindhard function, defined by

$$L(\vec{q},\omega)=\frac{2}{\Omega}\sum_{\vec{k}}\frac{f(E_{\vec{k}})-f(E_{\vec{k}+\vec{q}})}{\hbar\omega+E_{\vec{k}}-E_{\vec{k}+\vec{q}}}=\frac{2}{\Omega}\sum_{\vec{k}}f(E_{\vec{k}})\left\{\frac{1}{\hbar\omega+E_{\vec{k}}-E_{\vec{k}+\vec{q}}}-\frac{1}{\hbar\omega+E_{\vec{k}-\vec{q}}-E_{\vec{k}}}\right\}.$$

Using the effective mass approximation and assuming zero temperature, we find for 3D materials after introducing $w=2m_*\omega/\hbar$

$$L_{3D}(\vec{q},\omega)=\frac{m_*}{2\pi^3\hbar^2}\int f(\tfrac{\hbar^2k^2}{2m_*})\left\{\frac{1}{k^2-|\vec{k}+\vec{q}|^2+w}+\frac{1}{k^2-|\vec{k}-\vec{q}|^2-w}\right\}d^3k=-\frac{m_*k_F}{\pi^2\hbar^2}F_{3D}\left(\frac{q}{k_F},\frac{w}{k_F^2}\right),$$

$$F_{3D}(x,y)=\frac{1}{2}-\frac{x^4+y^2-2x^2(y+2)}{16x^3}\ln\frac{x^2+2x-y}{x^2-2x-y}-\frac{x^4+y^2+2x^2(y-2)}{16x^3}\ln\frac{x^2+2x+y}{x^2-2x+y}.$$

Here, k_F is the Fermi wave vector defined via $E_F=\hbar^2k_F^2/2m_*$. The function F_{3D} is plotted in Fig. 34.1.

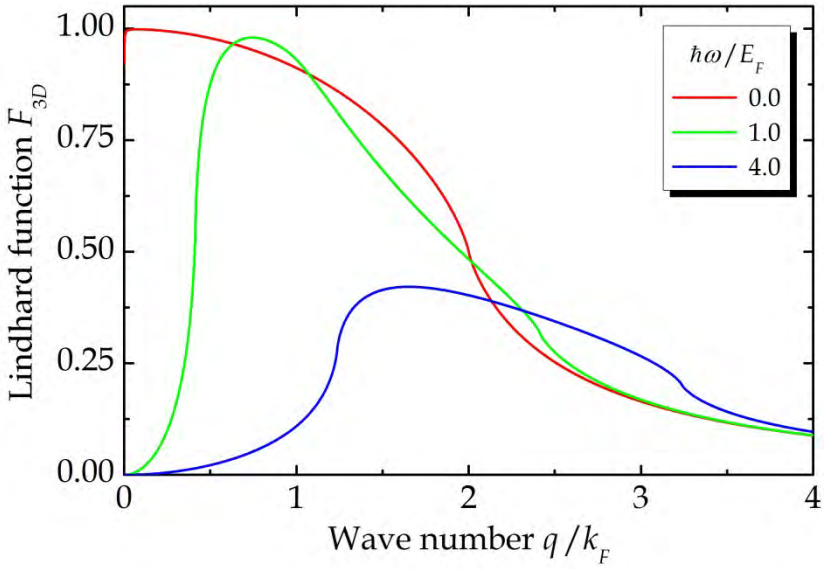


Figure 34.1. Lindhard function for the 3D case for different frequencies.

The static limit of the Lindhard function is

$$L_{3D}(\vec{q},0)=-\frac{m_*k_F}{2\pi^2\hbar^2}\left\{1+\frac{4k_F^2-q^2}{4qk_F}\ln\left|\frac{q+2k_F}{q-2k_F}\right|\right\}=-\frac{m_*k_F}{\pi^2\hbar^2}\left\{1-\frac{q^2}{12k_F^2}-\frac{q^4}{240k_F^4}+O(q^6)\right\}.$$

Also, for small q but finite ω it turns out that

$$L_{3D}(\vec{q},\omega)\approx\frac{m_*}{\pi^2\hbar^2}\frac{4k_F^3q^2}{3w^2}=\frac{nq^2}{m_*\omega^2},\quad n=\frac{k_F^3}{3\pi^2}.$$

Hence, the usual Drude formula $\bar{\varepsilon}(\omega)=1-\frac{e^2 n}{\varepsilon_0 m_* \omega^2}$ for a material with electron density n is recovered in this limit. For small q, the static dielectric constant becomes

$$\varepsilon(\vec{q},0)\approx 1+\frac{q_{3D}^2}{q^2},\quad q_{3D}^2=\frac{e^2 m_* k_F}{\pi^2\varepsilon_0\hbar^2}=\frac{e^2 D(E_F)}{\varepsilon_0},\quad D(E_F)=\frac{m_* k_F}{\pi^2\hbar^2},$$

where q_{3D} is the 3D Thomas-Fermi wave number and $D(E_F)$ is the density of states at the Fermi level. The relation between q_{3D} and $D(E_F)$ is, in fact, valid even without the effective mass approximation. The associated screened Coulomb interaction, also called the Yukawa potential, is then

$$V(\vec{r})=\frac{e^2}{8\pi^3\varepsilon_0}\int\frac{e^{i\vec{q}\cdot\vec{r}}}{q^2\varepsilon(\vec{q},0)}d^3q=\frac{e^2}{4\pi\varepsilon_0 r}e^{-q_{3D}r}. \tag{34.3}$$

In two dimensions, the Lindhard function becomes

$$L_{2D}(\vec{q},\omega)=\frac{m_*}{\pi^2\hbar^2}\int\left\{\frac{f(\frac{\hbar^2k^2}{2m_*})}{k^2-|\vec{k}+\vec{q}|^2+w}+\frac{f(\frac{\hbar^2k^2}{2m_*})}{k^2-|\vec{k}-\vec{q}|^2-w}\right\}d^2k=-\frac{m_*}{\pi\hbar^2}F_{2D}\left(\frac{q}{k_F},\frac{w}{k_F^2}\right),$$
$$F_{2D}(x,y)=1-\frac{1}{2}\sqrt{1+\frac{y^2}{x^4}-\frac{2(y+2)}{x^2}}-\frac{1}{2}\sqrt{1+\frac{y^2}{x^4}+\frac{2(y-2)}{x^2}}. \tag{34.4}$$

This function is illustrated in Fig. 34.2.

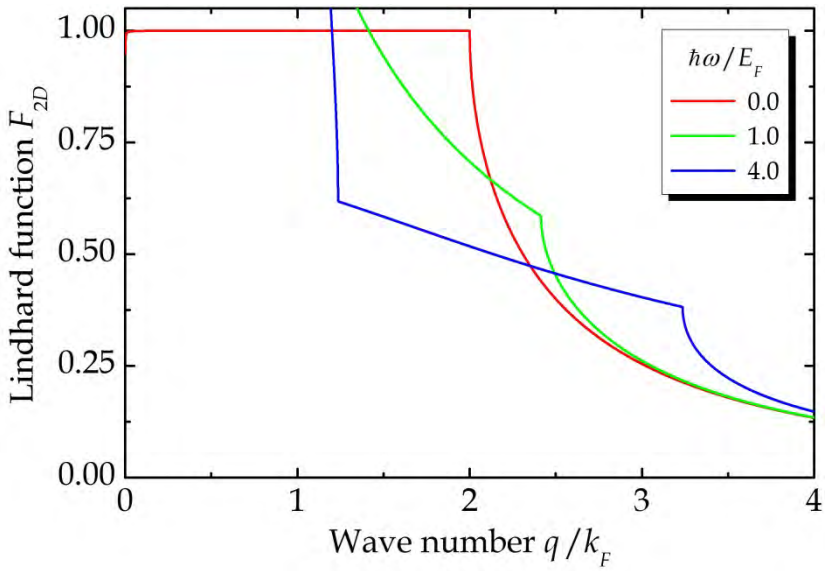


Figure 34.2. Lindhard function for the 2D case for different frequencies.

In two dimensions, the static limit of the Lindhard function is

$$L_{2D}(\vec{q},0) = -\frac{m_*}{\pi\hbar^2}\left\{1-\theta(q-2k_F)\sqrt{1-\frac{4k_F^2}{q^2}}\right\}.$$

Again, the prefactor is generally $-D(E_F)$ (calculated per area). In the last section, we will turn to a practical application of this result. The 2D Coulomb interaction $e^2/(2q\varepsilon_0)$ means that at small q

$$\varepsilon(\vec{q},0) \approx 1+\frac{q_{2D}}{q}, \quad q_{2D} = \frac{e^2 m_*}{2\pi\varepsilon_0\hbar^2},$$

where q_{2D} is the 2D Thomas-Fermi wave number. In the "plasmon-pole approximation", the inverse of the full space and frequency dependent response is approximated by

$$\frac{1}{\varepsilon(\vec{q},\omega)} \approx 1+\frac{\omega_p^2}{\omega^2-\omega_q^2}.$$

The pole ω_q is located by requiring the correct static behavior, so that $\omega_q^2 = \omega_p^2(1+q^2/q_{3D}^2)$ and $\omega_q^2 = \omega_p^2(1+q/q_{2D})$ in 3D and 2D, respectively.

34.2 Two-dimensional Coulomb Interaction

We will attempt to describe the special circumstance governing excitons in two-dimensional materials, such as the case illustrated in Fig. 34.3. Here, charges inside the layer interact via a Coulomb potential screened by the layer itself but, also, charges in the surrounding sub- and superstrates. Intuitively, one may imagine field-lines connecting the the charges. If the charges are far apart, the field-lines will predominantly permeate the surrounding dielectrics and only these will contribute to screening. Conversely, for short separations, charges in the 2D materials will contribute to screening.

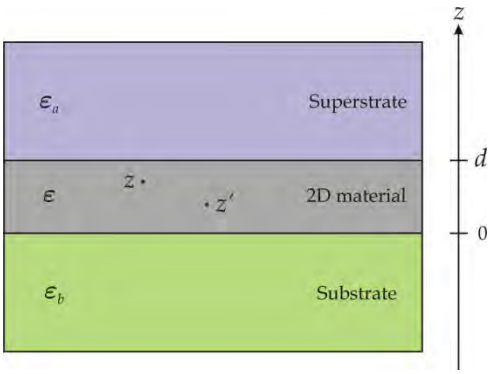


Figure 34.3. Layered geometry used in describing the screened Coulomb interaction between charges in encapsulated 2D materials.

We consider charged carriers in the general geometry illustrated in Fig. 34.3. Here, the 2D material is placed between materials with dielectric constants ε_a and ε_b above and below, respectively. In addition, the 2D layer is modelled as a homogeneous layer with dielectric constant ε and thickness d. As a starting point, we solve the Poisson equation for the potential at position z given an elementary charge unit located at position z' in such three-media geometries. With an in-plane separation $\vec{r}$, the full interaction is Fourier decomposed according to

$$V(r,z,z') = \frac{1}{4\pi^2}\int v(z,z';q)e^{i\vec{q}\cdot\vec{r}}d^2q\,.$$

The Fourier components satisfy the Poisson equation

$$\left(q^2\varepsilon(z) - \frac{d}{dz}\varepsilon(z)\frac{d}{dz}\right)v(z,z';q) = \frac{e^2}{\varepsilon_0}\delta(z-z')\,.$$

Provided point z' belongs to the 2D layer, this equation can be solved using the standard ansatz

$$v(z,z';q) = \frac{e^2}{2\varepsilon_0 q}\begin{cases} Ce^{-q|z-z'|} & z > d \\ \varepsilon^{-1}e^{-q|z-z'|} + Ae^{-q|z+z'|} + Be^{-q|2d-z-z'|} & 0 \le z \le d\,. \\ De^{-q|z-z'|} & z < 0 \end{cases}$$

Upon applying the appropriate boundary conditions, four equations are obtained for the four unknown coefficients. Eventually, for z and z' in the 2D layer, the full potential becomes

$$v(z,z';q) = \frac{e^2e^{-q|z+z'|}}{2q\varepsilon\varepsilon_0}\cdot\frac{\left(\varepsilon-\varepsilon_b+(\varepsilon+\varepsilon_b)e^{2qz_<}\right)\left(\varepsilon+\varepsilon_a+(\varepsilon-\varepsilon_a)e^{2q(z_>-d)}\right)}{(\varepsilon+\varepsilon_b)(\varepsilon+\varepsilon_a)-(\varepsilon-\varepsilon_b)(\varepsilon-\varepsilon_a)e^{-2qd}},$$

where $z_< = \min\{z,z'\}$ and $z_> = \max\{z,z'\}$. If $\varepsilon = \varepsilon_a$, the result above reduces to the usual two-media expression. We will now take the two-dimensional limit and approximate

$$v(z,z';q) \approx v(d/2,d/2;q) = \frac{e^2}{2q\varepsilon\varepsilon_0\varepsilon_{eff}(q)},$$

$$\varepsilon_{eff}(q) = \varepsilon\left\{1 - \frac{\varepsilon-\varepsilon_a}{\varepsilon-\varepsilon_a+(\varepsilon+\varepsilon_a)e^{qd}} - \frac{\varepsilon-\varepsilon_b}{\varepsilon-\varepsilon_b+(\varepsilon+\varepsilon_b)e^{qd}}\right\}. \tag{34.5}$$

It follows that screening is described by the effective dielectric constant $\varepsilon_{eff}(q)$. If, furthermore, the expression is expanded to first order in d we find

$$\varepsilon_{eff}(q) \approx \frac{\varepsilon_a + \varepsilon_b}{2} + \left\{ \frac{\varepsilon - 1}{2} - \frac{\varepsilon_a^2 + \varepsilon_b^2 - 2\varepsilon}{4\varepsilon} \right\} qd \,.$$

The first term in this expression is the usual image charge contribution for screening by semi-infinite sub- and superstrates. The last term in the effective dielectric constant is typically quite small, and we can approximate

$$\varepsilon_{eff}(q) \approx \bar{\varepsilon} + r_0 q, \quad \bar{\varepsilon} = \frac{\varepsilon_a + \varepsilon_b}{2}, \quad r_0 = \frac{\varepsilon - 1}{2} d \,, \tag{34.6}$$

where r_0 is the so-called screening length. In terms of this 2D dielectric function, the screened potential (in atomic units with $4\pi\varepsilon_0 = 1$) is given by the Keldysh form

$$V(r) = \frac{1}{2\pi} \int_0^{2\pi} \int_0^{\infty} \frac{e^{iqr\cos\phi}}{\bar{\varepsilon} + r_0 q} dq d\phi = \int_0^{\infty} \frac{J_0(qr)}{\bar{\varepsilon} + r_0 q} dq = \frac{\pi}{2r_0} \left\{ \mathrm{H}_0\left(\tfrac{\bar{\varepsilon} r}{r_0}\right) - Y_0\left(\tfrac{\bar{\varepsilon} r}{r_0}\right) \right\}. \tag{34.7}$$

Here, H_0 and Y_0 are Struve and Bessel functions, respectively. The Keldysh potential is compared to the unscreened $1/r$ interaction in Fig. 34.4. It is seen that the long-range behavior is similar but, for short distances, the divergence is only logarithmic.

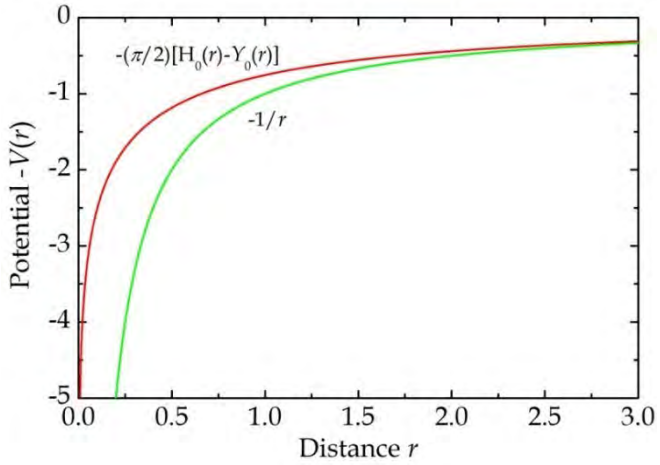


Figure 34.4. Comparison of Keldysh and bare Coulomb potentials.

This linearized form $\varepsilon(q) \approx \bar{\varepsilon} + r_0 q$ can be regarded as a first-order approximation to the physically more correct form

$$\varepsilon(q) = 1 + e^{-aq}\left[(\bar{\varepsilon} - 1)(1 + aq) + qr_0\right], \quad a = \frac{\sqrt{q_0 r_0\left[4(\bar{\varepsilon} - 1) + q_0 r_0\right]} - q_0 r_0}{2q_0(\bar{\varepsilon} - 1)}.$$

Here, q_0 is the q-value at maximum and $a = 1/q_0$ if $\bar{\varepsilon} = 1$. Some important material parameters for various monolayer transition-metal dichalcogenides are shown in the table below using parameters from [1].

Material	MoS_2	$MoSe_2$	WS_2	WSe_2
m_{eh}/m_0	0.28	0.27	0.22	0.23
m_e/m_0	0.55	0.49	0.46	0.48
r_0 [Å]	44.3	51.2	39.9	46.2

For suspended ($\bar{\varepsilon}=1$) and hBN-encapsulated ($\bar{\varepsilon}=4.5$) MoS_2, the dielectric constant is shown in Fig. 34.5 using $q_0 = 0.1$ Å^{-1}.

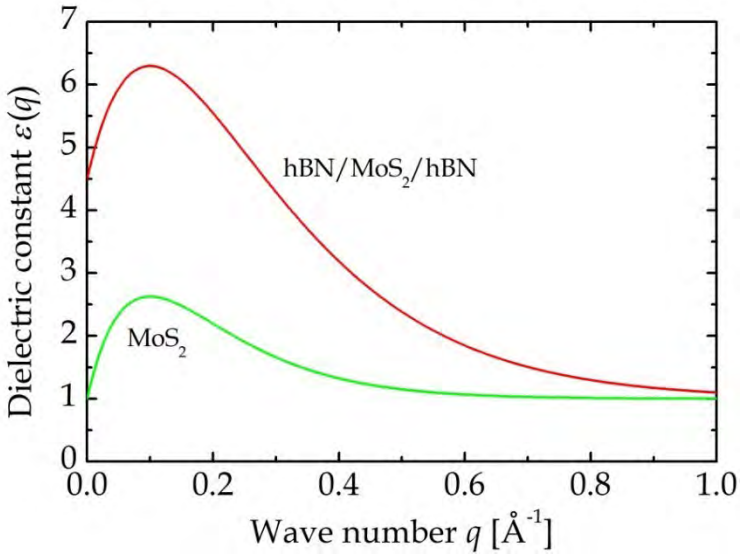


Figure 34.5. Dielectric constant of suspended and hBN-encapsulated MoS_2.

The plot below shows the 1S exciton binding energy for various TMDs using the Keldysh potential Eq.(34.7) as a function of screening by the surroundings. It is noted that very large values are reached in the low-screening regime.

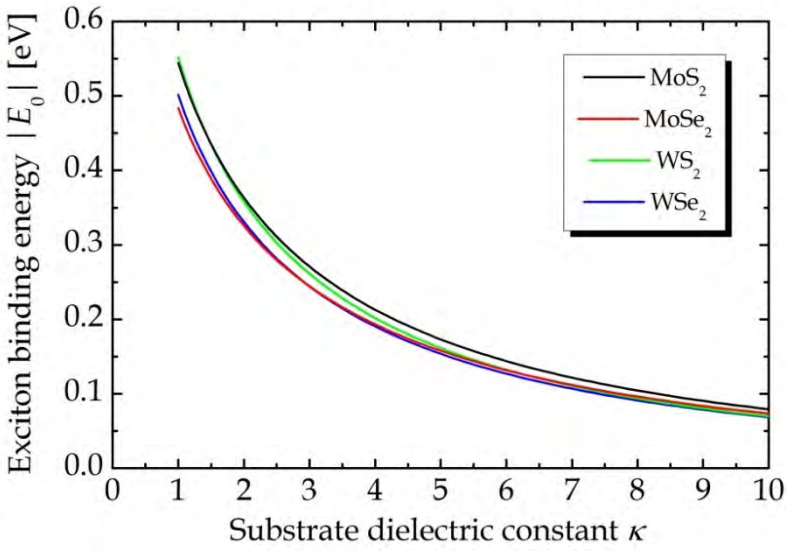


Figure 34.6. Exciton binding energy of four TMDs in various dielectric environments.

34.3 Doping in Two Dimensions

We now add screening by additional electrons in (say) the conduction band. Applying the 2D Fourier transform $v(q) = e^2/(2\varepsilon_0 q)$, this means that at $T = 0$

$$\varepsilon(q) = \varepsilon_{int}(q) + \frac{ge^2 m_e}{4\pi\varepsilon_0 \hbar^2 q}\left\{1 - \theta(q-2k_F)\sqrt{1-\frac{4k_F^2}{q^2}}\right\} = \varepsilon_{int}(q) + \frac{gm_e/m_0}{a_0 q}\left\{1 - \theta(q-2k_F)\sqrt{1-\frac{4k_F^2}{q^2}}\right\}.$$

Here, g is the product of spin and valley degeneracies, m_e is the effective mass, a_0 is the atomic Bohr radius, and ε_{int} is the intrinsic part. Moreover, the Fermi wave number is related to the doping density by $n = gk_F^2/4\pi$. Below, the inverse dielectric constant of MoS_2 is plotted for a range of doping densities using $\bar{\varepsilon} = 1$.

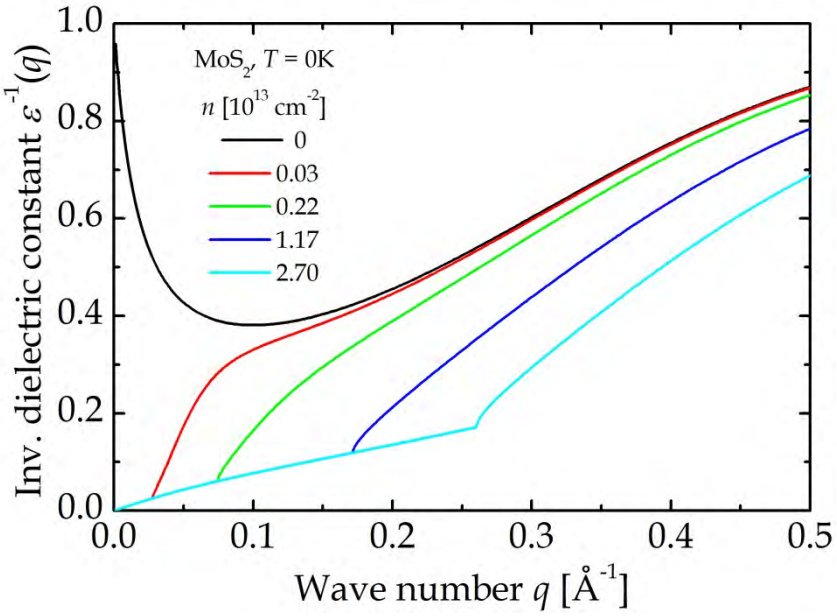


Figure 34.7. The inverse of the dielectric constant of suspended MoS_2 for various doping concentrations.

This figure is quite different from the *ab initio* results in Fig. 2 of Ref. [2]. However, agreement improves if finite temperatures are considered. First, we include broadening of the Lindhard function due to the smooth Fermi distribution. To this end, we define

$$L_T(z,\tau) = \frac{2}{z}\int_0^{z^2/4} \frac{dx}{\left\{\exp\left(\frac{x-\mathrm{sgn}\,\tau}{|\tau|}\right)+1\right\}\sqrt{z^2-4x}}.$$

In terms of this integral, the dielectric constant becomes

$$\varepsilon(q) = \varepsilon_{int}(q) + \frac{gm_e/m_0}{a_0 q} L_T\left(\frac{q}{|k_F|}, \frac{kT}{E_F}\right),$$

with $E_F = \hbar^2 k_F^2 / 2m_e$. The corresponding screening is shown in the top panel of Fig. 34.8 (using zero-temperature Fermi wave numbers). An additional temperature correction arises from the fact that the Fermi level is temperature dependent. Hence, Eq.(8.5) shows that the actual Fermi level E_F is related to the zero-temperature value E_F^0 via $E_F = kT \ln[\exp(E_F^0 / kT) - 1]$. Applying this correction leads to the plot in bottom panel in Fig. 34.8.

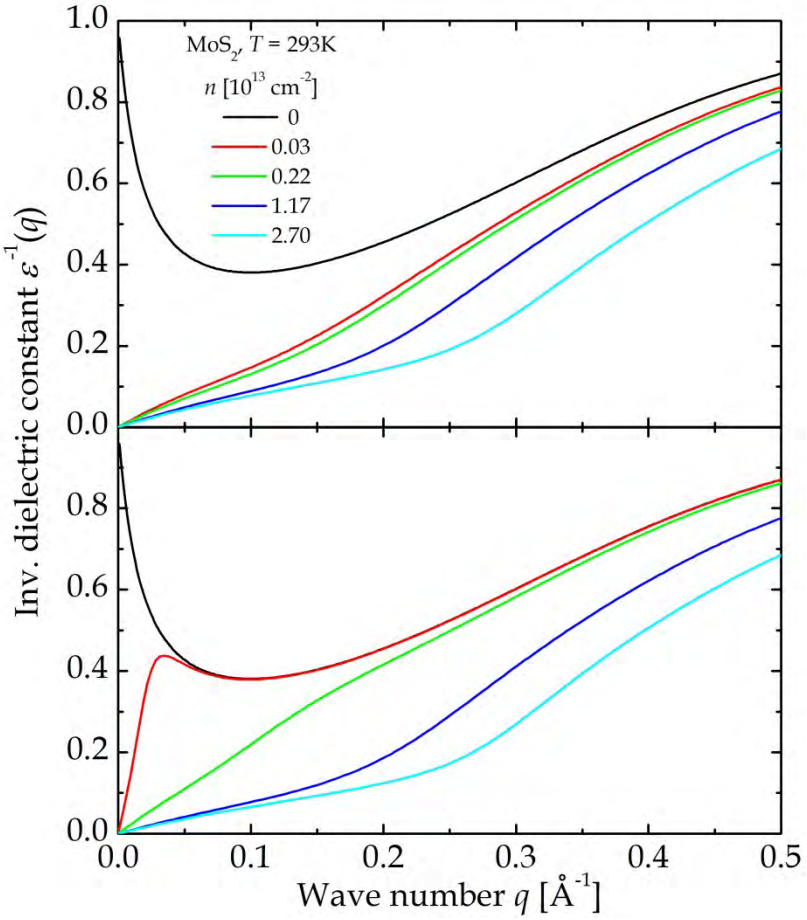


Figure 34.8. Finite-temperature corrected dielectric constant. In the top panel, only broadening of the Lindhard function is included. In the bottom, we include also the correction of the Fermi level.

Exercise: Eigenstates of the Yukawa potential

In natural exciton units, the eigenvalue problem for spherically symmetric solutions of the Yukawa potential Eq.(34.3) in three dimensions is given by

$$\left\{ -\frac{1}{2}\frac{\partial^2}{\partial r^2} - \frac{1}{r}\frac{\partial}{\partial r} - \frac{1}{r} e^{-qr} \right\} \varphi(r) = E\varphi(r) .$$

We will find *s*-type eigenstates by expanding a Gauss basis $|0,i\rangle = \exp(-b_i r^2)$ with exponents $b_i = b_1 \cdot 2^{(i-1)}$ for $i = 1,...,N$ with $b_1 = 10^{-3}$. The overlap and kinetic energy matrix elements are listed in App. 3 and for the screened Coulomb interaction

$$\langle 0,i|\frac{e^{-qr}}{r}|0,j\rangle = \frac{1}{2b_{ij}}\left\{1-\sqrt{\pi}\beta e^{\beta^2}\Phi_C(\beta)\right\}, \quad \beta = \frac{q}{2b_{ij}^{1/2}},$$

where Φ_C is the complementary error function.

a) Show that the three lowest eigenvalues depend on q as shown below.

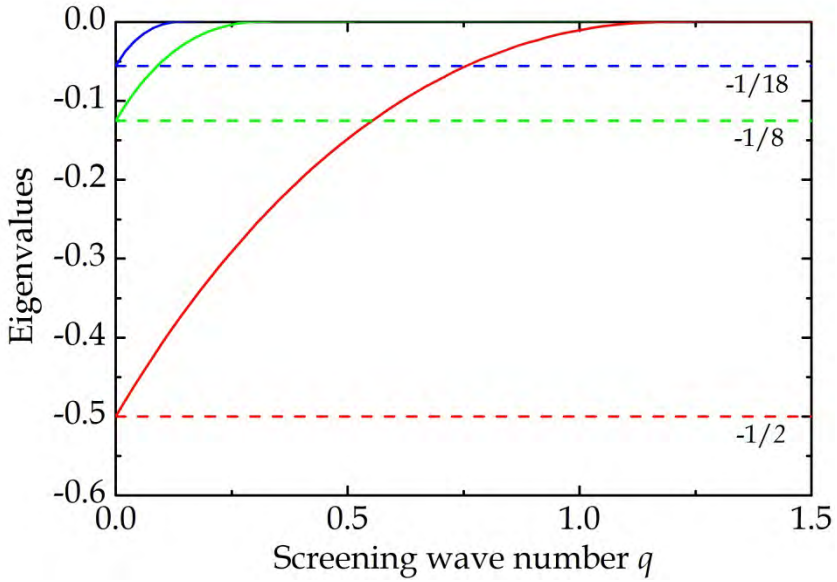


Figure 34.9. Lowest eigenvalues of the Yukawa potential.

References

[1] T. Olsen, S. Latini, F. Rasmussen, and K. S. Thygesen, Phys. Rev. Lett. 116, 056401 (2016).
[2] S. Gao *et al*. Nano Lett. 16, 5568 (2016).

35. Spontaneous Light Emission

We will apply some of the results derived in previous chapters to study the phenomenon of light emission. The focus will be on spontaneous emission, i.e. emission via spontaneous decay of an excited state, rather than stimulated (photon triggered) decay. The rate of emission will be analyzed for atomic as well as low-dimensional systems. In addition, we will briefly look at emission via transitions that are dipole forbidden.

We first investigate a small confined (zero-dimensional) system such as an atom or molecule. The decay rate is given by Eq.(25.7)

$$\Gamma_0 = \frac{e^2\omega_{n0}^3 n_r \,|\, D_{0n} \,|^2}{3\pi\varepsilon_0\hbar c^3} = \frac{e^2\omega_{n0} n_r \,|\, P_{0n} \,|^2}{3\pi\varepsilon_0\hbar m^2 c^3} \tag{35.1}$$

using the relation between dipole and momentum matrix elements. This result can also be expressed in terms of the oscillator strength

$$\Gamma_0 = \frac{e^2\omega_{n0}^2 n_r g_{n0}}{6\pi\varepsilon_0 m c^3}. \tag{35.2}$$

Moreover, introducing $\alpha \equiv e^2/(4\pi\varepsilon_0 c\hbar) \approx 1/137$ as the fine-structure constant we find

$$\hbar\Gamma_0 = \frac{2\alpha n_r g_{n0}}{3}\frac{E_{n0}^2}{mc^2}.$$

Thus, because $E_{n0} \ll mc^2$ we typically have $\hbar\Gamma_0 \ll E_{n0}$. Taking as an example the $2p \to 1s$ transition of hydrogen in vacuum, we have $g_{2p-1s} = 0.4162$ and so $\hbar\Gamma_0 \approx 410$ neV corresponding to a decay rate of $\Gamma_0 \approx 6.3\cdot 10^8$ s^{-1}. The reciprocal is the lifetime $\tau \approx 1.6$ ns. This result is in perfect agreement with experiments. Similarly, for the $1s2p \to 1s^2$ transition in helium we found earlier $g_{1s2p-1s^2} \approx 0.2633$ and $E_{1s2p-1s^2} = 0.78$ Ha, which means $\Gamma_0 \approx 1.72\cdot 10^9$ s^{-1} while the experimental value is around $1.8\cdot 10^9$ s^{-1} in agreement with the more accurate oscillator strength $g_{1s2p-1s^2} \approx 0.276$.

In a solid, a similar analysis can be made. If many-body effects are neglected, we can simply consider band to band transitions. Otherwise, excitonic effects should be considered. The starting point for this analysis is to "decompose" the decay rate into

contributions from different photon modes characterized by their wavevector $\vec{q}$. For an emitter in vacuum, this can be done as follows

$$\Gamma = \frac{e^2 \mid P_{0n} \mid^2}{8\pi^2 \omega_{n0} \varepsilon_0 \hbar m^2 c} \int \left(\frac{q_\perp}{q} \right)^2 \delta \left(\frac{\omega_{n0}}{c} - q \right) d^3 q \,. \tag{35.3}$$

Here, the factors under the integral sign take care of the two requirements for the emitted photons: The dispersion relation $\omega = qc$ ensuring energy conservation and the directional dependence, which states that emission follows a $(q_\perp / q)^2$ pattern with $q_\perp$ the $\vec{q}$ component perpendicular to the emitting dipole. Sometimes, this factor is explained in a seemingly different fashion: For a given photon emitted in the $\vec{q}$ direction, one should sum over the two polarization directions $\vec{e}_\lambda$ with $\lambda = 1,2$. These polarization vectors should be orthogonal to $\vec{q}$ and to each other. For a given electronic transition $n \to 0$ with momentum matrix element $\vec{P}_{0n}$ the projections onto $\vec{e}_\lambda$ should be taken. Hence, one would find

$$\Gamma \propto \sum_\lambda \left| \vec{e}_\lambda \cdot \vec{P}_{0n} \right|^2 = \vec{P}_{n0} \cdot \left(\vec{e}_1 \vec{e}_1 + \vec{e}_2 \vec{e}_2 \right) \cdot \vec{P}_{0n} \,.$$

Now, the two vectors $\vec{e}_\lambda$ together with the unit vector along $\vec{q}$, i.e. $\vec{e}_q$, form three orthogonal unit vectors, which means that $\vec{e}_1 \vec{e}_1 + \vec{e}_2 \vec{e}_2 + \vec{e}_q \vec{e}_q = \overleftrightarrow{I}$. Thus,

$$\Gamma \propto \vec{P}_{n0} \cdot \left(\overleftrightarrow{I} - \vec{e}_q \vec{e}_q \right) \cdot \vec{P}_{0n} = \mid P_{0n} \mid^2 \left(\frac{q_\perp}{q} \right)^2 .$$

For a simple dipole in vacuum, using a spherical coordinate system with the dipole axis as the polar axis, the latter factor simply adds a factor $\sin^2 \theta$ and doing the integral in spherical coordinates readily shows that Eqs.(35.1) and (35.3) agree. We now look at the modifications needed if the emitter is a low-dimensional semiconductor structure.

35.1 Low-dimensional Systems

We first consider a one-dimensional structure (quantum wire or nanotube) with the transition dipole oriented along the long-axis chosen to be the *x*-axis. In addition, because the system is translationally invariant along the *x*-axis, the emitting state has a certain center-of-mass momentum Q_x along this direction. This momentum must be conserved during emission and, consequently, $q_x = Q_x$ is required of the photon wavevector. Formally, we consider a piece of one-dimensional material of length *L*

and, eventually, should take the limit $L \to \infty$. The appropriate modification of Eq.(35.3) is then

$$\Gamma_{1D} = \frac{e^2 \mid P_{0n} \mid^2}{8\pi^2 \omega_{n0} \varepsilon_0 \hbar m^2 c} \left(\frac{2\pi}{L} \right) \sum_{q_x} \delta_{q_x, Q_x} \int \left(\frac{q_y^2 + q_z^2}{q^2} \right) \delta \left(\frac{\omega_{n0}}{c} - q \right) dq_y dq_z .$$

We write $dq_y dq_z = 2\pi q_{\parallel} dq_{\parallel}$ and use $q^2 = q_{\parallel}^2 + Q_x^2$. Hence, a simple calculation shows that

$$\Gamma_{1D} = \frac{e^2 \mid P_{0n} \mid^2}{2\omega_{n0}^2 \varepsilon_0 \hbar m^2 c^2 L} \left\{ \omega_{n0}^2 - Q_x^2 c^2 \right\} \theta \left(\omega_{n0} - Q_x c \right).$$

This result agrees with expressions derived using different but equivalent approaches [1,2]. In complete analogy, a 2*D* system with the *z*-axis perpendicular to the structure and the dipole along the *x*-axis is described by

$$\Gamma_{2D} = \frac{e^2 \mid P_{0n} \mid^2}{8\pi^2 \omega_{n0} \varepsilon_0 \hbar m^2 c} \left(\frac{2\pi}{L} \right)^2 \sum_{q_x, q_y} \delta_{q_x, Q_x} \delta_{q_y, Q_y} \int \left(\frac{q_y^2 + q_z^2}{q^2} \right) \delta \left(\frac{\omega_{n0}}{c} - q \right) dq_z .$$

In the general case of non-vanishing $\vec{Q}$, we find for a dipole along *x*

$$\Gamma_{2D}(Q \neq 0) = \frac{e^2 \mid P_{0n} \mid^2}{\omega_{n0}^2 \varepsilon_0 \hbar m^2 c L^2} \left\{ \sqrt{\omega_{n0}^2 - Q^2 c^2} + \frac{Q_y^2 c^2}{\sqrt{\omega_{n0}^2 - Q^2 c^2}} \right\} \theta \left(\omega_{n0} - Qc \right).$$

Hence, this expression diverges if $Q = \omega_{n0} / c$, a situation that corresponds to emission along the plane of the 2*D* structure. A similar result was found in Ref. [2]. Under the assumption $Q_x \approx Q_y \approx 0$, we find

$$\Gamma_{2D}(Q = 0) = \frac{e^2 \mid P_{0n} \mid^2}{\omega_{n0} \varepsilon_0 \hbar m^2 c L^2} .$$

The 3*D* case, in fact, is trickier. We cannot require, independently, momentum and energy conservation. Really, the decay rate vanishes identically unless $q = Q = \omega_{n0} / c$. However, in that particular case, and for a dipole along *x* we find

$$\begin{aligned} \Gamma_{3D}(Q = \omega_{n0} / c) &= \frac{e^2 \mid P_{0n} \mid^2}{8\pi^2 \omega_{n0}^2 \varepsilon_0 \hbar m^2} \left(\frac{2\pi}{L} \right)^3 \sum_{q_x, q_y, q_z} \delta_{q_x, Q_x} \delta_{q_y, Q_y} \delta_{q_z, Q_z} \left(1 - \frac{Q_x^2}{Q^2} \right) \\ &= \frac{e^2 \mid P_{0n} \mid^2 \pi}{\omega_{n0}^2 \varepsilon_0 \hbar m^2 L^3} \left(1 - \frac{Q_x^2 c^2}{\omega_{n0}^2} \right). \end{aligned}$$

35.2 Thermal Average

In a real experiment, states of all possible momenta $\vec{Q}$ exist in a thermal distribution. If the total mass of the state is *M*, the Boltzmann weight of that state is $\exp\{-\hbar^2 Q^2/(2MkT)\} \equiv \exp\{-\gamma Q^2\}$. In 1*D*, averaging over all Q_x leads to

$$\langle \Gamma_{1D} \rangle = \frac{e^2 \mid P_{0n} \mid^2}{2\omega_{n0}^2 \varepsilon_0 \hbar m^2 c^2 L} Z^{-1} \int_0^\infty \left\{ \omega_{n0}^2 - Q_x^2 c^2 \right\} \theta(\omega_{n0} - Q_x c) e^{-\gamma Q_x^2} dQ_x \, .$$

with partition function $Z = \int_0^\infty e^{-\gamma Q_x^2} dQ_x = \sqrt{\pi}/2\sqrt{\gamma}$. We always expect $\gamma \omega_{n0}^2 / c^2 \ll 1$ at normal temperatures and consequently Taylor expand in this quantity. The final result is

$$\langle \Gamma_{1D} \rangle = \frac{2e^2 \mid P_{0n} \mid^2 \omega_{n0}}{3\varepsilon_0 \hbar m^2 c^3 L} \sqrt{\frac{\gamma}{\pi}} \, .$$

In the 2*D* case, a similar analysis leads to

$$\langle \Gamma_{2D} \rangle = \frac{4e^2 \mid P_{0n} \mid^2 \omega_{n0}}{3\varepsilon_0 \hbar m^2 c^3 L^2} \gamma \, .$$

In the Wannier model, the momentum matrix of an 2D exciton is given by

$$\mid P_{0n} \mid^2 = 2L^2 \mid p_{vc} \mid^2 \mid \Psi(0) \mid^2 .$$

As this wave function is to be evaluated at the origin, only *s*-type excitations contribute. Restricting ourselves to the dominant 1*s* state for a material with effective Bohr radius a_B^*, we find (see Chapter 19): $\mid \Psi(0) \mid^2 = 8/\pi (a_B^*)^2$. Putting the pieces together, it follows that

$$\langle \Gamma_{2D} \rangle = \frac{64e^2 \mid p_{vc} \mid^2 \omega_{n0}}{3\pi\varepsilon_0 \hbar m^2 c^3 (a_B^*)^2} \gamma \, .$$

35.3 "Forbidden" Transitions

From the analysis above, it appears clear that only states having a non-vanishing momentum matrix element with the ground state should decay radiatively. In reality, all states can decay via photon emission but more advanced theory is needed to account for this. The most prominent sources of "forbidden" transitions are magnetic dipole emission and spin-orbit induced mixing.

The magnetic dipole correction follows from including the spatial variation of the electric field associated with the emitted photons. The field varies as $e^{i\vec{q}\cdot\vec{r}} \approx 1 + i\vec{q}\cdot\vec{r}$. Thus, we replace $P_{n0} = \langle n|\hat{P}_x|0\rangle \to \langle n|(1+i\vec{q}\cdot\vec{r})\hat{P}_x|0\rangle$. The new term can be symmetrized as follows

$$q\langle n|y\hat{P}_x|0\rangle = \frac{q}{2}\langle n|y\hat{P}_x - x\hat{P}_y|0\rangle + \frac{q}{2}\langle n|y\hat{P}_x + x\hat{P}_y|0\rangle .$$

The first of these is called the magnetic dipole interaction whereas the latter is the electric quadrupole. Since $q = n_r\omega_{n0}/c$ and $x\hat{P}_y - y\hat{P}_x = \hbar\hat{l}_z$, we find for the magnetic part

$$\Gamma_0^{MD} = \frac{e^2\hbar\omega_{n0}^3 n_r^3 |\langle n|\hat{l}_z|0\rangle|^2}{12\pi\varepsilon_0 m^2 c^5} . \tag{35.4}$$

The spontaneous decay via spin-orbit (SO) interaction is a separate mechanism (see Appendix 7) and the radiation is usually called phosphorescence to distinguish it from the dipole-allowed emission called fluorescence. As an example, we will consider the states formed from an *sp* two-electron configuration and their decay to the s^2 ground state. This ground state is a singlet and, so, spin-selection makes excited triplet states non-radiative. In the presence of SO interaction, however, excited singlets and triplets mix in the sense that actual eigenstates are mixtures of the two. The basis of singlet and triplet states consists of states denoted $|^{2S+1}L_{JM}\rangle$, where *S* and *L* are the total spin and angular momentum, respectively, whereas *J* and *M* are the magnitude and *z*-projection of the total angular momentum. The SO interaction only couples states with identical *J* and *M*. For the singlets, *J* = 1 and so *M* is maximally 1 and we focus on the *M* = 1 subset of states. By the usual construction

$$\begin{aligned}
|^1P_{1,1}\rangle &= \frac{1}{2}\{Y_{11}(1)Y_{00}(2) + Y_{00}(1)Y_{11}(2)\}\{\alpha(1)\beta(2) - \beta(1)\alpha(2)\} \\
|^3P_{1,1}\rangle &= \frac{1}{2}\{Y_{10}(1)Y_{00}(2) - Y_{00}(1)Y_{10}(2)\}\alpha(1)\alpha(2) \\
&\quad -\frac{1}{\sqrt{8}}\{Y_{11}(1)Y_{00}(2) - Y_{00}(1)Y_{11}(2)\}\{\alpha(1)\beta(2) + \beta(1)\alpha(2)\}.
\end{aligned}$$

Here, the radial parts have been suppressed and only angular and spin parts are shown. It is now relatively straightforward to shown that

$$\langle {}^3P_{1,1}|\hat{H}_{SO}|{}^1P_{1,1}\rangle = -\frac{\varsigma_{ST}}{\sqrt{2}}, \quad \varsigma_{ij} = \frac{\hbar^2}{4m^2c^2}\int_0^\infty R_i(r)\frac{dV}{dr}R_j(r)rdr .$$

In addition, we find the diagonal corrections

$$\begin{aligned}\left\langle {}^1P_{1,1}\middle|\hat{H}_{SO}\middle|{}^1P_{1,1}\right\rangle &= 0, \quad \left\langle {}^3P_{1,1}\middle|\hat{H}_{SO}\middle|{}^3P_{1,1}\right\rangle = -\frac{\varsigma_{TT}}{2},\\ \left\langle {}^3P_{2M}\middle|\hat{H}_{SO}\middle|{}^3P_{2M}\right\rangle &= \frac{\varsigma_{TT}}{2}, \quad \left\langle {}^3P_{00}\middle|\hat{H}_{SO}\middle|{}^3P_{00}\right\rangle = -\varsigma_{TT}\end{aligned}.$$

As long as the coupling is weak, we find states that are predominantly singlet or triplet in character. Denoting these ${}^1P_1^\circ$ and ${}^3P_1^\circ$, respectively, we find approximately

$$\begin{aligned}\left|{}^1P_1^\circ\right\rangle &= \left|{}^1P_{1,1}\right\rangle + \frac{\varsigma_{ST}}{\sqrt{2}(E_S - E_T)}\left|{}^3P_{1,1}\right\rangle, \quad E({}^1P_1^\circ) \approx E_S\\ \left|{}^3P_1^\circ\right\rangle &= -\frac{\varsigma_{ST}}{\sqrt{2}(E_S - E_T)}\left|{}^1P_{1,1}\right\rangle + \left|{}^3P_{1,1}\right\rangle, \quad E({}^3P_1^\circ) \approx E_T - \frac{\varsigma_{TT}}{2} - \frac{\varsigma_{ST}^2}{2(E_S - E_T)}.\end{aligned}$$

Taking the Ca atom as an example, it is found experimentally that

$E({}^1P_1^\circ) = 2.932512 \text{ eV}, \quad E({}^3P_2) = 1.898935 \text{ eV}, \quad E({}^3P_1^\circ) = 1.885807 \text{ eV}, \quad E({}^3P_0) = 1.879340 \text{ eV}.$
Hence, solving shows that $E_S - E_T = 1.04 \text{ eV}$, $\varsigma_{TT} = 13.1 \text{ meV}$ and $\varsigma_{ST} = 11.6 \text{ meV}$. In turn, the decay rates of the ${}^3P_1^\circ$ and ${}^1P_1^\circ$ states are related by

$$\frac{\Gamma\left({}^3P_1^\circ\right)}{\Gamma\left({}^1P_1^\circ\right)} \approx \left(\frac{E_T}{E_S}\right)^3 \frac{\varsigma_{ST}^2}{2(E_S - E_T)^2}. \tag{35.5}$$

Inserting, we find a ratio of $1.67 \cdot 10^{-5}$ while the experimental value is $1.2 \cdot 10^{-5}$.

35.4 Purcell Effect

If an emitter is placed near a surface, the spontaneous emission rate is modified. This is known as the Purcell effect and the modification depends on a range of factors. We consider a dipole placed in vacuum a distance d outside a surface. For a dipole oriented parallel or perpendicular to the surface, one finds [3]

$$\begin{aligned}\Gamma_\parallel &= \Gamma_0 + \frac{3\Gamma_0}{4k_0}\text{Re}\int_0^\infty \left\{r_s - r_p\frac{k_\perp^2}{k_0^2}\right\} e^{2ik_\perp d}\frac{k_\parallel}{k_\perp}dk_\parallel,\\ \Gamma_\perp &= \Gamma_0 + \frac{3\Gamma_0}{2k_0^3}\text{Re}\int_0^\infty r_p e^{2ik_\perp d}\frac{k_\parallel^3}{k_\perp}dk_\parallel.\end{aligned} \tag{35.6}$$

Here, the wave vector in vacuum has been decomposed into components parallel ($k_\parallel$) and perpendicular ($k_\perp$) to the surface. These are connected via $k_\parallel^2 + k_\perp^2 = k_0^2$. If the dipole sits in front of a homogeneous, isotropic medium with dielectric constant ε, the reflection coefficients are

$$r_s = \frac{k_\perp - q_\perp}{k_\perp + q_\perp}, \quad r_p = \frac{\varepsilon k_\perp - q_\perp}{\varepsilon k_\perp + q_\perp},$$

where $q_\perp = (\varepsilon k_0^2 - k_\parallel^2)^{1/2}$ is the perpendicular component of the wave vector inside the medium. To evaluate the integral, it is convenient to introduce $x = k_\perp / k_0$, $y = k_\parallel / k_0$, and $z = k_0 d$. We then split the integrals into "propagating" ($k_\parallel < k_0$) and "evanescent" ($k_\parallel > k_0$). It can be shown that by changing the integration variable from $k_\parallel$ to x and ix in these integrals, respectively,

$$\begin{aligned}
\Gamma_\parallel &= \Gamma_0 + \frac{3\Gamma_0}{4}\left\{\mathrm{Re}\int_0^1 \left\{r_s - r_p x^2\right\} e^{2ixz} dx + \mathrm{Im}\int_0^\infty \left\{r_s + r_p x^2\right\} e^{-2xz} dx\right\}, \\
\Gamma_\perp &= \Gamma_0 + \frac{3\Gamma_0}{2}\left\{\mathrm{Re}\int_0^1 r_p (1-x^2) e^{2ixz} dx + \mathrm{Im}\int_0^\infty r_p (1+x^2) e^{-2xz} dx\right\},
\end{aligned} \tag{35.7}$$

where the reflection coefficients are $r_s = (x - x')/(x + x')$ and $r_p = (\varepsilon x - x')/(\varepsilon x + x')$ and x' is given by either $x' = (\varepsilon - 1 + x^2)^{1/2}$ in the "propagating" integrals and $x' = (1 - \varepsilon + x^2)^{1/2}$ in the "evanescent" integrals. Numerical results for ideal and real Au mirrors are shown in the exercise.

The same idea can be extended to a dipole in arbitrary layered structures. Placing a parallel dipole at position z_0 in a particular slab, one finds [4]

$$\begin{aligned}
\Gamma_\parallel = \Gamma_0 + \frac{3\Gamma_0}{4k_0^3}\mathrm{Re}\int_0^\infty \Bigg\{ & k_0^2 \frac{r_s^{(L)} e^{2ik_\perp z_0} + r_s^{(R)} e^{2ik_\perp (d - z_0)} + e^{2ik_\perp d} 2 r_s^{(L)} r_s^{(R)}}{1 - r_s^{(L)} r_s^{(R)} e^{2ik_\perp d}} \\
& - k_\perp^2 \frac{r_p^{(L)} e^{2ik_\perp z_0} + r_p^{(R)} e^{2ik_\perp (d - z_0)} - e^{2ik_\perp d} 2 r_p^{(L)} r_p^{(R)}}{1 - r_p^{(L)} r_p^{(R)} e^{2ik_\perp d}} \Bigg\} \frac{k_\parallel}{k_\perp} dk_\parallel.
\end{aligned}$$

Similarly, for a perpendicular dipole,

$$\Gamma_\perp = \Gamma_0 + \frac{3\Gamma_0}{2k_0^3}\mathrm{Re}\int_0^\infty \frac{r_p^{(L)} e^{2ik_\perp z_0} + r_p^{(R)} e^{2ik_\perp (d - z_0)} + e^{2ik_\perp d} 2 r_p^{(L)} r_p^{(R)}}{1 - r_p^{(L)} r_p^{(R)} e^{2ik_\perp d}} \frac{k_\parallel^3}{k_\perp} dk_\parallel .$$

Here, $r_{s,p}^{(L,R)}$ are the coefficients for reflection from the interface left (L) and right (R) of the dipole. Also, d is the slab thickness and z_0 is the distance to the left interface, so that $d-z_0$ is the distance to the right interface. These distances are matched by the phase factors above. It is clear that these expressions reduce to the single-interface case if $r_{s,p}^{(R)}$ vanishes. Now, let's assume for simplicity that $r_{s,p}^{(L)}=r_{s,p}^{(R)}\equiv r_{s,p}$ and $z_0=d/2$ so that we place the dipole in the center of a symmetric cavity. In this case,

$$\Gamma_{\parallel}=\Gamma_0+\frac{3\Gamma_0}{2k_0^3}\mathrm{Re}\int_0^{\infty}\left\{\frac{k_0^2}{r_s^{-1}e^{-ik_{\perp}d}-1}-\frac{k_{\perp}^2}{r_p^{-1}e^{-ik_{\perp}d}+1}\right\}\frac{k_{\parallel}}{k_{\perp}}dk_{\parallel},$$

$$\Gamma_{\perp}=\Gamma_0+\frac{3\Gamma_0}{k_0^3}\mathrm{Re}\int_0^{\infty}\frac{1}{r_p^{-1}e^{-ik_{\perp}d}-1}\frac{k_{\parallel}^3}{k_{\perp}}dk_{\parallel}.$$

We will assume, furthermore, that $r_s=-1$ and $r_p=1$ as appropriate for a perfect reflector. It can then be shown that [5]

$$\Gamma_{\parallel}=\Gamma_0\frac{3\pi}{2z}\sum_{n=1}^{N}\left(1+\frac{n^2\pi^2}{z^2}\right)\sin^2\left(\tfrac{n\pi}{2}\right),\quad \Gamma_{\perp}=\Gamma_0\frac{3\pi}{2z}\left\{1+2\sum_{n=1}^{N}\left(1-\frac{n^2\pi^2}{z^2}\right)\cos^2\left(\tfrac{n\pi}{2}\right)\right\},$$

with $z=k_0d$ and N the integer part of z/π. The result based on these simple expressions can be compared to those for a realistic metal such as Au at an energy of 2.5 eV for which $\varepsilon\approx(1.04+i1.83)^2$. Such a comparison is shown in Fig. 35.1.

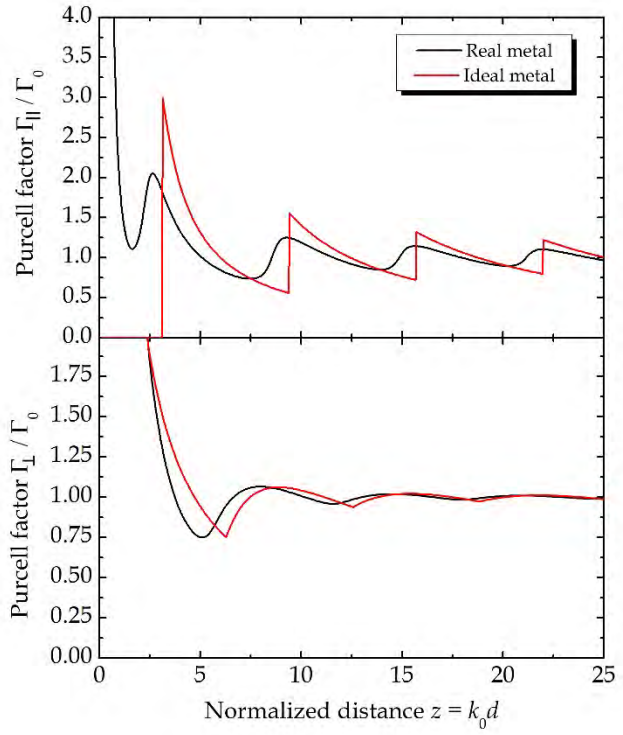


Figure 35.1. Purcell factors for parallel and perpendicular dipoles in real and ideal metal cavities.

Exercise: Purcell effect for single mirror

a) Consider first a highly conducting metal surface such that $\varepsilon \approx i\infty$ and $r_s = -1$ and $r_p = 1$. Show using Eq.(35.7) that for a dipole in front of a single mirror

$$\Gamma_{\parallel} = \Gamma_0 - \frac{3\Gamma_0}{2}\left\{\frac{\sin 2z}{2z} + \frac{\cos 2z}{(2z)^2} - \frac{\sin 2z}{(2z)^3}\right\}, \quad \Gamma_{\perp} = \Gamma_0 - 3\Gamma_0\left\{\frac{\cos 2z}{(2z)^2} - \frac{\sin 2z}{(2z)^3}\right\}.$$

b) Consider next a realistic model of Au taking $\varepsilon \approx (1.04 + i1.83)^2$. Write a computer program to evaluate the integral expressions numerically. The correct result is illustrated below and compared to the ideal metal. The strong Purcell enhancement close to the surface for real Au is due to Ohmic losses in the material.

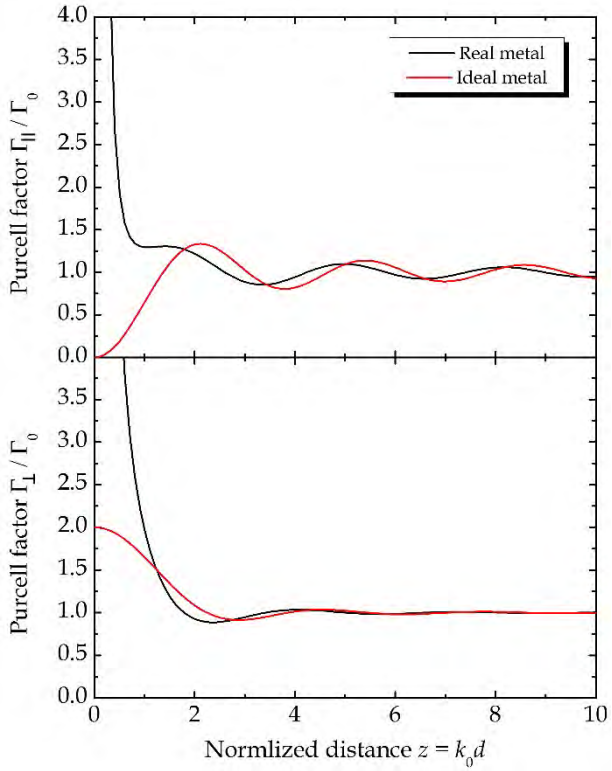


Figure 35.2. Purcell factors for parallel and perpendicular dipoles outside real and ideal metals.

References

[1] C.D. Spataru *et al.*, Phys. Rev. Lett. 95, 247402 (2005).
[2] Y.N. Chen and D.S. Chuu, Europhys. Lett. 54, 366 (2001).
[3] O. Kidwai, S.V. Zhukovsky, and J. E. Sipe, Phys. Rev. A85, 053842 (2012).
[4] R. L. Hartman, S. M. Cohen, and P. T. Leung, J. Chem. Phys. 110, 2189 (1999).
[5] G. S. Agarwal, Phys. Rev. A12, 1475 (1975).

36. Exciton Dissociation

As discussed in the previous chapters, bound excitons are very much like hydrogen atoms, living either in bulk (3D) or lower dimensions. The binding is due to the attractive Coulomb force between oppositely charged electrons and holes. Hence, the binding energy is large in structures that have little screening and are low-dimensional, i.e. with a large electron-hole overlap. In e.g. solar cells and photo-detectors, these excitons must be broken (dissociated) in order to extract the carriers at opposite electrodes. For loosely bound excitons, this can be accomplished simply by operating the device at room temperature. Thus, the average thermal energy $kT \approx 25$ meV at room temperature is sufficient to break up the bound pairs in materials such as bulk Si or GaAs. In other materials and, in particular, in lower dimensions, the exciton binding energy may be much larger than kT. In this case, a strong electric field may be required to dissociate the exciton. In this chapter, we will examine the field-assisted dissociation in detail.

36.1 One-Dimensional Excitons

The description of electron-hole pairs in an electric field $\vec{\mathcal{E}} = \mathcal{E}\vec{e}_x$ is tricky because the electrostatic potential $\mathcal{E}x$, with x the electron-hole separation, is unbounded. Physically, this means that the ground state is a state with infinitely low energy, in which electrons are located to the infinite left and holes to the infinite right, no matter the Coulomb attraction. But this, of course, is exactly what is meant by dissociation. We will start by studying some simple one-dimensional exciton models. Also, we will use natural units, in which $e = 4\pi\varepsilon_r\varepsilon_0 = \hbar = \mu = 1$ with ε_r the relative dielectric constant of the ambient medium. In general, the perturbed problem in 1D is of the form

$$\left\{-\frac{1}{2}\frac{d^2}{dx^2} + V(x) + \mathcal{E}x\right\}\varphi(x) = E\varphi(x)\,. \tag{36.1}$$

Our first model is a "hydrogen atom" with contact interaction. This means that the actual Coulomb interaction is replaced by the ultra-short-range delta-function approximation $-Z\delta(x)$, with "nuclear" charge Z. Obviously, for the exciton problem, Z = 1, but we keep this parameter because it adds generality without making the problem any harder. Hence, the problem without field is

$$\left\{-\frac{1}{2}\frac{d^2}{dx^2} - Z\delta(x)\right\}\varphi(x) = E\varphi(x)\,.$$

The delta-function potential is unusual in that it vanishes everywhere except for a single point. At this point, however, it is infinitely strong. This means that the wave function "bends" sharply at that point. By integrating the Schrödinger equations across the singularity and requiring that the wave function itself is continuous, we find a condition for the jump in the derivative $\varphi'(0_-)-\varphi'(0_+)=2Z\varphi(0)$. Hence, because the general solution for $x\neq 0$ is in the form of exponentials $\exp\{\pm Z|x|\}$, the normalized solution to the problem is

$$\varphi_0(x)=\sqrt{Z}\exp\{-Z|x|\},\quad E_0=-\frac{Z^2}{2}.$$

Note the very close resemblance to the three-dimensional hydrogen problem for both wave function and energy. With the field turned on, the problem becomes

$$\left\{-\frac{1}{2}\frac{d^2}{dx^2}-Z\delta(x)+\mathcal{E}x\right\}\varphi(x)=E\varphi(x).$$

The situation is illustrated in Fig. 36.1. Here, the full potential containing the delta-function attraction and the electrostatic potential are added. Also, in the right-hand panel, the case of a Coulomb potential is considered. The binding energy of the unperturbed problem E_0 is indicated in both plots. It is clear that if one moves sufficiently far to the left, a region of very low potential is reached and, eventually, the state will tunnel out of the central potential well.

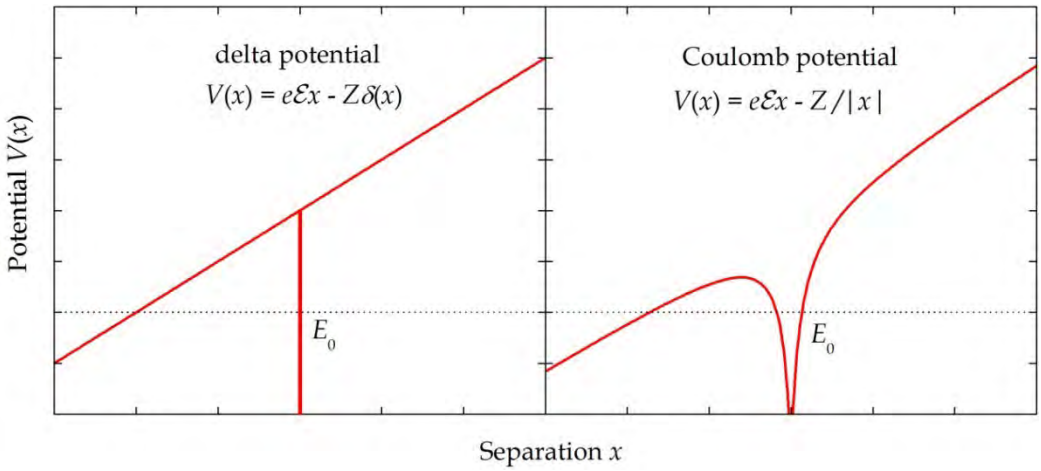


Figure 36.1. Illustration of the combined effect of the central potential well and the electrostatic potential.

Before attacking the full problem, let us apply Dalgarno-Lewis perturbation theory [1] to compute the second order correction to the energy due to the electric field. Using the fact that there is no first order energy correction, the first order wave function satisfies the inhomogeneous Dalgarno-Lewis equation (see Appendix 11)

$$\left\{-\frac{1}{2}\frac{d^2}{dx^2}-Z\delta(x)+\frac{Z^2}{2}\right\}\varphi_1(x)=-\mathcal{E}x\varphi_0(x),$$

having the solution

$$\varphi_1(x)=-\frac{\mathcal{E}x}{2Z^{3/2}}(1+Z|x|)\exp\{-Z|x|\}.$$

From this result, we compute the second order energy correction

$$E_2=\langle\varphi_1|\mathcal{E}x|\varphi_0\rangle=-\frac{5\mathcal{E}^2}{8Z^4}\equiv-\frac{1}{2}\alpha\mathcal{E}^2 \qquad (36.2)$$

leading to a polarizability of $\alpha=5/4Z^4$. Thus, as expected, the state has a lowered energy in the field. The full wave function assuming Z = 1 and $\mathcal{E}$ = 0.2 is illustrated in Fig. 36.2.

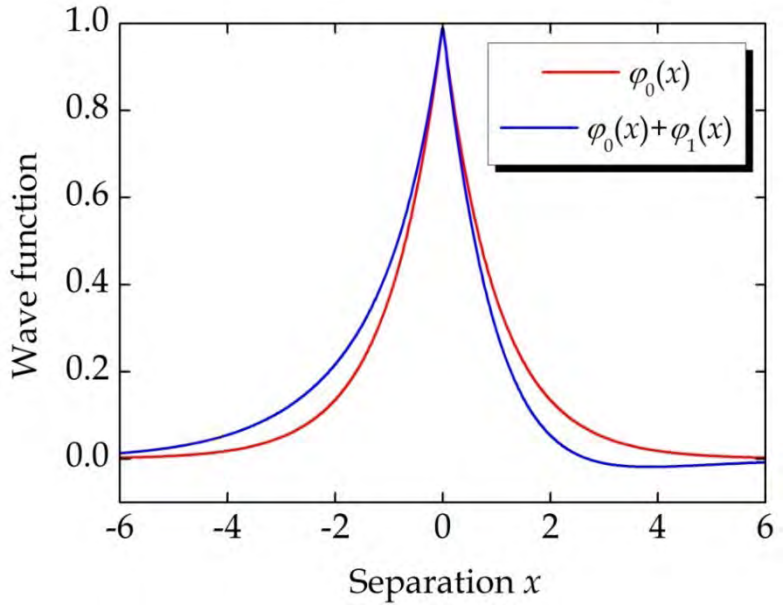


Figure 36.2. Unperturbed and perturbed wave functions in a field of magnitude $\mathcal{E}$ = 0.2.

It is seen that in the presence of the field, the wave function shifts its centre of gravity to the left. This is precisely what is meant by polarization. Eventually, if all high order corrections to the wave function are added, the state is shifted infinitely to the left no matter how small the electric field. This is the process of dissociation. If one imagines that the system is initially in the unperturbed ground state and the field is then suddenly turned on, the state gradually tunnels out of the central potential well. The tunnelling or decay rate Γ is such that the probability of finding the state in the initial state varies as $\exp\{-\Gamma t\}$. We know, however, that in an eigenstate with energy E the wave function varies in time according to $\exp\{-iEt\}$. Thus, because the probability follows from the absolute square of the wave function, we expect that after the field is turned on, the state evolves as $\exp\{-i(E-i\frac{1}{2}\Gamma)t\}$. We see that for such an unstable

state the real-valued energy is replaced by a complex resonance $E - i\frac{1}{2}\Gamma$. The important decay rate Γ provides the rate of exciton dissociation.

36.2 Exciton Decay Rate

There are several ways of computing the decay rate Γ. The simplest is the WKB approach introduced in Chapter 5. For instance, for the delta-function case in the left-hand panel of Fig. 36.1, the energy is $E_0 = -Z^2/2$ and the tunnelling barrier height is $V_0 - E = -E_0 = Z^2/2$. Thus, Eq.(5.5) in natural units means that

$$T_{WKB} = \exp\left\{-\frac{4}{3\mathcal{E}}\sqrt{2(Z^2/2)^3}\right\} = \exp\left\{-\frac{2Z^3}{3\mathcal{E}}\right\}. \tag{36.3}$$

Hence, in the WKB approach, we expect the decay rate to vary with field strength as $\Gamma = G(\mathcal{E})\exp\left\{-2Z^3/(3\mathcal{E})\right\}$ with some unknown prefactor $G(\mathcal{E})$ that cannot easily be determined. There is, however, two exact ways of computing the full result. The first one relies on the fact that for the simple delta-function model an exact wave function can be found even in the presence of the field. First, though, we examine the asymptotic behaviour of the general problem Eq.(36.1). For $x \to -\infty$, the potential can be ignored assuming that the asymptotic value vanishes and we then find approximately that $-\frac{1}{2}\varphi''(x) \approx (E - \mathcal{E}x)\varphi(x)$. If we write the wave function as $\varphi(x) = \exp\{if(x)\}$ we then see that $-\frac{i}{2}f''(x) + \frac{1}{2}[f'(x)]^2 = E - \mathcal{E}x$. Suppose we ignore the f'' term. It is not obvious that we can do this but *if* we do we can integrate to find $f(x) \approx [2(E - \mathcal{E}x)]^{3/2}/(3\mathcal{E})$. Thus, f'' and $[f']^2$ vary asymptotically as $|x|^{-1/2}$ and $|x|$, respectively, which means that we are correct in ignoring the f'' term asymptotically. It follows that to the far left, the state varies as $\varphi(x) \sim \exp\{i[2(E - \mathcal{E}x)]^{3/2}/(3\mathcal{E})\}$. A state of the form $\varphi(x) = \exp\{if(x)\}$ with f real is an outgoing wave. In particular, the local momentum $p = (\hat{p}\varphi)/\varphi = f'(x)$ is negative and, so, the state is a left-outgoing wave Hence, this solution represents precisely what we're looking for: a wave propagating from the central well to the far left.

The asymptotic wave function found above applies to any potential. For the delta-functional, the exact solution is easily found, however. Similarly to Chapter 5, the Schrödinger equation for $x \neq 0$ is recognized as the Airy equation. Thus, the general solution is

$$\varphi(x) = A\mathrm{Ai}\left\{(2\mathcal{E})^{1/3}\left(x - \frac{E}{\mathcal{E}}\right)\right\} + B\mathrm{Bi}\left\{(2\mathcal{E})^{1/3}\left(x - \frac{E}{\mathcal{E}}\right)\right\},$$

where Ai and Bi are the two linearly independent Airy functions. They have the following asymptotic forms

$$\mathrm{Ai}(x)\approx\begin{cases}\dfrac{1}{\pi^{1/2}(-x)^{1/4}}\sin\left(\dfrac{2}{3}(-x)^{3/2}+\dfrac{\pi}{4}\right) & x\to-\infty\\ \dfrac{1}{2\pi^{1/2}x^{1/4}}\exp\left(-\dfrac{2}{3}x^{3/2}\right) & x\to\infty.\end{cases}$$

and

$$\mathrm{Bi}(x)\approx\begin{cases}\dfrac{1}{\pi^{1/2}(-x)^{1/4}}\cos\left(\dfrac{2}{3}(-x)^{3/2}+\dfrac{\pi}{4}\right) & x\to-\infty\\ \dfrac{1}{\pi^{1/2}x^{1/4}}\exp\left(\dfrac{2}{3}x^{3/2}\right) & x\to\infty.\end{cases}$$

The divergent behaviour of Bi means that only the Ai solution can be used for $x>0$. For $x<0$, both solutions oscillate and any linear combination is, in principle, acceptable. That is,

$$\varphi(x)=\begin{cases}A\mathrm{Ai}\left\{(2\mathcal{E})^{1/3}\left(x-\dfrac{E}{\mathcal{E}}\right)\right\}+B\mathrm{Bi}\left\{(2\mathcal{E})^{1/3}\left(x-\dfrac{E}{\mathcal{E}}\right)\right\} & x<0\\ C\mathrm{Ai}\left\{(2\mathcal{E})^{1/3}\left(x-\dfrac{E}{\mathcal{E}}\right)\right\} & x>0.\end{cases}\tag{36.4}$$

Now, wave function continuity and the boundary condition $\varphi'(0_-)-\varphi'(0_+)=2Z\varphi(0)$ leads to (with $\lambda=-2^{1/3}E/\mathcal{E}^{2/3}$)

$$A\mathrm{Ai}(\lambda)+B\mathrm{Bi}(\lambda)=C\mathrm{Ai}(\lambda)$$
$$A\mathrm{Ai}'(\lambda)+B\mathrm{Bi}'(\lambda)-C\mathrm{Ai}'(\lambda)=\frac{2^{2/3}Z}{\mathcal{E}^{1/3}}C\mathrm{Ai}(\lambda).$$

Solving these simultaneous equations using the Wronskian $W[\mathrm{Ai},\mathrm{Bi}]=1/\pi$ it is found that

$$\frac{A}{B}=\frac{\dfrac{\mathcal{E}^{1/3}}{2^{2/3}\pi Z}-\mathrm{Ai}(\lambda)\mathrm{Bi}(\lambda)}{\mathrm{Ai}^2(\lambda)}.$$

Now, if the wave function is going to match the condition for the outgoing wave discussed above, the asymptotic forms for $x\to-\infty$ mean that $A/B=i$. Hence,

$$\frac{\mathcal{E}^{1/3}}{2^{2/3}\pi Z}-\mathrm{Ai}(\lambda)\mathrm{Bi}(\lambda)=i\mathrm{Ai}^2(\lambda).\tag{36.5}$$

The numerical solution to this problem for Z = 1 is shown as the solid lines in Fig. 36.3. The dots are the results of the complex scaling approach discussed below and the dashed blue line is the WKB result $\Gamma \approx \exp\left\{-2Z^3/(3\mathcal{E})\right\}$. We see that the first two approaches agree really well whereas the WKB formula is only accurate at very low fields. The numerical solution can be expanded for small fields and one then finds an improved approximation $\Gamma \approx Z^2(1-5\mathcal{E}/3Z^3)\exp\left\{-2Z^3/(3\mathcal{E})\right\}$. The real part, the Stark energy, starts out at $-1/2$ and then decreases following roughly the result of the polarizability $E_0 \approx -Z^2/2 - 5\mathcal{E}^2/8Z^4$, which is shown as the dashed red line. However, for moderate field strengths, the exact result stops decreasing as fast as the approximation would indicate. In fact, for sufficiently high fields, the energy goes through a minimum and then increases.

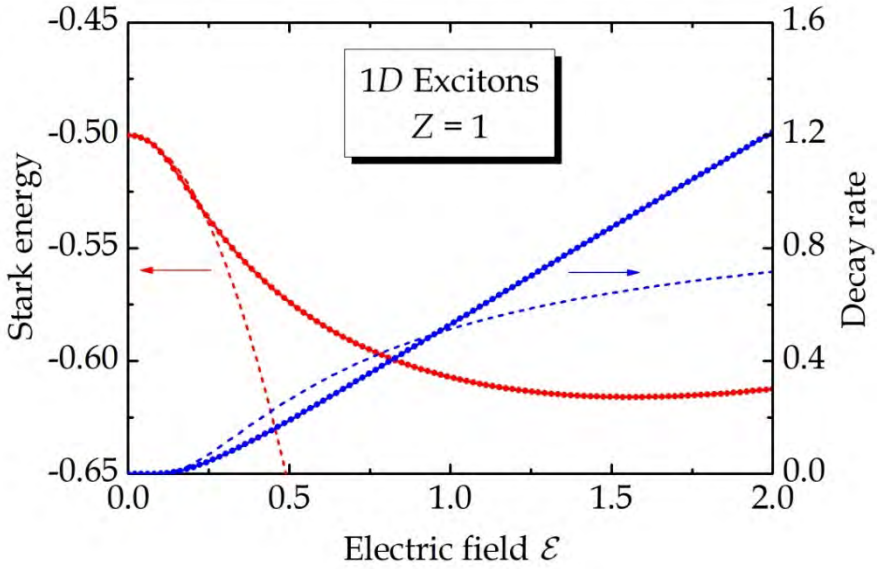


Figure 36.3. Real and imaginary parts of the resonance versus field strength. The full curves are the exact solutions and the dots are for complex scaling. The red and blue dashed lined are, respectively, the second order energy approximation and the WKB result for the decay rate.

The complex scaling method [2], relies on analytically continuing the Schrödinger problem into the complex plane. In short, one first scales every coordinate $\vec{r} \to \vec{r}e^{\theta}$. This yields

$$\left\{-\frac{e^{-2\theta}}{2}\frac{d^2}{dx^2} - e^{-\theta}Z\delta(x) + e^{\theta}\mathcal{E}x\right\}\varphi(x) = E\varphi(x).$$

Second, the scaled equation is then "rotated" into the complex plane by taking θ complex. In fact, most often the parameter is taken to be purely imaginary so that

$$\left\{-\frac{e^{-2i\theta}}{2}\frac{d^2}{dx^2} - e^{-i\theta}Z\delta(x) + e^{i\theta}\mathcal{E}x\right\}\varphi(x) = E(\theta)\varphi(x). \qquad (36.6)$$

Now, this is obviously a non-Hermitian problem and so the eigenvalues $E(\theta)$ will be complex. The magic of the complex scaling method is that: these complex eigenvalues are identical to the resonances $E - i\frac{1}{2}\Gamma$ discussed above! Moreover, the eigenfunctions are localized and normalizable. The exact solution is independent of θ as long as $\theta \neq 0$. In practice, however, this is only approximately true if, for instance, expansion in a finite basis is applied. To demonstrate the powerfulness of the method, we expand the eigenstate in a Laguerre basis of even and odd functions

$$|ns\rangle = \sqrt{\frac{k}{2}} L_n(k|x|)e^{-k|x|/2}, \quad |np\rangle = \frac{k^{3/2}}{2(n+1)} x L_n^1(k|x|)e^{-k|x|/2}.$$

They are normalized according to $\langle ns|ms\rangle = \delta_{nm}$ and $\langle np|mp\rangle = \delta_{nm} - \frac{1}{2}\delta_{n,m\pm1}$. The matrix elements of the potential follow from $L_n(0) = 1$ and for the kinetic energy with $n_< = \min(n,m)$ they are

$$\langle ns|-\frac{1}{2}\frac{d^2}{dx^2}|ms\rangle = \frac{k^2}{2}\left\{n_< + \frac{1}{2} - \frac{1}{4}\delta_{nm}\right\}, \quad \langle np|-\frac{1}{2}\frac{d^2}{dx^2}|mp\rangle = \frac{k^2}{8}\left\{\delta_{nm} + \frac{1}{2}\delta_{n,m\pm1}\right\}.$$

For the non-vanishing elements of the dipole integral we have

$$\langle ns|x|mp\rangle = \frac{1}{\sqrt{2}k}\begin{cases} 3n+2 & m=n \\ -(n+1) & m=n+1 \\ -(3n+1)' & m=n-1 \\ n & m=n-2. \end{cases}$$

Using this basis with $k = 2Z$ and 30 functions of each type, we find the dots in Fig. 36.3. It is clear that agreement with the exact results is practically perfect.

36.3 Higher-Dimensional Excitons

In two and three dimensions, the description of exciton dissociation is more complicated. Again, however, the complex scaling method can be applied quite successfully. We wish to solve the problem

$$\left\{-\frac{e^{-2i\theta}}{2}\nabla_\alpha^2 - \frac{e^{-i\theta}}{r} + e^{i\theta}\mathcal{E}z\right\}\psi = E(\theta)\psi.$$

Here, $\alpha = 2,3$ is the dimension of space. We will keep α as a variable parameter in the analysis because it is possible to formulate quite general results is this way. For an α-dimensional space, we will expand radial states with angular momentum l in a so-called Sturmian basis with $n = l+1, l+2, \ldots$ defined as

$$S_{nl}(r) = N_{nl} r^l e^{-kr} L_{n-l-1}^{2l+\alpha-2}(2kr), \quad k = \frac{2}{\alpha-1}, \quad N_{nl} = \left(\frac{(n-l-1)!}{(n+l+\alpha-3)!} \right)^{1/2} (2k)^{l+1/2}.$$

They are eigenstates of the operator

$$H = -\frac{1}{2r^{\alpha-1}} \frac{\partial}{\partial r} r^{\alpha-1} \frac{\partial}{\partial r} + \frac{l(l+\alpha-2)}{2r^2} - \frac{k(n-\frac{1}{2})}{r} + \frac{k^2}{2},$$

with eigenvalue 0. Also, they are normalized according to

$$\int_0^\infty S_{nl}(r) S_{ml}(r) r^{\alpha-2} dr = \delta_{nm}.$$

For the overlap we find

$$\langle nl | ml \rangle = \int_0^\infty S_{nl}(r) S_{ml}(r) r^{\alpha-1} dr = \begin{cases} (2n+\alpha-3)/2k & n = m \\ -\left[(n_< + l + \alpha - 2)(n_< - l) \right]^{1/2} / 2k & n = m \pm 1 \end{cases}.$$

Similarly, for the radial dipole integral

$$\langle nl | r | ml-1 \rangle^2 = \frac{1}{16k^4} \begin{cases} 9(2n+\alpha-3)^2 (n+\alpha+l-3)(n-l) & m = n \\ (4n+3\alpha+2l-7)^2 (n-l+1)(n-l) & m = n+1 \\ (4n+\alpha-2l-5)^2 (n+\alpha+l-4)(n+\alpha+l-3) & m = n-1 \\ (n+\alpha+l-2)(n-l+2)(n-l+1)(n-l) & m = n+2 \\ (n+\alpha+l-5)(n+\alpha+l-4)(n+\alpha+l-3)(n-l-1) & m = n-2. \end{cases}$$

Writing $z = r\cos\phi$, the angular part of the wave function $\chi_l(\phi)$ is a Gegenbauer polynomial and it can be shown that

$$\langle \chi_l | \cos\phi | \chi_{l-1} \rangle = \left(\frac{l(l+\alpha-3)}{(2l+\alpha-4)(2l+\alpha-2)} \right)^{1/2}.$$

Writing $\psi(r) = \sum_n c_{nl} S_{nl}(r)$, the unperturbed αD hydrogen eigenvalue problem for angular momentum l reads

$$\left(E + \frac{k^2}{2} \right) \sum_m \langle nl | ml \rangle c_{ml} = (n-1) k c_{nl}.$$

The perturbed problem including complex scaling is obviously more involved. The field couples angular momenta l and $l+1$ and so the Hamiltonian and overlap matrix have the structure

$$\overleftrightarrow{H}(\theta)=\begin{pmatrix} H_{00} & H_{01} & 0 & \cdots \\ H_{10} & H_{11} & H_{12} & \cdots \\ 0 & H_{21} & H_{22} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \quad \overleftrightarrow{S}=\begin{pmatrix} S_{00} & 0 & 0 & \cdots \\ 0 & S_{11} & 0 & \cdots \\ 0 & 0 & S_{22} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

The elements of the non-vanishing blocks are

$$S_{ll}(n,m)=\tfrac{1}{2k}(2n+\alpha-3)\delta_{nm}-\tfrac{1}{2k}\sqrt{(n_<+l+\alpha-2)(n_<-l)}\,\delta_{n,m\pm1}$$
$$H_{ll}(n,m)=\left\{e^{-2i\theta}\left((n-1)k+1\right)-e^{-i\theta}\right\}\delta_{nm}-\tfrac{k^2}{2}e^{-2i\theta}S_{ll}(n,m)$$
$$H_{l,l-1}(n,m)=e^{i\theta}\mathcal{E}\langle nl|r|ml-1\rangle\langle\chi_l|\cos\phi|\chi_{l-1}\rangle.$$

The result of a numerical diagonalization using angular momenta in the range $l\in[0,10]$ and 20 basis states for each l is shown in Fig. 36.4. Note the close resemblance to the 1D result in Fig. 36.3.

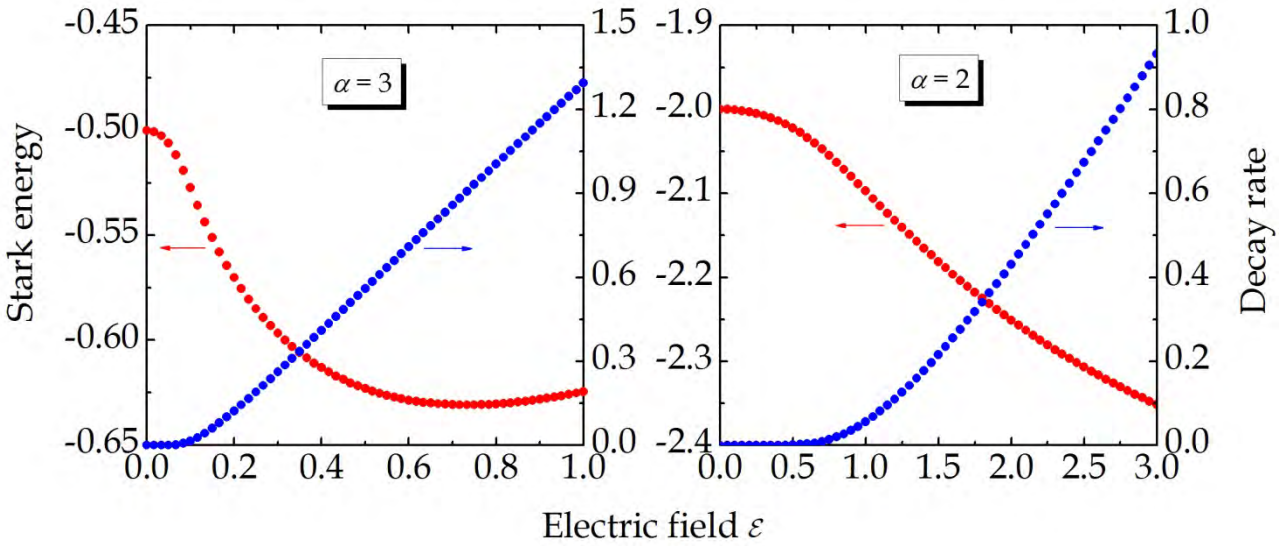


Figure 36.4. Complex scaling approach to the resonances in 3D and 2D excitons.

Exercise: One-dimensional Coulomb problem

In this exercise, we return to the problem of tunnelling out of a 1D Coulomb potential corresponding to the situation in the right-hand panel in Fig. 36.1. Actually, the full potential will be taken to be $V(x)=-Z/(|x|+d)$. This is a regularized potential with a finite binding energy, in contrast to the pure Coulomb potential found to $d=0$. Physically, d represents the diameter of the quantum wire or nanotube. For $Z=1$, the unperturbed eigenstates are so-called Whittaker functions given by [3]

$$\psi(x)=NW_{1/\kappa,1/2}\left\{2\kappa(|x|+d)\right\},\quad E=-\kappa^2/2.$$

The even states, which include the ground state, have vanishing derivative at the origin so that the eigenvalue condition is $W'_{1/\kappa,1/2}(2\kappa d)=0$.

a) Write a program that finds the ground state energy. The result should resemble the curve in Fig. 36.5.

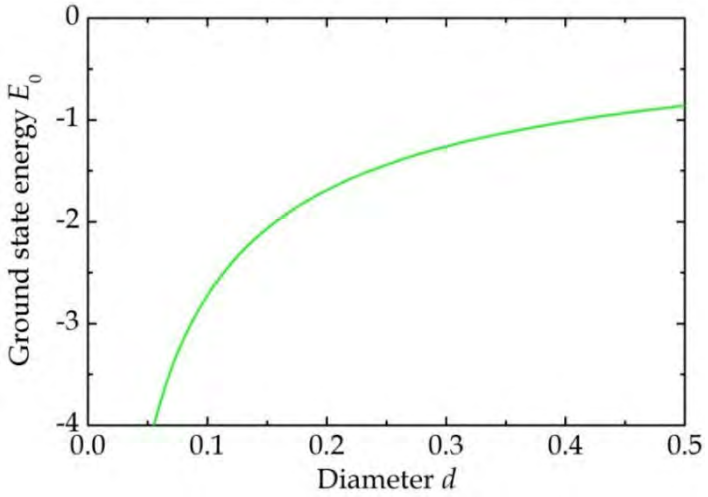


Figure 36.5. Ground state energy of the truncated Coulomb potential.

b) Use the 1D Laguerre basis to implement complex scaling for this potential. The result for $d=0.1$, $\theta=0.4$ using 70 s-type and 70 p-type functions with $k=2$ is illustrated in Fig. 36.6. Note that the unperturbed energy is around -2.72 in agreement with the result found above.

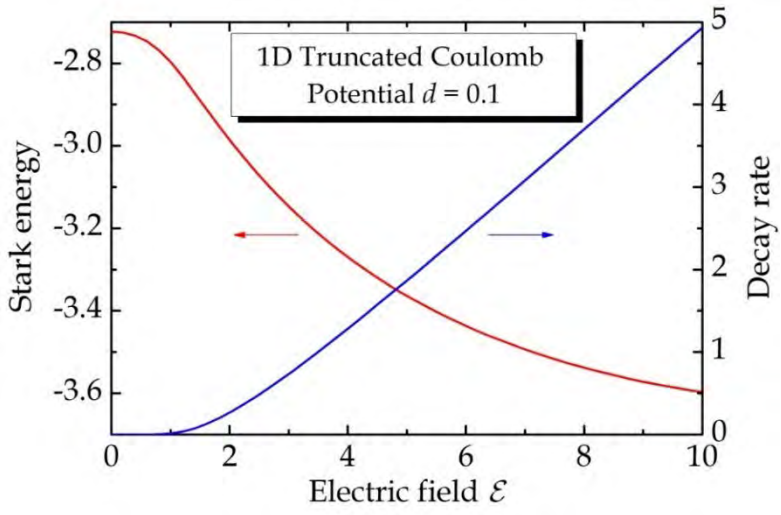


Figure 36.6. Stark shift and decay rate for the truncated Coulomb potential assuming d = 0.1.

We will now investigate the WKB approximation for this problem. After switching on the electric field, the classical turning points are defined by

$$Z/(|x|+d)+\mathcal{E}|x|=|E_0| \quad \Rightarrow \quad |x_{1,2}|=a-d\pm b, \quad a=\frac{|E_0|+d\mathcal{E}}{2\mathcal{E}}, \quad b=\frac{\sqrt{(|E_0|+d\mathcal{E})^2-4Z\mathcal{E}}}{2\mathcal{E}}.$$

The two turning points merge if the field exceeds a critical magnitude $\mathcal{E}_{crit} = (-d|E_0| + 2Z - 2\sqrt{Z^2 - Zd|E_0|})/d^2$. Hence, the approximate transmittance will be unity beyond this field strength.

c) Assuming a field below the critical value, show that the WKB transmittance becomes

$$T_{WKB} = \exp\left\{-2^{3/2}\mathcal{E}^{1/2}\int_{-b}^{b}\sqrt{\frac{b^2-x^2}{x+a}}dx\right\}.$$

This integral results in a complicated combination of elliptic integrals

$$T_{WKB} = \exp\left\{-\frac{8}{3}\sqrt{2\mathcal{E}(a+b)}\left[aE\left(\frac{2b}{a+b}\right)-(a-b)K\left(\frac{2b}{a+b}\right)\right]\right\}. \qquad (36.7)$$

This expression can be approximated for small fields as

$$T_{WKB} \approx \left(\frac{16|E_0|^2}{\mathcal{E}Z}\right)^{(2Z^2/|E_0|)^{1/2}} \exp\left\{-\frac{2(2|E_0|)^{3/2}}{3\mathcal{E}}-\frac{\sqrt{2}(2d|E_0|-Z)}{\sqrt{|E_0|}}\right\}. \qquad (36.8)$$

d) Plot and compare the full and approximate WKB expressions as in Fig. 36.7.

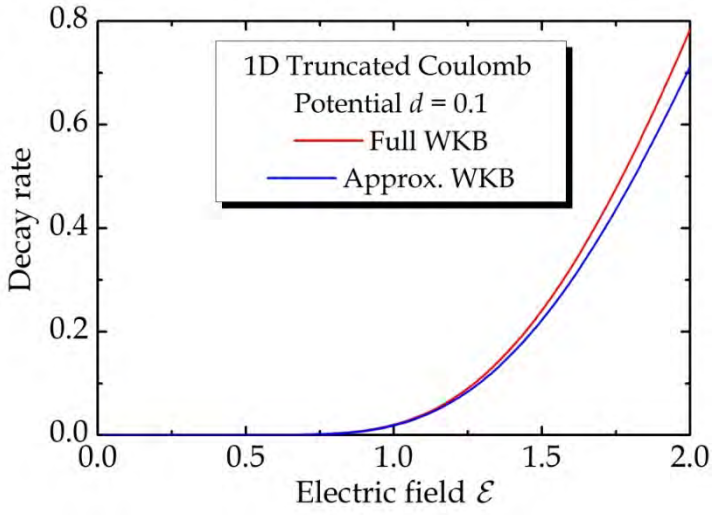


Figure 36.7. WKB decay rate for the truncated Coulomb potential assuming d = 0.1.

References

[1] A. Dalgarno and J.T. Lewis, Proc. R. Soc. Lond. A240, 284 (1957).
[2] I.W. Herbst and B. Simon, Phys. Rev. Lett. 41, 67 (1978).
[3] R. Loudon, Am. J. Phys. 27, 649 (1959).

37. Excitons in Molecular Chains

This chapter is intended as an example of excitons in simple geometries. The general theory derived in Chapter 18 will be used but many results are derived from scratch in order to provide a self-contained exposition. We will compare independent-particle spectra, Bethe-Salpeter spectra, and Wannier exciton results. The chain (polymethineimine, PMI) considered throughout is shown in Fig. 37.1.

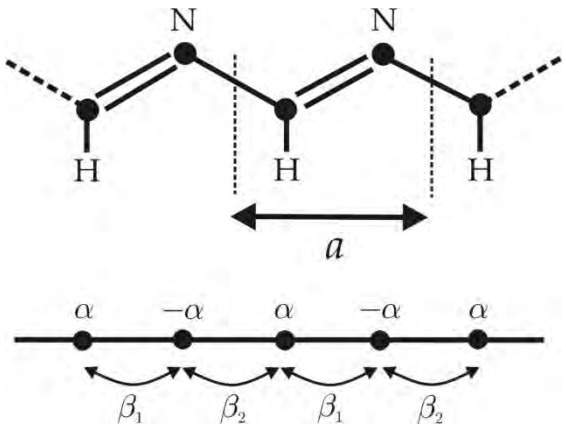


Figure 37.1. The PMI polymer and its tight-binding model.

It consists of alternating C and N atoms along the backbone and, in addition, is dimerized with alternating single and double bonds. We only consider the π - orbitals directed perpendicularly to the molecule. Only nearest neighbour interactions are included and the orbitals are taken to be orthogonal. We take the values $a = 4$ Å, $\alpha = 1.0$ eV, and $\beta_1 = \beta_2 \equiv \beta = -3.0$ eV, i.e. we ignore dimerization.

37.1 Independent Particle Approximation

If an infinite chain is considered in *k*-space the two orbitals per unit cell lead to the Hamiltonian for a single electron

$$\overleftrightarrow{H} = \begin{pmatrix} \alpha & 2\beta\cos(ka/2) \\ 2\beta\cos(ka/2) & -\alpha \end{pmatrix}.$$

The momentum operator becomes $\hat{p} = (m/\hbar)\, d\hat{H}/dk$ and we find the eigenvalues and matrix elements

$$E_{c,v}(k) = \pm\varepsilon, \quad \varepsilon = \sqrt{\alpha^2 + 4\beta^2\cos^2(ka/2)}$$

$$p_{cv}(k) = -\frac{ma\alpha\beta}{\hbar\varepsilon}\sin(ka/2).$$

The band gap is at $k=\pi/a$ and if we expand around this value we find

$$E_{cv}(k)=2\varepsilon\approx E_g+\frac{\hbar^2(k-\pi/a)^2}{2m_{eh}},\quad E_g=2|\alpha|,\quad m_{eh}=\frac{\hbar^2|\alpha|}{2a^2\beta^2}$$

$$p_{cv}(k)\approx p_{cv}\equiv-\frac{ma\beta}{\hbar}.$$

We wish to compute the linear optical response. To this end we use the expression Eq.(15.3) adapted to a 1D geometry with cross section A

$$\chi(\omega)=\frac{2e^2\hbar^2}{\pi\varepsilon_0 m^2 A}\int_0^{2\pi/a}\frac{|p_{cv}(k)|^2}{E_{cv}(k)(E_{cv}^2(k)-\hbar^2\omega^2)}dk. \tag{37.1}$$

We will compute this function normalized by $\chi_0=e^2\beta^2a/(\varepsilon_0E_g^3A)$. The result for the imaginary (absorptive) part obtained using the energies and momenta above is shown as the black curve in Fig. 37.2 using a broadening of 30 meV. It is seen to peak around the band gap of 2 eV.

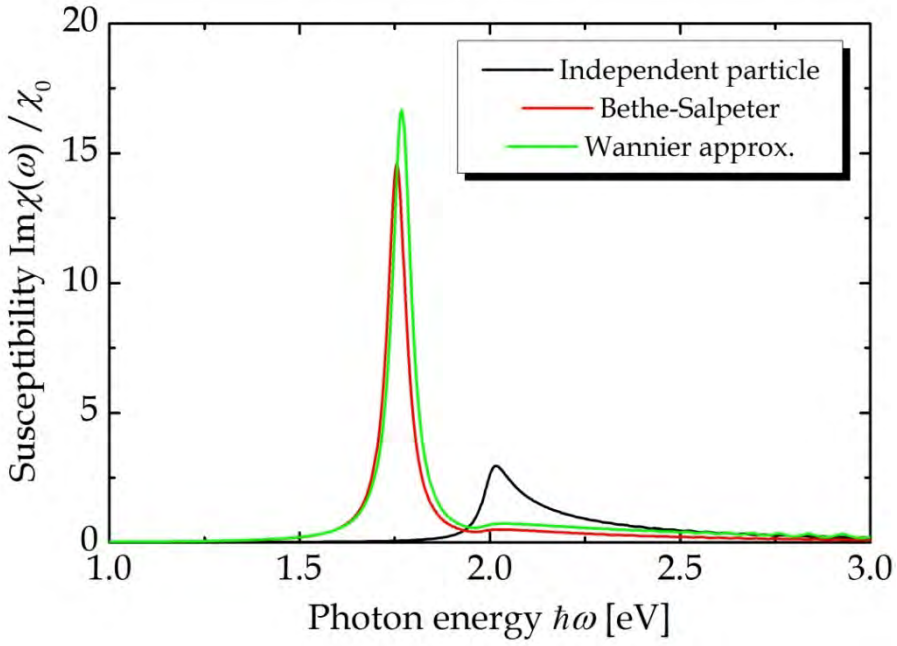


Figure 37.2. Absorption spectra within three computational schemes.

37.2 Bethe-Salpeter Equation

We now aim to include electron-hole interactions in the theory. The ground state $|0\rangle$ of the semiconductor is a Slater determinant constructed from all valence band single-electron states. Considering N k-vectors in each band and suppressing spin

$$|0\rangle=|v_{k_1},v_{k_2},...,v_{k_N}|.$$

Similarly, the excited states are constructed from Slater determinants with one or more valence band states replaced by conduction band states. If only a single state is replaced, we find single-excitations of the form $\left|v_{k_i} \to c_{k_i}\right\rangle = \left| v_{k_1}, v_{k_2}, ..., c_{k_i}, ..., v_{k_N} \right|$. If these excitations were taken to be the exact excited states, we would get back to the independent-particle approximation. Hence, to go beyond, we form the actual excited states $\left|n\right\rangle$ as linear combinations, i.e.

$$\left|n\right\rangle = \sum_i \psi_{k_i}^{(n)} \left|v_{k_i} \to c_{k_i}\right\rangle .$$

These states are the so-called excitons. They are determined from the Bethe-Salpeter eigenvalue problem. Ignoring exchange effects, one finds

$$E_{cv}(k)\psi_k - \sum_q V(k,q)\psi_q = E\psi_k . \tag{37.2}$$

Here, V is the screened Coulomb attraction between electron and holes. We now specialize to tight-binding band states. In general dimensions, the band states for $\alpha = c, v$ are of the form

$$\varphi_{\alpha\vec{k}}(\vec{r}) = \sum_n C_n^{\alpha\vec{k}} \chi_{n\vec{k}}(\vec{r}), \quad \chi_{n\vec{k}}(\vec{r}) = \frac{1}{\sqrt{N_{uc}}} \sum_{\vec{R}} e^{i\vec{k}\cdot\vec{R}} \phi_n(\vec{r} - \vec{R}) .$$

Here, N_{uc} is the number of unit cells and $\chi_{n\vec{k}}(\vec{r})$ is the n'th Bloch sum, i.e. the Bloch sum for the n'th atomic orbital $\phi_n(\vec{r})$ in the unit cell. The summation in $\chi_{n\vec{k}}(\vec{r})$ is over all unit cells. Next, we exploit the fact that the atomic orbitals are localized and orthogonal. Hence, a product of band states becomes

$$\varphi_{\alpha\vec{k}}^*(\vec{r})\varphi_{\alpha\vec{q}}(\vec{r}) \approx \frac{1}{N_{uc}} \sum_{\vec{R},n} C_n^{\alpha\vec{k}*} C_n^{\alpha\vec{q}} e^{i(\vec{q}-\vec{k})\cdot\vec{R}} \phi_n^2(\vec{r} - \vec{R}) .$$

If the real-space Coulomb interaction is $V(\vec{r}) = e^2 / (4\pi\varepsilon\varepsilon_0 r)$, with ε the (assumed isotropic) dielectric constant, we then find for the electron-hole interaction matrix element

$$V(k,q) = \iint \varphi_{c\vec{k}}^*(\vec{r})\varphi_{c\vec{q}}(\vec{r}) V(\vec{r} - \vec{r}') \varphi_{v\vec{q}}^*(\vec{r}')\varphi_{v\vec{k}}(\vec{r}') d^3r d^3r' .$$

We need integrals of the form

$$V_{nm}(\vec{R}, \vec{R}') = \iint \phi_n^2(\vec{r} - \vec{R}) V(\vec{r} - \vec{r}') \phi_m^2(\vec{r}' - \vec{R}') d^3r d^3r' .$$

Clearly, for large separations we find an interaction $V_{nm}(\vec{R},\vec{R}') \propto |\,\vec{R}-\vec{R}'\,|^{-1}$. On the other hand, for $\vec{R}=\vec{R}'$ and $n=m$ we find a finite result (the Hubbard U parameter). As an appropriate interpolation, we will assume the effective interaction

$$V_{nm}(\vec{R},\vec{R}') \approx V_{eff}(\vec{R}-\vec{R}') \equiv \frac{e^2}{4\pi\varepsilon\varepsilon_0\sqrt{l^2+|\,\vec{R}-\vec{R}'\,|^2}}.$$

This is sometimes called the Ohno form. Introducing then $I_{\alpha\vec{k},\alpha\vec{q}} \equiv \sum_n C_n^{\alpha\vec{k}*} C_n^{\alpha\vec{q}}$ we find

$$V(\vec{k},\vec{q}) \approx I_{c\vec{k},c\vec{q}} I_{v\vec{q},v\vec{k}} \frac{1}{N_{uc}} \sum_{\vec{R}} V_{eff}(\vec{R}) e^{i(\vec{q}-\vec{k})\cdot\vec{R}} \approx I_{c\vec{k},c\vec{q}} I_{v\vec{q},v\vec{k}} \frac{1}{\Omega} \int V_{eff}(\vec{R}) e^{i(\vec{q}-\vec{k})\cdot\vec{R}} d^3R\,.$$

The last form is obtained by converting the summation over a slowly varying function to an integral. In our 1D model, we find

$$V(k,q) \approx \frac{2\pi}{L} I_{ck,cq} I_{vq,vk} W(k-q), \quad W(k-q) = \frac{1}{2\pi}\int V_{eff}(x)\exp(i(q-k)x)dx = \frac{V_0}{\pi} K_0\left(|\,k-q\,|\,l\right),$$

$$V_0 = \frac{e^2}{4\pi\varepsilon\varepsilon_0} = \frac{\mathrm{Ha}\cdot a_B}{\varepsilon}.$$

We convert the q-sum into an integral

$$E_{cv}(k)\psi_k - \int_0^{2\pi/a} I_{ck,cq} I_{vq,vk} W(k-q)\psi_q dq = E\psi_k\,. \tag{37.3}$$

The eigenvalue problem can be discretized on a k-grid with the diagonal element

$$W(0) \approx \frac{V_0}{\pi}\left\{K_0(l\Delta k)+1\right\}.$$

The matrix elements of the many-body momentum operator $\hat{P} = \sum_i \hat{p}_i$ with a sum over all electrons become $P_n \equiv \langle 0|\hat{P}|n\rangle = \sum_i \psi_{k_i}^{(n)} p_{vc}(k_i)$. Normalization is such that $(2\pi/L)\sum_i |\,\psi_{k_i}^{(n)}\,|^2 = 1$ with $L = Na$. The many-body version of the susceptibility is given by the 1D version of Eq.(18.1)

$$\chi(\omega) = \frac{4e^2\hbar^2}{\varepsilon_0 m^2 L}\sum_n \frac{|\,P_n\,|^2}{E_n(E_n^2-\hbar^2\omega^2)}. \tag{37.4}$$

Taking $\varepsilon = 5$ and $l = 3$ Å and assuming $I_{\alpha k,\alpha q} \approx 1$, we find the red curve in Fig. 37.2. Note the large exciton binding energy of around 0.25 eV.

37.3 Wannier Approximation

The Wannier approximation to the Bethe-Salpeter equation is obtained by a series of approximations. Primarily, we use $I_{\alpha k,\alpha q} \approx 1$ but we also use the effective mass approximation for the band-to-band energy. Shifting the *k*-vector origin to the band gap location we then find

$$\left\{E_g + \frac{\hbar^2 k^2}{2m_{eh}}\right\}\psi_k - \frac{1}{2\pi}\int\int V_{eff}(x)\exp(i(k-q)x)dx\psi_q dq = E\psi_k .$$

This problem can be recast as a problem in real space rather than momentum space by introducing the Fourier transform, c.f. section 18.1

$$\psi(x) = \frac{1}{\sqrt{L}}\sum_k \psi_k e^{ikx} = \frac{\sqrt{L}}{2\pi}\int \psi_k e^{ikx} dk . \tag{37.5}$$

The prefactor is for normalization $\int |\psi(x)|^2 \, dx = 1$. Fourier transforming the approximate Bethe-Salpeter equation then yields

$$\left\{E_g - \frac{\hbar^2}{2m_{eh}}\frac{d^2}{dx^2}\right\}\psi(x) - V_{eff}(x)\psi(x) = E\psi(x) . \tag{37.6}$$

This is the famous Wannier equation, in which *x* can be interpreted as the electron-hole separation (relative coordinate). It is mathematically identical to the (one-dimensional) hydrogen atom. If we introduce natural exciton units such as $a_B^* = a_B \cdot \varepsilon(m_0 / m_{eh})$ and $\mathrm{Ha}^* = \mathrm{Ha}\cdot(m_{eh} / m_0)/\varepsilon^2$ for distance and energy, respectively, and measure energies relative to the band gap, we can write

$$\left\{-\frac{1}{2}\frac{d^2}{dx^2} - \frac{1}{\sqrt{x^2 + l^2}}\right\}\psi(x) = E\psi(x) . \tag{37.7}$$

Once the eigenstates are computed we need the momentum matrix elements. For this purpose, we ignore the *k*-dependence $p_{cv}(k) \approx p_{cv}$. Hence,

$$\langle 0|P|n\rangle = \sum_k \psi_k^{(n)} p_{vc}(k) \approx p_{vc}\sum_k \psi_k^{(n)} = \sqrt{L} p_{vc}\psi_n(0) .$$

Finally, the susceptibility Eq.(37.4) becomes

$$\chi(\omega)=\frac{4e^2\hbar^2\,|\,p_{vc}\,|^2}{\varepsilon_0 m^2}\sum_n\frac{|\,\psi_n(0)\,|^2}{(E_g+E_n)\left[(E_g+E_n)^2-\hbar^2\omega^2\right]}.$$

The fact that $|\,\psi_n(0)\,|^2$ enters is natural because it is a measure of the chance that electrons and holes overlap. For the present parameters, the exciton Bohr radius and Hartree are 100 Å and 29 meV, respectively. We solve for the eigenstates using the Laguerre basis in Chapter 36 taking k = 10. The corresponding absorption is plotted as the green curve in Fig. 37.2. If, finally, an electrostatic field is added we get the modified form

$$\left\{-\frac{1}{2}\frac{d^2}{dx^2}-\frac{1}{\sqrt{x^2+l^2}}+\mathcal{E}x\right\}\psi(x)=E\psi(x).$$

The exciton Stark energy and ionization rate derived from this equation are illustrated in Fig. 37.3 using the techniques discussed in Chapter 36.

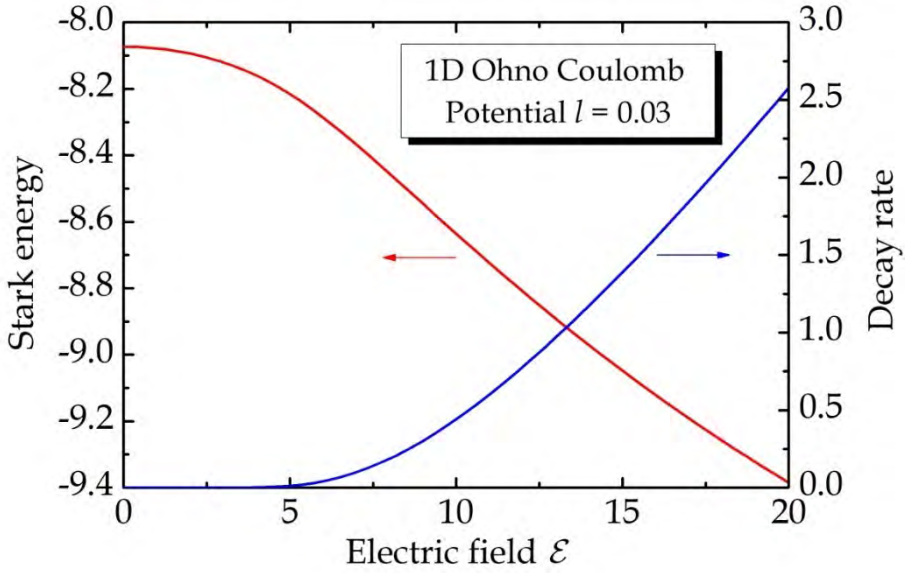


Figure 37.3. Stark energy and ionization rate for the Ohno potential.

Exercise: Variational estimate of the ground state

In this exercise, we will try to estimate the energy of the ground state of the unperturbed Wannier problem Eq.(37.7) using the ansatz

$$\psi(x)=N(1+a\,|\,x\,|)\exp(-q\,|\,x\,|).$$

a) Show that the normalization condition leads to $N=q^{3/2}/(q^2+aq+a^2/2)^{1/2}$.
The kinetic energy is a little tricky because of the absolute values.

b) Use the result $d^2f(|\,x\,|)/dx^2=f''(|\,x\,|)+2\delta(x)f'(0)$ to show that

$$-\frac{1}{2}\int_{-\infty}^{\infty}\psi(x)\frac{d^2}{dx^2}\psi(x)dx = N^2\frac{a^2-2aq+2q^2}{4q}.$$

The Coulomb energy is even trickier. We want the integral

$$V(l) = -2\int_0^{\infty}\frac{1}{\sqrt{x^2+l^2}}g(x)dx.$$

with $g(x)=\psi^2(x)$. The cutoff length l is small but we cannot just put it to 0 because the integral diverges.

c) Show using partial integration twice that

$$V(l) = 2g(0_+)\ln l + 2g'(0_+)l - 2\int_0^{\infty}\left\{x\ln\left(x+\sqrt{x^2+l^2}\right)-\sqrt{x^2+l^2}\right\}g''(x)dx.$$

By setting l = 0 in the last term it can be shown that approximately (with $\gamma \approx 0.577...$ Euler's constant.)

$$V(l) \approx -N^2\frac{a^2+4aq-4\gamma q^2+8lq^2(q-a)-4q^2\ln(ql)}{2q^2},$$

d) Finally, write a program that minimizes with respect to q and a. The result should be the curve below.

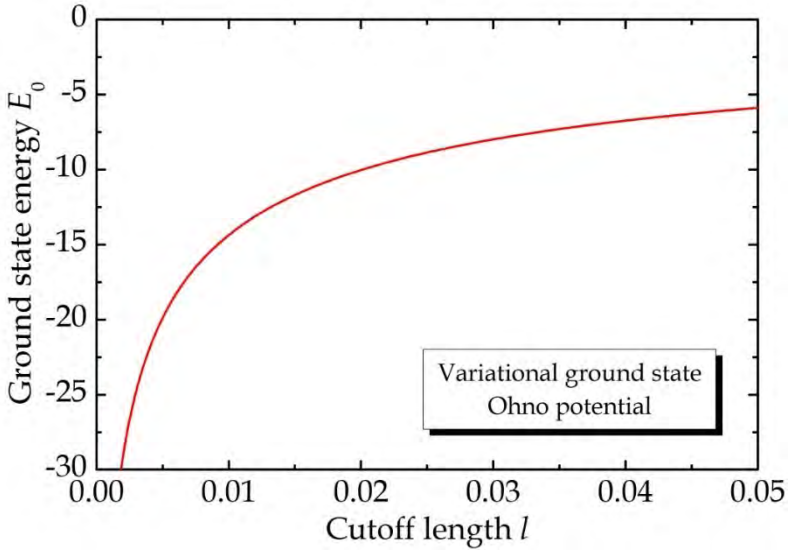


Figure 37.4. Variational ground state energy for the Ohno potential.

For l = 0.03, one finds an energy of -7.98 in good agreement with the numerically exact value of -8.07.

38. Hall Effect

In this chapter, we will try to describe the combined effects of electric and magnetic fields. When these are simultaneously present, new phenomena emerge. The most prominent situation is illustrated in Fig. 38.1, in which the fields are perpendicular with $\vec{\mathcal{E}} = \mathcal{E}\vec{e}_x$ and $\vec{\mathcal{B}} = \mathcal{B}\vec{e}_z$. We assume an electric current j_x is initially injected along the x-direction due to the electric field. In this way, an electron that moves along the (negative) x-axis, will experience a deflecting Lorentz force from the magnetic field $-e\vec{v}\times\vec{\mathcal{B}}$. Since $\vec{v} = -v\vec{e}_x$, the magnetic force is along the negative y-axis. Hence, a transverse current j_y is induced. In terms of conductivities, $j_y = \sigma_{yx}\mathcal{E}$ and, as usual, $j_x = \sigma_{xx}\mathcal{E}$. Note that if the electric field were along y, we would still find a current deflection to the left and, in this case, $j_y = \sigma_{yy}\mathcal{E}_y$ and $j_x = \sigma_{xy}\mathcal{E}_y$, where, for an isotropic material, $\sigma_{yy} = \sigma_{xx}$ and $\sigma_{yx} = -\sigma_{xy}$. This transverse current induced by the magnetic field is the Hall effect. In a sample with a finite extent along y, the deflected charge leads to a build-up of a transverse voltage, the Hall voltage.

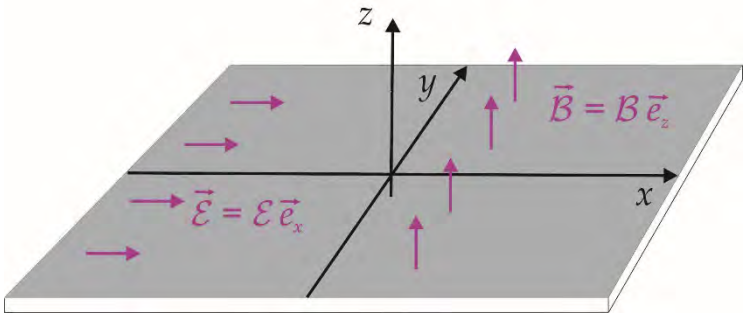


Fig. 38.1. Geometry of the Hall effect with perpendicular electric and magnetic fields.

38.1 Classical Picture

In order to describe the Hall effect, we will first adopt a completely classical picture. Newton's equation of motion for an electron including a relaxation term with relaxation time τ reads

$$m\frac{d\vec{v}}{dt} + \frac{m}{\tau}\vec{v} = -e\vec{\mathcal{E}} - e\vec{v}\times\vec{\mathcal{B}}\,.$$

We assume steady state, i.e. $d\vec{v}/dt = 0$. For the field directions in Fig. 38.1, it follows that

$$\frac{m}{\tau}\begin{pmatrix} v_x \\ v_y \end{pmatrix} = -e\mathcal{E}\begin{pmatrix} 1 \\ 0 \end{pmatrix} - e\mathcal{B}\begin{pmatrix} v_y \\ -v_x \end{pmatrix}.$$

The lower of these immediately implies $v_y = \omega_c \tau v_x$, where $\omega_c = e\mathcal{B}/m$ is the cyclotron frequency. This is precisely the deflection that we discussed above. Solving, we readily find

$$v_x = -\frac{e\tau}{m}\frac{\mathcal{E}}{1+(\omega_c\tau)^2}.$$

The current density is then found from $\vec{j} = -en\vec{v}$, with n the electron density, and writing $\vec{j} = \vec{\vec{\sigma}}\cdot\vec{\mathcal{E}}$ we find

$$\sigma_{xx} = \frac{e^2\tau}{m}\frac{n}{1+(\omega_c\tau)^2}, \quad \sigma_{yx} = -\sigma_{xy} = \frac{e^2\tau}{m}\frac{n\omega_c\tau}{1+(\omega_c\tau)^2}. \tag{38.1}$$

In ordinary materials of normal quality, the cyclotron frequency is small compared to $1/\tau$ and we recover the longitudinal conductivity $\sigma_{xx} = e^2\tau n/m$ and find the simple result for the Hall conductivity $\sigma_{xy} = -e^3\tau^2 n\mathcal{B}/m^2$. Note that, if electrons are replaced by holes of the same density so that the charge changes sign, σ_{xx} remains the same. In contrast, σ_{xy} changes sign. This means that, apparently, the sign of the Hall effect can be used to distinguish electrons from holes. This important feature will be confirmed in a more rigorous computation using the semiclassical picture below. Also, the mass m is expected to be replaced by some sort of effective mass in a more elaborate theory.

38.2 Semiclassical Picture

In the semiclassical picture, the Boltzmann equation for the electron distribution in both position and momentum space is solved. In the situation, in which no explicit time-dependence of the distribution is found and everything is homogeneous in space, Eqs. (12.2) and (12.3) show that the distribution function g obeys

$$g(\vec{k}) = f(E_{\vec{k}}) - \tau\nabla_{\vec{k}}g(\vec{k})\cdot\frac{d\vec{k}}{dt}.$$

Here, f is the unperturbed distribution, i.e. the Fermi function. Since $\vec{p} = \hbar\vec{k}$ is the momentum and $d\vec{p}/dt$ is the Lorentz force, we find

$$g(\vec{k}) = f(E_{\vec{k}}) + \frac{e\tau}{\hbar}\nabla_{\vec{k}}g(\vec{k})\cdot\left(\vec{\mathcal{E}} + \vec{v}\times\vec{\mathcal{B}}\right). \tag{38.2}$$

We now aim to solve this equation order by order in the external perturbations. The 0'th order is the unperturbed case $g_0(\vec{k}) = f(E_{\vec{k}})$ and the expression valid to first order in the force is then $g_1(\vec{k}) = f(E_{\vec{k}}) + (e\tau/\hbar)\nabla_{\vec{k}} f(E_{\vec{k}}) \cdot (\vec{\mathcal{E}} + \vec{v} \times \vec{\mathcal{B}})$. Using the chain rule and remembering that $\vec{v} = \hbar^{-1}\nabla_{\vec{k}} E_{\vec{k}}$ we have $g_1(\vec{k}) = f(E_{\vec{k}}) + e\tau f'(E_{\vec{k}})\vec{v} \cdot \vec{\mathcal{E}}$ with $f' = \partial f / \partial E$ because $\vec{v} \cdot (\vec{v} \times \vec{\mathcal{B}}) = 0$. This is in agreement with Eq.(12.5). Inserting this result back into Eq.(38.2) we find the distribution valid to first order in the magnetic field

$$g_2(\vec{k}) = f(E_{\vec{k}}) + e\tau f'(E_{\vec{k}})\vec{v} \cdot \vec{\mathcal{E}} + \frac{e^2\tau^2}{\hbar}\nabla_{\vec{k}}\left(f'(E_{\vec{k}})\vec{v} \cdot \vec{\mathcal{E}}\right) \cdot (\vec{v} \times \vec{\mathcal{B}}). \qquad (38.3)$$

Again, $\nabla_{\vec{k}} f'(E_{\vec{k}}) \propto \vec{v}$ and we find for the current

$$\vec{j} = -\frac{e}{4\pi^3}\int \vec{v} g_2(\vec{k}) d^3k = -\frac{e^2\tau}{4\pi^3}\int \vec{v} f'(E_{\vec{k}})(\vec{v} \cdot \vec{\mathcal{E}}) d^3k - \frac{e^3\tau^2}{4\pi^3\hbar}\int \vec{v} f'(E_{\vec{k}}) \nabla_{\vec{k}}(\vec{v} \cdot \vec{\mathcal{E}}) \cdot (\vec{v} \times \vec{\mathcal{B}}) d^3k .$$

For the geometry used here, we then get

$$\sigma_{xy} = \frac{e^3\tau^2\mathcal{B}}{4\pi^3\hbar}\int f'(E_{\vec{k}}) v_y \left\{\frac{\partial v_x}{\partial k_x} v_y - \frac{\partial v_x}{\partial k_y} v_x\right\} d^3k . \qquad (38.4)$$

In the low-temperature limit, we have $f' = -\delta(E - E_F)$ and the integral vanishes unless the Fermi level is inside a band. Hence, completely filled bands have zero contribution. In the isotropic effective mass approximation, $\vec{v} = \hbar\vec{k}/m_e$ and it follows that

$$\begin{aligned}\sigma_{xy} &= \frac{e^3\tau^2\hbar^2\mathcal{B}}{4\pi^3 m_e^3}\int f'(E_{\vec{k}}) k_y^2 d^3k = \frac{e^3\tau^2\mathcal{B}}{6\pi^3 m_e^2}\int f'(E_{\vec{k}}) E_{\vec{k}} d^3k \\ &= \frac{2e^3\tau^2\mathcal{B}}{3m_e^2}\int f'(E) E D(E) dE = -\frac{e^3\tau^2\mathcal{B}}{m_e^2} n.\end{aligned}$$

So this agrees with the classical picture. In deriving this result, we utilized the fact that the *k*-space integral can be averaged over all three directions for isotropic materials. The last reformulation uses the effective-mass density of states $D(E) = 3n/(2E)$ and the low-temperature limit. Note, also, that the result changes sign if a downward curving band is considered, i.e. if we replace $m_e \to -m_h$. This demonstrates that the sign of the Hall conductivity depends on the sign of the dominating carriers.

For graphene with $E_F > 0$ in the Dirac approximation, $\vec{v} = v_F \vec{k}/k$ so that $\partial v_x / \partial k_x = v_F k_y^2 / k^3$ and $\partial v_x / \partial k_y = -v_F k_x k_y / k^3$. A simple calculation of the Hall conductivity (normalized by area, not volume) then leads to $\sigma_{xy} = -e^3 v_F^2 \tau^2 \mathcal{B} / (\pi \hbar^2)$ after multiplication by the valley degeneracy. For $E_F < 0$, the sign changes. In Fig. 38.2, we compare this approximate result to a calculation using the full tight-binding dispersion (see Chapter 17) in Eq.(38.4) assuming a temperature of 50 K. We note that the response changes sign as the Fermi level crosses the Dirac point.

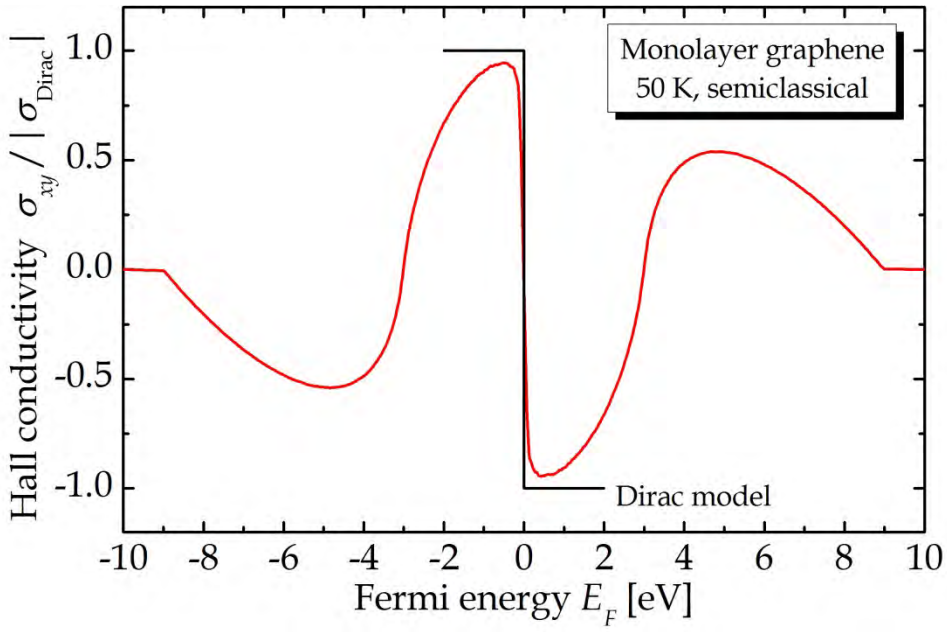


Fig. 38.2. Hall conductivity of monolayer graphene in a full tight-binding description divided by the Dirac approximation.

38.3 Quantum Picture for Massive Electrons

We now turn to a fully quantum mechanical analysis of the Hall effect. We begin by noting that such a treatment would be required in very high quality samples, in which transport is ballistic rather than diffusive. This corresponds to the limit $\tau \to \infty$ of infinite time between scattering events. In this case, Eq. (38.1) yields

$$\sigma_{xx} \approx 0, \quad \sigma_{xy} \approx -\frac{e^3 n \mathcal{B}}{m^2 \omega_c^2} = -\frac{en}{\mathcal{B}}. \tag{38.5}$$

The resistivity tensor is given by $\overleftrightarrow{\rho} = \overleftrightarrow{\sigma}^{-1}$ and one easily finds

$$\overleftrightarrow{\rho} = \begin{pmatrix} 0 & -en/\mathcal{B} \\ en/\mathcal{B} & 0 \end{pmatrix}^{-1} = \begin{pmatrix} 0 & \mathcal{B}/(en) \\ -\mathcal{B}/(en) & 0 \end{pmatrix}.$$

Hence, the longitudinal resistance ρ_{xx} actually vanishes in this limit. The question is, of course, to what extent a purely classical calculation can be trusted in this situation.

To answer this question, we begin by considering an effective-mass electron in a homogeneous material subject to a magnetic field. It is convenient to choose the Landau gauge $\vec{\mathcal{A}} = \mathcal{B}x\vec{e}_y$ so that the Schrödinger equation with momentum $\vec{\Pi} = \vec{p} + e\vec{\mathcal{A}}$ becomes

$$\frac{\Pi^2}{2m_e}\psi = \frac{1}{2m_e}\left(\frac{\hbar}{i}\nabla + e\mathcal{B}x\vec{e}_y\right)^2\psi = E\psi\,. \tag{38.6}$$

The magnetic field then only breaks translational symmetry along *x* and we can write (suppressing the *z*-dependence) $\psi(x,y) = L^{-1/2}\varphi(x)e^{iky}$ with the prefractor for normalization. We then find

$$\frac{1}{2m_e}\left\{-\hbar^2\frac{\partial^2}{\partial x^2} + \left(\hbar k + e\mathcal{B}x\right)^2\right\}\varphi(x) = E\varphi(x)\,.$$

This is readily transformed into a standard harmonic oscillator problem via the coordinate shift $\tilde{x} = x + \hbar k/(e\mathcal{B})$ that leads to

$$\left\{-\frac{\hbar^2}{2m_e}\frac{\partial^2}{\partial\tilde{x}^2} + \frac{1}{2}m_e\omega_c^2\tilde{x}^2\right\}\varphi(\tilde{x}) = E\varphi(\tilde{x})\,.$$

Thus, the eigenvalues are $E_\nu = \hbar\omega_c(\nu + \frac{1}{2})$ with ν a non-negative integer (the Landau index). These states are called Landau levels. The wave functions are harmonic oscillator functions in the shifted variable $\tilde{x}$. A large number of *k*-points in the states $\psi_{\nu k}(x,y) = L^{-1/2}\varphi_\nu(x)e^{iky}$ share the same energy E_ν and it can be shown that the degeneracy including spin is $g = e\mathcal{B}A/(\pi\hbar)$ [1].

Next, to compute the Hall conductivity we will use the expression

$$\sigma_{xy} = -\frac{ie^2\hbar}{A}\sum_{n,m} f_{nm}\frac{J^x_{nm}J^y_{mn}}{E_{mn}(E_{mn} - i\hbar/\tau)}\,.$$

Here, J_x is the current operator defined generally by the relation $J_x = \partial H/\partial\Pi_x$. Hence, for the ordinary Schrödinger Hamiltonian, we find $\vec{J} = \vec{\Pi}/m_e$, where, again, $\vec{\Pi} = \vec{p} + e\vec{\mathcal{A}}$. From the relation $\sigma_{yx} = -\sigma_{xy}$ it follows, in general, that

$$\sigma_{xy} = -\frac{2ie^2\hbar}{m_e^2 A}\sum_{n\in v, m\in c}\frac{\Pi_{nm}^x \Pi_{mn}^y}{E_{mn}^2 + (\hbar/\tau)^2}.$$

The action of position and momentum operators are given by

$$\tilde{x}\left|\psi_{mk}\right\rangle = \sqrt{\frac{\hbar}{2m_e\omega_c}}\left\{\sqrt{m+1}\left|\psi_{m+1,k}\right\rangle + \sqrt{m}\left|\psi_{m-1,k}\right\rangle\right\},$$
$$p_x\left|\psi_{mk}\right\rangle = i\sqrt{\frac{\hbar m_e\omega_c}{2}}\left\{\sqrt{m+1}\left|\psi_{m+1,k}\right\rangle - \sqrt{m}\left|\psi_{m-1,k}\right\rangle\right\}.$$

Thus, the matrix elements of $\vec{\Pi}$ are

$$\left\langle\psi_{nk}\left|\Pi_x\right|\psi_{mk}\right\rangle = i\sqrt{\frac{m_e\hbar\omega_c}{2}}\left\{\sqrt{m+1}\delta_{n,m+1} - \sqrt{m}\delta_{n,m-1}\right\},$$
$$\left\langle\psi_{nk}\left|\Pi_y\right|\psi_{mk}\right\rangle = \sqrt{\frac{m_e\hbar\omega_c}{2}}\left\{\sqrt{m+1}\delta_{n,m+1} + \sqrt{m}\delta_{n,m-1}\right\}.$$

The calculation simplifies enormously because coupling is only between states that are neighbors in Landau index. Hence, if the index of the highest occupied state is ν, one finds

$$\sigma_{xy} = -\frac{2ie^2\hbar g}{m_e^2 A}\frac{\Pi_{\nu,\nu+1}^x \Pi_{\nu+1,\nu}^y}{E_{\nu+1,\nu}^2 + (\hbar/\tau)^2}, \tag{38.7}$$

where the degeneracy is taken into account. Using the matrix elements as well as $E_{\nu+1,\nu} = \hbar\omega_c$ then yields

$$\sigma_{xy} = -\frac{eg}{\mathcal{B}A}\cdot\frac{(\nu+1)\omega_c^2}{\omega_c^2 + 1/\tau^2} = -\frac{e^2}{\pi\hbar}\frac{(\omega_c\tau)^2}{1+(\omega_c\tau)^2}(\nu+1).$$

In the limit of large τ, we have $\sigma_{xy} \approx -e^2(\nu+1)/(\pi\hbar)$. Hence, the Hall conductivity is clearly quantized in units of $e^2/(\pi\hbar)$. A connection to the classical result can be made if it is recalled that for a two-dimensional electron gas the density is $n = m_e E_F/(\pi\hbar^2)$. If we now set $E_F = E_\nu$ we find for large ν

$$\sigma_{xy} = -\frac{e^2}{\pi\hbar}\cdot\frac{E_F}{\hbar\omega_c} = -\frac{em_eE_F}{\pi\hbar^2\mathcal{B}} = -\frac{en}{\mathcal{B}}.$$

This agrees perfectly with the classical result.

38.4 Quantum Picture for Massless Dirac Electrons

For Dirac electrons in e.g. graphene, the wave equation for the two-component spinor $\Psi = (\psi, \psi')^T$ in the K - valley reads c.f. Chapter 17

$$v_F \begin{pmatrix} 0 & \Pi_x - i\Pi_y \\ \Pi_x + i\Pi_y & 0 \end{pmatrix} \cdot \begin{pmatrix} \psi \\ \psi' \end{pmatrix} = E \begin{pmatrix} \psi \\ \psi' \end{pmatrix}.$$

Isolating, we find

$$v_F^2 \left\{ \Pi^2 + i[\Pi_x, \Pi_y] \right\} \psi = E^2 \psi .$$

The commutator is $[\Pi_x, \Pi_y] = -i\hbar e\mathcal{B}$ and by comparison with Eq.(38.6) it is apparent that $\psi \propto \psi_{\nu k}$ and $E_\nu^{\pm} = \pm v_F \sqrt{2\hbar e\mathcal{B}|\nu|}$ with $\nu = 0, 1, 2, \ldots$ etc. The plus and minus solutions correspond to "electron" and "hole" states, respectively. The lower component is determined by the original equation $\psi' = (v_F / E)\{p_x + e\mathcal{B}\tilde{x}\}\psi$ as long as $E \neq 0$. Using results above it follows that $\psi_\nu' = \pm\psi_{\nu+1}$ and so the full, normalized state is $\Psi_\nu^{\pm} = (\psi_\nu, \pm\psi_{\nu+1})^T / \sqrt{2}$. For $E = 0$, the solution is $\Psi_0 = (0, \psi_0)^T$.

The current operator associated with the Dirac equation is $\vec{J} = \nabla_\Pi H = v_F \vec{\sigma}$. We now find the matrix elements

$$\left\langle \Psi_\nu^\sigma \middle| J_x \middle| \Psi_{\nu+1}^{\sigma'} \right\rangle = \sigma N_\nu v_F, \quad \left\langle \Psi_{\nu+1}^{\sigma'} \middle| J_y \middle| \Psi_\nu^\sigma \right\rangle = -i\sigma N_\nu v_F .$$

Here, $N_\nu = \frac{1}{2}\sqrt{1 + \delta_{\nu 0}}$. Assuming $E_F > 0$ and low temperature, a computation analogous to the massive case then yields

$$\sigma_{xy} = -\frac{e^2 \hbar g v_F^2}{A} \cdot \left\{ \frac{1}{(E_{\nu+1}^+ - E_\nu^+)^2 + \hbar^2 / \tau^2} + \frac{1}{(E_{\nu+1}^+ - E_\nu^-)^2 + \hbar^2 / \tau^2} \right\}.$$

The two terms in the braces correspond to electron-electron and hole-electron transitions, respectively. In the special case $\nu = 0$, only one term should really be kept. This term should be multiplied by two, however, because $N_0^2 = 2N_{\nu \neq 0}^2$. Hence, the above expression gives the correct result in all cases. In the limit $\tau \to \infty$, we find

$$\sigma_{xy} = -\frac{e^2}{\pi\hbar}(2\nu + 1) .$$

Similarly, for $\tau \to 0$

$$\sigma_{xy} = -\frac{e^3 v_F^2 \tau^2 \mathcal{B}}{\pi \hbar^2}.$$

This agrees perfectly with the semiclassical result. In graphene, $n = E_F^2/(\pi\hbar^2 v_F^2)$. Hence, if we put $E_\nu^+ = E_F$ one finds for the ballistic result $\sigma_{xy} \approx -en/\mathcal{B}$, similarly to the massive-electron case.

As an example, we will compare mono- and bilayer graphene. Monolayer graphene is described by massless Dirac electrons, whereas the carriers in the bilayer follow a normal (massive) dispersion at low energies. Solving for the "filling" ν one finds for the mono- and bilayer $\nu = [\pi\hbar n/(2e\mathcal{B})]$ and $\nu = [\pi\hbar n/(2e\mathcal{B}) - \frac{1}{2}]$, respectively, where the square brackets designate taking the integer part (because ν is integer). Moreover, for the mono- and bilayer, the Hall conductivities are $\sigma_{xy}^{\mathrm{ML}} \approx -2e^2(\nu + \frac{1}{2})/(\pi\hbar)$ and $\sigma_{xy}^{\mathrm{BL}} \approx -2e^2(\nu+1)/(\pi\hbar)$, where an extra factor of two is included in the bilayer case to account for the two sheets, each providing an energy band. Plugging the fillings into these expressions yields the plots in Fig. 38.3. It is clear that a significant difference is observed, with the monolayer displaying half-integer plateaux in units of $\sigma_0 = 2e^2/(\pi\hbar) = 4e^2/h$. In contrast, ordinary parabolic materials such as bilayer graphene display integer-valued plateaux. These predictions correspond very well with the experimental data obtained by Novoselov, Geim and coworkers in Ref. [2] and reproduced in Fig. 38.4. The experiments are at a field of $\mathcal{B} = 14$ T, for which the density unit is $n_0 = e\mathcal{B}/(\pi\hbar) = 6.8 \cdot 10^{11}$ cm^{-2}.

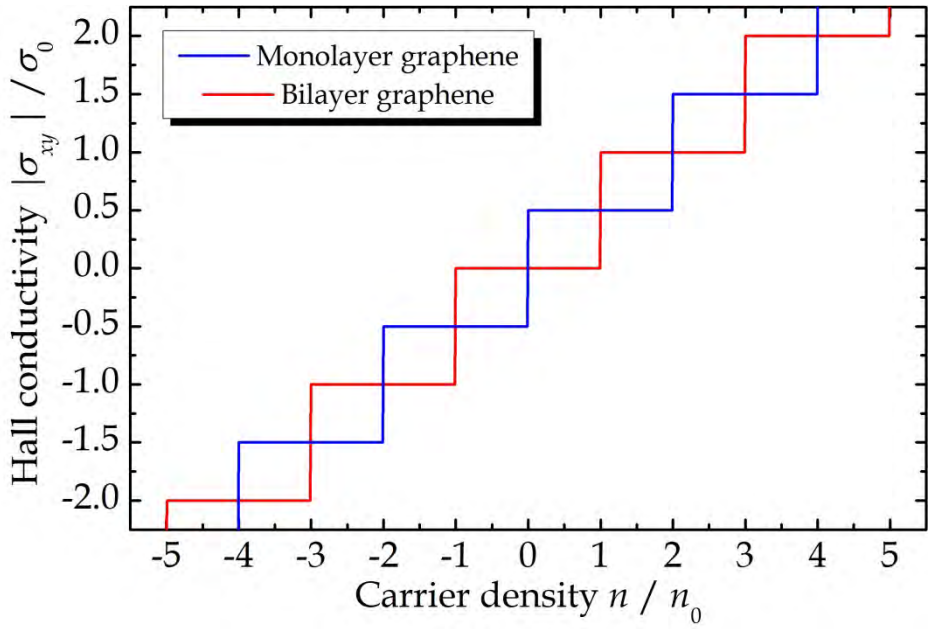


Fig. 38.3. Quantized Hall effect for mono- and bilayer graphene using the units $\sigma_0 = 2e^2/(\pi\hbar)$ and $n_0 = e\mathcal{B}/(\pi\hbar)$.

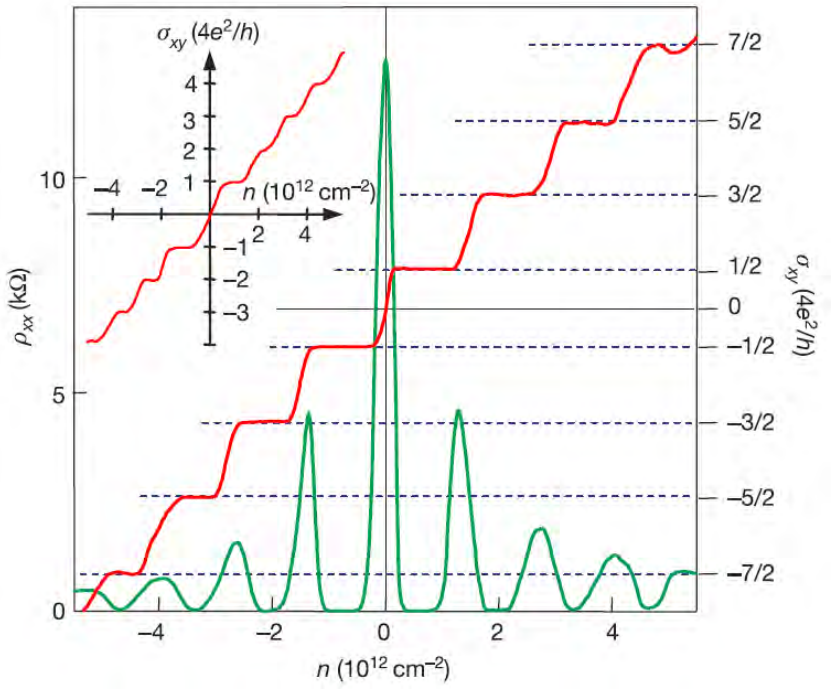


Fig. 38.4. Experimental quantum Hall effects for mono- and bilayer (inset) graphene. Taken from K.S. Novoselov *et al.*, Nature 438, 197 (2005).

Exercise: Shubnikov – de Haas oscillations

In this exercise, we will start by extending the results for massive electrons to finite temperature.

a) Show that the result in Eq.(38.7) becomes

$$\sigma_{xy} = -\frac{\sigma_0}{2}\frac{(\omega_c\tau)^2}{1+(\omega_c\tau)^2}\sum_{\nu=0}^{\infty}\{f(E_\nu)-f(E_{\nu+1})\}(\nu+1),$$

$$\sigma_{xx} = \frac{\sigma_0}{2}\frac{\omega_c\tau}{1+(\omega_c\tau)^2}\sum_{\nu=0}^{\infty}\{f(E_\nu)-f(E_{\nu+1})\}(\nu+1).$$

Here, $\sigma_0 = 2e^2/(\pi\hbar)$ and the lower result follows from $\sigma_{xy} = -\omega_c\tau\sigma_{xx}$. The resistivity tensor is $\overleftrightarrow{\rho} = \overleftrightarrow{\sigma}^{-1}$.

b) Show that $\rho_{xx} = \sigma_{xx}/(\sigma_{xx}^2+\sigma_{xy}^2) = \sigma_{xx}^{-1}/(1+(\omega_c\tau)^2)$.

The quantities here only depend on $\omega_c\tau$, the scaled temperature $t = kT\tau/\hbar$ and the scaled Fermi energy $E_F\tau/\hbar$.

c) Plot the longitudinal resistivity as a function of $\omega_c\tau$ by summing the first 100 Landau level contributions. The result is shown in Fig. 38.5. The oscillations are called Shubnikov – de Haas oscillations and a maximum occurs whenever a Landau level becomes occupied.

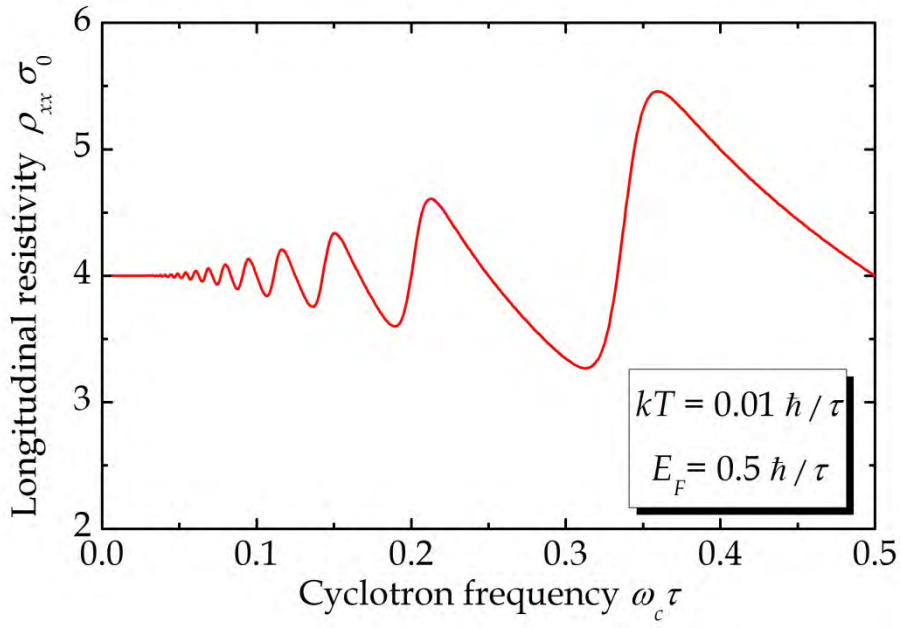


Fig. 38.5. Longitudinal resistivity versus scaled magnetic field for massive electrons.

References

[1] N.W. Ashcroft and N.D. Mermin *Solid State Physics* (Saunders College, Philadelphia, 1976).
[2] K.S. Novoselov *et al.*, Nature 438, 197 (2005).

39. Magneto-Optics

In the previous chapter, we studied the effects of magnetic fields in DC electronic transport. We now attempt to do the same for time-dependent electric field, keeping the magnetic field constant, though. Hence, all the results of Chapter 38 should emerge as the DC limit of a more comprehensive frequency-dependent result. We will start by analyzing the classical and semiclassical models and then turn to quantum models. Finally, the magneto-optical properties of semiconductors will be studied.

39.1 Classical and Semiclassical Pictures

In the time-dependent case, the electric field reads $\vec{\mathcal{E}}(t) = \frac{1}{2}\vec{\mathcal{E}}\, e^{-i\omega t} + c.c.$ Hence, $\vec{\mathcal{E}}$ now designates the amplitude of the field. It follows that Newton's equation of motion for an electron with effective mass m_e now reads

$$m_e \frac{d\vec{v}(t)}{dt} + \frac{m_e}{\tau}\vec{v}(t) = -e\left\{\tfrac{1}{2}\vec{\mathcal{E}}\, e^{-i\omega t} + c.c.\right\} - e\vec{v}(t)\times\vec{\mathcal{B}}\,.$$

The modified solution is easily found, however, if we write $\vec{v}(t) = \frac{1}{2}\vec{v}\, e^{-i\omega t} + c.c.$ so that

$$\frac{m_e}{\tau}(1-i\omega\tau)\begin{pmatrix} v_x \\ v_y \end{pmatrix} = -e\mathcal{E}\begin{pmatrix} 1 \\ 0 \end{pmatrix} - e\mathcal{B}\begin{pmatrix} v_y \\ -v_x \end{pmatrix}.$$

It is immediately clear that the frequency-dependent conductivity is obtained from the DC case by the substitution $\tau \to \tau/(1-i\omega\tau)$. This means that

$$\sigma_{xx}(\omega) = \frac{e^2 n\tau}{m_e}\frac{1-i\omega\tau}{(1-i\omega\tau)^2 + (\omega_c\tau)^2},\ \sigma_{yx}(\omega) = -\sigma_{xy}(\omega) = \frac{e^2 n\omega_c\tau^2}{m_e}\frac{1}{(1-i\omega\tau)^2 + (\omega_c\tau)^2}. \quad (39.1)$$

It is obvious that these results have the correct DC limit. In the semiclassical picture, modification is only slightly more complicated. We need to retain the time-derivative in Eq.(12.2) such that the Boltzmann equation becomes

$$g(\vec{k},t) = f(E_{\vec{k}}) + \frac{e\tau}{\hbar}\nabla_{\vec{k}} g(\vec{k},t)\cdot(\vec{\mathcal{E}} + \vec{v}\times\vec{\mathcal{B}}) - \tau\frac{\partial g(\vec{k},t)}{\partial t}.$$

The distribution function now has DC as well as frequency-dependent components. The 0'th order term is still $g_0(\vec{k}) = f(E_{\vec{k}})$. We write the amplitude of the first harmonic term as $g(\vec{k},\omega)$. For harmonic time variation, the time-derivative simply means

multiplication by a factor $-i\omega$. Hence, the first order term is now $g_1(\vec{k},\omega) = e\tau/(1-i\omega\tau) f'(E_{\vec{k}})\vec{v}\cdot\vec{\mathcal{E}}$. Eventually, at second order, we find

$$g_2(\vec{k},\omega) = \frac{e\tau}{1-i\omega\tau} f'(E_{\vec{k}})\vec{v}\cdot\vec{\mathcal{E}} + \frac{e^2\tau^2}{\hbar(1-i\omega\tau)^2}\nabla_{\vec{k}}\left(f'(E_{\vec{k}})\vec{v}\cdot\vec{\mathcal{E}}\right)\cdot(\vec{v}\times\vec{\mathcal{B}}). \quad (39.2)$$

Finally, the optical Hall conductivity therefore becomes

$$\sigma_{xy}(\omega) = \frac{e^3\tau^2\mathcal{B}}{4\pi^3\hbar(1-i\omega\tau)^2}\int f'(E_{\vec{k}})v_y\left\{\frac{\partial v_x}{\partial k_x}v_y - \frac{\partial v_x}{\partial k_y}v_x\right\}d^3k \approx -\frac{e^3\tau^2\mathcal{B}}{m_e^2(1-i\omega\tau)^2}n,$$

where the second, approximate equality is valid for effective-mass dispersion.

39.2 Quantum Picture

With a frequency-dependent perturbation the quantum expression for the optical Hall conductivity is

$$\sigma_{xy}(\omega) = -\frac{ie^2\hbar}{A}\sum_{n,m} f_{nm}\frac{J^x_{nm}J^y_{mn}}{E_{mn}(E_{mn}-\hbar\omega-i\hbar/\tau)}.$$

Hence, the required modification is precisely the substitution $\tau \to \tau/(1-i\omega\tau)$, exactly as we found in the classical and semiclassical expressions above. For massive and massless electrons in gapless systems we therefore find

$$\text{Massive: } \sigma_{xy}(\omega) = -\frac{e^2}{\pi\hbar}\frac{(\omega_c\tau)^2}{(1-i\omega\tau)^2+(\omega_c\tau)^2}(\nu+1).$$

$$\text{Massless: } \sigma_{xy}(\omega) = -\frac{e^2}{\pi\hbar}\frac{2\nu+1+\frac{(1-i\omega\tau)^2}{\omega_D^2\tau^2}}{\left(1+\frac{(1-i\omega\tau)^2}{\omega_D^2\tau^2}\right)^2+4\nu\frac{(1-i\omega\tau)^2}{\omega_D^2\tau^2}}, \quad \omega_D = v_F\sqrt{\frac{2e\mathcal{B}}{\hbar}}.$$

In a time-dependent field, it becomes relevant to study the magneto-optical response of gapped systems such as intrinsic semiconductors. As an example, we will study gapped graphene. Here, in fact, the gap appears in two locations (denoted valleys) that should be summed over. Denoting the "valley index" by $\eta = \pm 1$, the system is described by the massive Dirac Hamiltonian

$$\begin{pmatrix} \alpha & v_F(\eta\hat{\Pi}_x - i\hat{\Pi}_y) \\ v_F(\eta\hat{\Pi}_x + i\hat{\Pi}_y) & -\alpha \end{pmatrix}\cdot\begin{pmatrix}\psi \\ \psi'\end{pmatrix} = E\begin{pmatrix}\psi \\ \psi'\end{pmatrix}.$$

In this model, the band gap is 2α. Solving for the eigenstates, it can be shown that the new energies are $E_\nu^\pm = \pm(1-\delta_{\nu 0})\sqrt{\hbar^2\omega_D^2|\nu|+\alpha^2} - \delta_{\nu 0}\eta\alpha$ and that the massless matrix elements are modified into

$$\left\langle \Psi_\nu^\sigma \middle| \hat{J}_x \middle| \Psi_{\nu+1}^{\sigma'} \right\rangle = \sigma N_\nu v_F \left(1+\frac{\alpha}{E_\nu}\right), \quad \left\langle \Psi_{\nu+1}^{\sigma'} \middle| \hat{J}_y \middle| \Psi_\nu^\sigma \right\rangle = -i\sigma N_\nu v_F \left(1-\frac{\alpha}{E_\nu}\right).$$

In turn, after summing over valleys [1]

$$\sigma_{xy}(\omega) = -\sigma_0 \frac{2\hbar^2\omega_D^2}{\pi} \sum_{\nu=0}^{\infty} \left\{ f(E_{\nu+1}^-) - f(E_\nu^-) + f(E_{\nu+1}^+) - f(E_\nu^+) \right\} \times$$
$$\left\{ \frac{1}{(E_{\nu+1}^+ - E_\nu^+)^2 - \hbar^2(\omega + i/\tau)^2} \left(1 - \frac{\alpha^2}{E_\nu^+ E_{\nu+1}^+}\right) + \frac{1}{(E_{\nu+1}^+ - E_\nu^-)^2 - \hbar^2(\omega + i/\tau)^2} \left(1 - \frac{\alpha^2}{E_\nu^- E_{\nu+1}^+}\right) \right\}.$$

with $\sigma_0 = e^2/4\hbar$ the graphene DC conductivity and the understanding that $E_0^\pm = \alpha$. This expression actually vanishes identically if $E_F = 0$. This is a consequence of electron-hole symmetry, which means that electron and hole contributions cancel exactly in the symmetric case $E_F = 0$. For finite E_F, however, a finite response is obtained. At $T = 0$, the requirement is $|E_F| \geq \alpha$. Below, we plot the response for $2\alpha = 1.8$ eV and E_F at the band edge taking $\hbar/\tau = 20$ meV.

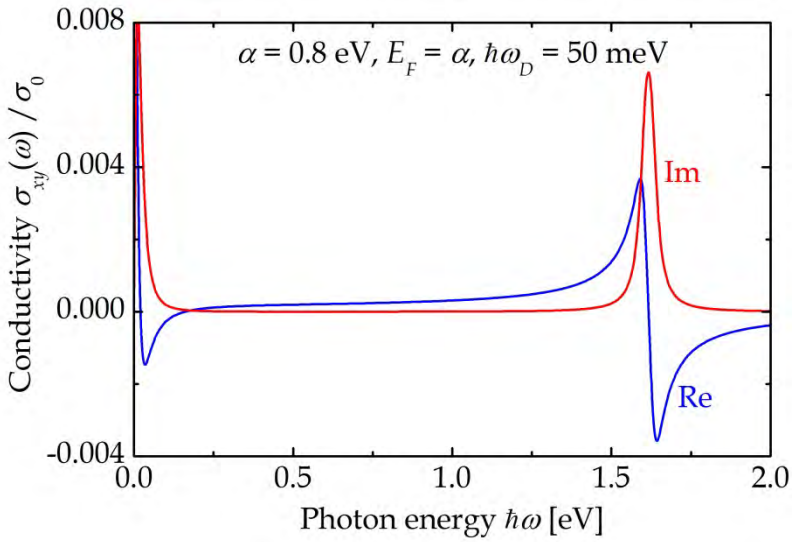


Fig. 39.1. Hall conductivity in the electron-hole symmetric gapped graphene model.

39.3 Magneto-Excitons

We now aim to describe the magneto-optical response of semiconductors in the presence of excitonic effects. We will work within the Wannier picture and consider a simple band structure composed of a valence band with effective mass $-m_h$ and a

conduction band with effective mass m_e. These bands are separated by a band gap E_g and the Fermi level lies in the gap. This means that the Wannier Hamiltonian relative to the gap is

$$\hat{H} = \frac{1}{2m_e}\left(\hat{\vec{p}}_e + e\vec{\mathcal{A}}_e\right)^2 + \frac{1}{2m_h}\left(\hat{\vec{p}}_h - e\vec{\mathcal{A}}_h\right)^2 + v(\vec{r}_e - \vec{r}_h).$$

Here, v is the electron-hole Coulomb attraction. We will use the symmetric gauge such that $\vec{\mathcal{A}}_i = \frac{1}{2}\vec{\mathcal{B}}\times\vec{r}_i$ with $i = e,h$. We also introduce the usual relative and centre of mass coordinates $\vec{r} = \vec{r}_e - \vec{r}_h$ and $\vec{R} = (m_e\vec{r}_e + m_h\vec{r}_h)/M$ with $M = m_e + m_h$. It can then be shown that under the unitary transformation $U = \exp\{-\frac{ie}{2\hbar}\vec{\mathcal{B}}\cdot(\vec{R}\times\vec{r})\}$, the Hamiltonian becomes $\hat{H} \to U^{-1}\hat{H}U \equiv \hat{H}_r + \hat{H}_R$ with $\hat{H}_R = \hat{P}^2/2M$ and relative motion part

$$\hat{H}_r = \frac{\hat{p}^2}{2m_{eh}} + \frac{e\hbar\gamma}{2m_{eh}}\vec{\mathcal{B}}\cdot\hat{\vec{l}} + \frac{e^2}{8m_{eh}}(\vec{\mathcal{B}}\times\vec{r})^2 + v(\vec{r}).$$

Here, $\gamma = (m_e - m_h)/(m_e + m_h)$ is a measure of mass imbalance and we assume $\gamma \approx 1$. The two field-induced modifications are the para- and diamagnetic contributions, respectively. We take the magnetic field along the *z*-directions and, from now on, specialize to two dimensions so that $\vec{r} = x\vec{e}_x + y\vec{e}_y$. Hence,

$$H_r = \frac{\hat{p}^2}{2m_{eh}} + \frac{e\hbar}{2m_{eh}}\mathcal{B}\hat{l}_z + \frac{e^2\mathcal{B}^2r^2}{8m_{eh}} + v(\vec{r}).$$

This Hamiltonian unfortunately doesn't have analytical eigenfunctions. We will therefore attempt to expand in a basis of Landau levels, i.e. solution to the problem taking $v = 0$. These Landau states are of the form $e^{il\theta}\varphi_{nl}(r)/\sqrt{2\pi}$ with l integer and

$$\varphi_{nl}(r) = \frac{1}{L^{|l|+1}}\sqrt{\frac{n!}{2^{|l|}(n+|l|)!}}r^{|l|}e^{-r^2/4L^2}L_n^{|l|}(r^2/2L^2).$$

Here, $L = \sqrt{\hbar/e\mathcal{B}}$ is the magnetic length and n is a non-negative integer. The associated energy is

$$E_{nl} = \hbar\omega_c\left\{n + \tfrac{1}{2}(1 + l + |l|)\right\}$$

with $\omega_c = e\mathcal{B}/m_{eh} = e\mathcal{B}/m_e + e\mathcal{B}/m_h$ the sum of electron and hole cyclotron frequencies. In this basis, we need matrix elements of the Coulomb interaction. Previously, we have only considered optically active *s*-type excitons with $l = 0$. In

the general case, however, we need $V^l_{nm} = \langle \varphi_{nl} | L/r | \varphi_{ml} \rangle$ that turns out to involve a hypergeometric function [2]

$$V^l_{nm} = \frac{\Gamma(l+\frac{1}{2})\Gamma(m+\frac{1}{2})}{l!}\sqrt{\frac{(n+l)!}{2\pi m!n!(m+l)!}}\,{}_3F_2[-n, l+\tfrac{1}{2}, \tfrac{1}{2}; l+1, \tfrac{1}{2}-m; 1].$$

After a few manipulations, the matrix elements in this basis can then be shown to be of the form $H_{nm} = \hbar\omega_c\{n+\frac{1}{2}(1+l+|l|)\}\delta_{nm} - \sqrt{2Ry^*\hbar\omega_c}V^l_{nm}$. Finally, writing the radial eigenstates in the form $\psi_{exc}(r) = \sum_n c_{nl}\varphi_{nl}(r)$, the exciton oscillator strength is determined by $\psi_{exc}(0) = L^{-1}\sum_n c_{n0}$. In Fig. 39.2 below, we have plotted the three lowest energy eigenvalues as a function of magnetic field for $Ry^* = 0.1$ eV. It is seen that at zero field they follow the expected $-Ry^*\{4, \frac{4}{9}, ...\}$ 2D Rydberg series, c.f. Chapter 19. In a small field, first order perturbation theory shows that the energy correction is $\Delta E_l = \frac{l}{2}\hbar\omega_c$.

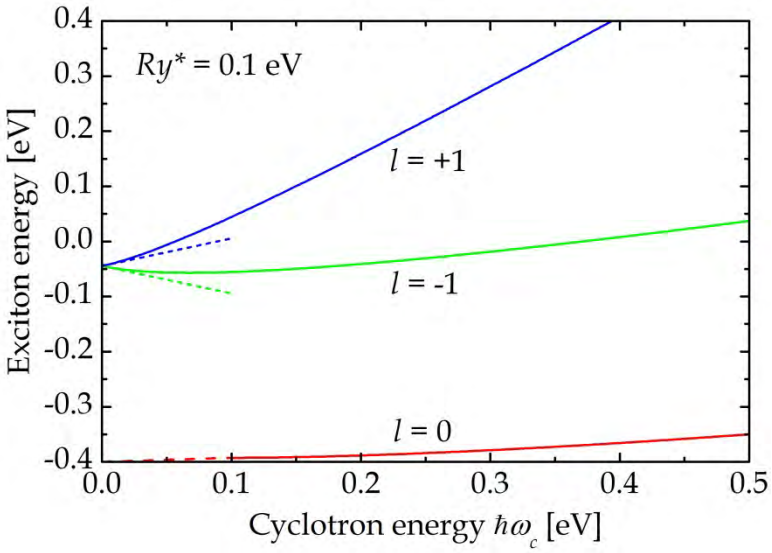


Fig. 39.2. Magneto-exciton energies versus magnetic field in energy units. Dashed lines indicate perturbation results.

As usual, the diagonal excitonic optical response is evaluated using

$$\sigma_{xx}(\omega) = \frac{2e^2\hbar^2\omega}{im^2A}\sum_{exc}\frac{|P^x_{exc}|^2}{E_{exc}[E^2_{exc} - \hbar^2(\omega+i\Gamma)^2]}.$$

In the Wannier approximation, this becomes

$$\sigma_{xx}(\omega) = \frac{4e^2\hbar^2\omega|p^x_{vc}|^2}{im^2\hbar\omega_c L^2E_g^2}X_2(\omega), \quad X_2(\omega) = \hbar\omega_c L^2E_g^2\sum_{exc}\frac{|\psi_{exc}(0)|^2}{E_{exc}[E^2_{exc} - \hbar^2(\omega+i\Gamma)^2]}. \quad (39.3)$$

Here, $X_2(\omega)$ is a conveniently normalized susceptibility function and the factor $\hbar\omega_c L^2$ simplifies to $\hbar\omega_c L^2 = \hbar^2 / m_{eh}$. In the plots below, we consider the interplay between magnetic and Coulomb effects in the optical response using a basis of 600 Landau states. Similarly to Fig. 19.2, we take a gap of 1.6 eV and add broadening of $\hbar\Gamma = 20$ meV.

In Fig. 39.3, it is seen that a small field, meaning $\hbar\Gamma \sim \hbar\omega_c$, leads to spectra that are very similar to the field-free results in Fig. 19.2. On the other hand, with $\hbar\omega_c = 50$ meV the Landau level clearly show, and a significant modulation is observed in the spectra.

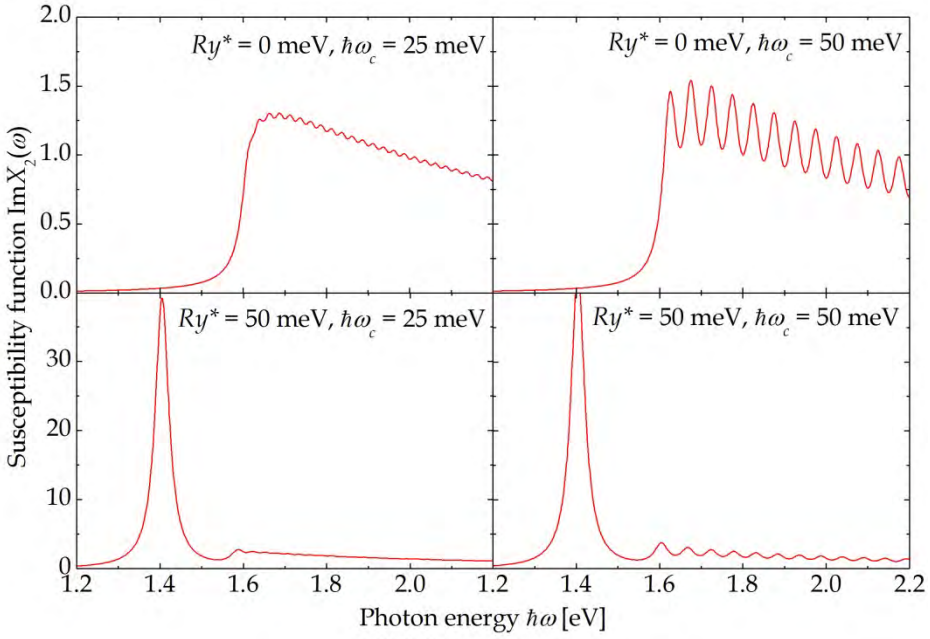


Fig. 39.3. Magneto-excitonic optical response illustrating the competing effects.

We would now like to consider a specific model in order to extend the response to the off-diagonal conductivity. To this end, we examine again the gapped graphene model, for which the momenta in the $\eta = \pm 1$ valleys are (c.f. Chapter 17)

$$p_{vc}^x = mv_F\left(\frac{\alpha}{E}\cos\varphi - i\eta\sin\varphi\right), \quad p_{vc}^y = mv_F\left(\frac{\alpha}{E}\sin\varphi + i\eta\cos\varphi\right) \qquad (39.4)$$

with energy $E = (\alpha^2 + \hbar^2 v_F^2 k^2)^{1/2}$ and φ the polar angle of the wave vector. Clearly, at the position of the gap $k = 0$, one has $|p_{vc}^x|^2 = |p_{vc}^y|^2 = m^2 v_F^2$ in agreement with a non-vanishing isotropic response. Consider now, however, the off-diagonal response, for which we would need factors like $p_{vc}^x p_{vc}^{y*} = -i\eta m^2 v_F^2$. When summed over valleys, this will vanish. Hence, $\sigma_{xy}(\omega) = 0$, which must clearly be so in the case of no magnetic field. A non-vanishing off-diagonal response should reflect the presence of the field.

To be more accurate, we should go back to the original Bethe-Salpeter formulation, so that the correct exciton momentum becomes $P^x_{exc} = \int p^x_{vc}(\vec{k})\psi_{\vec{k}} d^2k$. It is readily shown that with cylindrical symmetry $\psi_{\vec{k}} = e^{il\varphi}\psi_l(k)$ and approximating $E \approx \alpha$ in Eq.(39.4) leads to

$$P^x_{exc} = m v_F \delta_{l\eta} i^l \int_0^\infty \psi_l(k) k dk, \quad P^y_{exc} = i\eta m v_F \delta_{l\eta} i^l \int_0^\infty \psi_l(k) k dk, \quad \psi_l(k) = \int_0^\infty \psi_{exc}(r) J_l(kr) r dr .$$

Hence, the valleys "pick out" different l quantum numbers. It follows that whenever the energies and/or momenta of the $l = \pm 1$ states differ, the two contributions will not cancel. This is the origin of the optical Hall effect in the excitonic theory. In the relevant cases,

$$\int_0^\infty \psi_l(k) k dk = \begin{cases} \psi_{exc}(0), & l = 0 \\ \int_0^\infty \psi_{exc}(r) r^{-1} dr, & l = 1. \end{cases}$$

In the Landau basis, we find

$$Q_{n,\pm 1} \equiv \int_0^\infty \varphi_{n,\pm 1}(r) L / r dr = \sqrt{\frac{\pi(n+1)}{2}} \times \begin{cases} (n-1)!!/\, n!!, & n \text{ even} \\ n!!/\,(n+1)!!, & n \text{ odd.} \end{cases}$$

It follows that $P^x_{exc,\pm 1} P^{y*}_{exc,\pm 1} = \mp i m^2 v_F^2 \left[\sum_n c_{n,\pm 1} Q_{n,\pm 1}\right]^2 / L^2$. The fact that $P^x_{exc,\pm 1} P^{y*}_{exc,\pm 1}$ is purely imaginary guarantees that $\sigma_{yx} = -\sigma_{xy}$. Again, we introduce a normalized susceptibility function

$$\begin{aligned} \sigma_{xy}(\omega) &= \frac{2e^2\hbar}{im^2 A} \sum_{exc,l} \frac{P^x_{exc,l} P^{y*}_{exc,l}}{E^2_{exc,l} - \hbar^2(\omega + i\Gamma)^2} = \frac{2e^2 v_F^2}{\pi \omega_c L^2 E_g} Y_2(\omega), \\ Y_2(\omega) &= -\hbar\omega_c E_g \sum_{\substack{exc \\ l=\pm 1}} \frac{l\left[\sum_n c^{exc}_{n,l} Q_{n,l}\right]^2}{E^2_{exc,l} - \hbar^2(\omega + i\Gamma)^2}. \end{aligned} \tag{39.5}$$

In Fig. 39.4, this off-diagonal susceptibility function is illustrated, demonstrating how the response grows with magnetic field (note that the prefactor $\omega_c L^2$ does not depend on $\mathcal{B}$).

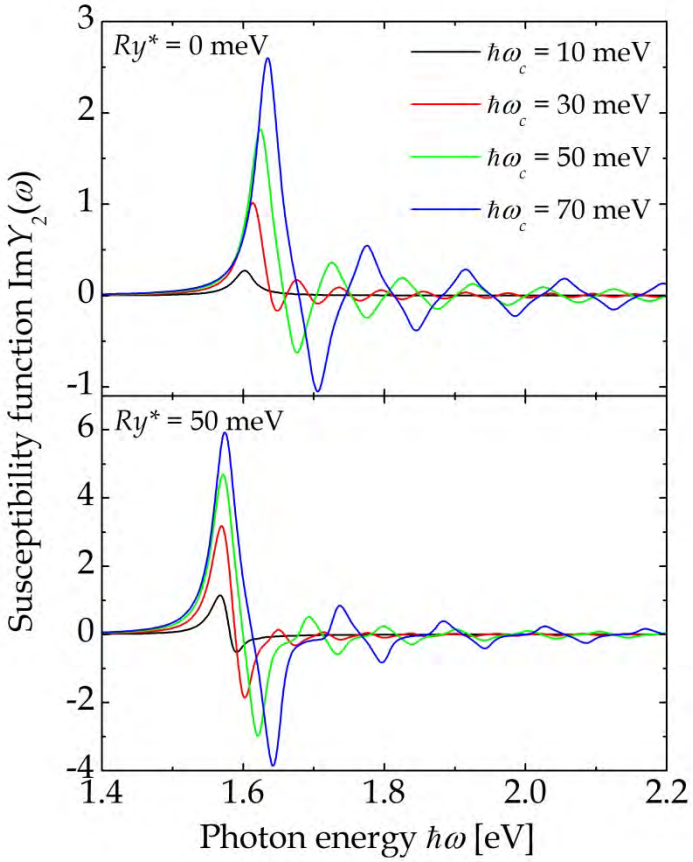


Fig. 39.4. Off-diagonal magneto-excitonic response.

Exercise: Faraday rotation

In this exercise, we examine the consequences of an off-diagonal component in the conductivity. In the (*x*,*y*) plane, the optical response is given by the dielectric tensor

$$\overleftrightarrow{\varepsilon} = \begin{pmatrix} \varepsilon_{xx} & \varepsilon_{xy} \\ -\varepsilon_{xy} & \varepsilon_{xx} \end{pmatrix}$$

with $\varepsilon_{xx} = 1 + i\sigma_{xx} / \varepsilon_0 \omega$ and $\varepsilon_{xy} = i\sigma_{xy} / \varepsilon_0 \omega$.

a) Show that the wave equation $(\nabla^2 + \overleftrightarrow{\varepsilon} k_0^2) \cdot \vec{\mathcal{E}}(z)$ has circularly polarized solutions

$$\vec{\mathcal{E}}_{\pm}(z) = \mathcal{E}_0 \begin{pmatrix} 1 \\ \pm i \end{pmatrix} e^{iq_{\pm} z}, \quad q_{\pm}^2 = k_0^2 (\varepsilon_{xx} \pm i\varepsilon_{xy}).$$

Now, suppose a linearly *x*-polarized field of amplitude $\mathcal{E}_0$ enters the material at $z = 0$.

b) Using the fact that $\vec{e}_x = \frac{1}{2}\left\{(\vec{e}_x + i\vec{e}_y) + (\vec{e}_x - i\vec{e}_y)\right\}$ show that the transmitted field after propagating a distance d is

$$\vec{\mathcal{E}}(d) = \tfrac{1}{2}\mathcal{E}_0 \begin{pmatrix} e^{iq_+ d} + e^{iq_- d} \\ i(e^{iq_+ d} - e^{iq_- d}) \end{pmatrix}.$$

If we write $q_\pm = q_0 \pm \delta q$, it follows that $\mathcal{E}_y(d)/\mathcal{E}_x(d) = \tan(\delta q d)$. Such a field is elliptically polarized with a polarization rotated by the (complex) angle $\theta = \delta q d$.

c) Show that to first order in $\mathcal{B}$, the rotation is $\theta = -\sigma_{xy} d/(2n\varepsilon_0 c)$, where n is the unperturbed refractive index.

References

[1] J. G. Pedersen and T. G. Pedersen, Phys. Rev. B84, 115424 (2011).
[2] A. H. MacDonald and D. S. Ritchie, Phys. Rev. B33, 8336 (1986).

40. Topological Hall Effect

We have seen that a static magnetic field leads to off-diagonal elements in the DC and optical conductivity. Recently, it was discovered that, actually, such effects can arise without any magnetic field, leading to the so-called anomalous Hall effect. This discovery is intimately related to the mathematical subject of topology and very elegantly explains the quantization of the Hall conductivity that we found in e.g. graphene, c.f. Chapter 38. Some of the central concepts in these developments are the Berry connection and Berry curvature and we start with an introduction to this subject.

40.1 Berry Connection and Curvature

First, a word on naming: the Berry connection $\vec{\Omega}$ is sometimes called the Berry (vector) potential and the Berry curvature $\vec{\mathcal{C}}$ is also known as the Berry field. In addition, different symbols are used for these quantities in the literature. We consider a periodic system with Bloch states of the form $\psi_{n\vec{k}} = u_{n\vec{k}} e^{i\vec{k}\cdot\vec{r}}$ with $u_{n\vec{k}}$ normalized within a unit cell. In the most general case, the relevant definitions are $\vec{\Omega}_{nm} = i\left\langle u_{n\vec{k}} \middle| \nabla_{\vec{k}} \middle| u_{m\vec{k}} \right\rangle$ and $\vec{\mathcal{C}}_{nm} = \nabla_{\vec{k}} \times \vec{\Omega}_{nm}$. The interband Berry connection is just the transition dipole moment, as can be demonstrated by differentiating $0 = \left\langle \psi_{n\vec{k}'} \middle| \psi_{m\vec{k}} \right\rangle$ for $n \neq m$:

$$0 = i\nabla_{\vec{k}} \left\langle u_{n\vec{k}'} e^{i\vec{k}'\cdot\vec{r}} \middle| u_{m\vec{k}} e^{i\vec{k}\cdot\vec{r}} \right\rangle = -\left\langle u_{n\vec{k}'} \middle| e^{i(\vec{k}-\vec{k}')\cdot\vec{r}} \vec{r} \middle| u_{m\vec{k}} \right\rangle + i\left\langle u_{n\vec{k}'} \middle| e^{i(\vec{k}-\vec{k}')\cdot\vec{r}} \nabla_{\vec{k}} \middle| u_{m\vec{k}} \right\rangle.$$

Hence, taking the limit $\vec{k}' \to \vec{k}$ we obtain

$$\vec{\Omega}_{nm} = \left\langle u_{n\vec{k}} \middle| \vec{r} \middle| u_{m\vec{k}} \right\rangle, \quad \text{for } n \neq m. \tag{40.1}$$

For the present purposes, we will restrict ourselves to two-dimensional materials, for which $\vec{\Omega}$ is in the (*x*,*y*) plane and $\vec{\mathcal{C}}_{nm} = \vec{e}_z \mathcal{C}_{nm}$ with

$$\mathcal{C}_{nm} = i\left\langle \frac{\partial u_{n\vec{k}}}{\partial k_x} \middle| \frac{\partial u_{m\vec{k}}}{\partial k_y} \right\rangle - i\left\langle \frac{\partial u_{n\vec{k}}}{\partial k_y} \middle| \frac{\partial u_{m\vec{k}}}{\partial k_x} \right\rangle.$$

Apart from the definition, another important relation exists between Berry connection and curvature. Using completeness

$$\begin{aligned}\mathcal{C}_{nn} &= i\sum_m \left\langle \frac{\partial u_{n\vec{k}}}{\partial k_x}\middle| u_{m\vec{k}}\right\rangle\left\langle u_{m\vec{k}}\middle|\frac{\partial u_{n\vec{k}}}{\partial k_y}\right\rangle - i\sum_m \left\langle \frac{\partial u_{n\vec{k}}}{\partial k_y}\middle| u_{m\vec{k}}\right\rangle\left\langle u_{m\vec{k}}\middle|\frac{\partial u_{n\vec{k}}}{\partial k_x}\right\rangle \\ &= i\sum_m \Omega^{x*}_{mn}\Omega^{y}_{mn} - i\sum_m \Omega^{y*}_{mn}\Omega^{x}_{mn} \\ &= i\sum_m \Omega^{x}_{nm}\Omega^{y}_{mn} - i\sum_m \Omega^{y}_{nm}\Omega^{x}_{mn} \\ &= 2i\sum_m \Omega^{x}_{nm}\Omega^{y}_{mn}.\end{aligned} \tag{40.2}$$

Finally, the integral of the diagonal components is essentially the *Chern number* $c_n \equiv \int \mathcal{C}_{nn} d^2k / 2\pi$. As we will see, c_n is a topological integer. The quantity $2\pi c_n$ is called the *Berry phase*.

To illustrate all these concepts, we will study a concrete case: The Haldane model [1]. This celebrated model is actually gapped graphene with a twist: next-nearest neighbor interactions of the form $\lambda e^{i\phi\nu_{ab}} e^{i\vec{k}\cdot\vec{r}_{ab}}$ are added. Here, λ, ϕ are fixed parameters and $\vec{r}_{ab}$ is a vector joining the sites. Finally, $\nu_{ab} = \pm 1$ depending on the rotation sense (clockwise or counter-clockwise) within the graphene hexagon. With little loss of generality, we will take $\phi = \pi / 2$, which leads to the Hamiltonian

$$\overleftrightarrow{H} = \begin{pmatrix} \alpha + \lambda s & -\gamma f \\ -\gamma f^* & -\alpha - \lambda s \end{pmatrix}$$

with

$$\begin{aligned} f(\vec{k}) &= e^{ik_x a/\sqrt{3}} + 2e^{-ik_x a/2\sqrt{3}} \cos(k_y a / 2), \\ s(\vec{k}) &= 2\left\{\sin(k_x a\sqrt{3}/2 + k_y a/2) - \sin(k_y a) - \sin(k_x a\sqrt{3}/2 - k_y a/2)\right\}. \end{aligned}$$

Hence, the energies are $E_{v,c} = \mp\varepsilon$ where $\varepsilon = [\gamma^2 |f|^2 + \bar{\alpha}^2]^{1/2}$ with $\bar{\alpha} = \alpha + \lambda s$. For convenience, we will assume $\alpha, \lambda \geq 0$. The band gaps are at the *K* and *K'* points, at which $s = \mp 3\sqrt{3}$, respectively. This means that the gap closes if $\lambda = \lambda_C$ with the critical interaction strength $\lambda_C = \alpha / 3\sqrt{3}$. Writing $f = |f| e^{i\theta}$, the "standard" eigenvectors are

$$\vec{v}_c = \frac{1}{\sqrt{2}} \begin{pmatrix} -e^{i\theta}\sqrt{(\varepsilon + \bar{\alpha})/\varepsilon} \\ \sqrt{(\varepsilon - \bar{\alpha})/\varepsilon} \end{pmatrix}, \quad \vec{v}_v = \frac{1}{\sqrt{2}} \begin{pmatrix} e^{i\theta}\sqrt{(\varepsilon - \bar{\alpha})/\varepsilon} \\ \sqrt{(\varepsilon + \bar{\alpha})/\varepsilon} \end{pmatrix}.$$

In this way, the valence band Berry connection turns out to be

$$\vec{\Omega}_{vv} = -\frac{\gamma^2 a}{2\sqrt{3}\varepsilon(\varepsilon+\bar{\alpha})}\begin{pmatrix}\cos(\frac{\sqrt{3}}{2}k_x a)\cos(\frac{1}{2}k_y a)-\cos(k_y a)\\ \sin(\frac{\sqrt{3}}{2}k_x a)\sin(\frac{1}{2}k_y a)\end{pmatrix}.$$

Importantly, however, this result is certainly not unique. Any other phase choice would be allowed for the eigenvectors and if this phase is k-dependent, the Berry connection will change. As an example, we could have chosen $\vec{v}'_v = e^{-i\theta}\vec{v}_v$. Such a choice obviously implies $\vec{\Omega}'_{vv} = \vec{\Omega}_{vv} + \nabla_{\vec{k}}\theta$. In much of the literature, terms are borrowed from electromagnetism and the phase choice is called a "gauge". Hence, $\vec{\Omega}_{vv}$ is gauge-dependent whereas $\mathcal{C}_{vv}$ is always gauge-invariant because $\nabla_{\vec{k}}\times(\nabla_{\vec{k}}\theta)=0$. In Fig. 40.1, we plot the magnitudes of these Berry connections for a quasi-random set of parameters, but for which $\lambda = 2\lambda_C$. The crucial observation is that $\vec{\Omega}_{vv}$ diverges at the K-points, whereas $\vec{\Omega}'_{vv}$ diverges at the K'-points.

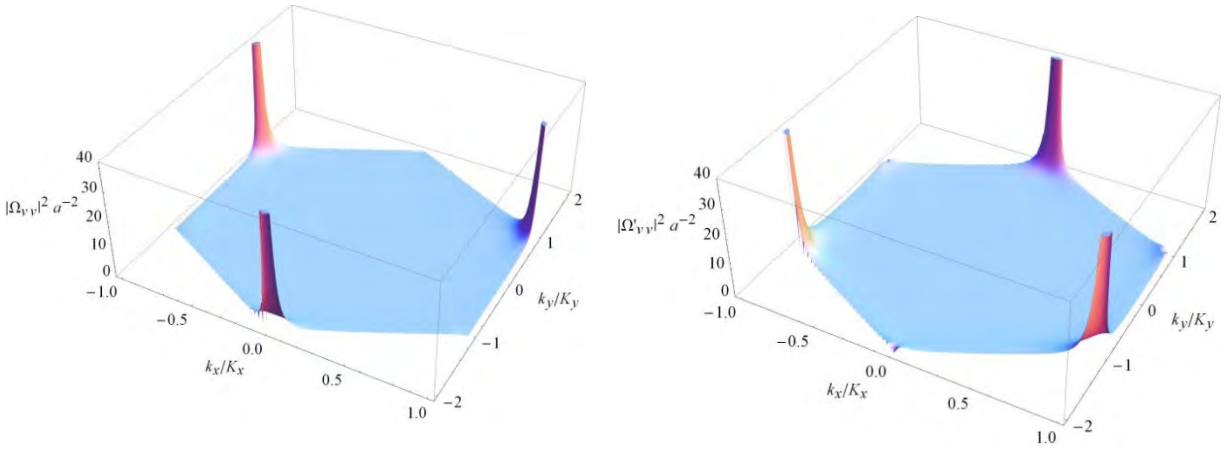


Figure 40.1. Berry connection using "standard" (left) and phase-shifted (right) eigenstates.

The diverging factor in $\vec{\Omega}_{vv}$ is $(\varepsilon+\bar{\alpha})^{-1}$ since, at the K-point, $\varepsilon = |\bar{\alpha}| = -\bar{\alpha}$. By phase-shifting into $\vec{\Omega}'_{vv} = \vec{\Omega}_{vv} + \nabla_{\vec{k}}\theta$, this divergence is cancelled by the numerator, leading to a finite value. This, however, happens at the cost of a divergence at K' in the $\nabla_{\vec{k}}\theta$ term, as can readily be demonstrated. In fact, there is no single choice of phase that will keep the Berry connection finite as long as $\lambda \geq \lambda_C$ since this means that points exist, for which $\bar{\alpha} \leq 0$. On the other hand, if $\lambda < \lambda_C$, $\vec{\Omega}_{vv}$ can be made regular everywhere with a global phase convention. This quantitative difference has extremely important consequences, as we shall see. Irrespective of the phase choice, the valence band Berry curvature is

$$\mathcal{C}_{vv} = \frac{\gamma^2 a^2}{4\sqrt{3}\varepsilon^3}\Big\{\alpha\Big[\sin(k_y a) - 2\cos(\tfrac{\sqrt{3}}{2}k_x a)\sin(\tfrac{1}{2}k_y a)\Big] + \lambda\Big[-8\cos(\tfrac{\sqrt{3}}{2}k_x a)\cos(\tfrac{1}{2}k_y a) + 2(\cos(\sqrt{3}k_x a)-2)\cos(k_y a) + \cos(2k_y a) + 9\Big]\Big\}.$$

In Fig. 40.2, we plot $\mathcal{C}_{vv}$ for the characteristic cases $\lambda = 0$ and $\lambda = 2\lambda_C$ and see something characteristic: If $\lambda = 0$, $\mathcal{C}_{vv}$ has equal positive and negative regions but for $\lambda = 2\lambda_C$ it is everywhere positive.

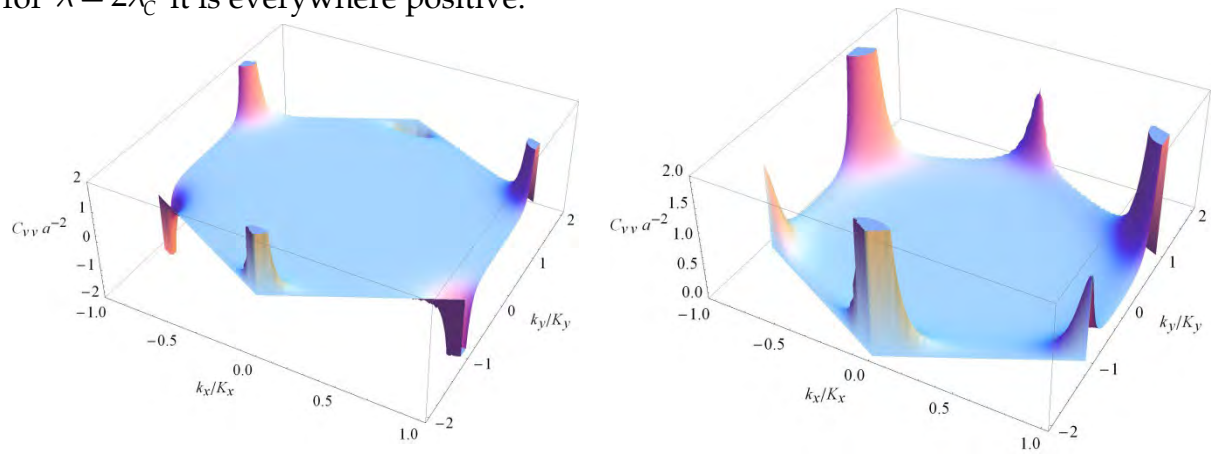


Figure 40.2. Berry curvature with $\lambda = 0$ (left) and $\lambda = 2\lambda_C$ (right).

This leads us directly to the valence band Chern number $c_v \equiv \int \mathcal{C}_{vv} d^2k / 2\pi$. We might immediately anticipate that $c_v = 0$ and $c_v > 0$ for the two cases in Fig. 40.2, respectively. The less obvious fact is, however, that c_v is always an integer and, in fact, $c_v = 1$ in the right-hand case $\lambda \geq \lambda_C$.

In mathematical geometry, a famous theorem by Gauss and Bonnet says that the integral of the curvature $\mathcal{C}$ over a closed surface is $\int \mathcal{C}\, dS = 4\pi(1-g)$, where g is the number of "handles" or punctures in the surface. For instance, a sphere has $\int \mathcal{C}\, dS = 4\pi$, whereas a torus has $\int \mathcal{C}\, dS = 0$. In our case, the usual hexagonal Brillouin zone applied in Figs. 40.1 and 40.2 can be deformed into a rectangle $\{-K_x < k_x \leq K_x, -\frac{3}{2}K_y < k_y \leq \frac{3}{2}K_y\}$. If any *globally smooth* quantity is considered on the Brillouin zone, the rectangle has precisely the topology of a torus because of the periodic boundary conditions in k-space that identify opposite edges of Brillouin zone with each other. Hence, in our case, the smooth region $\lambda < \lambda_C$ is topologically equivalent to a torus and $c_v = 0$. The singularities found whenever $\lambda \geq \lambda_C$ mean that there is no globally smooth Berry connection in this case.

The integer values of the Chern number can be understood as follows: Suppose we divide the torus into two parts, one in which $\vec{\Omega}_{vv}$ is regular and one in which $\vec{\Omega}'_{vv} = \vec{\Omega}_{vv} + \nabla_{\vec{k}}\theta$ is regular. The integral of $\nabla_{\vec{k}} \times \vec{\Omega}_{vv}$ vanishes but on the curve separating the two regions we get via Stokes' theorem $c_v = \int \nabla_{\vec{k}}\theta \cdot d\vec{k} / 2\pi$. The phase

$e^{-i\theta}$ must repeat after one trip around the closed curve and, therefore, c_v can only be integer. The same argument applies, if more than two separate regions are required.

40.2 Hall Effect

We are now ready to study the impact of Berry connections and curvatures on the Hall conductivity. If we consider the case of no external magnetic field we have $\vec{\Pi}_{nm} = \vec{P}_{nm}$ and, hence, if we return to Eq.(36.7) in the spin-less case

$$\sigma_{xy} = -\frac{ie^2\hbar}{4\pi^2 m_e^2}\sum_{n,m}\int f_{nm}(\vec{k})\frac{P^x_{nm}(\vec{k})P^y_{mn}(\vec{k})}{E^2_{mn}(\vec{k})+(\hbar/\tau)^2}d^2k\,.$$

The intraband contribution $n=m$ vanishes due to the f_{nm} factor. For the interband part, Eq.(40.1) and the commutator Eq.(2.2) shows that $\vec{P}_{nm}(\vec{k}) = im_e E_{nm}(\vec{k})\vec{\Omega}_{nm}(\vec{k})/\hbar$ and we can write

$$\sigma_{xy} = -\frac{ie^2}{4\pi^2\hbar}\sum_{n,m}\int f_{nm}(\vec{k})\Omega^x_{nm}(\vec{k})\Omega^y_{mn}(\vec{k})\frac{E^2_{mn}(\vec{k})}{E^2_{mn}(\vec{k})+(\hbar/\tau)^2}d^2k\,.$$

Now assuming $\tau\approx\infty$, splitting $f_{nm} = f_n - f_m$ and using Eq.(40.2) then demonstrates that

$$\sigma_{xy} = -\frac{e^2}{4\pi^2\hbar}\sum_{n}\int f_n(\vec{k})\mathcal{C}_{nn}(\vec{k})\,d^2k\,.$$

Hence, we obtain the celebrated TKNN (after Thouless, Kohmoto, Nightingale, and Nijs [2]) formula

$$\sigma_{xy} = -\frac{e^2}{h}\sum_{n\ \mathrm{occ}} c_n\,. \tag{40.3}$$

The formula says that, for materials with completely filled or empty bands, the Hall conductivity is quantized as an integer multiple of e^2/h. At a finite frequency, the optical conductivity is

$$\sigma_{xy}(\omega) = -\frac{ie^2\hbar}{4\pi^2 m_e^2}\sum_{n,m}\int f_{nm}(\vec{k})\frac{P^x_{nm}(\vec{k})P^y_{mn}(\vec{k})}{E^2_{mn}(\vec{k})-\hbar^2(\omega+i/\tau)^2}d^2k\,.$$

Below, we plot the spectra for some representative values of λ for $\alpha = 0.2$ eV. We see that the TKNN result is exactly obeyed in the static limit.

At this point, one can raise the question whether the Haldane model is a realistic description of an actual physical solid-state system. In fact, it is not (but it can be realized with particles trapped in optical lattices). The reason is that electronic systems in the absence of magnetic fields respect time-reversal symmetry. Among other things, this implies that energies are degenerate under a simultaneous flip of k-vector and spin s_z, i.e. $E_n(-\vec{k},-s_z)=E_n(\vec{k},s_z)$ for the n'th band. This is not respected by the Haldane model since $s(-\vec{k})=-s(\vec{k})$. It may be remedied, however, by turning it into the Kane-Mele model [3]

$$\overleftrightarrow{H}=\begin{pmatrix}\alpha+\lambda s_z s & -\gamma f\\ -\gamma f^* & -\alpha-\lambda s_z s\end{pmatrix}.$$

This extended model is time-reversal symmetric and implies that our previous results should be summed over $\lambda=\pm|\lambda|$. Hence, $\sigma_{xy}(\omega)$ vanishes identically in this model. One may consider a *spin* Hall effect, however. Since the two spin directions experience opposite $\sigma_{xy}(\omega)$, they will be deflected to opposite sides in a Hall geometry such as Fig. 36.1. Hence, opposite spins collect at the two edges of the sample, similarly to the opposite charges in the normal Hall effect. Consequently, the spin Hall conductivity is just two times the Haldane Hall conductivity.

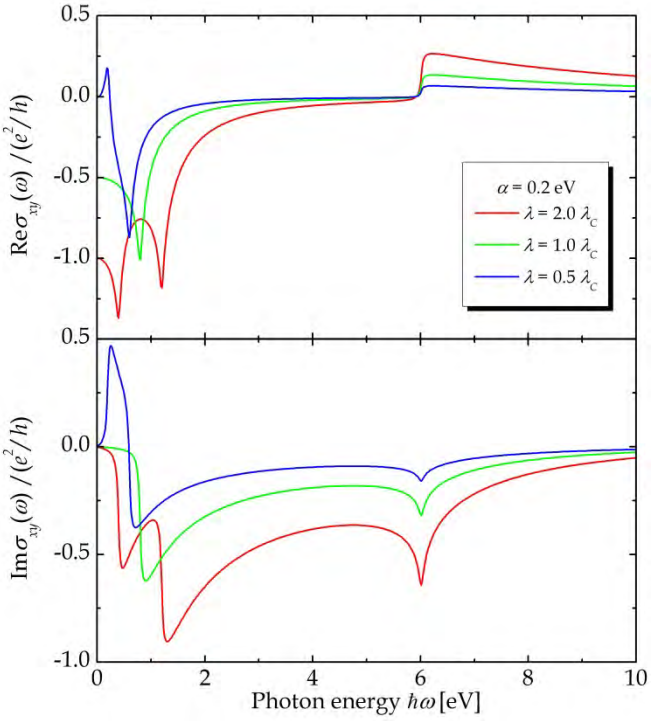


Figure 40.3. Optical Hall conductivity in the Haldane model with various λ - values .

40.3 Semiclassical Picture

The topological effects can be understood at a semiclassical level, to a certain extent. In the absence of a real magnetic field, we know that the distribution function g obeys

$$g(\vec{k},t) = f(E_{\vec{k}}) + \frac{e\tau}{\hbar}\nabla_{\vec{k}} g(\vec{k},t)\cdot\vec{\mathcal{E}}(t) - \tau\frac{\partial g(\vec{k},t)}{\partial t}.$$

We consider a single band with energy $E_{\vec{k}}$ and Berry curvature $\vec{\mathcal{C}}$. The new effects appear due to an "anomalous" velocity contribution such that

$$\vec{v} = \hbar^{-1}\left\{\nabla_{\vec{k}}E_{\vec{k}} + \vec{\mathcal{C}}\times\frac{d\vec{p}}{dt}\right\} = \hbar^{-1}\left\{\nabla_{\vec{k}}E_{\vec{k}} - e\vec{\mathcal{C}}\times\vec{\mathcal{E}}(t)\right\}.$$

The new, second term involving $\vec{\mathcal{C}}$ is the anomalous one and it leads to an "anomalous Hall effect" described by the current (not summed over spin)

$$\vec{j}^{\,AHE}(t) = \frac{e^2}{8\pi^3\hbar}\int g(\vec{k},t)\vec{\mathcal{C}}\times\vec{\mathcal{E}}(t)d^3k\,.$$

In a time-harmonic field $\vec{\mathcal{E}}(t) = \frac{1}{2}\vec{\mathcal{E}}\,e^{-i\omega t} + c.c.$, the distribution function can be expanded in harmonics as well as field strength. The 0'th order is the unperturbed case $g_0(\vec{k}) = f(E_{\vec{k}})$ the first order term is $g_1(\vec{k},\omega) = e\tau/(1-i\omega\tau)f'(E_{\vec{k}})\vec{v}\cdot\vec{\mathcal{E}}$. However, because the anomalous velocity is itself time-dependent, we find first and second harmonic currents given by

$$\vec{j}_{\omega}^{\,AHE} = \frac{e^2}{8\pi^3\hbar}\int g_0(\vec{k})\vec{\mathcal{C}}\times\vec{\mathcal{E}}\,d^3k,$$
$$\vec{j}_{2\omega}^{\,AHE} = \frac{e^2}{8\pi^3\hbar}\int g_1(\vec{k},\omega)\vec{\mathcal{C}}\times\vec{\mathcal{E}}\,d^3k.$$

As above, we now specialize to two-dimensional materials, for which $\vec{\mathcal{C}} = \mathcal{C}\vec{e}_z$, and we take $\vec{\mathcal{E}} = \mathcal{E}\vec{e}_x$. Hence, we find linear and second-harmonic conductivities

$$\sigma_{yx}^{AHE}(\omega) = \frac{e^2}{4\pi^2\hbar}\int f(E_{\vec{k}})\mathcal{C}\,d^2k,$$
$$\sigma_{yxx}^{AHE}(2\omega) = \frac{e^3\tau}{4\pi^2\hbar(1-i\omega\tau)}\int f'(E_{\vec{k}})v_x\mathcal{C}\,d^2k.$$

The first of these clearly agrees with the TKNN formula. The second is proportional to the Berry curvature dipole

$$\mathcal{D} \equiv -\frac{1}{4\pi^2}\int f'(E_{\vec{k}})v_x \mathcal{C}\, d^2k = \frac{1}{4\pi^2}\int f(E_{\vec{k}})\frac{d\mathcal{C}}{dk_x}d^2k$$

using integration by parts. There are strict requirements for obtaining a non-vanishing $\mathcal{D}$ since it must vanish if symmetry is high. Time-reversal invariance means that $\vec{\mathcal{C}}(-\vec{k}) = -\vec{\mathcal{C}}(\vec{k})$ (if separate valleys related by inversion in k-space are considered, interchange of these is implied). The largest symmetry allowing for $\mathcal{D}$ is a single symmetry line [4], for instance the mirror symmetry $k_y \to -k_y$. As a consequence, $\mathcal{D}$ vanishes for the Kane-Mele model. With the y mirror symmetry $k_y \to -k_y$, $\vec{j}^{AHE}_{2\omega}$ will be along y, i.e. $\sigma^{AHE}_{yxx}(2\omega) \neq 0$. Hence, as a better example, we will investigate gapped graphene with a slight twist. In the Dirac approximation, we will take

$$\overleftrightarrow{H} = \begin{pmatrix} \alpha & \hbar v(\eta k_x - ik_y) \\ \hbar v(\eta k_x + ik_y) & -\alpha \end{pmatrix} + \begin{pmatrix} \beta\eta k_x & 0 \\ 0 & \beta\eta k_x \end{pmatrix}.$$

Here, $\eta = \pm 1$ is the valley index and the second term is added to produce interesting results. It corresponds physically to a tilt of the Dirac cones along the x-direction. The extra term only corresponds to a local k_x - dependent shift of both bands. Hence, there is no change to neither transition energies nor momentum matrix elements and the Berry curvature for the conduction band becomes

$$\mathcal{C} = -\frac{\alpha\eta\hbar^2 v^2}{2(\alpha^2 + \hbar^2 v^2 k^2)^{3/2}},$$

clearly unaffected by the tilt term. In fact, the only effect of having $\beta \neq 0$ is via the Fermi level E_F. Assuming $E_F > 0$ and low temperature, we have on the Fermi surface

$$(\alpha^2 + \hbar^2 v^2 k^2)^{1/2} + \beta\eta k_x = E_F.$$

The associated "warping" of the Fermi surface is the reason for a non-vanishing Berry curvature dipole. The general integral is rather difficult but to first order in β the result turns out to be

$$\mathcal{D} = \frac{3\alpha\beta(E_F^2 - \alpha^2)}{4\pi E_F^4}.$$

Exercise: Tight-binding with magnetic field

In tight-binding, the simplest way of including a magnetic field is via the Peierls substitution, which says that any hopping matrix element γ_{12} between sites $\vec{r}_1$ and $\vec{r}_2$ should be replaced by $\gamma_{12}\exp(i\Phi_{12})$, where Φ_{12} is called the Peierls phase. It is given by

$$\Phi_{12} = \frac{e}{\hbar}\int_{\vec{r}_1}^{\vec{r}_2} \vec{\mathcal{A}}(\vec{r})\cdot d\vec{r}\,,$$

where the line integral is along the line joining the sites.

a) Show that using the symmetric gauge $\vec{\mathcal{A}}(\vec{r}) = \frac{1}{2}\mathcal{B}(x\vec{e}_y - y\vec{e}_x)$, the Peierls phase becomes $\Phi_{12} = \frac{e}{\hbar}\mathcal{B}S_{12}$ with $S_{12} = \frac{1}{2}(x_1y_2 - x_2y_1)$ such that $|S_{12}|$ is the hatched area shown to the left in Fig. 40.4. The sign of S_{12} is given by the right-hand rule. The quantity $\mathcal{B}S_{12}$ is the (oriented) magnetic flux through the hatched triangle and $\Phi_0 = h/e$ is frequently called the flux quantum.

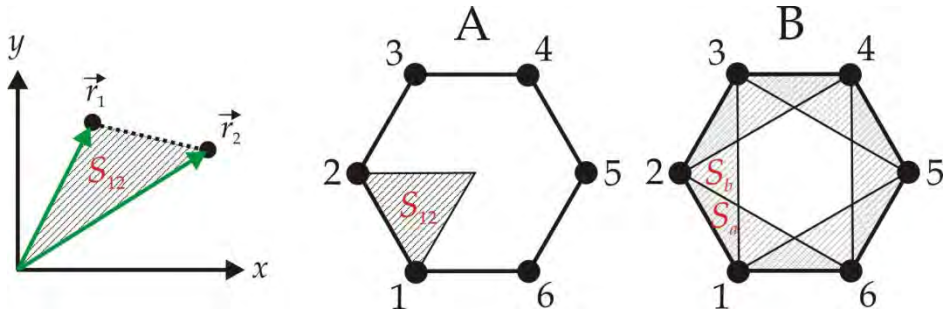


Figure 40.4. Geometry of two tight-binding sites (left) and two magnetic configurations of benzene (right).

We now consider a benzene ring as shown to the right in the figure using the graphene lattice constant *a*, i.e. the distance between e.g. sites 1 and 3. In case A, the oriented area $S_{12} = -\frac{a^2}{4\sqrt{3}}$ is clearly negative according to the right-hand rule. We also have $S_{21} = -S_{12}$ and $S_{23} = S_{34} = ... = S_{12}$.

b) In a nearest-neighbor model with zero-field hopping element $-\gamma$ and taking $\Phi \equiv \frac{e}{\hbar}\mathcal{B}S_{12}$, show that the new eigenvalues are

$$E = -\gamma\left\{\pm 2\cos\Phi, \cos\Phi \pm \sqrt{3}\sin\Phi, -\cos\Phi \pm \sqrt{3}\sin\Phi\right\}.$$

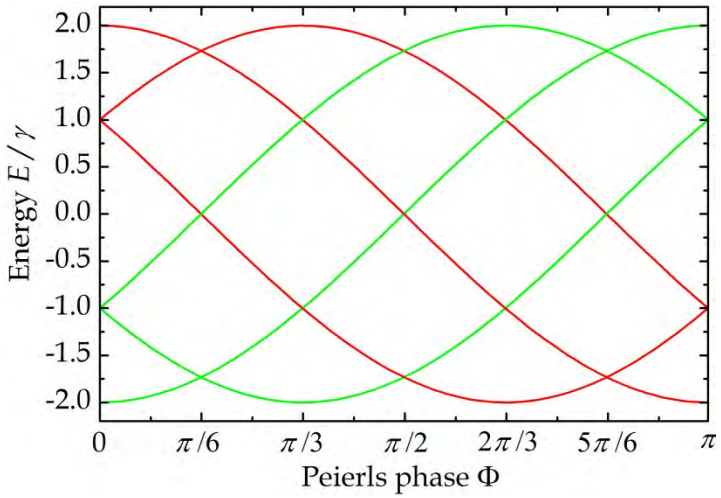


Figure 40.5. Eigenvalues E/γ versus Peierls phase Φ.

These eigenvalues are illustrated in Fig. 40.5. Note that values repreat with a period of $\pi/3$.

The hoppings in the Haldane and Kane-Mele models can be realized with the curious flux pattern in case B of Fig. 40.4. Here, the magnetic field in areas *a* and *b* are assumed equal but opposite in sign and so the fluxes cancel. Hence, the nearest-neighbor matrix elements are unchanged by the field. The next-nearest-neighbor phase, however, becomes $\Phi_{13} = \frac{e}{\hbar}\mathcal{B}S_a$. It is then clear that the hoppings are of the form $\lambda e^{\pm i\Phi}$, where λ is the zero-field next-nearest-neighbor coupling, $\Phi = \frac{e}{\hbar}\mathcal{B}S_a$, and the sign is determined by the orientation of the bond according to the right-hand rule.

References

[1] F. D. M. Haldane, Phys. Rev. Lett. 61, 2015 (1988).
[2] D. J. Thouless, M. Kohmoto, M. P. Nightingale, and M. den Nijs, Phys. Rev. Lett. 49, 405 (1982).
[3] C. L. Kane and E. J. Mele, Phys. Rev. Lett. 95, 226801 (2005).
[4] I. Sodemann and L. Fu, Phys. Rev. Lett. 115, 216806 (2015).

41. Dielectric and Transport Properties of Phonons

In this chapter, we will study the optical and transport response of the ionic lattice of crystals. In ionic (or polar) crystals such as NaCl, the individual atoms are charged because electrons have been transferred between them. These ions respond to optical fields in the terahertz (THz) range. In particular, resonances are found when the optical frequency coincides with a transverse-optical vibration of the lattice. Hence, we begin by analyzing the band structure of lattice vibration, i.e. phonons. We consider atoms placed at positions $\vec{R}+\vec{\tau}_\alpha+\vec{u}_\alpha(\vec{R})$, where $\vec{R}$ is the lattice vector, $\vec{\tau}_\alpha$ is the position of the α'th atom relative to the unit cell center and $\vec{u}_\alpha(\vec{R})$ is its displacement from equilibrium. By definition, the positions $\vec{R}+\vec{\tau}_\alpha$ are the equilibrium positions, for which all forces vanish. We wish to find the total elastic energy per unit cell as the atoms are displaced. To this end, we fix the attention on a particular reference cell at $\vec{R}=0$ and determine its interaction with all other cells. Thus, if we expand to second order in the small displacements we find

$$U_{tot}(\{\vec{u}_\alpha\})\approx U_{tot}(\{0\})+\frac{1}{2}\sum_{\vec{R},\alpha,\beta,i,j} d_{\alpha i,\beta j}(\vec{R})u_{\alpha i}(0)u_{\beta j}(\vec{R}),$$

where *i* and *j* indicate spatial directions *x*, *y* or *z* and

$$d_{\alpha i,\beta j}(\vec{R})=\left.\frac{\partial^2 U_{tot}(\vec{R}+\vec{\tau}_\beta+\vec{u}_\beta(\vec{R})-\vec{\tau}_\alpha-\vec{u}_\alpha(0))}{\partial u_{\alpha i}(0)\partial u_{\beta j}(\vec{R})}\right|_{\vec{u}_\beta(\vec{R})=\vec{u}_\alpha(0)=0}.$$

A phonon mode is characterized by the Bloch condition $\vec{u}_\alpha(\vec{R})=\vec{u}_\alpha e^{i\vec{k}\cdot\vec{R}}$ with $\vec{u}_\alpha\equiv\vec{u}_\alpha(0)$ a constant displacement amplitude. We collect all the displacements in a single super-vector $\vec{u}=(\vec{u}_1,\vec{u}_2,...)$. If the time dependence of all displacements is of the form $e^{-i\omega t}$, Newton's equations mean that the phonon dispersion relations are given as eigenvalues of the equation

$$\left\{\overleftrightarrow{D}(\vec{k})-\overleftrightarrow{M}\omega^2\right\}\cdot\vec{u}=0. \qquad (41.1)$$

Here, the dynamical and mass matrices are given by

$$\overleftrightarrow{D}(\vec{k}) = -\sum_{\vec{R}} e^{i\vec{k}\cdot\vec{R}} \begin{pmatrix} \overleftrightarrow{d}_{11} & \overleftrightarrow{d}_{12} & \cdots \\ \overleftrightarrow{d}_{21} & \overleftrightarrow{d}_{22} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}, \quad \overleftrightarrow{M} = \begin{pmatrix} M_1\overleftrightarrow{I} & 0 & 0 \\ 0 & M_2\overleftrightarrow{I} & 0 \\ 0 & 0 & \ddots \end{pmatrix}.$$

41.1 Zinc-Blende Crystals

In order to provide a specific example, we now focus on zinc-blende crystals, which include diamond structures as a sub-class. In these crystals, each unit cell contains two atoms A and B. We will choose the intra-cell position vectors as $\vec{\tau}_A = 0$ and $\vec{\tau}_B \equiv \vec{\tau} = a(1,1,1)/4$, with a the lattice constant. For two atoms having a relative vector $\vec{\tau} = \tau(1,1,1)$ symmetry dictates that

$$\overleftrightarrow{d} = \begin{pmatrix} \alpha & \beta & \beta \\ \beta & \alpha & \beta \\ \beta & \beta & \alpha \end{pmatrix}.$$

If some of the components are inverted such that $\vec{\tau} = \tau(r,s,t)$ with $r,s,t = \pm 1$ we have

$$\overleftrightarrow{d} = \begin{pmatrix} \alpha & rs\beta & rt\beta \\ rs\beta & \alpha & st\beta \\ rt\beta & st\beta & \alpha \end{pmatrix}.$$

For a zinc-blende crystal with atoms in the corners of a tetrahedron we find

$$\sum_{\vec{R}} e^{i\vec{k}\cdot\vec{R}} \overleftrightarrow{d}_{12}(\vec{R}) = \begin{pmatrix} \alpha f & \beta g_{xy} & \beta g_{xz} \\ \beta g_{xy} & \alpha f & \beta g_{yz} \\ \beta g_{xz} & \beta g_{yz} & \alpha f \end{pmatrix}, \quad \begin{aligned} f &= 1 + e^{ia(k_x+k_y)} + e^{ia(k_x+k_z)} + e^{ia(k_y+k_z)} \\ g_{xy} &= 1 + e^{ia(k_x+k_y)} - e^{ia(k_x+k_z)} - e^{ia(k_y+k_z)} \end{aligned}.$$

A rigid translation of the entire crystal must be an eigenmode corresponding to $\vec{k} = 0$ with zero eigenvalue and constant eigenvector $\vec{u} = u(1,1,...)$. This leads to the acoustic sum rule

$$\sum_{\vec{R}} \begin{pmatrix} \overleftrightarrow{d}_{11} & \overleftrightarrow{d}_{12} & \cdots \\ \overleftrightarrow{d}_{21} & \overleftrightarrow{d}_{22} & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \cdot \begin{pmatrix} (a,b,c)^T \\ (a,b,c)^T \\ \vdots \end{pmatrix} = 0.$$

This means that

$$\sum_{\vec{R}} e^{i\vec{k}\cdot\vec{R}} \overleftrightarrow{d}_{11}(\vec{R}) = \begin{pmatrix} -4\alpha & 0 & 0 \\ 0 & -4\alpha & 0 \\ 0 & 0 & -4\alpha \end{pmatrix}.$$

In addition to the short-range interactions, ionic crystals have a long-range Coulomb interaction. For a lattice with two ions per unit cell having charges $\pm Ze$, the electrostatic energy is

$$U_C = C\left\{\sum_{\vec{R}\neq 0}\frac{1}{R} - \sum_{\vec{R}}\frac{1}{|\vec{R}+\vec{\tau}|}\right\}, \quad C = \frac{Z^2e^2}{4\pi\varepsilon_0\varepsilon_\infty}.$$

The Ewald technique consists in writing 1 as the sum $\Phi(\varepsilon R)+\Phi_C(\varepsilon R)$. Here, ε is the Ewald convergence parameter, and Φ and Φ_C are the regular and complementary error functions, respectively. The term involving Φ_C converges rapidly and by Fourier-transforming the Φ term, the same rapid convergence is found there. Hence, the electrostatic energy becomes

$$U_C = U_{AA} - U_{AB}, \quad U_{AA} = C\left\{f(0) - \frac{2\varepsilon}{\sqrt{\pi}} + \sum_{\vec{R}\neq 0}\frac{\Phi_C(\varepsilon R)}{R}\right\}, \quad U_{AB} = C\left\{f(\vec{\tau}) + \sum_{\vec{R}}\frac{\Phi_C(\varepsilon|\vec{R}+\vec{\tau}|)}{|\vec{R}+\vec{\tau}|}\right\}$$

Here, we have used $\lim_{R\to 0}\Phi(\varepsilon R)/R = 2\varepsilon/\sqrt{\pi}$ and introduced

$$f(\vec{\tau}) = \sum_{\vec{R}}\frac{\Phi(\varepsilon|\vec{R}+\vec{\tau}|)}{|\vec{R}+\vec{\tau}|} = \frac{4\pi}{\Omega_{UC}}\sum_{\vec{G}}\frac{1}{G^2}e^{-G^2/4\varepsilon^2}e^{-i\vec{G}\cdot\vec{\tau}}.$$

The divergent $\vec{G}=0$ can be omitted because it cancels with the identical term in $f(0)$. For a zinc-blende lattice with lattice constant *a* we find

$$U_C = \frac{C}{a}M, \quad M = -3.782926....$$

However, taking only $M \approx a\left(f(0) - f(\vec{\tau}) - 2\varepsilon/\sqrt{\pi}\right)$ and limiting the *G*-vector sum to $G \le 16\pi/a$ and taking $\varepsilon = 6/a$ we find as an excellent approximation $M \approx -3.781$. Hence, the force constants can be approximated by taking the reciprocal space sum only. The second order change in electrostatic energy is then

$$\Delta U_{AA} \approx C\Delta f(0), \quad \Delta U_{AB} \approx C\Delta f(\vec{\tau})$$

with

$$\Delta f(\vec{\tau}) = \frac{1}{2}\sum_{\vec{R},\alpha,\beta,i,j}\left[\frac{d^2}{dR_i dR_j}\frac{\Phi(\varepsilon\,|\,\vec{R}+\vec{\tau}\,|)}{|\,\vec{R}+\vec{\tau}\,|}\right]u_{\alpha i}(0)u_{\beta j}(\vec{R})\,.$$

The associated dynamical matrices become

$$\overleftrightarrow{D}_{AA}(\vec{k}) = -C\left\{\frac{4\pi}{\Omega_{UC}}\sum_{\vec{G}}\frac{(\vec{k}+\vec{G})(\vec{k}+\vec{G})}{(\vec{k}+\vec{G})^2}e^{-(\vec{k}+\vec{G})^2/4\varepsilon^2} - \frac{4\varepsilon^3}{3\sqrt{\pi}}\overleftrightarrow{I}\right\}$$

and

$$\overleftrightarrow{D}_{AB}(\vec{k}) = C\frac{4\pi}{\Omega_{UC}}\sum_{\vec{G}}\frac{(\vec{k}+\vec{G})(\vec{k}+\vec{G})}{(\vec{k}+\vec{G})^2}e^{-(\vec{k}+\vec{G})^2/4\varepsilon^2}e^{-i(\vec{k}+\vec{G})\cdot\vec{\tau}}\,.$$

Here, we used $\lim_{R\to 0}\frac{d^2}{dR_i dR_j}\frac{\Phi(\varepsilon R)}{R} = -\frac{4\varepsilon^3}{3\sqrt{\pi}}\delta_{ij}$. These results can be expressed via the scalars

$$h_{ij} = C\left\{\frac{4\pi}{\Omega_{UC}}\sum_{\vec{G}}\frac{(\vec{k}+\vec{G})_i(\vec{k}+\vec{G})_j}{(\vec{k}+\vec{G})^2}e^{-(\vec{k}+\vec{G})^2/4\varepsilon^2} - \frac{4\varepsilon^3}{3\sqrt{\pi}}\delta_{ij}\right\}$$

$$k_{ij} = -C\frac{4\pi}{\Omega_{UC}}\sum_{\vec{G}}\frac{(\vec{k}+\vec{G})_i(\vec{k}+\vec{G})_j}{(\vec{k}+\vec{G})^2}e^{-(\vec{k}+\vec{G})^2/4\varepsilon^2}e^{-i(\vec{k}+\vec{G})\cdot\vec{\tau}}.$$

Collecting all contributions, it then follows that the full 6 by 6 dynamical matrix is given by

$$\overleftrightarrow{D} = -\begin{pmatrix} -4\alpha+h_{xx} & h_{xy} & h_{xz} & \alpha f+k_{xx} & \beta g_{xy}+k_{xy} & \beta g_{xz}+k_{xz} \\ h_{xy} & -4\alpha+h_{yy} & h_{yz} & \beta g_{xy}+k_{xy} & \alpha f+k_{yy} & \beta g_{yz}+k_{yz} \\ h_{xz} & h_{yz} & -4\alpha+h_{zz} & \beta g_{xz}+k_{xz} & \beta g_{yz}+k_{yz} & \alpha f+k_{zz} \\ \alpha f^*+k^*_{xx} & \beta g^*_{xy}+k^*_{xy} & \beta g^*_{xz}+k^*_{xz} & -4\alpha+h_{xx} & h_{xy} & h_{xz} \\ \beta g^*_{xy}+k^*_{xy} & \alpha f^*+k^*_{yy} & \beta g^*_{yz}+k^*_{yz} & h_{xy} & -4\alpha+h_{yy} & h_{yz} \\ \beta g^*_{xz}+k^*_{xz} & \beta g^*_{yz}+k^*_{yz} & \alpha f^*+k^*_{zz} & h_{xz} & h_{yz} & -4\alpha+h_{zz} \end{pmatrix}. \quad (41.2)$$

The required constants for a few important materials are listed in the table below.

Material	α [N/m]	β [N/m]	C/Ω_{UC} [N/m]	a [Å]	M_1 [u]	M_2 [u]
Si [1]	65.87	45.73	0	5.43	28	28
Ge [2]	49.54	30.82	0	5.66	72.6	72.6
GaAs	40.6	30.1	2.41	5.65	69.7	74.9

Band structures of apolar (Si) and polar (GaAs) materials are compared below.

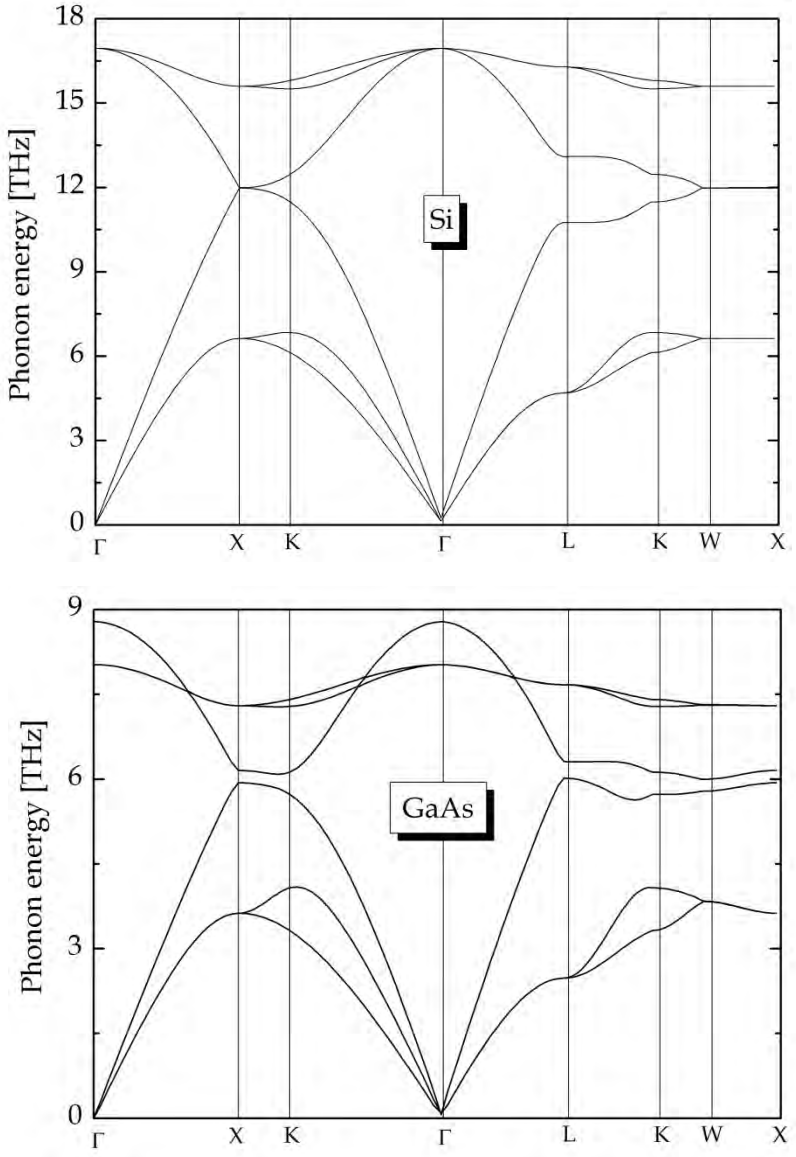


Figure 41.1. Phonon band structures of Si and GaAs.

The most notable difference between the two is the splitting of optical modes near the zone center. Hence, the two transverse-optical modes are lower in frequency than the longitudinal one for GaAs but degenerate for Si. This tendency holds generally and has important consequences for the optical response discussed below.

41.2 Properties of Phonons

We now turn to the thermodynamical and dielectric properties of phonons. The most straightforward quantity is the heat capacity that is derived from the internal energy

$$U_{int} = \int_0^\infty \hbar\omega n(\omega) D(\omega) d\omega, \quad n(\omega) = \frac{1}{\exp\{\hbar\omega / kT\} - 1}.$$

Here, n is the Bose-Einstein distribution and D is the phonon density of states. Differentiating we find

$$C_V = \frac{dU_{int}}{dT} = \int_0^\infty \hbar\omega \frac{dn(\omega)}{dT} D(\omega) d\omega, \quad \frac{dn(\omega)}{dT} = \frac{\exp\{\hbar\omega / kT\}}{\left(\exp\{\hbar\omega / kT\} - 1\right)^2} \frac{\hbar\omega}{kT^2}. \quad (41.3)$$

The density of states and phonon heat capacity of GaAs are plotted below.

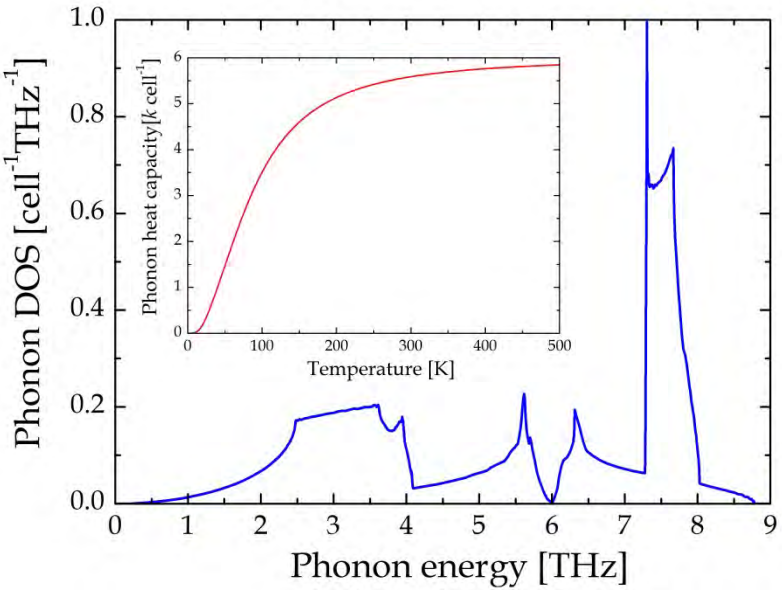


Figure 41.2. Phonon density of states and heat capacity (inset) of GaAs.

In addition, the heat conductance of phonons can be computed. This proceeds very much by analogy with the electrical conductance of electrons. Recall Eq.(10.2) for electrical conduction

$$I_e = \frac{2e}{h} \sum_n \int_{-\infty}^{\infty} \{f_L(E) - f_R(E)\} T_{n\to n}(E) dE.$$

To convert this into an expression for the heat current, we notice that (*i*) the factor 2 is from electron spin and must be omitted, (*ii*) an electron carries a charge -*e* whereas

a phonon carries an energy $\hbar\omega$, and (*iii*) the electron Fermi factor should be replaced by phonon Bose-Einstein factors. Then,

$$I_Q = \frac{1}{h}\sum_n \int_0^\infty \{n_L(\omega) - n_R(\omega)\} T_{n\to n}(\omega)\hbar\omega d(\hbar\omega)\,.$$

For electrons, we assumed a bias difference ΔV between left and right so that

$$I_e \approx -\frac{2e^2}{h}\sum_n \int_{-\infty}^\infty \frac{df(E)}{dE} T_{n\to n}(E) dE \Delta V\,.$$

Similarly, for phonons we assume a temperature difference ΔT between left and right

$$I_Q \approx \frac{1}{h}\sum_n \int_0^\infty \frac{dn(\omega)}{dT} T_{n\to n}(\omega)\hbar\omega d(\hbar\omega)\Delta T\,.$$

But inspection of Eq.(41.3) shows that the temperature derivative can be replaced by a frequency one instead. We then find a heat conductance $\kappa \equiv I_Q / \Delta T$

$$\kappa = -\frac{\hbar}{2\pi T}\sum_n \int_0^\infty \frac{dn(\omega)}{d\omega} T_{n\to n}(\omega)\omega^2 d\omega\,.$$

For electrons we had the normalized "window" function

$$W_e(E) = -\frac{df(E)}{dE}, \quad \int_{-\infty}^\infty W_e(E) dE = 1\,.$$

Similarly, for phonons we can define [4]

$$W_{ph}(\omega) = -\frac{3}{\pi^2}\left(\frac{\hbar\omega}{kT}\right)^2 \frac{dn(\omega)}{d\omega}, \quad \int_0^\infty W_{ph}(\omega) d\omega = 1\,.$$

And so eventually

$$\kappa = \frac{\pi^2 k^2 T}{3h}\sum_n \int_0^\infty T_{n\to n}(\omega) W_{ph}(\omega) d\omega\,.$$

Finally, the dielectric constant is given by

$$\varepsilon(\omega) \approx \varepsilon_\infty + \sum_n \frac{S_n}{\omega_n^2 - (\omega + i\Gamma)^2}.$$

Here, S_n is the phonon oscillator strength. It is basically determined by

$$S_n \propto \left| \left\langle 1_{\gamma,\vec{q}}; 0_{n,\vec{k}} \left| \vec{e} \cdot \vec{d} \right| 0_{\gamma,\vec{q}}; 1_{n,\vec{k}} \right\rangle \right|^2,$$

where $\left| 1_{\gamma,\vec{q}}; 0_{n,\vec{k}} \right\rangle$ is a state with one photon (momentum $\vec{q}$) and zero phonons and $\left| 0_{\gamma,\vec{q}}; 1_{n,\vec{k}} \right\rangle$ is a state with one phonon (momentum $\vec{k}$) and zero photons. Also, $\vec{d} = -e\sum_\alpha Z_\alpha \vec{u}_\alpha$ is the dipole operator and $\vec{e}$ is a polarization vector perpendicular to $\vec{q}$. Since $\vec{q}$ is very small, only zone center phonons are excited and because of $\vec{e}$ only transverse ones. Finally, at the zone center, only optical phonons have a dipole moment. Hence, the transverse-optical (TO) response reads

$$\varepsilon(\omega) \approx \varepsilon_\infty + \frac{S_{TO}}{\omega_{TO}^2 - (\omega + i\Gamma)^2} = \varepsilon_\infty + \frac{\omega_{TO}^2 (\varepsilon_{DC} - \varepsilon_\infty)}{\omega_{TO}^2 - (\omega + i\Gamma)^2}.$$

For GaAs, $\varepsilon_\infty = 10.92$, $\varepsilon_{DC} = 12.9$ and $\Gamma = 0.007\omega_T$ and the result obtained using these values is shown in Fig. 41.3.

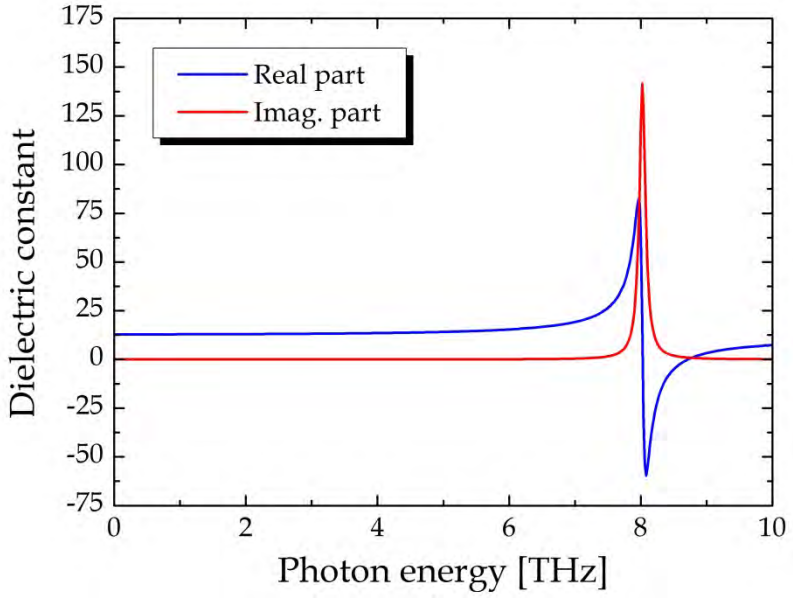
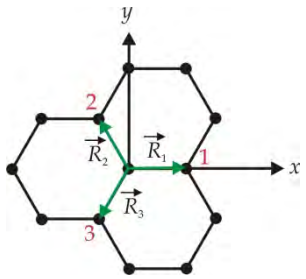


Figure 41.3. Terahertz dielectric constant of GaAs.

Exercise: Graphene phonons

In the planar geometry of graphene, the coupling of the reference atom at the origin to the atom labeled 1 is given by

$$\overleftrightarrow{d}_1 = \begin{pmatrix} \alpha_1 & 0 & 0 \\ 0 & \alpha_2 & 0 \\ 0 & 0 & \alpha_3 \end{pmatrix}.$$



These force constants correspond to in-plane stretch, in-plane bend, and out-of-plane bend, respectively. Similarly, the coupling to the atoms labeled 2 and 3 are given by rotations around the 3-axis by $\pm 120^\circ$, i.e.

$$\overleftrightarrow{d}_2 = \begin{pmatrix} \frac{1}{4}(\alpha_1+3\alpha_2) & \frac{\sqrt{3}}{4}(\alpha_1-\alpha_2) & 0 \\ \frac{\sqrt{3}}{4}(\alpha_1-\alpha_2) & \frac{1}{4}(3\alpha_1+\alpha_2) & 0 \\ 0 & 0 & \alpha_3 \end{pmatrix}, \quad \overleftrightarrow{d}_3 = \begin{pmatrix} \frac{1}{4}(\alpha_1+3\alpha_2) & \frac{\sqrt{3}}{4}(\alpha_2-\alpha_1) & 0 \\ \frac{\sqrt{3}}{4}(\alpha_2-\alpha_1) & \frac{1}{4}(3\alpha_1+\alpha_2) & 0 \\ 0 & 0 & \alpha_3 \end{pmatrix}.$$

a) Show that the dynamical matrix becomes

$$\overleftrightarrow{D} = -\begin{pmatrix} -(\overleftrightarrow{d}_1+\overleftrightarrow{d}_2+\overleftrightarrow{d}_3) & \overleftrightarrow{d}_1 e^{i\vec{k}\cdot\vec{R}_1} + \overleftrightarrow{d}_2 e^{i\vec{k}\cdot\vec{R}_2} + \overleftrightarrow{d}_3 e^{i\vec{k}\cdot\vec{R}_3} \\ \overleftrightarrow{d}_1 e^{-i\vec{k}\cdot\vec{R}_1} + \overleftrightarrow{d}_2 e^{-i\vec{k}\cdot\vec{R}_2} + \overleftrightarrow{d}_3 e^{-i\vec{k}\cdot\vec{R}_3} & -(\overleftrightarrow{d}_1+\overleftrightarrow{d}_2+\overleftrightarrow{d}_3) \end{pmatrix}.$$

b) Using $\{\alpha_1,\alpha_2,\alpha_3\} = \{418,152,102\}$ N/m [3] show that the band structure is as shown below (left).

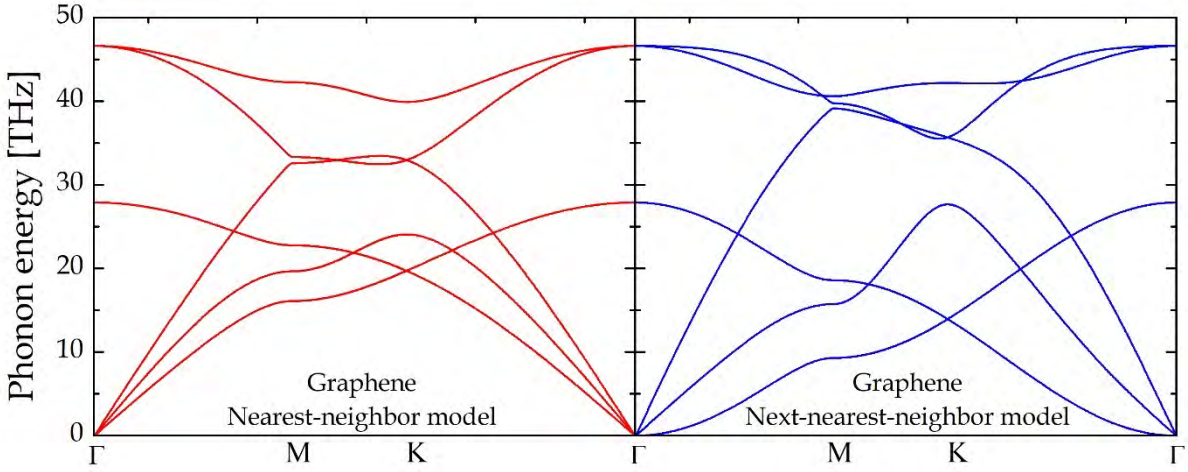


Figure 41.4. Phonon band structure of graphene in nearest- and next-nearest neighbor model left and right.

Next-nearest neighbor forces can be included as a simple extension of the model by adding an un-rotated force matrix of the form $\mathrm{diag}(\tilde{\alpha}_1, \tilde{\alpha}_2, \tilde{\alpha}_3)$.

c) Show that this adds the following block in the upper left and lower right blocks of the dynamical matrix

$$\overleftrightarrow{d}_{NNN} = \begin{pmatrix} 3\tilde{\alpha}_1(1-c_{xy}) + \tilde{\alpha}_2(3-c_{xy}-2c_y) & \sqrt{3}(\tilde{\alpha}_1 - \tilde{\alpha}_2)s_{xy} & 0 \\ \sqrt{3}(\tilde{\alpha}_1 - \tilde{\alpha}_2)s_{xy} & 3\tilde{\alpha}_2(1-c_{xy}) + \tilde{\alpha}_1(3-c_{xy}-2c_y) & 0 \\ 0 & 0 & 2\tilde{\alpha}_3\left\{3-2c_{xy}-c_y\right\} \end{pmatrix}$$

$$c_{xy} = \cos(3k_x a/2)\cos(\sqrt{3}k_y a/2), \quad s_{xy} = \sin(3k_x a/2)\sin(\sqrt{3}k_y a/2), \quad c_y = \cos(\sqrt{3}k_y a).$$

The next-nearest neighbor force constants fitted to a full DFT band structure in Ref. [3] are $\{\tilde{\alpha}_1, \tilde{\alpha}_2, \tilde{\alpha}_3\} = \{76, -43.5, -10.8\}$ N/m. However, experimentally it is found that the lowest (so-called ZA) band behaves as k^2 rather than linearly near the Γ point. This is clearly not found in the nearest neighbor model and, neither, in the next-nearest neighbor using these values. In fact, it can be shown that the requirement for a parabolic band is $\tilde{\alpha}_3 = -\alpha_3/6$. This value is also required to fulfill the "rotational sum rule" that says that a rigid rotation must be a zero-frequency eigenmode [3]. Hence, we use this value in the right-hand plot of Fig. 41.4.

References

[1] A. Mazur and J. Pollmann, Phys. Rev. B39, 5261 (1989).
[2] F. Herman, J. Phys. Chem. Solids 8, 405 (1959).
[3] J. Zimmermann, P. Pavone, and G. Cuniberti, Phys. Rev. B78, 045410 (2008).
[4] C. Jeong, S. Datta, and M. Lundstrom, J. Appl. Phys. 109, 073718 (2011).

42. Polarizability of Molecules

In this chapter, we will look at models of molecules and their polarizabilities. In a simple picture, molecular states are linear combinations of atomic orbitals (LCAOs) also known as tight-binding states (see Appendix 2). This means that molecular states are of the form $\psi_n(\vec{r}) = \sum_i c_i^{(n)} \varphi_i(\vec{r})$ with φ_i the i'th atomic orbital and $\{c_i^{(n)}\}$ the coefficient eigenvector of molecular state n. Within the independent electron model, the lowest N molecular states are occupied in an N electron molecule. We will study transitions associated with excited states, where one of the initially occupied molecular states $\psi_n(\vec{r})$ is replaced by an empty one $\psi_m(\vec{r})$ with $m > N$. A particularly simple situation is found, if all atomic orbitals φ_i are identical except for their position. This is the case, to a good approximation, in molecular hydrogen H_2 and well as planar (conjugated) carbon molecules such as benzene. In the former, the atomic states are hydrogen orbitals, while in the latter they are π - orbitals. We will use the notation

$$H_{ij} = \langle \varphi_i | \hat{H} | \varphi_j \rangle, \quad S_{ij} = \langle \varphi_i | \varphi_j \rangle,$$

for Hamilton and overlap matrix elements, respectively. Each atomic orbital is assumed to have definite parity and, if all φ_i are identical except for center position, it is readily shown that $x_{ij} = \langle \varphi_i | x | \varphi_j \rangle = \overline{x}_{ij} S_{ij}$ with $\overline{x}_{ij} = \frac{1}{2}(x_i + x_j)$ the average of the atomic centers. Hence, $\langle \psi_n | x | \psi_m \rangle = \sum_{i,j} c_i^{(n)*} c_j^{(m)} \overline{x}_{ij} S_{ij}$. A particularly simple model is the nearest-neighbor approximation, in which $H_{ii} = \varepsilon$ and $S_{ii} = 1$, while $H_{ij} = -\gamma$ and $S_{ij} = s$ if atoms i and j are nearest neighbors and otherwise vanish. This leads to eigenvalue problems of the form $\overleftrightarrow{H} \cdot \vec{c}_n = E_n \overleftrightarrow{S} \cdot \vec{c}_n$ with Hamilton and overlap matrices

$$\overleftrightarrow{H} = \begin{pmatrix} \varepsilon & -\gamma & 0 & \cdots \\ -\gamma & \varepsilon & -\gamma & \\ 0 & -\gamma & \varepsilon & \\ \vdots & & & \ddots \end{pmatrix}, \quad \overleftrightarrow{S} = \begin{pmatrix} 1 & s & 0 & \cdots \\ s & 1 & s & \\ 0 & s & 1 & \\ \vdots & & & \ddots \end{pmatrix}.$$

An even simpler version is found in the orthogonal approximation $s \approx 0$.

42.1. Hydrogen Molecular Ion

We will start by investigating the hydrogen molecular ion H_2^+ with just a single electron and then turn to conjugated molecules. Hence, we consider the H_2^+ Hamiltonian with two protons A, B at $\vec{R}_{A,B}$

$$\hat{h}(\vec{r}) = -\frac{1}{2}\nabla^2 - \frac{1}{r_A} - \frac{1}{r_B},$$

where $r_{A,B} = |\vec{r} - \vec{R}_{A,B}|$ is the distance to the nuclei. The atomic orbitals are $\varphi_{A,B}(\vec{r}) = \varphi(\vec{r}_{A,B})$ and we have $\varepsilon = \langle \varphi_A | \hat{h} | \varphi_A \rangle$ and $-\gamma = \langle \varphi_A | \hat{h} | \varphi_B \rangle$. If overlap is retained and only one orbital at each site is included, the eigenstates and eigenvalues are

$$E_\pm = \frac{\varepsilon \mp \gamma}{1 \pm s}, \quad \vec{c}_\pm = \frac{1}{\sqrt{2(1 \pm s)}} \begin{pmatrix} 1 \\ \pm 1 \end{pmatrix}. \tag{42.1}$$

If $\gamma > 0$, the $+/-$ solutions are ground and excited states, respectively. We now assume the atomic orbitals $\varphi_{A,B}$ to be hydrogen-type with principle quantum number n, i.e. eigenstates of $-\frac{1}{2}\nabla^2 - \zeta / r$ with eigenvalue $-\zeta^2 / (2n^2)$. These include $1s$, $2p_\sigma$ and $2p_\pi$ types (taking the molecular axis along x).

$$s(\vec{r}) = \left(\frac{\zeta_S^3}{\pi}\right)^{1/2} e^{-\zeta_S r}, \quad p^\sigma(\vec{r}) = \left(\frac{\zeta_P^5}{2^5 \pi}\right)^{1/2} x e^{-\zeta_P r/2}, \quad p^\pi(\vec{r}) = \left(\frac{\zeta_P^5}{2^5 \pi}\right)^{1/2} y e^{-\zeta_P r/2}.$$

In such cases, the virial theorem leads to $\langle \varphi_A | r_A^{-1} | \varphi_A \rangle = \zeta / n^2$ and

$$\langle \varphi_A | \hat{h} | \varphi_j \rangle = -\frac{\zeta^2}{2n^2} \langle \varphi_A | \varphi_j \rangle - \langle \varphi_A | \frac{1 - \zeta}{r_A} + \frac{1}{r_B} | \varphi_j \rangle.$$

This means that

$$\varepsilon = \frac{\zeta^2 - 2\zeta}{2n^2} - \langle \varphi_A | \frac{1}{r_B} | \varphi_A \rangle, \quad \gamma = \frac{\zeta^2}{2n^2} s + (2 - \zeta) \langle \varphi_A | \frac{1}{r_A} | \varphi_B \rangle.$$

If the nuclei are separated by a distance R, we find the characteristic integrals for s states

$$S_{ss\sigma} = \langle s_A | s_B \rangle = (1 + x + \frac{1}{3}x^2) e^{-x}, \langle s_A | \frac{1}{r_B} | s_A \rangle = \frac{\zeta_S}{x} \left\{ 1 - (1 + x) e^{-2x} \right\}, \langle s_A | \frac{1}{r_A} | s_B \rangle = \zeta_S (1 + x) e^{-x},$$

where $x = \zeta_S R$. Similarly, for p_σ states

$$S_{pp\sigma} = \left\langle p_A^\sigma \middle| p_B^\sigma \right\rangle = \frac{1}{15}(15+15x+3x^2-2x^3-x^4)e^{-x}, \quad \left\langle p_A^\sigma \middle| \frac{1}{r_A} \middle| p_B^\sigma \right\rangle = \frac{\zeta_P}{4}(1+x-\frac{1}{3}x^3)e^{-x},$$
$$\left\langle p_A^\sigma \middle| \frac{1}{r_B} \middle| p_A^\sigma \right\rangle = \frac{\zeta_P}{4x^3}\left\{6+2x^2-(6+12x+14x^2+11x^3+6x^4+2x^5)e^{-2x}\right\},$$

and p_π states

$$S_{pp\pi} = \left\langle p_A^\pi \middle| p_B^\pi \right\rangle = \frac{1}{15}(15+15x+6x^2+x^3)e^{-x}, \quad \left\langle p_A^\pi \middle| \frac{1}{r_A} \middle| p_B^\pi \right\rangle = \frac{\zeta_P}{4}(1+x+\frac{1}{3}x^2)e^{-x},$$
$$\left\langle p_A^\pi \middle| \frac{1}{r_B} \middle| p_A^\pi \right\rangle = \frac{\zeta_P}{2x^3}\left\{-\frac{3}{2}+x^2+(3+6x+4x^2+x^3)e^{-2x}\right\},$$

where $x = \zeta_P R/2$. For s states, this means that

$$\varepsilon_S = \frac{e^{-2\zeta_S R}-1}{R} + \zeta_S\left\{\tfrac{1}{2}\zeta_S + e^{-2\zeta_S R}-1\right\}, \quad \gamma_S = 2\zeta_S e^{-\zeta_S R} + \frac{\zeta_S^2}{6}\left\{3(4R-1)-3R\zeta_S+R^2\zeta_S^2\right\}e^{-\zeta_S R}.$$

The exponent (actually screened nuclear charge) ζ_S should be found by minimizing the ground state energy. Ground and excited states formed from 1*s* states are denoted $1\sigma_g$ and $1\sigma_u$, respectively, with *g* and *u* for *gerade* and *ungerade* (even and odd). To find the total molecular energy, the nuclear repulsion $1/R$ must be added to the electronic energy E_+ from Eq.(42.1). In addition, we shift the zero-point upwards by $1/2$ such that $\lim_{R\to\infty} E_+(R) = 0$. The energies $E(1\sigma_g)$ and $E(1\sigma_u)$, as well as the optimal exponent, are shown in Fig. 42.1. It is seen that ζ_S varies between 2 and 1 as R is increased, as expected. For p_σ states, the lower (E_-) and higher (E_+) states are $E(2\sigma_g)$ and $E(2\sigma_u)$, respectively, and the optimal exponent is denoted ζ_P^σ. These states are not really proper eigenstates as they, in reality, will hybridize with $1\sigma_g$ and $1\sigma_u$ to form linear combinations. We also show p_π states $E(2\pi_g)$ and $E(2\pi_u)$ including their optimal exponent ζ_P^π. The equilibrium distance for the ground state is at $R \approx 2$ Bohr with a binding energy of $-2.35\,\text{eV}$.

The nuclei perform vibrations around the equilibrium separation. By fitting the ground state curve to a parabola near the minimum $E_+(R) \approx E_+(R_0) + \frac{1}{2}k(R-R_0)^2$, one finds $k = 2.58\ \text{eV/Bohr}^2$. For H_2 (see exercise), the value is $k = 9.72\ \text{eV/Bohr}^2$. In both cases, the vibration frequency is $\omega_{vib} = \sqrt{k/M_r}$ with the reduced mass given by half the proton mass $M_r = M_p/2$. For H_2^+ and H_2, we then get $\omega_{vib} = 10.2\ \text{mHa} = 2232\ \text{cm}^{-1}$ and $\omega_{vib} = 19.7\ \text{mHa} = 4331\ \text{cm}^{-1}$, respectively.

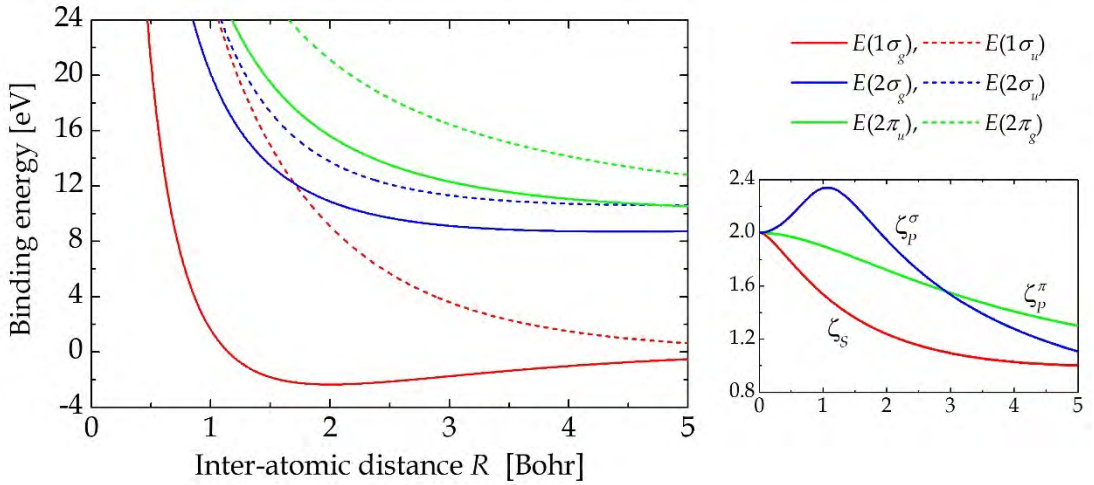


Figure 42.1. Ground and excited state energies for the H_2^+ molecular ion. The small plot shows the optimal exponents.

The transition dipole moment between $1\sigma_g$ and $2\sigma_u$ is easily seen to be $\langle 1\sigma_g|x|1\sigma_u\rangle = -\frac{1}{2}R/(1-S_{ss\sigma}^2)^{1/2}$. Moreover, for the *sp* onsite matrix element

$$X_{sp} = \langle s_A|x|p_A^\sigma\rangle = \langle s_A|y|p_A^\pi\rangle = 2^{15/2}\frac{\zeta_S^{3/2}\zeta_P^{5/2}}{(2\zeta_S+\zeta_P)^5}.$$

Transitions to $2\sigma_u$ and $2\pi_u$ are a little more involved and yield

$$\langle 1\sigma_g|x|2\sigma_u\rangle = \frac{X_{sp}+\frac{1}{2}\langle s_A|x|p_B^\sigma\rangle+\frac{1}{2}\langle s_B|x|p_A^\sigma\rangle}{\sqrt{(1+S_{ss\sigma})(1+S_{pp\sigma})}},\quad \langle 1\sigma_g|y|2\pi_u\rangle = \frac{X_{sp}+\langle s_A|y|p_B^\pi\rangle}{\sqrt{(1+S_{ss\sigma})(1+S_{pp\pi})}}.$$

The remaining integrals are related by $\langle s_B|x|p_A^\sigma\rangle = \langle s_A|x|p_B^\sigma\rangle - RS_{sp\sigma}$ with $S_{sp\sigma} = \langle s_A|p_B^\sigma\rangle$. The general integrals are complicated but in the simpler case $\zeta_P = \zeta_S$ (in the *R*-dependence) we find with $x = \zeta_S R$

$$S_{sp\sigma} = \frac{16\sqrt{2}}{27x^2}\left\{(64+32x-16x^2+3x^3)e^{-x/2}-8(8+8x+x^2)e^{-x}\right\},$$

$$\langle s_A|x|p_B^\sigma\rangle = -\frac{3X_{sp}}{x^3}\left\{(512+256x+64x^2-30x^3+3x^4)e^{-x/2}-\right.$$
$$\left.(512+512x+256x^2+45x^3+3x^4)e^{-x}\right\},$$

$$\langle s_A|y|p_B^\pi\rangle = -\frac{3X_{sp}}{x^3}\left\{(128+64x-24x^2+3x^3)e^{-x/2}-(256+256x+48x^2+3x^3)e^{-x}\right\}.$$

The polarizability is obviously anisotropic with different components parallel and perpendicular to the bond. As a simple approximation, we will only include the lowest transition in each case, so that

$$\alpha_{\parallel}(\omega,R)\approx\frac{2\,|\langle 1\sigma_g|x|1\sigma_u\rangle|^2\,[E(1\sigma_u)-E(1\sigma_g)]}{[E(1\sigma_u)-E(1\sigma_g)]^2-\omega^2},$$

$$\alpha_{\perp}(\omega,R)\approx\frac{2\,|\langle 1\sigma_g|x|2\pi_u\rangle|^2\,[E(2\pi_u)-E(1\sigma_g)]}{[E(2\pi_u)-E(1\sigma_g)]^2-\omega^2}.$$

The static values are shown in Fig. 42.2 as functions of the inter-atomic distance. The values at the equilibrium distance are found to be 6.06 and 1.79, respectively, in rather good agreement with highly accurate model values 5.08 and 1.76 [1]. The parallel component grows without limit as R increases, while the perpendicular one goes through a maximum and, in fact, settles at the hydrogen atom value at large separations.

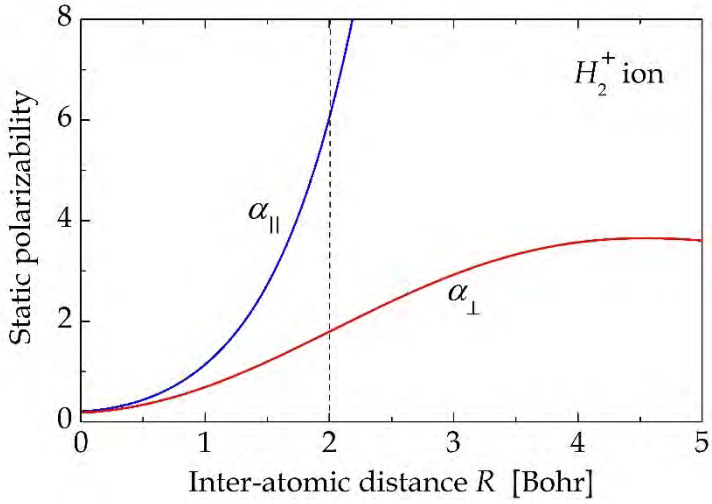


Figure 42.2. Cartesian components of the static polarizability of the H_2^+ molecular ion. The equilibrium separation is indicated by the dashed line.

42.2. Conjugated Carbon Molecules

Conjugated carbon molecules are planar structures with alternating single and double bonds between carbon atoms (see App. 2). All atoms lie within the same plane, taken as (x,y), and, hence, parity with respect to reflection in this plane is a symmetry. This means that eigenstates are either even or odd functions of z. As a consequence, p_z states decouple from the rest and form delocalized molecular π-orbitals. Models of such states are rather simple and, as an example, the six carbon benzene ring is described by Eq.(6.1):

$$\overleftrightarrow{H} = -\begin{pmatrix} 0 & \gamma & 0 & 0 & 0 & \gamma \\ \gamma & 0 & \gamma & 0 & 0 & 0 \\ 0 & \gamma & 0 & \gamma & 0 & 0 \\ 0 & 0 & \gamma & 0 & \gamma & 0 \\ 0 & 0 & 0 & \gamma & 0 & \gamma \\ \gamma & 0 & 0 & 0 & \gamma & 0 \end{pmatrix}.$$

Note that, compared to Chapter 6, we use a different ordering of atoms here, according to their sequence around the ring. The eigenvectors are of the form $c_k^{(n)} = e^{ik(n-1)\pi/3}/\sqrt{6}$ with $k,n \in [0,5]$. Here, n is the state index and the associated eigenvalues are $E_n = \gamma\{-1,-2,-1,1,2,1\}$. Six electrons occupy π - states, and, hence, with this labelling, the states 0, 1, and 2 are doubly occupied, while 3, 4, and 5 are empty. We can now compute the dipole matrix elements using $x_k = a\cos\frac{k\pi}{3}$ and $y_k = a\sin\frac{k\pi}{3}$ with C-C bond length a. The absolute values of the dipole matrix elements $X_{nm} = \langle\psi_n|x|\psi_m\rangle$ and $Y_{nm} = \langle\psi_n|y|\psi_m\rangle$ are identical and, when collected as a matrix, given by

$$|\overleftrightarrow{X}| = |\overleftrightarrow{Y}| = \frac{a}{2}\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

The structure is seen to be identical to the original Hamiltonian (but with a different meaning of row and column indices). Now, we look for allowed transitions between occupied and empty states. From the matrix above, these are seen to be $|\psi_0\rangle \to |\psi_5\rangle$ and $|\psi_3\rangle \to |\psi_4\rangle$, both with a transition energy of 2γ. It follows that the polarizability is isotropic and has a single resonance

$$\alpha_{xx}(\omega) = \alpha_{yy}(\omega) = \frac{4\gamma a^2}{4\gamma^2 - \omega^2}.$$

If N benzene rings are joined along the x-axis at their *para* positions, we form N-paraphenylenes, such as those shown in Fig. 42.3. These are anisotropic and, in the limit of large N, form the conjugated polymer poly-paraphenylene. The evolution of the polarizabilities (normalized by N) is shown in Fig. 42.3, in which broadening of $i0.02\gamma$ is added to the frequency. For the isolated ring, a single isotropic resonance at $\omega = 2\gamma$ is observed, as expected. For chains, the long-axis absorption onset shifts to low frequencies, while short-axis absorption sets in around $\sim 1.5\gamma$. For essentially

infinite chains, the absorption edge is at the band gap that can be shown to be $E_g = 2\gamma(3-\sqrt{8})^{1/2} \approx 0.828\gamma$.

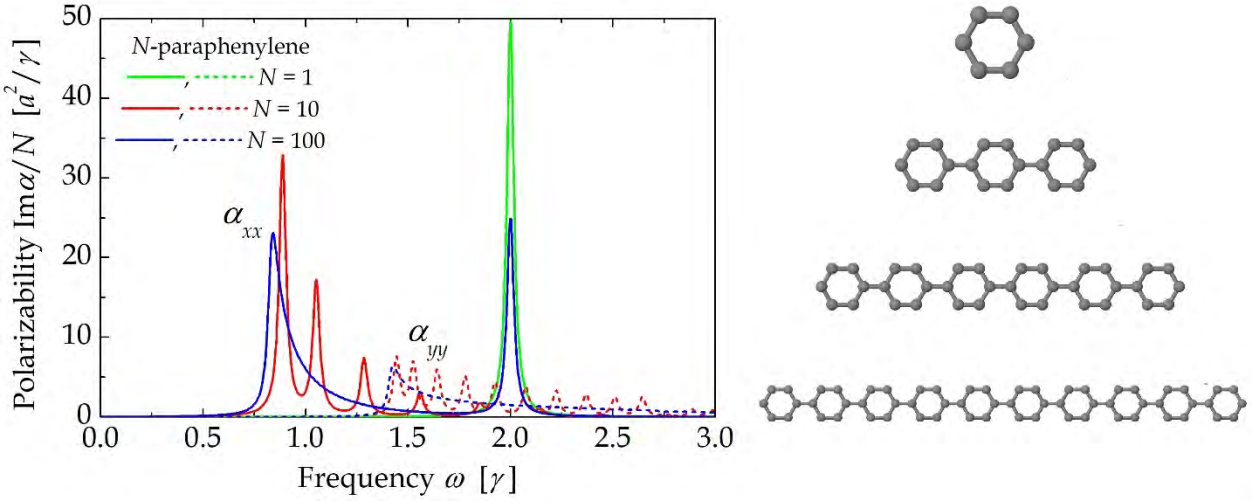


Figure 42.3. Normalized absorption spectra for different *N*-paraphenylenes. Examples with *N* = 1, 3, 6, and 10 are shown to the right.

The vibrational modes of large molecules are much more complicated that small ones. Even if only planar vibrations are considered, we have 12 degrees of freedom and, thus, 12 eigenmodes in benzene. To model these, we adopt a next-nearest neighbor force-field model. Since all masses are identical, we need to solve the problem

$$\left\{\vec{\vec{D}} - M\omega^2\right\}\cdot\vec{u} = 0\,, \tag{42.2}$$

where the dynamical matrix is $\vec{\vec{D}}$. We only include bond stretching with spring constants K and k for nearest and next-nearest neighbors, respectively. For a bond along the x-axis, the interaction is given by $\vec{\vec{d}} = \mathrm{diag}(1,0)$ times a spring constant. At other angles, we apply a rotation $\vec{\vec{d}}_v = \vec{\vec{R}}(v)\cdot\vec{\vec{d}}\cdot\vec{\vec{R}}(-v)$ with angle v. In this manner, it can be shown that

$$\vec{\vec{D}} = -\begin{pmatrix} \vec{\vec{D}}_0 & K\vec{\vec{d}}_{-120} & k\vec{\vec{d}}_{30} & 0 & k\vec{\vec{d}}_{-30} & K\vec{\vec{d}}_{120} \\ K\vec{\vec{d}}_{-120} & \vec{\vec{D}}_1 & K\vec{\vec{d}}_0 & k\vec{\vec{d}}_{-30} & 0 & k\vec{\vec{d}}_{90} \\ k\vec{\vec{d}}_{30} & K\vec{\vec{d}}_0 & \vec{\vec{D}}_2 & K\vec{\vec{d}}_{120} & k\vec{\vec{d}}_{90} & 0 \\ 0 & k\vec{\vec{d}}_{-30} & K\vec{\vec{d}}_{120} & \vec{\vec{D}}_3 & K\vec{\vec{d}}_{-120} & k\vec{\vec{d}}_{30} \\ k\vec{\vec{d}}_{-30} & 0 & k\vec{\vec{d}}_{90} & K\vec{\vec{d}}_{-120} & \vec{\vec{D}}_4 & K\vec{\vec{d}}_0 \\ K\vec{\vec{d}}_{120} & k\vec{\vec{d}}_{90} & 0 & k\vec{\vec{d}}_{30} & K\vec{\vec{d}}_0 & \vec{\vec{D}}_5 \end{pmatrix}. \tag{42.3}$$

In each row, the acoustic sum rule implies that the diagonal element equals the negative of the sum of the other terms so, for instance, $\overleftrightarrow{D}_0 = -K(\overleftrightarrow{d}_{-120} + \overleftrightarrow{d}_{120}) - k(\overleftrightarrow{d}_{30} + \overleftrightarrow{d}_{-30})$. The model is clearly quite approximate, since Hydrogen atoms are omitted. We fit the spring constants K and k to match the breathing and dimerization modes (7 and 12 in Fig. 42.4 with resonances $\omega_{vib}^2 = (3k+K)/M$ and $3K/M$) using accurate data from [2]. This leads to $K/M = 631 \cdot 10^3$ cm^{-2} and $k/M = 132.5 \cdot 10^3$ cm^{-2}. The 12 modes and their eigenfrequencies are shown in Fig. 42.4. It is noted that pure translational and rotational modes have zero frequency. Also, degeneracies between (4, 5), (8, 9), and (10, 11) appear. The high number of degeneracies reflect the high symmetry of the benzene molecule.

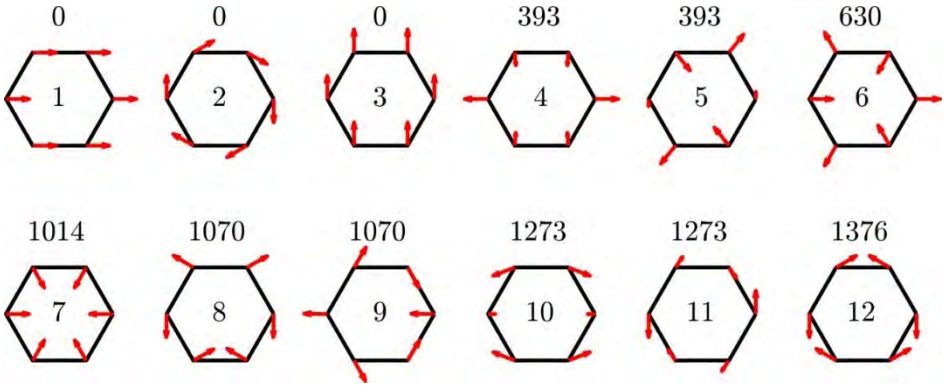


Figure 42.4. Vibrational modes of the benzene ring with frequencies in cm^{-1} added above.

In a general polygon with radius R, the breathing mode frequency is $\omega_{\text{breathing}} = (Ka^2/M + ka_{NNN}^2/M)^{1/2}/R$, where a_{NNN} is the distance between next-nearest neighbors. This clearly agrees with the benzene case above, for which $a = R$ and $a_{NNN} = \sqrt{3}R$. Similarly to graphene in the previous chapter, one could add nearest- and next-nearest-neighbor tangential force constants $K_\perp$ and $k_\perp$ to improve the fit to more accurate models or experiments. In this case, the condition that the rigid rotation (mode 2 above) has a vanishing eigenfrequency is $k_\perp = -K_\perp/3$, which can be shown from the rotational sum rule.

42.3. Raman Scattering

In the approach above, we simply treated the nuclear positions as variable, but fixed parameters. In reality, excitation by light could cause the molecule to vibrate. In turn, the scattered light would have lost energy compared to the incident beam and emission would be at the *Stokes* frequency $\omega_s = \omega - \omega_{vib}$. If the molecule is already vibrating before photon absorption, the scattered light may gain this extra energy and produce emission at the *anti-Stokes* frequency $\omega_{as} = \omega + \omega_{vib}$. This can only occur, if a phonon is already present, due to e.g. temperature. Such *Raman* scattering, therefore, provides a useful optical probe of vibrational frequencies. Bose-statistical occupation factors $1+f$ and f with $f = 1/[\exp(\hbar\omega_{vib}/kT) - 1]$ appear in the Stokes

and anti-Stokes terms, such that at low temperature, only the former occurs. A simple, semiclassical way of describing the Raman effect is by including a harmonic vibration around the equilibrium value R_0, for instance $R(t) = R_0 + \delta R \cos(\omega_{vib} t)$ in the Hydrogen molecule. If the oscillation amplitude δR is small, we may Taylor expand to first order. Hence, the induced dipole moment becomes

$$\vec{p}(t) = \tfrac{1}{2}\overleftrightarrow{\alpha}(\omega, R(t)) \cdot \vec{\mathcal{E}}(t) \approx \tfrac{1}{2}\overleftrightarrow{\alpha}(\omega, R_0) \cdot \vec{\mathcal{E}}_0 e^{-i\omega t} +$$
$$\tfrac{1}{4}\overleftrightarrow{\alpha}'(\omega, R_0) \cdot \vec{\mathcal{E}}_0 e^{-i(\omega - \omega_{vib})t} \delta R + \tfrac{1}{4}\overleftrightarrow{\alpha}'(\omega, R_0) \cdot \vec{\mathcal{E}}_0 e^{-i(\omega + \omega_{vib})t} \delta R + c.c.$$

This demonstrates that the new terms oscillating at the shifted frequency are proportional to the distance-derivative of the polarizability $\overleftrightarrow{\alpha}'(\omega, R_0)$ at the equilibrium position. Terms oscillating as $e^{-i(\omega - \omega_{vib})t}$ and $e^{-i(\omega + \omega_{vib})t}$ represent scattered photons with decreased and increased energy, respectively, corresponding to creation and absorption of a vibration quantum ("phonon"). The amplitude δR is found by equating the classical energy $\frac{1}{2}k(\delta R)^2$ to the quantum value $\hbar\omega_{vib}$ so that $\delta R = (2\hbar / M_r \omega_{vib})^{1/2}$. When plotting the Raman response, we introduce the Raman shift $\delta\omega = \omega - \omega_s$ and use $\omega_s = \omega_0 - \omega_{vib}$ with ω_0 the equilibrium resonance to write $\omega = \omega_0 - \omega_{vib} + \delta\omega$. We then plot the imaginary part of α_{Raman} versus $\delta\omega$ such that resonances occur if $\delta\omega = \omega_{vib}$. Thus, we finally find the Raman polarizability

$$\overleftrightarrow{\alpha}_{Raman}(\omega_s) = \tfrac{1}{2}\overleftrightarrow{\alpha}'(\omega, R_0)\delta R = \left(\frac{\hbar}{2M_r\omega_{vib}}\right)^{1/2} \overleftrightarrow{\alpha}'(\omega_0 - \omega_{vib} + \delta\omega, R_0).$$

If several vibrational modes $\{Q_k\}$ with eigenfrequencies ω_k exist, the nuclear positions are displaced by $\delta\vec{R} = \sum_k Q_k \vec{u}_k$, where $\vec{u}_k$ is the normalized eigenvector of the k'th mode found from Eq.(42.2), and the Raman polarizability becomes

$$\overleftrightarrow{\alpha}_{Raman}(\omega_s) = \sum_k \left(\frac{\hbar}{2M_r\omega_k}\right)^{1/2} \frac{\partial}{\partial Q_k} \overleftrightarrow{\alpha}(\omega_0 - \omega_k + \delta\omega, \vec{R})\Big|_{\{Q_k\}=0}.$$

The notation $\{Q_k\} = 0$ means that the derivatives are to be evaluated in the equilibrium configuration. Not all modes lead to peaks, however, since only some are Raman active. In Fig. 42.5, we see that this is the case in benzene, for which only modes 4/5, 7, and 10/11 are Raman active. These are the modes, for which a deformation leads to a linear change in polarizability. It is obvious that this is not the case for modes 1-3. Also, modes like 6 are Raman-inactive because the polarizability is clearly an even function of the amplitude, in this case Q_6. A similar argument holds for modes 8/9 and 12.

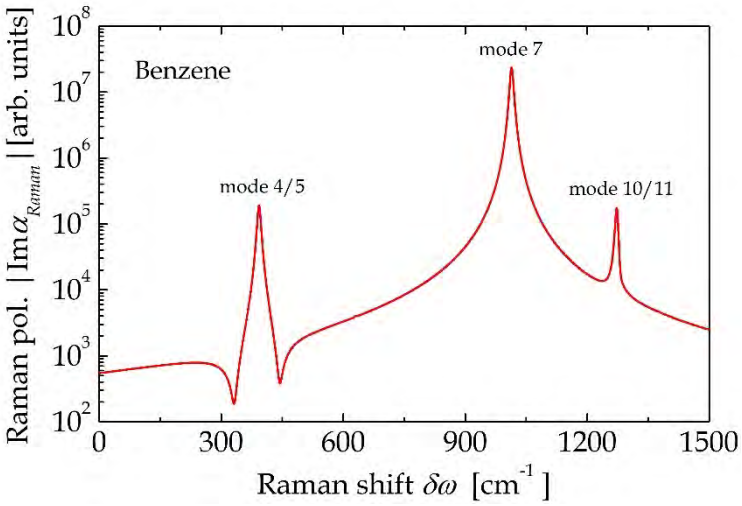


Figure 42.5. Raman spectrum of benzene with modes referring to Fig. 42.4. The intense breathing mode 7 has been down-scaled by 10^{-5} and broadening is 5 cm^{-1}.

Exercise: Hydrogen molecule

In this exercise, we use variational calculus to estimate the ground state energy of the two-electron molecule H_2 described by the Hamiltonian

$$\hat{H} = -\frac{1}{2}\nabla_1^2 - \frac{1}{2}\nabla_2^2 - \frac{1}{r_{1A}} - \frac{1}{r_{1B}} - \frac{1}{r_{2A}} - \frac{1}{r_{2B}} + \frac{1}{r_{12}} = \hat{h}(\vec{r}_1) + \hat{h}(\vec{r}_2) + \frac{1}{r_{12}}.$$

The last relation shows that the H_2 Hamiltonian consists of separate H_2^+ terms for each electron plus an interaction term. In order to set up matrix elements, we need the integrals

$$V_{ijkl} = \left\langle \varphi_i(\vec{r}_1)\varphi_j(\vec{r}_2)\left|\frac{1}{r_{12}}\right|\varphi_k(\vec{r}_1)\varphi_l(\vec{r}_2)\right\rangle.$$

In particular, we need the Coulomb and exchange integrals

$$J \equiv V_{ABAB} = \left\langle \varphi_A(\vec{r}_1)\varphi_B(\vec{r}_2)\left|\frac{1}{r_{12}}\right|\varphi_A(\vec{r}_1)\varphi_B(\vec{r}_2)\right\rangle, \quad K \equiv V_{ABBA} = \left\langle \varphi_A(\vec{r}_1)\varphi_B(\vec{r}_2)\left|\frac{1}{r_{12}}\right|\varphi_B(\vec{r}_1)\varphi_A(\vec{r}_2)\right\rangle.$$

Using the same 1*s* states as above, we get

$$J = \frac{\zeta_S}{x} - \frac{\zeta_S}{24x}\left\{24 + 33x + 18x^2 + 4x^3\right\}e^{-2x}, K = \frac{\zeta_S}{120}(75 - 138x - 72x^2 - 8x^3)e^{-2x} +$$
$$\frac{2\zeta_S}{15x}\left\{e^{2x}(3 - 3x + x^2)^2\,\mathrm{Ei}(-4x) - 2(9 - 3x^2 + x^4)\mathrm{Ei}(-2x) + e^{-2x}(3 + 3x + x^2)^2(\gamma_E + \ln x)\right\},$$

where $\gamma_E = 0.57721566....$ is Euler's constant, $x = \zeta_S R$ and Ei denotes the exponential integral. It turns out that an excellent approximation to K is

$$K \approx \zeta_S \frac{5}{8} \exp(0.285x - 0.763x^{3/2} + 0.0805x^2).$$

An accurate approximation for the ground state is the Heitler-London variational ansatz designed to avoid double occupancy on either atom

$$\psi(\vec{r}_1, \vec{r}_2) = \frac{1}{\sqrt{2(1 \pm s^2)}} \{\varphi_A(\vec{r}_1)\varphi_B(\vec{r}_2) \pm \varphi_B(\vec{r}_1)\varphi_A(\vec{r}_2)\}.$$

(a) Show that plugging in leads to the energy estimate

$$E_\pm = \frac{2\alpha \mp 2s\gamma + J \pm K}{1 \pm s^2}.$$

(b) Use the 1*s* hydrogen orbitals and vary the exponent to minimize the energy. The result should look like Fig. 42.6. The equilibrium distance is seen to be $R \approx 1.39$ Bohr with a binding energy of $-4.05\,\text{eV}$.

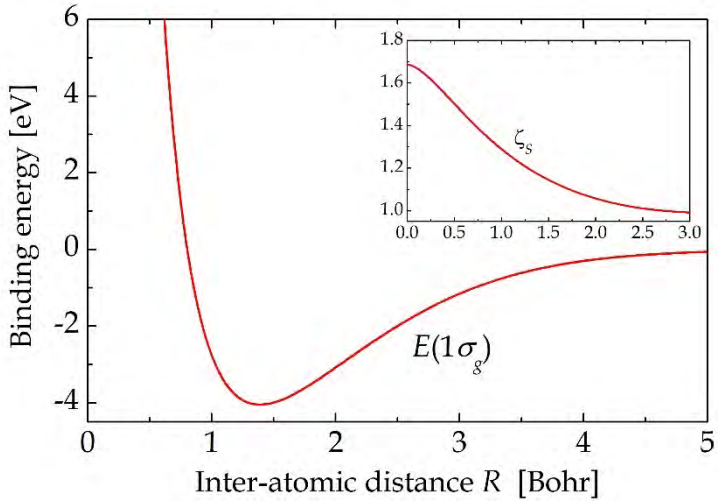


Figure 42.6. Variational ground state energy of H_2 with optimal exponent in the inset.

References

[1] D.M. Bishop and L.M. Cheung, J. Phys. B: At. Mol. Phys. 11, 3133 (1978).
[2] M. Preuss and F. Bechstedt, Phys. Rev. B73, 155413 (2006).

43. Optical Properties of Two-Level Systems

In this chapter, we study a very simple atomic system comprised of only two quantum levels. The simplicity of the system makes it possible to study highly non-perturbative effects of an optical field. The system is described by the Hamiltonian

$$H = H_0 + e\mathcal{E}x\cos(\omega t)\,.$$

We assume that the two levels $|1\rangle$ and $|2\rangle$ have definite and opposite parity. Moreover, taking these states to be real-valued, $x_{12} = \langle 1|x|2\rangle = x_{21}$ is the only non-vanishing dipole matrix element. It follows that Eq.(1.4) yields

$$\dot{a}_1 = -ia_2\Omega\cos(\omega t)e^{-i\omega_{21}t}\,,\quad \dot{a}_2 = -ia_1\Omega\cos(\omega t)e^{i\omega_{21}t}\,, \tag{43.1}$$

where $\Omega = e\mathcal{E}x_{12}/\hbar$ is the Rabi frequency. Below we will compute the population of the excited state $|2\rangle$ and the induced dipole moment as a function of applied field strength. Both approximate analytical and purely numerical methods will be applied.

43.1 Rotating Wave Approximation

Despite the innocent appearance, the set of coupled equations Eq.(43.1) cannot be solved in closed form. To provide an approximate solutions, we split the cosine according to $\cos(\omega t) = \frac{1}{2}(e^{i\omega t} + e^{-i\omega t})$. We then notice that these have to be combined with $e^{\pm i\omega_{21}t}$. Near resonance $\omega \sim \omega_{21}$ we will find very rapidly varying terms $e^{\pm i(\omega+\omega_{21})t}$ and slowly varying terms $e^{\pm i\Delta t}$, where $\Delta = \omega - \omega_{21}$ is the detuning. It may be assumed that the effect of the rapidly varying terms averages to nearly zero. Hence, we only keep the slow terms, which leads to the rotating wave approximation (RWA):

$$\dot{a}_1 = -\tfrac{i}{2}a_2\Omega e^{i\Delta t}\,,\quad \dot{a}_2 = -\tfrac{i}{2}a_1\Omega e^{-i\Delta t}\,. \tag{43.2}$$

These coupled equations are easily solved if combined so that

$$\ddot{a}_2 + i\Delta\dot{a}_2 + \tfrac{1}{4}\Omega^2 a_2 = 0\,.$$

Introducing the effective Rabi frequency $\tilde{\Omega} = (\Omega^2 + \Delta^2)^{1/2}$ we find the general solution

$$a_2(t) = e^{-\frac{i}{2}\Delta t}\left(A_+ e^{\frac{i}{2}\tilde{\Omega}t} + A_- e^{-\frac{i}{2}\tilde{\Omega}t}\right).$$

We will assume as the initial condition that the system at time $t=0$ is in the ground state. Thus, $a_2(0)=0$ and we may write $a_2(t)=Ae^{-\frac{i}{2}\Delta t}\sin(\frac{1}{2}\tilde{\Omega}t)$. To determine A, we need a_1 and from Eq.(43.2) we have $a_1(t)=A/\Omega e^{\frac{i}{2}\Delta t}\left[\Delta\sin(\frac{1}{2}\tilde{\Omega}t)+i\tilde{\Omega}\cos(\frac{1}{2}\tilde{\Omega}t)\right]$. We then readily find $|a_1|^2+|a_2|^2=A^2\tilde{\Omega}^2/\Omega^2$ so that we may take $A=\Omega/\tilde{\Omega}$. Eventually, the full solution reads

$$\begin{aligned} a_1(t) &= e^{\frac{i}{2}\Delta t}\left[\Delta\sin(\tfrac{1}{2}\tilde{\Omega}t)+i\tilde{\Omega}\cos(\tfrac{1}{2}\tilde{\Omega}t)\right]/\tilde{\Omega} \\ a_2(t) &= (\Omega/\tilde{\Omega})e^{-\frac{i}{2}\Delta t}\sin(\tfrac{1}{2}\tilde{\Omega}t). \end{aligned} \tag{43.3}$$

This is the famous RWA solution to the Rabi problem. It is seen that the probability of finding the system in the excited state is $|a_2(t)|^2=(\Omega/\tilde{\Omega})^2\sin^2(\frac{1}{2}\tilde{\Omega}t)$. Hence, this probability oscillates with the effective Rabi frequency and can get close to one if the field is sufficiently strong that $\Omega/\tilde{\Omega}\approx 1$. In Fig. 43.1, we compare the approximate solution to the numerical result obtained by directly solving Eq.(43.1) numerically. It is seen that the RWA misses the finer oscillations but captures the slower features. In the example, the effective Rabi frequency is $\tilde{\Omega}=(0.3^2+0.1^2)^{1/2}\omega_{21}\approx 0.316\omega_{21}$ and the slow period is $2\pi/\tilde{\Omega}\approx 19.7/\omega_{21}$ in agreement with the plot.

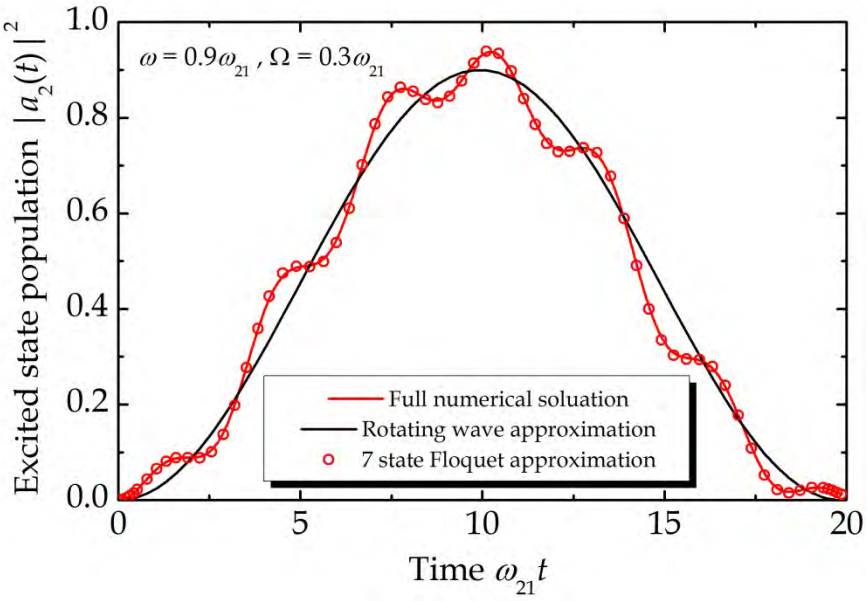


Fig. 43.1. Comparison of the full numerical solution to the rotating wave and Floquet approximations.

43.2 Floquet Solution

If the RWA is abandoned, a numerical solution can be used instead. Obviously, the coupled equations Eq.(43.1) can simply be solved numerically as shown in Fig. 43.1. Often, however, it is advantageous to find a semi-analytical solution. To this end, the method adopted here is based on so-called Floquet states. The name derives from the

Floquet theorem, which is very similar to the Bloch theorem from solid state physics except that time replaces space. For crystals that are periodic in the x direction, the Bloch theorem says that states are of the form $e^{ikx}u(x)$ with $u(x)$ lattice-periodic. Similarly, if a Hamiltonian is periodic in time such as in the presence of monochromatic light of frequency ω, the time dependence of the states is of the form $e^{i\lambda t}u(t)$ with $u(t)$ periodic with period $T=2\pi/\omega$. Hence, we write completely general solutions in the form

$$a_1(t)=e^{i(\lambda+\omega_1)t}\sum_n A_n^{(1)}e^{-in\omega t},\quad a_2(t)=e^{i(\lambda+\omega_2)t}\sum_n A_n^{(2)}e^{-in\omega t}.$$

When inserted into Eq.(43.1), we find

$$(n\omega-\lambda-\omega_1)A_n^{(1)}=\tfrac{1}{2}\Omega(A_{n-1}^{(2)}+A_{n+1}^{(2)}),\quad (n\omega-\lambda-\omega_2)A_n^{(2)}=\tfrac{1}{2}\Omega(A_{n-1}^{(1)}+A_{n+1}^{(1)}).$$

This, of course, is an infinite set of coupled equations. To be of practical use, it must be truncated to a reasonable size. In the limit of vanishing field, we have $A_0^{(1)}=1$ so we better keep $A_0^{(1)}$ in the set. From the coupled equations, we see that $A_0^{(1)}$ couples to $A_{\pm1}^{(2)}$. Keeping only $A_1^{(2)}$ reproduces the rotating wave solution as we will now demonstrate. In matrix form, we have in this case

$$\begin{pmatrix}-\omega_1 & -\frac{1}{2}\Omega\\ -\frac{1}{2}\Omega & \omega-\omega_2\end{pmatrix}\cdot\begin{pmatrix}A_0^{(1)}\\ A_1^{(2)}\end{pmatrix}=\lambda\begin{pmatrix}A_0^{(1)}\\ A_1^{(2)}\end{pmatrix}.$$

Hence, λ is an eigenvalue and the unknown coefficients are the corresponding eigenvectors. Solving, it is seen that $\lambda_\pm=\frac{1}{2}(\Delta\pm\tilde{\Omega})-\omega_1$ are the eigenvalues (see red lines in Fig. 43.2). The corresponding eigenvectors are

$$\begin{pmatrix}A_0^{(1)}\\ A_1^{(2)}\end{pmatrix}_\pm=N_\pm\begin{pmatrix}\Delta\mp\tilde{\Omega}\\ \Omega\end{pmatrix}.$$

The final, general solution is a linear combination of the eigenvectors, i.e.

$$\begin{aligned}a_1(t)&=e^{i(\lambda_++\omega_1)t}N_+(\Delta-\tilde{\Omega})+e^{i(\lambda_-+\omega_1)t}N_-(\Delta+\tilde{\Omega})\\ a_2(t)&=e^{i(\lambda_++\omega_2-\omega)t}N_+\Omega+e^{i(\lambda_-+\omega_2-\omega)t}N_-\Omega.\end{aligned}$$

After a little rewriting it follows that

$$\begin{aligned}a_1(t)&=e^{\frac{i}{2}\Delta t}\left[\Delta\left(N_+e^{\frac{i}{2}\tilde{\Omega}t}+N_-e^{-\frac{i}{2}\tilde{\Omega}t}\right)+\tilde{\Omega}\left(-N_+e^{\frac{i}{2}\tilde{\Omega}t}+N_-e^{-\frac{i}{2}\tilde{\Omega}t}\right)\right]\\ a_2(t)&=e^{-\frac{i}{2}\Delta t}\Omega\left(N_+e^{\frac{i}{2}\tilde{\Omega}t}+N_-e^{-\frac{i}{2}\tilde{\Omega}t}\right).\end{aligned}$$

Again, the initial conditions mean that $N_- = -N_+ \equiv iN$ and so

$$a_1(t) = 2Ne^{\frac{i}{2}\Delta t}\left[\Delta \sin(\tfrac{1}{2}\tilde{\Omega}t) + i\tilde{\Omega}\cos(\tfrac{1}{2}\tilde{\Omega}t)\right], \quad a_2(t) = 2Ne^{-\frac{i}{2}\Delta t}\Omega \sin(\tfrac{1}{2}\tilde{\Omega}t).$$

Requiring normalization easily proves that $2N = \tilde{\Omega}^{-1}$ and, consequently, the obtained solution is identical to Eq.(43.3), i.e. the rotating wave result. We may now include more Floquet states. As an example, 7 states lead to a Floquet equation

$$\begin{pmatrix} -3\omega-\omega_2 & -\frac{1}{2}\Omega & 0 & 0 & 0 & 0 & 0 \\ -\frac{1}{2}\Omega & -2\omega-\omega_1 & -\frac{1}{2}\Omega & 0 & 0 & 0 & 0 \\ 0 & -\frac{1}{2}\Omega & -\omega-\omega_2 & -\frac{1}{2}\Omega & 0 & 0 & 0 \\ 0 & 0 & -\frac{1}{2}\Omega & -\omega_1 & -\frac{1}{2}\Omega & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{2}\Omega & \omega-\omega_2 & -\frac{1}{2}\Omega & 0 \\ 0 & 0 & 0 & 0 & -\frac{1}{2}\Omega & 2\omega-\omega_1 & -\frac{1}{2}\Omega \\ 0 & 0 & 0 & 0 & 0 & -\frac{1}{2}\Omega & 3\omega-\omega_2 \end{pmatrix} \cdot \begin{pmatrix} A_{-3}^{(2)} \\ A_{-2}^{(1)} \\ A_{-1}^{(2)} \\ A_0^{(1)} \\ A_1^{(2)} \\ A_2^{(1)} \\ A_3^{(2)} \end{pmatrix} = \lambda \begin{pmatrix} A_{-3}^{(2)} \\ A_{-2}^{(1)} \\ A_{-1}^{(2)} \\ A_0^{(1)} \\ A_1^{(2)} \\ A_2^{(1)} \\ A_3^{(2)} \end{pmatrix}.$$

It would appear that 7 eigenstates are found. However, only two should fall within the "Brillouin zone", i.e. the range $[-\omega, \omega]$. This is illustrated in Fig. 43.2, where a large basis of 40 states has been used. The repeating pattern in the figure follows from the fact that if λ is an eigenvalue of the full (un-truncated) set of equations, then so is $\lambda \pm 2\omega$ and higher multiples. If, however, the equations are truncated this is only approximately true as illustrated for the 2 state approximation.

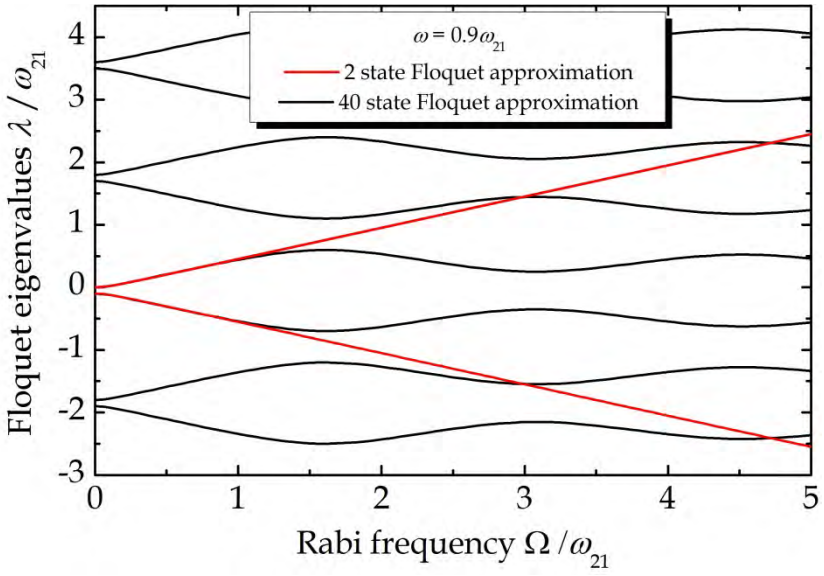


Fig. 43.2. Numerically computed eigenvalues for a small and a large Floquet basis.

In Fig. 43.1, the result for the population of the excited state calculated using a 7 state Floquet approximation is shown. It is clear that it is practically identical to the direct solution of the coupled differential equations. Thus, the accuracy of the Floquet approach is demonstrated.

43.3 Density Matrix Formalism

So far, we have assumed perfectly lossless conditions. This is always a requirement when a direct solution of the Schrödinger equation is attempted because losses would lead to a non-Hermitian operator. In reality, losses are always present. For an atomic system these will originate from coupling to the environment including the possibility that an excited atom decays spontaneously by emitting a photon. To include such mechanisms, we will introduce the density matrix formalism, which allows for losses in a straightforward manner. Writing $c_i = a_i e^{-i\omega_i t}$ the elements of the density matrix are defined by $\rho_{ij} = c_i c_j^*$. The time-dependent Schrödinger equation states that

$$\dot{c}_i = \frac{1}{i\hbar}\sum_k H_{ik} c_k, \quad \dot{c}_j^* = -\frac{1}{i\hbar}\sum_k H_{kj} c_k^* .$$

Combining these, it follows that

$$\dot{\rho}_{ij} = \dot{c}_i c_j^* + c_i \dot{c}_j^* = \frac{1}{i\hbar}\sum_k \left(H_{ik}\rho_{kj} - \rho_{ik} H_{kj} \right) = \frac{1}{i\hbar}[H,\rho]_{ij} .$$

In general then, the density matrix obeys the so-called Liouville equation $\dot{\rho} = (i\hbar)^{-1}[H,\rho]$. Solving this equation provides precisely the same information as solving the Schrödinger equation. However, we can now add terms that model losses. We want to do this in a manner, which preserves normalization, i.e. for a two-level system $\rho_{11} + \rho_{22} = 1$. For this simple system, it can be shown [1] that the form

$$\dot{\rho} = \frac{1}{i\hbar}[H,\rho] - \gamma \begin{pmatrix} -2\rho_{22} & \rho_{12} \\ \rho_{21} & 2\rho_{22} \end{pmatrix}, \quad \rho = \begin{pmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{pmatrix}$$

does precisely this. For the Hamiltonian in Eq.(43.1), the dynamical equation becomes

$$\dot{\rho} = \begin{pmatrix} i\Omega(\rho_{12}-\rho_{21})\cos(\omega t) + 2\gamma\rho_{22} & i\Omega(\rho_{11}-\rho_{22})\cos(\omega t) + i\omega_{21}\rho_{12} - \gamma\rho_{12} \\ -i\Omega(\rho_{11}-\rho_{22})\cos(\omega t) - i\omega_{21}\rho_{21} - \gamma\rho_{21} & -i\Omega(\rho_{12}-\rho_{21})\cos(\omega t) - 2\gamma\rho_{22} \end{pmatrix}. \quad (43.4)$$

If the field is suddenly switched off, the density matrix will decay to its equilibrium form with $\rho_{11} = 1$ and $\rho_{22} = \rho_{12} = \rho_{21} = 0$. This set of four coupled equations is even more difficult to handle than the original Schrödinger equation, but is more realistic because losses are included. In Fig. 43.3, we show a few traces for the population of the excited state ρ_{22} with different amounts of loss. We note that the initial large oscillations die out and eventually the population reaches a quasi-stationary level with smaller scale oscillation superimposed. The dashed curves, which roughly capture the slow dynamics and the steady state level, are derived from the RWA

version of the Liouville equation. This simplified equation is obtained by writing the density matrix elements as $\rho_{ij} = \tilde{\rho}_{ij} e^{i\omega_{ji}t}$, where $\tilde{\rho}_{ij} = a_i a_j^*$ is expected to be slowly varying. By introducing this form in Eq.(43.4) and omitting rapidly varying terms similarly to the RWA for the wave function, we find

$$\dot{\tilde{\rho}} = \begin{pmatrix} \frac{i}{2}\Omega(\tilde{\rho}_{12}e^{-i\Delta t} - \tilde{\rho}_{21}e^{i\Delta t}) + 2\gamma\tilde{\rho}_{22} & \frac{i}{2}\Omega(\tilde{\rho}_{11} - \tilde{\rho}_{22})e^{i\Delta t} - \gamma\tilde{\rho}_{12} \\ -\frac{i}{2}\Omega(\tilde{\rho}_{11} - \tilde{\rho}_{22})e^{-i\Delta t} - \gamma\tilde{\rho}_{21} & -\frac{i}{2}\Omega(\tilde{\rho}_{12}e^{-i\Delta t} - \tilde{\rho}_{21}e^{i\Delta t}) - 2\gamma\tilde{\rho}_{22} \end{pmatrix}. \quad (43.5)$$

It is readily demonstrated that the steady-state solution is of the form $\tilde{\rho}_{ii} = \tilde{\rho}_{ii}^{(0)}$ as well as $\tilde{\rho}_{12} = \tilde{\rho}_{12}^{(0)} e^{i\Delta t}$ and $\tilde{\rho}_{21} = \tilde{\rho}_{12}^*$ with the superscript "(0)" indicating a constant amplitude. Plugging in and solving yields

$$\tilde{\rho}_{12}^{(0)} = \frac{\frac{1}{2}(\Delta + i\gamma)\Omega}{\Delta^2 + \gamma^2 + \frac{1}{2}\Omega^2}, \quad \tilde{\rho}_{22}^{(0)} = \frac{\frac{1}{4}\Omega^2}{\Delta^2 + \gamma^2 + \frac{1}{2}\Omega^2} \quad (43.6)$$

in addition to $\tilde{\rho}_{21}^{(0)} = (\tilde{\rho}_{12}^{(0)})^*$ and $\tilde{\rho}_{11}^{(0)} = 1 - \tilde{\rho}_{22}^{(0)}$. We see that $\tilde{\rho}_{22}^{(0)} \approx \tilde{\rho}_{11}^{(0)} \approx 1/2$ if $\Omega \gg \Delta, \gamma$ in agreement with Fig. 43.3.

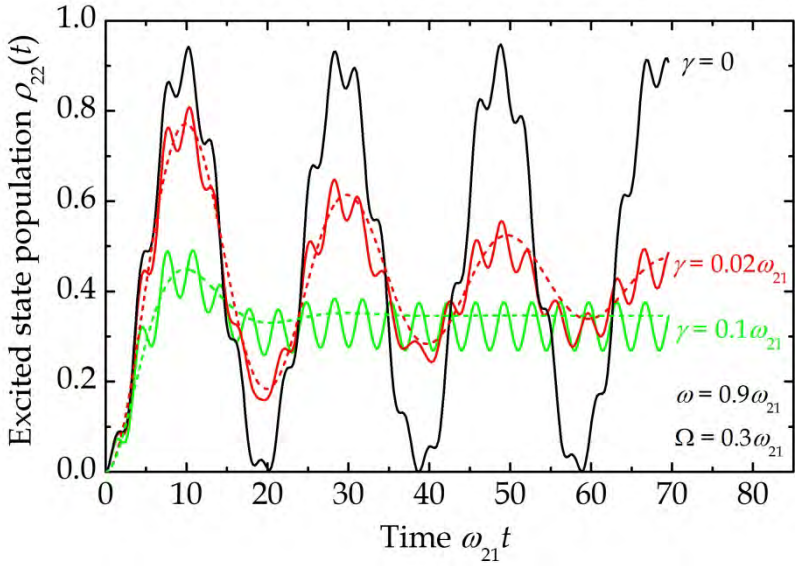


Fig. 43.3. Population of the upper level as predicted by the Liouville equation including losses. The dashed curves are the RWA results.

43.4 Induced Dipole Moment

Having determined the density matrix from the solution we can now compute measurable quantities. The obvious one is the induced dipole moment $d_x = -e\langle x \rangle$. Using Eq.(43.6) it is readily established that

$$d_x = -e(x_{21}\rho_{12} + x_{12}\rho_{21}) = -ex_{12}\frac{\frac{1}{2}(\Delta - i\gamma)\Omega}{\Delta^2 + \gamma^2 + \frac{1}{2}\Omega^2}e^{-i\omega t} + c.c.$$

If we use the definition of the Rabi frequency $\Omega = e\mathcal{E}x_{12}/\hbar$ and rewrite in the form $d_x = \frac{1}{2}\alpha(\omega)\mathcal{E}e^{-i\omega t} + c.c.$ we find a generalized (field dependent) polarizability $\alpha(\omega)$ of

$$\alpha(\omega) = -\frac{e^2 |x_{12}|^2}{\hbar}\frac{\Delta - i\gamma}{\Delta^2 + \gamma^2 + \frac{1}{2}\Omega^2}.$$

In the weak field limit $\Omega \to 0$, this result agrees with the resonant part of the perturbation result of Chapters 2 and 30

$$\alpha_{pert}(\omega) = \frac{e^2 |x_{12}|^2}{\hbar\omega_{21} - \hbar(\omega + i\gamma)}.$$

In Fig. 43.4, we plot the imaginary part of the generalized polarizability normalized by the amplitude of the perturbative result $\alpha_0 = e^2 |x_{12}|^2 / \hbar\gamma$. Clearly a large Rabi frequency saturates the absorption. Physically, this is consequence of the fact that atoms, which are already excited, cannot absorb further. Hence, absorption saturates at high field.

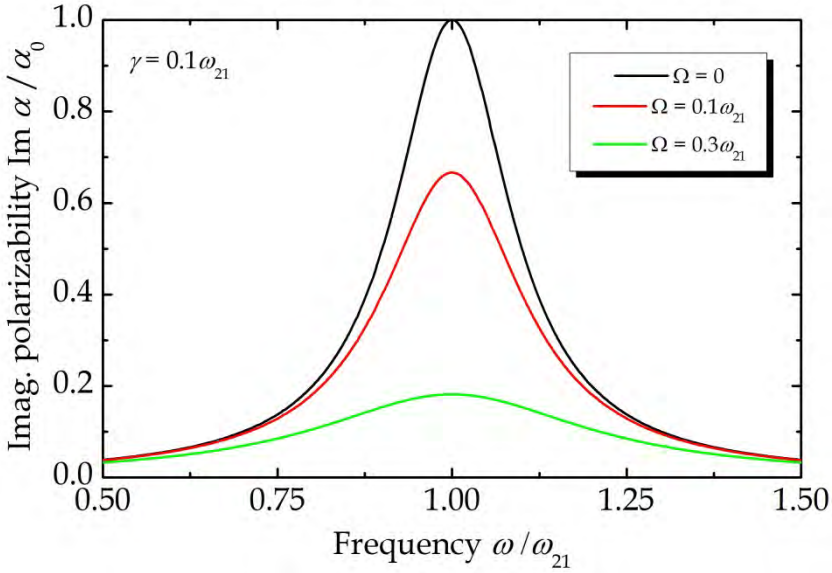


Fig. 43.4. The absorptive part of the generalized polarizability for different values of the Rabi frequency. Here, the RWA is applied.

The saturable absorption in Fig. 43.4 is an example of a nonlinearity, i.e. an effect of the finite field strength. However, as the field increases, the RWA becomes increasingly inaccurate. Hence, the full Liouville equation Eq.(43.4) rather than the RWA version must be considered. To obtain the full polarizability, one must (1) solve Eq.(43.4) numerically from $t = 0$ to some final $t \gg \gamma^{-1}$, (2) average the steady-state

time trace of $\rho_{12}e^{-i\omega t}$ to find the component oscillating at the applied frequency, and (3) compute $d=-2e\,\mathrm{Re}\{x_{21}\langle\rho_{12}e^{-i\omega t}\rangle\}$, where the angular brackets indicate time average. The result of this procedure is illustrated in Fig. 43.5. We observe that in addition to the saturation, a shift of the resonance appears. This so-called Bloch-Siegert shift [2] is an additional nonlinearity. It is shown below that to lowest order the shift is $\Omega^2/4\omega_{21}$.

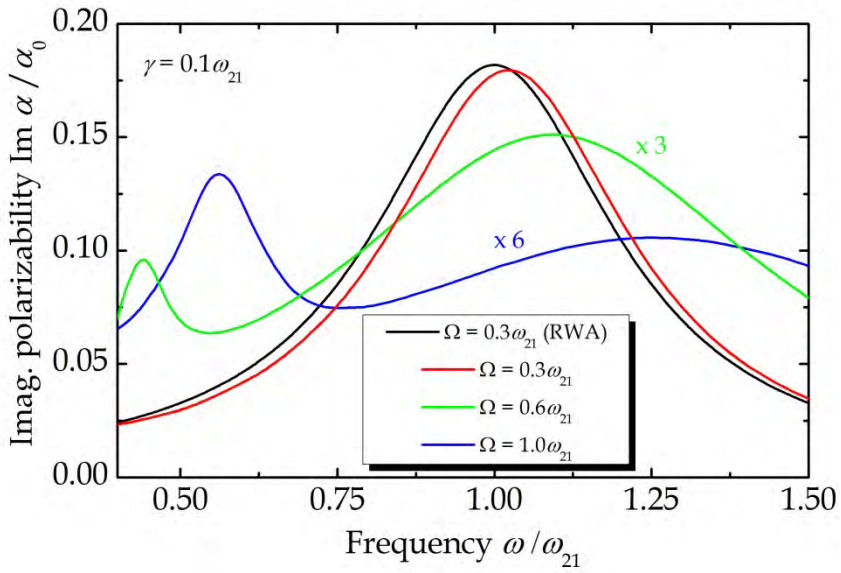


Figure 43.5. Imaginary part of the full polarizability obtained without adopting the RWA. Note that some curves have been magnified.

It is possible to perform a Floquet-type analysis of the density matrix as well. At a first glance, this looks difficult because we deal with four unknowns. To facilitate the analysis, however, one should remember that $\rho_{11}+\rho_{22}=1$ and, moreover, introduce the new variables $d=\rho_{11}-\rho_{22}$, $f=\rho_{12}-\rho_{21}$, and $g=\rho_{12}+\rho_{21}$. In terms of these, the dynamical equations can be written

$$\dot{d}=2i\Omega f\cos(\omega t)-2\gamma d+2\gamma,\quad \dot{f}=2i\Omega d\cos(\omega t)+i\omega_{21}g-\gamma f,\quad \dot{g}=i\omega_{21}f-\gamma g. \quad (43.7)$$

To solve, we expand in Fourier series $d(t)=\sum_n d_n e^{-in\omega t}$ and similarly for f and g. The last of the equations above means that $g_n=-\omega_{21}f_n/(n\omega+i\gamma)$ and with the abbreviation $D_n=n\omega+i\gamma-\omega_{21}^2/(n\omega+i\gamma)$ we therefore find

$$D_n f_n=-\Omega(d_{n-1}+d_{n+1}),\quad (n\omega+2i\gamma)d_n=-\Omega(f_{n-1}+f_{n+1})+2i\gamma\delta_{n0}.$$

In general, $f_{-n}=-f_n^*$ and $g_{-n}=g_n^*$. The coupled equations lead to a matrix problem very similar to the one for the wave function found above. If we truncate to the set $\{d_{-2},f_{-1},d_0,f_1,d_2\}$ we find a matrix equation

$$\begin{pmatrix} -2\omega+2i\gamma & \Omega & 0 & 0 & 0 \\ \Omega & D_{-1} & \Omega & 0 & 0 \\ 0 & \Omega & 2i\gamma & \Omega & 0 \\ 0 & 0 & \Omega & D_1 & \Omega \\ 0 & 0 & 0 & \Omega & 2\omega+2i\gamma \end{pmatrix} \cdot \begin{pmatrix} d_{-2} \\ f_{-1} \\ d_0 \\ f_1 \\ d_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 2i\gamma \\ 0 \\ 0 \end{pmatrix}.$$

The normalized polarizability is $\alpha / \alpha_0 = g_1 \gamma / \Omega$ and after some manipulations

$$\alpha(\omega) / \alpha_0 = \frac{2\omega_{21}\gamma((\omega+i\gamma)^2 - \omega_{21}^2 - \frac{1}{2}\Omega^2)}{(\omega^2-\omega_{21}^2)^2 + 2\omega_{21}^2\Omega^2 + 2\gamma^2(\omega_{21}^2+\omega^2+\Omega^2) + \frac{3}{4}\Omega^4 + \gamma^4}.$$

Differentiating the imaginary part to find the maximum, one then finds $\omega_{res} \approx \omega_{21} + \Omega^2 / 4\omega_{21}$ in agreement with Fig. 43.5. Similarly, the excited state population is given by

$$\rho_{22} = \frac{\frac{1}{2}(\omega_{21}^2+\omega^2+\gamma^2)\Omega^2 + \frac{1}{4}\Omega^4}{(\omega^2-\omega_{21}^2)^2 + 2\omega_{21}^2\Omega^2 + 2\gamma^2(\omega_{21}^2+\omega^2+\Omega^2) + \frac{3}{4}\Omega^4 + \gamma^4}.$$

The resonance frequency of ρ_{22} is the same as for the polarizability.

Exercise: Two-level system with permanent dipole moments.

If the system does not have inversion symmetry, both ground and excited state will generally possess non-vanishing dipole moments $x_{11} = \langle 1|x|1\rangle$ and $x_{22} = \langle 2|x|2\rangle$. If we denote the corresponding Rabi frequencies by Ω_{11} and Ω_{22}, we find

$$H = \hbar \begin{pmatrix} \omega_1 + \Omega_{11}\cos(\omega t) & \Omega\cos(\omega t) \\ \Omega\cos(\omega t) & \omega_2 + \Omega_{22}\cos(\omega t) \end{pmatrix}.$$

a) Show using the Liouville equation that the new dynamical equation for ρ_{12} is

$$\dot{\rho}_{12} = i\Omega(\rho_{11}-\rho_{22})\cos(\omega t) + i[\omega_{21} + \delta\Omega\cos(\omega t)]\rho_{12} - \gamma\rho_{12}, \qquad (43.8)$$

where $\delta\Omega = \Omega_{22} - \Omega_{11}$. The diagonal elements are unchanged and $\dot{\rho}_{21} = \dot{\rho}_{12}^*$. The revised equations for *f* and *g* become

$$\dot{f} = 2i\Omega d\cos(\omega t) + i[\omega_{21} + \delta\Omega\cos(\omega t)]g - \gamma f, \quad \dot{g} = i[\omega_{21} + \delta\Omega\cos(\omega t)]f - \gamma g.$$

Hence, the Fourier decomposition yields

$$(n\omega + 2i\gamma)d_n = -\Omega(f_{n-1} + f_{n+1}) + 2i\gamma\delta_{n0}$$
$$(n\omega + i\gamma)g_n = -\omega_{21}f_n - \tfrac{1}{2}\delta\Omega(f_{n-1} + f_{n+1})$$
$$(n\omega + i\gamma)f_n = -\Omega(d_{n-1} + d_{n+1}) - \omega_{21}g_n - \tfrac{1}{2}\delta\Omega(g_{n-1} + g_{n+1}).$$

b) Show that the averaged dipole moment $d_x = -e\langle x\rangle$ is determined by

$$\langle x\rangle = \tfrac{1}{2}(x_{11} + x_{22}) + \tfrac{1}{2}(x_{11} - x_{22})d + x_{12}g\,.$$

The above set of coupled equations is very demanding. As a shortcut, we will apply perturbation theory to extract the solution to a given order in the applied field. We wish to compute the dipole moment up to second order in the field. We will denote quantities that are proportional to the p'th power of the field by a superscript (p). Thus, the only zero order quantity is $d_0^{(0)} = 1$. Accordingly, the permanent dipole moment at lowest order is $d_x^{(0)} = -ex_{11}$ as expected.

c) Show that, in the next order, we find the non-vanishing coefficients

$$f_{\pm 1}^{(1)} = \frac{-\Omega(\pm\omega + i\gamma)}{(\omega \pm i\gamma)^2 - \omega_{21}^2}, \quad g_{\pm 1}^{(1)} = \frac{\Omega\omega_{21}}{(\omega \pm i\gamma)^2 - \omega_{21}^2}\,.$$

The second order is trickier. After some algebra, it turns out that

$$f_{\pm 2}^{(2)} = -\frac{\delta\Omega\Omega(\pm 3\omega + 2i\gamma)\omega_{21}}{2[(2\omega \pm i\gamma)^2 - \omega_{21}^2][(\omega \pm i\gamma)^2 - \omega_{21}^2]}$$
$$g_{\pm 2}^{(2)} = \frac{\delta\Omega\Omega[\omega_{21}^2 - (\omega \pm i\gamma)(2\omega \pm i\gamma)]}{2[(2\omega \pm i\gamma)^2 - \omega_{21}^2][(\omega \pm i\gamma)^2 - \omega_{21}^2]}, \quad d_{\pm 2}^{(2)} = \frac{\Omega^2}{2[(\omega \pm i\gamma)^2 - \omega_{21}^2]}.$$

Eventually, the second order hyperpolarizability becomes

$$\alpha_2(\omega) = \frac{e^3}{\hbar^2}\frac{(x_{22} - x_{11})\,|\,x_{12}\,|^2\,(3\omega_{21}^2 + i\gamma(2\omega + i\gamma))}{[(2\omega + i\gamma)^2 - \omega_{21}^2][(\omega + i\gamma)^2 - \omega_{21}^2]}.$$

References

[1] R. Loudon, *The Quantum Theory of Light* (Oxford University Press, New York, 1991)
[2] F. Bloch and A. Siegert, Phys. Rev. 57, 522 (1940).

44. Nonlinear Response Theory

All chapters of this book, so far, have relied on linear perturbation theory. Thus, we have obtained the perturbed quantum states by expanding in a basis of unperturbed states and retaining only first-order corrections. This approach, however, can be extended to arbitrarily high order. In this manner, we can compute the *nonlinear* response, i.e. corrections that are or second or higher order in the perturbation. As a starting point we write the unknown wave function as

$$\psi = \sum_n a_n \varphi_n e^{-iE_n t/\hbar},$$

where φ_n and E_n are unperturbed eigenstates and energies, respectively. Importantly, a_n is a time-dependent coefficient that is a polynomial in the perturbation $a_n = \sum_p a_n^{(p)}$ so that $a_n^{(p)}$ is proportional to the p'th power of the perturbation. We have previously shown that

$$\frac{\partial a_m^{(p)}}{\partial t} = \frac{1}{2i\hbar}\sum_n a_n^{(p-1)} H_{mn}\left\{e^{i(\omega_{mn}-\omega)t} + e^{i(\omega_{mn}+\omega)t}\right\},$$

where $H_{mn} = \left\langle \varphi_m \middle| \hat{H}_1 \middle| \varphi_n \right\rangle$ and the perturbation is assumed Hermitian. We further restrict ourselves to low-temperatures such that $a_n^{(0)} = \delta_{n0}$. By repeated usage of this relation we find

$$a_m^{(1)} = -\frac{1}{2\hbar} H_{m0} \sum_{s=\pm 1} \frac{e^{i(\omega_{m0}-s\omega)t}}{\omega_{m0} - s\omega}. \tag{44.1}$$

$$a_m^{(2)} = \frac{1}{(2\hbar)^2}\sum_n H_{mn}H_{n0} \sum_{s,s'=\pm 1} \frac{e^{i(\omega_{m0}-(s+s')\omega)t}}{\left(\omega_{n0} - s\omega\right)\left(\omega_{m0} - (s+s')\omega\right)}. \tag{44.2}$$

$$a_m^{(3)} = -\frac{1}{(2\hbar)^3}\sum_{n,l} H_{mn}H_{nl}H_{l0} \sum_{s,s',s''=\pm 1} \frac{e^{i(\omega_{m0}-(s+s'+s'')\omega)t}}{\left(\omega_{l0} - s\omega\right)\left(\omega_{n0} - (s+s')\omega\right)\left(\omega_{m0} - (s+s'+s'')\omega\right)}. \tag{44.3}$$

The pattern is readily continued to higher order if needed. We how proceed to calculating the actual response.

$$X(t) = \left\langle \psi \middle| \hat{X} \middle| \psi \right\rangle = \sum_{m,n} a_n^* a_m X_{nm} e^{i\omega_{nm}t}.$$

In particular, we are interested in terms proportional to a specified power p of the perturbation. Hence,

$$X_p(t) = \sum_{m,n,q} a_n^{(q)*} a_m^{(p-q)} X_{nm} e^{i\omega_{nm}t}.$$

Resolving this response in frequency components, we furthermore find a term that oscillates at the maximal frequency $p\omega$. For this term we have

$$X_p(t) = \frac{1}{2^p} X_p(\omega) e^{-ip\omega t} + \frac{1}{2^p} X_p^*(\omega) e^{ip\omega t}.$$

The first few expressions are

$$X_1(\omega) = -\frac{1}{\hbar} \sum_m \left\{ \frac{X_{0m} H_{m0}}{\omega_{m0} - \omega} + \frac{X_{m0} H_{0m}}{\omega_{m0} + \omega} \right\},$$

$$X_2(\omega) = \frac{1}{\hbar^2} \sum_{m,n} \left\{ \frac{X_{0m} H_{mn} H_{n0}}{(\omega_{n0} - \omega)(\omega_{m0} - 2\omega)} + \frac{H_{0n} H_{nm} X_{m0}}{(\omega_{n0} + \omega)(\omega_{m0} + 2\omega)} + \frac{H_{0n} X_{nm} H_{m0}}{(\omega_{n0} + \omega)(\omega_{m0} - \omega)} \right\}.$$

We now specialize to a particular case: optical response. In this case, if the field is along x, $\hat{H}_1 = e\mathcal{E}\sum x_i$, where the sum is over all electrons. Moreover, the response is the induced dipole moment $\hat{X} = -e\sum x_i$. Hence, in this case $\hat{H}_1 = -\mathcal{E}\hat{X}$. We then introduce the generalized polarizability via $X_p(\omega) = (-1)^p \alpha_p(\omega) \mathcal{E}^p$. This demonstrates that

$$\alpha_1(\omega) = \frac{e^2}{\hbar} \sum_m x_{0m} x_{m0} \left\{ \frac{1}{\omega_{m0} - \omega} + \frac{1}{\omega_{m0} + \omega} \right\},$$

$$\alpha_2(\omega) = \frac{e^3}{\hbar^2} \sum_{m,n} x_{0m} x_{mn} x_{n0} \left\{ \frac{1}{(\omega_{n0} - \omega)(\omega_{m0} - 2\omega)} + \frac{1}{(\omega_{n0} + 2\omega)(\omega_{m0} + \omega)} + \frac{1}{(\omega_{n0} - \omega)(\omega_{m0} + \omega)} \right\}.$$

A general prescription is given by [1]

$$\alpha_p(\omega) = \frac{e^{p+1}}{\hbar^p} \sum_{m_1 \ldots m_p} \sum_{k=1}^{p+1} \frac{x_{0m_1} x_{m_1 m_2} \cdots x_{m_p 0}}{\left(\omega_{m_1 0} + s_{k1}\omega\right)\left(\omega_{m_2 0} + s_{k2}\omega\right) \cdots \left(\omega_{m_p 0} + s_{kp}\omega\right)} \qquad (44.4)$$

with $s_{k1} = 1 - (p+1)\delta_{k1}$ and $s_{k,j} = s_{k,j-1} + 1 - (p+1)\delta_{kj}$. Defined in this way, the static value of the induced dipole moment of an inversion-symmetric systems is $d = \alpha_1(0)\mathcal{E} + \alpha_3(0)\mathcal{E}^3 + \cdots$. Hence, upon integration it follows that the field-induced change in the ground-state energy is

$$\Delta E = -\int d(\mathcal{E})d\mathcal{E} = -\frac{1}{2}\alpha_1(0)\mathcal{E}^2 - \frac{1}{4}\alpha_3(0)\mathcal{E}^4 + \cdots$$

A note on the choice of origin of the x-axis: This choice should, of course, be immaterial to any measurable quantity. For matrix elements like x_{mn} with $m \neq n$, the choice is of no consequence because of orthogonality. However, for x_{mm} the choice matters and the *correct* choice is the one, in which the permanent molecular dipole moment vanishes.

As an example, we will again use hydrogen for which [2] $\alpha_1(0) = 9/2$ and $\alpha_3(0) = 3555/16$ in atomic units (see also exercise). We already computed the first polarizability in Chapter 30 and here we will study the third order response. As the ground state is s-type there are two contributions to the response

$$\begin{aligned}\alpha_3(\omega) = &\frac{e^4}{\hbar^3}\sum_{mnq}\sum_{k=1}^{4}\frac{\langle 1s|x|np\rangle\langle np|x|ms\rangle\langle ms|x|qp\rangle\langle qp|x|1s\rangle}{\left(\omega_{np-1s}+s_{k1}\omega\right)\left(\omega_{ms-1s}+s_{k2}\omega\right)\left(\omega_{qp-1s}+s_{k3}\omega\right)} \\ &+\frac{e^4}{\hbar^3}\sum_{mnq}\sum_{k=1}^{4}\frac{\langle 1s|x|np\rangle\langle np|x|md\rangle\langle md|x|qp\rangle\langle qp|x|1s\rangle}{\left(\omega_{np-1s}+s_{k1}\omega\right)\left(\omega_{md-1s}+s_{k2}\omega\right)\left(\omega_{qp-1s}+s_{k3}\omega\right)}.\end{aligned} \tag{44.5}$$

The dipole moments are then given by

$$\langle ms|x|np\rangle = \frac{1}{\sqrt{3}}\int_0^\infty R_{ms}(r)R_{np}(r)r^3dr, \quad \langle md|x|np\rangle = \frac{2}{\sqrt{15}}\int_0^\infty R_{md}(r)R_{np}(r)r^3dr.$$

We find the required states using the sine basis of App. 4. The final result is plotted below.

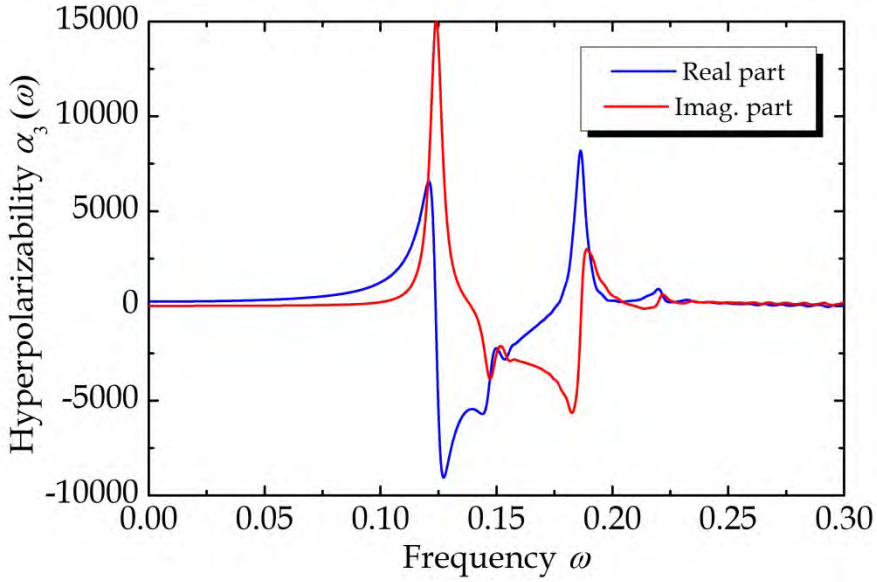


Figure 44.1. Third-order hyperpolarizability of a hydrogen atom.

The DC value is found to be approximately $\alpha_3(0) \approx 230$ is reasonable agreement with the precise value. Also, the main resonances are located at $3\omega = \omega_{2p-1s} = 3/8$ and $2\omega = 3/8$, as expected.

44.1 Generalizations

We now attempt to abandon some of the restrictions, under which the nonlinear response functions were derived. Primarily, these are (i) the low-temperature assumption and (ii) the many-body picture. In turn, we will also be able to handle extended, periodic structures. As a first step, abandoning the low-temperature assumption is easily carried out. It can in a straightforward manner be shown that

$$\alpha_p(\omega) = \frac{e^{p+1}}{\hbar^p} \sum_{m_0 \ldots m_p} \sum_{k=1}^{p+1} f_{m_0} \frac{x_{m_0 m_1} x_{m_1 m_2} \cdots x_{m_p m_0}}{\left(\omega_{m_1 m_0} + s_{k1}\omega\right)\left(\omega_{m_2 m_0} + s_{k2}\omega\right) \cdots \left(\omega_{m_p m_0} + s_{kp}\omega\right)}, \quad (44.6)$$

where f_{m_0} is the probability that the state $|m_0\rangle$ is occupied in the absence of perturbations. To ease the notation, we introduce

$$D_{m_0 m_1 \ldots m_p}(\omega) = \sum_{k=1}^{p+1} \frac{1}{\left(\omega_{m_1 m_0} + s_{k1}\omega\right)\left(\omega_{m_2 m_0} + s_{k2}\omega\right) \cdots \left(\omega_{m_p m_0} + s_{kp}\omega\right)},$$

so that

$$\alpha_p(\omega) = \frac{e^{p+1}}{\hbar^p} \sum_{m_0 \ldots m_p} f_{m_0} x_{m_0 m_1} x_{m_1 m_2} \cdots x_{m_p m_0} D_{m_0 m_1 \ldots m_p}(\omega).$$

Next, we can perform a series of cyclic permutations $P(m_0 m_1 \ldots m_{p-1} m_p) = m_1 \ldots m_{p-1} m_p m_0$ and so on. The product of matrix elements is invariant under such permutations. Hence, we can "average" over permutations

$$\alpha_p(\omega) = \frac{e^{p+1}}{(p+1)\hbar^p} \sum_{m_0 \ldots m_p} x_{m_0 m_1} x_{m_1 m_2} \cdots x_{m_p m_0} \left\{ f_{m_0} D_{m_0 m_1 \ldots m_p}(\omega) + f_{m_1} D_{m_1 \ldots m_p m_0}(\omega) + \ldots \right\}.$$

It can be shown that under the action of all such permutations

$$D_{m_0 m_1 \ldots m_p}(\omega) + D_{m_1 \ldots m_p m_0}(\omega) + \ldots = 0.$$

Thus, if the initial occupation is the same for all states the response vanishes. This result can be used to rewrite the general expression as

$$\alpha_p(\omega)=\frac{e^{p+1}}{(p+1)\hbar^p}\sum_{m_0\ldots m_p} x_{m_0m_1}x_{m_1m_2}\cdots x_{m_pm_0}\left\{\left(f_{m_1}-f_{m_0}\right)D_{m_1\ldots m_pm_0}(\omega)+\left(f_{m_2}-f_{m_0}\right)D_{m_2\ldots m_pm_0m_1}(\omega)+\ldots\right\}.$$

For instance, the first and second order results now read

$$\alpha_1(\omega)=\frac{e^2}{\hbar}\sum_{m,n} f_{nm}\,|\,x_{nm}\,|^2\,\frac{\omega_{mn}}{\omega_{mn}^2-\omega^2}\,, \tag{44.7}$$

and

$$\begin{aligned}\alpha_2(\omega)&=\frac{e^3}{\hbar^2}\sum_{m,n,l} x_{lm}x_{mn}x_{nl}\left\{f_{ml}D_{mnl}+f_{nl}D_{nlm}\right\},\\ D_{mnl}&=\frac{\omega\omega_{nl}+\omega_{nm}\omega_{lm}}{\left(\omega_{nm}-2\omega\right)\left(\omega_{nm}+\omega\right)\left(\omega_{lm}+2\omega\right)\left(\omega_{lm}-\omega\right)}\end{aligned} \tag{44.8}$$

After some further manipulations

$$\alpha_2(\omega)=\frac{e^3}{\hbar^2}\sum_{m,n,l}\frac{x_{lm}x_{mn}x_{nl}}{\omega_{ml}-2\omega}\left\{\frac{f_{ln}}{\omega_{nl}-\omega}+\frac{f_{mn}}{\omega_{mn}-\omega}\right\}. \tag{44.9}$$

In the limit $|\,\omega_{mn}\,|\ll\omega$, we can expand using $f_{nm}\approx -f'(\omega_n)\hbar\omega_{mn}-\frac{1}{2}f''(\omega_n)\hbar^2\omega_{mn}^2$ and find

$$\alpha_1(\omega)\approx\frac{e^2}{\omega^2}\sum_{m,n} f'(\omega_n)\,|\,x_{nm}\,|^2\,\omega_{mn}^2\,,$$

as well as

$$\alpha_2(\omega)\approx\frac{e^3}{8\omega^3}\sum_{m,n,l} f''(\omega_n)x_{lm}x_{mn}x_{nl}\omega_{lm}\omega_{mn}\omega_{nl}\,.$$

So far, all results have been expressed in terms of many-body states. If, now, electron-electron interaction is ignored, each many-body state is simply a Slater determinant. The ground state Slater determinant is built by all occupied states (v for "valence band") and in the excited ones one or several of these are replaced by empty states (labeled c for "conduction band"). We find for the first and second order responses (including spin summation)

$$\alpha_1(\omega)=\frac{4e^2}{\hbar}\sum_{v,c} f_{vc}\,|\,x_{vc}\,|^2\,\frac{\omega_{cv}}{\omega_{cv}^2-\omega^2},\quad \alpha_2(\omega)=-\frac{6e^3}{\hbar^2}\sum_{v,c,l} x_{vc}x_{cl}x_{lv}\left\{f_{lc}D_{clv}+f_{lv}D_{vcl}\right\}. \tag{44.10}$$

44.2. Gauge Invariance

At this point, we would like to discuss a rather technical issue: gauge invariance. So far, we have relied on the dipole approximation for the computation of optical response. This is perfectly fine as long as the field can be considered spatially constant. If, on the other hand, the field varies in space, we need to go back to the full Hamiltonian. The generalized result, however, should agree with the dipole approximation when specialized to a constant field. This requirement is known as gauge invariance (in response-theory context).

The dipole approximation framework goes by the name "Length Gauge" in contrast to the non-approximate alternative known as the "Velocity Gauge". In the former, the perturbation is simply $\hat{H}_{pert}(\vec{r},t)=e\vec{r}\cdot\vec{\mathcal{E}}(t)$, as should be well known by now. If the field varies in space, we return to the full Hamiltonian, in which the optical perturbation is contained in a time- and space-dependent vector potential $\vec{\mathcal{A}}(\vec{r},t)$ while keeping the scalar potential $V(\vec{r})$ static. The optical perturbation is then included by replacing the usual momentum $\hat{\vec{p}}$ by the canonical momentum $\hat{\vec{\Pi}}=\hat{\vec{p}}+e\vec{\mathcal{A}}(\vec{r},t)$. The Hamiltonian is therefore

$$\hat{H}=\frac{1}{2m}\left\{\hat{\vec{p}}+e\vec{\mathcal{A}}(\vec{r},t)\right\}^2+V(\vec{r}).$$

It follows that the perturbation is

$$\hat{H}_{pert}(\vec{r},t)=\frac{e}{2m}\left\{\hat{\vec{p}}\cdot\vec{\mathcal{A}}(\vec{r},t)+\vec{\mathcal{A}}(\vec{r},t)\cdot\hat{\vec{p}}\right\}+\frac{e^2}{2m}\mathcal{A}^2(\vec{r},t).$$

This clearly looks rather different from the dipole perturbation. The time-dependent fields are related via $\vec{\mathcal{E}}(\vec{r},t)=-\partial\vec{\mathcal{A}}(\vec{r},t)/\partial t$. We will now specialize to the space-independent case $\vec{\mathcal{A}}(\vec{r},t)\to\vec{\mathcal{A}}(t)$ in order to discuss gauge invariance. In this case,

$$\hat{H}_{pert}(\vec{r},t)=\frac{e}{m}\vec{\mathcal{A}}(t)\cdot\hat{\vec{p}}+\frac{e^2}{2m}\mathcal{A}^2(t).$$

The presence of the last, purely time-dependent term makes no complications since it simply adds a phase to the wave function, which becomes $\Psi(\vec{r},t)=\psi(\vec{r},t)\exp\{-i\frac{e^2}{2m\hbar}\int^t\mathcal{A}^2(t')dt'\}$ such that $\psi(\vec{r},t)$ is governed by the first term $\hat{H}_{pert}(\vec{r},t)=\frac{e}{m}\vec{\mathcal{A}}(t)\cdot\hat{\vec{p}}$. Now, if the electric field is of the form $\vec{\mathcal{E}}(t)=\frac{1}{2}\vec{\mathcal{E}}_0e^{-i\omega t}+c.c.$ it follows that the vector potential is $\vec{\mathcal{A}}(t)=\frac{1}{2}\vec{\mathcal{A}}_0e^{-i\omega t}+c.c.$ with $\vec{\mathcal{A}}_0=\vec{\mathcal{E}}_0/(i\omega)$. Hence, in the language of Chapter 1, the interaction Hamiltonian is $\hat{H}_1=\frac{e}{i\omega m}\vec{\mathcal{E}}_0\cdot\hat{\vec{p}}$.

Comparing to the dipole case, we see that this amounts to the substitution $\vec{r} \to \hat{\vec{p}} / (i\omega m)$.

In fact, the full implications of the velocity gauge are slightly more subtle. This has to do with the fact that we also need to apply the substitution $\hat{\vec{p}} \to \hat{\vec{p}} + e\vec{\mathcal{A}}(\vec{r},t)$ when we introduce observables. Hence, the current operator now becomes $\hat{\vec{j}} = -\frac{e}{m}\{\hat{\vec{p}} + e\vec{\mathcal{A}}(\vec{r},t)\}$ and the current is $\vec{\mathcal{J}} = \langle\psi|\hat{\vec{j}}|\psi\rangle$. With a space-independent field we see that $\vec{\mathcal{J}} = -\frac{e}{m}\langle\psi|\hat{\vec{p}}|\psi\rangle - \frac{e^2}{m}\langle\psi|\psi\rangle\vec{\mathcal{A}}(t)$. In the independent-electron picture, $\langle\psi|\psi\rangle = \sum_n f_n = N$ equals the number of electrons. Thus, the diamagnetic $\mathcal{A}(t)$ terms simply adds a $ie^2N/m\omega$ term to the linear conductivity without affecting any higher non-linearities. It follows that the current density operator can be taken as the usual $-\frac{e}{m}\hat{\vec{p}}$ if we just remember to add $ie^2N/m\omega$ to the linear response. Finally, whenever we want the polarization $\vec{\mathcal{P}}(\omega)$ at frequency ω, we find it from the current $\vec{\mathcal{J}}(\omega)$ via $\vec{\mathcal{P}}(\omega) = \frac{i}{\omega}\vec{\mathcal{J}}(\omega)$.

We can now put the pieces together. For simplicity, we will restrict ourselves to a single dimension so that both field and responses are in the *x*-direction. At first order, we get a term of the form $f_{nm}p_{mn}p_{nm}$. Hence, combining with the diamagnetic contribution above, we obtain the total conductivity $\sigma_1(\omega)$ and, in turn, the polarizability $\alpha_1(\omega) = i\sigma_1(\omega)/\omega$ is given by

$$\alpha_1(\omega) = -\frac{e^2N}{m\omega^2} + \frac{e^2}{\hbar m^2\omega^2}\sum_{m,n} f_{nm}\,|\,p_{nm}\,|^2\,\frac{\omega_{mn}}{\omega_{mn}^2 - \omega^2}. \tag{44.11}$$

The first term *looks* like a Drude term for a partly filled parabolic band, c.f. Chapters 3 and 8. However, it involves the *total* number of electrons (not just valence ones) and the effective mass is the free-electron one. Continuing, the velocity gauge response at second order becomes

$$\alpha_2(\omega) = \frac{ie^3}{2\hbar^2m^3\omega^3}\sum_{m,n,l}\frac{p_{lm}p_{mn}p_{nl}}{\omega_{ml} - 2\omega}\left\{\frac{f_{ln}}{\omega_{nl} - \omega} + \frac{f_{mn}}{\omega_{mn} - \omega}\right\}. \tag{44.12}$$

The task now consists in proving that these expressions are, in fact, identical to Eqs.(44.7) and (44.9). This would establish gauge invariance.

A key to these proofs is the commutator relation $p_{mn} = -im\omega_{mn}x_{mn}$. A second important relation is $(E-\omega)^{-1} = E\omega^{-1}(E-\omega)^{-1} - \omega^{-1}$. We rewrite Eq.(44.11) as

$$\alpha_1(\omega) = -\frac{e^2 N}{m\omega^2} + \frac{e^2}{\hbar m^2 \omega^2} \sum_{m,n} f_{nm} \,|\, p_{nm} \,|^2 \, \frac{1}{\omega_{mn}} \left\{ 1 + \frac{\omega^2}{\omega_{mn}^2 - \omega^2} \right\}$$
$$= \frac{e^2}{\hbar} \sum_{m,n} f_{nm} \,|\, x_{nm} \,|^2 \, \frac{\omega_{mn}}{\omega_{mn}^2 - \omega^2} - \frac{e^2 N}{m\omega^2} + \frac{e^2}{\hbar m^2 \omega^2} \sum_{m,n} f_{nm} \frac{|\, p_{nm} \,|^2}{\omega_{mn}}.$$

Here, the first term is clearly the dipole result. The last term is transformed using $|\, p_{nm} \,|^2 / \omega_{mn} = m[x_{nm} p_{mn} - p_{nm} x_{mn}] / 2i$ and splitting $f_{nm} = f_n - f_m$. Hence, performing the m summation for the f_n term and vice versa, we find using completeness and $[x, p]_{nn} = i\hbar$

$$\sum_{m,n} f_{nm} \frac{|\, p_{nm} \,|^2}{\omega_{mn}} = \frac{m}{i} \sum_n f_n [x, p]_{nn} = m\hbar N \,.$$

This establishes that the two extra terms above cancel. Note, however, that we needed to assume a complete set of states. If only a truncated set is used, gauge invariance is lost [3].

Gauge invariance at second order is trickier. Essentially, we need to "convert" $i p_{lm} p_{mn} p_{nl} / (2\omega^3 m^3)$ in the velocity gauge to $x_{lm} x_{mn} x_{nl}$ to prove equivalence with the length gauge. We will use the x's and p's to designate the different expressions. Hence, Eq.(44.9) and (44.12) are , respectively,

$$\alpha_2^{ppp}(\omega) = \frac{ie^3}{2\hbar^2 m^3 \omega^3} \sum_{m,n,l} \frac{p_{lm} p_{mn} p_{nl}}{\omega_{ml} - 2\omega} \left\{ \frac{f_{ln}}{\omega_{nl} - \omega} + \frac{f_{mn}}{\omega_{mn} - \omega} \right\}.$$

$$\alpha_2^{xxx}(\omega) = \frac{e^3}{\hbar^2} \sum_{m,n,l} \frac{x_{lm} x_{mn} x_{nl}}{\omega_{ml} - 2\omega} \left\{ \frac{f_{ln}}{\omega_{nl} - \omega} + \frac{f_{mn}}{\omega_{mn} - \omega} \right\}.$$

First, we use $(\omega_{ml} - 2\omega)^{-1} = 2\omega(\omega_{ml} - 2\omega)^{-1} / \omega_{ml} + 1 / \omega_{ml}$ in the *ppp* form. The extra bit from the last term vanishes, i.e.

$$\sum_{m,n,l} \frac{p_{lm} p_{mn} p_{nl}}{\omega_{ml}} \left\{ \frac{f_{ln}}{\omega_{nl} - \omega} + \frac{f_{mn}}{\omega_{mn} - \omega} \right\} = im \sum_{m,n,l} x_{lm} p_{mn} p_{nl} \left\{ \frac{f_{ln}}{\omega_{nl} - \omega} + \frac{f_{mn}}{\omega_{mn} - \omega} \right\}$$
$$= im \sum_{n,l} (xp)_{ln} p_{nl} \frac{f_{ln}}{\omega_{nl} - \omega} + im \sum_{m,n} p_{mn} (px)_{nm} \frac{f_{mn}}{\omega_{mn} - \omega}$$
$$= im \sum_{n,l} [x, p]_{ln} p_{nl} \frac{f_{ln}}{\omega_{nl} - \omega} = 0.$$

The last line follows because $[x,p]_{ln} = i\hbar\delta_{ln}$ and $f_{nn} = 0$. Hence, for starters $\alpha_2^{ppp}(\omega) = \alpha_2^{xpp}(\omega)$ with

$$\alpha_2^{xpp}(\omega) = \frac{-e^3}{\hbar^2 m^2 \omega^2} \sum_{m,n,l} \frac{x_{lm} p_{mn} p_{nl}}{\omega_{ml} - 2\omega} \left\{ \frac{f_{ln}}{\omega_{nl} - \omega} + \frac{f_{mn}}{\omega_{mn} - \omega} \right\}.$$

Similarly, applying $(\omega_{ml} - 2\omega)^{-1} = \omega_{ml}(\omega_{ml} - 2\omega)^{-1} / 2\omega - 1/2\omega$ to the *xxx* form, we find $\alpha_2^{xxx}(\omega) = \alpha_2^{pxx}(\omega)$ with

$$\alpha_2^{pxx}(\omega) = \frac{-ie^3}{2\hbar^2 m\omega} \sum_{m,n,l} \frac{p_{lm} x_{mn} x_{nl}}{\omega_{ml} - 2\omega} \left\{ \frac{f_{ln}}{\omega_{nl} - \omega} + \frac{f_{mn}}{\omega_{mn} - \omega} \right\}.$$

Using $\omega_{ml} = \omega_{mn} + \omega_{nl}$, the terms in braces found in all forms can be expanded according to

$$\frac{f_{ln}}{\omega_{nl} - \omega} + \frac{f_{mn}}{\omega_{mn} - \omega} = \frac{f_{ln}(\omega_{mn} - \omega) + f_{mn}(\omega_{nl} - \omega)}{\omega^2} - \frac{\omega_{nl}\omega_{mn} - \omega\omega_{ml}}{\omega^2} \left\{ \frac{f_{ln}}{\omega_{nl} - \omega} + \frac{f_{mn}}{\omega_{mn} - \omega} \right\}.$$

Now, a key identity is that

$$S \equiv \sum_{m,n,l} \frac{x_{lm} x_{mn} x_{nl}}{\omega_{ml} - 2\omega} \{ f_{ln}(\omega_{mn} - \omega) + f_{mn}(\omega_{nl} - \omega) \} = 0.$$

To prove this, one needs "time-reversal symmetry", which simply says that it is possible to choose phases such that $x_{ba} = x_{ab}$ (for finite systems one simply uses real-valued wave functions). Then, in a couple of steps,

$$S = \sum_{m,n,l} x_{lm} x_{mn} x_{nl} \left\{ f_{ln} + f_{ml} \frac{\omega_{nl} - \omega}{\omega_{ml} - 2\omega} \right\} = \sum_{m,n,l} x_{lm} x_{mn} x_{nl} \omega_{ml} f_{ml} \frac{\omega_{nl} - \omega_{mn}}{\omega_{ml}^2 - 4\omega^2}$$
$$= m^2 \sum_{m,n,l} p_{lm} f_{ml} \frac{x_{mn} p_{nl} - p_{mn} x_{nl}}{\omega_{ml}^2 - 4\omega^2} = m^2 \sum_{m,l} p_{lm} f_{ml} \frac{[x,p]_{ml}}{\omega_{ml}^2 - 4\omega^2} = 0.$$

Here, the f_{ln} contribution in the first form vanishes and time-reversal symmetry was applied in the second reformulation. Exploiting this identity and the expansion above to the *xxx* form leads directly to $\alpha_2^{xxx}(\omega) = -\alpha_2^{xpp}(\omega) + 2\alpha_2^{pxx}(\omega)$. Combining all these versions eventually demonstrates that $\alpha_2^{xxx}(\omega) = \alpha_2^{pxx}(\omega) = \alpha_2^{xpp}(\omega) = \alpha_2^{ppp}(\omega)$.

44.3. Gauge Invariance with Spin-Orbit Interaction

In the presence of spin-orbit interaction, the unperturbed Hamiltonian of a single electron is of the form $\hat{H} = \hat{H}_0 + \hat{H}_{SO}$ with

$$\hat{H}_{SO} = -\frac{e\hbar}{4m^2c^2}(\vec{\sigma}\times\vec{\mathcal{E}}_V)\cdot\hat{\vec{p}}\,.$$

Here, the field is the static one related to the potential via $e\vec{\mathcal{E}}_V = -\nabla V(\vec{r})$. The associated current density operator is $\hat{\vec{j}} = \frac{e}{i\hbar}[\hat{H},\vec{r}]$ and we should apply the minimal substitution $\hat{\vec{p}} \to \hat{\vec{p}} + e\vec{\mathcal{A}}(\vec{r},t)$. It is readily found that

$$\hat{\vec{j}} = -\tfrac{e}{m}\{\hat{\vec{p}} + e\vec{\mathcal{A}}(\vec{r},t)\} + \frac{e^2\hbar}{4m^2c^2}\vec{\sigma}\times\vec{\mathcal{E}}_V\,.$$

However, if the minimal substitution is applied directly to the Hamiltonian and we assume a space-independent field, we get the perturbation

$$\hat{H}_{pert}(\vec{r},t) = \frac{e}{m}\vec{\mathcal{A}}(t)\cdot\hat{\vec{p}} + \frac{e^2}{2m}\mathcal{A}^2(t) - \frac{e^2\hbar}{4m^2c^2}(\vec{\sigma}\times\vec{\mathcal{E}}_V)\cdot\vec{\mathcal{A}}(t)\,.$$

As usual, the $\mathcal{A}^2(t)$ simply adds a time-dependent phase to the wave function and can be ignored. It follows that the interaction Hamiltonian becomes

$$\hat{H}_1 = \tfrac{i}{\omega}\vec{\mathcal{E}}_0\cdot\left\{-\frac{e}{m}\hat{\vec{p}} + \frac{e^2\hbar}{4m^2c^2}\vec{\sigma}\times\vec{\mathcal{E}}_V\right\}.$$

The extra bit here is exactly the same as found for the current operator. It follows that all results derived so far hold if we make the substitution

$$\hat{\vec{p}} \to \hat{\vec{p}} - \frac{e\hbar}{4mc^2}\vec{\sigma}\times\vec{\mathcal{E}}_V\,.$$

This also includes gauge invariance because $[\vec{\sigma}\times\vec{\mathcal{E}}_V,\vec{r}] = 0$.

Exercise: Stark effect in hydrogen

In this exercise, based on Ref. [4], we consider again hydrogen in a static field. We will, however, go to higher order in the perturbation, thereby computing static hyper-polarizabilities. The perturbation, as usual, is $V = \mathcal{E}z$. The trick to a systematic treatment of this problem is to work in parabolic coordinates $\xi = r + z$ and $\eta = r - z$. In these coordinates, the Laplacian is

$$\nabla^2 = \frac{4}{\xi+\eta}\left\{\frac{\partial}{\partial\xi}\xi\frac{\partial}{\partial\xi}+\frac{\partial}{\partial\eta}\eta\frac{\partial}{\partial\eta}\right\}.$$

Moreover, the potential energy terms are $V=-1/r=-2/(\xi+\eta)$ for the Coulomb potential and $V=\mathcal{E}z=\frac{1}{2}\mathcal{E}(\xi-\eta)$ for the electrostatic potential.

a) Show that this leads to the eigenvalue problem

$$\left\{\frac{\partial}{\partial\xi}\xi\frac{\partial}{\partial\xi}+\frac{\partial}{\partial\eta}\eta\frac{\partial}{\partial\eta}-\mathcal{E}\frac{\xi^2-\eta^2}{4}+E\frac{\xi+\eta}{2}+1\right\}\psi=0.$$

To solve, we write $\psi=f(\xi)g(\eta)$ and introduce $\beta=(E/E_0)^{1/2}$, $E_0=-\frac{1}{2}$, $F=\mathcal{E}/4\beta^3$, $x=\beta\xi$, and $y=\beta\eta$. We now separate the problem into *x*- and *y*-dependent parts. To this end, we introduce two separation constants satisfying $\beta_1+\beta_2=1/\beta$ so that

$$\left\{\frac{\partial}{\partial x}x\frac{\partial}{\partial x}+\left(\beta_1-Fx^2-\frac{x}{4}\right)\right\}f=0$$
$$\left\{\frac{\partial}{\partial y}y\frac{\partial}{\partial y}+\left(\beta_2+Fy^2-\frac{y}{4}\right)\right\}g=0.$$

b) It is convenient to introduce $z\equiv f'/f$. Show that the *x*-equation becomes

$$z+x\left(z'+z^2\right)+\beta_1-Fx^2-\frac{x}{4}=0.$$

We focus exclusively on the ground state and write the unknowns as Taylor series in *F*, i.e. $\beta_1=\sum_{n=0}^{\infty}a_nF^n$, $z(x)=\sum_{n=0}^{\infty}z_n(x)F^n$. At zeroth order, $\beta=1$ and by symmetry $\beta_1=\beta_2=1/2$. It is readily shown by inspection that $z_0(x)=-1/2$ and thereby $f_0(x)=\exp\left(-\frac{x}{2}\right)$. Generally, for $k>0$ the *k*′th order term in $z^2(x)$ is

$$\left[z^2(x)\right]_k=2z_0z_k+\sum_{n=1}^{k-1}z_nz_{k-n},\quad k>0.$$

The first order equation for *z* therefore reads as $\{1+2xz_0\}z_1+xz_1'=-a_1+x^2$. Via the auxiliary function *h*

$$\frac{h'(x)}{h(x)}=-\frac{1}{x}\{1+2xz_0\}$$

we find $z_1(x) = h(x)\int_x^\infty (a_1 - t^2)/[t\,h(t)]dt$ as can be checked by inspection . The solution is $h(x) = C/[xf_0^2(x)]$ and so $z_1(x) = \int_x^\infty f_0^2(t)(a_1 - t^2)dt/[xf_0^2(x)]$. In order for the solution to be regular at zero, $a_1 = \int_0^\infty x^2 f_0^2(x)dx / \int_0^\infty f_0^2(x)dx = 2$. In turn, $z_1(x) = -(2+x)$. For the general *k*′th order term with $k > 1$ we find

$$\left\{\frac{1}{x} + 2z_0\right\} z_k + z_k' = -\frac{a_k}{x} - \sum_{n=1}^{k-1} z_n z_{k-n}, \quad k > 1,$$

so that

$$a_k = -\frac{\int_0^\infty x f_0^2(x) \sum_{n=1}^{k-1} z_n(x) z_{k-n}(x) dx}{\int_0^\infty f_0^2(x)dx}.$$

c) Show by iterating the above procedure that $a_2 = -18$ and $z_2(x) = 18 + 7x + x^2$. Write a program to help show that $a_3 = 356$ and $a_4 = -10026$. For the *y*-dependent function *g* a similar series solution can be constructed. Since the two equations only differ by the sign of *F*, the series for β_2 will be $\beta_2 = \sum_{n=0}^\infty (-1)^n a_n F^n$. Eventually, adding the two provides the energy. At second order, $1/\beta = 1 + a_2\left(\mathcal{E}/4\beta^3\right)^2$. Remembering that $\beta = (-2E)^{1/2}$ we find that $E \approx -1/2 - 9\mathcal{E}^2/4$.

d) Repeat this at fourth order and show that $E \approx -1/2 - 9\mathcal{E}^2/4 - 3555\mathcal{E}^4/64$.

Thus, the static (hyper-) polarizabilities are $\alpha_1(0) = 9/2$ and $\alpha_3(0) = 3555/16$.

References

[1] L. Pan, K.T. Taylor, and C.W. Clark, Phys. Rev. A39, 4894 (1989).
[2] G.L. Sewell, Proc. Camb. Phil. Soc., 45, 678 (1949).
[3] A. Taghizadeh, F. Hipolito, and T. G. Pedersen, Phys. Rev. B96, 195413 (2017).
[4] V. Privman, Phys. Rev. A22, 1833 (1980).

45. Nonlinear Response of Periodic Structures

In this chapter, we depart from the highly general discussion of the previous chapter and focus on a particular example of special importance: Periodic structures. To simplify the analysis, we will ignore excitonic effects so that all states are taken to be single-electron band states. Accordingly, each state index above is replaced by a collection of band, *k*-point and spin index, i.e. $n \to \{n, \vec{k}, \sigma\}$. Summation of spin is always straightforward if spin-orbit coupling is ignored and simply leads to multiplicative factors of two.

Before attempting actual computations we need to address a subtle point related to *intraband* transitions in infinite structures, that is, transitions from $\vec{k}$ to $\vec{k}'$ within a particular band. To facilitate the discussion, we formally separate the dipole operator $\vec{r}$ into intraband $\vec{r}^{\,i}$ and interband $\vec{r}^{\,e}$ parts, i.e. $\vec{r} = \vec{r}^{\,e} + \vec{r}^{\,i}$. Then, $\vec{r}^{\,i}$ only has diagonal matrix elements in the band indices and $\vec{r}^{\,e}$ only has off-diagonal ones. As long as we restrict ourselves to interband transitions between non-degenerate bands, everything is simple because we rely on the usual reformulation in terms of momentum matrix elements $\langle m\vec{k}'|\vec{r}^{\,e}|n\vec{k}\rangle = \langle m\vec{k}'|\hat{\vec{p}}|n\vec{k}\rangle \hbar / [im(E_{m\vec{k}'} - E_{n\vec{k}})]$. Such a procedure is tricky whenever $n = m$ because we end up taking the limit $\vec{k}' \to \vec{k}$ so that the energy difference vanishes. Therefore, a more systematic approach is required.

We write the band states as $|n\vec{k}\rangle = \Omega^{-1/2} u_{n\vec{k}} \exp(i\vec{k}\cdot\vec{r})$, where $u_{n\vec{k}}$ is the lattice-periodic part with the normalization $\delta_{nm} = \Omega_{UC}^{-1} \int u_{n\vec{k}}^* u_{m\vec{k}} d^3r$. We now define the important Berry connection as $\vec{\Omega}_{nm\vec{k}} = i\Omega_{UC}^{-1} \int u_{n\vec{k}}^* \nabla_{\vec{k}} u_{m\vec{k}} d^3r$. To proceed, we follow Blount [1] and write

$$-i\nabla_{\vec{k}} |n\vec{k}\rangle = \frac{-i}{\Omega^{1/2}} e^{i\vec{k}\cdot\vec{r}} \nabla_{\vec{k}} u_{n\vec{k}} + \vec{r}\,|n\vec{k}\rangle .$$

Thus, because $\langle m\vec{k}'|n\vec{k}\rangle = \delta_{nm}\delta_{\vec{k}\vec{k}'}$ the inner product with $\langle m\vec{k}'|$ yields

$$\langle m\vec{k}'|\vec{r}|n\vec{k}\rangle = \vec{\Omega}_{mn\vec{k}}\delta_{\vec{k}\vec{k}'} - i\delta_{nm}\nabla_{\vec{k}}\delta_{\vec{k}\vec{k}'} .$$

Similarly, interchanging quantum numbers, we find $\langle n\vec{k}|\vec{r}|m\vec{k}'\rangle = \vec{\Omega}_{nm\vec{k}}\delta_{\vec{k}\vec{k}'} - i\delta_{nm}\nabla_{\vec{k}'}\delta_{\vec{k}\vec{k}'}$. Comparison with the complex conjugate of the result above then shows that $\nabla_{\vec{k}'}\delta_{\vec{k}\vec{k}'} = -\nabla_{\vec{k}}\delta_{\vec{k}\vec{k}'}$. Note, that $\vec{\Omega}_{mn\vec{k}} = \hbar\vec{p}_{mn\vec{k}} / (imE_{mn\vec{k}})$ with $E_{mn\vec{k}} = E_{m\vec{k}} - E_{n\vec{k}}$ for $n \neq m$ as

one can demonstrate using perturbation theory to compute $u_{n,\vec{k}+\delta\vec{k}}$. For the intraband dipole operator we then find

$$\left\langle n\vec{k}\left|\vec{r}^{\,i}\right|m\vec{k}'\right\rangle = \delta_{nm}\left\{\vec{\Omega}_{nn\vec{k}}\delta_{\vec{k}\vec{k}'} - i\nabla_{\vec{k}'}\delta_{\vec{k}\vec{k}'}\right\}.$$

The problem is how to make sense of the last term. To solve this, we now consider the commutator with an arbitrary operator $\hat{O}$ and apply the completeness relation

$$\begin{aligned}
\left\langle n\vec{k}\left|[\vec{r}^{\,i},\hat{O}]\right|m\vec{k}'\right\rangle &= \sum_{l,\vec{k}''}\left\{\left\langle n\vec{k}\left|\vec{r}^{\,i}\right|l\vec{k}''\right\rangle\left\langle l\vec{k}''\left|\hat{O}\right|m\vec{k}'\right\rangle - \left\langle n\vec{k}\left|\hat{O}\right|l\vec{k}''\right\rangle\left\langle l\vec{k}''\left|\vec{r}^{\,i}\right|m\vec{k}'\right\rangle\right\} \\
&= \sum_{\vec{k}''}\left\{\left[\vec{\Omega}_{nn\vec{k}}\delta_{\vec{k}\vec{k}''} - i\nabla_{\vec{k}''}\delta_{\vec{k}\vec{k}''}\right]\left\langle n\vec{k}''\left|\hat{O}\right|m\vec{k}'\right\rangle - \left\langle n\vec{k}\left|\hat{O}\right|m\vec{k}''\right\rangle\left[\vec{\Omega}_{mm\vec{k}'}\delta_{\vec{k}'\vec{k}''} - i\nabla_{\vec{k}''}\delta_{\vec{k}'\vec{k}''}\right]\right\} \\
&= \delta_{\vec{k}\vec{k}'}\left[\vec{\Omega}_{nn\vec{k}} - \vec{\Omega}_{mm\vec{k}}\right]O_{nm\vec{k}} - i\sum_{\vec{k}''}\left\{\left(\nabla_{\vec{k}''}\delta_{\vec{k}\vec{k}''}\right)\left\langle n\vec{k}''\left|\hat{O}\right|m\vec{k}'\right\rangle - \left\langle n\vec{k}\left|\hat{O}\right|m\vec{k}''\right\rangle\left(\nabla_{\vec{k}'}\delta_{\vec{k}'\vec{k}''}\right)\right\} \\
&= \delta_{\vec{k}\vec{k}'}\left[\vec{\Omega}_{nn\vec{k}} - \vec{\Omega}_{mm\vec{k}}\right]O_{nm\vec{k}} - i\sum_{\vec{k}''}\left\{\left(\nabla_{\vec{k}''}\delta_{\vec{k}\vec{k}''}\right)\left\langle n\vec{k}''\left|\hat{O}\right|m\vec{k}'\right\rangle + \left\langle n\vec{k}\left|\hat{O}\right|m\vec{k}''\right\rangle\left(\nabla_{\vec{k}''}\delta_{\vec{k}'\vec{k}''}\right)\right\}.
\end{aligned}$$

We now use "integration by parts" to transfer the derivatives of the Kronecker deltas to the matrix elements (a more rigorous would involve converting the sum to an integral and the Kronecker deltas to delta functions before integration). After subsequently summing over k'' we find

$$\left\langle n\vec{k}\left|[\vec{r}^{\,i},\hat{O}]\right|m\vec{k}'\right\rangle = \delta_{\vec{k}\vec{k}'}\left\{\left[\vec{\Omega}_{nn\vec{k}} - \vec{\Omega}_{mm\vec{k}}\right]O_{nm\vec{k}} + i\nabla_{\vec{k}}O_{nm\vec{k}}\right\}.$$

This important result can be expressed in a brief form by introducing the "generalized derivative" defined by

$$\left\langle n\vec{k}\left|[\vec{r}^{\,i},\hat{O}]\right|m\vec{k}'\right\rangle = i\delta_{\vec{k}\vec{k}'}(O_{nm})_{;\vec{k}}, \quad (O_{nm})_{;\vec{k}} \equiv \nabla_{\vec{k}}O_{nm\vec{k}} - i\left[\vec{\Omega}_{nn\vec{k}} - \vec{\Omega}_{mm\vec{k}}\right]O_{nm\vec{k}}. \quad (45.1)$$

Here, "$;\vec{k}$" is the symbol for generalized derivative [3].

45.1 Response Functions

For simplicity, we now restrict ourselves to one-dimensional systems. Applying the generalized derivative formalism to the linear intraband polarizability Eq.(44.7) we find

$$\alpha_1^{intra}(\omega)=\frac{e^2}{\hbar}\sum_{n,k,k'}\left(f_{nk}-f_{nk'}\right)|\,x^i_{nk,nk'}\,|^2\,\frac{\omega_{nk',nk}}{\omega^2_{nk',nk}-\omega^2}$$
$$=\frac{e^2}{\hbar^2}\sum_{n,k,k'}\left\langle nk\right|[x^i,f(\hat{H})]\left|nk'\right\rangle\left\langle nk'\right|[x^i,\hat{H}]\left|nk\right\rangle\frac{1}{\omega^2_{nk',nk}-\omega^2}.$$

Accordingly, we find

$$\alpha_1^{intra}(\omega)=\frac{e^2}{\hbar^2\omega^2}\sum_{n,k}\frac{df_{nk}}{dk}\frac{dE_{nk}}{dk}. \tag{45.2}$$

Converting to an integral and using $df/dk=f'\cdot dE/dk$, this expression is seen to be equivalent to the Drude formulas discussed in Chapter 3 and 8. We now consider $\alpha_2(\omega)$ in the special case of a two-band semiconductor. When only two-band contributions are retained, the general expression Eq.(44.9) can be reduced to

$$\alpha_2(\omega)=\frac{3e^3}{\hbar^2}\sum_{m\neq n,k,k'}f_{nk}\frac{|\,x_{nk,mk'}\,|^2\,(x^i_{nk,nk'}-x^i_{mk,mk'})\omega^2_{mk',nk}}{\left(\omega^2_{mk',nk}-4\omega^2\right)\left(\omega^2_{mk',nk}-\omega^2\right)}$$
$$=\frac{3e^3}{\hbar^2m^2}\sum_{m\neq n,k,k'}f_{nk}\frac{|\,p_{nk,mk'}\,|^2\,(x^i_{nk,nk'}-x^i_{mk,mk'})}{\left(\omega^2_{mk',nk}-4\omega^2\right)\left(\omega^2_{mk',nk}-\omega^2\right)}.$$

Because the intraband dipole operator has only diagonal elements we can rewrite as the commutator $\lim_{k'\to k}p_{nk,mk'}(x^i_{nk,nk'}-x^i_{mk,mk'})=\lim_{k'\to k}\left\langle nk\right|[x^i,\hat{p}_x]\left|mk'\right\rangle$. Thus, with $p_{nmk}=\left\langle nk\right|p\left|mk\right\rangle$ and $\omega_{mnk}=(E_{mk}-E_{nk})/\hbar$

$$\alpha_2(\omega)=\frac{3ie^3}{\hbar^2m^2}\sum_{m\neq n,k}f_{nk}\frac{p_{nmk}(p_{mnk})_{;k}}{\left(\omega^2_{mnk}-4\omega^2\right)\left(\omega^2_{mnk}-\omega^2\right)} \tag{45.3}$$

in agreement with Refs. [2,4]. Finally, a purely interband contribution is found

$$\alpha_2^{inter}(\omega)=-\frac{2ie^3}{\hbar^2m^3}\sum_{\substack{m,n,l,k\\ m\neq n\neq l}}\frac{p_{lmk}p_{mnk}p_{nlk}}{\omega_{mlk}\omega_{mnk}\omega_{lnk}\left(\omega_{mlk}-2\omega\right)}\left\{\frac{f_{lnk}}{\omega_{nlk}-\omega}+\frac{f_{mnk}}{\omega_{mnk}-\omega}\right\}. \tag{45.4}$$

In a periodic material, $p_{mn,-k}=-p_{nmk}$ and, hence, averaging over $\pm k$ shows that only the imaginary part of the triple matrix product contributes. The energy denominators yield delta-function contributions to the imaginary part of the response precisely at resonance and after collecting and relabeling terms

$$\mathrm{Im}\,\alpha_2^{inter}(\omega)=\frac{2\pi e^3}{\hbar^2 m^3}\sum_{\substack{m\in c,n\in v\\ k,l\neq m\neq n}}\frac{\mathrm{Im}\{p_{nmk}p_{mlk}p_{lnk}\}}{\omega_{mnk}\omega_{mlk}\omega_{lnk}}\times \\ \left\{\frac{\delta(\omega_{mnk}-2\omega)}{\omega_{lnk}-\omega}+\left[\frac{1}{\omega_{mlk}+\omega}+\frac{1}{\omega_{lnk}+\omega}\right]\delta(\omega_{mnk}-\omega)\right\}. \tag{45.5}$$

A similar analysis applies to the third order response. Here, even for a two-band material, both mixed (*iie* and *iee*) and purely interband (*eee*) terms appear. One finds

$$\alpha_3^{inter}(\omega)=\frac{4e^4\hbar^4}{m^4}\sum_k\frac{|\,p_{vck}\,|^4}{E_{cvk}^3(E_{cvk}^2-9\hbar^2\omega^2)(E_{cvk}^2-\hbar^2\omega^2)},$$

$$\alpha_3^{(iie)}(\omega)=\frac{4e^4\hbar^2}{m^2}\sum_k\left(\frac{p_{vck}}{E_{cvk}-3\hbar\omega}\right)_{;k}\frac{1}{E_{cvk}-2\hbar\omega}\left(\frac{p_{cvk}}{E_{cvk}(E_{cvk}-\hbar\omega)}\right)_{;k}+(\omega\to-\omega)^*,$$

$$\alpha_3^{(iee)}(\omega)=\frac{4e^4\hbar^2}{m^2}\sum_k\frac{p_{vvk}-p_{cck}}{(E_{cvk}^2-\hbar^2\omega^2)(E_{cvk}^2-4\hbar^2\omega^2)} \\ \times\left\{\Omega_{vck}\Omega'_{cvk}(2\hbar\omega-E_{cvk})+\Omega'_{vck}\Omega_{cvk}(2\hbar\omega+E_{cvk})-\frac{6\hbar\omega E_{cvk}E'_{cvk}\,|\,\Omega_{vck}\,|^2}{E_{cvk}^2-\hbar^2\omega^2}\right\}.$$

45.2 Gapped Graphene

We now specialize to gapped graphene, for which analytical results can be found provided appropriate simplifications are made. In particular, we expand k-dependent functions in the vicinity of a Dirac point. If only first order terms are retained, this would constitute the normal Dirac approximation. However, non-vanishing even-order nonlinearities require broken symmetry, which is missing in the Dirac model. Consequently, we need to expand to higher order, retaining so-called trigonal warping terms.

The transition energy and momentum matrix elements of gapped graphene have already been discussed in Chapter 17 within the Dirac model. In this case, the band energies are $E_{v,c}=\mp\varepsilon$ with $\varepsilon=\sqrt{\alpha^2+(\hbar v_F k)^2}$, where on-site energies $\pm\alpha$ yield a band gap of $E_g=2\alpha$. Going one order higher means that

$$E_{cv}\approx 2\varepsilon\left\{1+2\zeta_{TW}\eta ka\left(1-\frac{\alpha^2}{\varepsilon^2}\right)\sin 3\varphi\right\}.$$

Here, $\eta=\pm1$ is the valley index, $\zeta_{TW}=1/(8\sqrt{3})$ is the trigonal-warping strength and φ the polar angle of $\vec{k}$. The name "trigonal warping" refers to the $\sin 3\varphi$ factor, which warps the energy surface in a trigonal fashion. Similarly, momentum matrix elements become

$$\frac{p_{cv}^x}{mv_F} \approx \frac{\alpha}{\varepsilon}\cos\varphi + i\eta\sin\varphi + \zeta_{TW} ka\left(-3i\cos 2\varphi + i\cos 4\varphi + \eta\frac{\alpha}{\varepsilon}\left(4+\frac{\alpha^2}{\varepsilon^2}\right)\sin 2\varphi - \eta\frac{\alpha}{\varepsilon}\left(2-\frac{\alpha^2}{\varepsilon^2}\right)\sin 4\varphi\right)$$

$$\frac{p_{cv}^y}{mv_F} \approx \frac{\alpha}{\varepsilon}\sin\varphi - i\eta\cos\varphi + \zeta_{TW} ka\left(3i\sin 2\varphi + i\sin 4\varphi + \eta\frac{\alpha}{\varepsilon}\left(4+\frac{\alpha^2}{\varepsilon^2}\right)\cos 2\varphi + \eta\frac{\alpha}{\varepsilon}\left(2-\frac{\alpha^2}{\varepsilon^2}\right)\cos 4\varphi\right).$$

In addition, we need Berry connections

$$\Omega_{cc}^x - \Omega_{vv}^x \approx \frac{a\alpha}{\varepsilon}\left\{-\eta\frac{\sin\varphi}{ka} + \zeta_{TW}\left(5-\frac{\alpha^2}{\varepsilon^2}\right)\cos 2\varphi - \zeta_{TW}\left(3-\frac{\alpha^2}{\varepsilon^2}\right)\cos 4\varphi\right\},$$

$$\Omega_{cc}^y - \Omega_{vv}^y \approx \frac{a\alpha}{\varepsilon}\left\{\eta\frac{\cos\varphi}{ka} - \zeta_{TW}\left(5-\frac{\alpha^2}{\varepsilon^2}\right)\sin 2\varphi - \zeta_{TW}\left(3-\frac{\alpha^2}{\varepsilon^2}\right)\sin 4\varphi\right\}.$$

The motivation for introducing ζ_{TW} even though this is a fixed numerical factor is that it enables systematic expansions in orders of trigonal warping.

Next, we turn to linear and nonlinear optical response functions. Our aim is to apply the above expansions in order to enable analytical integrations over *k*-space. In turn, analytical optical conductivities can be found. Note that we compute conductivities rather than polarizabilties because we're dealing with an infinite two-dimensional material. First, we look at the diagonal linear response

$$\sigma_{xx}(\omega) = \frac{e^2\hbar^2\omega}{i\pi^2 m^2}\int \frac{|\, p_{cv}^x\,|^2}{E_{cv}(E_{cv}^2 - \hbar^2\omega^2)}d^2k\,.$$

Without approximations, the integral is over the entire Brillouin zone. However, with the expansion introduced above, it means (1) integration over the infinite Dirac cone and (2) sum over valley index. To zeroth order in ζ_{TW}, we find $E_{cv} = 2\varepsilon$ and

$$\frac{1}{2\pi}\sum_{\eta=\pm 1}\int_0^{2\pi}|\, p_{vc}^x\,|^2 d\varphi = (mv_F)^2\left(1+\frac{\alpha^2}{\varepsilon^2}\right) = (mv_F)^2\left(1+\frac{E_g^2}{E_{cv}^2}\right).$$

Using $\hbar^2 v_F^2 k dk = \varepsilon d\varepsilon = \frac{1}{4}E_{cv}dE_{cv}$ it follows that

$$\begin{aligned}\sigma_{xx}(\omega) &= \frac{e^2\omega}{2i\pi}\int_{E_g}^{\infty}\left(1+\frac{E_g^2}{E_{cv}^2}\right)\frac{1}{E_{cv}^2-\hbar^2\omega^2}dE_{cv} \\ &= \frac{2i\sigma_0}{\pi}\left\{\frac{E_g}{\hbar\omega} - \left(1+\frac{E_g^2}{\hbar^2\omega^2}\right)\tanh^{-1}\left(\frac{\hbar\omega}{E_g}\right)\right\}.\end{aligned} \tag{45.6}$$

Here, we introduced the pristine graphene conductivity $\sigma_0 = e^2/4\hbar$. We now turn to the second order nonlinear response. In the general case of two fields with frequencies ω_p and ω_q, whose sum is $\omega_{pq} \equiv \omega_p + \omega_q$, the second order response in the length gauge consists of inter- and intraband contributions

$$\sigma^{(inter)}_{\lambda\alpha\beta}(\omega_p,\omega_q) = \frac{e^3\hbar^2}{4m^3\Omega}\sum_{\substack{m,n,l,\vec{k}\\ m\neq n\neq l}} \frac{p^{\lambda}_{nm}}{E_{ln}E_{ml}[E_{mn}-\hbar\omega_{pq}]}\left\{\frac{f_{nl}p^{\beta}_{ml}p^{\alpha}_{ln}}{E_{ln}-\hbar\omega_p} - \frac{f_{lm}p^{\alpha}_{ml}p^{\beta}_{ln}}{E_{ml}-\hbar\omega_p}\right\} + (p\leftrightarrow q),$$

$$\sigma^{(intra)}_{\lambda\alpha\beta}(\omega_p,\omega_q) = \frac{e^3\hbar}{4m^2\Omega}\sum_{\substack{m,n,\vec{k}\\ m\neq n}} \frac{p^{\lambda}_{nm}}{E_{mn}[E_{mn}-\hbar\omega_{pq}]}\frac{f_{nm}(p^{\alpha}_{mn})_{;k_\beta}}{E_{mn}-\hbar\omega_p} + (p\leftrightarrow q).$$

In the symmetrization $(p\leftrightarrow q)$, both positive and negative frequencies should be given a positive imaginary broadening. In a clean two-band semiconductor, only the intraband response survives and for the diagonal we find

$$\sigma^{(intra)}_{xxx}(\omega_p,\omega_q) = \frac{e^3\hbar^2}{2\pi^2m^2}\int p^{x}_{vc}(p^{x}_{cv})_{;k_x}\frac{\omega_p+\omega_{pq}}{(E^2_{cv}-\hbar^2\omega^2_p)(E^2_{cv}-\hbar^2\omega^2_{pq})}d^2k + (p\leftrightarrow q).$$

We're interested in the response to first order in ζ_{TW}. Thus, there are contributions from both momentum $p^{x}_{vc}(p^{x}_{cv})_{;k_x}$ and energy E_{cv} factors. However, due to time-reversal symmetry, $E_{cv}(-\vec{k}) = E_{cv}(\vec{k})$ and, therefore, any term containing $p^{x}_{vc}(p^{x}_{cv})_{;k_x}$ expanded to zeroth order vanishes upon integration (and summation over valley). We are consequently allowed to set $E_{cv} = 2\varepsilon$ throughout. This means that valley index and angle only appear in $p^{x}_{vc}(p^{x}_{cv})_{;k_x}$. It is readily demonstrated that

$$\frac{1}{2\pi}\sum_{\eta=\pm1}\int_0^{2\pi} p^{x}_{vc}(p^{x}_{cv})_{;k_x}d\varphi = \zeta_{TW}(mv_F)^2\frac{8a\alpha}{i\varepsilon} = \zeta_{TW}(mv_F)^2\frac{8aE_g}{iE_{cv}}.$$

We again use $\hbar^2v_F^2kdk = \varepsilon d\varepsilon = \frac{1}{4}E_{cv}dE_{cv}$ and with $\sigma_2 = e^3a/(4\hbar E_g)$ we find

$$\sigma^{(intra)}_{xxx}(\omega_p,\omega_q) = i\sigma_2\frac{4\zeta_{TW}}{\pi}E_g^2\int_{E_g}^{\infty}\frac{\hbar\omega_p+\hbar\omega_{pq}}{(E^2_{cv}-\hbar^2\omega^2_q)(E^2_{cv}-\hbar^2\omega^2_{pq})}dE_{cv} + (p\leftrightarrow q).$$

The factor $\sigma_2 = e^3a/(4\hbar E_g) = ea\sigma_0/E_g$ has a simple interpretation: Since first and second order currents vary as $\mathcal{E}$ and $\mathcal{E}^2$, respectively, the two contributions are of the same order, i.e. $\sigma_2\mathcal{E}^2 = \sigma_0\mathcal{E}$, at a field strength given by $ea\mathcal{E} = E_g$. Hence, a field of this (huge) magnitude is required to make linear and nonlinear terms of the same order. Physically, $ea\mathcal{E}$ is the dipole energy of an electron-hole pair separated by the

lattice constant. Hence, if $ea\mathcal{E} = E_g$, we're clearly outside the regime of small perturbations.

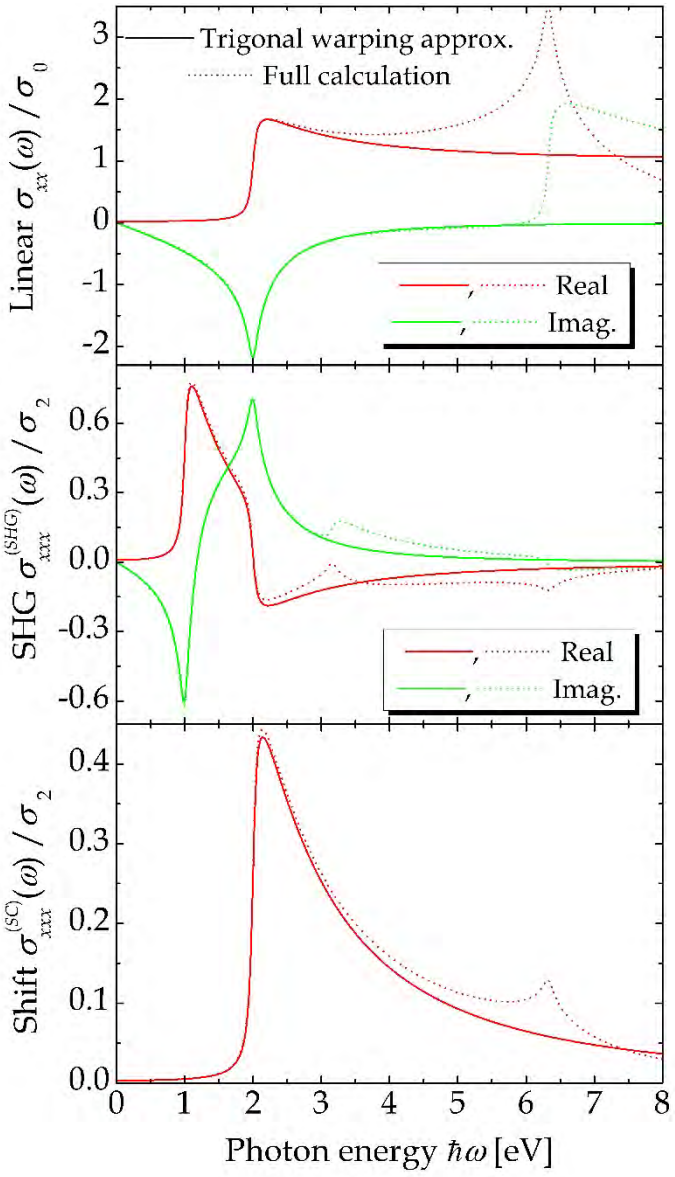


Figure. 45.1. Comparison between full gapped graphene response (dashed curves) and trigonal warping approximation (solid curves).

We now specialize to SHG and shift current (SC) responses. After integrations,

$$\sigma^{(SHG)}_{xxx}(\omega) \equiv \sigma^{(intra)}_{xxx}(\omega,\omega) = i\sigma_2 \frac{16\zeta_{TW}}{\pi}\left(\frac{E_g}{\hbar\omega}\right)^2 \left\{\tanh^{-1}\left(\frac{\hbar\omega}{E_g}\right) - \frac{1}{2}\tanh^{-1}\left(\frac{2\hbar\omega}{E_g}\right)\right\}, \quad (45.7)$$

and

$$\sigma^{(SC)}_{xxx}(\omega) \equiv \sigma^{(intra)}_{xxx}(\omega,-\omega^*) = \sigma_2 \frac{16\zeta_{TW}}{\pi}\,\mathrm{Im}\left\{\frac{E_g^2}{\hbar^2\omega^2}\tanh^{-1}\left(\frac{\hbar\omega}{E_g}\right) - \frac{E_g}{\hbar\omega}\right\}. \quad (45.8)$$

It is seen that, remarkably, both linear and nonlinear conductivities depend essentially only on band gap (the nonlinear ones also depend on a but only as a prefactor). We can assess the quality of Eqs.(45.6), (45.7) and (45.8) by comparing to full spectra found by integrating numerical over full Brillouin zones without approximating energies and matrix elements. This is done in Fig. 45.1, where we consider gapped graphene with $E_g = 2$ eV and applying a broadening of $\hbar\Gamma = 0.05$ eV. Also, for the full model, we use nearest neighbor hopping $\gamma = -3$ eV.

The comparison in the figure demonstrates that the trigonal warping model is an excellent approximation in the vicinity of the band gap. It fails near the van Hove singularity, i.e. around $\hbar\omega = n\sqrt{\alpha^2+\gamma^2}$, with $n = 2$ for linear and shift current response and $n = 1$ for SHG. This is not surprising as the van Hove singularity is completely absent in the simplified model.

Exercise: Oligo- and polymethineimine

The molecule oligo-methineimine (OMI) and the infinite polymer poly-methineimine (PMI) are illustrated in Fig. 45.2.

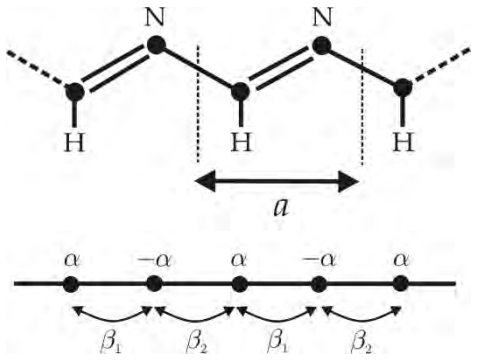


Figure 45.2. poly-methineimine and its tight-binding representation.

It consists of alternating C and N atoms along the backbone and, in addition, is dimerized with alternating single and double bonds. Hence, the tight-binding Hamiltonian is of the form

$$\overleftrightarrow{H} = \begin{pmatrix} \alpha & \beta_1 & 0 & \cdots \\ \beta_1 & -\alpha & \beta_2 & \cdots \\ 0 & \beta_2 & \alpha & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

We take the values $\alpha = 0.5$ eV, $\beta_1 = -2.0$ eV and $\beta_2 = -2.5$ eV and the lattice constant a = 2Å. All bond lengths are assumed identical and the number of unit cells in the chain is N.

a) Write a program that computes the nonlinear polarizability per unit cell according to Eq.(44.9). If successful, the result becomes as shown in Fig. 45.3.

For the infinite polymer we can use the two-band expression Eq.(45.3). The required quantities are

$$E_{v,c} = \mp E, \quad E = \sqrt{\alpha^2 + F^2}, \quad F = \sqrt{\beta_1^2 + \beta_2^2 + 2\beta_1\beta_2 \cos(ka)}$$

$$p_{cv} = \frac{ma}{\hbar}\left\{-2\alpha\beta_1\beta_2 \sin(ka) + i\left(\beta_1^2 - \beta_2^2\right)E\right\} / (2FE)$$

$$\Omega_{cc} - \Omega_{vv} = a\alpha\left(\beta_1^2 - \beta_2^2\right) / (2F^2E).$$

b) Compute the nonlinear response for the infinite chain. The result, shown in black in Fig. 45.3, agrees nicely with the dipole expression for the longest chain.

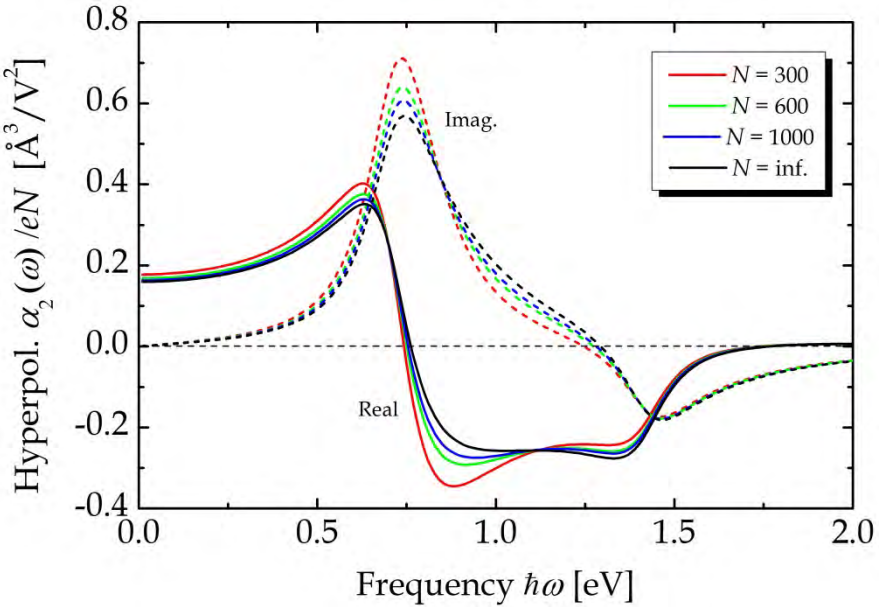


Figure 45.3. Second-order hyperpolarizability for methineimine finite chains and polymer.

References

[1] E.I. Blount, Solid State Physics 13, Ed. F. Seitz and D. Turnbull (Academic Press Inc., New York, 1962).
[2] B. Kirtman, F.L. Gu, and D.M. Bishop, J. Chem. Phys. 113, 1294 (2000).
[3] C. Aversa and J.E. Sipe, Phys. Rev. B52, 14636 (1995).
[4] T.G. Pedersen, Phys. Rev. B92, 235432 (2015).

46. Linear and Nonlinear Response of Bulk Crystals

The previous chapters have presented expressions for optical response functions of periodic systems. In this chapter, we will reformulate the expressions for optical response functions into forms that are suitable for a highly efficient integration scheme called the tetrahedron method. We will focus on the velocity gauge second order response Eq.(44.12) applied to a periodic system. Below, we also show that this result is identical to the length gauge expression, thus demonstrating gauge invariance.

When summed over spin and symmetrized over polarization indices (suppressing $\vec{k}$ - dependence of all quantities) the velocity gauge nonlinear susceptibility is

$$\chi^{(2)}_{ijk}(\omega) = \frac{ie^3}{(2\pi)^3 \varepsilon_0 m^3 \hbar^2 \omega^3} \sum_{lmn} \int \frac{\left\langle p^i_{lm} p^j_{mn} p^k_{nl} \right\rangle}{\omega_{ml} - 2\omega} \left\{ \frac{f_{ln}}{\omega_{nl} - \omega} + \frac{f_{mn}}{\omega_{mn} - \omega} \right\} d^3k \,. \tag{46.1}$$

This can be rewritten

$$\chi^{(2)}_{ijk}(\omega) = -\frac{3ie^3}{(2\pi)^3 \varepsilon_0 m^3 \hbar^2 \omega^3} \int \left\{ \sum_{vcv'} \left\langle p^i_{vc} p^j_{cv'} p^k_{v'v} \right\rangle D_{cv'v}(\omega) - \sum_{vcc'} \left\langle p^i_{vc} p^j_{cc'} p^k_{c'v} \right\rangle D_{vcc'}(\omega) \right\} d^3k \,, \tag{46.2}$$

where v's and c's denote valence and conduction bands, respectively, and

$$D_{cv'v}(\omega) = \frac{\omega\omega_{v'v} + \omega_{cv'}\omega_{cv}}{(\omega_{cv} - 2\omega)(\omega_{cv} + \omega)(\omega_{cv'} + 2\omega)(\omega_{cv'} - \omega)},$$
$$D_{vcc'}(\omega) = \frac{\omega\omega_{cc'} + \omega_{c'v}\omega_{cv}}{(\omega_{cv} - 2\omega)(\omega_{cv} + \omega)(\omega_{c'v} + 2\omega)(\omega_{c'v} - \omega)}.$$

We aim to extract the imaginary parts in the limit of vanishing broadening and find

$$\pi^{-1} \operatorname{Im} D_{cv'v}(\omega) = -\frac{\omega_{v'v} + \omega_{cv}}{3(\omega_{cv'} + \omega_{vv'})(\omega_{cv'} + \omega_{cv})} \delta(\omega_{cv'} - \omega) + \frac{2}{3(\omega_{cv'} + \omega_{vv'})} \delta(\omega_{cv} - 2\omega),$$
$$\pi^{-1} \operatorname{Im} D_{vcc'}(\omega) = \frac{\omega_{cc'} + \omega_{cv}}{3(\omega_{cc'} + \omega_{vc'})(\omega_{cc'} + \omega_{cv})} \delta(\omega_{c'v} - \omega) - \frac{2}{3(\omega_{cc'} + \omega_{vc'})} \delta(\omega_{cv} - 2\omega).$$

To proceed, we interchange v and v' in the first part of $D_{cv'v}$ as well as c and c' in the first part of $D_{vcc'}$. Since $\left\langle p^i_{vc\vec{k}} p^j_{cv'\vec{k}} p^k_{v'v\vec{k}} \right\rangle$ can be taken purely imaginary (employing time-reversal symmetry), it then follows that

$$\mathrm{Im}\,\chi^{(2)}_{ijk}(\omega) = -\frac{\pi e^3}{(2\pi)^3 \varepsilon_0 m^3 \hbar^2 \omega^3} \times$$
$$\int \sum_{vcl} \mathrm{Im}\left\langle p^i_{vc} p^j_{cl} p^k_{lv} \right\rangle \left\{ \frac{\omega_{cl} - \omega_{lv}}{(\omega_{cv} + \omega_{lv})(\omega_{cv} + \omega_{cl})} \delta(\omega_{cv} - \omega) + \frac{2}{\omega_{cl} - \omega_{lv}} \delta(\omega_{cv} - 2\omega) \right\} d^3k.$$

Here, the l summation is now over all states. In practice, small imaginary parts must be added to the remaining denominators to avoid divergences. For numerical purposes, it is advantageous to introduce

$$\Pi^{(\omega)}_{ijk} = \sum_l \mathrm{Im}\left\langle p^i_{vc} p^j_{cl} p^k_{lv} \right\rangle \frac{\omega_{cl} - \omega_{lv}}{(\omega_{cv} + \omega_{lv})(\omega_{cv} + \omega_{cl})}, \quad \Pi^{(2\omega)}_{ijk} = \sum_l \mathrm{Im}\left\langle p^i_{vc} p^j_{cl} p^k_{lv} \right\rangle \frac{2}{\omega_{cl} - \omega_{lv}} \tag{46.3}$$

such that

$$\mathrm{Im}\,\chi^{(2)}_{ijk}(\omega) = -\frac{\pi e^3}{(2\pi)^3 \varepsilon_0 m^3 \hbar^2 \omega^3} \int \sum_{vc} \left\{ \Pi^{(\omega)}_{ijk} \delta(\omega_{cv} - \omega) + \Pi^{(2\omega)}_{ijk} \delta(\omega_{cv} - 2\omega) \right\} d^3k. \tag{46.4}$$

This expression is well suited for numerical evaluation, as we will discuss below. The linear response Eq.(15.3) is

$$\chi_{ij}(\omega) = \frac{4e^2}{(2\pi)^3 \varepsilon_0 m^2 \hbar} \int \sum_{vc} \frac{\left\langle p^i_{vc} p^j_{cv} \right\rangle}{\omega_{cv}(\omega^2_{cv} - \omega^2)} d^3k. \tag{46.5}$$

Hence, the nonlinear expression Eq.(46.4) is obviously quite similar to the linear equivalent in the limit of vanishing broadening

$$\mathrm{Im}\,\chi_{ij}(\omega) = \frac{2\pi e^2}{(2\pi)^3 \varepsilon_0 m^2 \hbar \omega^2} \int \sum_{vc} \left\langle p^i_{vc} p^j_{cv} \right\rangle \delta(\omega_{cv} - \omega) d^3k. \tag{46.6}$$

46.1 Tetrahedron Integration

To evaluate both linear and nonlinear optical response functions, we would like an efficient way of computing integrals of the form

$$\chi(\omega) = \int F(\vec{k}) \delta[E(\vec{k}) - \omega] d^3k. \tag{46.7}$$

Such a tool is provided by the tetrahedron method and we will discuss the particular implementation described in Ref. [1]. First, the total integration volume is chopped into tetrahedra with corners $\vec{k}_i$, $i \in [1,4]$. The energy $E(\vec{k})$ is evaluated in all corners providing $E_i \equiv E(\vec{k}_i)$ and sorted according to $E_1 \le E_2 \le E_3 \le E_4$. We also need

$F_i \equiv F(\vec{k}_i)$. By linearizing both $E(\vec{k})$ and $F(\vec{k})$ inside each tetrahedron, the integral Eq.(46.7) is then approximated by the sum over tetrahedra T with volumes V_T

$$\chi(\omega) = \sum_T \sum_{j=1}^{4} G_j(\omega) F_j V_T \,. \tag{46.8}$$

We now introduce $E_{ij} \equiv E_i - E_j$ and $f_{ij} \equiv (\omega - E_j)/E_{ij}$ for $i \neq j$ and $f_{ii} = 0$. The weights in Eq.(46.8) depend on ω and $G_j(\omega) = 0$ for $\omega < E_1$ and $\omega > E_4$. Otherwise,

$$G_j(\omega) = \begin{cases} f_{21} f_{31} f_{41} / (\omega - E_1) \left\{ (f_{12} + f_{13} + f_{14}) \delta_{j1} + f_{j1} \right\}, & E_1 < \omega < E_2 \\ \quad E_{41}^{-1} \left\{ f_{14} f_{23} f_{31} + f_{14} f_{32} f_{24} + f_{13} f_{31} f_{23} \right\} \delta_{j1} + & \\ \quad E_{41}^{-1} \left\{ f_{23}^2 f_{31} + f_{23} f_{32} f_{24} + f_{24}^2 f_{32} \right\} \delta_{j2} + & E_2 < \omega < E_3 \\ \quad E_{41}^{-1} \left\{ f_{32} f_{23} f_{31} + f_{32}^2 f_{24} + f_{31}^2 f_{23} \right\} \delta_{j3} + & \\ \quad E_{41}^{-1} \left\{ f_{41} f_{23} f_{31} + f_{41} f_{32} f_{24} + f_{42} f_{24} f_{32} \right\} \delta_{j4}, & \\ f_{14} f_{24} f_{34} / (E_4 - \omega) \left\{ (f_{41} + f_{42} + f_{43}) \delta_{j4} + f_{j4} \right\}, & E_3 < \omega < E_4. \end{cases}$$

It is convenient to use normalized quantities $k = (2\pi/a)\tilde{k}$ and $p = \hbar(2\pi/a)\tilde{p}$ with lattice constant a. Evaluating frequencies in energy units, one has

$$\begin{aligned} \mathrm{Im}\, \chi_{ij}(\omega) &= \frac{32\pi^4}{\omega^2} \mathrm{Ha}^3 \left(\frac{a_0}{a} \right)^5 \int \sum_{vc} \left\langle \tilde{p}_{vc}^i \tilde{p}_{cv}^j \right\rangle \delta(\omega_{cv} - \omega) d^3\tilde{k}, \\ \mathrm{Im}\, \chi_{ijk}^{(2)}(\omega) &= -\chi_0^{(2)} \frac{128\pi^6}{\omega^3} \mathrm{Ha}^5 \left(\frac{a_0}{a} \right)^6 \int \sum_{vc} \left\{ \tilde{\Pi}_{ijk}^{(\omega)} \delta(\omega_{cv} - \omega) + \tilde{\Pi}_{ijk}^{(2\omega)} \delta(\omega_{cv} - 2\omega) \right\} d^3\tilde{k}. \end{aligned} \tag{46.9}$$

Here, $\chi_0^{(2)} = a_0^2 \varepsilon_0 / e \approx 0.155$ pm/V is a universal constant.

To illustrate the power of the tetrahedron method, we investigate the linear and nonlinear optical response of zinc-blende GaAs modelled using the empirical pseudopotential method, as in App. 3, using 65 G-vectors and including 15 bands. The linear response is isotropic $\chi_{xx} = \chi_{yy} = \chi_{zz}$ and the non-vanishing nonlinear tensor elements are $\chi_{xyz}^{(2)}$ and associated permutations. Integration can be restricted to the irreducible Brillouin zone if the symmetrized momentum products are taken as

$$\begin{aligned} \left\langle p^x p^x \right\rangle &= \tfrac{1}{3} (p^x p^x + p^y p^y + p^z p^z), \\ \left\langle p^x p^y p^z \right\rangle &= \tfrac{1}{6} (p^x p^y p^z + p^y p^x p^z + p^z p^y p^x + p^x p^z p^y + p^z p^x p^y + p^y p^z p^x). \end{aligned}$$

We use Matlab's *generateMesh* to partition the irreducible Brillouin zone into tetrahedra. Setting the maximal edge length to 0.075, 0.05, or 0.03 (in units of $2\pi / a$) leads to 1501, 5328 or 24742 tetrahedra, respectively. The broadening is taken as 20 meV throughout. We will compare

A: Tetrahedron integration evaluating Eq.(46.8) and (46.9).
B: Point sampling evaluation of Eq.(46.2) and (46.5) using averages of the four corners of the tetrahedra for energy differences and momentum products.

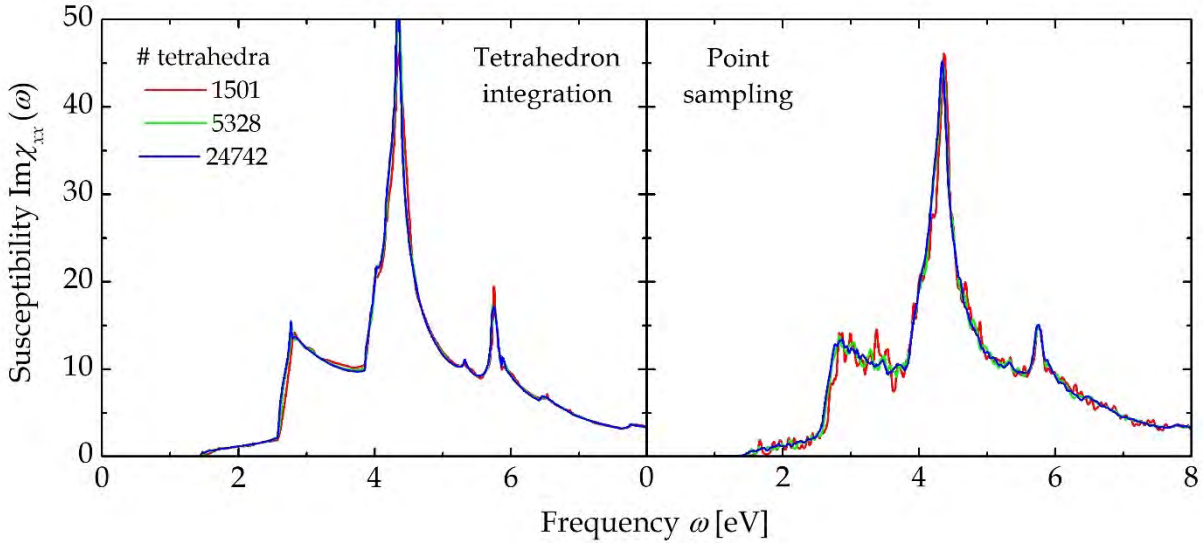


Figure 46.1. Linear response of GaAs computed using tetrahedron integration (left) and point sampling (right).

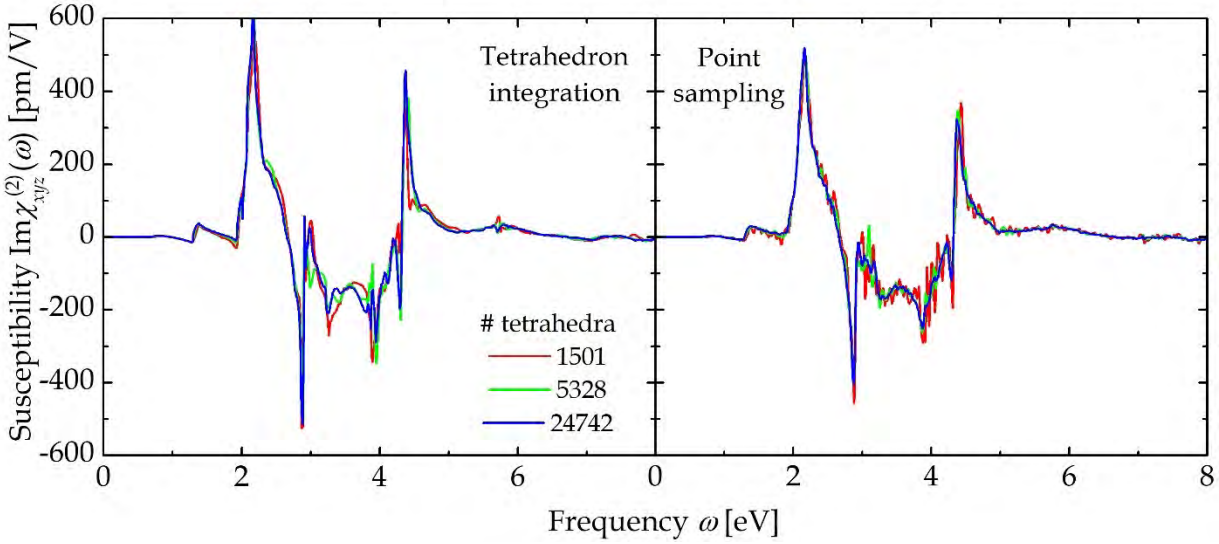


Figure 46.2. Same as Fig. 46.1 but for the second order nonlinear response.

It is readily seen that the tetrahedron method converges much faster than point sampling in both the linear and nonlinear cases. Also, it should be noted that, for a given number of tetrahedra, point sampling is much slower than tetrahedron integration. The tetrahedron method provides the imaginary part of the response in the limit of vanishing broadening. In fact, the full (real and imaginary) response

including broadening $\chi_\Gamma(\omega)$ can be obtained from the complex Kramers-Kronig transformation [2]

$$\chi_\Gamma(\omega) = \frac{2}{\pi}\int_0^\infty \mathrm{Im}\,\chi_{\Gamma=0}(\nu)\frac{\nu}{\nu^2-(\omega+i\Gamma)^2}d\nu\,. \tag{46.10}$$

Note that $\mathrm{Im}\,\chi_{\Gamma=0}(\nu)$ must be known to high frequencies in order to do this integral accurately. Here, an upper limit of 20 eV is used. The Kramers-Kronig convolution applies to both linear and nonlinear response. In Figs. 46.3 and 46.4, the results for zinc-blende GaAs using 5328 tetrahedra and $\Gamma = 20$ meV are illustrated.

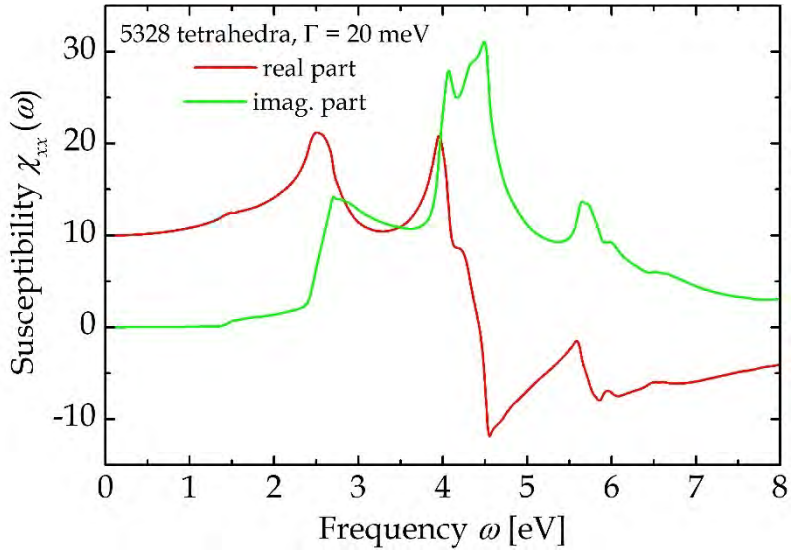


Figure 46.3. GaAs susceptibility obtained from the broadened Kramers-Kronig transform.

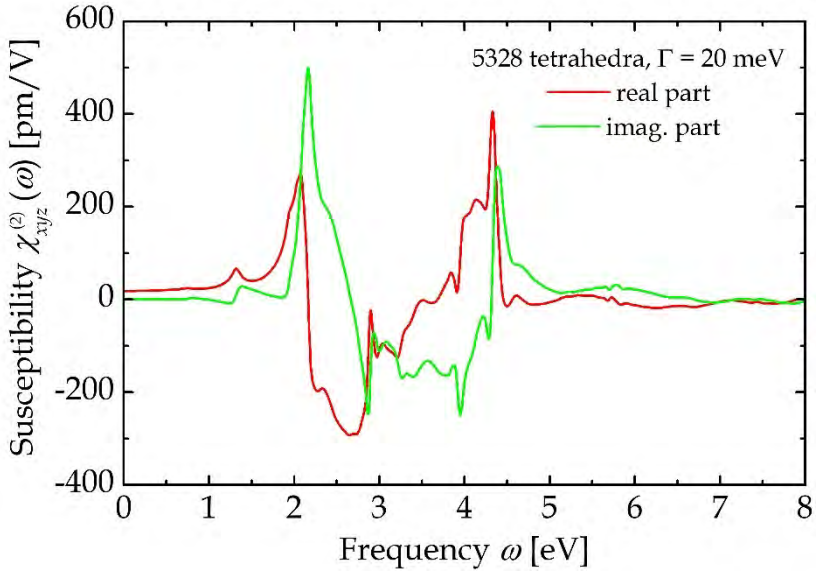


Figure 46.4. Same as Fig. 46.3 but for the nonlinear response.

Next, we will discuss gauge invariance. An alternative to the above results is provided by the length gauge expression $\chi^{(2)}_{ijk}(\omega)=\chi^{(2e)}_{ijk}(\omega)+\chi^{(2i)}_{ijk}(\omega)$ with interband contribution

$$\chi^{(2e)}_{ijk}(\omega)=\frac{2ie^3}{(2\pi)^3\varepsilon_0 m^3\hbar^2}\sum_{\substack{lmn\\ l\neq m\neq n}}\int\frac{p^i_{lm}p^j_{mn}p^k_{nl}}{\omega_{mn}\omega_{nl}\omega_{ml}(\omega_{ml}-2\omega)}\left\{\frac{f_{ln}}{\omega_{nl}-\omega}+\frac{f_{mn}}{\omega_{mn}-\omega}\right\}d^3k$$

and mixed intra/interband term

$$\chi^{(2i)}_{ijk}(\omega)=\frac{2ie^3}{(2\pi)^3\varepsilon_0 m^2\hbar^2}\sum_{\substack{mn\\ m\neq n}}\int\frac{f_{nm}p^i_{nm}}{\omega_{mn}(\omega_{mn}-2\omega)}\left\{\frac{p^j_{mn}}{\omega_{mn}(\omega_{mn}-\omega)}\right\}_{;k_k}d^3k\,.$$

The interband part resembles the velocity gauge case and, similarly,

$$\mathrm{Im}\,\chi^{(2e)}_{ijk}(\omega)=-\frac{2\pi e^3}{(2\pi)^3\varepsilon_0 m^3\hbar^2}\int\sum_{vc}\left\{\Pi^{(\omega,e)}_{ijk}\delta(\omega_{cv}-\omega)+\Pi^{(2\omega,e)}_{ijk}\delta(\omega_{cv}-2\omega)\right\}d^3k\,,$$

where, now,

$$\Pi^{(\omega,e)}_{ijk}=\sum_{l\neq v,c}\frac{\mathrm{Im}\left\langle p^i_{vc}p^j_{cl}p^k_{lv}\right\rangle}{\omega_{cv}\omega_{cl}\omega_{lv}}\frac{\omega_{cl}-\omega_{lv}}{(\omega_{cv}+\omega_{lv})(\omega_{cv}+\omega_{cl})},\quad \Pi^{(2\omega,e)}_{ijk}=-\sum_{l\neq v,c}\frac{\mathrm{Im}\left\langle p^i_{vc}p^j_{cl}p^k_{lv}\right\rangle}{\omega_{cv}\omega_{cl}\omega_{lv}}\frac{2}{\omega_{cl}-\omega_{lv}}.$$

Using sum rules [3], the generalized derivative is

$$\left\{\frac{p^j_{mn}}{\omega_{mn}(\omega_{mn}-\omega)}\right\}_{;k_k}=-\frac{p^j_{mn}\Delta^k_{mn}(2\omega_{mn}-\omega)}{\omega^2_{mn}(\omega_{mn}-\omega)^2}+\frac{1}{m\omega_{mn}(\omega_{mn}-\omega)}\left\{\hbar m\delta_{mn}\delta_{jk}+\sum_{l\neq m}\frac{p^k_{ml}p^j_{ln}}{\omega_{ml}}-\sum_{l\neq n}\frac{p^j_{ml}p^k_{ln}}{\omega_{ln}}\right\}.$$

The contribution from the first term vanishes in cubic crystals and eventually

$$\chi^{(2i)}_{ijk}(\omega)=\frac{2e^3}{(2\pi)^3\varepsilon_0 m^2\hbar^2}\sum_{\substack{v,c\\ l\neq v,c}}\int\frac{(\omega_{cl}-\omega_{lv})\,\mathrm{Im}\left\langle p^i_{vc}p^j_{cl}p^k_{lv}\right\rangle}{\omega^2_{cv}(\omega_{cv}-\omega)(\omega_{cv}-2\omega)\omega_{cl}\omega_{lv}}d^3k\,.$$

Extracting the imaginary part, the *e* and *i* terms can be grouped so that

$$\Pi^{(\omega,e+i)}_{ijk}=\sum_{l\neq v,c}\frac{\mathrm{Im}\left\langle p^i_{vc}p^j_{cl}p^k_{lv}\right\rangle}{\omega_{cv}\omega_{cl}\omega_{lv}}\left\{\frac{\omega_{cl}-\omega_{lv}}{(\omega_{cv}+\omega_{lv})(\omega_{cv}+\omega_{cl})}-\frac{\omega_{cl}-\omega_{lv}}{2\omega^2_{cv}}\right\},$$

$$\Pi^{(2\omega,e+i)}_{ijk}=-2\sum_{l\neq v,c}\frac{\mathrm{Im}\left\langle p^i_{vc}p^j_{cl}p^k_{lv}\right\rangle}{\omega_{cv}\omega_{cl}\omega_{lv}}\left\{\frac{1}{\omega_{cl}-\omega_{lv}}-\frac{\omega_{cl}-\omega_{lv}}{\omega^2_{cv}}\right\}.$$

Applying the simplifications resulting from the respective delta functions, it is readily shown that the two gauges produce identical results. For instance, $\Pi_{ijk}^{(\omega,e+i)}\Big|_{\omega_{cv}=\omega} = -\Pi_{ijk}^{(\omega)}/(2\omega^3)$ and similarly for the 2ω term. Note that the restriction $l \neq v,c$ can be applied to the velocity gauge as well, as $\mathrm{Im}\left\langle p_{vc}^i p_{cv}^j p_{vv}^k \right\rangle = \mathrm{Im}\left\langle p_{vc}^i p_{cc}^j p_{cv}^k \right\rangle = 0$.

Exercise: Wurtzite crystals

Wurtzite crystals can be handled using tetrahedron integration as well. Results are slightly more complicated because the crystal is anistropic. Hence, $\chi_{xx} = \chi_{yy} \neq \chi_{zz}$ and the distinct nonlinear tensor elements are $\chi_{xzx}^{(2)}$, $\chi_{zxx}^{(2)}$ and $\chi_{zzz}^{(2)}$. We will start by finding the appropriate symmetrized momentum products for integration over the irreducible Brillouin zone (IBZ) shown in Fig. A3.1. Suppose we pick a random *k*-vector in the IBZ $\vec{k} = (k_x, k_y, k_z)$. There are 24 equivalent *k*-vectors given by

$$\vec{k}_{pqr} = (k_x \cos\tfrac{p\pi}{3} + k_y \sin\tfrac{p\pi}{3}, q[k_y \cos\tfrac{p\pi}{3} - k_x \sin\tfrac{p\pi}{3}], rk_z)$$

where $p = \{0,1,2,3,4,5\}$, $q = \{-1,1\}$ and $r = \{-1,1\}$. The original vector is $\vec{k} = \vec{k}_{011}$. In the following, we arbitrarily pick $\vec{k} = (0.3, 0.1, 0.2\frac{a}{c})\frac{2\pi}{a}$ and bands $(n,m,l) = (3,5,8)$ using GaN parameters from Ref. [4] and 147 *G*-vectors. A code for Wurtzite band structure and eigenvectors is a prerequisite for this exercise.

a) Write a computer program demonstrating that energies at all 24 *k*-points are degenerate.

b) Form the products $S_{ij}(\vec{k}) = p_{nm}^i(\vec{k}) p_{mn}^j(\vec{k})$ and $S_{ijk}(\vec{k}) = p_{nm}^i(\vec{k}) p_{ml}^j(\vec{k}) p_{ln}^k(\vec{k})$ as well as the averages over IBZ's $\left\langle S_{ij} \right\rangle = \frac{1}{24}\sum_{\vec{k}} S_{ij}(\vec{k})$ and $\left\langle S_{ijk} \right\rangle = \frac{1}{24}\sum_{\vec{k}} S_{ijk}(\vec{k})$ and show that

$$\left\langle S_{xx} \right\rangle = \tfrac{1}{2}(p_{nm}^x p_{mn}^x + p_{nm}^y p_{mn}^y), \quad \left\langle S_{zz} \right\rangle = p_{nm}^z p_{mn}^z, \quad \left\langle S_{zzz} \right\rangle = \mathrm{Im}\, p_{nm}^z p_{ml}^z p_{ln}^z,$$
$$\left\langle S_{xzx} \right\rangle = \tfrac{1}{2}\mathrm{Im}(p_{nm}^x p_{ml}^z p_{ln}^x + p_{nm}^y p_{ml}^z p_{ln}^y), \quad \left\langle S_{zxx} \right\rangle = \tfrac{1}{2}\mathrm{Im}(p_{nm}^z p_{ml}^x p_{ln}^x + p_{nm}^z p_{ml}^y p_{ln}^y).$$

Here, the right-hand sides can be evaluated in *any* IBZ, i.e. they are invariant under the symmetry operations of the crystal.

c) Apply tetrahedron integration to compute all distinct tensor elements of both linear and nonlinear response in GaN. The result is shown below.

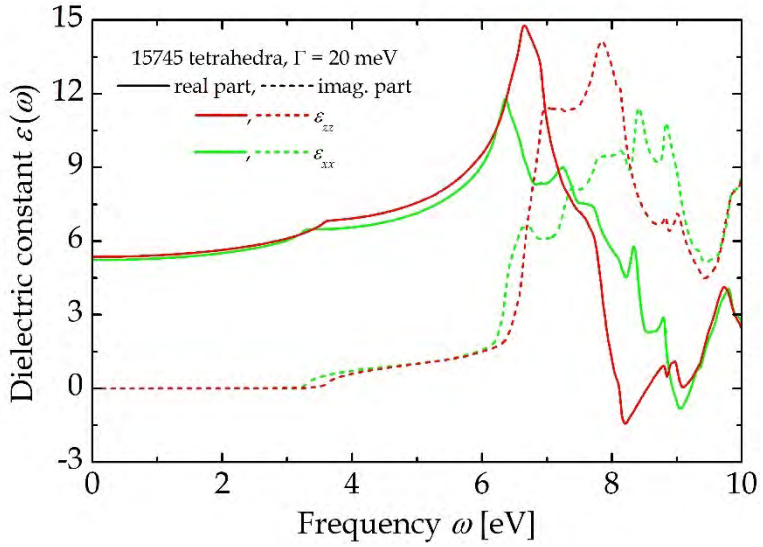


Figure 46.5. Anisotropic dielectric constant of Wurtzite GaN.

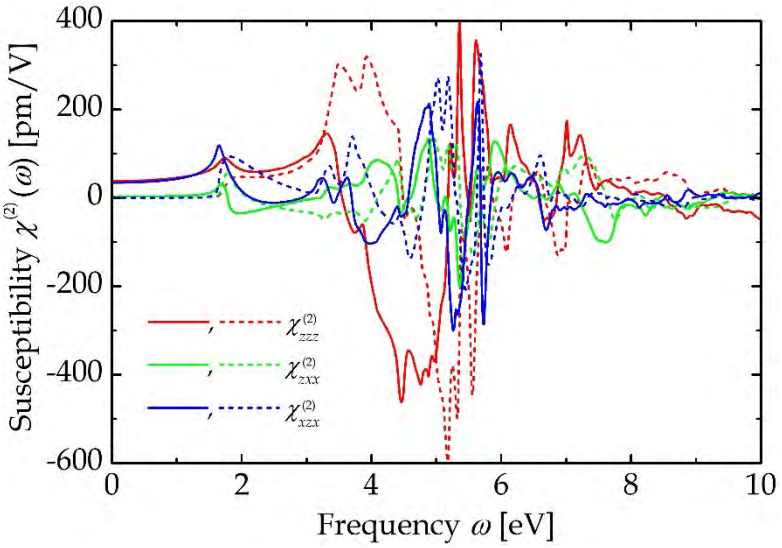


Figure 46.6. Same as Fig. 46.5 but for the nonlinear tensor elements.

The above results have been obtained with a maximal tetrahedron edge of $0.03\times 2\pi/a$ and a small portion near the zone center has been refined using $0.01\times 2\pi/a$ producing a total of 15745 tetrahedra.

References

[1] A. H. MacDonald, S. H. Vosko, and P. T. Coleridge, J. Phys. C12, 2991 (1979).
[2] T. G. Pedersen and T. B. Lynge, Phys. Rev. B65, 085201 (2002).
[3] A. Taghizadeh, F. Hipolito, and T. G. Pedersen, Phys. Rev. B96, 195413 (2017).
[4] R. Wang, P. Ruden, J. Kolnik, I. Oguzman, and K. Brennan, J. Phys. Chem. Solids 58, 913 (1997).

47. Spatial Dispersion and Nonlocal Response

We are perfectly used to temporal dispersion: this simply means that properties of matter depend on frequency. In linear optics, for instance, $\mathcal{P}(\omega)=\varepsilon_0\chi(\omega)\mathcal{E}(\omega)$ with a frequency-dependent susceptibility $\chi(\omega)$. In general, this relation is between Fourier components of field and polarization and holds even for a non-monochromatic field. Since the right-hand side is a product in frequency-domain, the time-domain expression becomes a convolution

$$\mathcal{P}(t)=\varepsilon_0\int_{-\infty}^{t}\chi(t-t')\mathcal{E}(t')dt'.$$

This results clearly shows that the present polarization depends on the field at previous times $t'<t$. Hence, the response in *nonlocal* in time. A more unusual concept is that of spatial dispersion or, equivalently, spatial nonlocality. In space-domain, this means that the response at $\vec{r}$ depends on perturbations at other locations $\vec{r}'$. In translationally invariant cases, we may Fourier-transform in space as well as time to find a response that depends on wave vector $\vec{q}$. Thus, in the optical case, we find a susceptibility $\chi(\omega,\vec{q})$. Without spatial invariance, we get the space-domain expression $\chi(\omega,\vec{r},\vec{r}')$. In fact, examples of this kind were considered in Chapter 30.

Spatial dispersion effects are most important for the optical response, and we will focus on this case. A crucial point is that we cannot apply the dipole approximation, as we have done in previous chapters. The reason is that we want to treat response and perturbation at separate locations. Hence, the spatial variation of the perturbation becomes important and this is precisely what is neglected in the dipole approximation. We may choose a gauge, in which the optical perturbation is contained in a time- and space-dependent vector potential $\vec{\mathcal{A}}(\vec{r},t)$ while keeping the scalar potential $V(\vec{r})$ static. The single-electron Hamiltonian is therefore

$$\hat{H}=\frac{1}{2m}\left\{\hat{\vec{p}}+e\vec{\mathcal{A}}(\vec{r},t)\right\}^2+V(\vec{r})$$

and it follows that the perturbation is

$$\hat{H}_{pert}(\vec{r},t)=\frac{e}{2m}\left\{\hat{\vec{p}}\cdot\vec{\mathcal{A}}(\vec{r},t)+\vec{\mathcal{A}}(\vec{r},t)\cdot\hat{\vec{p}}\right\}+\frac{e^2}{2m}\mathcal{A}^2(\vec{r},t).$$

The rule for deriving the current density operator is by taking the functional derivative $\hat{j}_x(\vec{r},\vec{r}',t)=-\delta\hat{H}_{pert}(\vec{r},t)/\delta\mathcal{A}_x(\vec{r}',t)$ so that

$$\hat{\vec{j}}(\vec{r},\vec{r}',t) = -\frac{e}{2m}\left\{\hat{\vec{p}}'\delta(\vec{r}-\vec{r}') + \delta(\vec{r}-\vec{r}')\hat{\vec{p}}'\right\} - \frac{e^2}{m}\vec{\mathcal{A}}(\vec{r},t)\delta(\vec{r}-\vec{r}').$$

The associated charge density operator is simply $\hat{\rho}(\vec{r},\vec{r}') = -e\delta(\vec{r}-\vec{r}')$. In a monochromatic electric field $\vec{\mathcal{E}}(\vec{r},t) = \frac{1}{2}\vec{\mathcal{E}}(\vec{r})e^{-i\omega t} + c.c.$, the vector potential is $\vec{\mathcal{A}}(\vec{r},t) = \frac{1}{2}\vec{\mathcal{A}}(\vec{r})e^{-i\omega t} + c.c.$ with $\vec{\mathcal{A}}(\vec{r}) = \vec{\mathcal{E}}(\vec{r})/(i\omega)$. Hence, if we write the current as $\vec{\mathcal{J}}(\vec{r},t) = \frac{1}{2}\vec{\mathcal{J}}(\vec{r})e^{-i\omega t} + c.c.$, we can then use standard perturbation theory to write the nonlocal response in the form

$$\vec{\mathcal{J}}(\vec{r}) = \int \overleftrightarrow{\sigma}_{nl}(\vec{r},\vec{r}')\cdot\vec{\mathcal{E}}(\vec{r}')d^3r'$$

with the nonlocal conductivity (ignoring broadening)

$$\overleftrightarrow{\sigma}_{nl}(\vec{r},\vec{r}') = \overleftrightarrow{\sigma}_D(\vec{r})\delta(\vec{r}-\vec{r}') + \frac{2e^2}{im^2\omega}\sum_{m,n} f_{nm}\frac{E_{mn}\vec{j}_{nm}(\vec{r})\vec{j}_{mn}(\vec{r}')}{E_{mn}^2 - \hbar^2\omega^2}. \tag{47.1}$$

Here, $\overleftrightarrow{\sigma}_D(\vec{r}) = ie^2 n(\vec{r})\overleftrightarrow{I}/(m\omega)$ and we introduced the transition momentum density defined by

$$\vec{j}_{mn}(\vec{r}) = \tfrac{1}{2}\left\langle\varphi_m(\vec{r}')\right|\left\{\hat{\vec{p}}'\delta(\vec{r}-\vec{r}') + \delta(\vec{r}-\vec{r}')\hat{\vec{p}}'\right\}\left|\varphi_n(\vec{r}')\right\rangle = \tfrac{1}{2}\left\{\varphi_m^*(\vec{r})\hat{\vec{p}}\varphi_n(\vec{r}) - \varphi_n(\vec{r})\hat{\vec{p}}\varphi_m^*(\vec{r})\right\}.$$

It follows that $\int\vec{j}_{mn}(\vec{r})d^3r = \vec{p}_{mn}$ and by averaging $\overleftrightarrow{\sigma} = \Omega^{-1}\iint\overleftrightarrow{\sigma}_{nl}(\vec{r},\vec{r}')d^3rd^3r'$ we recover the usual conductivity.

47.1 Periodic Systems

We will now specialize to the crystalline case, in which states are of the Bloch form $\varphi_n(\vec{r}) = \Omega^{-1/2}u_{n\vec{k}}(\vec{r})e^{i\vec{k}\cdot\vec{r}}$ and $\varphi_m(\vec{r}) = \Omega^{-1/2}u_{m\vec{k}'}(\vec{r})e^{i\vec{k}'\cdot\vec{r}}$. Furthermore, the field is taken to be a plane wave $\vec{\mathcal{E}}(\vec{r}) = \vec{\mathcal{E}}_0 e^{i\vec{q}\cdot\vec{r}}$, where we assume that q is small compared to all non-vanishing reciprocal lattice vectors. Then, to evaluate the current, we need the integral

$$\int\vec{j}_{m\vec{k}'n\vec{k}}(\vec{r}')\cdot\vec{\mathcal{E}}(\vec{r}')d^3r' = \delta_{\vec{k}',\vec{k}+\vec{q}}\vec{\mathcal{E}}_0\cdot\vec{p}_{m\vec{k}'n\vec{k}}, \quad \vec{p}_{m\vec{k}'n\vec{k}} = \tfrac{1}{2}\hbar(\vec{k}'+\vec{k})\left\langle u_{m\vec{k}'}\middle|u_{n\vec{k}}\right\rangle + \left\langle u_{m\vec{k}'}\middle|\hat{\vec{p}}\middle|u_{n\vec{k}}\right\rangle.$$

It is readily seen that the response will vary as the phase factor $e^{i\vec{q}\cdot\vec{r}}$ times a lattice-periodic term. If the latter is averaged over a unit cell, we obtain the macroscopic response. In terms of the polarization, we then have $\vec{\mathcal{P}}(\vec{r}) = \varepsilon_0\overleftrightarrow{\chi}(\omega,\vec{q})\cdot\vec{\mathcal{E}}_0 e^{i\vec{q}\cdot\vec{r}}$ with

$$\overleftrightarrow{\chi}(\omega,\vec{q}) = -\frac{e^2 n}{\varepsilon_0 m\omega^2}\overleftrightarrow{I} + \frac{2e^2}{\varepsilon_0 m^2\omega^2\Omega}\sum_{m,n,\vec{k}} f_{n\vec{k},m\vec{k}+\vec{q}}\frac{E_{m\vec{k}+\vec{q},n\vec{k}}\vec{p}_{n\vec{k},m\vec{k}+\vec{q}}\vec{p}_{m\vec{k}+\vec{q},n\vec{k}}}{E^2_{m\vec{k}+\vec{q},n\vec{k}} - \hbar^2\omega^2}. \quad (47.2)$$

Here, the density is $n = (2/\Omega)\sum_{n,\vec{k}} f_{n\vec{k}}$. We now apply sum-rule techniques similar to Chapter 43. Hence, by expanding the second term and using the Thomas-Reiche-Kuhn sum rule as derived in Chapter 2

$$\frac{2}{m}\sum_{n,\vec{k}}\frac{\vec{p}_{n\vec{k},m\vec{k}+\vec{q}}\vec{p}_{m\vec{k}+\vec{q},n\vec{k}}}{E_{m\vec{k}+\vec{q},n\vec{k}}} = -\overleftrightarrow{I},$$

it can be shown that

$$\overleftrightarrow{\chi}(\omega,\vec{q}) = \frac{2e^2\hbar^2}{\varepsilon_0 m^2\Omega}\sum_{m,n,\vec{k}} f_{n\vec{k},m\vec{k}+\vec{q}}\frac{\vec{p}_{n\vec{k},m\vec{k}+\vec{q}}\vec{p}_{m\vec{k}+\vec{q},n\vec{k}}}{E_{m\vec{k}+\vec{q},n\vec{k}}[E^2_{m\vec{k}+\vec{q},n\vec{k}} - \hbar^2\omega^2]}. \quad (47.3)$$

This expression clearly reduces to the standard one Eq.(15.3) in the limit $\vec{q}\to 0$ and, hence, agrees with Chapter 15. The limit is

$$\overleftrightarrow{\chi}(\omega,0) = \frac{2e^2}{\varepsilon_0 m^2\omega^2\Omega}\sum_{n,\vec{k}} f'(E_{n\vec{k}})\vec{p}_{nn}\vec{p}_{nn} + \frac{2e^2\hbar^2}{\varepsilon_0 m^2\Omega}\sum_{m\neq n,\vec{k}} f_{nm}\frac{\vec{p}_{nm}\vec{p}_{mn}}{E_{mn}[E^2_{mn} - \hbar^2\omega^2]}.$$

Next, we consider the second order response. For simplicity, only the fully diagonal term $\chi^{(2)} = \chi^{(2)}_{xxx}$ will be studied. We find $\chi^{(2)} = \chi^{(2)}_A + \chi^{(2)}_B$ with

$$\chi^{(2)}_A(\omega,\vec{q}) = \frac{e^3}{2i\varepsilon_0 m^3\omega^3\Omega}\sum_{m,n,l,\vec{k}}\frac{p_{n\vec{k}-\vec{q},m\vec{k}+\vec{q}}p_{m\vec{k}+\vec{q},l\vec{k}}p_{l\vec{k},n\vec{k}-\vec{q}}}{E_{m\vec{k}+\vec{q},n\vec{k}-\vec{q}} - 2\hbar\omega}\left\{\frac{f_{n\vec{k}-\vec{q},l\vec{k}}}{E_{l\vec{k},n\vec{k}-\vec{q}} - \hbar\omega} + \frac{f_{m\vec{k}+\vec{q},l\vec{k}}}{E_{m\vec{k}+\vec{q},l\vec{k}} - \hbar\omega}\right\}, \quad (47.4)$$

$$\chi^{(2)}_B(\omega,\vec{q}) = \frac{e^3}{\varepsilon_0 m^2\omega^2\Omega}\sum_{m,n,\vec{k}} f_{n\vec{k},m\vec{k}+\vec{q}}\frac{E_{m\vec{k}+\vec{q},n\vec{k}}\left\{p_{n\vec{k},m\vec{k}+\vec{q}}s_{m\vec{k}+\vec{q},n\vec{k}} + s_{n\vec{k},m\vec{k}+\vec{q}}p_{m\vec{k}+\vec{q},n\vec{k}}\right\}}{E^2_{m\vec{k}+\vec{q},n\vec{k}} - \hbar^2\omega^2},$$

where $s_{m\vec{k}'n\vec{k}} = \left\langle u_{m\vec{k}'}\middle| u_{n\vec{k}}\right\rangle$. A significant simplification results if we (1) assume a small q and (2) only include the q-dependence in $f_{n\vec{k}-\vec{q}}$ and $f_{m\vec{k}+\vec{q}}$, which turn out to be dominant in metallic systems. In this case, $\chi^{(2)}_B$ vanishes because $E_{m\vec{k}n\vec{k}}s_{m\vec{k}n\vec{k}} = E_{m\vec{k}n\vec{k}}\delta_{mn} = 0$ and we have

$$\chi^{(2)}_A(\omega,\vec{q}) \approx \chi^{(2)}_A(\omega,0) + \frac{ie^3}{2\varepsilon_0 m^3\omega^3\Omega}\sum_{m,n,l,\vec{k}}\frac{p_{nm}p_{ml}p_{ln}}{E_{mn} - 2\hbar\omega}\left\{\frac{\vec{q}\cdot\nabla_k f_{n\vec{k}}}{E_{ln} - \hbar\omega} - \frac{\vec{q}\cdot\nabla_k f_{m\vec{k}}}{E_{ml} - \hbar\omega}\right\}.$$

At low temperatures, using $\vec{p}_{nn} = (m/\hbar)\nabla_k E_{n\vec{k}}$, this is further reduced to

$$\chi_A^{(2)}(\omega,\vec{q}) \approx \chi_A^{(2)}(\omega,0) + \frac{e^3\hbar}{2i\varepsilon_0 m^4\omega^3\Omega}\sum_{m,n,l,\vec{k}}\frac{p_{nm}p_{ml}p_{ln}}{E_{mn}-2\hbar\omega}\left\{\frac{\vec{q}\cdot\vec{p}_{nn}\delta(E_{n\vec{k}}-E_F)}{E_{ln}-\hbar\omega} - \frac{\vec{q}\cdot\vec{p}_{mm}\delta(E_{m\vec{k}}-E_F)}{E_{ml}-\hbar\omega}\right\}.$$

Switching from $\vec{k}$ to $-\vec{k}$ amounts to replacing $p_{nm}p_{ml}p_{ln}$ by $-p_{mn}p_{lm}p_{nl}$. Also, $\vec{p}_{mm}$ is an odd function of $\vec{k}$. It follows that (1) averaging over $\vec{k}$ and $-\vec{k}$ and (2) interchanging m and n in the last term leads to

$$\chi_A^{(2)}(\omega,\vec{q}) \approx \chi_A^{(2)}(\omega,0) + \frac{e^3\hbar^2}{i\varepsilon_0 m^4\omega^2\Omega}\sum_{m,n,l,\vec{k}}\frac{\vec{q}\cdot\vec{p}_{nn}p_{nm}p_{ml}p_{ln}(E_{mn}+2E_{ln})}{(E_{mn}^2-4\hbar^2\omega^2)(E_{ln}^2-\hbar^2\omega^2)}\delta(E_{n\vec{k}}-E_F). \quad (47.5)$$

47.2 Spatial Dispersion in Gapped Graphene

We would like to compute the purely x response of gapped graphene, i.e. taking all matrix elements as well as $\vec{q}$ along x. Within the Dirac approximation, the K-valley of gapped graphene is characterized by

$$E_{c,v} = \pm E, \quad E = \sqrt{\alpha^2 + (\hbar v_F k)^2},$$
$$p_{cv}^x = p_{vc}^{x*} = m v_F\left(\frac{\alpha}{E}\cos\varphi + i\sin\varphi\right), \quad p_{cc}^x = -p_{vv}^x = m v_F\frac{\hbar v_F k}{E}\cos\varphi.$$

Here, φ is the polar angle of the wave vector and 2α the band gap. In the Dirac approximation, $\chi_A^{(2)}(\omega,0)=0$ so we are left with the nonlocal contribution. We assume $E_F>0$ and so $n=c$, whereas $\{m,l\}\in\{v,c\}$ are free in Eq.(47.5). Upon integration and multiplying by valley degeneracy, we then find the susceptibility per area

$$\chi_{xxx}^{(2)} = q\frac{3e^3v_F^2}{i8\pi\varepsilon_0\hbar^2\omega^4}\frac{(E_F^2-\alpha^2)(E_F^2+3\alpha^2)}{(E_F^2-\hbar^2\omega^2)(4E_F^2-\hbar^2\omega^2)}\mathrm{sgn}(E_F)\theta(|E_F|-\alpha).$$

In Fig. 47.1 below, we compare this result to a full pseudopotential calculation for pristine graphene with $\alpha=0$. It is seen that the Dirac model captures the low-frequency contribution whereas the high-energy range disagrees with the full calculation. This missing high-energy response is resonant at the interband transition and, in fact, practically independent of Fermi level.

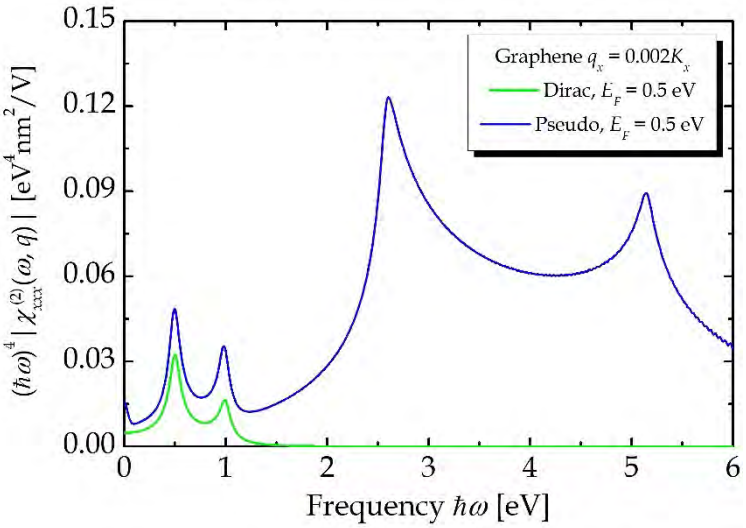


Figure 47.1. Nonlocal second-harmonic susceptibility of graphene in a full pseudopotential model and the Dirac approximation.

Exercise: Semiclassical nonlocal response of graphene

In this exercise, we will adopt the semiclassical Boltzmann description of the optical response. The time-dependent case has already been treated in Chapter 39, which also included magnetic effects. Here, no magnetic field is present but complications arise because we include the spatial variation of the electric field. Thus, the Boltzmann equation reads

$$g(\vec{k},\vec{r},t) = f(E_{\vec{k}}) + \frac{e\tau}{\hbar}\nabla_{\vec{k}} g(\vec{k},\vec{r},t)\cdot\vec{\mathcal{E}}(\vec{r},t) - \tau\frac{\partial g(\vec{k},\vec{r},t)}{\partial t} - \tau\vec{v}\cdot\nabla_{\vec{r}} g(\vec{k},\vec{r},t)\,.$$

Here, the last term is present because the distribution depends on position, c.f. Chapter 12, and the band velocity is $\vec{v} = \nabla_{\vec{k}} E_{\vec{k}} / \hbar$. The field is a plane wave $\vec{\mathcal{E}}(\vec{r},t) = \frac{1}{2}\vec{\mathcal{E}}_0 e^{i(\vec{q}\cdot\vec{r}-\omega t)} + c.c.$ and we write the first-order perturbation as $g_1(\vec{k},\vec{r},t) = \frac{1}{2}G_1 e^{i(\vec{q}\cdot\vec{r}-\omega t)} + c.c.$

a) Show that

$$G_1 = e\tau\frac{f'(E_{\vec{k}})\vec{v}\cdot\vec{\mathcal{E}}_0}{1 - i(\omega - \vec{q}\cdot\vec{v})\tau}\,.$$

In turn, the conductivity of a two-dimensional system becomes

$$\overleftrightarrow{\sigma} = -\frac{e^2\tau}{2\pi^2}\int f'(E_{\vec{k}})\frac{\vec{v}\vec{v}}{1 - i(\omega - \vec{q}\cdot\vec{v})\tau}d^2k\,.$$

In n-doped graphene, $E_{\vec{k}} = \hbar v_F k$, $\vec{v} = v_F \vec{k}/k$ and we will take $\vec{q} = q\vec{e}_x$. Hence, at low temperature, remembering valley degeneracy and using polar coordinates

$$\sigma_{xx} = \frac{e^2\tau v_F^2}{\pi^2}\int \frac{\delta(E_F - \hbar v_F k)\cos^2\theta}{1 - i(\omega - q v_F \cos\theta)\tau} d^2k, \quad \sigma_{yy} = \frac{e^2\tau v_F^2}{\pi^2}\int \frac{\delta(E_F - \hbar v_F k)\sin^2\theta}{1 - i(\omega - q v_F \cos\theta)\tau} d^2k,$$

while the off-diagonal components vanish. The angular integrals can be evaluated using

$$\int_0^{2\pi} \frac{\cos^2\theta}{1 + x\cos\theta} d\theta = \frac{2\pi}{x^2}\left(-1 + \frac{1}{\sqrt{1-x^2}}\right) \approx \pi + \frac{3\pi}{4}x^2,$$
$$\int_0^{2\pi} \frac{\sin^2\theta}{1 + x\cos\theta} d\theta = \frac{2\pi}{x^2}\left(1 - \sqrt{1-x^2}\right) \approx \pi + \frac{\pi}{4}x^2.$$

b) Show that $\sigma_{xx} = \sigma + 3\sigma_{NL}$ and $\sigma_{yy} = \sigma + \sigma_{NL}$ with

$$\sigma = \frac{e^2\tau E_F}{\pi\hbar^2(1 - i\omega\tau)}, \quad \sigma_{NL} = -\frac{\sigma}{4}\left(\frac{q v_F \tau}{1 - i\omega\tau}\right)^2.$$

By another iteration, the second order conductivity becomes

$$\sigma^{(2)}_{xxx} = \frac{e^3\tau^2(1 - 2i\omega\tau)}{2\pi^2\hbar}\int f'(E_{\vec{k}}) \frac{v_x dv_x/dk_x}{\left[1 - i(\omega - \vec{q}\cdot\vec{v})\tau\right]\left[1 - 2i(\omega - \vec{q}\cdot\vec{v})\tau\right]^2} d^2k.$$

We expand to first order in q and find

$$\sigma^{(2)}_{xxx} \approx \frac{e^3\tau^3(-5 + 6i\omega\tau)q}{2\pi^2\hbar(1 - i\omega\tau)^2(1 - 2i\omega\tau)^2}\int f'(E_{\vec{k}}) v_x^2 \frac{dv_x}{dk_x} d^2k.$$

c) Show that $\int f'(E_{\vec{k}}) v_x^2 \frac{dv_x}{dk_x} d^2k = -\frac{\pi v_F^2}{4\hbar}$ for graphene, such that

$$\sigma^{(2)}_{xxx} \approx \frac{e^3\tau^3(5 - 6i\omega\tau)v_F^2 q}{8\pi\hbar^2(1 - i\omega\tau)^2(1 - 2i\omega\tau)^2}.$$

Since $\chi^{(2)}_{xxx} = -\sigma^{(2)}_{xxx}/(2i\varepsilon_0\omega)$, this agrees with the quantum result if we take the limits $\tau, E_F \to \infty$.

Appendix 1. Nanostructures

Much of these lecture notes presupposes some degree of familiarity with the basic properties of nanostructures. It is assumed that electronic states of quantum wells, wires and dots in the infinite step barrier and effective mass approximations are known. For completeness, however, and to refresh the reader's memory, we present in this appendix a brief overview of some common nanostructures, shown in Fig. A1.1. The effective mass approximation is assumed throughout and only simple geometries are considered. We consider the quantum states in various dimensions and confinement potentials as well as their accompanying density of states. We focus on electrons rather than holes since states for the latter follows easily from the former.

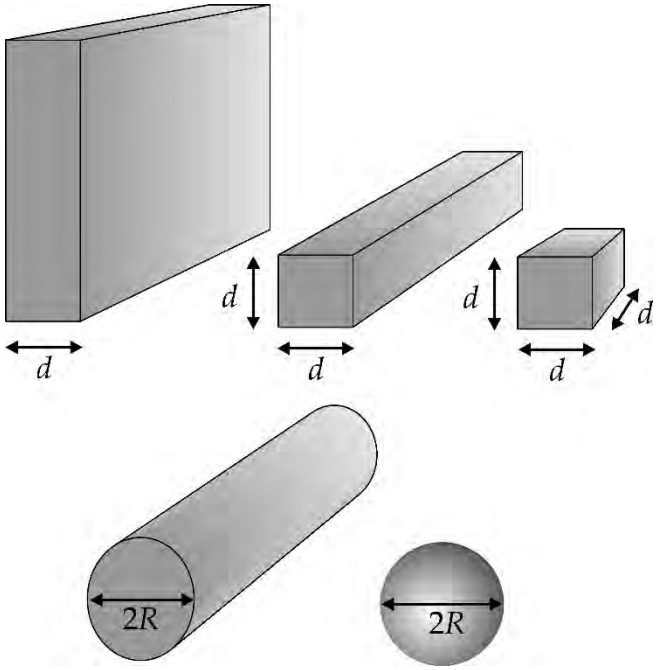


Figure A1.1. A selection of nanostructures. Top from left to right: quantum well, wire and dot with rectangular confinement. Bottom from left to right: quantum wire and dot with circular confinement.

A1.1 Quantum Wells

Quantum wells are 2D structures, for which the bound states are characterized by standing waves in the confined directions and running in the other two. Taking the confinement to be along z we can write the Hamiltonian in the effective mass approximation as

$$\hat{H} = -\frac{\hbar^2 \nabla^2}{2m_e} + V(z)$$

with eigenstates

$$\psi_{n\vec{k}}(\vec{r}) = \frac{1}{\sqrt{A}} e^{i\vec{k}\cdot\vec{r}} \varphi_n(z), \quad \vec{k} = (k_x, k_y, 0). \tag{A1.1}$$

Here, A the area of the sample viewed head-on. The 2D wave vector is $\vec{k}$ and the standing wave is labeled by the discrete index n. The total energy is then $E_{n\vec{k}} = \hbar^2 k^2 / 2m_e + E_n^e$ with E_n^e determined by the eigenvalue problem

$$\left\{ -\frac{\hbar^2}{2m_e} \frac{d^2}{dz^2} + V(z) \right\} \varphi_n(z) = E_n^e \varphi_n(z).$$

In the case of an infinite step potential,

$$V(z) = \begin{cases} 0 & 0 < z < d \\ \infty & \text{otherwise} \end{cases}$$

the solution is simply

$$\varphi_n(z) = \sqrt{\frac{2}{d}} \sin\left(\frac{n\pi z}{d}\right), \quad E_n^e = \frac{\hbar^2}{2m_e}\left(\frac{n\pi}{d}\right)^2, \quad n = 1, 2, 3, ...$$

A more realistic model is that of a finite step

$$V(z) = \begin{cases} 0 & -d/2 < z < d/2 \\ V_0 & \text{otherwise,} \end{cases}$$

where V_0 is the barrier height. Note that we shifted the origin to the center of the well. In this case, the bound states are sine/cosine inside the well and exponentially decaying outside. Due to the inversion symmetry of the potential, the states are either even or odd functions of z. Hence, we can write

$$\text{even:} \quad \varphi_n(z) = A_{even} \begin{cases} \cos(k_n z) & -d/2 < z < d/2 \\ \cos(k_n \frac{d}{2}) \exp\{-q_n(|z| - d/2)\} & \text{otherwise,} \end{cases}$$

$$\text{odd:} \quad \varphi_n(z) = A_{odd} \begin{cases} \sin(k_n z) & -d/2 < z < d/2 \\ \text{sign}(z) \sin(k_n \frac{d}{2}) \exp\{-q_n(|z| - d/2)\} & \text{otherwise.} \end{cases}$$

In both cases, the energy is given by $E_n^e = \hbar^2 k_n^2 / 2m_e = V_0 - \hbar^2 q_n^2 / 2m_e$. This relation provides a link between k_n and q_n. Note that even or odd does not refer to the value of n but simply to the dependence on z. The purpose of writing the exponentials with the above constant prefactors is to automatically ensure continuity. Requiring, in addition, continuity of the derivative leads to the trancendental equations

$$\text{even:}\quad k_n \sin(k_n \tfrac{d}{2}) = q_n \cos(k_n \tfrac{d}{2})$$
$$\text{odd:}\quad k_n \cos(k_n \tfrac{d}{2}) = -q_n \sin(k_n \tfrac{d}{2}).$$

Moreover, the normalization constant can be written as simple analytical expressions

$$\text{even:}\quad A_{even}^2 = 2k_n q_n / \left[k_n(1+q_n d) + k_n \cos(k_n d) + q_n \sin(k_n d)\right]$$
$$\text{odd:}\quad A_{odd}^2 = 2k_n q_n / \left[k_n(1+q_n d) - k_n \cos(k_n d) - q_n \sin(k_n d)\right].$$

As an example, we consider a 300 Å wide GaAs quantum well embedded in $Al_xGa_{1-x}As$ (x = 0.3) barriers. For electron states in this system, we may take $V_0 \approx 0.243$ eV and $m_e = 0.067 m_0$. The squared quantum states are shown in Fig. A1.2, where the vertical displacement corresponds to the energy.

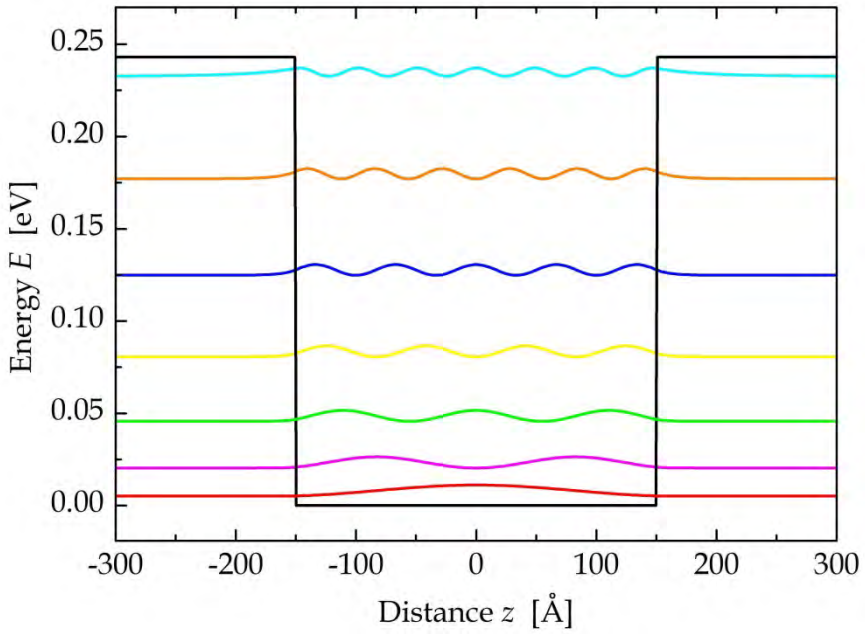


Figure A1.2. Squared eigenstates of a 300 Å GaAs quantum well. The states are shown displaced above zero by their energy.

We want to compute the density of states for various low-dimensional structures. These can then be compared to the usual 3D result $D_3(E) = \sqrt{2} m_e^{3/2} / (\pi^2 \hbar^3) \sum_n (E - E_n)^{1/2} \theta(E - E_n)$, where the sum is over all bands with band edges E_n. For our 2D system, we now calculate the density of states per volume with the volume given by $\Omega = Ad$. Hence, we write

$$D_2(E) = \frac{2}{Ad} \sum_{n,\vec{k}} \delta(E_{n\vec{k}} - E),$$

where the factor of 2 is for spin. Converting to an integral, we find

$$D_2(E) = \frac{1}{2\pi^2 d}\sum_n \int \delta\left(\frac{\hbar^2 k^2}{2m_e} + E_n^e - E\right) d^2k\,,$$

To evaluate the integral, we use $x = \hbar^2 k^2 / 2m_e$ so $d^2k = 2\pi k dk = \pi d(k^2) = \pi \frac{2m_e}{\hbar^2} dx$. In turn,

$$D_2(E) = \frac{m_e}{d\pi\hbar^2}\sum_n \int \delta\left(x + E_n^e - E\right) dx = \frac{m_e}{d\pi\hbar^2}\sum_n \theta(E - E_n^e)\,, \qquad \text{(A1.2)}$$

where $\theta(E)$ is the Heaviside step function. Thus, in 2D the density of states is a series of steps of height $m_e / d\pi\hbar^2$. This result is illustrated in Fig. A1.3.

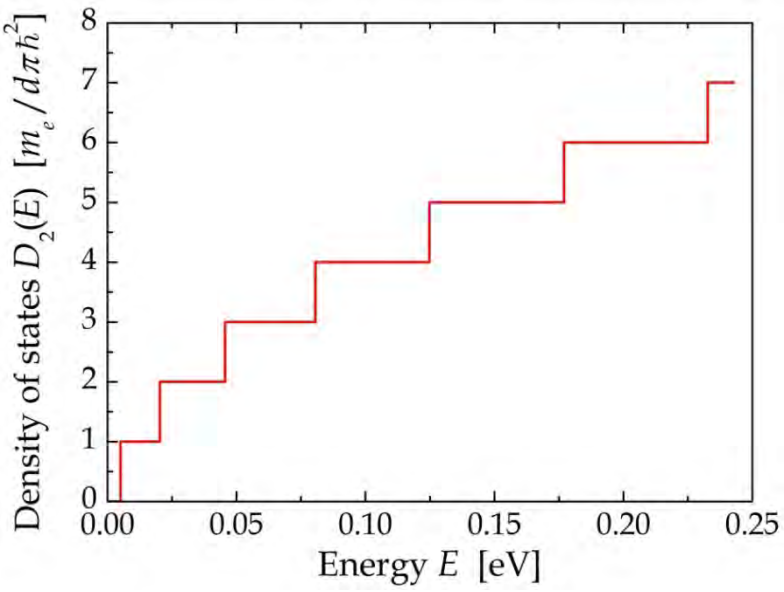


Figure A1.3. Density of states for the 300 Å GaAs quantum well shown in Fig. A1.2.

A1.2 Quantum Wires

One-dimensional systems with the extended direction along z are governed by a Hamiltonian

$$\hat{H} = -\frac{\hbar^2\nabla^2}{2m_e} + V(x,y)$$

and have eigenstates (for a wire of length L)

$$\psi_{nk}(\vec{r}) = \frac{1}{\sqrt{L}} e^{ikz}\varphi_n(x,y)\,. \qquad \text{(A1.3)}$$

Again, the simplest case is an infinite step potential of the sort

$$V(x,y)=\begin{cases}0 & 0<x,y<d\\ \infty & \text{otherwise.}\end{cases}$$

The eigenstates in this case are double copies of the corresponding quantum well state

$$\varphi_n(x,y)=\frac{2}{d}\sin\left(\frac{n_x\pi x}{d}\right)\sin\left(\frac{n_y\pi y}{d}\right),\quad E_n^e=\frac{\hbar^2}{2m_e}\left(\frac{n_x\pi}{d}\right)^2+\frac{\hbar^2}{2m_e}\left(\frac{n_y\pi}{d}\right)^2,\quad n_x,n_y=1,2,3,....$$

Hence, n should be regarded as a composite index $n=\{n_x,n_y\}$. There exists a related model that is a more realistic description of many vapor-grown quantum wires. Such wires (Fig. A1.1, bottom) typically have a roughly circular cross section and using polar coordinates (r,ϕ) a better model for the confinement is

$$V(r)=\begin{cases}0 & 0<r<R\\ \infty & \text{otherwise.}\end{cases}$$

Hence, the potential has cylindrical symmetry and the eigenstates are separable as $\varphi_{m,n}(r,\phi)=R_{m,n}(r)e^{im\phi}/\sqrt{2\pi}$. For $r<R$ the radial functions are determined by

$$-\frac{\hbar^2}{2m_e}\left\{\frac{d^2}{dr^2}+\frac{1}{r}\frac{d}{dr}-\frac{m^2}{r^2}\right\}R_{m,n}(r)=E_{m,n}^eR_{m,n}(r).$$

Introducing the radial wave number $k_{m,n}=\sqrt{2m_eE_{m,n}^e/\hbar^2}$, the problem turns into Bessel's equation

$$r^2R_{m,n}''(r)+rR_{m,n}'(r)+(k_{m,n}^2r^2-m^2)R_{m,n}(r)=0$$

with the solution

$$R_{m,n}(r)=N_{m,n}J_m(k_{m,n}r)$$

in which J_m is the m'th Bessel function. The boundary condition $R_{m,n}(R)=0$ means that $k_{m,n}=\lambda_{m,n}/R$, where $\lambda_{m,n}$ is the n'th zero-point (root) of the m'th Bessel function, i.e. $J_m(\lambda_{m,n})=0$. For the 0'th Bessel function these are $\lambda_{0,n}=\{2.405,5.520,8.654,...\}$. The full, normalized solution becomes

$$R_{m,n}(r)=\frac{\sqrt{2}J_m(\lambda_{m,n}r/R)}{RJ_{m+1}(\lambda_{m,n})}.$$

In complete analogy with the 2D case, we find a 1D density of states given by

$$D_1(E) = \frac{1}{A\pi}\sum_n \int \delta\left(\frac{\hbar^2 k^2}{2m_e} + E_n^e - E\right) dk.$$

Here, A is the cross sectional area of the wire and the sum is over all standing wave modes. By introducing again $x = \hbar^2 k^2 / 2m_e$ we find

$$D_1(E) = \frac{\sqrt{2m_e}}{A\pi\hbar}\sum_n \frac{\theta(E - E_n^e)}{\sqrt{E - E_n^e}}, \tag{A1.4}$$

This result shows that the density of states is a series of inverse square roots and, in fact, diverges precisely at the quantization energies. This is illustrated for a cylindrical GaAs quantum wire (in vacuum) with R = 100 Å in Fig. A1.4.

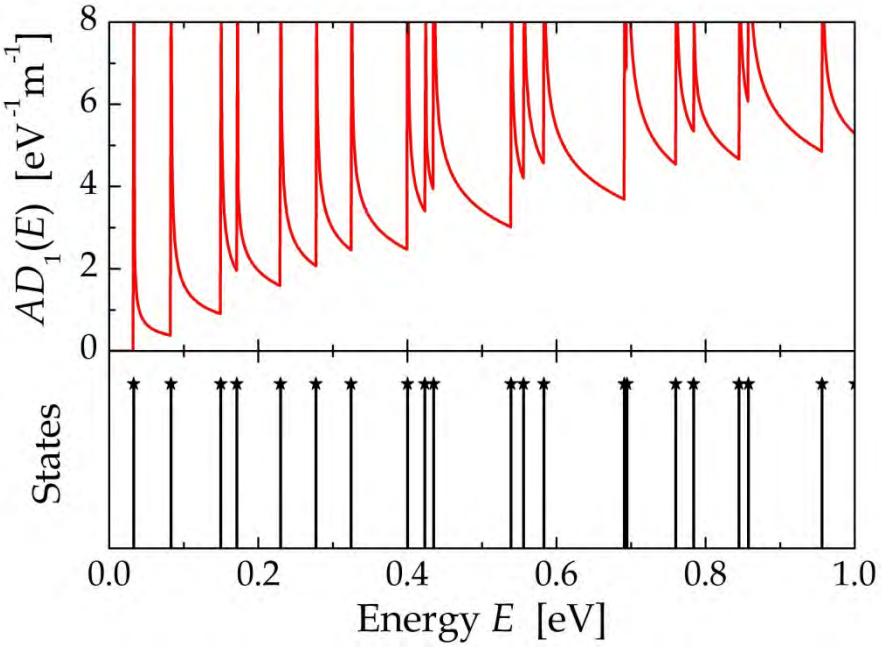


Figure A1.4. Eigenvalue location and density of states for an R = 100 Å GaAs quantum wire.

A1.3 Quantum Dots

If the confinement is present in all spatial direction, the nanostructure is known as a quantum dot. The simplest case is again infinite step confinement

$$V(x,y,z) = \begin{cases} 0 & 0 < x,y,z < d \\ \infty & \text{otherwise} \end{cases}$$

with fully confined bound states

$$\varphi_n(x,y,z)=\left(\frac{2}{d}\right)^{3/2}\sin\left(\tfrac{n_x\pi x}{d}\right)\sin\left(\tfrac{n_y\pi y}{d}\right)\sin\left(\tfrac{n_z\pi z}{d}\right),\quad E_n^e=\frac{\hbar^2\pi^2}{2m_e d^2}\left(n_x^2+n_y^2+n_z^2\right),\quad n_x,n_y,n_z=1,2,3,....$$

But so-called colloidal quantum dots are typically nearly spherical (Fig. A1.1, bottom) so that a realistic confinement in spherical coordinates (r,θ,ϕ) is

$$V(r)=\begin{cases}0 & 0<r<R\\ \infty & \text{otherwise}\end{cases}$$

just like for the cylindrical wire. In this case, the angular part of the eigenstates is a spherical harmonic $Y_{lm}(\theta,\phi)$ and the radial part $R_{l,n}(r)$ for $r<R$ is determined by

$$-\frac{\hbar^2}{2m_e}\left\{\frac{d^2}{dr^2}+\frac{2}{r}\frac{d}{dr}-\frac{l(l+1)}{r^2}\right\}R_{l,n}(r)=E_{l,n}^e R_{l,n}(r).$$

Now, the eigenfunctions are spherical Bessel functions denoted j_l with roots $\beta_{l,n}$, i.e. $j_l(\beta_{l,n})=0$. In full, we find

$$R_{l,n}(r)=\frac{\sqrt{2}j_l(\beta_{l,n}r/R)}{R^{3/2}j_{l+1}(\beta_{l,n})}.$$

The 1D density of states is simply a series of delta-function peaks located at the discrete energy eigenvalues $E_{l,n}=\hbar^2\beta_{l,n}^2/(2m_eR^2)$.

A1.4 Carbon Nanostructures

The previous examples in this Appendix were semiconductor structures confined by square barriers and described within the effective mass approximation. Such "traditional" nanostructures have recently been supplemented by a host of carbon-based ones. These are either planar molecules or hollow geometries formed by rolling up sheets. The class of hollow carbon nanostructures famously includes Fullerenes and nanotubes. If curvature effects can be ignored, the dominant states are formed by π- orbitals localized on the different atoms. In a simple picture, only nearest-neighbor atomic π- orbitals couple, which allows for analytic solutions in many cases. Here, we will describe the geometries and properties of C_{60} Fullerenes and carbon nanotubes.

The C_{60} molecule shown in Fig. A1.5 looks complicated but has many symmetries simplifying the geometry. The basic pattern is that of pentagons surrounded by five hexagons. The six atoms highlighted in red contain one from the topmost pentagon plus an entire additional pentagon. The locations of the remaining 24 atoms in the top half ($z>0$) are found by rotations by $2\pi/5$ around the z-axis. Finally, the 30

atoms in the lower half follow from reflection through the center, i.e. $\vec{r}_{61-i} = -\vec{r}_i$ with $i = 1...30$. In units of the sphere radius R divided by $\sqrt{1090}$, the coordinates of the red atoms can be taken as

Atom:	1	2	3	4	5	6
x	$2a_1$	$4a_1$	a_4	a_4	a_6	a_6
y	0	0	$-a_5$	a_5	$-a_7$	a_7
z	a_2	a_3	$3a_1$	$3a_1$	a_1	a_1

$$a_1^2 = 50 - 8\sqrt{5}, \quad a_2^2 = 890 + 32\sqrt{5}, \quad a_3^2 = 290 + 128\sqrt{5}, \quad a_4^2 = 535 + 67\sqrt{5},$$
$$a_5^2 = 105 + 5\sqrt{5}, \quad a_6^2 = 895 + 53\sqrt{5}, \quad a_7^2 = 145 - 45\sqrt{5}.$$

All coordinates should be scaled to the ratio between nearest-neighbor distance and radius $a^2 / R^2 = (58 - 18\sqrt{5}) / 109 \approx 0.4035^2$.

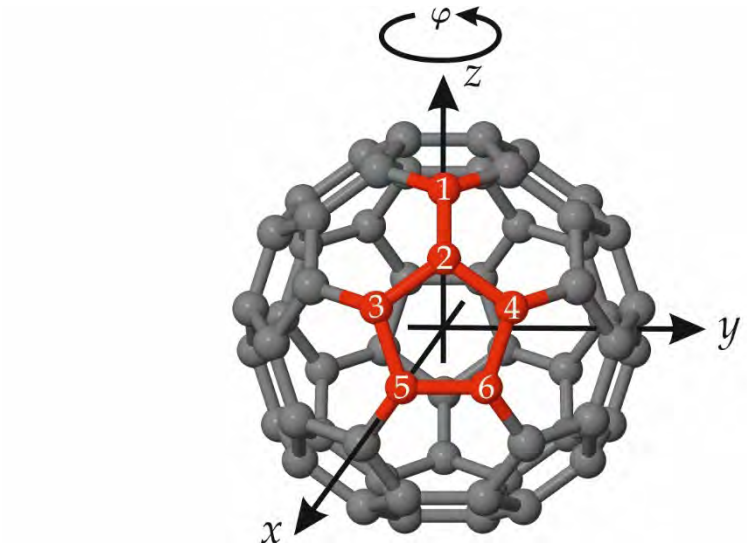


Figure. A1.5. Geometry of the C_{60} molecule with the 6-atom basic motif shown in red.

The tight-binding model of electronic structure in carbon-based material is described in detail in appendix 2. Here, we will apply this model keeping only a single π-orbital per atom. Also, only nearest-neighbors are assumed to interact as described by the hopping matrix element γ. Hence, the full Hamiltonian in the π- orbital model is a 60×60 matrix $\vec{\vec{H}}$. In can, however, be block-diagonalized into ten 6×6 blocks using the symmetries above. Thus, suppose the full eigenvector is $\vec{u}$ with 60 elements out of which the first six form a separate vector $\vec{v}$. We will apply a super-vector notation $\vec{u}_{qp} = (\vec{v}, \vec{v}_2, \vec{v}_3, \vec{v}_4, \vec{v}_5, \vec{w}_1, \vec{w}_2, \vec{w}_3, \vec{w}_4, \vec{w}_5)_{qp}$. Here, $q = 0...4$ describes rotational phase around the z-axis and $p = \pm$ the parity upon reflection through the center. With $\hat{R}$ designating such reflection, the symmetries then mean that

$$\vec{v}_k = \exp(iqk\tfrac{2\pi}{5})\vec{v}, \quad \vec{w}_k = p\exp(iqk\tfrac{2\pi}{5})\hat{R}(\vec{v}).$$

In matrix notation, $\vec{u}_{qp} = \overleftrightarrow{M}_{qp} \cdot \vec{v}$ with

$$\overleftrightarrow{M}_{qp} = \left\{ \overleftrightarrow{I}, \exp(iq\tfrac{2\pi}{5})\overleftrightarrow{I}, ..., p\overleftrightarrow{R}, p\exp(iq\tfrac{2\pi}{5})\overleftrightarrow{R}, ... \right\}^T, \quad \overleftrightarrow{I} = \begin{pmatrix} 1 & 0 & \\ 0 & \ddots & 0 \\ & 0 & 1 \end{pmatrix}, \quad \overleftrightarrow{R} = \begin{pmatrix} & 0 & 1 \\ 0 & \iddots & 0 \\ 1 & 0 & \end{pmatrix}.$$

The ten blocks are then found as $\overleftrightarrow{H}_{qp} = \overleftrightarrow{M}_{qp}^{-1} \cdot \overleftrightarrow{H} \cdot \overleftrightarrow{M}_{qp}$. The eigenvalues found in this way are shown in Fig. A1.6. It is clearly seen that the lowest ones follow the $2l+1$ sequence expected for angular momenta on a sphere: 1, 3, 5, …For $l \geq 3$, levels split because the symmetry in only approximately spherical. Thus, the seven $l = 3$ states split into blocks of 3 and 4. Also, the $l = 5$ block splits into blocks of 5, 3, and 3, while the $l = 4$ levels remain intact. The splitting can be partially understood from the coupling of spherical harmonic by the potential. Thus, the lowest harmonic compatible with the symmetry (apart from Y_{00}) can be shown to be

$$V(\theta,\varphi) \propto \sqrt{\frac{11}{7}} Y_{50}(\theta,\varphi) - iY_{55}(\theta,\varphi) - iY_{5,-5}(\theta,\varphi).$$

This form is shown as the inset in Fig. A1.6. Adding a potential of this form, leads to precisely the correct splitting of the $l = 3$ and $l = 5$ blocks, although not the energetic ordering.

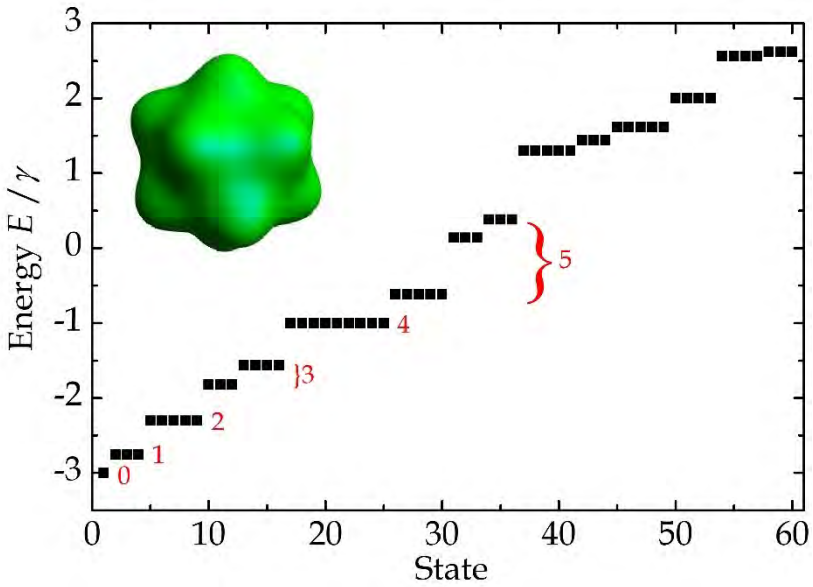


Figure. A1.6. Tight-binding energies of C_{60} in units of the hopping integral. The approximate angular momenta are indicated in red and the perturbing potential is shown in the inset.

An extremely important class of carbon nanostructures is nanotubes. These come in a wide range of diameters and chiralities ("twists") that are essential for their properties. Also, individual nanotubes (single-walled carbon nanotubes) as well as

multi-walled nanotubes (tubes within tubes) exist. Here, we only describe the former. Although this is not how they are actually formed experimentally, it is extremely useful to picture them as rolled up strips of graphene, as shown in Fig. A1.7. In the figure, the highlighted rectangle forms the nanotube unit cell after being rolled into a cylinder. The long-axis of the cylinder is along $\vec{T}$, which is the lattice vector of the infinite tube formed by joining unit cells along the cylinder. The perpendicular vector $\vec{C}$ is the "roll-up" vector that becomes the circumference after rolling up. Note that the atoms located in all four corners of the rectangle are equivalent (belong to the same sublattice). Thus, upon rolling, the atoms at the origin and the one displaced by $\vec{C}$ will coincide. Hence, atoms located on the sides opposite $\vec{C}$ and $\vec{T}$ are not included in the unit cell. All vectors pointing to equivalent atoms can be expressed as integer multiples of the graphene lattice vectors $\vec{a}_1$ and $\vec{a}_2$. Thus, importantly, we write $\vec{C} = n\vec{a}_1 + m\vec{a}_2$ defining the *chiral indices* n and m. In this way, any particular nanotube can be specified by (n,m) and, for instance, the one shown in Fig. A1.7 is (6,2). A simple geometric exercise demonstrates that

$$\vec{T} = -\frac{2m+n}{g}\vec{a}_1 + \frac{2n+m}{g}\vec{a}_2, \quad g = \text{greatest common divisor}(2m+n, 2n+m).$$

This is the shortest lattice vector perpendicular to $\vec{C}$. It may also the shown that $C = La$ and $T = \sqrt{3}La/g$ with $L = \sqrt{n^2+m^2+nm}$. Thus, the number of hexagons (half the number of atoms) in the unit cell is $N = 2L^2/g$ and the angle θ between $\vec{C}$ and the x-axis can be expressed as

$$\cos\theta = \frac{\sqrt{3}(n+m)}{2L}, \quad \sin\theta = \frac{n-m}{2L}.$$

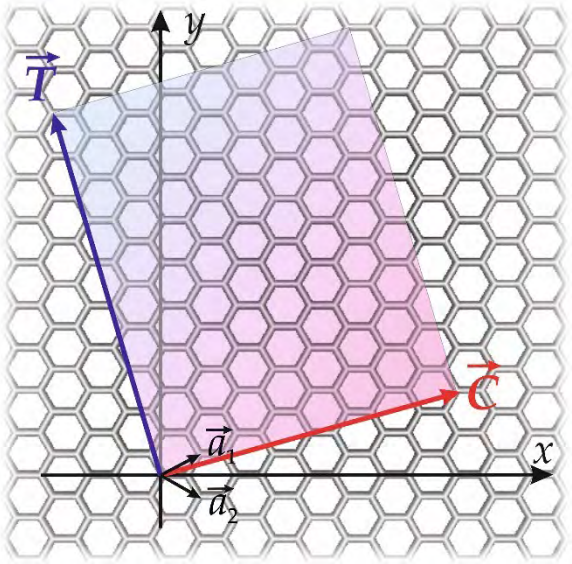


Figure. A1.7. Nanotube unit cell defined as a rectangular piece of graphene that is subsequently rolled into a cylinder.

The number of atoms in the unit cell easily runs into the hundreds. This means that accurate models of the electronic structure are generally rather demanding. However, if the curvature of the nanotube is ignored, the model is exceedingly simple. Thus, the band structure of a nanotube is *exactly* the same as for a graphene sheet except that additional conditions must be imposed on the allowable *k*-vectors. Rolling up a strip of graphene amounts to identifying the atoms on the right side with those on the left side. This corresponds to imposing periodic boundary conditions in the $\vec{C}$ direction. As a consequence, the band structure of a nanotube can be described by the two-dimensional *k* vector $\vec{k}$ but with the condition $\vec{k}\cdot\vec{C} = 2p\pi$ with *p* an integer. This is technically known as *zone folding*. We now write $\vec{k} = k\vec{e}_T + q\vec{e}_C$ in terms of unit vectors along $\vec{C}$ and $\vec{T}$. The $\vec{T}$ component *k* is continuous and restricted by $-\pi/T < k < \pi/T$, while $q = 2p\pi/C$ is discrete with $q = 1,...,N$. Eventually, this means that the nanotube band structure consists of *N* bands derived from each graphene band $E(\vec{k})$ as

$$E_p(k) = E(k_x^{(p)}, k_y^{(p)}), \quad k_x^{(p)} = k\cos\theta - p(2\pi/L)\sin\theta, \quad k_y^{(p)} = k\sin\theta + p(2\pi/L)\cos\theta\,.$$

In all π - orbital carbon models, the lower half of the bands are occupied. Below, we illustrate the results for (5,0) and (5,5) nanotubes. It is seen that the former is a semiconductor, while the latter is metallic. In fact, all geometries, for which $n-m$ is divisible by 3, are metals and the rest are semiconductors. The reason is that structures with $n-m = 0,3,6,...$ have a *k*-point hitting the $\vec{K}$-point of the graphene band structure.

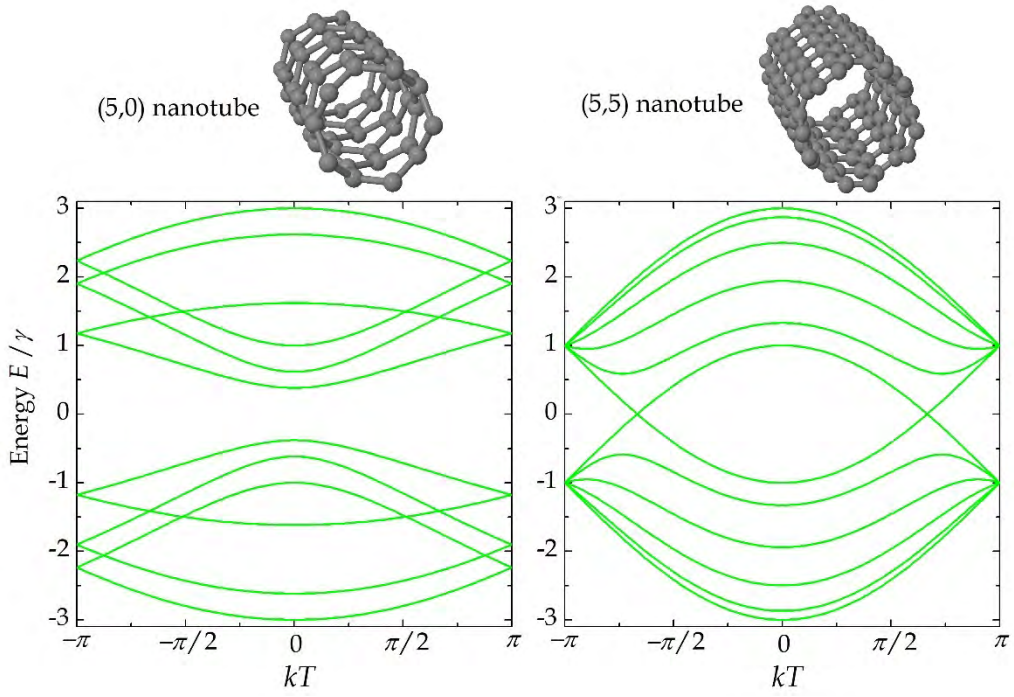


Figure. A1.8. Band structures of (5,0) and (5,5) carbon nanotubes calculated from zone-folding of the graphene bands.

Appendix 2. Tight-Binding Formalism

This appendix describes the tight-binding or LCAO (linear combination of atomic orbitals) formalism applied to simple molecules and periodic systems. We initially study a finite molecule and set up the corresponding finite matrix equation. Then, the analysis is extended to infinitely periodic structures in 1, 2, and 3 dimensions. Thus, we begin by considering a DFT-type Hamiltonian that can be written as a sum of single-electron terms:

$$H_{total} = \sum_{n=1}^{N} H(\vec{r}_n),$$

where $\vec{r}_n$ is the position of the n'th electron. The idea, now, is to expand the eigenstates in a basis of localized "atomic" orbitals $\varphi_i(\vec{r} - \vec{R}_\alpha)$. Here, i denotes the symmetry of the orbital i.e. $i \in \{s, p_x, p_y, p_z, d_{xy}, d_{xz}, ...\}$ as illustrated in Fig. A2.1 and α labels the atoms so that $\vec{R}_\alpha$ is the position of the α'th atom (nucleus). Correspondingly, we expand the single-electron wave function as follows

$$\left|\psi(\vec{r})\right\rangle = \sum_{i,\alpha} c_{i\alpha} \left|\varphi_i(\vec{r} - \vec{R}_\alpha)\right\rangle. \tag{A2.1}$$

We then need the following matrix elements:

$$H_{i\alpha,j\beta} = \left\langle \varphi_i(\vec{r} - \vec{R}_\alpha) \middle| H(\vec{r}) \middle| \varphi_j(\vec{r} - \vec{R}_\beta) \right\rangle, \quad S_{i\alpha,j\beta} = \left\langle \varphi_i(\vec{r} - \vec{R}_\alpha) \middle| \varphi_j(\vec{r} - \vec{R}_\beta) \right\rangle \tag{A2.2}$$

to set up the matrix equation

$$\left(\overleftrightarrow{H} - E\overleftrightarrow{S}\right) \cdot \vec{c} = 0,$$

where $\vec{c}$ is a vector containing the expansion coefficients $c_{i\alpha}$. As an example, we will look at a two-atomic molecule described with an sp^3 basis. This very important case will be discussed in detail. For each atom we take an *s*-orbital and the three *p*-orbitals $\{p_x, p_y, p_z\}$. We may take the vector $\vec{R}$ joining the two atoms to point along the *x*-axis. With this convention, the *s* and p_x orbitals belong to the so-called σ states that do not change under a rotation around the molecular bond. In contrast, p_y and p_z change from plus to minus and back to plus under a full rotation. These states are therefore called π orbitals. Collectively, we will refer to p_x as p_σ and to p_y and p_z as p_π.

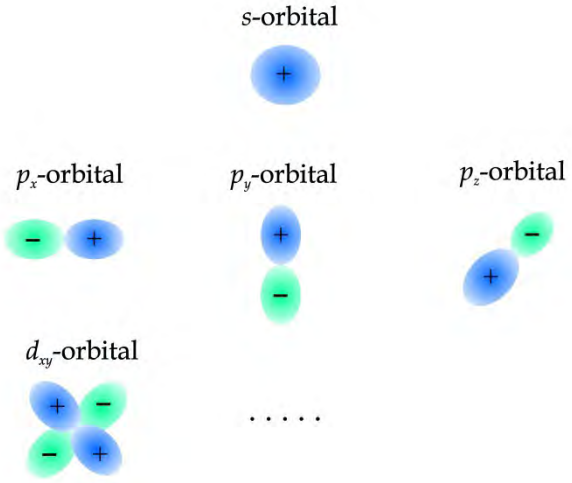


Figure A2.1. The lowest atomic orbitals and their symmetry.

First, we need the Hamilton matrix elements for orbitals on the same site, the so-called *on-site* matrix elements. We will assume that the atoms are identical. Also, as an approximation, the orthogonality between s, p_x, p_y and p_z will allow us to neglect on-site term involving orbitals of different symmetry. Finally, differences between the radial parts of p_σ and p_π will be ignored. It follows that the only non-zero on-site terms are

$$\begin{aligned} E_s &\equiv \langle s(\vec{r})|H(\vec{r})|s(\vec{r})\rangle = \langle s(\vec{r}-\vec{R})|H(\vec{r})|s(\vec{r}-\vec{R})\rangle \\ E_p &\equiv \langle p_i(\vec{r})|H(\vec{r})|p_i(\vec{r})\rangle = \langle p_i(\vec{r}-\vec{R})|H(\vec{r})|p_i(\vec{r}-\vec{R})\rangle. \end{aligned} \tag{A2.3}$$

In addition to the on-site elements, we need the Hamilton matrix elements for orbitals centred on *different* atoms. From symmetry considerations we find

$$\begin{gathered} H_{ss\sigma} \equiv \langle s(\vec{r})|H(\vec{r})|s(\vec{r}-\vec{R})\rangle, \quad H_{sp\sigma} \equiv \langle s(\vec{r})|H(\vec{r})|p_\sigma(\vec{r}-\vec{R})\rangle, \quad H_{ps\sigma} = -H_{sp\sigma} \\ H_{pp\sigma} \equiv \langle p_\sigma(\vec{r})|H(\vec{r})|p_\sigma(\vec{r}-\vec{R})\rangle, \quad H_{pp\pi} \equiv \langle p_\pi(\vec{r})|H(\vec{r})|p_\pi(\vec{r}-\vec{R})\rangle \\ \langle s(\vec{r})|H(\vec{r})|p_\pi(\vec{r}-\vec{R})\rangle = 0, \quad \langle p_\sigma(\vec{r})|H(\vec{r})|p_\pi(\vec{r}-\vec{R})\rangle = 0 \end{gathered} \tag{A2.4}$$

and similarly for the overlap matrix elements. As an approximation, the potential energy is often written as a sum of spherical terms centered at each atom. This leads to very simple angular dependencies for the matrix elements. The general results can be found in Ref. [1]. We can now set up the Hamilton matrix using the following basis:

$$|s(\vec{r})\rangle, |p_x(\vec{r})\rangle, |p_y(\vec{r})\rangle, |p_z(\vec{r})\rangle, |s(\vec{r}-\vec{R})\rangle, |p_x(\vec{r}-\vec{R})\rangle, |p_y(\vec{r}-\vec{R})\rangle, |p_z(\vec{r}-\vec{R})\rangle$$

which leads to

$$\overleftrightarrow{H}=\begin{pmatrix} E_s & 0 & 0 & 0 & H_{ss\sigma} & H_{sp\sigma} & 0 & 0 \\ 0 & E_p & 0 & 0 & -H_{sp\sigma} & H_{pp\sigma} & 0 & 0 \\ 0 & 0 & E_p & 0 & 0 & 0 & H_{pp\pi} & 0 \\ 0 & 0 & 0 & E_p & 0 & 0 & 0 & H_{pp\pi} \\ H_{ss\sigma} & -H_{sp\sigma} & 0 & 0 & E_s & 0 & 0 & 0 \\ H_{sp\sigma} & H_{pp\sigma} & 0 & 0 & 0 & E_p & 0 & 0 \\ 0 & 0 & H_{pp\pi} & 0 & 0 & 0 & E_p & 0 \\ 0 & 0 & 0 & H_{pp\pi} & 0 & 0 & 0 & E_p \end{pmatrix}.$$

The $\overleftrightarrow{S}$ matrix looks perfectly similar, apart from the changes that it has 1 everywhere in the diagonal and that $H_{ss\sigma}$ should be replaced by $S_{ss\sigma}$ and so on. If we ignore overlap, i.e. set $\overleftrightarrow{S}$ equal to the unit matrix, we find a standard eigenvalue problem. Four of the eight eigenvalues are then $E=E_p\pm H_{pp\pi}$ each of which is two-fold degenerate. These states are clearly π states. The remaining eigenvalues belong to the σ state sub-matrix

$$\overleftrightarrow{H}_\sigma=\begin{pmatrix} E_s & 0 & H_{ss\sigma} & H_{sp\sigma} \\ 0 & E_p & -H_{sp\sigma} & H_{pp\sigma} \\ H_{ss\sigma} & -H_{sp\sigma} & E_s & 0 \\ H_{sp\sigma} & H_{pp\sigma} & 0 & E_p \end{pmatrix}$$

with eigenvalues

$$E=\frac{1}{2}\begin{cases} E_s+E_p+H_{ss\sigma}-H_{pp\sigma}\pm\sqrt{\left(E_s-E_p+H_{ss\sigma}+H_{pp\sigma}\right)^2+4H_{sp\sigma}^2} \\ E_s+E_p-H_{ss\sigma}+H_{pp\sigma}\pm\sqrt{\left(E_p-E_s+H_{ss\sigma}+H_{pp\sigma}\right)^2+4H_{sp\sigma}^2} \end{cases}.$$

As a numerical example, we can consider the C-C molecule with parameters [2]:

Energy integrals [eV]	Overlap integrals
$E_s=-8.868,\ E_p=0$	$S_s=1,\ S_p=1$
$H_{ss\sigma}=-6.769$	$S_{ss\sigma}=0.212$
$H_{sp\sigma}=5.580$	$S_{sp\sigma}=-0.102$
$H_{pp\sigma}=5.037$	$S_{pp\sigma}=-0.146$
$H_{pp\pi}=-3.033$	$S_{pp\pi}=0.129$

and the corresponding eigenvalues

$$E = \{\text{-18.03 eV, -5.15 eV, -3.03 eV, -3.03 eV, -2.64 eV, 3.03 eV, 3.03eV, 8.09 eV}\}.$$

To accommodate a total of eight valence electrons, the four lowest states are all doubly occupied. Hence, in usual molecular notation [3] we can write the electronic configuration as $(\sigma_g 2s)^2(\sigma_u 2s)^2(\pi_u 2p)^4$, where subscripts g and u mean even ("gerade") and odd ("ungerade") under inversion through the center of the molecule, respectively.

If the vector $\vec{R}$ does not coincide with the x-axis, the formulas are a little more complicated. In general, we can write $\vec{R} = R(l,m,n)$, such that l, m, and n are the Cartesian components of the unit vector. We then find that

$$\begin{aligned} &\left\langle s(\vec{r})\middle|H(\vec{r})\middle|s(\vec{r}-\vec{R})\right\rangle = H_{ss\sigma}, \quad \left\langle s(\vec{r})\middle|H(\vec{r})\middle|p_x(\vec{r}-\vec{R})\right\rangle = lH_{sp\sigma} \\ &\left\langle p_x(\vec{r})\middle|H(\vec{r})\middle|p_x(\vec{r}-\vec{R})\right\rangle = l^2H_{pp\sigma} + (1-l^2)H_{pp\pi}, \left\langle p_x(\vec{r})\middle|H(\vec{r})\middle|p_y(\vec{r}-\vec{R})\right\rangle = lm(H_{pp\sigma} - H_{pp\pi}). \end{aligned} \tag{A2.5}$$

The remaining elements can be found by symmetry.

A2.1 Tight-Binding in Periodic Structures

We now extend the analysis to infinitely periodic structures characterized by unit cells that are repeated periodically in 1, 2, or 3 dimensions. Consider a linear combination of *identical* orbitals:

$$\left|\chi_i(\vec{r})\right\rangle = \sum_{\vec{R}} c_{\vec{R}} \left|\varphi_i(\vec{r}-\vec{R})\right\rangle, \tag{A2.6}$$

where $\vec{R}$ denotes the location of the unit cell to which the orbital belongs. Note that $\vec{R}$ is also used as a summation index. The Bloch theorem requires that $\left|\chi_i(\vec{r}+\vec{R}')\right\rangle = e^{i\vec{k}\cdot\vec{R}'}\left|\chi_i(\vec{r})\right\rangle$. This leads to the condition

$$\begin{aligned} &\sum_{\vec{R}} c_{\vec{R}} \left|\varphi_i(\vec{r}+\vec{R}'-\vec{R})\right\rangle = e^{i\vec{k}\cdot\vec{R}'}\sum_{\vec{R}} c_{\vec{R}} \left|\varphi_i(\vec{r}-\vec{R})\right\rangle \\ &\Downarrow \\ &\sum_{\vec{R}''} c_{\vec{R}''+\vec{R}'} \left|\varphi_i(\vec{r}-\vec{R}'')\right\rangle = e^{i\vec{k}\cdot\vec{R}'}\sum_{\vec{R}} c_{\vec{R}} \left|\varphi_i(\vec{r}-\vec{R})\right\rangle, \end{aligned}$$

where we have introduced $\vec{R}'' = \vec{R} - \vec{R}'$ so that $\vec{R} = \vec{R}'' + \vec{R}'$. If, finally, we "relabel"

the summation index from $\vec{R}''$ to $\vec{R}$ in the last line we find $c_{\vec{R}+\vec{R}'} = e^{i\vec{k}\cdot\vec{R}'} c_{\vec{R}}$ or, equivalently, $c_{\vec{R}} = e^{i\vec{k}\cdot\vec{R}} c_{\vec{0}}$. The constant $c_{\vec{0}}$ can be chosen as $c_{\vec{0}} = 1/\sqrt{N}$, where N is the number of unit cells. A sum like Eq.(A2.6) is called a "Bloch sum" because it fulfils the Bloch condition. If we have a number of different atomic orbitals in the unit cell we can write the eigenstates as linear combinations of Bloch sums, i.e.

$$\left|\psi(\vec{r})\right\rangle = \sum_i c_i \left|\chi_i(\vec{r})\right\rangle \tag{A2.7}$$

and we still get a matrix equation $\left(\overleftrightarrow{H} - E\overleftrightarrow{S}\right)\cdot\vec{c} = 0$, where now

$$H_{ij} = \left\langle \chi_i(\vec{r})\middle|H(\vec{r})\middle|\chi_j(\vec{r})\right\rangle, \quad S_{ij} = \left\langle \chi_i(\vec{r})\middle|\chi_j(\vec{r})\right\rangle.$$

Using the Bloch sum expressions we find

$$H_{ij} = \frac{1}{N}\sum_{\vec{R},\vec{R}'} e^{i\vec{k}\cdot(\vec{R}-\vec{R}')} \left\langle \varphi_i(\vec{r}-\vec{R}')\middle|H(\vec{r})\middle|\varphi_j(\vec{r}-\vec{R})\right\rangle$$
$$S_{ij} = \frac{1}{N}\sum_{\vec{R},\vec{R}'} e^{i\vec{k}\cdot(\vec{R}-\vec{R}')} \left\langle \varphi_i(\vec{r}-\vec{R}')\middle|\varphi_j(\vec{r}-\vec{R})\right\rangle.$$

Hence, due to the periodicity, one of the summations simply cancels the factor N and we can write

$$H_{ij} = \sum_{\vec{R}} e^{i\vec{k}\cdot\vec{R}} \left\langle \varphi_i(\vec{r})\middle|H(\vec{r})\middle|\varphi_j(\vec{r}-\vec{R})\right\rangle$$
$$S_{ij} = \sum_{\vec{R}} e^{i\vec{k}\cdot\vec{R}} \left\langle \varphi_i(\vec{r})\middle|\varphi_j(\vec{r}-\vec{R})\right\rangle. \tag{A2.8}$$

A2.2 One-Dimensional Periodic Structures

As an example of a 1D structure, we consider polyacetylene:

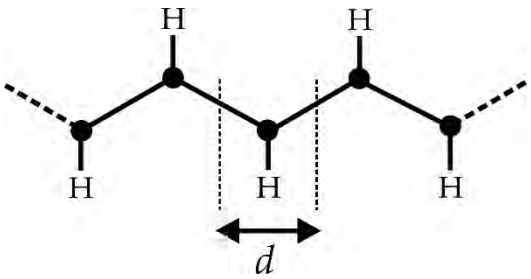


Figure A2.2. Atomic structure of the polyacetylene zigzag chain composed of carbon atoms (dots) and hydrogen.

The rules in Eq.(A2.5) show that a p orbital that is perpendicular to all lattice vectors in the structure will be completely decoupled from the other p orbitals as well as s orbitals. Thus, for planar structures such as this one, π and σ orbitals decouple and we will consider only π - states. As far as the p_z orbitals are concerned, polyacetylene is just a very long chain of atomic sites and neither the zig-zag geometry nor the H atoms play any role. With a single p_z orbital on each carbon atom and only nearest-neighbor interactions we then find the energy matrix element

$$H_{zz} = \alpha + \beta\left(e^{ikd} + e^{-ikd}\right) = \alpha + 2\beta\cos(kd),$$

where the Hückel notation $\alpha = E_p$ and $\beta = H_{pp\pi}$ has been used. Here, k is the wavenumber and d is the length of the unit cell (effective unit cell for p_z orbitals). If nearest-neighbor overlap is neglected, we then immediately find $E(k) = \alpha + 2\beta\cos(kd)$ as shown below for $\alpha = 0$ and $\beta = -3.033\text{eV}$ within the Brillouin zone $k \in [-\pi/d, \pi/d]$.

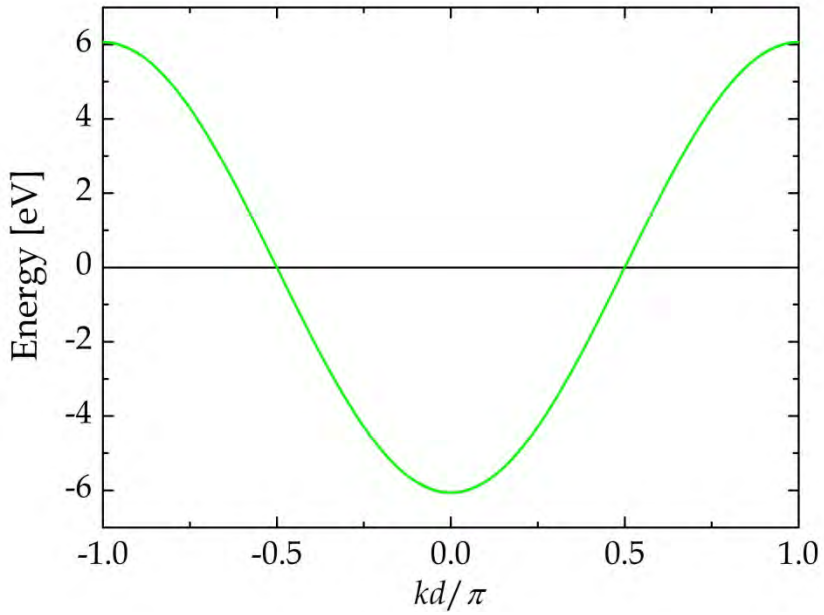


Figure A2.3. Band structure of π – states in polyacetylene.

Now, suppose we had used a unit cell twice as large. We would then have a basis containing Bloch sums for the two p_z orbitals belonging to the two atoms. If we call them $|1\rangle$ and $|2\rangle$ we would get the matrix elements

$$H_{11} = H_{22} = \alpha, \quad H_{12} = H_{21}^{*} = \beta(1 + e^{ik2d})$$

Hence, the energy eigenvalues are

$$E_{\pm}(k) = \alpha \pm |\beta(1 + e^{ik2d})| = \alpha \pm 2\beta\cos(kd). \qquad \text{(A2.9)}$$

But now we should remember that the Brillouin zone is twice as small as before: $k \in [-\pi/2d, \pi/2d]$. Thus, the band structure is as shown by the dashed lines in Fig. A2.4 below. We see that the bands are "folded" into the new Brillouin zone. In actual polyacetylene, the structure is "dimerized" so that the bond alternates between short and long bonds. If the two energy integrals are denoted β_1 and β_2 for the long and short bond, respectively, we find the matrix elements

$$H_{11} = H_{22} = \alpha, \quad H_{12} = H_{21}^* = \beta_1 + \beta_2 e^{ik2d}$$

and the eigenvalues

$$E_{\pm}(k) = \alpha \pm |\beta_1 + \beta_2 e^{ik2d}| = \alpha \pm \sqrt{\beta_1^2 + \beta_2^2 + 2\beta_1\beta_2 \cos(2kd)}. \qquad \text{(A2.10)}$$

For $\beta_1 = -2.5\text{eV}$ and $\beta_2 = -3.5\text{eV}$ the result is shown as the solid lines in Fig. A2.4. Notice the appearance of a band gap in the dimerized structure.

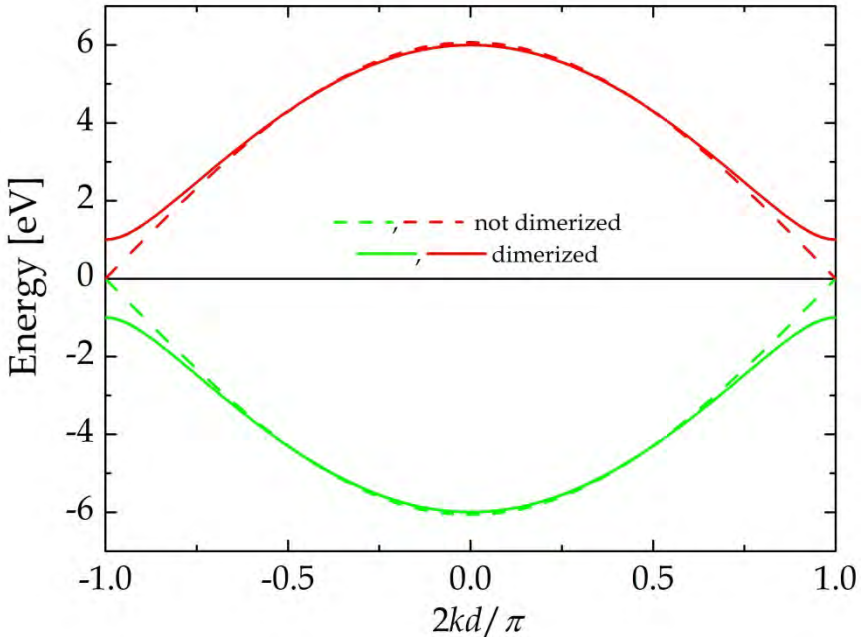


Figure A2.4. Band structure of π – states in polyacetylene folded once by doubling the unit cell (dashed) and when dimerized (solid).

A2.3 Two-Dimensional Periodic Structures

As our 2D example, we consider graphene consisting of a honeycomb lattice of carbon atoms. The lattice and Brillouin zone are illustrated in Fig. A2.5.

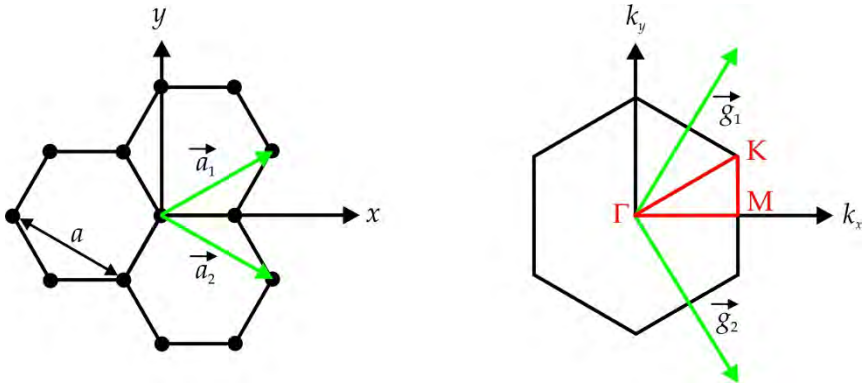


Figure A2.5. Graphene lattice incl. lattice vectors (left) and Brillouin zone incl. reciprocal lattice vectors (right).

The lattice constant of graphene is $a = 2.46\,\text{Å}$ and the elementary lattice vectors are

$$\vec{a}_1 = \frac{a}{2}\begin{pmatrix}\sqrt{3}\\ 1\end{pmatrix}, \quad \vec{a}_2 = \frac{a}{2}\begin{pmatrix}\sqrt{3}\\ -1\end{pmatrix}.$$

The two reciprocal lattice vectors are then determined via $\vec{g}_i \cdot \vec{a}_j = 2\pi\delta_{ij}$ to be

$$\vec{g}_1 = \frac{2\pi}{\sqrt{3}a}\begin{pmatrix}1\\ \sqrt{3}\end{pmatrix}, \quad \vec{g}_2 = \frac{2\pi}{\sqrt{3}a}\begin{pmatrix}1\\ -\sqrt{3}\end{pmatrix}.$$

As in the case of polyacetylene, graphene is planar and so p_z- states completely decouple from the rest. Again, we assume that each p_z- orbital is coupled to its nearest neighbors only with energy integral $H_{pp\pi} \equiv -\gamma$. Also, we form a basis from the two Bloch sums given by summing p_z- orbitals belonging to the two atoms A (leftmost) and B (rightmost) in the unit cell (so-called sublattices), separately. Focusing on an A atom, the vectors reaching to the three nearest B atoms are

$$\vec{R}_1 = a\begin{pmatrix}1/\sqrt{3}\\ 0\end{pmatrix}, \quad \vec{R}_2 = \frac{a}{2}\begin{pmatrix}-1/\sqrt{3}\\ 1\end{pmatrix}, \quad \vec{R}_3 = \frac{a}{2}\begin{pmatrix}-1/\sqrt{3}\\ -1\end{pmatrix}.$$

The matrix element is then

$$H_{12} = -\gamma h(\vec{k}), \quad h(\vec{k}) = \sum_{i=1}^{3} e^{i\vec{k}\cdot\vec{R}_i} = e^{ik_x a/\sqrt{3}} + 2e^{-ik_x a/2\sqrt{3}}\cos\left(k_y a/2\right),$$

and $H_{21} = H_{12}^*$. Taking the on-site p_z energy as the zero-point again, $H_{11} = H_{22} = 0$. It follows the eigenvalues are $E_\pm = \pm\gamma\,|h(\vec{k})|$. The full band structure is shown in Fig. A2.6.

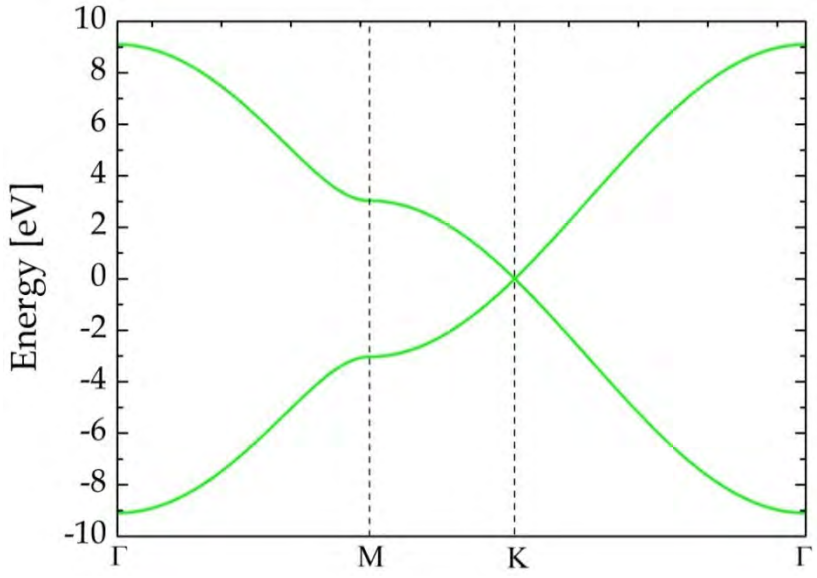


Figure A2.6. Band structure of π – states in graphene.

A2.4 Three-Dimensional Periodic Structures

Finally, as a 3D example, we will look at the simple cubic crystal:

Figure A2.7. Simple cubic lattice.

For an atom in (0,0,0) the six nearest neighbors are (1,0,0), (-1,0,0), (0,1,0), (0,-1,0), (0,0,1) and (0,0,-1). We will only include nearest-neighbor interactions. Hence, if we call the nearest neighbor distance d, it is clear that

$$\sum_{\vec{R}\in\{n.n.\}} e^{i\vec{k}\cdot\vec{R}} = 2\cos(k_x d) + 2\cos(k_y d) + 2\cos(k_z d).$$

Now, if we take again s, p_x, p_y and p_z as our basis set orbitals, we immediately find

$$H_{ss} = E_s + 2H_{ss\sigma}\left\{\cos(k_x d) + \cos(k_y d) + \cos(k_z d)\right\}.$$

For s and p_x we find

$$H_{sx} = H_{sp\sigma}\left\{\exp(ik_x d) - \exp(-ik_x d)\right\} = 2iH_{sp\sigma}\sin(k_x d),$$

and similarly for H_{sy} and H_{sz}. For p_x with itself we have

$$H_{xx} = E_p + 2H_{pp\sigma}\cos(k_x d) + 2H_{pp\pi}\left\{\cos(k_y d) + \cos(k_z d)\right\}$$

and similarly for H_{yy} and H_{zz}. Finally, we find that H_{xy} and similar terms vanish. The total matrix then reads as

$$\overleftrightarrow{H} = \begin{pmatrix} H_{ss} & H_{sx} & H_{sy} & H_{sz} \\ H_{sx}^* & H_{xx} & 0 & 0 \\ H_{sy}^* & 0 & H_{yy} & 0 \\ H_{sz}^* & 0 & 0 & H_{zz} \end{pmatrix}.$$

If we use the carbon parameters given above we find the band structure shown below. Here, the eigenvalues are plotted for the $\Gamma \to M$ and $\Gamma \to X$ directions with $M:(\pi/d, \pi/d, 0)$ and $X:(\pi/d, 0, 0)$.

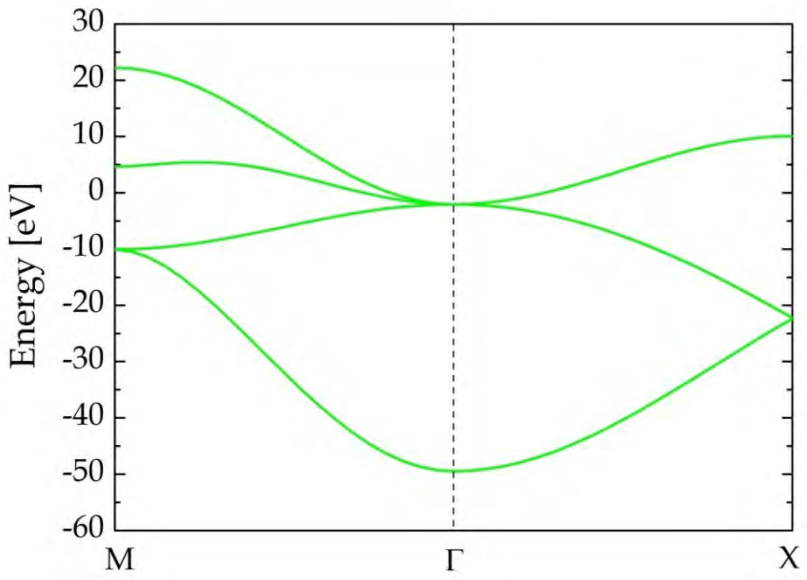


Figure A2.8. Band structure of the simple cubic lattice.

References

[1] J.C. Slater and C.F. Koster, Phys. Rev. 94, 1498 (1954).
[2] R. Saito, G. Dresselhaus, and M.S. Dresselhaus *Physical Properties of Carbon Nanotubes* (Imperial College, London, 1998).
[3] J.P. Dahl *Introduction to the Quantum World of Atoms and Molecules* (World Scientific, Singapore, 2001).

Appendix 3. Pseudopotential Method

In this appendix, we present the basic ideas of the pseudopotential method, sometimes also known as the "empirical pseudopotential" (EMP) method. One advantage of the EMP method is that the actual wave function is available, while in tight-binding everything is expressed in an atomic basis, which is not explicitly known. The starting point is an expansion of the potential as a Fourier series $V(\vec{r}) = \sum_{\vec{G}} V_{\vec{G}} e^{i\vec{G}\cdot\vec{r}}$. This is, of course, always possible. The name "pseudopotential" refers to the fact that this potential is typically not the actual potential but, rather, a smoothed version with the rapid variations near the atomic cores omitted. In Fourier space, this amounts to removing Fourier components $V_{\vec{G}}$ with large *G*. As a consequence, the core states formed by the tightly bound core atomic orbitals are absent in the band structure, which only contains valence electron bands.

The wave function is expanded similarly to the potential and therefore reads $\varphi_{n,\vec{k}}(\vec{r}) = \Omega^{-1/2} \sum_{\vec{G}} C_{n,\vec{k}}(\vec{G}) e^{i(\vec{G}+\vec{k})\cdot\vec{r}}$, where Ω is the crystal volume. Normalization is such that $\sum_{\vec{G}} |C_{n,\vec{k}}(\vec{G})|^2 = 1$. In this basis, the matrix elements of the Hamiltonian are

$$H_{\vec{G},\vec{G}'}(\vec{k}) = \frac{\hbar^2}{2m}(\vec{k}+\vec{G})^2 \delta_{\vec{G},\vec{G}'} + V_{\vec{G}-\vec{G}'}.$$

If spin-orbit interation is included, one should add matrix elements

$$\left\langle \sigma \left| H^{SO}_{\vec{G},\vec{G}'} \right| \sigma' \right\rangle = \frac{i\hbar^2}{4m^2c^2} V_{\vec{G}-\vec{G}'} [(\vec{k}+\vec{G}) \times (\vec{k}+\vec{G}')] \cdot \left\langle \sigma \left| \hat{\vec{\sigma}} \right| \sigma' \right\rangle.$$

Computing inter-band momentum matrix elements is easy in the Fourier basis and they are simply given by $\vec{p}_{cv\vec{k}} = \hbar \sum_{\vec{G}} C^*_{c,\vec{k}}(\vec{G}) \vec{G} C_{v,\vec{k}}(\vec{G})$.

We will now provide some useful details for specific lattices. The idea is the following: An atom *i* is characterized by a certain potential $v^{(i)}(r)$, which can be assumed spherically symmetric. Importantly, its Fourier components $v^{(i)}_G$ are then functions of *G*, the magnitude of $\vec{G}$, only. Summing over all atoms in the unit cell, the total potential is $V(\vec{r}) = \sum_i v^{(i)}(|\vec{r} - \vec{\tau}_i|)$, where $\vec{\tau}_i$ is the position of the *i*'th atom. In turn, this means $V_{\vec{G}} = \sum_i v^{(i)}_G e^{i\vec{G}\cdot\vec{\tau}_i}$, as can easily be shown. The lattices considered are zinc-blende and wurtzite with Brillouin zones as shown in Fig. A3.1

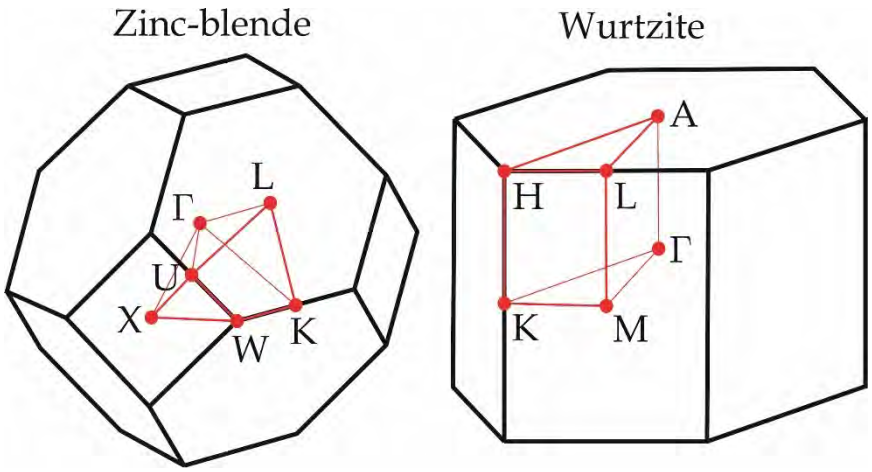


Figure A3.1. Brillouin zones of zinc-blende and wurzite lattices including high symmetry points and irreducible zones enclosed by red lines.

A3.1 Zinc-Blende Lattice

In the zinc-blende (ZB) case with lattice constant a, there are two atoms per unit cell:

$$\text{One cation } C \text{ at } \vec{\tau}_1 = a(1,1,1)/8 \equiv \vec{\tau} \text{ and one anion } A \text{ at } \vec{\tau}_2 = -\vec{\tau}\,.$$

This case also includes the diamond lattice, in which A and C atoms are identical. We also define symmetric and antisymmetric Fourier coefficients $V_G^{S,A} = v_G^{(C)} \pm v_G^{(A)}$. This leads to a total of

$$V_{\vec{G}}^{ZB} = V_G^S F_S(\vec{G}) + V_G^A F_A(\vec{G}), \quad F_S(\vec{G}) = \cos(\vec{G}\cdot\vec{\tau}), \quad F_A(\vec{G}) = i\sin(\vec{G}\cdot\vec{\tau})\,.$$

The great advantage here is that the angular dependence is only in the structure factors $F_{S,A}(\vec{G})$. The ZB lattice consists of two inter-penetrating FCC lattices. Hence, the reciprocal lattice is a BCC lattice with reciprocal lattice vectors $\vec{G} = \frac{2\pi}{a}(k,l,m)$ with k,l,m either all even or all odd. We can then tabulate the properties of the smallest $\vec{G}$-vectors:

$\vec{G}(\frac{a}{2\pi})$	$G^2(\frac{a}{2\pi})^2$	$\vec{G}\cdot\vec{\tau}$
(0,0,0)	0	0
(1,1,1) class	3	$\pm\frac{\pi}{4}, \pm\frac{3\pi}{4}$
(2,0,0) class	4	$\pm\frac{\pi}{2}$
(2,2,0) class	8	$0, \pm\pi$
(1,1,3) class	11	$\pm\frac{\pi}{4}, \pm\frac{3\pi}{4}, \pm\frac{5\pi}{4}$

An important consequence is that one may ignore $V^A_{(0,0,0)}$, $V^S_{(2,0,0)}$ and $V^A_{(2,2,0)}$ because the associated structure factors vanish. Also, we may take $V^S_{(0,0,0)} = 0$ as a choice of zero point. Hence, truncating our potential at the $\vec{G}$-vector set in the table, we are left with

$$V^S_{(1,1,1)}, V^S_{(2,2,0)}, V^S_{(1,1,3)}, V^A_{(1,1,1)}, V^A_{(2,0,0)}, V^A_{(1,1,3)}.$$

Moreover, for the diamond lattice, all antisymmetric coefficients vanish. As two classic examples, one has (in Rydberg energy units) [1]

$$\text{Si: } V^S_{(1,1,1)} = -0.21, V^S_{(2,2,0)} = 0.04, V^S_{(1,1,3)} = 0.08, V^A_{(1,1,1)} = V^A_{(2,0,0)} = V^A_{(1,1,3)} = 0,$$
$$\text{GaAs: } V^S_{(1,1,1)} = -0.23, V^S_{(2,2,0)} = 0.01, V^S_{(1,1,3)} = 0.06, V^A_{(1,1,1)} = 0.07, V^A_{(2,0,0)} = 0.05, V^A_{(1,1,3)} = 0.01.$$

The respective lattice constants a are 5.43 Å and 5.64 Å. In the examples below, we limit the $\vec{G}$-vectors in the basis set to those, for which $G(\frac{a}{2\pi}) \leq 5$, leading to a total of 137 basis functions. The computed band structures of Si and GaAs (without spin-orbit) are illustrated below, where bands are shifted such that the highest valence band state is at zero.

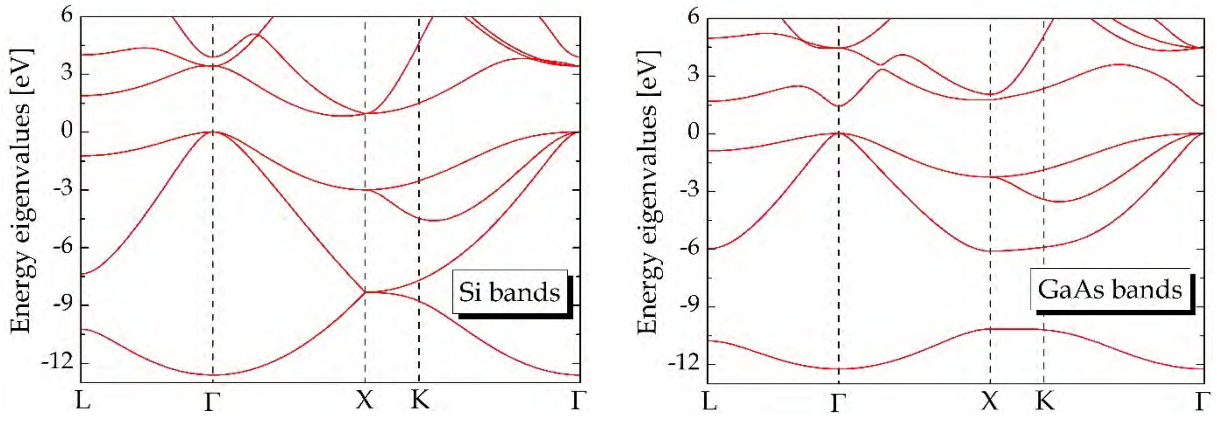


Figure A3.2. Pseudopotential band structures of Si and zinc-blende GaAs.

A3.2 Wurtzite Lattice

The wurtzite (W) system has four atoms per unit cell and lattice constants a and c:

$$\text{Cations } C \text{ at } \vec{\tau}_1 = (0,0,0) \text{ and } \vec{\tau}_2 = (\tfrac{1}{\sqrt{3}}a, 0, \tfrac{1}{2}c) \text{ and}$$
$$\text{anions } A \text{ at } \vec{\tau}_3 = (\tfrac{1}{\sqrt{3}}a, 0, (\tfrac{1}{2}-u)c) \text{ and } \vec{\tau}_4 = (0,0,(1-u)c).$$

This means that the Fourier potential is

$$V^W_{\vec{G}} = V^S_G F_S(\vec{G}) + V^A_G F_A(\vec{G}), \quad F_{S,A}(\vec{G}) = \tfrac{1}{2}\left\{1 + e^{i\vec{G}\cdot\vec{\tau}_2} \pm (e^{i\vec{G}\cdot\vec{\tau}_3} + e^{i\vec{G}\cdot\vec{\tau}_4})\right\}.$$

Note that many references define the wurtzite structure factors with a factor of ¼ instead of ½. In the "ideal" case, all nearest-neighbor distances are identical and we have $c/a=\sqrt{8/3}$ as well as $u=3/8$. The wurtzite lattice is formed as two interpenetrating HCP lattices and the reciprocal lattice is spanned by the vectors

$$\vec{b}_1=\frac{2\pi}{a}(\tfrac{1}{\sqrt{3}},-1,0),\quad \vec{b}_2=\frac{2\pi}{a}(\tfrac{1}{\sqrt{3}},1,0),\quad \vec{b}_3=\frac{2\pi}{c}(0,0,1).$$

Reciprocal lattice vectors $\vec{G}=k\vec{b}_1+l\vec{b}_2+m\vec{b}_3$ have a magnitude

$$G^2=\frac{4}{3}(k^2+l^2-kl)\left(\frac{2\pi}{a}\right)^2+m^2\left(\frac{2\pi}{c}\right)^2.$$

When defining the Fourier components, it is advantageous to refer to the ideal lattice, in which $g^2\equiv 24G^2(\tfrac{a}{2\pi})^2=32(k^2+l^2-kl)+9m^2$. Listed in increasing order, the smallest vectors are

$$g^2\in\{0,9,32,36,41,68,81,96,105,113,128,132,137,144,164,176,...\}.$$

Among these, both structure factors vanish for $g^2\in\{9,81,105\}$, while $F_S(\vec{G})=0$ for $g^2\in\{144,176\}$ and $F_A(\vec{G})=0$ for $g^2\in\{0,32,96,128\}$. The remaining Fourier components for ZnO and GaN can be found in Refs. [2,3] and the lattice constants a are 3.25 Å and 3.189 Å, respectively. Limiting the $\vec{G}$-vectors to $G(\tfrac{a}{2\pi})\le 3$ leads to a total of 147 basis functions.

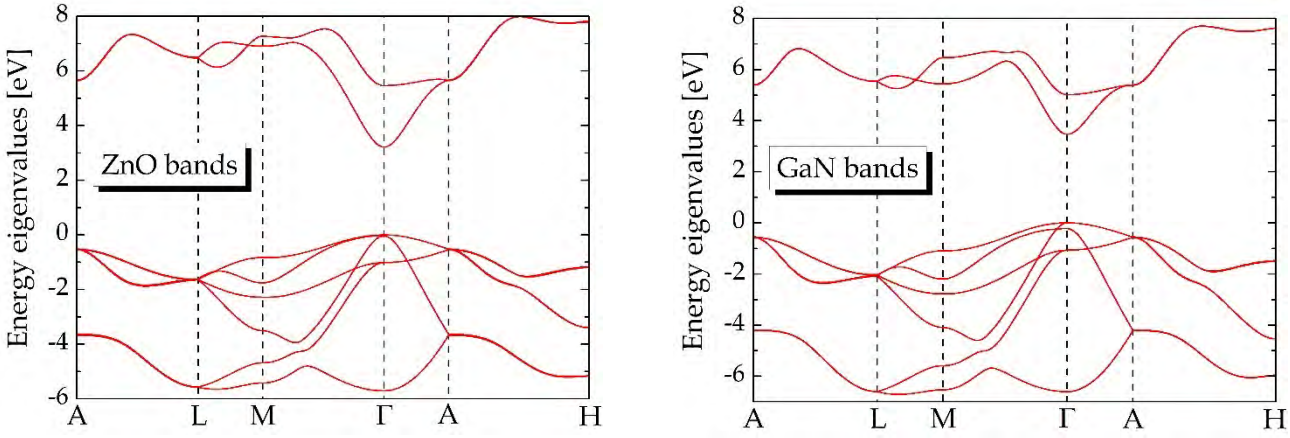


Figure A3.3. Pseudopotential band structures of wurtzite ZnO and GaN.

References

[1] M.L. Cohen and T.K. Bergstresser, Phys. Rev. 141, 789 (1966).
[2] S. Bloom and I. Ortenburger, Phys. Stat. Sol. (b) 58, 561 (1973).
[3] Y.C. Yeo, T.C. Chong, and M.F. Li, J. Appl. Phys. 83, 1429 (1998).

Appendix 4. Density Functional Theory

Density functional theory (DFT) is arguably the most important method in solid state electronic structure calculations. It is also rapidly becoming a standard tool of chemistry and as methods and computers evolve, DFT is able to tackle increasingly large structures. Thus, DFT is now heavily used in nanostructure simulations as well. In this appendix, we will briefly explain the basics of DFT. As a concrete implementation, the Perdew-Zunger parameterization of the exchange-correlation energy functional will be discussed along with the concepts of self-consistency and density mixing. We will then turn to a practical application: the DFT simulation of atomic states using a Gaussian basis.

The formal foundation of DFT is the Hohenberg-Kohn theorem, which states that there exists a one to one correspondence between a system (specified by the arrangement of atomic nuclei) and its spatially varying electron density [1]. The potential produced by the nuclei is given by

$$V_N(\vec{r}) = -\sum_I \frac{Z_I}{|\vec{r} - \vec{R}_I|}, \tag{A4.1}$$

where atomic units (distances and energies measured in Bohrs and Hartrees, respectively) are applied and the sum is over the nuclei having charges Z_I and positions $\vec{R}_I$. In standard quantum theory, the electrons would be described by a many-electron wave function $\psi(\vec{r}_1,....,\vec{r}_N)$ assuming a total of N electrons and suppressing spin. This wave function allows for the calculation of all relevant quantities including the electron density

$$n(\vec{r}) = \int |\psi(\vec{r},\vec{r}_2,....,\vec{r}_N)|^2 \, d^3r_2 .... d^3r_N .$$

The key insight of Kohn and Sham is that a one can introduce a fictitious system of *independent* electrons having the *same density* as the true one [2]. For independent electrons subject to the Pauli principle only, the exact wave function is a Slater determinant

$$\psi(\vec{r}_1,....,\vec{r}_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \varphi_1(\vec{r}_1) & \varphi_1(\vec{r}_2) & ... & \varphi_1(\vec{r}_N) \\ \varphi_2(\vec{r}_1) & \varphi_2(\vec{r}_2) & ... & \varphi_2(\vec{r}_N) \\ \vdots & \vdots & & \vdots \\ \varphi_N(\vec{r}_1) & \varphi_N(\vec{r}_2) & ... & \varphi_N(\vec{r}_N) \end{vmatrix},$$

for which the electron density is simply

$$n(\vec{r}) = \sum_i |\varphi_i(\vec{r})|^2 .$$

By means of the variational principle, the single-electron states $\varphi_i(\vec{r})$ for the ground state are found to be eigenstates of the Kohn-Sham Hamiltonian

$$\hat{H} = -\frac{1}{2}\nabla^2 + V_N(\vec{r}) + V_H(\vec{r}) + V_{xc}(\vec{r}), \tag{A4.2}$$

where V_N is the nuclear potential given by Eq.(A4.1), V_H is the Hartree potential

$$V_H(\vec{r}) = \int \frac{n(\vec{r}')}{|\vec{r} - \vec{r}'|} d^3r', \tag{A4.3}$$

and V_{xc} is the so-called exchange-correlation potential. This potential is, in fact, unknown except for the fact that it must be possible to express it in terms of the density $n(\vec{r})$ and nothing else. Excellent guesses exist, however. In particular, very accurate numerical results based on quantum Monte Carlo techniques have been used to parameterize V_{xc}. In the *local density approximation* (*LDA*), V_{xc} is assumed to depend on $n(\vec{r})$ but not on its gradient or any higher derivatives. This means that Monte Carlo data for the homogeneous electron gas are enough to parameterize V_{xc}. This can then be done for various degrees of spin polarization, in particular, unpolarized and fully polarized cases. Perdew and Zunger [3] adopted the following parameterization for the exchange-correlation energy E_{xc} and potential V_{xc}

$$E_{xc}[n(\vec{r})] = \int n(\vec{r})\varepsilon_{xc} d^3r, \quad V_{xc}[n(\vec{r})] = \frac{\delta E_{xc}}{\delta n},$$

$$\varepsilon_{xc} = -\frac{3}{4}\left(\frac{9}{4\pi^2}\right)^{1/3}\frac{1}{r_s} + \begin{cases} \dfrac{\gamma}{1+\beta_1\sqrt{r_s}+\beta_2 r_s} & r_s > 1 \\ A\ln r_s + B + Cr_s \ln r_s + Dr_s & r_s < 1 \end{cases} \tag{A4.4}$$

$$V_{xc} = -\left(\frac{9}{4\pi^2}\right)^{1/3}\frac{1}{r_s} + \begin{cases} \gamma\dfrac{1+\frac{7}{6}\beta_1\sqrt{r_s}+\frac{4}{3}\beta_2 r_s}{(1+\beta_1\sqrt{r_s}+\beta_2 r_s)^2} & r_s > 1 \\ A\ln r_s + B - \frac{1}{3}A + \frac{2}{3}Cr_s\ln r_s + \left(\frac{2}{3}D - \frac{1}{3}C\right)r_s & r_s < 1, \end{cases}$$

where the density dependence is via the Seitz radius $r_s = (3/4\pi n(\vec{r}))^{1/3}$. For the unpolarized and fully polarized cases, the parameters are

	γ	β_1	β_2	A	B	C	D
Unpol.	-0.1423	1.0529	0.3334	0.0311	-0.048	0.002	-0.0116
Pol.	-0.0843	1.3981	0.2611	0.01555	-0.0269	0.0007	-0.0048

In the situation, where the spin polarization $\zeta = (n_\uparrow - n_\downarrow)/n$ is intermediate, interpolation between the two limits is done via the expression

$$f(\zeta) = \frac{(1+\zeta)^{4/3} + (1-\zeta)^{4/3} - 2}{2^{4/3} - 2} \tag{A4.5}$$

with the limits $f(0) = 0, f(1) = 1$. The eigenvalues of the Kohn-Sham Hamiltonian are denoted E_i. In terms of these, we find a total energy of the entire electronic system of

$$E_{tot} = \sum_{i=1}^{N} E_i - \frac{1}{2}\int V_H(\vec{r})n(\vec{r})d^3r + E_{xc} - \int V_{xc}(\vec{r})n(\vec{r})d^3r\,. \tag{A4.6}$$

Because the potential depends on the density that, in turn, is computed from the wave functions it is necessary to solve using iterations. Thus, an initial guess at the density can be used and subsequently updated until convergence.

A4.1 Atomic States

In a spherically symmetric problem, we base the analysis on spherical coordinates r, θ, ϕ and the eigenstates are of the form $\varphi_i(\vec{r}) = Y_{lm}(\theta,\phi)R_n(r)$, where Y_{lm} is a spherical harmonic and R contains the radial dependence. However, the electron density found by summing squares of these functions is typically not spherically symmetric. The reason is that the spherical harmonics lead to a net angular dependence of the density unless the atom has a closed valence shell. Consider, as an example, the C atom with two electrons in 2*p* orbitals. If these orbitals are taken to be, for instance, Y_{10} and Y_{11} the density contributions from these terms $|Y_{10}(\theta,\phi)|^2 + |Y_{11}(\theta,\phi)|^2$ will not be symmetric, i.e. independent of angles. On the other hand, for a closed shell atom such as Ne, the extra four electrons produce a total density without angular dependence because, quite generally,

$$\sum_{m=-l}^{l} |Y_{lm}(\theta,\phi)|^2 = \frac{2l+1}{4\pi}.$$

Thus, we can enforce spherical symmetry by making a "restricted" calculation in the sense that we force the occupation of such valence states to be equal. For instance, in the case of C, the occupancy of each of the 2*p* orbitals will be taken to be 2/3 so that the total density including *s*-orbitals becomes

$$
\begin{aligned}
n(\vec{r}) &= 2\,|\,Y_{00}(\theta,\phi)\,|^2\;R_{1s}^2(r) + 2\,|\,Y_{00}(\theta,\phi)\,|^2\;R_{2s}^2(r) \\
&\quad + \frac{2}{3}\left\{|\,Y_{1-1}(\theta,\phi)\,|^2 + |\,Y_{10}(\theta,\phi)\,|^2 + |\,Y_{11}(\theta,\phi)\,|^2\right\} R_{2p}^2(r) \\
&= \frac{1}{2\pi}R_{1s}^2(r) + \frac{1}{2\pi}R_{2s}^2(r) + \frac{1}{2\pi}R_{2p}^2(r).
\end{aligned}
$$

For the N atom, the weight of the 2*p* term would be $3/4\pi$ and so on. With such a restricted calculation, the potential $V_N(\vec{r}) + V_H(\vec{r}) + V_{xc}(\vec{r})$ and thereby the entire problem has full spherical symmetry. In fact, it is readily demonstrated that

$$
V_H(r) = 4\pi\int_0^\infty \frac{n(r')}{\max(r,r')}r'^2 dr'\,. \tag{A4.7}
$$

Accordingly, the radial eigenfunctions are solutions to

$$
\left\{-\frac{1}{2r}\frac{d^2}{dr^2}r + \frac{l(l+1)}{2r^2} + V_N(r) + V_H(r) + V_{xc}(r)\right\} R(r) = ER(r)\,. \tag{A4.8}
$$

In order to find the eigenstate, we expand in a basis so that $\left|R(r)\right\rangle = \sum_i c_i \left|l,i\right\rangle$. A Gaussian basis function for this problem is of the form $\left|l,i\right\rangle = r^l \exp\{-b_i r^2\}$. We will restrict the discussion to *s*, *p*, and *d* states, i.e. to *l* = 0, 1, and 2. To simplify notation, we introduce $b_{ij} = b_i + b_j$. Primarily, we need the overlap integrals:

$$
\left\langle 0,i\middle|0,j\right\rangle = \frac{\pi^{1/2}}{4b_{ij}^{3/2}},\quad \left\langle 1,i\middle|1,j\right\rangle = \frac{3\pi^{1/2}}{8b_{ij}^{5/2}},\quad \left\langle 2,i\middle|2,j\right\rangle = \frac{15\pi^{1/2}}{16b_{ij}^{7/2}}\,.
$$

Also, we need matrix elements of the kinetic energy $\hat{T} = -\frac{1}{2r}\frac{d^2}{dr^2}r + \frac{l(l+1)}{2r^2}$

$$
\left\langle 0,i\middle|\hat{T}\middle|0,j\right\rangle = \frac{3\pi^{1/2}b_i b_j}{4b_{ij}^{5/2}},\quad \left\langle 1,i\middle|\hat{T}\middle|1,j\right\rangle = \frac{15\pi^{1/2}b_i b_j}{8b_{ij}^{7/2}},\quad \left\langle 2,i\middle|\hat{T}\middle|2,j\right\rangle = \frac{105\pi^{1/2}b_i b_j}{16b_{ij}^{9/2}}
$$

and Coulomb operator

$$
\left\langle 0,i\middle|\frac{1}{r}\middle|0,j\right\rangle = \frac{1}{2b_{ij}},\quad \left\langle 1,i\middle|\frac{1}{r}\middle|1,j\right\rangle = \frac{1}{2b_{ij}^2},\quad \left\langle 2,i\middle|\frac{1}{r}\middle|2,j\right\rangle = \frac{1}{b_{ij}^3}\,.
$$

Optimized values of the exponents b_i have been tabulated by Koga and coworkers [4].

A4.2 Example: Carbon Atom

As an important and illustrative example, we will study the carbon atom in detail. The electronic configuration is $1s^2 2s^2 2p^2$ in the neutral atom and $2p$ electrons are removed and added in the positive and negative ions, respectively. We use the Koga Gauss basis [4] but in order to describe delocalized $2p$ electrons more accurately, we add two additional p Gaussians with coefficients $b = 38 \cdot 10^{-5}$ and $b = 19 \cdot 10^{-5}$. Also, we use an equidistant radial grid with 2000 points between 0 and 15. As a starting point, we may totally ignore electron-electron interaction terms. In this case, the eigenstates are easily obtained from the generalized eigenvalue problem described above. Using the Gauss basis, we find eigenvalues of $E_{1s} = -17.9999, E_{2s} = -4.4979, E_{2p} = -4.4999$ Hartrees that should be compared to the exact values of $E_{1s} = -18, E_{2s} = -9/2, E_{2p} = -9/2$. The associated electron density is shown as the black curve in Fig. A4.1.

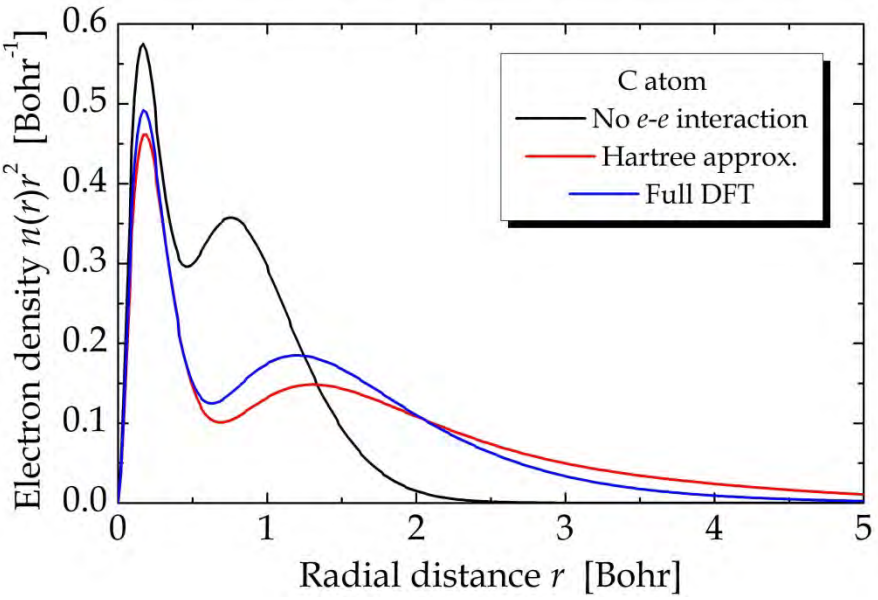


Figure A4.1. Radial densities for the C atom in various approximations.

If we next turn on the Hartree term, we need to iterate until convergence. This procedure leads to the density plotted in red in Fig. A4.1 and eigenvalues $E_{1s} = -8.5520, E_{2s} = -0.2634, E_{2p} = -0.0106$. If we finally include the full xc potential, we find $E_{1s} = -9.9469, E_{2s} = -0.5008, E_{2p} = -0.1992$ and the density shown in blue. The total energy calculated using Eq.(A4.5) is then $E_{tot} = -37.423$ Ha $= -1018.3$ eV. We notice several things. First, ignoring *e-e* interactions vastly overestimates the binding energies and localizes the electrons much too close to the nucleus. On the other hand, the Hartree approximation underestimates electron binding and leads to a $2p$ state that is hardly bound at all. Accordingly, the density is too delocalized. Finally, the full DFT increases binding and localization of the electrons somewhat compared to the Hartree approximation.

We can then go on to looking at positive and negative ions, C^+ and C^-, found by removing or adding one $2p$ electron, respectively. For the two ions we find $E_{tot}(\mathrm{C}^+) = -37.019\ \mathrm{Ha} = -1007.3\ \mathrm{eV}$ and $E_{tot}(\mathrm{C}^-) = -37.471\ \mathrm{Ha} = -1019.6\ \mathrm{eV}$. It follows that the ionization energy and electron affinity are 11.0 eV and 1.3 eV, respectively. Both of these values agree quite accurately with experiments. The electron densities of the ions are compared to the neutral case in Fig. A4.2. Note that the core region is virtually unaffected.

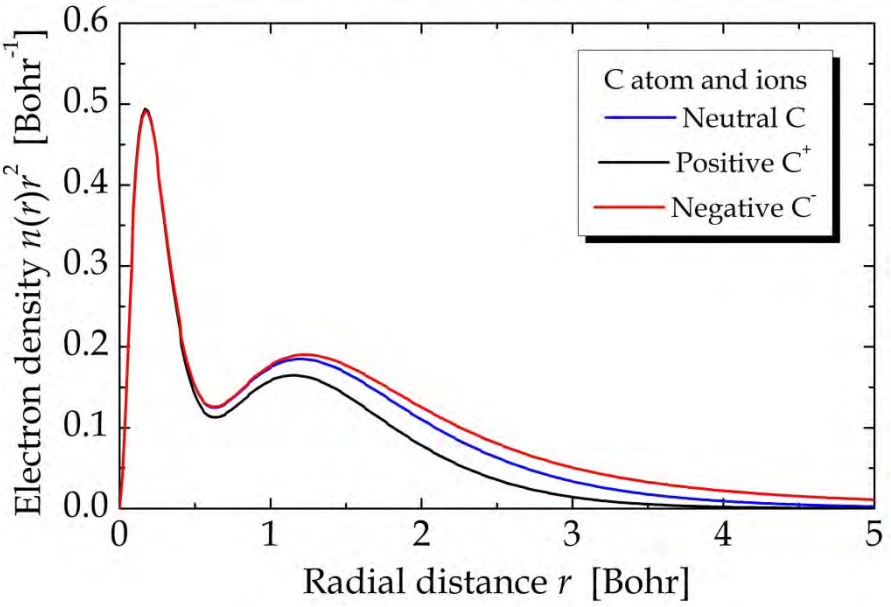


Figure A4.2. Radial densities for the neutral and charged C atoms.

A4.3 Density-Functional Based Tight-Binding

In order to explain the idea behind the DFTB approach [5], we now switch to a slightly more complicated example: A two-atomic molecule consisting of nuclei having charges Z_A and Z_B at positions $\vec{R}_A = 0$ and $\vec{R}_B = \vec{R}$, respectively. In this case, the Kohn-Sham Hamiltonian is

$$\hat{H} = -\frac{1}{2}\nabla^2 - \frac{Z_A}{r} - \frac{Z_B}{|\vec{r} - \vec{R}|} + V_H(\vec{r}) + V_{xc}(\vec{r}). \qquad \text{(A4.9)}$$

We would like to write this expression in terms of Hamiltonians for the individual atoms. However, this is clearly not rigorously possible. The physical reason is that electrons accumulate in the region between the nuclei because of binding. Hence, the Hartree and exchange-correlation terms are different from the simple sum of atomic terms. The Hartree potential for an isolated A atom will be denoted $V_H^A(\vec{r})$ and similarly for the xc term and for the B atom. Thus, *if* we ignored the binding effect, the two-atom Hamiltonian would be approximated by

$$\hat{H}_{\text{isolated atoms}} = \hat{H}_A + \hat{H}_B + \frac{1}{2}\nabla^2,$$
$$\hat{H}_A = -\frac{1}{2}\nabla^2 - \frac{Z_A}{r} + V_H^A(\vec{r}) + V_{xc}^A(\vec{r}), \quad \hat{H}_B = -\frac{1}{2}\nabla^2 - \frac{Z_B}{|\vec{r}-\vec{R}|} + V_H^B(\vec{r}-\vec{R}) + V_{xc}^B(\vec{r}-\vec{R}). \tag{A4.10}$$

The extra $\nabla^2/2$ term is to avoid double-counting the kinetic energy. Simply replacing $\hat{H}$ by $\hat{H}_{\text{isolated atoms}}$ would be much too inaccurate. However, rather than trying to improve $\hat{H}_{\text{isolated atoms}}$, a different approach is followed in DFTB. The effect of binding is built into the basis states instead. What we need is pseudo-atomic states that effectively shift electron density into the binding region between the nuclei. If the distance between the nuclei is about r_0^A, this is achieved by means of the pseudo-atomic Hamiltonian

$$\hat{H}_A^{ps} = -\frac{1}{2}\nabla^2 - \frac{Z_A}{r} + V_H^A(\vec{r}) + V_{xc}^A(\vec{r}) + \left(\frac{r}{r_0^A}\right)^2. \tag{A4.11}$$

Here, the compression term $(r/r_0^A)^2$ serves to move electrons into the bond region. The pseudo-atomic eigenstates are easily obtained from the matrix elements

$$\langle 0,i|\frac{r^2}{r_0^2}|0,j\rangle = \frac{3\pi^{1/2}}{8r_0^2 b_{ij}^{5/2}}, \quad \langle 1,i|\frac{r^2}{r_0^2}|1,j\rangle = \frac{15\pi^{1/2}}{16r_0^2 b_{ij}^{7/2}}, \quad \langle 2,i|\frac{r^2}{r_0^2}|2,j\rangle = \frac{105\pi^{1/2}}{32r_0^2 b_{ij}^{9/2}}.$$

We can now assume that we have computed pseudo-atomic eigenstates for *A* and *B* pseudo-atoms $\varphi_i^A(\vec{r})$ and $\varphi_i^B(\vec{r})$. The overlap matrix elements needed for the tight-binding parameterization of compounds containing *A* and *B* atoms are then

$$S_{i\alpha,j\beta}^{AB} = \left\langle \varphi_i^A(\vec{r}-\vec{R}_\alpha) \middle| \varphi_j^B(\vec{r}-\vec{R}_\beta) \right\rangle. \tag{A4.12}$$

As the effect of binding has been put into the eigenstates, we subsequently use these states along with the isolated-atoms Hamiltonian to compute the Hamilton matrix elements

$$H_{i\alpha,j\beta}^{AB} = \left\langle \varphi_i^A(\vec{r}-\vec{R}_\alpha) \middle| \hat{H}_{\text{isolated atoms}} \middle| \varphi_j^B(\vec{r}-\vec{R}_\beta) \right\rangle.$$

We see that if the pseudo-atomic eigenvalues are ε_i^A and ε_j^B, respectively, we have

$$H_{i\alpha,j\beta}^{AB} = \left(\varepsilon_i^A + \varepsilon_j^B\right) S_{i\alpha,j\beta}^{AB} - \left\langle \varphi_i^A(\vec{r}-\vec{R}_\alpha) \middle| -\frac{1}{2}\nabla^2 + \left(\frac{|\vec{r}-\vec{R}_\alpha|}{r_0^A}\right)^2 + \left(\frac{|\vec{r}-\vec{R}_\beta|}{r_0^B}\right)^2 \middle| \varphi_j^B(\vec{r}-\vec{R}_\beta) \right\rangle. \tag{A4.13}$$

For C and N atoms, the compression radii can be taken as 1.42Å [5] and 1.38Å [6], respectively. Thus, to compute the TB parameters, we need matrix elements between basis states located on *different* atoms. In general, these are quite complicated. As a relatively simple example, we consider the expression for *s*-states:

$$\left\langle 0,i,\vec{0}\middle|0,j,\vec{R}\right\rangle = \frac{\pi^{1/2}}{4b_{ij}^{3/2}}e^{-\beta_{ij}R^2}, \quad \left\langle 0,i,\vec{0}\middle|-\frac{\nabla^2}{2}\middle|0,j,\vec{R}\right\rangle = \frac{3\pi^{1/2}b_ib_j(1-\frac{2}{3}\beta_{ij}R^2)}{4b_{ij}^{5/2}}e^{-\beta_{ij}R^2},$$

$$\left\langle 0,i,\vec{0}\middle|r^2\middle|0,j,\vec{R}\right\rangle = \frac{3\pi^{1/2}(1+\frac{2b_j\beta_{ij}R^2}{3b_i})}{8b_{ij}^{5/2}}e^{-\beta_{ij}R^2}, \quad \left\langle 0,i,\vec{0}\middle|(\vec{r}-\vec{R})^2\middle|0,j,\vec{R}\right\rangle = \frac{3\pi^{1/2}(1+\frac{2b_i\beta_{ij}R^2}{3b_j})}{8b_{ij}^{5/2}}e^{-\beta_{ij}R^2}.$$

Here, $\beta_{ij} = b_ib_j/(b_i+b_j)$ and if two different atoms are considered b_i and b_j should be taken as coefficients of the *A* and *B* pseudo-atoms, respectively. In Fig. A4.3, we have taken the carbon orbitals as an example. The overlap and Hamilton matrix element for the 2*s* case are what is usually known as $S_{ss\sigma}$ and $H_{ss\sigma}$, see previous appendix.

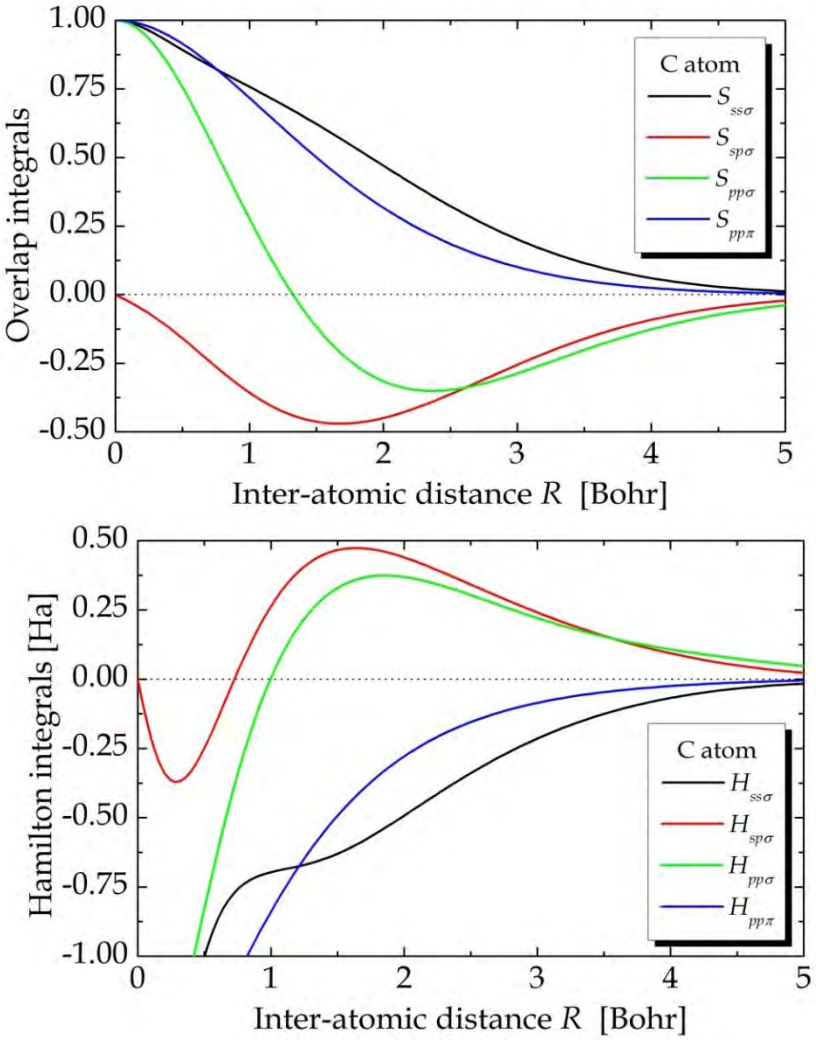


Figure A4.3. Overlap (top) and Hamilton (bottom) integrals for a pair of C atoms.

The final aspect of DFTB is the calculation of total energies. As in full DFT, the total energy is not just the sum of occupied energy eigenvalues, c.f. Eq. (A4.6). There are corrections for double counting etc. but also the Coulomb repulsion between nuclei. In DFTB, all these addition terms are collected in a repulsive potential $V_{rep}^{AB}(R)$ that depends on the inter-atomic distance R. Hence, the total energy is

$$E_{tot} = \sum_{i=1}^{N} E_i + \sum_{pairs} V_{rep}^{AB}(R) .$$

Two examples of total energies are shown in Fig. A4.4 along with repulsive potentials in the inset. Here, CC and NN dimers are considered and the energies are relative to those of the dissociated molecules. Hence, the curves actually display the binding energy.

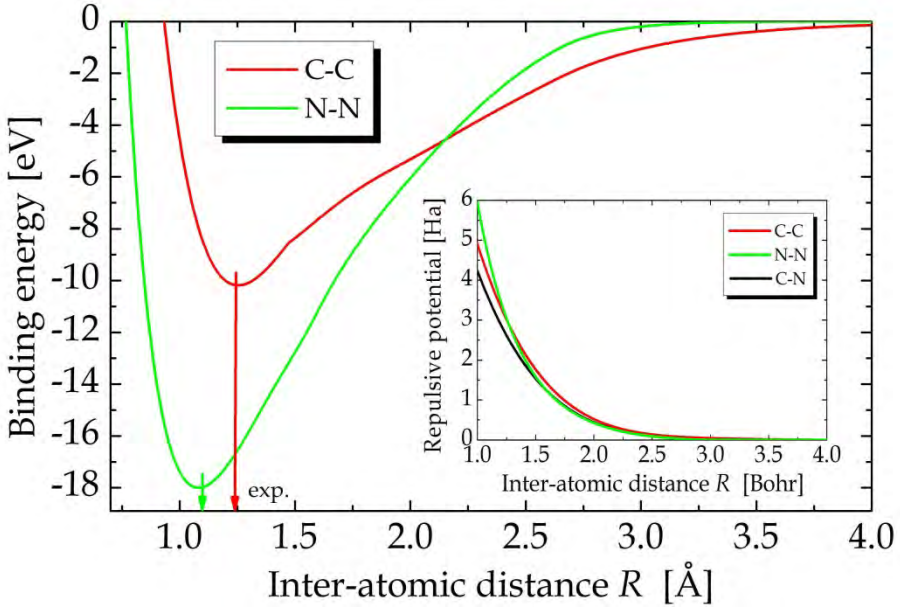


Figure A4.4. Binding energy of CC and NN dimers. The repulsive potentials are shown in the inset.

Note that the potential minima are close to the experimental equilibrium distances. Also, the binding energies are found to agree with experiments, in particular, if spin-polarization effects in the dissociated molecules are taken into account.

References

[1] P. Hohenberg and W. Kohn, Phys. Rev. 136, B864 (1964).
[2] W. Kohn and L.J. Sham, Phys. Rev. 140, A1133 (1965).
[3] J.P. Perdew and A. Zunger, Phys. Rev. B23, 5048 (1981).
[4] T. Koga, T. Tatewaki, and T. Shimazaki, Chem. Phys. Lett. 328, 473 (2000).
[5] D. Porezag *et al.*, Phys. Rev. B51, 12947 (1995).
[6] J. Widany *et al.*, Phys. Rev. B53, 4443 (1996).

Appendix 5. Hartree-Fock Theory

The full Schrödinger equation for an atom or molecule is easy to state but impossible to solve, generally. The problem lies primarily in the electron-electron interaction, which implies that the full wave function is a complicated function of all electron coordinates $\psi(\vec{r}_1,....,\vec{r}_N)$. In contrast, the wave function for *N* independent electrons is simply an antisymmetrized product of *N* single-electron orbitals, i.e. a Slater determinant. The idea of the Hartree-Fock approach is to *approximate* the full wave function by a Slater determinant. Applying the variational technique, the orbitals used in constructing the Slater determinant are found by minimizing the energy expectation value for the ground state. Hence, in this sense, the best possible Slater determinant is obtained. In this appendix, we explain how this is achieved in practice and illustrate the general principle through some examples.

The full Schrödinger equation for a molecule can be formulated as

$$\hat{H}\psi = E\psi, \quad \hat{H} = \sum_i \hat{h}(\vec{r}_i) + \sum_{i<j} \frac{1}{|\vec{r}_i - \vec{r}_j|} \tag{A5.1}$$

where $\hat{h}(\vec{r}_i)$ is the Hamiltonian of the *i*'th electron in the absence of e-e interactions, i.e.

$$\hat{h}(\vec{r}) = -\frac{1}{2}\nabla^2 + V_N(\vec{r}), \quad V_N(\vec{r}) = -\sum_I \frac{Z_I}{|\vec{r} - \vec{R}_I|}. \tag{A5.2}$$

Here, V_N is the potential produced by the nuclei with charges $Z_I e$ and positions $\vec{R}_I$. Note that atomic units (distances and energies measured in Bohrs and Hartrees, respectively) are applied. Now, as explained above, we proceed by approximating the exact wave function by a Slater determinant

$$\psi(\vec{r}_1,....,\vec{r}_N) \approx \frac{1}{\sqrt{N!}} \begin{vmatrix} \varphi_1(\vec{r}_1) & \varphi_1(\vec{r}_2) & ... & \varphi_1(\vec{r}_N) \\ \varphi_2(\vec{r}_1) & \varphi_2(\vec{r}_2) & ... & \varphi_2(\vec{r}_N) \\ \vdots & \vdots & & \vdots \\ \varphi_N(\vec{r}_1) & \varphi_N(\vec{r}_2) & ... & \varphi_N(\vec{r}_N) \end{vmatrix}.$$

More precisely, the general case is that a linear combination of a few distinct Slater determinants is applied. Hence, we will need matrix elements between ψ and a separate slater determinant ψ' constructed by replacing all the orbitals φ_i above by new ones φ_i'. It may then happen, of course, that $\varphi_i' = \varphi_i$ for certain *i* but generally some of the orbitals will differ. Using the antisymmetry and the orthogonality between orbitals it can then be shown that

$$\langle\psi|\sum_i \hat{h}(\vec{r}_i)|\psi'\rangle = \begin{cases} \sum_i h_{ii} & \text{if } \psi = \psi' \\ \langle\varphi_k|\hat{h}(\vec{r})|\varphi_k'\rangle & \text{if only } \varphi_k \neq \varphi_k' \\ 0 & \text{more than one orbital differ.} \end{cases}$$

Here, $h_{ij} = \langle\varphi_i|\hat{h}(\vec{r})|\varphi_j\rangle$. Similarly, for the e-e term

$$\langle\psi|\sum_{i<j}\frac{1}{r_{ij}}|\psi'\rangle = \begin{cases} \sum_{i<j}\left(J_{ij} - K_{ij}\right) & \text{if } \psi = \psi' \\ \sum_l \left(\langle\varphi_k\varphi_l|\frac{1}{|\vec{r}-\vec{r}'|}|\varphi_k'\varphi_l\rangle - \langle\varphi_k\varphi_l|\frac{1}{|\vec{r}-\vec{r}'|}|\varphi_l\varphi_k'\rangle\right) & \text{if only } \varphi_k \neq \varphi_k' \\ \langle\varphi_k\varphi_l|\frac{1}{|\vec{r}-\vec{r}'|}|\varphi_k'\varphi_l'\rangle - \langle\varphi_k\varphi_l|\frac{1}{|\vec{r}-\vec{r}'|}|\varphi_l'\varphi_k'\rangle & \text{if only } \varphi_k \neq \varphi_k' \text{ and } \varphi_l \neq \varphi_l' \\ 0 & \text{more than two orbitals differ.} \end{cases}$$

Here, we introduced the compact notation $J_{ij} = V_{ijij}, K_{ij} = V_{ijji}$ with

$$V_{ijkl} = \langle\varphi_i\varphi_j|\frac{1}{|\vec{r}-\vec{r}'|}|\varphi_k\varphi_l\rangle \equiv \iint \frac{\varphi_i^*(\vec{r})\varphi_j^*(\vec{r}')\varphi_k(\vec{r})\varphi_l(\vec{r}')}{|\vec{r}-\vec{r}'|}d^3r d^3r'$$

The *J* and *K* integrals are known as Coulomb and exchange integrals, respectively. It follows that the energy expectation value for a single slater determinant is

$$E \approx E_{HF} \equiv \langle\psi|\hat{H}|\psi\rangle = \sum_i h_{ii} + \sum_{i<j}\left(J_{ij} - K_{ij}\right). \tag{A5.3}$$

Minimizing this expression with respect to the orbitals $\varphi_i(\vec{r})$ is a classic exercise in variational calculus. Eventually, it is found that the orbitals satisfy the Hartree-Fock equations $\hat{F}\varphi_i = E_i\varphi_i$, where $\hat{F}$ is the Fock operator

$$\hat{F} = -\frac{1}{2}\nabla^2 + V_N(\vec{r}) + V_H(\vec{r}) + \hat{V}_x, \tag{A5.4}$$

where V_N is the nuclear potential given by Eq.(A4.1), V_H is the Hartree potential determined by the electron density $n(\vec{r}) = \sum_i |\varphi_i(\vec{r})|^2$ via

$$V_H(\vec{r}) = \int \frac{n(\vec{r}')}{|\vec{r}-\vec{r}'|}d^3r',$$

and $\hat{V}_x$ is the exchange operator defined by its action on a state

$$\hat{V}_x\varphi_i(\vec{r}) = -\sum_j \int \frac{\varphi_j^*(\vec{r}')\varphi_i(\vec{r}')}{|\vec{r}-\vec{r}'|} d^3r' \varphi_j(\vec{r}) .$$

It is noted that the Hartree and exchange contributions from the *i*'th orbital to the equation for the same orbital cancel and can therefore be left out.

So far, we have completely suppressed the spin degrees of freedom. In fact, all integrals over space should be supplemented by additional "integrals" over spin. This is of no consequence for the Coulomb integrals and Hartree potential as orbitals only appear in absolute squares. However, the exchange operator terms and exchange integrals are only non-zero if orbitals φ_i and φ_j have identical spins.

A5.1 Two-Electron Systems

For systems having only two electrons, such as the He atom, the ground state is a space symmetric spin singlet, i.e. the two orbitals constructing the Slater determinant have identical spatial parts but opposite spin parts. We write these as φ_1^+ and φ_1^- with the understanding that φ_1^+ means $\varphi_1(\vec{r})|\uparrow\rangle$ and φ_1^- means $\varphi_1(\vec{r})|\downarrow\rangle$. In compressed notation, the ground state Slater determinant is then written $\psi_0 = |\varphi_1^+\varphi_1^-|$. The Hartree-Fock equations for the two orbitals are identical and given by

$$\left\{-\frac{1}{2}\nabla^2 + V_N(\vec{r}) + V_H(\vec{r})\right\}\varphi_1(\vec{r}) = E_1\varphi_1(\vec{r}), \quad V_H(\vec{r}) = \int \frac{|\varphi_1(\vec{r}')|^2}{|\vec{r}-\vec{r}'|} d^3r' . \quad \text{(A5.5)}$$

This Hartree potential, in fact, only includes the effect of one electron because of the cancellation between Hartree and exchange self-interaction terms discussed above. The total energy, in turn, is given by

$$E = E_1 + h_{11}, \quad h_{11} = \langle\varphi_1|\hat{h}|\varphi_1\rangle, \quad \hat{h} = -\frac{1}{2}\nabla^2 + V_N(\vec{r}) .$$

For the He atom, one actually finds $E_1 = -0.9176$ and $E = -2.8612$, both in Hartree units. This result is found using the Gaussian matrix elements of the previous appendix as well as the Coulomb integral for *s*-orbitals $\psi_i = \exp\{-b_i r^2\}$

(s,s) Coulomb integral $$V_{ijkl}^{ssss} \equiv \iint \frac{\psi_i(\vec{r})\psi_j(\vec{r})\psi_k(\vec{r}')\psi_l(\vec{r}')}{|\vec{r}-\vec{r}'|} d^3r d^3r' = \frac{2\pi^{5/2}}{b_{ij}b_{kl}\left(b_{ij}+b_{kl}\right)^{1/2}} .$$

Next, we turn to the excited states. If only one of the ground state orbitals is replaced by an excited one, we find four options, i.e.

$$|\varphi_2^+\varphi_1^-|,\quad |\varphi_2^-\varphi_1^+|,\quad |\varphi_2^+\varphi_1^+|,\quad |\varphi_2^-\varphi_1^-|.$$

These four states actually form one triplet T (with three members of equal energy) and one singlet S given by

$$\psi_T = \begin{cases} |\varphi_2^+\varphi_1^+| \\ \left\{|\varphi_2^+\varphi_1^-| + |\varphi_2^-\varphi_1^+|\right\}/\sqrt{2} \\ |\varphi_2^-\varphi_1^-| \end{cases},\quad \psi_S = \left\{|\varphi_2^+\varphi_1^-| - |\varphi_2^-\varphi_1^+|\right\}/\sqrt{2}\,.$$

The three triplet components correspond to the three possible spin projections 1, 0, -1 of a state with a total spin 1 while the singlet is a spin 0 state. Using the matrix element rules above, it is found that the energies are given by

$$E_T = h_1 + h_2 + J_{12} - K_{12},\quad E_S = h_1 + h_2 + J_{12} + K_{12}\,.$$

Expanding the Koga basis with two diffuse Gaussians $b \in \{0.04, 0.01\}$ to accurately describe the 2s state in helium, one find in this case $E_T = -2.1241$ and $E_S = -2.0786$. Relative to the ground state, we then find energy differences of $E_T - E_0 = 20.06$ eV and $E_S - E_0 = 21.29$ eV. The corresponding experimental values are 19.82 eV and 20.62 eV, respectively. The absolute energies found experimentally are $E_0 = -2.9037$, $E_T = -2.1752$, and $E_S = -2.1459$. In a similar manner, excited states of the types $|\varphi_s\varphi_p|$ and $|\varphi_p\varphi_p|$ can be obtained. To this end, we need integrals

(s,p) Coulomb integral $$V_{ijkl}^{sspp} = \frac{\pi^{5/2}}{b_{ij}b_{kl}\left(b_{ij}+b_{kl}\right)^{3/2}}\left(3+2\frac{b_{ij}}{b_{kl}}\right)$$

(s,p) Exchange integral $$W_{ijkl}^{spsp} = \frac{\pi^{5/2}}{b_{ij}b_{kl}\left(b_{ij}+b_{kl}\right)^{3/2}}$$

(p,p) Coulomb integral $$V_{ijkl}^{pppp} = \frac{3\pi^{5/2}\left(2b_{ij}^2+2b_{kl}^2+5b_{ij}b_{kl}\right)}{2b_{ij}^2b_{kl}^2\left(b_{ij}+b_{kl}\right)^{5/2}}$$

(p,p) Exchange integral $$W_{ijkl}^{pppp} = \frac{\pi^{5/2}\left(2b_{ij}^2+2b_{kl}^2+7b_{ij}b_{kl}\right)}{2b_{ij}^2b_{kl}^2\left(b_{ij}+b_{kl}\right)^{5/2}}.$$

A5.2 Configuration Interaction

It is instructive to see how these excited states emerge from a more general matrix problem. To this end, we will try expanding the wave function in a large basis

$$\psi = \sum_{n,m} c_{nm} |\varphi_n^+ \varphi_m^-|.$$

Hence, we need the matrix elements

$$|\varphi_n^+ \varphi_m^-| \hat{H} |\varphi_k^+ \varphi_l^-| = \langle \varphi_n | \hat{h} | \varphi_k \rangle \delta_{ml} + \langle \varphi_m | \hat{h} | \varphi_l \rangle \delta_{nk} + V_{nmkl}.$$

We first consider the matrix element coupling the ground state to a singly excited state, i.e.

$$|\varphi_n^+ \varphi_1^-| \hat{H} |\varphi_1^+ \varphi_1^-| = \langle \varphi_n | \hat{h} | \varphi_1 \rangle + V_{n111}.$$

However, if we consider the Hartree-Fock equation Eq.(A1.5) and multiply from the left by φ_n^* we immediately see that this matrix element vanishes. This is a particular instance of a completely general theorem (Brillouin's theorem) that matrix elements between the ground state and singly excited configurations vanish. If we limit ourselves to the basis states $|\varphi_1^+ \varphi_1^-|, |\varphi_1^+ \varphi_2^-|, |\varphi_2^+ \varphi_1^-|, |\varphi_2^+ \varphi_2^-|$ we find the configuration interaction matrix

$$H_{CI} = \begin{pmatrix} 2h_{11} + V_{1111} & 0 & 0 & V_{1122} \\ 0 & h_{11} + h_{22} + V_{1212} & V_{1221} & h_{12} + V_{1222} \\ 0 & V_{1221} & h_{11} + h_{22} + V_{1212} & h_{12} + V_{1222} \\ V_{1122} & h_{12} + V_{1222} & h_{12} + V_{1222} & 2h_{22} + V_{2222} \end{pmatrix}.$$

We see that if the last column and row are erased we get precisely the ground state and singlet and triplet excited states as eigenstates, realizing that $V_{1212} = J_{12}$ and $V_{1221} = K_{12}$. If, however, the full matrix is retained we find corrections to both the ground and excited states. In this simple example, we only considered the two lowest orbitals. A much better result is found if we use all 8 orbitals found above. In this manner, we find $E_0 = -2.8785$, $E_T = -2.1734$, and $E_S = -2.1431$.

If, eventually, the basis is extended to include *p*-type orbitals as well, the agreement with experiments becomes really impressive. In fact, the *s*-state cannot couple to an arbitrary *p*-state. The reason is that the ground state is a singlet with vanishing total angular momentum and, naturally, vanishing projection. In the standard $|J, M\rangle$ notation it is of the type $|0,0\rangle$. Now forming states from two *p*-orbital leads to 9

combination in total. A general un-coupled member of this basis is $Y_{1m_1}(\theta_1,\phi_1)Y_{1m_2}(\theta_2,\phi_2)$, which we will denote $|m_1,m_2|$ for brevity. In the coupled representation, we find a *D*, a *P* and an *S* term. The *D* term includes the states $|2,M=-2,\cdots,2\rangle$ while the *P* and *S* terms are $|1,M=-1,\cdots,1\rangle$ and $|0,0\rangle$. Among these nine states, only the last couples to the ground state. We find it using the usual procedure of stepping down from a known state (here $|2,2\rangle=|1,1|$) and requiring orthogonality. It then follows that $|p_0\rangle\equiv|0,0\rangle=3^{-1/2}\left(|0,0|-|1,-1|-|-1,1|\right)$. In terms of polar angles this state is $|0,0\rangle=3^{1/2}/(4\pi)\left(\cos\theta_1\cos\theta_2+\sin\theta_1\sin\theta_2\cos(\phi_1-\phi_2)\right)$. For this particular state, we need the integral

$$(\mathrm{p}_0,\mathrm{p}_0)\text{ Coulomb integral}\quad V_{ijkl}^{p_0p_0p_0p_0}=\frac{3\pi^{5/2}\left(10b_{ij}^2+10b_{kl}^2+29b_{ij}b_{kl}\right)}{10b_{ij}^2b_{kl}^2\left(b_{ij}+b_{kl}\right)^{5/2}}.$$

Thus, full configuration interaction using the lowest 8 *s*- and *p*-orbitals leads to $E_0=-2.9005$, $E_T=-2.1746$, and $E_S=-2.1452$.

A5.3 Highly Excited States

While the Gauss basis is excellent for localized states such as the ground state, it is not ideal for delocalized states such as highly excited ones or ionized states. Moreover, the very low number of basis states means that only a few states can be obtained from the Gauss basis. A better approach for delocalized states is to imagine the system embedded in a spherical quantum well with infinite barriers and use the quantum well eigenfunctions or similar functions to construct a basis. The states are taken to be

$$|l,n\rangle=\psi_n^{(l)}(r)=r^{l-1}\sin\left(\frac{n\pi r}{R}\right),\quad n=1,2,3,...$$

Here, *l* is the angular momentum quantum number. The overlaps for *s*, *p*, and *d* states are (with the notation $\overline{\delta}_{nm}\equiv1-\delta_{nm}$)

$$\langle0,n|0,m\rangle=\frac{R}{2}\delta_{nm},\quad\langle1,n|1,m\rangle=\frac{R^3}{12}\left(2-\frac{3}{\pi^2n^2}\right)\delta_{nm}+\frac{4(-1)^{n+m}mnR^3}{\pi^2(n^2-m^2)^2}\overline{\delta}_{nm}$$

$$\langle2,n|2,m\rangle=\frac{R^5(15-10\pi^2n^2+2\pi^4n^4)}{20\pi^4n^4}\delta_{nm}+\frac{8(-1)^{n+m}mnR^5\left[\pi^2(n^2-m^2)^2-12(n^2+m^2)\right]}{\pi^4(n^2-m^2)^4}\overline{\delta}_{nm}.$$

Similarly, for the matrix elements of the kinetic energy $\hat{T}=-\frac{1}{2r}\frac{d^2}{dr^2}r+\frac{l(l+1)}{2r^2}$

$$\langle 0,n|\hat{T}|0,m\rangle = \frac{\pi^2 n^2}{4R}\delta_{nm}, \quad \langle 1,n|\hat{T}|1,m\rangle = \frac{R}{24}\left(15+2\pi^2 n^2\right)\delta_{nm} + \frac{(-1)^{n+m} mn(n^2+m^2)R}{(n^2-m^2)^2}\bar{\delta}_{nm},$$

$$\langle 2,n|\hat{T}|2,m\rangle = \frac{R^3(-105+70\pi^2 n^2+6\pi^4 n^4)}{120\pi^2 n^2}\delta_{nm} + \frac{2(-1)^{n+m} mnR^3\left[\pi^2(n^2-m^2)^2(n^2+m^2)-2(m^4+n^4+22n^2m^2)\right]}{\pi^2(n^2-m^2)^4}\bar{\delta}_{nm}.$$

And for the Coulomb potential (with Ci as the cosine integral)

$$\langle 0,n|\frac{1}{r}|0,m\rangle = \frac{1}{2}(\gamma - \mathrm{Ci}(2\pi n) + \ln(2\pi n))\delta_{nm} - \frac{1}{2}\left(\mathrm{Ci}(\pi|n+m|) - \mathrm{Ci}(\pi|n-m|) + \ln\left|\frac{n-m}{n+m}\right|\right)\bar{\delta}_{nm},$$

$$\langle 1,n|\frac{1}{r}|1,m\rangle = \frac{R^2}{4}\delta_{nm} + \frac{2[(-1)^{n+m}-1]mnR^2}{\pi^2(n^2-m^2)^2}\bar{\delta}_{nm},$$

$$\langle 2,n|\frac{1}{r}|2,m\rangle = \frac{R^4}{8}\left(1-\frac{3}{\pi^2 n^2}\right)\delta_{nm} - \frac{6mnR^4\left[4[(-1)^{n+m}-1](n^2+m^2)-\pi^2(-1)^{n+m}(n^2-m^2)^2\right]}{\pi^4(n^2-m^2)^4}\bar{\delta}_{nm}.$$

As an application, we show below the density of singlet p-type states $|\varphi_s\varphi_p|$ for a He atom calculated using the basis above with 600 members and R = 1000 Bohr.

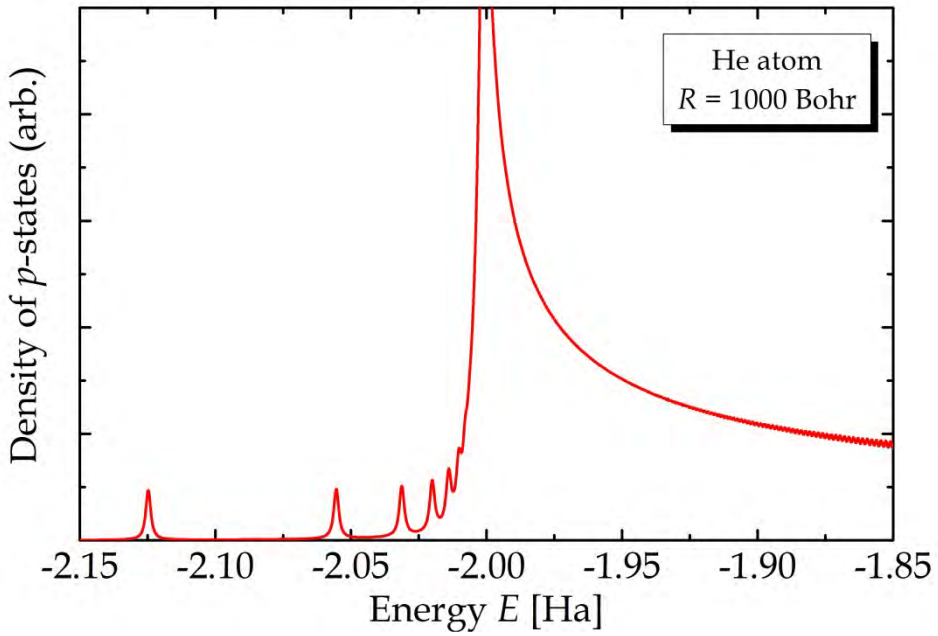


Figure A5.1. Density of p-type states in He using a spherical well basis with 600 members. The plot is broadened by 1 mHa.

Appendix 6. Jellium Model of Nanostructures

The problem of finding electron eigenstates in nanostructures can be daunting if electron-electron (e-e) interactions cannot be ignored. In the previous appendices we saw some approaches to the problem and applied these to atomic systems. In a metallic nanostructure, e-e interactions are expected to be important because of the high electron density. On the other hand, many electron states are expected to be highly delocalized in a good metal. Hence, electrons in such states don't see the details of the attractive nuclear potentials but, rather, a smoothed out version. This is the background for the so-called jellium model.

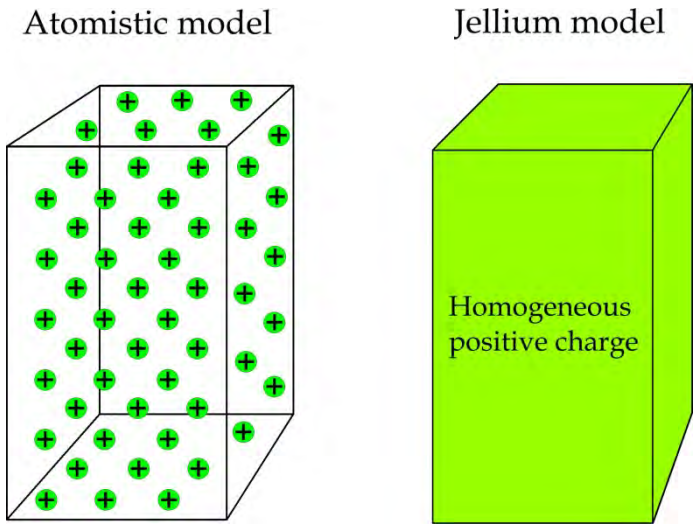


Figure A6.1. An atomistic ion charge distribution and the jellium approximation.

We may think of valence electrons as moving in an effective potential created by the positive ion cores. If the actual distribution of positive charges determined by the crystal lattice is replaced by a homogeneous positive charge density we find the picture in Fig. A6.1. In the jellium approach, this approximation is made for the ion charges and the electron charge is determined from the self-consistent quantum problem of electrons moving in the approximate potential. Hence, e-e interactions are retained in the model. If each atom supplies Z_v valence electrons, the homogeneous, positive jellium charge density is given by $n_+ = Z_v N_+$, where N_+ is the atomic density, i.e. the number of atoms per volume. If exchange and correlation is included, electrons are governed by a Schrödinger equation

$$\left\{-\frac{1}{2}\nabla^2 + V(\vec{r}) + V_{xc}(\vec{r})\right\}\varphi(\vec{r}) = E\varphi(\vec{r}), \qquad \text{(A6.1)}$$

where $V(\vec{r})$ is the total Coulomb potential. We can now introduce the characteristic function of our nanostructure defined by

$$\chi(\vec{r}) = \begin{cases} 1 & \vec{r} \text{ inside the structure} \\ 0 & \vec{r} \text{ outside the structure.} \end{cases}$$

In this manner, the total charge density is $\rho(\vec{r}) = n_+\chi(\vec{r}) - n(\vec{r})$, where $n(\vec{r})$ is the electron charge density. In turn, the Coulomb potential is derived from the Poisson equation $\nabla^2 V(\vec{r}) = 4\pi\rho(\vec{r})$. The sign implies that V is the potential acting on negatively charged electrons. Below, we look in detail at the potential for some typical nanostructures in various dimensions: the sphere, the cylinder, and the slab.

Before looking into various low-dimensional applications, we briefly recall the results for a bulk material in the jellium picture. In this case, the electron density n is homogeneous and equal to n_+ everywhere to ensure neutrality. Thus, all classical Coulomb interactions cancel. The kinetic and exchange energies, however, are non-zero and the values per volume Ω are given by the expressions [1]

$$E_{kin}/\Omega = \frac{3^{5/3}\pi^{4/3}}{10} n^{5/3}, \quad E_x/\Omega = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3} n^{4/3}. \tag{A6.2}$$

The exchange potential in DFT is simply the density-derivative of the energy, i.e. $V_x(\vec{r}) = -\left(3n(\vec{r})/\pi\right)^{1/3}$ in agreement with Appendix 3.

A6.1 Low-Dimensional Poisson Equations

For 0-dimensional spheres, symmetry dictates that charges and potential are radially symmetric, i.e. functions of r alone. Thus, the Poisson equation reads

$$\frac{1}{r}\frac{d^2}{dr^2} rV(r) = 4\pi\rho(r).$$

Integrating twice immediately shows that

$$V(r) = \frac{4\pi}{r}\int_0^r \int_0^{r'} \rho(r'')\, r''\, dr''\, dr'$$

The innermost integral may be viewed as a function of r' and upon integration by parts we find

$$V(r)=\frac{4\pi}{r}\left\{\left[r'\int_0^{r'}\rho(r'')r''dr''\right]_{r'=0}^{r'=r}-\int_0^r\rho(r')(r')^2dr'\right\}$$
$$=\frac{4\pi}{r}\left\{r\int_0^r\rho(r')r'dr'-\int_0^r\rho(r')(r')^2dr'\right\}=4\pi\int_0^r\rho(r')\left\{r'-\frac{(r')^2}{r}\right\}dr'.$$

We can now evaluate the ion part of the potential $V_+(\vec{r})$. If the radius of the sphere is *R* we have

$$V_+(r)=4\pi n_+\int_0^{r_<}\left\{r'-\frac{(r')^2}{r}\right\}dr',\quad r_<=\min\{r,R\}.$$

Thus,

$$V_+(r)=4\pi n_+\left\{\frac{r_<^2}{2}-\frac{r_<^3}{3r}\right\}.$$

For 1-dimensional wires, the analysis is almost completely identical. All quantities now only depend on the radial coordinate *r* perpendicular to the wire axis. However, the starting point is a slightly different Poisson equation

$$\frac{1}{r}\frac{d}{dr}r\frac{d}{dr}V(r)=4\pi\rho(r).$$

After a few manipulations including, again, integration by parts we then find

$$V(r)=-4\pi\int_0^r\ln\frac{r'}{r}\rho(r')r'dr'.$$

For a wire of radius *R*, it follows that the ion part is

$$V_+(r)=-4\pi n_+\int_0^{r_<}\ln\frac{r'}{r}r'dr'=\pi n_+r_<^2\left\{1-2\ln\frac{r_<}{r}\right\},\quad r_<=\min\{r,R\}.$$

Finally, we look at the 2D slab or quantum well geometry. Taking the quantization dimension to be the *z*-axis we find

$$\frac{d^2}{dz^2}V(z)=4\pi\rho(z).$$

Hence, integrating this expression twice and using the integration by parts trick again leads to

$$V(z) = 4\pi \int_{-\infty}^{z} (z - z')\rho(z')dz'.$$

In this case, this ion charge can be written as $n_+\theta(d/2 - |z|)$ for a slab of width d. We then find

$$V_+(z) = 4\pi n_+ \begin{cases} 0 & z < -d/2 \\ (2z+d)^2/8 & |z| < d/2 \\ zd & z > d/2. \end{cases}$$

This asymmetric potential arises because the ion-system is not neutral. The total potential including both electrons and ions will be symmetrical in z, however. The ion potentials of 0D and 1D structures are illustrated in Fig. A6.2.

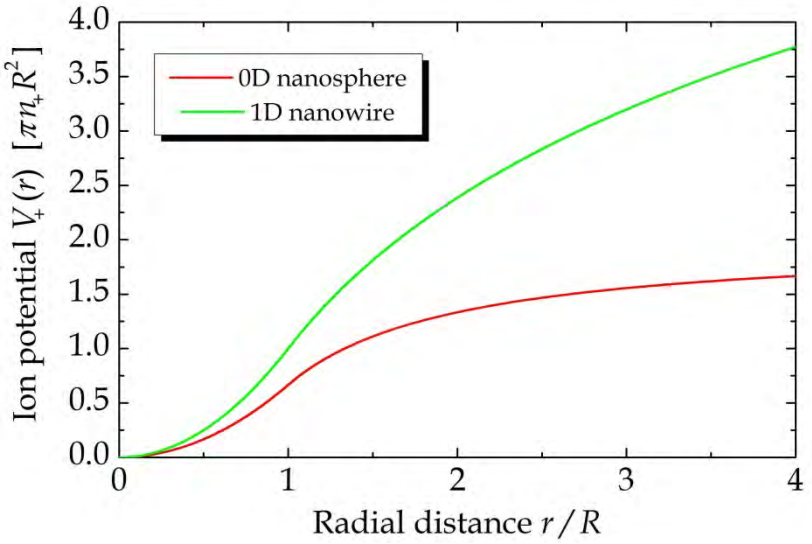


Figure A6.2. Ion Coulomb potentials of 0D and 1D cases in normalized units.

For the 0D potential the limiting value far from the particle is $2\pi n_+ R^2$ while the 1D potential keeps increasing logarithmically.

A6.2 Jellium Model of Nanowires

We now go into some detail in the nanowire case. In particular, we wish to see how quantities such as work function and surface energy vary with wire radius and approach the bulk values for large wires. The jellium approach clearly applies better to free electron like metals than e.g. transition metals. For this reason we will consider Na wires with a bulk charge density of $n_+ = 3.93 \cdot 10^{-3}$ Bohr^{-3}. The cylindrical

symmetry of the problem means that we adopt cylindrical coordinates (r,ϕ,z) and that eigenstates for a wire of length L are of the form

$$\psi_{kmn}(\vec{r}) = \frac{1}{\sqrt{2\pi L}} e^{ikz} e^{im\phi} R_{m,n}(r).$$

Here, k is the continuous k-vector, m is the angular quantum number and n labels the radial eigenstates for a particular m. If the corresponding eigenvalues are denoted $E_{m,n}$, integration of the density of states Eq. (A1.2) shows that the (zero temperature) density is

$$n(r) = \frac{\sqrt{2}}{\pi^2} \sum_{m,n} |R_{m,n}(r)|^2 \sqrt{E_F - E_{m,n}}\,\theta(E_F - E_{m,n}). \qquad \text{(A6.3)}$$

The radial functions can be obtained in many different ways and the results in the present appendix were found by expanding in a Gauss-Laguerre basis and using the accompanying quadrature for numerical integration [2]. We note that in addition to the Coulomb potential above, we also include the exchange term $V_x(r) = -(3n(r)/\pi)^{1/3}$. Finally, we add the so-called stabilized jellium correction [3], which amounts to the term $V_S(r) = \left[\frac{1}{5}(3\pi^2 n_+)^{2/3} - \frac{1}{4}(3n_+/\pi)^{1/3}\right]\theta(R-r)$. Similarly to standard DFT, the jellium approach requires an iterative procedure in order to obtain self-consistency. In general, convergence can be difficult to achieve and "mixing" schemes will be required during iteration. As an example, the final converged electron density and potential profiles are shown in Fig. A6.3.

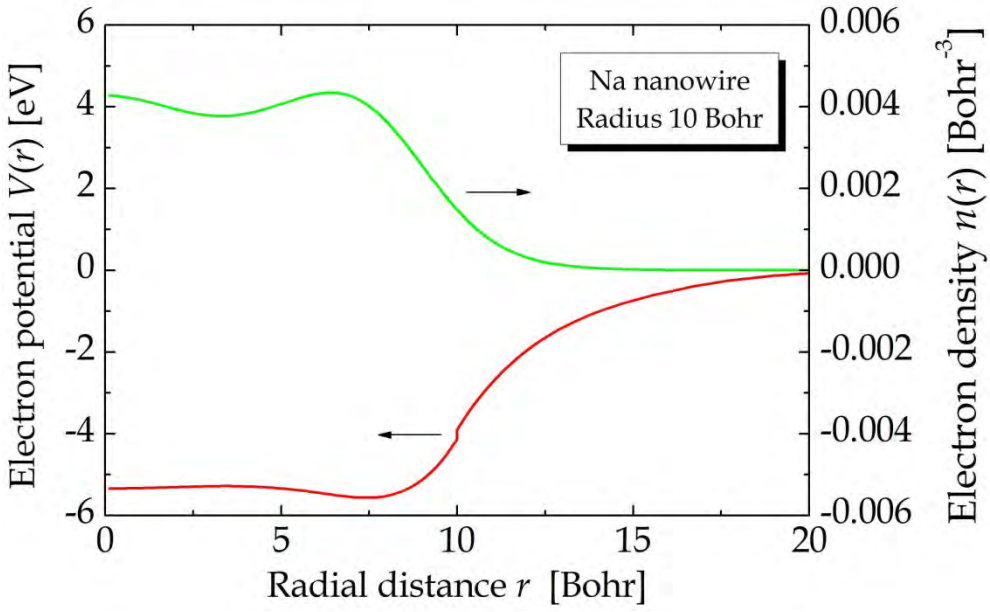


Figure A6.3. Converged total potential and density profiles for an R = 10 Bohr Na nanowire.

During the iterations the Fermi level must be continuously updated to ensure a correct integrated electron density, i.e. $\int_0^\infty n(r) r dr = n_+ R^2 / 2$. Eventually, the converged Fermi level E_F can be determined in this manner. In Fig. A6.4, the dependence of E_F on nanowire radius is illustrated. It is seen to vary in a series of peaks and troughs. A peak appears whenever a new subband (*n*,*m*) becomes occupied, i.e. when $E_{m,n}$ moves below E_F. For large enough wires, the Fermi level converges towards something like -2.2 eV. The work function in this limit is therefore 2.2 eV and compares reasonably to the experimental value of bulk Na of 2.35 eV. In addition, the surface energy E_S can be obtained. This quantity is obtained by looking at the difference between the energies of the actual nanowire and a sample with the same shape but energy density given by the bulk expression. Writing this difference as a coefficient E_S times area $2\pi RL$ we eventually find

$$E_S = \frac{1}{2\pi RL}\left(E_{tot} - E_{bulk}\right),$$

where E_{bulk} is

$$E_{bulk} = \pi R^2 L\left[\frac{1}{10}\left(3\pi^2\right)^{2/3} n_+^{5/3} - \frac{3}{4}\left(3/\pi\right)^{1/3} n_+^{4/3}\right].$$

The total nanowire energy is mainly determined by the sum of occupied energy levels E_{occ} that follows from an integral over the density of states

$$\begin{aligned} E_{occ} &= \pi R^2 L \int_0^{E_F} E D_1(E) dE = L\frac{\sqrt{2}}{\pi}\sum_{m,n}\int_0^{E_F} E\frac{\theta(E-E_{m,n})}{\sqrt{E-E_{m,n}}}dE \\ &= L\frac{2^{3/2}}{3\pi}\sum_{m,n}\left(E_F + 2E_{m,n}\right)\sqrt{E_F - E_{m,n}}\,\theta(E_F - E_{m,n}). \end{aligned}$$

However, as always in DFT, we need to make corrections for double counting of Coulomb and exchange terms. The final *R*- dependence is shown in Fig. A6.4 as the green curve. We see that the bulk limit is around $3\ \mathrm{meV/Bohr^2}$. Experimentally, the values is more like $5\ \mathrm{meV/Bohr^2}$. Hence, forming a surface is predicted to cost energy, which is clearly correct since otherwise bulk Na would spontaneously disintegrate into nanoparticles.

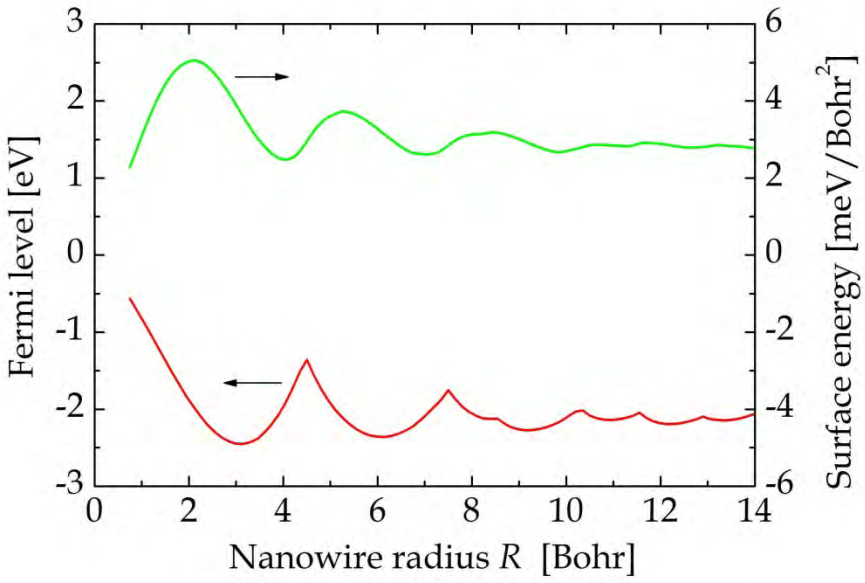


Figure A6.4. Fermi level and surface energy versus Na nanowire radius.

Particularly low values of the surface energy are found around R = 4, 7, and 9.5 Bohr and these "magic" radii are expected to correspond to relatively stable nanowire geometries. Similar effects are seen for nanospheres and have been verified experimentally.

References

[1] R. G. Parr and W. Yang *Density-functional Theory of Atoms and Molecules* (Oxford, New York, 1989).
[2] W.H. Press *et al. Numerical Recipes* (Cambridge, New York, 1987).
[3] J.P. Perdew, H. Q. Tran, and E.D. Smith, Phys. Rev. B42, 11627 (1990).
[4] W. A. de Haar, Rev. Mod. Phys. 65, 611 (1993).

Appendix 7. Spin-Orbit Interaction

The spin-orbit (SO) interaction is an important relativistic correction to the Schrödinger equation that leads to several notable consequences. Primarily, the SO interaction splits degenerate levels and is therefore observed in e.g. absorption spectra. Secondly, the interaction mixes levels of definite spin, which means that selection rules are relaxed. One effect of this is the phenomenon of phosphorescence discussed in Chapter 33. The correction to the Schrödinger equation can be derived heuristically based on arguments treating the electron as affected by an effective magnetic field produced by the nucleus. The trick is to choose a coordinate system, in which the electron is stationary and the nucleus orbits the electron. Unfortunately, this approach does not properly treat relativistic requirements for the change in observation point and leads to a result that is precisely twice as large as the correct one. For this reason, we will use the full relativistic Dirac equation as our starting point and show how the SO term emerges in the non-relativistic limit.

The time-independent Dirac equation for single electron in the absence of magnetic fields can be written as

$$\begin{Bmatrix} E-V & -c\vec{\sigma}\cdot\vec{p} \\ -c\vec{\sigma}\cdot\vec{p} & E+2mc^2-V \end{Bmatrix}\cdot\begin{pmatrix}\varphi_A \\ \varphi_B\end{pmatrix}=0\,. \tag{A7.1}$$

Here, V is electrostatic potential, $\vec{\sigma}$ is the vector of Pauli spin matrices, and E is the usual (non-relativistic) energy that does not include the rest mass term mc^2. Also, both components of the wave function are, in fact, two-component "spinors" themselves. The lower of these equations can be solved to give

$$\varphi_B=\frac{1}{E+2mc^2-V}c\vec{\sigma}\cdot\vec{p}\varphi_A=\frac{1}{2mc}\cdot\frac{1}{1+\dfrac{E-V}{2mc^2}}\vec{\sigma}\cdot\vec{p}\varphi_A\approx\frac{1}{2mc}\left\{1-\frac{E-V}{2mc^2}\right\}\vec{\sigma}\cdot\vec{p}\varphi_A\,.$$

The last approximation relies on the rest mass energy being much larger than the other terms. When substituted back into the upper equation we find

$$E\varphi_A=V\varphi_A+\frac{1}{2m}\vec{\sigma}\cdot\vec{p}\left\{1-\frac{E-V}{2mc^2}\right\}\vec{\sigma}\cdot\vec{p}\varphi_A\,. \tag{A7.2}$$

The spin matrices have the following multiplication properties: $\sigma_x^2=\sigma_y^2=\sigma_z^2=1$ and $\sigma_x\sigma_y=-\sigma_y\sigma_x=i\sigma_z$ as well as cyclic permutations. Using these, it can then be shown for general spatial operators $\vec{a}$ and $\vec{b}$ that $(\vec{\sigma}\cdot\vec{a})(\vec{\sigma}\cdot\vec{b})=\vec{a}\cdot\vec{b}+i\vec{\sigma}\cdot(\vec{a}\times\vec{b})$. Hence $(\vec{\sigma}\cdot\vec{p})^2=p^2$, and we immediately see that ignoring the second term in the braces leads to usual Schrödinger equation. It is clear that, in general, several corrections are

found. The relevant term for us is $\vec{\sigma}\cdot\vec{p}V\vec{\sigma}\cdot\vec{p}$. The first p factor operates on the V as well as on the wave function. We judiciously select the first term, i.e. $\vec{\sigma}\cdot(\vec{p}V)\vec{\sigma}\cdot\vec{p}$, where the parenthesis is meant to indicate that the operator only acts inside. Since $\vec{p}V=-ie\hbar\vec{\mathcal{E}}$, with $\vec{\mathcal{E}}$ the electric field, this is $-ie\hbar(\vec{\sigma}\cdot\vec{\mathcal{E}})(\vec{\sigma}\cdot\vec{p})=-ie\hbar\vec{\mathcal{E}}\cdot\vec{p}+e\hbar\vec{\sigma}\cdot(\vec{\mathcal{E}}\times\vec{p})$. The second term leads to the interaction

$$H_{SO}=\frac{e\hbar}{4m^2c^2}\vec{\sigma}\cdot(\vec{\mathcal{E}}\times\vec{p}).$$

This is the celebrated spin-orbit interaction. For a spherically symmetric potential we have $e\vec{\mathcal{E}}=\nabla V(r)=(dV/dr)\vec{r}/r$ and so

$$H_{SO}=\frac{\hbar^2}{4m^2c^2}\cdot\frac{1}{r}\frac{dV}{dr}\vec{\sigma}\cdot\vec{l},\quad \hbar\vec{l}=\vec{r}\times\vec{p}\,.$$

Also, in a spherically symmetric systems, the atomic eigenstates are of the form $\psi_{nlm\sigma}=R_{nl}(r)Y_{lm}(\theta,\varphi)\chi_\sigma$, where χ_σ is the spin part. It is readily demonstrated that among all the corrections stemming from Eq.(A7.2), only H_{SO} mixes states of different angular momentum projections m and m'. Without mixing, these are degenerate and therefore H_{SO} has important consequences for the corrected atomic energy levels. Remembering that $\vec{\sigma}=2\vec{s}$, the matrix elements are given by

$$\left\langle\psi_{nlm\sigma}\left|H_{SO}\right|\psi_{n'l'm'\sigma'}\right\rangle=\varsigma\left\langle Y_{lm}\left|\vec{l}\right|Y_{l'm'}\right\rangle\cdot\left\langle\chi_\sigma\left|\vec{s}\right|\chi_{\sigma'}\right\rangle,\quad \varsigma=\frac{\hbar^2}{2m^2c^2}\int_0^\infty R_{nl}(r)\frac{dV}{dr}R_{n'l'}(r)rdr\,.$$

The $\vec{l}$ operator only couples states of identical angular momentum l. Hence, matrix elements are only non-vanishing for p states coupling to other p states etc. and obviously s states are unaffected. Their calculation can be carried out in various ways but a direct one involves introducing $j_\pm\equiv j_x\pm ij_y$ and writing

$$\vec{l}\cdot\vec{s}=\tfrac{1}{2}l_+s_-+\tfrac{1}{2}l_-s_++l_zs_z\,.$$

Using the properties of the operators $j_\pm\left|j,m\right\rangle=\sqrt{j(j+1)-m(m\pm1)}\left|j,m\pm1\right\rangle$ we easily find the matrix elements. For the p-p block using the order $Y_x\uparrow,Y_y\uparrow,Y_z\uparrow,Y_x\downarrow,Y_y\downarrow,Y_z\downarrow$ we get

$$\overleftrightarrow{H}_{pp}=\frac{\varsigma}{2}\begin{pmatrix}\overleftrightarrow{H}_{\uparrow\uparrow} & \overleftrightarrow{H}_{\uparrow\downarrow}\\ \overleftrightarrow{H}_{\downarrow\uparrow} & \overleftrightarrow{H}_{\downarrow\downarrow}\end{pmatrix},\quad \overleftrightarrow{H}_{\uparrow\uparrow}=-\overleftrightarrow{H}_{\downarrow\downarrow}=\begin{pmatrix}0 & -i & 0\\ i & 0 & 0\\ 0 & 0 & 0\end{pmatrix},\quad \overleftrightarrow{H}_{\uparrow\downarrow}=\overleftrightarrow{H}_{\downarrow\uparrow}^{\dagger}=\begin{pmatrix}0 & 0 & 1\\ 0 & 0 & -i\\ -1 & i & 0\end{pmatrix}.$$

Similarly for *d*-*d* coupling with ordering $Y_{xy}, Y_{yz}, Y_{zx}, Y_{x^2-y^2}, Y_{3z^2-r^2}$

$$\overleftrightarrow{H}_{\uparrow\uparrow} = -\overleftrightarrow{H}_{\downarrow\downarrow} = \begin{pmatrix} 0 & 0 & 0 & 2i & 0 \\ 0 & 0 & i & 0 & 0 \\ 0 & -i & 0 & 0 & 0 \\ -2i & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad \overleftrightarrow{H}_{\uparrow\downarrow} = \overleftrightarrow{H}_{\downarrow\uparrow}^{\dagger} = \begin{pmatrix} 0 & 1 & -i & 0 & 0 \\ -1 & 0 & 0 & -i & -i\sqrt{3} \\ i & 0 & 0 & -1 & \sqrt{3} \\ 0 & i & 1 & 0 & 0 \\ 0 & i\sqrt{3} & -\sqrt{3} & 0 & 0 \end{pmatrix}.$$

The eigenvalues of H_{pp} are $-\varsigma$ (degeneracy 2) and $\varsigma/2$ (degeneracy 4). For H_{dd} they are $-3\varsigma/2$ (degeneracy 4) and ς (degeneracy 6). The general pattern for angular momentum l is a set of eigenvalues $-\varsigma(l+1)/2$ and $\varsigma l/2$. Hence, the spin-orbit splitting is $E_{SO} = \varsigma(l+\frac{1}{2})$. In the hydrogen-like case with $V = -Ze^2/4\pi\varepsilon_0 r$, we find

$$\varsigma = \frac{Ze^2\hbar^2}{8\pi\varepsilon_0 m^2c^2}\int_0^\infty R_{nl}(r)\frac{1}{r}R_{n'l'}(r)dr\,.$$

We saw above that only the $l = l'$ case is relevant. If, in addition, we take $n = n'$ we find

$$\varsigma_{nl} = \frac{Ze^2\hbar^2}{8\pi\varepsilon_0 m^2c^2}\int_0^\infty R_{nl}^2(r)\frac{1}{r}dr = \frac{Z^4\alpha^2}{n^3 l(l+1)(2l+1)}Ha\,.$$

In the general case, the Kohn-Sham potential of density functional theory can be applied to estimate the spin-orbit coupling. As an example, for Al we find $\varsigma_{2p} = 11.5\ \text{meV}$ corresponding to a splitting of $E_{SO} = 17.1\ \text{meV}$ while the experimental splitting is 13.9 meV. Similarly for the *d*-electron case of Sc we calculate $\varsigma_{3d} = 12.5\ \text{meV}$ meaning a splitting of $E_{SO} = 31.3\ \text{meV}$ with the experiment at 21 meV. We therefore see that our values seem to systematically overestimate the true ones.

Appendix 8. Fermi's Golden Rule and Scattering Processes

In this appendix, we tie the connection to some other important formulations and results from perturbation theory. Primarily, we will derive Fermi's golden rule as the transition probability from an initial state into various final states. Then, this result will be applied to compute the cross section for electrons scattering off a screened Coulomb potential. This leads, in the limit on vanishing screening, to the Rutherford scattering cross section. As a starting point, we consider the perturbation results from Chapter 1 in the case, where only a particular initial state $|\varphi_i\rangle$ is occupied before the perturbation, i.e. $a_n^{(0)} = \delta_{ni}$. To first order, we have for a particular final state $|\varphi_f\rangle$

$$\dot{a}_f^{(1)} = \frac{1}{2i\hbar}\left\{\langle\varphi_f|\hat{H}_1|\varphi_i\rangle e^{-i\omega t} + \langle\varphi_f|\hat{H}_1^\dagger|\varphi_i\rangle e^{i\omega t}\right\} e^{iE_{fi}t/\hbar}.$$

Integrating this result and introducing broadening then determines the correction to the expansion coefficient a_f, c.f. Eq.(1.7)

$$a_f^{(1)} = -\frac{1}{2}\left\{\frac{\langle\varphi_f|\hat{H}_1|\varphi_i\rangle e^{-i\omega t} e^{iE_{fi}t/\hbar}}{E_{fi} - \hbar\omega - i\hbar\Gamma} + \frac{\langle\varphi_f|\hat{H}_1^\dagger|\varphi_i\rangle e^{i\omega t} e^{iE_{fi}t/\hbar}}{E_{fi} + \hbar\omega - i\hbar\Gamma}\right\}.$$

We will be interested in computing the probability $P_{fi} = |a_f|^2$ that the final state is occupied due to the perturbation. The rate, at which the probability changes, gives the scattering rate from initial to final state, i.e. $dP_{fi}/dt = a_f\dot{a}_f^* + \dot{a}_f a_f^* = 2\,\mathrm{Re}\{\dot{a}_f a_f^*\}$. Combining the results above and assuming $\hat{H}_1^\dagger = \hat{H}_1$ we find

$$\frac{dP_{fi}}{dt} = \frac{1}{2\hbar}\mathrm{Re}\left\{\frac{i(1+e^{2i\omega t})}{E_{fi} - \hbar\omega + i\hbar\Gamma} + \frac{i(1+e^{-2i\omega t})}{E_{fi} + \hbar\omega + i\hbar\Gamma}\right\}\left|\langle\varphi_f|\hat{H}_1|\varphi_i\rangle\right|^2.$$

We now assume a time-independent $(\omega = 0)$ perturbation and take the limit $\Gamma \to 0$ to find

$$\frac{dP_{fi}}{dt} = \frac{2\pi}{\hbar}\left|\langle\varphi_f|\hat{H}_1|\varphi_i\rangle\right|^2 \delta(E_{fi}).$$

This is the standard form of Fermi's golden rule. We wish to apply this result to scattering of electrons from a static potential $\hat{H}_1 = V(\vec{r})$. If both initial and final states are assumed to be free-electron states $\varphi_n(\vec{r}) = \Omega^{-1/2} e^{i\vec{k}_n\cdot\vec{r}}$ we find $\langle\varphi_f|\hat{H}_1|\varphi_i\rangle = \Omega^{-1}\tilde{V}(\vec{k}_f - \vec{k}_i)$ with $\tilde{V}(\vec{k})$ the Fourier transform of the potential. We now

consider a flux I of electrons moving with initial speed v_i, so that $I = nv_i$, where n is the density that, considering just a single electron, equals $n = \Omega^{-1}$. The cross section σ for scattering out of the initial state is defined by stating that the probability rate dP_{fi}/dt summed over all final states equals $nv_i\sigma$ for each electron. Hence,

$$\sigma = \frac{\Omega}{v_i}\sum_f \frac{dP_{fi}}{dt} = \frac{2\pi}{\hbar v_i \Omega}\sum_f \left|\tilde{V}(\vec{k}_f - \vec{k}_i)\right|^2 \delta(E_{fi}).$$

The sum over final states is actually a sum over wave vectors and applying the standard technique

$$\sigma = \frac{1}{4\pi^2\hbar v_i}\int \left|\tilde{V}(\vec{k}_f - \vec{k}_i)\right|^2 \delta(E_{fi})\, d^3k_f. \qquad \text{(A8.1)}$$

For a start, we will consider central potentials of the form $V(r)$. Their Fourier transform only depends on the magnitude of $\vec{k}$ so we write $\tilde{V}(\vec{k}) = \tilde{V}(k)$ with

$$\tilde{V}(k) = \int V(r) e^{i\vec{k}\cdot\vec{r}} d^3r = \frac{4\pi}{k}\int_0^\infty V(r)\sin(kr) r dr.$$

Due to the energy conserving delta-function in Eq.(A8.1), the initial and final wave vectors have the same magnitude k_i and therefore $|\vec{k}_f - \vec{k}_i| = 2k_i \sin\frac{\theta}{2}$, where θ is the *scattering angle*, i.e. the angle between initial and final directions. Further simplification follows from the facts that $\delta(E_{fi}) = m_e/(\hbar^2 k_i)\delta(k_f - k_i)$ and $v_i = \hbar k_i / m_e$. Finally, $d^3k_f = k_f^2 dk_f d\Omega_f$ with $d\Omega_f$ an infinitesimal solid angle element around the scattering direction. Defining the differential cross section as $d\sigma/d\Omega_f$ we then find

$$\frac{d\sigma}{d\Omega_f} = \frac{m_e^2}{4\pi^2\hbar^4}\left|\tilde{V}(2k_i \sin\tfrac{\theta}{2})\right|^2. \qquad \text{(A8.2)}$$

For the case of a screened Coulomb potential of the Yukawa form $V(r) = V_0 e^{-\mu r}/r$ one then finds $\tilde{V}(k) = 4\pi V_0/(k^2 + \mu^2)$. Hence, in the un-screened case taking $V_0 = Ze^2/(4\pi\varepsilon_0)$

$$\frac{d\sigma}{d\Omega_f} = \frac{m_e^2}{4\pi^2\hbar^4}\left\{\frac{Ze^2}{\varepsilon_0(2k_i \sin\frac{\theta}{2})^2}\right\}^2 = \left\{\frac{Ze^2}{16\pi\varepsilon_0 E_i \sin^2\frac{\theta}{2}}\right\}^2. \qquad \text{(A8.3)}$$

Appendix 9. Response Theory and Green's Functions

In this appendix, we turn to a seemingly rather technical subject: Green's functions. However, Green's functions will provide an important link between some of the separate results obtained so far. In addition, using Green's functions can in certain cases lead to great simplicity in the evaluation of responses. We consider a system with an unperturbed Hamiltonian $\hat{H}$ with eigenstates $|n\rangle$ defined by $\hat{H}|n\rangle = E_n|n\rangle$. The corresponding Green's functions is

$$\hat{G}(z) = (z - \hat{H})^{-1} \tag{A9.1}$$

Since $\hat{H}^p|n\rangle = (E_n)^p|n\rangle$ the Green's function can be written in the equivalent manner

$$\hat{G}(z) = \sum_n \frac{|n\rangle\langle n|}{z - E_n}. \tag{A9.2}$$

In fact, the Green's function itself is intimately related to an important quantity, viz. the density of states $D(E)$. To see this, we note that

$$\mathrm{Im}\{\mathrm{Tr}\hat{G}(z - i\eta)\} = \mathrm{Im}\sum_n \frac{1}{z - i\eta - E_n} = \sum_n \frac{\eta}{(z - E_n)^2 + \eta^2}.$$

Thus,

$$\lim_{\eta \to 0} \mathrm{Im}\{\mathrm{Tr}\hat{G}(z - i\eta)\} = \pi \sum_n \delta(z - E_n).$$

From the definition of $D(E)$ (including spin summation) we then see that

$$D(E) = \frac{2}{\pi\Omega} \lim_{\eta \to 0} \mathrm{Im}\{\mathrm{Tr}\hat{G}(E - i\eta)\}. \tag{A9.3}$$

That is, the density of states is essentially the imaginary part of the Green's function. The general principle can be illustrated by considering gapped graphene with the Hamiltonian for the wave vector $\vec{q}$

$$\overleftrightarrow{H}_{\vec{q}} = \begin{pmatrix} \alpha & \hbar v_F (q_x - i q_y) \\ \hbar v_F (q_x + i q_y) & -\alpha \end{pmatrix}.$$

It is readily shown that $\overleftrightarrow{G}_{\vec{q}}(z) = (z\overleftrightarrow{I} - \overleftrightarrow{H}_{\vec{q}})^{-1}$ is

$$\overleftrightarrow{G}_{\vec{q}}(z)=\frac{1}{z^2-\alpha^2-\hbar^2v_F^2q^2}\begin{pmatrix} z+\alpha & \hbar v_F(q_x-iq_y) \\ \hbar v_F(q_x+iq_y) & z-\alpha \end{pmatrix}.$$

In this case, the total Hamiltonian $\overleftrightarrow{H}=\text{diag}(\overleftrightarrow{H}_{\vec{q}_1},\overleftrightarrow{H}_{\vec{q}_2},...,\overleftrightarrow{H}_{\vec{q}_N})$ is diagonal in the wave vector index $\vec{q}$ and so taking the trace means (i) summing over $\vec{q}$ and (ii) summing over the diagonal matrix elements in $\overleftrightarrow{G}_{\vec{q}}(z)$. Thus,

$$\frac{1}{A}\text{Tr}\hat{G}(z)=\frac{1}{4\pi^2}\int\frac{2z}{z^2-\alpha^2-\hbar^2v_F^2q^2}d^2q.$$

Note that for this 2D material we replace volume by area A. Introducing $\varepsilon=\sqrt{\alpha^2+\hbar^2v_F^2q^2}$ we see that

$$\frac{1}{A}\text{Tr}\hat{G}(z)=\frac{1}{4\pi^2}\int\frac{z}{\varepsilon}\left\{\frac{1}{z-\varepsilon}-\frac{1}{z+\varepsilon}\right\}d^2q$$

and so

$$\frac{1}{A}\text{Im}\{\text{Tr}\hat{G}(E-i\eta)\}=\frac{1}{4\pi^2}\int\left\{\frac{\eta}{(E-\varepsilon)^2+\eta^2}+\frac{\eta}{(E+\varepsilon)^2+\eta^2}\right\}d^2q$$

with the limiting behavior

$$\frac{1}{A}\lim_{\eta\to0}\text{Im}\{\text{Tr}\hat{G}(E-i\eta)\}=\frac{1}{4\pi}\int\{\delta(E-\varepsilon)+\delta(E+\varepsilon)\}d^2q.$$

The integral is easily performed using $d^2q=2\pi\varepsilon d\varepsilon/(\hbar^2v_F^2)$ and finally

$$D(E)=\frac{2|E|}{\pi\hbar^2v_F^2}\theta(|E|-\alpha) \tag{A9.4}$$

in agreement with Chapter 17.

A9.1 Response Functions

We will now proceed to the calculation of response functions. Recall that

$$X(\omega)=-\sum_{m,n}[f(E_n)-f(E_m)]\frac{\langle n|\hat{X}|m\rangle\langle m|\hat{H}_1|n\rangle}{E_m-E_n-\hbar\omega}.$$

Here, ω in general has an imaginary part. By simple relabeling, we can write this as

$$X(\omega)=\sum_{m,n} f(E_m)\left\{\frac{\langle n|\hat{X}|m\rangle\langle m|\hat{H}_1|n\rangle}{E_m-E_n-\hbar\omega}+\frac{\langle n|\hat{H}_1|m\rangle\langle m|\hat{X}|n\rangle}{E_m-E_n+\hbar\omega}\right\}.$$

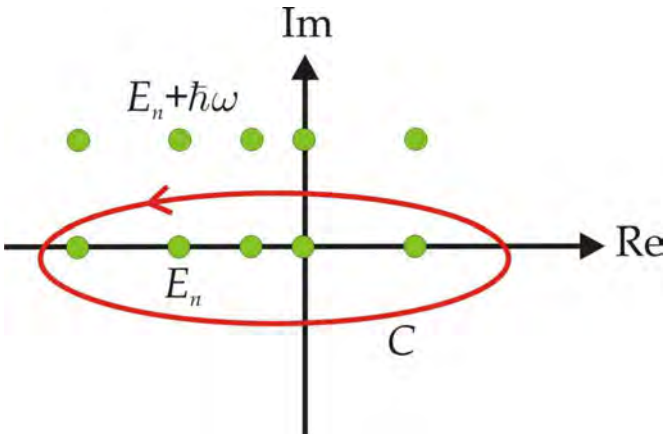


Figure A9.1. Integration contour for response functions.

We now consider a range of eigenvalues $E_1 ... E_N$ and introduce the integration contour defined in Fig. A9.1. We will require that all points E_n lie enclosed by the contour but that $E_n \pm \hbar\omega$ are not. This allows us to write

$$\frac{f(E_m)}{E_m-E_n\pm\hbar\omega}=\frac{1}{2\pi i}\oint\frac{f(z)}{(z-E_m)(z-E_n\pm\hbar\omega)}dz.$$

As a consequence

$$X(\omega)=\frac{1}{2\pi i}\oint f(z)\sum_{m,n}\left\{\frac{\langle n|\hat{X}|m\rangle\langle m|\hat{H}_1|n\rangle}{(z-E_m)(z-E_n-\hbar\omega)}+\frac{\langle n|\hat{H}_1|m\rangle\langle m|\hat{X}|n\rangle}{(z-E_m)(z-E_n+\hbar\omega)}\right\}dz.$$

We now use the definition Eq.(A9.2) to write

$$X(\omega)=\frac{1}{2\pi i}\oint f(z)\sum_{n}\left\{\frac{\langle n|\hat{X}\hat{G}(z)\hat{H}_1|n\rangle}{z-E_n-\hbar\omega}+\frac{\langle n|\hat{H}_1\hat{G}(z)\hat{X}|n\rangle}{z-E_n+\hbar\omega}\right\}dz.$$

Then, both terms can be expressed as traces

$$X(\omega)=\frac{1}{2\pi i}\oint f(z)\mathrm{Tr}\left\{\hat{X}\hat{G}(z)\hat{H}_1\hat{G}(z-\hbar\omega)+\hat{H}_1\hat{G}(z)\hat{X}\hat{G}(z+\hbar\omega)\right\}dz.$$

Finally, we use the invariance of the trace under cyclic permutations and shift the argument

$$X(\omega)=\frac{1}{2\pi i}\oint f(z)\Big\{\mathrm{Tr}\big[\hat{X}\hat{G}(z)\hat{H}_1\hat{G}(z-\hbar\omega)\big]+(z\to z+\hbar\omega)\Big\}dz\,.$$

As an important example, we will look at the electrical conductivity $\sigma(\omega)$, which describes the current density induced by an electric field. Here, the observable is $\hat{j}_x=-e\hat{p}_x/(m\Omega)$ and the perturbation is given by $\hat{H}_1=e\hat{p}_x\mathcal{E}/(im\omega)$. In total,

$$\sigma(\omega)=\frac{e^2}{\pi m^2\omega\Omega}\oint f(z)\Big\{\mathrm{Tr}\big[\hat{p}_x\hat{G}(z)\hat{p}_x\hat{G}(z-\hbar\omega)\big]+(z\to z+\hbar\omega)\Big\}dz\,,$$

where spin has already been counted. For graphene, the sheet conductivity (replacing volume by area) after conversion of the q – part of the trace to an integral and addition of valley degeneracy is given by

$$\sigma(\omega)=\frac{e^2}{2\pi^3 m^2\omega}\int\oint f(z)\Big\{\mathrm{Tr}\big[\hat{p}_x\hat{G}(z)\hat{p}_x\hat{G}(z-\hbar\omega)\big]+(z\to z+\hbar\omega)\Big\}dzd^2q\,.$$

As shown in Chapter 17, the matrix representation of $\hat{p}_x$ is $mv_F\begin{pmatrix}0&1\\1&0\end{pmatrix}$ and using standard contour integration it follows that

$$\oint f(z)\Big\{\mathrm{Tr}\big[\hat{p}_x\hat{G}(z)\hat{p}_x\hat{G}(z-\hbar\omega)\big]+(z\to z+\hbar\omega)\Big\}dz=8\pi i m^2v_F^2\frac{(\varepsilon^2-\hbar^2v_F^2q_x^2)\{f(\varepsilon)-f(-\varepsilon)\}}{\varepsilon(4\varepsilon^2-\hbar^2\omega^2)}$$

Converting again to an energy integral similarly to the density of states we then find

$$\begin{aligned}\sigma(\omega)&=\mathrm{Re}\frac{4ie^2}{\pi\hbar^2\omega}\int_\alpha^\infty\frac{(\varepsilon^2+\alpha^2)\{f(\varepsilon)-f(-\varepsilon)\}}{4\varepsilon^2-\hbar^2\omega^2}d\varepsilon\\&=\frac{e^2}{2\hbar^2\omega}\int_\alpha^\infty\frac{(\varepsilon^2+\alpha^2)\{f(-\varepsilon)-f(\varepsilon)\}}{\varepsilon}\delta(\varepsilon-\hbar\omega/2)d\varepsilon\\&=\frac{e^2(\hbar^2\omega^2+4\alpha^2)}{4\hbar^3\omega^2}\{f(-\hbar\omega/2)-f(\hbar\omega/2)\}\theta(\hbar\omega-2\alpha).\end{aligned}$$

in agreement with Eq.(17.10).

For calculations of the excitonic response, the Greens' function approach is even more direct, at least in principle. To appreciate this fact, we consider the imaginary part of the many-body susceptibility Eq.(18.1)

$$\chi_I(\omega)=\frac{\pi e^2}{\varepsilon_0 m^2\omega^2\Omega}\sum_{exc}\left|\left\langle 0\left|\hat{P}_z\right|exc\right\rangle\right|^2\delta(E_{exc}-\hbar\omega).$$

Here, $|0\rangle$ is the ground state and $|exc\rangle$ is an eigenstate of the exciton Hamiltonian $\hat{H}|exc\rangle = E_{exc}|exc\rangle$ in which $\hat{H}$ is written in a basis of single excitations, i.e. $H_{ij,kl} = \langle v_i \to c_j | \hat{H} | v_k \to c_l \rangle$. We now introduce the many-body Green's function $\hat{G}(z) = (z - \hat{H})^{-1}$ so that

$$\chi_I(\omega) = \frac{e^2}{\varepsilon_0 m^2 \omega^2 \Omega} \mathrm{Im} \langle 0 | \hat{P}_z \hat{G}(\hbar\omega - i\hbar\eta) \hat{P}_z | 0 \rangle$$

with the understanding that $\hbar\eta$ should be taken to zero after evaluation. In fact, this result can be formulated even more briefly by introducing

$$|P\rangle \equiv \hat{P}_z |0\rangle = \sum_{ij} \langle v_i \to c_j | \hat{P}_z | 0 \rangle | v_i \to c_j \rangle = \sqrt{2} \sum_{ij} \langle v_i | \hat{p}_z | c_j \rangle | v_i \to c_j \rangle .$$

Hence,

$$\chi_I(\omega) = \frac{e^2}{\varepsilon_0 m^2 \omega^2 \Omega} \mathrm{Im} \langle P | \hat{G}(\hbar\omega - i\hbar\eta) | P \rangle .$$

Diagonal matrix elements of the Green's function such as $\langle P | \hat{G}(\hbar\omega - i\hbar\eta) | P \rangle$ can be computed very efficiently, as we now demonstrate.

Any matrix element $\langle u | \hat{G}(z) | u \rangle$ can be generated via the Lanczos-Haydock routine [1]. Here, we tri-diagonalize the Hamiltonian approximately by setting $|q_0\rangle := 0, |q_1\rangle := |u\rangle$ and generating

$$|Q_{j+1}\rangle = (\hat{H} - a_j)|q_j\rangle - b_j |q_{j-1}\rangle, \quad a_j = \langle q_j | \hat{H} | q_j \rangle, \quad b_j = \langle Q_j | Q_j \rangle^{1/2}, \quad |q_{j+1}\rangle = |Q_{j+1}\rangle / b_{j+1}$$

.

In this rotated basis, the tri-diagonal Hamiltonian H_{TD} is

$$H_{TD} = \begin{pmatrix} a_1 & b_2 & 0 & \cdots & \\ b_2 & a_2 & b_3 & \cdots & \\ 0 & b_3 & a_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & b_N \\ & & 0 & b_N & a_N \end{pmatrix}.$$

In this manner, we find recursively

$$\left\langle u\left|\hat{G}(z)\right|u\right\rangle = \cfrac{1}{z-a_1-\cfrac{b_2^2}{z-a_2-\cfrac{b_3^2}{\cdots}}}. \tag{A9.5}$$

Here, the sign of the square root has been chosen in accordance with Chapter 7. This method prides a very simple way of computing the excitonic response without diagonalizing large matrices.

References

[1] R. Haydock, Solid State Phys. 35, 215 (1980).

Appendix 10. Stationary Perturbation Theory

In this appendix, we will study the effect of static perturbations on the eigenstates. In several chapters, we have investigated time-dependent perturbations by solving the time-dependent Schrödinger equation. In the presence of a static perturbation, however, one still has a set of stationary eigenstates and, hence, we turn to the time-independent Schrödinger equation. Typically, we encounter scenarios such as atoms exposed to static electric or magnetic fields. In this case, the zero'th order Hamiltonian can usually be regarded as completely dominating over the perturbations. It follows that the true eigenstates are close to the unperturbed ones and so are the energies. An exception is the case of degenerate unperturbed states that may be split by the perturbation. We shall study both standard and degenerate perturbation theory below.

We assume, as usual, that the total Hamiltonian can be written as $\hat{H} = \hat{H}_0 + \hat{H}_1$, where $\hat{H}_1$ is the perturbation. Hence, were trying to solve

$$\left(\hat{H}_0 + \hat{H}_1\right)\psi_n = E_n\psi_n .$$

Next, similarly to the time-dependent case, we indicate by a superscript p the order of the perturbation. Thus, $\psi_n = \psi_n^{(0)} + \psi_n^{(1)} + \psi_n^{(2)} + \ldots$ and $E_n = E_n^{(0)} + E_n^{(1)} + E_n^{(2)} + \ldots$, where the p'th term is proportional to the p'th power of the perturbation. When inserted in the eigenvalue problem, we can then collect terms of identical order and find

$$\begin{aligned} \hat{H}_0\psi_n^{(0)} &= E_n^{(0)}\psi_n^{(0)} \\ \hat{H}_0\psi_n^{(1)} + \hat{H}_1\psi_n^{(0)} &= E_n^{(0)}\psi_n^{(1)} + E_n^{(1)}\psi_n^{(0)} \\ \hat{H}_0\psi_n^{(2)} + \hat{H}_1\psi_n^{(1)} &= E_n^{(0)}\psi_n^{(2)} + E_n^{(1)}\psi_n^{(1)} + E_n^{(2)}\psi_n^{(0)}. \\ &\vdots \end{aligned} \tag{A10.1}$$

By solving these equations order by order, we can systematically include the effects of the perturbation on wave functions and energies. We need to supplement with the normalization requirement, which reads

$$\left\langle\psi_n\middle|\psi_n\right\rangle = 1 + \left\langle\psi_n^{(0)}\middle|\psi_n^{(1)}\right\rangle + \left\langle\psi_n^{(1)}\middle|\psi_n^{(0)}\right\rangle + \left\langle\psi_n^{(0)}\middle|\psi_n^{(2)}\right\rangle + \left\langle\psi_n^{(1)}\middle|\psi_n^{(1)}\right\rangle + \left\langle\psi_n^{(2)}\middle|\psi_n^{(0)}\right\rangle + \ldots = 1$$

Here, we assumed orthonormalized unperturbed states $\left\langle\psi_n^{(0)}\middle|\psi_m^{(0)}\right\rangle = \delta_{nm}$. It follows that the normalization corrections at each order must vanish. We now exploit the fact that the unperturbed states form a complete set, so that we can write

$\psi_n^{(p)} = \sum_m c_{nm}^{(p)} \psi_m^{(0)}$. If we insert this expansion into the first-order equation in Eq.(A10.1), we immediately find

$$\sum_m c_{nm}^{(1)} E_m^{(0)} \psi_m^{(0)} + \hat{H}_1 \psi_n^{(0)} = E_n^{(0)} \sum_m c_{nm}^{(1)} \psi_m^{(0)} + E_n^{(1)} \psi_n^{(0)} .$$

Now, if we multiply from the left with $\left\langle \psi_n^{(0)} \right|$ and use orthonormality, we find $E_n^{(1)} = \left\langle \psi_n^{(0)} \middle| \hat{H}_1 \middle| \psi_n^{(0)} \right\rangle$. Alternatively, we multiply with $\left\langle \psi_m^{(0)} \right|$ with $m \neq n$. In this case, one readily sees that *provided* $E_n^{(0)} \neq E_m^{(0)}$ one has $c_{nm}^{(1)} = \left\langle \psi_m^{(0)} \middle| \hat{H}_1 \middle| \psi_n^{(0)} \right\rangle / (E_n^{(0)} - E_m^{(0)})$. This does not determine $c_{nn}^{(1)}$ but the normalization requirement evaluated up to first order quickly leads to $c_{nn}^{(1)} = 0$ since $\left\langle \psi_n^{(0)} \middle| \psi_n^{(1)} \right\rangle = 0$. Hence, the first-order perturbation becomes

$$\psi_n^{(1)} = \sum_{m \neq n} \frac{\left\langle \psi_m^{(0)} \middle| \hat{H}_1 \middle| \psi_n^{(0)} \right\rangle}{E_n^{(0)} - E_m^{(0)}} \psi_m^{(0)}, \quad E_n^{(1)} = \left\langle \psi_n^{(0)} \middle| \hat{H}_1 \middle| \psi_n^{(0)} \right\rangle . \tag{A10.2}$$

In turn, multiplying the second order equation in Eq.(A10.1) by $\left\langle \psi_n^{(0)} \right|$ and using $\left\langle \psi_n^{(0)} \middle| \psi_n^{(1)} \right\rangle = 0$, we find $E_n^{(2)} = \left\langle \psi_n^{(0)} \middle| \hat{H}_1 \middle| \psi_n^{(1)} \right\rangle$ or

$$E_n^{(2)} = \sum_{m \neq n} \frac{\left| \left\langle \psi_m^{(0)} \middle| \hat{H}_1 \middle| \psi_n^{(0)} \right\rangle \right|^2}{E_n^{(0)} - E_m^{(0)}} . \tag{A10.3}$$

One may note that, for the ground state, we have $E_n^{(2)} \leq 0$ irrespective of the form of the perturbation.

In many cases, such as the Stark effect in hydrogen studied below, all odd corrections to the energy vanish. In the Stark case, this is because the direction of the perturbing electric field plays no role for the energy. We will find the fourth order correction for this situation using the notation $h_n^{(p,q)} = \left\langle \psi_n^{(p)} \middle| \hat{H}_0 \middle| \psi_n^{(q)} \right\rangle$, $H_n^{(p,q)} = \left\langle \psi_n^{(p)} \middle| \hat{H}_1 \middle| \psi_n^{(q)} \right\rangle$ and $S_n^{(p,q)} = \left\langle \psi_n^{(p)} \middle| \psi_n^{(q)} \right\rangle$. Importantly, interchanging p and q just amounts to a complex conjugation. Also, $h_n^{(0,q)} = E_n^{(0)} S_n^{(0,q)}$ and $S_n^{(0,0)} = 1$. We then find from Eq.(A10.1)

$$h_n^{(p,q)} + H_n^{(p,q-1)} = \sum_{r=0}^{q} E_n^{(r)} S_n^{(p,q-r)} .$$

The relations needed follow by setting $(p,q) = \{(0,4),(1,3),(3,1)\}$, i.e.

$$H_n^{(0,3)} = E_n^{(2)} S_n^{(0,2)} + E_n^{(4)},$$
$$h_n^{(1,3)} + H_n^{(1,2)} = E_n^{(0)} S_n^{(1,3)} + E_n^{(2)} S_n^{(1,1)},$$
$$h_n^{(3,1)} + H_n^{(3,0)} = E_n^{(0)} S_n^{(3,1)}.$$

Conjugating the last relation and combining all of these, we eventually find $E_n^{(4)} = H_n^{(1,2)} - E_n^{(2)}(S_n^{(0,2)} + S_n^{(1,1)})$. It is seen that, in fact, only second-order information is required to compute the fourth-order energy correction. A similar calculation shows that $E_n^{(6)} = H_n^{(2,3)} - E_n^{(2)}(S_n^{(1,3)} + S_n^{(2,2)}) - E_n^{(4)}(S_n^{(0,2)} + S_n^{(1,1)} + S_n^{(2,0)})$.

A10.1 Stark Effect

A classic application of perturbation theory is the Stark effect, i.e. the change in energy due to a constant electric field $\vec{\mathcal{E}} = \mathcal{E}\vec{e}_z$. In particular, the static polarizability associated with the second-order energy can be calculated. Hence, in this case the perturbation is $\hat{H}_1 = e\mathcal{E}z$. We will assume that the system is symmetric in the z-direction so that the unperturbed states have definite z-parity. Then, it is readily shown that $\left\langle \psi_n^{(0)} \middle| \hat{H}_1 \middle| \psi_n^{(0)} \right\rangle = 0$ so that the first-order energy vanishes. This is only valid for non-degenerate states, however, as we will demonstrate below. For these states, the first correction is of second order and, in terms of the oscillator strength $g_{mn} = \frac{2m_e}{\hbar^2}\left|\left\langle \psi_m^{(0)} \middle| z \middle| \psi_n^{(0)} \right\rangle\right|^2 (E_m^{(0)} - E_n^{(0)})$, one may write

$$E_n^{(2)} = -\frac{1}{2}\alpha_n \mathcal{E}^2, \quad \alpha_n = \frac{\hbar^2 e^2}{m_e} \sum_{m \neq n} \frac{g_{mn}}{(E_m^{(0)} - E_n^{(0)})^2}, \tag{A10.4}$$

where α_n is the polarizability of the n'th state. This result agrees with the static limit of the frequency-dependent result in Chapter 2. For the hydrogen ground state, using atomic units and oscillator strengths from Chapter 30, one finds contributions from transitions to bound and free states, i.e. $\alpha_n = \alpha_n^{\text{Bound}} + \alpha_n^{\text{Free}}$ with

$$\alpha_{1s}^{\text{Bound}} = 4\sum_{n=2}^{\infty} \frac{g_{np1s}}{(1-n^{-2})^2}, \quad g_{np1s} = \frac{2^8 n^5 (n-1)^{2n-4}}{3(n+1)^{2n+4}},$$

and

$$\alpha_{1s}^{\text{Free}} = 4\int_0^{\infty} g'(k) \frac{k}{(1+k^2)^2} dk, \quad g'(k) = \frac{2^8}{3(1+k^2)^4} \exp\left\{-4\frac{\tan^{-1} k}{k}\right\} / \left(1 - \exp\{-2\pi/k\}\right).$$

When evaluated, $\alpha_{1s}^{\text{Bound}} \approx 3.663$ and $\alpha_{1s}^{\text{Free}} \approx 0.837$ while $\alpha_{1s}^{\text{Bound}} + \alpha_{1s}^{\text{Free}} = 9/2$ exactly.

The perturbation of the hydrogen ground state is an excellent example of non-degenerate perturbation theory. If, however, we consider the $\{2s, 2p_x, 2p_y, 2p_z\}$ manifold, we need to worry about degeneracies. One can show that $\left\langle \psi_{2s}^{(0)} \middle| \hat{H}_1 \middle| \psi_{2p_z}^{(0)} \right\rangle = -3\mathcal{E}$ while all other matrix elements within this block vanish. We now attempt to find perturbed states in the $n=2$ block by writing a general state as $\psi = c_s \psi_{2s}^{(0)} + c_p \psi_{2p_z}^{(0)}$. With the full Hamiltonian, this leads to the matrix eigenvalue problem

$$\begin{pmatrix} E_{2s}^{(0)} & -3\mathcal{E} \\ -3\mathcal{E} & E_{2p_z}^{(0)} \end{pmatrix} \cdot \begin{pmatrix} c_s \\ c_p \end{pmatrix} = E \begin{pmatrix} c_s \\ c_p \end{pmatrix}.$$

Generally, the eigenvalues of this matrix would be $E_{\pm} = \frac{1}{2}(E_{2s}^{(0)} + E_{2p_z}^{(0)}) \pm [\frac{1}{4}(E_{2s}^{(0)} - E_{2p_z}^{(0)})^2 + 9\mathcal{E}^2]^{1/2}$. However, precisely because $E_{2s}^{(0)} = E_{2p_z}^{(0)} \equiv E_2^{(0)}$, diagonalization of the matrix produces the eigenstates

$$\psi_{\pm} = \frac{1}{\sqrt{2}} \left(\psi_{2s}^{(0)} \pm \psi_{2p_z}^{(0)} \right), \quad E_{\pm} = E_2^{(0)} \mp 3\mathcal{E}\,.$$

Hence, in this case, the correction is actually linear in the electric field.

A10.2 Zeeman Effect

A second classic application is the Zeeman effect describing the energy change in response to a constant magnetic field $\vec{\mathcal{B}} = \mathcal{B}\vec{e}_z$. In atomic units taking the magnetic field in units of $2\hbar / ea_0^2$, the perturbation with $r_\perp = r\sin\theta$ reads $\hat{H}_1 = \mathcal{B}\hat{l}_z + \frac{1}{2}\mathcal{B}^2 r_\perp^2$. We will apply perturbation theory to compute the magnetic corrections. The first (paramagnetic) term is very simple for a spherical system with quantum number *nlm* and simply leads to $E_{nlm}^{(1,para)} = m\mathcal{B}$. However, the degenerate version is generally required for the diamagnetic term. It can be shown that, within the block sharing a given principle quantum number *n*, the only non-vanishing couplings are [1]

$$\left\langle nlm \middle| r_\perp^2 \middle| nlm \right\rangle = \frac{n^2[5n^2 - 3l(l+1) + 1][l(l+1) + m^2 - 1]}{(2l-1)(2l+3)},$$

and

$$\left\langle nlm \middle| r_\perp^2 \middle| nl-2m \right\rangle = -\frac{5n^2}{2}\sqrt{\frac{[n^2-(l-1)^2][(l-1)^2-m^2](n^2-l^2)(l^2-m^2)}{(2l-3)(2l-1)^2(2l+1)}}\,.$$

Hence, at first order, a particular state is only coupled to states shifted by ± 2 in angular momentum. This means, in fact, that the five lowest states are uncoupled. We then find the energy including first-order Zeeman shifts for the hydrogen states

$$E_{nlm}^{(0)} + E_{nlm}^{(1)} = \begin{cases} -\frac{1}{2} + \mathcal{B}^2 & 1s \\ -\frac{1}{8} + 14\mathcal{B}^2 & 2s \\ -\frac{1}{8} + 6\mathcal{B}^2 & 2p_0 \\ -\frac{1}{8} - \mathcal{B} + 12\mathcal{B}^2 & 2p_{-1} \\ -\frac{1}{8} + \mathcal{B} + 12\mathcal{B}^2 & 2p_{+1}. \end{cases}$$

For the $n = 3$ block with $m = 0$, we have a three-fold degeneracy and a matrix given by

$$H_1 = \begin{pmatrix} -\frac{1}{18} + 69\mathcal{B}^2 & 0 & -15\sqrt{2}\mathcal{B}^2 \\ 0 & -\frac{1}{18} + 36\mathcal{B}^2 & 0 \\ -15\sqrt{2}\mathcal{B}^2 & 0 & -\frac{1}{18} + 30\mathcal{B}^2 \end{pmatrix}.$$

The coupling between s and d states then leads to the eigenstates

$$E_{nlm}^{(0)} + E_{nlm}^{(1)} = -\frac{1}{18} + \begin{cases} \frac{9}{2}(11 - \sqrt{41})\mathcal{B}^2 & -(13 - 3\sqrt{41})(3s) + 10\sqrt{2}(3d_0) \\ 36\mathcal{B}^2 & , \quad 3p_0 \\ \frac{9}{2}(11 + \sqrt{41})\mathcal{B}^2 & -(13 + 3\sqrt{41})(3s) + 10\sqrt{2}(3d_0). \end{cases}$$

References

[1] E.A. Solov'ev, Sov. Phys. JETP 55, 1017 (1982).

Appendix 11. Dalgarno-Lewis Perturbation Theory

In this appendix, we introduce the Dalgarno-Lewis perturbation technique [1] for the calculation of static polarizabilities of centrosymmetric systems. We start by analyzing one-dimensional systems and then extend to higher dimensions. The starting point is the perturbed problem in a static field $\mathcal{E}$ is then

$$\{H_0 + \mathcal{E}x\}\varphi(x) = E\varphi(x), \quad H_0 = -\frac{1}{2}\frac{d^2}{dx^2} + V(x).$$

The normalized solution to the H_0 problem is $H_0\varphi_0(x) = E_0\varphi_0(x)$. We note that since the unperturbed system is inversion symmetric, the energy is of the form $E = E_0 + E_2 + ...$ even if the wave function contains all orders $\varphi(x) = \varphi_0(x) + \varphi_1(x) + \varphi_2(x) + ...$ Using the fact that there is no first order energy correction, the first order wave function satisfies the inhomogeneous Dalgarno-Lewis equation

$$\{H_0 - E_0\}\varphi_1(x) = -\mathcal{E}x\varphi_0(x).$$

A clever trick for solving this problem is to write $\varphi_1(x) = f(x)\varphi_0(x)$, which means that

$$-\frac{1}{2}f''(x) - f'(x)\frac{\varphi_0'(x)}{\varphi_0(x)} = -\mathcal{E}x. \tag{A11.1}$$

Even if this is a second order differential equation, the fact that no un-differentiated term appears allows us to find an analytical solution

$$f'(x) = \frac{2\mathcal{E}}{\varphi_0^2(x)}\int_a^x \varphi_0^2(x')x'dx'. \tag{A11.2}$$

The undetermined integration limit a plays the role of integration constant. Integrating once more, we find

$$f(x) = -2\mathcal{E}\int_b^x \frac{1}{\varphi_0^2(x')}\int_{x'}^a \varphi_0^2(x'')x''dx''dx'. \tag{A11.3}$$

The fact that f must be odd means that $b = 0$. The constant a must the determined so that appropriate boundary conditions are obeyed. The second order equation reads

$$H_0\varphi_2(x) + \mathcal{E}x\varphi_1(x) = E_0\varphi_2(x) + E_2\varphi_0(x).$$

So, multiplying by $\varphi_0(x)$ and integrating leads to the second order energy

$$E_2 = \mathcal{E}\langle\varphi_0|x|\varphi_1\rangle = \mathcal{E}\int_{-\infty}^{\infty}\varphi_0^2(x)f(x)xdx\,.$$

Using partial integration and Eq.(A11.2), this is also

$$E_2 = -\mathcal{E}\int_{-\infty}^{\infty}\left[\int_{-\infty}^{x}\varphi_0^2(x')x'dx'\right]f'(x)dx = -\frac{1}{2}\int_{-\infty}^{\infty}\varphi_0^2(x)[f'(x)]^2dx\,. \quad \text{(A11.4)}$$

In turn, the polarizability α follows by writing $E_2 = -\frac{1}{2}\alpha\mathcal{E}^2$. The Dalgarno-Lewis method is connected to logarithmic perturbation theory [2]. There, one introduces $\chi(x) = \varphi'(x)/\varphi(x)$. It's readily shown that $\varphi''/\varphi = \chi' + \chi^2$. Hence,

$$\chi_0(x) = \frac{\varphi_0'(x)}{\varphi_0(x)}, \quad \chi_1(x) = \frac{\varphi_1'(x)}{\varphi_0(x)} - \frac{\varphi_0'(x)\varphi_1(x)}{\varphi_0^2(x)} = \frac{d}{dx}\left[\frac{\varphi_1(x)}{\varphi_0(x)}\right] = f'(x)\,,$$

and

$$E_2 = -\frac{1}{2}\int_{-\infty}^{\infty}\varphi_0^2(x)[\chi_1(x)]^2dx\,.$$

The function $\chi(x) = \varphi'(x)/\varphi(x)$ obeys the equation

$$-\frac{1}{2}\{\chi' + \chi^2\} + V(x) + \mathcal{E}x = E\,.$$

The first order version is $\chi_1' + 2\chi_0\chi_1 - 2\mathcal{E}x = 0$, which is identical to Eq.(A11.1).

A11.1 One-dimensional Examples

We now illustrate the general approach by applying it to a number of representative cases.

Harmonic Oscillator

In this case, $V(x) = \frac{1}{2}\omega^2x^2$ and $\varphi_0(x) = \left(\frac{\omega}{\pi}\right)^{1/4}\exp(-\frac{\omega}{2}x^2)$. Then, $f(x) = -\mathcal{E}x/\omega$ is found. In turn, the second order energy becomes $E_2 = -\frac{1}{2}\mathcal{E}^2/\omega^2$ and the polarizability is $\alpha = 1/\omega^2$.

Dirac-δ potential

Here, $V(x)=-Z\delta(x)$ and $\varphi_0(x)=\sqrt{Z}\exp\{-Z|x|\}$. Then, $f(x)=-\mathcal{E}x\left(1+Z|x|\right)/2Z^2$. From this result, we compute the second order energy correction $E_2=-5\mathcal{E}^2/(8Z^4)$ leading to a polarizability of $\alpha=5/4Z^4$.

Square well

We take the potential to vanish in the range $|x|<L/2$ and be infinite otherwise. Then, the ground state is $\varphi_0(x)=\sqrt{2/L}\cos(\frac{\pi x}{L})$. Hence, $f(x)=-\mathcal{E}L\left\{\pi(L^2-4x^2)\tan(\frac{\pi x}{L})-4Lx\right\}/(8\pi^2)$ and $E_2=-\mathcal{E}^2L^4(15-\pi^2)/(24\pi^4)$ so that $\alpha=L^4(15-\pi^2)/(12\pi^4)$.

Softened Coulomb potential

Here, $V(x)=-Z/(|x|+1/Z)$ and $\varphi_0(x)=\sqrt{2Z/5}(1+Z|x|)\exp\{-Z|x|\}$. Then, $f(x)=-\mathcal{E}\left\{xZ\left(6+2Z|x|+1/(1+Z|x|)\right)+2\ln(1+Z|x|)\right\}/(4Z^3)$. From this result, we compute the second order energy correction $E_2=-\mathcal{E}^2\left\{82-e^2\mathrm{Ei}(-2)\right\}/(20Z^4)$ leading to a polarizability of $\alpha=\left\{82-e^2\mathrm{Ei}(-2)\right\}/(10Z^4)$.

Linear well

We take the potential to be $V(x)=F|x|$. Then, the ground state is $\varphi_0(x)=F^{1/6}\mathrm{Ai}\{(2F)^{1/3}x-\beta\}/\left[2^{1/3}\beta^{1/2}\mathrm{Ai}\{-\beta\}\right]$, where $\beta\approx 1.01879...$ is the first zero of $\mathrm{Ai}'(x)$. Numerical integration yields $E_2=-\mathcal{E}^2 0.449231/F^{4/3}$ so that $\alpha=0.898462/F^{4/3}$.

A11.2 Two-dimensional Examples

In two-dimensional systems with cylindrical symmetry, the Schrödinger equation reads (pulling a factor $r^{-1/2}$ out of all wave functions to simplify the radial Laplacian)

$$\left\{H_0+\mathcal{E}r\cos\theta\right\}\varphi(r,\theta)=E\varphi(r,\theta),\quad H_0=-\frac{1}{2}\frac{d^2}{dr^2}-\frac{1}{8r^2}-\frac{1}{2r^2}\frac{d^2}{d\theta^2}+V(r). \qquad \text{(A11.5)}$$

We now expand $\varphi(r,\theta)=\varphi_0(r)+\varphi_1(r,\theta)...$ and write $\varphi_1(r,\theta)=\cos\theta f(r)\varphi_0(r)$. In this manner, the first order equation becomes

$$-\frac{1}{2}f''(r)-f'(r)\frac{\varphi_0'(r)}{\varphi_0(r)}+\frac{1}{2r^2}f(r)=-\mathcal{E}r\,.$$

Moreover, with $\varphi_0(r)=R_0(r)/\sqrt{2\pi}$ the second order energy becomes

$$E_2=\frac{\mathcal{E}}{2}\int_0^\infty R_0^2(r)f(r)rdr\,, \tag{A11.6}$$

where the factor $1/2$ comes from the angular integration.

Cylindrical quantum well

Here, the potential vanishes $0<r<a$ and is infinite otherwise. Hence, $R_0(r)=\sqrt{2r}J_0(\frac{\beta r}{a})/[aJ_1(\beta)]$, where β is the first zero of J_0. Hence, $f(r)=\mathcal{E}a(r^2-a^2)J_1(\frac{\beta r}{a})/[2\beta J_0(\frac{\beta r}{a})]$. Integrating leads to the energy $E_2=-\mathcal{E}^2a^4(4+\beta^2)/(12\beta^4)$ and polarizability $\alpha=a^4(4+\beta^2)/(6\beta^4)$.

2D hydrogen atom

With the potential $V(r)=-Z/r$ and ground state $R_0(r)=4Zr^{1/2}\exp(-2Zr)$ the general procedure yields $f(r)=-\mathcal{E}r(3+4rZ)/(16Z^2)$. This produces an energy $E_2=-\mathcal{E}^2 21/(256Z^4)$ and a polarizability $\alpha=21/(128Z^4)$.

A11.3 Three-dimensional Examples

In three-dimensional systems with spherical symmetry, the Schrödinger equation reads (pulling a factor r^{-1} out of all wave functions to simplify the radial Laplacian)

$$\left\{\hat{H}_0+\mathcal{E}r\cos\theta\right\}\psi(r,\theta)=E\psi(r,\theta),\quad \hat{H}_0=-\frac{1}{2}\frac{d^2}{dr^2}+\frac{\hat{l}^2}{2r^2}+V(r)\,. \tag{A11.7}$$

We now expand $\varphi(r,\theta)=\varphi_0(r)+\varphi_1(r,\theta)...$ and write $\varphi_1(r,\theta)=\cos\theta f(r)\varphi_0(r)$. In this manner, the first order equation becomes

$$-\frac{1}{2}f''(r)-f'(r)\frac{\varphi_0'(r)}{\varphi_0(r)}+\frac{1}{r^2}f(r)=-\mathcal{E}r\,.$$

Moreover, with $\varphi_0(r)=R_0(r)/\sqrt{4\pi}$ the second order energy becomes

$$E_2 = \frac{\mathcal{E}}{3}\int_0^\infty R_0^2(r) f(r) r dr\,, \tag{A11.8}$$

where the factor $1/3$ comes from the angular integration.

Spherical quantum well

Here, the potential vanishes $0 < r < a$ and is infinite otherwise. Hence, $R_0(r) = \sqrt{2/a}\sin(\frac{\pi r}{a})$ and $f(r) = -\mathcal{E}a\left\{a^3 + \pi r(r^2 - a^2)\cot(\frac{\pi r}{a})\right\}/(2\pi^2 r)$. Integrating leads to the energy $E_2 = -\mathcal{E}^2 a^4(3+4\pi^2)/(24\pi^4)$ and polarizability $\alpha = a^4(3+4\pi^2)/(12\pi^4)$.

3D hydrogen atom

With the potential $V(r) = -Z/r$ and ground state $R_0(r) = 2\sqrt{Z^3}\, r\exp(-Zr)$ the general procedure yields $f(r) = -\mathcal{E}r(2+rZ)/(2Z^2)$. This produces an energy $E_2 = -\mathcal{E}^2 9/(4Z^4)$ and a polarizability $\alpha = 9/(2Z^4)$

References

[1] A. Dalgarno and J.T. Lewis, Proc. R. Soc. Lond. A 233, 70 (1955).
[2] Y. Aharonov and C.K. Au, Phys. Rev. Lett. 42, 1582 (1979).