

**ELECTRON TRANSPORT
THROUGH
ONE-DIMENSIONAL QUANTUM SYSTEMS**

**A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
ATILIM UNIVERSITY
BY
LATIFA ALGADI**

**IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
PHYSICS
AT
THE DEPARTMENT OF APPLIED PHYSICS MASTER PROGRAM**

NOVEMBER 2017

Approval of the Graduate School of Natural and Applied Sciences, Atılım University.

Prof. Dr. Ali Kara

Director

I certify that this thesis satisfies all the requirements as a thesis for the degree of Master of Science in Physics at the Department of Applied Physics Master Program.

Prof. Dr. Ramazan Aydın

Head of Department

This is to certify that we have read the thesis “Electron Transport Through One-Dimensional Quantum Systems” submitted by “Latifa Algadi” and that in our opinion it is fully adequate, in scope and quality, as a thesis for the degree of Master of Science.

Assoc. Prof. Dr. Hüseyin

Oymak

Supervisor

Examining Committee Members:

Prof. Dr. Ramazan Aydın
Atılım University

Prof. Dr. Enver Bulur
Middle East Technical University

Assoc. Prof. Dr. Kemal Efe Eseller
Atılım University

Assoc. Prof. Dr. Emre Taşçı
Hacettepe University

Assoc. Prof. Dr. Hüseyin Oymak
Atılım University

Date: November 1, 2017



I declare and guarantee that all data, knowledge and information in this document has been obtained, processed and presented in accordance with academic rules and ethical conduct. Based on these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: Latifa Algadi

Signature:

XXXXXXXXXX

ABSTRACT

ELECTRON TRANSPORT THROUGH ONE-DIMENSIONAL QUANTUM SYSTEMS

Algadi, Latifa

M.S., Department of Applied Physics Master Program

Supervisor: Assoc. Prof. Dr. Hüseyin Oymak

November 2017, 234 pages

We study the quantum transport of an electron through some one-dimensional (1D) quantum systems (or seemingly 2D systems which are practically reducible to 1D systems) coupled to two ideal leads. After reviewing the rudiments of electricity phenomena, leading to the concept of current, and some magnetism phenomena in terms of classical mechanics, we provide the basic formalism of quantum mechanics by elucidating its working machinery in terms of Dirac notation. We next deal with 1D scattering problems and study in detail some examples containing delta-function potentials. We then focus on the so-called two-state quantum systems, like the ammonia molecule, the hydrogen molecular ion, and the hydrogen molecule. Later on we investigate the propagation of an electron in a 1D crystal lattice with all of its fine details. After settling down the rudiments of 1D finite linear atomic-chain and atomic-ring structures, we come to the subject of resonant transport that we investigate through some illustrative problems. We then present the mathematics and physics of the Aharonov-Bohm (AB) rings, for which we touch the classical Lagrangian and Hamiltonian dynamics for a charged particle moving in an electromagnetic field and the relevant gauge transformations. Finally, for obtaining transport properties of 1D quantum systems, we demonstrate how to use a pruning technique, called Riccati-ratios method, which changes the time-independent discrete Schrödinger equation into a two-point recursive equation. We test the method in some 1D well-known

systems as well as in single and double AB rings by simply reducing them to one dimension with the help of a decimation method.

Keywords: Electron transport, quantum transport, one-dimensional quantum systems, scattering problems, transmission and reflection, delta-function potentials, Dirac notation, atomic chains, atomic rings, resonance, antiresonance, Aharonov-Bohm rings, Riccati-ratios method



ÖZ

BİR BOYUTLU KUANTUM SİSTEMLERDE ELEKTRON İLETİMİ

Algadi, Latifa

Yüksek Lisans, Uygulamalı Fizik Yüksek Lisans Bölümü

Tez Yöneticisi: Doç. Dr. Hüseyin Oymak

Kasım 2017, 234 sayfa

İki ideal iletkenle bağlanmış, bir boyutlu kuantum sistemlerde (veya bir boyutlu sistemlere indirgenebilen iki boyutlu sistemlerde) bir elektronun kuantum iletim problemini çalıştık. Klasik mekaniğe göre, akım kavramına yönelik en temel elektrik ve bazı diğer manyetik olguları inceledikten sonra, kuantum mekaniğin temel formalizmini verdik. Bunun için kuantum mekaniğin çalışma prensiplerini Dirac notasyonu kullanarak formüle ettik. Sonra bir boyutlu saçılma problemleri ile ilgilendik ve delta fonksiyonu içeren bazı örnekleri detaylı olarak çalıştık. Daha sonra, amonyak molekülü, hidrojen molekül iyonu ve hidrojen molekülü gibi iki durumlu kuantum sistemler üzerine yoğunlaştık. Devamında, bir elektronun bir boyutlu kristal örgü içindeki yayılımını tüm ince detayları ile araştırdık. Bir boyutlu doğrusal atom zinciri ve atom halkası yapılarının temellerini oturttukten sonra, açıklayıcı problemler yoluyla incelediğimiz rezonant iletim konusuna geldik. Ardından ise Aharonov-Bohm (AB) halkalarının matematik ve fiziğini verdik; bunun için, elektromanyetik alan içinde hareket eden yükü bir parçacıkla ilgili klasik Lagrange ve Hamilton dinamikleri ve bununla alakalı ayar dönüştürme konularına dokunduk. Son olarak, bir boyutlu kuantum sistemlerin iletim özelliklerini elde edebilmek için, zamandan bağımsız Schrödinger denklemini, iki noktadan yinelemeli başka bir denkleme çeviren, “Ricatti oranlar yöntemi” olarak adlandırılan bir budama/kesme tekniğinin nasıl kul-

lanılacağını gösterdik. Bu yöntemi, basit bir seyreltme/kaldırma tekniği yardımı ile, sonuçları iyi bilinen bazı bir boyutlu sistemlerde ve ayrıca tekli ve çiftli AB halkalarında test ettik.

Anahtar Kelimeler: Elektron iletimi, kuantum iletimi, bir boyutlu kuantum sistemler, saçılma problemleri, iletim ve yansıma, delta fonksiyon potansiyelleri, Dirac gösterimi, atomik zincirler, atomik halkalar, rezonans, antirezonans, Aharonov–Bohm halkaları, Ricatti oranlar yöntemi





To My Beloved Country, Libya, the Beautiful.
To My Lovely Daughters, Rahaf, Ranim, and Laila.



ACKNOWLEDGMENTS

First I wish to express my praise and thanks to Allah. Then, I want to express my gratitude to my supervisor Assoc. Prof. Dr. Hüseyin Oymak; it is so hard to put into words to express my thankfulness. His office's door was always open whenever I ran into trouble, or I had a question about my research. I always say that I am so lucky that he is my supervisor. I'll forever be grateful to him. And I should express my gratitude to everyone who has given immense practical help. Finally, I must give my profound gratitude to my parents, my husband Salem Rajab Falefla, and my relatives for providing me with unfailing support and continuous encouragement throughout my years of study.



TABLE OF CONTENTS

ABSTRACT	iv
ÖZ	vi
DEDICATION	viii
ACKNOWLEDGMENTS	ix
TABLE OF CONTENTS	x
LIST OF FIGURES	xv
CHAPTERS	
1 INTRODUCTION	1
2 CLASSICAL REVIEW	4
2.1 Electric Charge	4
2.2 Coulomb's Law	5
2.3 Electric Field	6
2.4 Gauss' Law	7
2.5 Electric Potential Energy and Electric Potential	8
2.6 Current	10
2.7 Current Density	12
2.8 Resistance and Resistivity	13
2.9 Variation of Resistivity with Temperature	15
2.10 Ohm's Law	15
2.11 A Rudimentary Microscopic Account for Ohm's Law	16
2.12 A Charged Particle in a Magnetic Field	21
2.12.1 Lorentz Force	23
2.12.2 Magnetic Flux	24
3 THE WORKING MACHINERY OF QUANTUM MECHANICS	25
3.1 Postulate of Quantum Mechanics	26

3.1.1	Postulate I	26
3.1.1.1	The Linear Momentum Operator	26
3.1.1.2	The Energy Operator	28
3.1.2	Postulate II	30
3.1.2.1	Position Measurement	30
3.1.3	Postulate III	31
3.1.4	Postulate IV	33
3.2	Quantum Dynamics	37
4	SCATTERING PROBLEMS IN ONE DIMENSION	39
4.1	General Features of the 1D Schrödinger Equation	39
4.2	Bound and Unbound States	41
4.2.1	Continuity Equation	44
4.2.2	Transmission and Reflection Coefficients	46
4.2.3	The Simple Step Problem: A Worked Example	48
5	DELTA-FUNCTION POTENTIALS	52
5.1	Delta-Function Well	52
5.1.1	Bound States	53
5.1.2	Unbound States	56
5.2	Delta-Function Barrier	60
5.3	Scattering Matrix	62
5.3.1	The S-Matrix of a Delta-function Well: A Worked Example	64
6	DIRAC NOTATION FOR QUANTUM STATES	66
6.1	Dirac Notation for Continuous States	66
6.2	Dirac Notation for Discrete, Path-Like, States	67
6.2.1	Time Evolution of Amplitudes in Time	70
6.3	Dirac Notation for Discrete States (Continued)	72
6.3.1	Measurements in Dirac Notation	77
6.4	The Time Evolution of States in the Dirac Notation	79
6.5	The Hamiltonian (or, Energy) Matrix	82

7	TWO-STATE PROBLEMS	85
7.1	A Fully Worked Example: The Ammonia Molecule	85
7.2	The Hydrogen Molecular Ion	92
7.3	The Hydrogen Molecule	98
8	PROPAGATION OF AN ELECTRON IN A CRYSTAL LATTICE . .	103
8.1	The Base States for an Electron in a 1D Chain of Atoms . . .	105
8.2	The States Possessing Definite Energy	108
8.3	The Time-Dependent States	114
8.4	An Electron in a 3D (Perfect) Crystalline Lattice	117
8.5	Other Probable States in a Crystalline Lattice Structure . . .	120
8.6	Scattering from Irregularities (or Imperfections) in a 1D Lattice	122
8.7	An Electron Trapped by an Impurity Atom	127
8.8	Bound States, Again, and Their Relationship to Scattering Amplitudes	129
9	1D FINITE ATOMIC-CHAIN and ATOMIC-RING STRUCTURES .	131
9.1	Introduction	131
9.2	A 1D (Linear) Chain of N Atoms	132
9.2.1	An Easy Solution Leading to (9.17) and (9.18) . .	138
9.3	A Ring-Like Chain of N Atoms	140
10	RESONANT TRANSPORT: A COLLECTION OF PROBLEMS . . .	144
10.1	Transport Through a Single Central Site	144
10.2	Transport Through Two Horizontal Sites	157
10.3	Transport Through Three Horizontal Sites	166
10.4	Transport Through Four Horizontal Sites	177
10.5	Transport Through a Four-Site Ring Structure	181
11	AHARONOV–BOHM RINGS: A DETAILED EXPOSITION	186
11.1	Introduction	186
11.2	A Review of Classical Mechanics	187
11.2.1	The Lagrangian Dynamics	187

11.2.2	The Hamiltonian Dynamics	188
11.2.3	Hamiltonian of a Charged Particle in an Electro- magnetic Field	189
11.3	Gauge Transformations	191
11.3.1	Double–Slit Diffraction	193
11.3.2	Aharonov–Bohm Effect	194
11.4	An Isolated Aharonov–Bohm Ring	197
11.4.1	Field–Free Case	197
11.4.2	Constant Magnetic Field Case	198
11.4.3	Gauge–Invariance Treatment	200
11.5	The Aharonov–Bohm Ring as a Part of an External Circuit	202
11.5.1	Field–Free Case	202
11.5.2	Constant Magnetic Field Case	203
11.6	An Aharonov–Bohm Ring Exposed to a Time–Dependent Magnetic Field	210
11.6.1	Oscillating Magnetic Flux	211
12	THE RICATTI-RATIOS METHOD IN DISCRETE TRANSPORT PROBLEMS	216
12.1	The Ricatti System	216
12.2	Models Which Are Practically Reducible to Ricatti Systems: the Decimation Technique	222
12.2.1	Single Aharonov-Bohm Ring	222
12.2.2	Double AB Ring	227
	REFERENCES	232

LIST OF FIGURES

FIGURES

Figure 1.1 A symbolic depiction of a 1D mesoscopic system represented by a number of central sites coupled to 1D left and right leads.	2
Figure 2.1 A portion of uniform wire.	11
Figure 2.2 A symbolic depiction to represent the complicated zigzag motion of a negative charge carrier within a conductor.	12
Figure 2.3 A scetch for the calculation of resistance.	14
Figure 2.4 The voltage V versus current i graph for an ohmic device.	16
Figure 2.5 A mechanical model for a crystalline bulk material.	17
Figure 2.6 A diagram for the rapid motion of two representative free electrons in a metal for the zero electric field case in which the drift velocity of the electrons is zero.	18
Figure 2.7 A diagram similar to Fig. 2.6 for the case in which there is now an electric field; in this case free electrons have non-zero drift velocities.	18
Figure 2.8 A mechanical analogue for the electrical conduction in a metal.	19
Figure 3.1 A wave packet at some time t	36
Figure 4.1 Two distinct domains of a particle that is in motion in a 1D potential (energy) field $V(x)$. Domain I is classically <i>allowed region</i> , while domain II is classically <i>forbidden region</i> . The dashed border passes through the so-called classical <i>turning point</i> , where $T = V$	40
Figure 4.2 A representative wave function for the energy profile shown in Fig. 4.1.	41

Figure 4.3	A representative potential (energy) profile of a 1D scattering problem. An incident beam with current density $\mathbb{J}_{\text{inc}}$ comes from $x = -\infty$. It encounters the central potential barrier $V(x)$; some of it is transmitted towards $x = +\infty$ with $\mathbb{J}_{\text{trans}}$ and the remaining is reflected back again towards $x = -\infty$ with $\mathbb{J}_{\text{refl}}$. Before and after the central scattering region, the potential $V(x)$ is constant at some specific values.	43
Figure 4.4	The geometry for the continuity equation (4.14).	45
Figure 4.5	The simple step potential (energy) barrier treated in the worked example.	48
Figure 4.6	The simple step barrier with representative solutions in both regions.	50
Figure 5.1	The (Dirac) delta-function (potential energy) well.	52
Figure 5.2	The wave function $\varphi(x)$ of the only bound state of the delta-function well.	54
Figure 5.3	All the possible components of the wave function $\varphi(x)$ for unbound states of the delta-function well.	58
Figure 5.4	The (Dirac) delta-function barrier.	60
Figure 5.5	A depiction for a generalized scattering problem. Note that $V(x) = 0$ in Region I and II.	62
Figure 6.1	The double-slit experiment.	67
Figure 6.2	A multi-slit experiment.	69
Figure 7.1	The two states of the ammonia molecule.	85
Figure 7.2	Probability curves for the ammonia molecule. P_1 (P_2) is the probability that the molecule is found in state $ 1\rangle$ ($ 2\rangle$).	90
Figure 7.3	The base states for the H_2^+ molecule.	92
Figure 7.4	The stationary-state energies E_I and E_{II} of the hydrogen molecular ion, plotted against the distance D between its two protons, for the case in which the repulsive proton-proton electrostatic energy is <i>not</i> included.	95
Figure 7.5	The same as Fig. 7.4, but this time for the case in which the repulsive proton-proton electrostatic energy <i>is</i> included.	96

Figure 7.6	The two bases states for the hydrogen molecule.	98
Figure 7.7	The stationary-state energies of the hydrogen molecular, plotted against the distance D between its two protons, where the repulsive proton- proton electrostatic energy is included.	100
Figure 8.1	(a) A 1D (lattice) composed of a very long chain of atoms. (b)-(d) The base states of the extra electron.	106
Figure 8.2	The graph of the real part of the amplitude $C_n(t)$ at a fixed time t . .	110
Figure 8.3	The graph of $E = E_0 - 2A \cos kb$, $A > 0$, as a function of the wave number k	111
Figure 8.4	The graphs of the real part of the space-part amplitude $a(x_n)$ for $k = \pi/4b$ (curve 1) and $k = 7\pi/4b$ (curve 2).	113
Figure 8.5	Formation of beats as a superposition of a two waves with <i>slightly</i> different wavelengths. The upper plot is for $f_1(x) = \sin[2\pi(10)x]$, the middle plot is for $f_2(x) = \sin[2\pi(10.1)x]$, and the lower plot is their sum, $f_1(x) + f_2(x)$	115
Figure 8.6	The real part of the total amplitude $C(x_n, t)$ versus the atomic po- sition x . The plot is said to be in the shape of a wave packet because it is the superposition of various states whose wave numbers are only slightly different.	115
Figure 8.7	1D chain of atoms containing a single impurity atom at $n = 0$. Also indicated are the incident, scattered (or reflected), and transmitted waves. .	125
Figure 8.8	The curves for the probabilities that the single “trapped” electron is found at any atoms near the central impurity atom.	129
Figure 9.1	A 1D (linear) chain composed of N hydrogen-like atoms.	132
Figure 9.2	The eigenenergy spectra for 1D (linear) chains containing $N = 1-5$, and ∞ atoms, depicted using (9.18).	138
Figure 9.3	A ring-like chain composed of N hydrogen-like atoms.	140
Figure 9.4	The eigenenergy spectra for atomic ring structures containing $N =$ 1, 3–5, and ∞ atoms, depicted using (9.30).	143
Figure 9.5	The band structure of the ring containing an infinite number of atoms.	143

Figure 10.1 A 1D (linear) atomic chain containing a single “impurity” at site $j = 0$	144
Figure 10.2 A plot for the number of energy states $N(E)$, per unit volume of a 3D metallic crystalline structure, versus energy E at $T = 0$ and $T = 300$ K. The Fermi energy is taken as $E_F = 3$ eV.	147
Figure 10.3 (a) The transmission coefficient $T(\varepsilon, \varepsilon')$ and (b) the reflection coefficient $R(\varepsilon, \varepsilon)$ versus the incoming electron energy ε for the 1D linear chain with single impurity. The common system parameters are $t = 1.0$, $t_1 = 0.2$, $v_c \equiv v_0 = 0$, and $v = v' = 0$	154
Figure 10.4 The transmission coefficient $T(\varepsilon, \varepsilon')$ versus the incoming electron energy ε for the 1D linear chain with single impurity for different values of the central gate potential $v_c \equiv v_0$ (see Fig. 10.1). The other common system parameters are $t = 1.0$, $t_1 = 0.2$, and $v = v' = 0$	155
Figure 10.5 The transmission coefficient $T(\varepsilon, \varepsilon')$ versus the incoming electron energy ε for the 1D linear chain with single impurity for different values the hopping constant t_1 mediating the transmission between the central site and the left and right leads (see Fig. 10.1). The other common system parameters are $t = 1.0$, $v_c \equiv v_0 = 0$, and $v = v' = 0$	156
Figure 10.6 A 1D (linear) atomic chain containing two “impurities” at sites $j = 0$ and $j = 1$	157
Figure 10.7 (a) The transmission coefficient $T(\varepsilon, \varepsilon')$ and (b) the reflection coefficient $R(\varepsilon, \varepsilon)$ versus the incoming electron energy ε for the 1D linear chain with two impurities. The common system parameters are $t = 1.0$, $t_1 = 0.2$, $t_2 = 1.0$, $v_c \equiv v_0 = v_1 = 0$, and $v = v' = 0$	161
Figure 10.8 The transmission coefficient $T(\varepsilon, \varepsilon')$ versus the incoming electron energy ε for the 1D linear chain with two central impurities for different values the hopping constant t_2 mediating the transmission between the two central sites (see Fig. 10.6) and the gate potentials $v_c \equiv v_0 = v_1$. The other common system parameters are $t = 1.0$, $t_1 = 0.2$ and $v = v' = 0$	164

Figure 10.9 The transmission coefficient $T(\varepsilon, \varepsilon')$ versus the incoming electron energy ε for the 1D linear chain with two central impurities for different values the hopping constant t_1 mediating the transmission between the central impurity region and the left and right leads (see Fig. 10.6). The other common system parameters are $t = 1.0$, $t_2 = 1.0$, $v_c \equiv v_0 = v_1 = 0$, and $v = v' = 0$	165
Figure 10.10A 1D (linear) atomic chain containing three “impurities” at sites $j = 0$, $j = 1$, and $j = 2$	166
Figure 10.11(a) The transmission coefficient $T(\varepsilon, \varepsilon')$ and (b) the reflection coefficient $R(\varepsilon, \varepsilon)$ versus the incoming electron energy ε for the 1D linear chain with three impurities. The common system parameters are $t = 1.0$, $t_1 = 0.2$, $t_2 = 1.0$, $v_c \equiv v_0 = v_1 = v_2 = 0.0$, and $v = v' = 0$	173
Figure 10.12The transmission coefficient $T(\varepsilon, \varepsilon')$ versus the incoming electron energy ε for the 1D linear chain with three central impurities for different values the hopping constant t_2 mediating the transmission between any of the next-neighboring three central sites (see Fig. 10.10) and the gate potentials $v_c \equiv v_0 = v_1 = v_2$. The other common system parameters are $t = 1.0$, $t_1 = 0.2$ and $v = v' = 0$	175
Figure 10.13The transmission coefficient $T(\varepsilon, \varepsilon')$ versus the incoming electron energy ε for the 1D linear chain with three impurities for different values the hopping constant t_1 mediating the transmission between the central impurity region and the left and right leads (see Fig. 10.10). The other common system parameters are $t = 1.0$, $t_2 = 1.0$, $v_c \equiv v_0 = v_1 = v_2 = 0$, and $v = v' = 0$	176
Figure 10.14A 1D (linear) atomic chain containing four “impurities” at sites $j = 0-4$	177
Figure 10.15(a) The transmission coefficient $T(\varepsilon, \varepsilon')$ and (b) the reflection coefficient $R(\varepsilon, \varepsilon)$ versus the incoming electron energy ε for the 1D linear chain with four impurities. The common system parameters are $t = 1.0$, $t_1 = 0.2$, $t_2 = 1.0$, $v_c \equiv v_0 = v_1 = v_2 = v_3 = 0.0$, and $v = v' = 0$	178

Figure 10.16 The transmission coefficient $T(\varepsilon, \varepsilon')$ versus the incoming electron energy ε for the 1D linear chain with four central impurities for different values the hopping constant t_2 mediating the transmission between any of the next-neighboring three central sites (see Fig. 10.14) and the gate potentials $v_c \equiv v_0 = v_1 = v_2 = v_3$. The other common system parameters are $t = 1.0$, $t_1 = 0.2$ and $v = v' = 0$ 179

Figure 10.17 The transmission coefficient $T(\varepsilon, \varepsilon')$ versus the incoming electron energy ε for the 1D linear chain with four impurities for different values the hopping constant t_1 mediating the transmission between the central impurity region and the left and right leads (see Fig. 10.14). The other common system parameters are $t = 1.0$, $t_2 = 1.0$, $v_c \equiv v_0 = v_1 = v_2 = v_3 = 0$, and $v = v' = 0$ 180

Figure 10.18 Two 1D (linear) atomic chains connected to a central ring-like structure containing four “impurities” at sites $j = 0-4$ 181

Figure 10.19 (a) The transmission coefficient $T(\varepsilon, \varepsilon')$ and (b) the reflection coefficient $R(\varepsilon, \varepsilon)$ versus the incoming electron energy ε for two 1D (linear) atomic chains connected to a central four-site ring-like “impurity” structure. The common system parameters are $t = 1.0$, $t_1 = 0.2$, $t_2 = 0.9$, $v_c \equiv v_0 = v_1 = v_2 = v_3 = 0.0$, and $v = v' = 0$ 182

Figure 10.20 The transmission coefficient $T(\varepsilon, \varepsilon')$ versus the incoming electron energy ε for two 1D (linear) atomic chains connected to a central four-site ring-like “impurity” structure for different values the hopping constant t_2 mediating the transmission between any of the next-neighboring three central sites (see Fig. 10.18) and the gate potentials $v_c \equiv v_0 = v_1 = v_2 = v_3$. The other common system parameters are $t = 1.0$, $t_1 = 0.2$ and $v = v' = 0$. 183

Figure 10.21 The transmission coefficient $T(\varepsilon, \varepsilon')$ versus the incoming electron energy ε for two 1D (linear) atomic chains connected to a central four-site ring-like “impurity” structure for different values the hopping constant t_1 mediating the transmission between the central impurity region and the left and right leads (see Fig. 10.18). The other common system parameters are $t = 1.0$, $t_2 = 0.9$, $v_c \equiv v_0 = v_1 = v_2 = v_3 = 0$, and $v = v' = 0$ 185

Figure 11.1 Double–slit diffraction experiment.	193
Figure 11.2 A longitudinally magnetized iron.	195
Figure 11.3 Double–slit diffraction around a magnetic flux tube.	195
Figure 11.4 An Aharonov–Bohm ring with no magnetic field.	197
Figure 11.5 An Aharonov–Bohm ring exposed to a magnetic field $\mathbf{B} = B_0\hat{z}$. . .	198
Figure 11.6 An isolated Aharonov–Bohm ring which is now part of an external circuit.	202
Figure 11.7 An Aharonov–Bohm ring, which is now part of an external circuit, exposed to a constant magnetic field $\mathbf{B} = B_0\hat{z}$	203
Figure 11.8 The same system as that in Fig. 11.7.	204
Figure 11.9 The same system as that in Fig. 11.7 with the values of the wave functions at the two junctions.	206
Figure 11.10A double Aharonov–Bohm ring system exposed to a magnetic field $\mathbf{B} = B_0\hat{z}$ and connected to two external leads.	208
Figure 11.11Transmission coefficient T versus magnetic flux f for $k\ell = 0.01\pi$ (solid curve) and $k\ell = 0.49\pi$ (dotted curve) for the single and double AB ring structures. It is seen that both resonance and antiresonance phenom- ena exist in the double AB rings case, whereas only the resonance phe- nomenon exists in the single AB ring case. These are the plots of (11.53) and (11.54) that we have produced with the help of a MATHEMATICA code). The reader should compare these plots with the original ones reported in the work of Yi <i>et al.</i> [72].	209
Figure 11.12An Aharonov–Bohm ring exposed to a time–dependent magnetic field $\mathbf{B}(t) = B_0(t)\hat{z}$	210
Figure 11.13An Aharonov–Bohm ring exposed to an oscillating magnetic field $\mathbf{B}(t) = B_0(t)\hat{z}$	211
Figure 12.1 The Riccati system, R_s , under study.	216
Figure 12.2 The transmission probability $T(\varepsilon, \varepsilon')$ versus the incoming electron energy ε for the R_1 – R_4 systems	220
Figure 12.3 An imitation of a single AB ring.	222

Figure 12.4 The transformed form of the single AB ring in Fig. 12.3.	224
Figure 12.5 The transmission probability $T(\varepsilon, \varepsilon')$ versus the incoming electron energy ε for the single AB ring shown in Fig. 12.3.	225
Figure 12.6 An imitation of a double AB ring.	227
Figure 12.7 The transformed form of the double AB ring in Fig. 12.6.	228
Figure 12.8 The transmission probability $T(\varepsilon, \varepsilon')$ versus the incoming electron energy ε for the double AB ring shown in Fig. 12.6.	229
Figure 12.9 The transmission probability $T(\varepsilon, \varepsilon')$ versus the magnetic flux f for the single and double AB rings at $k = 0.05\pi$ (solid lines) and at $k = 0.495\pi$ (dashed lines).	230



CHAPTER 1

INTRODUCTION

In this study we deal generally with “one-dimensional (1D) mesoscopic systems” which are connected to some ideal “1D conductors.” With a “mesoscopic system” we mean a small aggregate of, say, semiconductor quantum dots, or single-atomic or single-molecular devices. The adjective “1D” applies also to some (quasi) 2D mesoscopic systems, like Aharonov-Bohm (AB) ring systems, which can be shown to be reducible to 1D ones. “1D conductors” are any materials whose electrical conductivity is substantially higher only along one specific direction than those along other directions [1]. If $\ell_{\parallel}$ and $d_{\parallel}$ ($\ell_{\perp}$ and $d_{\perp}$) are respectively the electron mean free path and the interatomic distance along (perpendicular to) the 1D axis, then we call a material a 1D conductor when the conditions $\ell_{\parallel} \geq d_{\parallel}$ and $\ell_{\perp} \ll d_{\perp}$ are satisfied; consequently, we can speak of a material possessing a “1D conduction band.” The notion of such a 1D conduction band can be best visualized by considering a 1D atomic chain in which interatomic distances are small enough for the overlap of electron wave functions to be large so that a conduction band can be formed [1].

The most important reason for studying a 1D system, or any system which can be practically reduced to one dimension, rather than its 3D or higher versions, is to facilitate the mathematics involved. There are a good number of interesting physical problems which are exactly soluble in one dimension. Although they sometimes seem to be oversimplified, investigations of 1D systems surprisingly give invaluable insights into the understanding of the system at hand, usually giving rise to results which can be ultimately applicable to 3D systems. Examples for such physically interesting systems are polymers, magnetic molecules, long chain molecules, etc. [2]. It seems in order to note before passing that, despite the fact that the 1D modeling greatly sim-

plifies the mathematics, we still do not have a full understanding of the properties 1D electronic systems [1].

The present manuscript is an introductory (and, at the same time, somehow complementary) detailed exposition, with the possible finest details, to two previous works of Oymak, Shin, and Hong carried out some years ago [3, 4], where they aimed to focus on a general 1D lattice whose sites have only two nearest neighbors. In these past works and in the present one, we consider specifically, in the tight-binding approximation, a 1D mesoscopic system represented by a number of central sites which are coupled to two noninteracting reservoirs (i.e., via left and right leads) through resonant tunneling, like the one depicted in Fig. 1.1. The two noninteracting reservoirs are to be represented by two 1D conductors. The 1D central mesoscopic system under question will contain no specific interaction, i.e., including left and right leads, we are interested in a nominally 1D quantum wire of some kind. In the previous two works [3, 4], the central sites contained electron-phonon interaction and/or electron-photon interaction (via an applied time-periodic magnetic flux or time-periodic voltage bias), and/or similar excitations. Our main aim in this present work is to probe the transport properties of an electron while it resonantly tunnels the central mesoscopic system, say, from the left to the right lead.

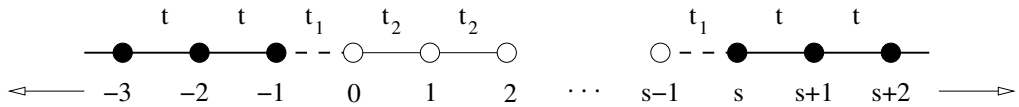


Figure 1.1: A symbolic depiction of a 1D mesoscopic system represented by a number of central sites coupled to 1D left and right leads.

Since it naturally contains the conduction band phenomenon, a tight-binding (TB) formulation in its simplest form is to be used in this work. The TB methods were originally developed as a technique for calculating the electronic properties of crystalline solids from atomic wave functions [5], and since then they have been the central techniques of many theoretical solid state physics problems. Yet even today, because they are simple and physically intuitive, the TB methods continue to be one of the most widely used methods, especially in modeling of materials [6]. In particular, they have

proven useful to the utmost in studying the electronic transport through molecules and small devices [7]. One of the subjects of this work is again a TB method which is exploited in considering resonant tunneling of an electron through a central mesoscopic system.

Our ultimate aim in this study is to solve the time-independent Schrödinger equation by using the Ricatti-ratios method, after the works of Heinrichs [8] and Faizabadi and Ebrahimi [9], which is a very efficient “pruning” [10] technique. We shall follow a single-electron picture, hence the treatment in this manuscript will not take account of many-body effects like the Coulomb repulsion, spin-spin interaction, electron-electron interaction, etc.

The present manuscript may seem at first unusually voluminous, especially if it is to be pertinent to a master of science study. So it is, indeed. We just desire that a novice prospective researcher who is about to embark on a study about quantum transport may not suffer from the inadequacy of a good literature source introducing and explaining the necessary details from ground-zero to a fairly high level. We also wish that s/he may not be disheartened by the highly advanced and complicated language of today’s literature papers or books which usually adress to already expert researchers. In these respects, the present manuscript is believed to contain all the necessary materials needed to give a quick start to quantum transport phenomena which are elucidated by *not* using so much unnecessarily sophisticated language.

Before plunging into the realm of quantum mechanics, it is our very debt to pay homage to R. P. Feynman, or *Feynman, The Great Master*; as the reader will appreciate shorlty, without his long-enduring masterworks, we could have accomplished simply nothing.

CHAPTER 2

CLASSICAL REVIEW¹

In this chapter we shall review mainly the rudiments of electricity phenomena in terms of classical mechanics. This short survey will follow such a way that leads to the concept of *current*, on which we shall place much more emphasis than the other concepts, because current is the central subject of this study. In the last part, there will be a quite small section for some magnetism phenomena which will be of great interest when we investigate the transport properties of *Aharonov-Bohm rings*. The principal aim of this chapter is to make the serious reader acquainted with the vocabulary of the subject that we will use in the rest of the manuscript. In the following, the most important outcomes will be written in **coloured** equations.

2.1 Electric Charge

It is an empirical fact that all kinds of matter in the universe contain intrinsically the so-called (*electric*) *charges*, which are simply of the order of Avogadro's number, $N_A = 6.022 \times 10^{23} \text{ mol}^{-1}$.

We say that the total charges in the universe is *conserved*, because charge cannot be created nor destroyed.

There are two kinds of charges; out of mathematical convenience they are referred to as *positive* (+) and *negative* (−) charges. Experiments show that similar (dissimilar) charges repel (attract) each other.

¹ This chapter might be skipped in first reading without losing the integrity and continuity of the manuscript.

The general mathematical symbol that is used for a point charge is q and for a continuous charge distribution is Q . The SI unit of charge is *Coulomb*, C, which is the total charge of approximately 6.24×10^{18} *charge carriers* of the same kind.

Charge is said to be *quantized*, meaning that the amount of charge of any object is actually an integral number of the magnitude of an electron's charge, in the form

$$q = ne, \quad n = 0, \pm 1, \pm 2, \pm 3 \dots$$

In other words, the smallest unit of charge is the charge of an electron, $-e$ (or that of a proton, $+e$); the numerical value of e is

$$e = 1.602 \times 10^{-19} \text{ C}$$

In most general terms we classify matters into two: *conductors*, throughout which charges can wander almost freely, and *insulators*, in which charges cannot or does not move, i.e., they are fixed around the same regions of space.

2.2 Coulomb's Law

The attraction or repulsion force between two point particles of charges q_1 and q_2 , separated by a distance r_{12} in vacuum, is given by *Coulomb's law* as

$$\mathbf{F}_{12} = k \frac{q_1 q_2}{r_{12}^2} \hat{\mathbf{r}}_{12} \quad (\text{N}) \quad (2.1)$$

where $\mathbf{F}_{12}$ is the force on charge q_2 because of q_1 , $\hat{\mathbf{r}}_{12}$ is the unit vector from q_1 to q_2 along the imaginary line connecting the two charges, and k is the *Coulomb (electrostatic) constant*, with the numerical value

$$k = 8.99 \times 10^9 \text{ N} \cdot \text{m}^2/\text{C}^2$$

There is a close relationship between the Coulomb constant and the *permittivity constant* ϵ_0 of vacuum:

$$k = \frac{1}{4\pi\epsilon_0}$$

The numerical value of ϵ_0 is

$$\epsilon_0 = 8.85 \times 10^{-12} \text{ C}^2/\text{N} \cdot \text{m}^2$$

If there are more than two charges, say four, the net force on one of them, say on the third charge, due to the remaining charges is found by summing vectorially the pairwise forces on it:

$$\mathbf{F}_3 = \mathbf{F}_{13} + \mathbf{F}_{23} + \mathbf{F}_{43}$$

This pairwise additivity of the Coulomb law is called *superposition principle*.

2.3 Electric Field

Coulomb's law does not give a proper answer to the question: "How do any two charges know the presence of each other so that they can apply forces on each other?" In order to clarify this issue, the *electric field* concept, which is a somewhat visual mathematical tool, has been invented. We say, in its simplest form, that a particle of charge q carries inherently always its own electric field around itself. Whenever another charge q_0 , usually called a *test charge*, is brought from infinity and placed in some vicinity of charge q , it will be affected by the electric field $\mathbf{E}$ of charge q at that place, so that there will be an electric force on q_0 due to q . This very force is *defined* as

$$\mathbf{F}_0 \equiv q_0 \mathbf{E} \quad (2.2)$$

Comparing this with the form of Coulomb's law (2.1), we see that the electric field of charge q at a point radial point r from itself is

$$\mathbf{E} = k \frac{q}{r^2} \hat{\mathbf{r}} \quad (\text{N/C}) \quad (2.3)$$

where $\hat{\mathbf{r}}$ is the unit vector directed along the line joining the charge to the point of interest.

Due to the definition (2.2) above, the *electric field lines* starts from positive charges and terminate at negative charges, both of them might sometimes be at infinity. They are drawn commensurate with the magnitude of the electric field in the space of interest: the larger the magnitude of the electric field (of some charges) is, the more electric field lines are depicted. It should be noted that electric field lines never cross each other, as a result of (2.3).

If there happens to be a group of charges, their net electric field at some field point is found by applying the superposition principle as

$$\mathbf{E} = k \sum_n \frac{q_n}{r_n^2} \hat{\mathbf{r}}_n \quad (2.4)$$

which for a continuous charge distribution becomes

$$\mathbf{E} = k \int \frac{dq}{r^2} \hat{\mathbf{r}} \quad (2.5)$$

The dynamics of a point charge q , with mass m , within an electric field $\mathbf{E}$ is determined from the concurrent application Newton's second law and Coulomb's law, through (2.2), as

$$\mathbf{F} = m\mathbf{a} = q\mathbf{E}$$

whence the acceleration of charge q is

$$\mathbf{a} = \frac{q}{m} \mathbf{E} \quad (\text{m/s}^2) \quad (2.6)$$

2.4 Gauss' Law

For determining the electric field of some charge distribution, discrete or continuous, at a field point in space, there is a very powerful method, which uses *Gauss' law*, which states that the total electric flux Φ_E penetrating the surface of *any* (imaginary) closed *Gaussian surface* amounts to the total charge q_{enc} enclosed by this surface divided by the permittivity constant ϵ_0 :

$$\Phi_E = \oint \mathbf{E} \cdot d\mathbf{A} = \frac{q_{\text{enc}}}{\epsilon_0} \quad (2.7)$$

A rudimentary application of Gauss' law to a conductor (or metal) which is in *electrostatic equilibrium* leads to three important conclusions:

1. The electric field inside the conductor is everywhere zero, $\mathbf{E}_{\text{in}} = \mathbf{0}$.
2. An excess amount of charge distributes itself entirely on the surface of the conductor, i.e., there is no excess charge within the conductor, $q_{\text{in}} = 0$.
3. Electric field lines are always perpendicular to the surface of the conductor.

2.5 Electric Potential Energy and Electric Potential

We know that its inverse-square law makes the Colomb force (2.1) a *conservative* force. Therefore, we can associate a potential energy for systems of charges. If a test charge q_0 moves from some initial point i to a final point f within and under the effect of an electric field $\mathbf{E}$, we can associate a potential energy change ΔU for the charge-electric field system to the work W done by the electric force $q_0\mathbf{E}$, by making use of the defining equation for the potential energy, as

$$\Delta U = U_f - U_i \equiv -W = - \int_i^f \mathbf{F} \cdot d\mathbf{s} = - \int_i^f q_0\mathbf{E} \cdot d\mathbf{s} \quad (2.8)$$

At this point we define the so-called (*electric*) *potential* as

$$V \equiv \frac{U}{q_0} \quad (\text{J/C}) \quad (2.9)$$

It is customary to use *volt* (V) for the unit of potential:

$$1 \text{ J/C} \equiv 1 \text{ V}$$

With the definition (2.9), we recast (2.8) as

$$\Delta V = V_f - V_i = - \int_i^f \mathbf{E} \cdot d\mathbf{s} \quad (2.10)$$

An immediate application of (2.10) to a charged conductor reveals that the potential everywhere inside the conductor is constant and amounts to that at its surface.

Another practical result obtained from the use of (2.10) is that the potential difference between any two points, separated by a distance d , within a uniform electric field $\mathbf{E}$, like the one between two oppositely charged (large) plates, is equal to

$$\Delta V = -Ed$$

which is most of the time written as

$$V = Ed \quad (2.11)$$

with the understanding that V here actually stands for the potential difference and that the positively charged plate is at a higher potential than the negatively charged plate which is kept at *zero* potential.

Before passing we note from (2.11) that another unit for electric field, which is widely used, is volts per meter:

$$1 \text{ V/m} \equiv 1 \text{ N/C}$$

If every points on a surface are all at the same potential, such a surface is called an *equipotential surface*. Three properties of them are important:

1. If a charge is moved on an equipotential surface by an electric field or by an external agent, no work is done on the charge. This property is a direct consequence of (2.8)–(2.10).
2. Electric field lines are always perpendicular to equipotential surfaces.
3. A conductor surface is always an equipotential surface, so that, to repeat, electric field lines are always perpendicular to conductor surfaces.

Taking the potential $V_f = 0$ at *infinity* in (2.10), we always have such a freedom, we find the potential of a point charge q at a radial point, which is a distance r from the charge itself, as

$$V = k \frac{q}{r} \quad (\text{w.r.t. } \infty) \quad (2.12)$$

The *potential energy* U of system of two charges q_1 and q_2 separated by a distance r_{12} with respect to infinity is calculated from

$$U = k \frac{q_1 q_2}{r_{12}} \quad (\text{w.r.t. } \infty) \quad (2.13)$$

which is equal also to the work W_a done by an external agent in bringing these charges from infinity to their final positions, as is seen from (2.8). If we are speaking of more than two charges, say four, the above equation is generalized into

$$U = W_a = k \left(\frac{q_1 q_2}{r_{12}} + \frac{q_1 q_3}{r_{13}} + \frac{q_1 q_4}{r_{14}} + \frac{q_2 q_3}{r_{23}} + \frac{q_2 q_4}{r_{24}} + \frac{q_3 q_4}{r_{34}} \right) \quad (\text{w.r.t. } \infty)$$

If the potential V at a point in space happens to be known, we can determine the electric field vector at that point. It follows from (2.10) that

$$\mathbf{E} = -\nabla V \quad (2.14)$$

which becomes for the usual rectangular Cartesian coordinates, for example,

$$\mathbf{E} = -\hat{\mathbf{x}} \frac{\partial V}{\partial x} - \hat{\mathbf{y}} \frac{\partial V}{\partial y} - \hat{\mathbf{z}} \frac{\partial V}{\partial z}$$

2.6 Current

Until so far we were interested in *electrostatics* in which the charges considered were not moving. What happens when charges are allowed to move? We start with the phenomenon of (*electric*) *current*. If a *net* amount of charge dq passes through an (imaginary) cross-sectional area within, say, a conducting wire in a time dt , we define the current mathematically as

$$i \equiv \frac{dq}{dt} \quad (\text{C/s}) \quad (2.15)$$

In other words, the current is the rate of flow at which charges pass *definitely* through a cross-sectional area.

The name of the unit of current is adopted as *ampère*, A:

$$1 \text{ A} \equiv 1 \text{ C/s}$$

Out of some historical reasons, the direction of current has been adopted that of *positive* charge carriers. But, later on it has been revealed that it was electrons' motion that constitutes current in solids. We now simply say that the direction of current is opposite that of electrons within the wire.

What is the relationship between the current and the motion of charges within a conductor? To answer this we think of the current in a uniform wire of cross-sectional area A , as shown in Fig. 2.1, where the shaded portion has a length Δx so that it has a volume $A \Delta x$. If we let n be the number of mobile charge carriers per volume; then, the total charge within this portion will be equal to $nA \Delta x$. Thus, in turn, the total charge

contained in the portion will be $\Delta q = n(qA \Delta x)$, where q is the charge of each carrier. If v_d is the speed of these charge carriers, we must have $v_d = \Delta x / \Delta t$, or $\Delta x = v_d \Delta t$, where Δt is the time required for the charge carriers to cover the distance Δx . With this we rewrite Δq as

$$\Delta q = nqv_d A \Delta t$$

But dividing both sides by Δt and using (2.15) we obtain the current through this portion of the wire as

$$i = nqv_d A \quad (2.16)$$

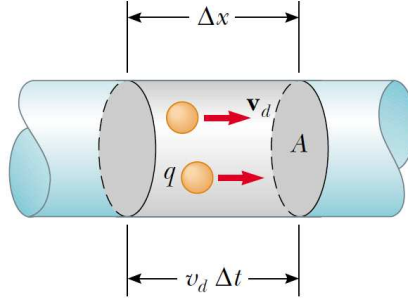


Figure 2.1: A portion of uniform wire; taken from Ref. [11].

In the above result v_d is called *drift velocity* of charge carriers. It is not equal to the “true” velocity of charge carriers; it is incomparably smaller. In order to understand this point, we consider again a portion of metallic conductor where the mobile charge carriers are electrons. If there is not an external voltage bias, these nearly free electrons go through completely random motion all around the conductor, with no net displacement on the average and there results no current at all. But if we apply an external voltage bias to the ends of the conductor, an electric field will be established instantly, because of which there will be electric force on the electrons. Under the effect of this force, electrons will (1) continue their random motion, and at the same time (2) “drift” in the direction opposite to that of electric field, as depicted in Fig. 2.2. Because of these, drift velocities, which are of the order of 10^{-5} or 10^{-4} m/s, are much smaller than the random-motion speeds of electrons in a metallic conductor, which are of the order of 10^6 m/s [12]. During their random motions, electrons made innumerable many collisions with atoms, so that there is a constant energy transfer

from electrons to atoms. This, in turn, increases the vibrational energy of atoms and, as a result, the temperature of the conductor increases, as well. This simple picture somehow, though inadequately, accounts for the fact that metals are not only good conductors of electric charge, but they are also good conductors of heat. (We shall expound this discussion shortly near the end of this chapter.)

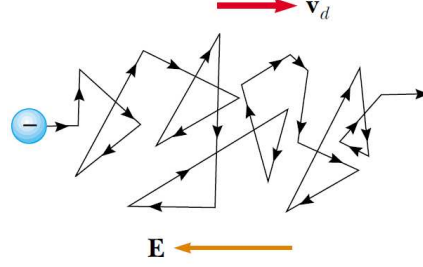


Figure 2.2: A symbolic depiction to represent the complicated zigzag motion of a negative charge carrier within a conductor; taken from Ref. [11].

2.7 Current Density

Consider a current-carrying conductor of any shape. We can relate current i , which is a *scalar* quantity, to *current density* $\mathbf{J}$, which is a *vector* quantity, through the relation

$$i = \int \mathbf{J} \cdot d\mathbf{A} \quad (2.17)$$

where $d\mathbf{A}$ directed differential, cross-sectional surface area element through which current might pass. The direction of $\mathbf{J}$ is adopted as the same as that of current (and that of the drift velocity of the charge carriers if they are positive, and, lastly, that of the electric field). As its name suggests, the current density equals, in simple speaking, current per area:

$$J = \frac{i}{A} \quad (\text{A/m}^2) \quad (2.18)$$

which is valid for a conductor through which a uniform current i flows across a cross-sectional surface area of A in perpendicular manner.

Using (2.16), we express current velocity in terms of drift velocity as

$$J = nqv_d \quad (2.19)$$

or, in vector form for positive charge carriers as

$$\mathbf{J} = nq\mathbf{v}_d \quad (2.20)$$

2.8 Resistance and Resistivity

Whenever a potential difference V is applied to the ends of a material, almost instantly a current i starts to flow. The *resistance* R of material is *defined* as

$$R = \frac{V}{i} \quad (\text{V/A}) \quad (2.21)$$

and it is a measure of difficulty shown by the material against the current flow: the larger the resistance of a material, the smaller the current throughout it.

The unit of resistance has, as usual, a special name, the “ohm,” Ω :

$$1 \text{ V/A} \equiv 1 \Omega$$

A *resistor* is a special conductor fabricated to have a particular resistance for use in a circuit.

We say that the resistance R in (2.21) is *extrinsic*, meaning that it is absolutely “tailorable” according to the special needs or uses. This equation has an *intrinsic* counterpart, as well:

$$\rho = \frac{E}{J} \quad (2.22)$$

where ρ is the so-called *resistivity* of the material under question. As the label “intrinsic” suggests, the resistivity of material, at constant temperature, is constant; it cannot be tailored out for special uses.

The unit of resistivity is “ohm-meter”, which can be seen readily as

$$\frac{\text{V/m}}{\text{A/m}^2} = \frac{\text{V}}{\text{m}} \frac{\text{m}^2}{\text{A}} = \frac{\text{V}}{\text{A}} \cdot \text{m} = \Omega \cdot \text{m}$$

The electric field and current density in (2.22) are actually two parallel vector quantities. Consequently, we can rewrite this equation in the most general vector form as

$$\mathbf{E} = \rho \mathbf{J} \quad (2.23)$$

which is another form of Ohm's law. It seems in order to point out here that the two preceding equations, (2.22) and (2.23), are valid for *isotropic* materials, which possess the same electrical properties in all directions.

The inverse of resistivity is referred to as *conductivity*:

$$\sigma = \frac{1}{\rho} \quad (\Omega \cdot \text{m})^{-1} \quad (2.24)$$

As its name implies, it might be seen as the measure of “ease” or “readiness” exhibited by a material to the flow of current throughout itself.

We can calculate the resistance of an object which is made from a material of known resistivity. To be specific, as shown in Fig. 2.3, we consider an object, of length L and cross-sectional area A , which carries a uniform longitudinal current i . An external voltage difference V is applied to the longitudinal ends of the object so that an electric field and, in turn, a current density, with respective magnitudes E and J , are already established between the same ends.

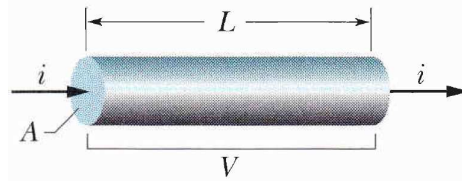


Figure 2.3: A sketch for the calculation of resistance; taken from Ref. [13].

The use of (2.11) for this setup gives $E = V/L$. Then with (2.18), $J = i/A$, we have

$$\rho = \frac{E}{J} = \frac{V/L}{i/A} = \frac{V A}{i L} = R \frac{A}{L}$$

where we have used the definition of resistance (2.21). (As a check we see directly from this expression that the unit of resistivity is $\Omega \cdot \text{m}$, indeed.) By recasting this result, we obtain the desired equation for the resistance as

$$R = \rho \frac{L}{A} \quad (2.25)$$

This equation says that you can make any resistor with a desired value of R (*extrinsic* property) out of a material with a fixed resistivity ρ (*intrinsic* property) (for the use of some definite temperature interval).

2.9 Variation of Resistivity with Temperature

In order for the discussion to be complete (and for our later studies), we will give here the dependence of resistivity on temperature. We shall see shortly in one of the following sections that the resistance of a material has a strong dependence upon the collisions of the charge carriers with the constituent atoms of the material. Therefore, it is not unexpected that resistance increases with increasing temperature, since the higher temperature, the more (and frequent) collisions.

It is an experimental finding that the variation of resistivity $\Delta\rho$, for especially metals, is almost linear with temperature change, ΔT , over a wide range of temperature values. We write this mathematically as

$$\frac{\Delta\rho}{\rho_0} = \alpha\Delta T \quad (2.26)$$

which is most of the time written in the form

$$\rho - \rho_0 = \rho_0\alpha(T - T_0) \quad (2.27)$$

where T_0 is a specific temperature and ρ_0 is the resistivity of the metallic material at that specific temperature. For practical purposes, T_0 is usually taken to be the room temperature, 293 K.

2.10 Ohm's Law

The expression in (2.21) is actually the characterizing equation for the famous Ohm's law. Rewriting it as $i = V/R$, we speak of an *ohmic device* provided that the current i through it is directly proportional to the external potential difference V applied to the ends of it (see Fig.2.4). This proportionality does not depend on the magnitude or polarity of V . (This means, of course, in intrinsic terms, that a metallic material is

ohmic if its resistivity does not depend on the magnitude and direction of the electric field established within it.)

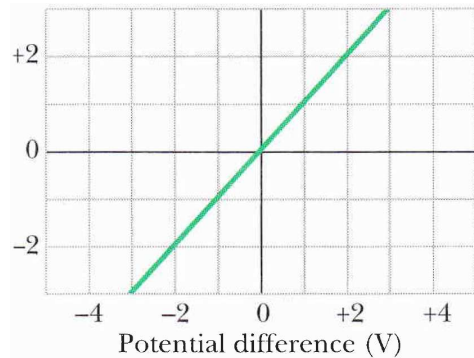


Figure 2.4: The voltage V versus current i graph for an ohmic device; taken from Ref. [14].

2.11 A Rudimentary Microscopic Account for Ohm's Law [15]

Here we depict a model for the (electrical) conduction in metallic conductors. We shall see that this model naturally results in Ohm's law and displays beautifully that resistivity is associated with the motion of electrons in metallic conductors.

We can think of a metallic conductor (or, any bulk material with a well-defined crystal structure) as a highly regular arrangements of atoms which are connected to each other with very stiff springs (so stiff that atoms cannot wander freely, they can only vibrate around themselves); Fig. 2.5 depicts such a mechanical model. Among the atoms, and confined by the edges of the metal, are innumerable many nearly free (conduction) electrons; their number is roughly of the order of that of atoms. They are practically almost free to move around “fixed” atoms, throughout the metal. If there is no electric field, these free electrons undergo motion in completely random directions; their average speeds are very large, in the order of 10^6 m/s, as we mentioned before. (It is possible only with quantum mechanical methods to estimate the speeds of free electrons correctly.) This motion of electrons is analogous to that of gas molecules within a container, so that these free conduction electrons are usually referred to as an *electron gas*. At every moment the conduction electrons undergo repeated ceaseless

collisions with atoms, which is the main mechanism for the resistivity of any metallic conductor at moderate (normal) temperatures. We must indicate here that if there exists no electric field within the metal, there is not any current through it, because the *average velocity* of the wandering free electrons (*not* their average speed) is zero, on the average. This means that the average number of electrons going to some direction at any time is, for the most part, just equal to that of electrons going exactly to the opposite direction; therefore, the net electric charge flow, an indispensable condition for a current to be established, is zero.

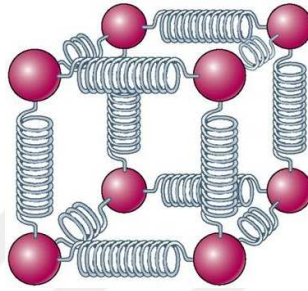


Figure 2.5: A mechanical model for a crystalline bulk material; taken from Ref. [16].

Whenever there *does* exist an electric field through the metal, perhaps because of a voltage difference applied to it externally, this situation changes. Free electrons still have the same delirious random motion we described above. In addition, they “drift” very slowly in the opposite direction of the electric field. As we said above, the average speed v_d of free electrons in a metal are of the order of 10^{-4} m/s, a value far smaller than the average speed between any two collisions a free electron makes with atoms.

The following three figures describe roughly the situations without and with an applied electric field $\mathbf{E}$. If $\mathbf{E} = 0$, the net displacement of a free electron after many collisions with atoms is zero (Fig. 2.6). But if $\mathbf{E} \neq 0$, the free electrons are constantly forced to move in the opposite direction of $\mathbf{E}$, by the electric forces $\mathbf{F} = -e\mathbf{E}$, with $-e$ being the elementary charge of an electron. The net effect is that the free electrons are now caused to “drift” through the metal very slowly, as depicted in Fig. 2.7. A mechanical analogue for the electrical conduction, in the presence of a non-zero electric field, in a metal is illustrated in Fig. 2.8. In this depiction the inclined plane repre-

sents the voltage difference applied to the ends of the metal; the red ball represents a free electron which is rolling down under the effect of the electric force, which is represented, in turn, by that component of the gravitational along the inclined plane; and the pegs represent both the nearly fixed atoms and the opposition shown against the free electron to roll down.

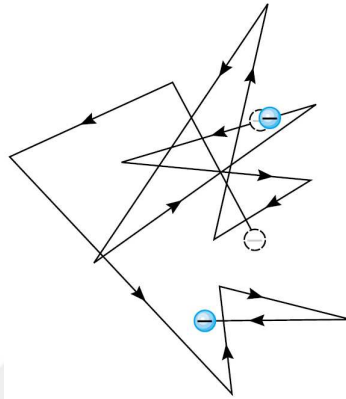


Figure 2.6: A diagram for the rapid motion of two representative free electrons in a metal for the zero electric field case in which the drift velocity of the electrons is zero; taken from Ref. [17].

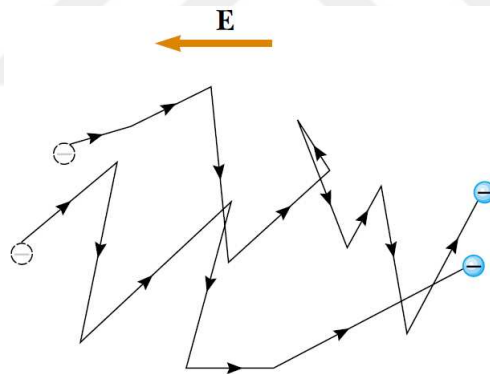


Figure 2.7: A diagram similar to Fig. 2.6 for the case in which there is now an electric field; in this case free electrons have non-zero drift velocities; taken from Ref. [17].

This model assumes that the extra energy gained by the electrons within the electric field is spent up through the collision processes. This energy is acquired by atoms and their vibrational motion about themselves increases. This means that, as a result, the temperature of the metal increases, as well. We also assume in this model that the path of a free electron after a collision is absolutely random, it cannot be predicted from the information before the collision.

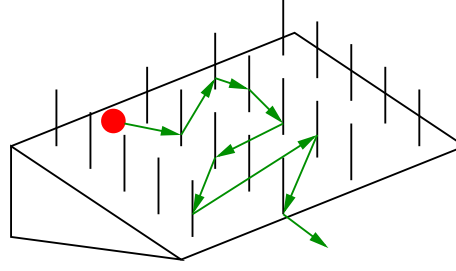


Figure 2.8: A mechanical analogue for the electrical conduction in a metal; adapted from Ref. [17].

We now consider a general charge carrier of mass m and charge q . If it is within an electric field $\mathbf{E}$, there will be an electric force acting on it, which is given by $\mathbf{F} = q\mathbf{E} = m\mathbf{a}$, so that it will have an acceleration

$$\mathbf{a} = \frac{q\mathbf{E}}{m} \quad (2.28)$$

It is this acceleration that causes free electrons to gain a *small* drift velocity, because it is at work only for a very short time interval between frequent successive collisions with atoms. Let t be the time elapsed since the last collision and $\mathbf{v}_0$ be the initial velocity of the charge carrier; then its velocity at later time t will be

$$\mathbf{v} = \mathbf{v}_0 + \mathbf{a}t = \mathbf{v}_0 + \frac{q\mathbf{E}}{m} t \quad (2.29)$$

The time average of the initial velocity $\mathbf{v}_0$ is zero, because they are supposed to be distributed absolutely randomly in all directions. The second term $\frac{q\mathbf{E}}{m} t$ is the small velocity that is added to the initial velocity of the charge carrier at the end of its journey from one atom to another. If τ is the *average time* between any two collisions and if the charge carrier starts with zero initial velocity, then the average of this second term will amount to its drift velocity in some definite direction (i.e., in the direction of the electric field $\mathbf{E}$):

$$\mathbf{v}_d = \frac{q\mathbf{E}}{m} \tau \quad (2.30)$$

Feeding this into (2.19), the magnitude of the current density $\mathbf{J}$ is found as

$$J = nqv_d = \frac{nq^2E}{m} \tau \quad (2.31)$$

Here the average time τ between collisions is associated with the average thermal speed v_{avg} of charge carriers and the average distance ℓ between collisions (which is actually referred to as *the mean free path*):

$$\tau = \frac{\ell}{v_{\text{avg}}} \quad (2.32)$$

Let us compare the result in (2.31), in magnitude, with (2.23), the second form of Ohm's law, $E = \rho J$. We see that the resistivity of a material is expressed as

$$\rho = \frac{m}{nq^2\tau} \quad (2.33)$$

or, via (2.24), its conductivity is given by

$$\sigma = \frac{1}{\rho} = \frac{nq^2\tau}{m} \quad (2.34)$$

We say after these results that the resistivity (and the conductivity) is independent of the applied electric field. This is a distinguishing character of an ohmic metallic conductor. According to this (classical) model, the resistivity can be readily obtained if we know the number of mobile charge carriers per volume, their mass and charge, and finally the average time between electron-atom collisions.

As we have seen in the preceding paragraphs, this classical model, commensurate with its assumptions, yields automatically Ohm's law (2.23). However it cannot account for some other well known phenomena. For instance, this model makes use of the *ideal-gas approximation* in calculating the average thermal speed v_{avg} of charge carriers and its results are roughly 10 times smaller than the true ones. Also, our present model predicts that the resistivity ρ should be proportional to v_{avg} , as is readily seen from (2.32) and (2.33), which in turn is directly proportional to $\sqrt{T}$, as stated by the ideal-gas approximation. But this is in complete contrast to what we have learnt from (2.27); experiments say that the resistivity ρ for pure metals should have a linear dependence with temperature T .

The correct explanations for these and similar shortcomings of classical mechanics, we must resort to quantum mechanics. It says that electrons are inherited with wave-like characteristics. If the atoms of a metallic conductor are perfectly *periodic* (i.e., if they are ordered regularly in space), the wavelike property of free electrons allows them to wander freely throughout the metal. As a result, it is a highly unlikely event that a free electron will collide with a (fixed) atom. In other words, if we are to speak for an ideal pure metal, there must be no collision between free electrons and atoms; the mean free path must then be infinite. The result in this case would be a pure metal with zero resistivity. But these are all for idealized pure metals. In reality they contain, although in very small quantity, structural impurities and/or defects so that their arrays

of atoms are not perfectly regular or periodic. These kinds of irregularities / impurities can readily scatter electron “waves.” There are two thermal extrema for the resistivity of metals: (1) at low temperatures, resistivity results primarily from the scatterings due to the collisions between the free electrons and the impurities contained in the metal; while (2) at high temperatures, resistivity is mainly because of the collisions that take place between, again, the free electrons and the metal’s atoms which ceaselessly vibrate owing to *thermal agitation*. At high temperatures, the structure of the metal is quite irregular if we compare it with that of the metal at low temperatures and this situation is a result of thermal motion of the atoms. This irregularity decreases the mean free path of the electron, hence the resistivity of the metal increases, as is neatly expressed in (2.32) and (2.33).

2.12 A Charged Particle in a Magnetic Field

We now know what happens if a particle of charge q is within an electric field $\mathbf{E}$. According to (2.2), which repeat here as

$$\mathbf{F}_E = q\mathbf{E}, \quad (2.35)$$

there will be an electric force $\mathbf{F}_E$ acting on the particle which is in the same (opposite) direction of $\mathbf{E}$ if its charge is positive (negative), and it does not matter that the particle might be at rest or not. This equation is sometimes described as being *operational* in the sense that if we want to confirm the presence of an electric field in some part of space, we just place a charged particle there and check that if there is a force acting on it. Further, measuring the magnitude of this force and determining its direction, we can readily determine the magnitude and direction of the electric field in that part of space by making use of (2.35).

What about the same charged particle within a magnetic field $\mathbf{B}$? To answer this question we again put the particle within the magnetic field and detect and measure the force, if any, on it. The result of such an experiment can be compactly expressed as

$$\mathbf{F}_B = q\mathbf{v} \times \mathbf{B} \quad (2.36)$$

which is the *operational* equation for the magnetic force on a particle of charge q which moves within a magnetic field $\mathbf{B}$ with a velocity $\mathbf{v}$. In terms of magnitude, we

write the above expression as

$$F_B = qvB \sin \theta \quad (2.37)$$

where θ is the smaller angle between the direction of the velocity and that of the magnetic field vectors.

We observe that:

1. If the charged particle is at rest, $v = 0$, there is no magnetic force acting on it, $F_B = 0$.
2. If the charged particle moves parallel and anti-parallel to the magnetic field, so that $\theta = 0$ or $\theta = 180^\circ$, then the magnetic force acting on it is zero, $F_B = 0$.
3. Provided that $v \neq 0$ and $\theta \neq 0, 180^\circ$, the magnitude of the magnetic force on the charged particle is directly proportional its charge q and speed v and the absolute value of the sine of the angle between $\mathbf{v}$ and $\mathbf{B}$.
4. The magnitude of the magnetic force on the charged particle is maximum, $F_B = qvB$, when it moves perpendicular to the magnetic field, i.e, when $\theta = 90^\circ$.
5. The direction of the magnetic force $\mathbf{F}_B$ acting on a *positive* charge is perpendicular to the plane formed by $\mathbf{v}$ and $\mathbf{B}$, and is found by the application of the usual *right-hand* rule: with our right hand, we imagine sweeping $\mathbf{v}$ into $\mathbf{B}$ through the smaller angle between them; the outstretched thumb then shows the direction of $\mathbf{F}_B$. This direction is just reversed for the magnetic force acting on a *negative* charge.

The SI unit of magnetic field is Wb/m^2 (weber per square meter), which is called *tesla*, T:

$$1 \text{ T} = 1 \frac{\text{Wb}}{\text{m}^2} = 1 \frac{\text{N}}{\text{C} \cdot \text{m/s}} = \frac{\text{N}}{\text{A} \cdot \text{m}}$$

Because *tesla* is an unusually large unit, the cgs unit *gauss* G, defined as

$$1 \text{ T} = 10^4 \text{ G},$$

is also frequently used.

Using (2.35) and (2.36), we can differentiate the operational effects of electric and magnetic fields on a charged particle placed within them:

1. While the electric force is always parallel or anti-parallel to the electric field, the magnetic force is always perpendicular to the magnetic field.
2. For a non-zero magnetic force, the charge must not be at rest; it must be in motion within the magnetic field. For the electric force there is no such condition, it always acts on the charge, no matter it is at rest or not.
3. If the charged particle is not fixed in place, the electric force will move it and therefore *does* work on it. On the other hand, the magnetic force due to a constant magnetic field does *no* work when the charged particle undergoes a net displacement. Thus, according to the work-energy theorem, $\Delta K = W$, we conclude that a magnetic field alone *cannot* change the kinetic energy, and, in turn, the velocity of a charged particle.

We can prove the last fact by calculating the differential work done on the particle when it undergoes a displacement $d\mathbf{r}$:

$$dW = \mathbf{F}_B \cdot d\mathbf{r} = \left(\mathbf{F}_B \cdot \frac{d\mathbf{r}}{dt} \right) dt = (\mathbf{F}_B \cdot \mathbf{v}) dt = 0$$

where at the last part we have used the fact that the magnetic force is perpendicular to the velocity vector.

2.12.1 Lorentz Force

Now we can consider the most general situation: we have a particle of charge q in some part of space where there exist both an electric, $\mathbf{E}$, and a magnetic field, $\mathbf{B}$. The charged particle will be affected by both the electric and magnetic forces. Because they are *additive*, so that the *superposition principle* is at work, the net *electromagnetic* force on the charged particle is the vector sum of (2.35) and (2.36):

$$\mathbf{F} = q\mathbf{E} + q\mathbf{v} \times \mathbf{B} \quad (2.38)$$

which is referred to as the *Lorentz force*. We shall later on make use of this equation when we try to derive the quantum mechanical Hamiltonian of a charged particle moving in an electro-magnetic field.

2.12.2 Magnetic Flux

When we deal with Aharonov-Bohm rings, one of the main subjects of this manuscript, threaded by a magnetic field (*time-independent* in this present M.S. study and *time-periodic* in a possible future Ph.D. study as a continuation of the former), we shall need the concept of *magnetic flux*. Just like its electric counterpart, it is given by

$$\Phi_B = \int \mathbf{B} \cdot d\mathbf{A} \quad (2.39)$$

It is a measure of the amount of magnetic field lines piercing through a particular surface in space. In the above equation $d\mathbf{A}$ is the directed differential area element on the surface, $\mathbf{B}$ is the magnetic field at that element.

The unit of magnetic flux is $\text{T} \cdot \text{m}^2$, but *weber* (Wb) is used more frequently:

$$1 \text{ T} \cdot \text{m}^2 = 1 \text{ Wb}$$

CHAPTER 3

THE WORKING MACHINERY OF QUANTUM MECHANICS

We shall provide in this chapter the basic formalism of quantum mechanics. By this we would have settled down the nomenclature that shall be used throughout the manuscript. Our treatment will not be mathematically rigorous and far from complete; our primary aim is to pave the way for the *(nearly) free electron* concept that is central to the present work. In the following, the vocabulary of quantum mechanics, which might appear new for the novice reader, will be written in *italic* form. Also, we have made **colorful** the most important results. The material here follows closely the lines the third chapter of Liboff's invaluable gigantic masterpiece [18]. We have cited nonetheless some additional sources whenever it seems necessary and appropriate.

In this chapter there is more stuff than that is needed. Especially the last section, titled "Quantum Dynamics" is added to the manuscript not only for the sake of completeness and integrity. We have a strong desire to continue the present M.S. level study in a possibly (and hopefully) subsequent Ph.D. study where we are planning to work with one-dimensional systems with some time-periodic potential excitations. Because of this future intention it seemed reasonable to include into the manuscript a section to deal with the time-evolution of quantum mechanical systems.

3.1 Postulate of Quantum Mechanics

3.1.1 Postulate I

For every physical *observable* (or, *measurable*) quantity A , such as linear momentum, energy, or position, there is a corresponding *operator* $\hat{A}$ in such a way that the measurement of A results in values a , which are mathematically the *eigenvalues* of $\hat{A}$. In other words, $\hat{A}$ and a satisfy the so-called *eigenvalue equation*

$$\hat{A}\varphi = a\varphi \quad (3.1)$$

where φ , referred to as the *eigenfunction* of $\hat{A}$, is the solution of the above equation.

3.1.1.1 The Linear Momentum Operator

The operator for the most general three-dimensional (3D) *linear momentum* observable is

$$\hat{\mathbf{p}} = -i\hbar\nabla \quad (3.2)$$

where

$$\hbar = \frac{h}{2\pi} \quad (3.3)$$

and ∇ is the usual gradient operator:

$$\nabla = \frac{\partial}{\partial x} \hat{\mathbf{x}} + \frac{\partial}{\partial y} \hat{\mathbf{y}} + \frac{\partial}{\partial z} \hat{\mathbf{z}}$$

For a single *free particle* confined to move along, say, an x -direction, (3.2) reduces to the 1D linear momentum operator

$$\hat{p}_x = -i\hbar \frac{\partial}{\partial x} \quad (3.4)$$

for which the eigenvalue equation (3.1) becomes

$$-i\hbar \frac{\partial \varphi}{\partial x} = p_x \varphi \quad (3.5)$$

Here the eigenvalues p_x are the likely values we might possibly obtain if we measure the x -component of the linear momentum of the particle. For the free particle at hand there is no boundary condition so that the solution to (3.5) is seen to be

$$\varphi(x) = A \exp\left(\frac{ip_x x}{\hbar}\right) \quad (3.6)$$

which is the eigenfunction that corresponds to a definite linear momentum p_x . We elucidate the very meaning of the eigenfunction $\varphi(x)$ in accord with the 1927 *Copenhagen interperation* [19] (or, sometimes the *Born postulate*):

The quantity $|\varphi(x)|^2 dx = \varphi(x)^* \varphi(x) dx$ gives the probability that the free particle with a specific linear momentum p_x is found in the interval $\{x, x + dx\}$.

If we delete the subscript x in p_x and let k denote the *wavenumber* as

$$k = \frac{p}{\hbar} \quad (3.7)$$

we rewrite (3.6) in a more compact way:

$$\varphi(x) = A e^{ikx} \quad (3.8)$$

Because of its particular mathematical form, this eigenfunction is periodic in x . If its *wavelength* is λ , it must be correct that

$$\begin{aligned} e^{ikx} &= e^{ik(x+\lambda)} = e^{ikx} e^{ik\lambda} \\ 1 &= \cos k\lambda + i \sin k\lambda \end{aligned}$$

which is equivalent to

$$\cos k\lambda = 1 \quad \text{and} \quad \sin k\lambda = 0$$

The (first) nontrivial solution to these is

$$k\lambda = 2\pi$$

which, with (3.3) and (3.7), amounts to the famous *de Broglie relation*

$$p = \frac{h}{\lambda} \quad (3.9)$$

Put in other words, the linear momentum operator for a definite eigenvalue p has an eigenfunction that has the de Broglie wavelenght h/p .

It is somehow cumbersome to work, in quantum mechanics, in terms of linear momentum p ; it is a common practise to study with wavenumber k , instead. Consequently, we will write from now on the eigenfunctions and eigenvalues of the linear momentum operator, via (3.7), as

$$\varphi_k(x) = Ae^{ikx}, \quad p = \hbar k \quad (3.10)$$

where the subscript k on φ_k designates the *continuum* of eigenvalues $\hbar k$.

3.1.1.2 The Energy Operator

In quantum mechanics the energy operator is taken from its classical counterpart; it is the *Hamiltonian* of the system under question. The only change is that we replace the linear momentum $\mathbf{p}$ with its operator form. If we have a single particle, of mass m moving within an environment described by a potential field $V(\mathbf{r})$, we write its Hamiltonian operator as

$$\hat{H} = \frac{\hat{p}^2}{2m} + V(\mathbf{r}) = -\frac{\hbar^2}{2m} \nabla^2 + V(\mathbf{r}) \quad (3.11)$$

where ∇^2 is the *Laplacian operator*

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

The eigenvalue equation (3.1) for $\hat{H}$ is written down as

$$\hat{H}\varphi(\mathbf{r}) = E\varphi(\mathbf{r}) \quad (3.12)$$

which is the celebrated *time-independent Schrödinger equation*. Here the energy eigenvalues E are the likely energies that a measurement made on the single particle may result in.

Let us return back to our 1D free particle problem in which there is no potential energy so that Hamiltonian (3.11) will be, with $\hat{p}_x = \hat{p}$,

$$\hat{H} = \frac{\hat{p}^2}{2m} = -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} \quad (3.13)$$

for which the time independent Schrödinger equation (3.12) takes the form

$$-\frac{\hbar^2}{2m} \frac{\partial^2 \varphi}{\partial x^2} = E\varphi \quad (3.14)$$

with the eigenfunction being a function of x only: $\varphi = \varphi(x)$. Using the abbreviation

$$k^2 = \frac{2mE}{\hbar^2} \quad (3.15)$$

we rewrite (3.14) as

$$\frac{\partial^2 \varphi}{\partial x^2} + k^2 \varphi = 0 \quad (3.16)$$

This is the well-known elementary wave equation. Since we do not have any boundary conditions for the free particle, the most general solution to (3.16) is

$$\varphi_k(x) = Ae^{ikx} + Be^{-ikx} \quad (3.17)$$

which is the eigenfunction of Hamiltonian (3.13) for the specific eigenvalue, via (3.15),

$$E_k = \frac{\hbar^2 k^2}{2m} \quad (3.18)$$

At this point we just look again at (3.10); we see that for $B = 0$, the eigenfunction (3.10) of Hamiltonian operator $\hat{H}$ is also that of the linear momentum operator $\hat{p}$. That is, for a free particle, $\hat{H}$ and $\hat{p}$ possess common eigenfunctions. To see this more clearly, we just first let

$$\hat{p} \varphi = \hbar k \varphi$$

Then, for the Hamiltonian operator (3.13) we have

$$\hat{H} \varphi = \frac{\hat{p}^2}{2m} \varphi = \frac{1}{2m} \hat{p} (\hat{p} \varphi) = \frac{1}{2m} \hat{p} \cdot \hbar k \varphi = \frac{\hbar k}{2m} \hat{p} \varphi = \frac{\hbar k}{2m} \cdot \hbar k \varphi = \frac{\hbar^2 k^2}{2m} \varphi$$

Thus, we conclude for a free particle that the eigenfunction

$$\varphi_k(x) = Ae^{ikx} \quad (3.19)$$

is common for both the Hamiltonian and linear momentum operators with the respective eigenvalues

$$E_k = \frac{\hbar^2 k^2}{2m} \quad \text{and} \quad p_k = \hbar k \quad (3.20)$$

both of which comprise *continuous*, not *discrete*, values. If the free particle is in the state (3.19) we know for certain that its energy and linear momentum measurements will result in $\hbar^2 k^2 / 2m$ and $p = \hbar k$, respectively.

What if we are to measure the position x of the free particle? Where is the particle found most probably? Again the Born postulate says that if our free particle is in the state $\varphi_k(x)$, then the *probability density* of the finding it in the interval $\{x, x + dx\}$ is

$$|\varphi_k|^2 = \varphi_k^* \varphi_k = A^2 = \text{const.}$$

which means that the free particle might be found at *any* point from $-\infty$ to $+\infty$ with equal probability. We note that this is expected according to the *Heisenberg uncertainty principle*

$$\Delta x \Delta p \geq \hbar/2$$

If the free particle is in the state $\varphi_k(x)$, its linear momentum is $\hbar k$ for sure without any uncertainty, $\Delta p = 0$, which, or course, means a maximum uncertainty in its position, $\Delta x = \infty$.

3.1.2 Postulate II

If we measure a physical observable A on a system and if we obtain as a result the specific value a , this measurement process leaves the system in the definite state φ_a , which is the eigenfunction of $\hat{A}$ for the specific eigenvalue a .

We consider, for example, a 1D free particle. At the outset we do not know in which state it is. At some time we measure its linear momentum and find the specific value $p = \hbar k$. We say that this measurement process *collapses* the system into the definite state φ_k and *leaves* it there, so that an immediate successive measurement of the linear momentum results again in $p = \hbar k$ for sure.

3.1.2.1 Position Measurement

Assume that we are to measure the position of our 1D free particle; this means, according to Postulate I, that we have an operator $\hat{x}$ for the position measurement. Assume also that, as the result of the measurement, we see that the particle is at position $x = x'$. Postulate II says that this measurement leaves the particle at this very position $x = x'$ (so that an immediate successive measurement of the position finds the particle still

at $x = x'$), in the eigenfunction of $\hat{x}$ for the eigenvalue x' :

$$\hat{x} \delta(x - x') = x' \delta(x - x') \quad (3.21)$$

where $\delta(x - x')$ is the *Dirac delta function*. In other words, the Dirac delta function is the eigenfunction of the position operator $\hat{x}$.

The most important two properties of the Dirac delta function are

$$\begin{aligned} \int_{-\infty}^{\infty} \delta(x - x') dx &= 1 \\ \int_{-\infty}^{\infty} f(x) \delta(x - x') dx &= f(x') \end{aligned} \quad (3.22)$$

where the last one emphasizes the *sifting property* of the Dirac delta function with which the eigenvalue equation (3.21) for the position measurement assumes its very meaning.

Before passing we should be aware of the fact that when the position operator $\hat{x}$ operates from the left on a function $f(x)$ merely produces multiplication by x :

$$\hat{x} f(x) = x f(x) \quad (3.23)$$

3.1.3 Postulate III

The state of any system at any time might be represented by a *wave function* (or, *state function*) ψ which is both (1) continuous and single-valued and (2) differentiable everywhere. We say that all the information pertaining to the system at that time is embodied within this wave function ψ and can be readily extracted at will.

Here the reason for the compulsory differentiability property of the wave function comes only from momentum considerations, because momentum operators are just first-order partial derivatives [20], as the one in (3.4).

Consider a general system (say, a system of a single particle) which is in a state $\psi(\mathbf{r}, t)$. As we designated it before, the expression $|\psi|^2$ is called *probability density*; it is always positive and real. If

$$\int |\psi|^2 d\mathbf{r} = 0$$

we say that the particle is *nowhere*, or that it does *not* exist at time t . In the above expression $d\mathbf{r}$ designates the differential volume element for a general 3D system and the integral is taken over all space.

If the free particle is to be found somewhere, it must be correct at time t that

$$\int |\psi|^2 d\mathbf{r} = 1 \quad (3.24)$$

Any wave function with this property is referred to as being *normalized* or *normalizable*. It is more usual to interpret $|\psi|^2 d\mathbf{r}$ in (3.24) as the probability of finding the particle in a differential volume $d\mathbf{r}$ located at $\mathbf{r}$ at time t . For 1D problem (3.24) reduces to

$$\int_{-\infty}^{\infty} |\psi|^2 dx = 1 \quad (3.25)$$

Therefore, if we have a normalized 1D wave function for a single particle confined to move in the x -direction, the probability to detect it in the interval between x_1 and x_2 at time t is evaluated via

$$P(x_1, x_2) = \int_{x_1}^{x_2} |\psi|^2 dx \quad (3.26)$$

The normalizability condition (3.24) or (3.25) cannot always be observed by wave functions. For example, the eigenfunction of the 1D linear momentum operator, given in (3.10), cannot be normalized. In the next chapter we shall say more on this issue.

Consider, again, a system which is in a state $\psi(\mathbf{r}, t)$. The *average* (or, *mean*) of any physically measurable quantity C related to the system at any time t is calculated through

$$\langle C \rangle = \int \psi^* \hat{C} \psi d\mathbf{r} \quad (3.27)$$

where $d\mathbf{r}$ designates the differential volume element for a general 3D system. We call this average, $\langle C \rangle$, as the *expectation value* of observable C .

Before passing we must emphasize that the wave function ψ of system at any time is not necessarily a *single* eigenfunction φ_a of any observable. Most of the time it is a linear combination these eigenfunctions, just like expressing a vector in terms of, say, the rectangular Cartesian unit vectors $(\hat{x}, \hat{y}, \hat{z})$:

$$\psi = \sum_n a_n \varphi_n$$

3.1.4 Postulate IV

The wave function $\psi(\mathbf{r}, t)$ of a system evolves in time according to the *time-dependent Schrödinger equation*

$$i\hbar \frac{\partial \psi}{\partial t} = \hat{H} \psi \quad (3.28)$$

where $\hat{H}$ is the Hamiltonian operator.

If $\hat{H}$ is independent of time,

$$\hat{H} = \hat{H}(\mathbf{r})$$

we can then find a general form for the solution to (3.28) by using the technique of *separation of variables*. We just assume that the solution is of the form

$$\psi(\mathbf{r}, t) = \varphi(\mathbf{r}) T(t) \quad (3.29)$$

Feeding this into (3.28) we obtain

$$i\hbar \varphi \frac{\partial T}{\partial t} = \hat{H} \varphi T = T \hat{H} \varphi$$

Dividing both sides by φT ,

$$i\hbar \frac{1}{T} \frac{\partial T}{\partial t} = \frac{1}{\varphi} \hat{H} \varphi \quad (3.30)$$

We see that the right-hand side of this equation is a function of $\mathbf{r}$ only while its left-hand side is that of t only. If this equation is to be satisfied, both its right- and left-hand sides must be equal to the same constant; we let it be E (at this point we do not know yet that it is the energy):

$$i\hbar \frac{1}{T} \frac{\partial T}{\partial t} = E = \frac{1}{\varphi} \hat{H} \varphi \quad (3.31)$$

whence

$$\hat{H}\varphi(\mathbf{r}) = E\varphi(\mathbf{r}) \quad (3.32)$$

$$\left(\frac{\partial}{\partial t} + \frac{iE}{\hbar}\right)T(t) = 0 \quad (3.33)$$

We immediately notice that the first equation is the time-independent Schrödinger equation (3.12) that we have seen before. This fact allows us to identify E in (3.31) as the energy of the system, indeed.

The solution to (3.33) is elementary; we have

$$T(t) = A \exp\left(-\frac{iEt}{\hbar}\right) \quad (3.34)$$

which has a oscillatory behaviour.

Let us assume that we have solved the time-independent Schrödinger equation, (3.12) or equivalently (3.32), for a particular somehow “confined” system whose energy eigenvalues are *discrete*, not *continuous*; we obtain the *spatial* eigenfunctions and eigenvalues

$$\hat{H}\varphi_n = E_n\varphi_n \quad n = 1, 2, 3, \dots \quad (3.35)$$

It is clear from (3.29) that for each spatial *eigensolution* for a specific value of n there corresponds an eigensolution to the time-dependent Schrödinger equation (3.28), with the *temporal* part of the form (3.34):

$$\varphi_n(\mathbf{r}, t) = A \varphi_n(\mathbf{r}) \exp\left(-\frac{iE_n t}{\hbar}\right) \quad (3.36)$$

Now let's get back to our main actor, the free particle. We have said that, because it is perfectly unconfined, it has a *continuum* of eigenfunctions $\varphi_k(x)$ and eigenvalues E_k . Here we repeat them again:

$$\varphi_k = Ae^{ikx}, \quad \hat{H}\varphi_k = E_k\varphi_k, \quad E_k = \frac{\hbar^2 k^2}{2m}$$

With this $\varphi_k(x)$, we write the solution to the time-dependent Schrödinger equation as

$$\psi_k(x, t) = A\varphi_k(x)e^{-iE_k t/\hbar} = Ae^{ikx}e^{-iE_k t/\hbar}$$

or, with

$$E_k = \hbar\omega \quad (3.37)$$

we have

$$\psi_k(x, t) = Ae^{i(kx - \omega t)} \quad (3.38)$$

This is one of our target results: the wave function for the free particle has the distinguishing 1D periodic *traveling* (or, *propagating*) wave character [21]. We know from any elementary physics course [22] that any function f of the form

$$f(x, t) = f(x \mp vt) \quad (3.39)$$

stands for a wave, real or imaginary, moving in the positive (negative) x -direction, with v being the speed of the propagation of the wave.

In order to obtain the “speed” of the wave propagation of the free particle wave function (3.38), we rewrite it as

$$\psi_k(x, t) = A \exp \left[ik \left(x - \frac{\omega}{k} t \right) \right] \quad (3.40)$$

If we compare this with (3.39), we conclude that $\psi_k(x, t)$ is a traveling wave moving to the right in the positive x -direction with the speed

$$v = \frac{\omega}{k} = \frac{\hbar\omega}{\hbar k} = \frac{E}{p} = \frac{p^2/2m}{p} = \frac{1}{2} \frac{p}{m} = \frac{1}{2} \frac{mv_{\text{CL}}}{m} = \frac{1}{2} v_{\text{CL}} \quad (3.41)$$

where v_{CL} is the classical speed of a particle with mass m and linear momentum p . Thus the conclusion to be drawn is that the “speed” of the wave propagation of the free particle wave function, with a specific momentum $p = \hbar k$, is equal to half the classical speed, $v_{\text{CL}} = p/m$.

At first glance the above result might seem unreasonable. Is there any discrepancy between the classical and quantum mechanical points of view? In order to clarify this issue, let us calculate the probability density for the state (3.38):

$$|\psi_k|^2 = \psi_k^* \psi_k = A^2 = \text{const.}$$

This means that we can find our free particle anywhere from $x = -\infty$ to $x = +\infty$ with equal probability. But this result opposes to the very idea of (idealized) particle of in classical mechanics. A better way to represent a *localized* particle is the use of a *wave packet* concept, as shown in Fig. 3.1. We may readily construct such a wave packet by the combination of sinusoidal waves; this is called *Fourier synthesis*, based on *Fourier series* concept [23]. Consequently, a wave packet like shown that in Fig. 3.1 can be

obtained by summing many eigenstates of the form (3.38). The velocity of such a wave packet is referred to as the *group velocity* and given by [24]

$$v_g = \frac{\partial \omega}{\partial k} \quad (3.42)$$

(In this respect, the velocity $v = \omega/k$ in (3.41) is called the *phase velocity* of an individual component of the wave packet [24].)

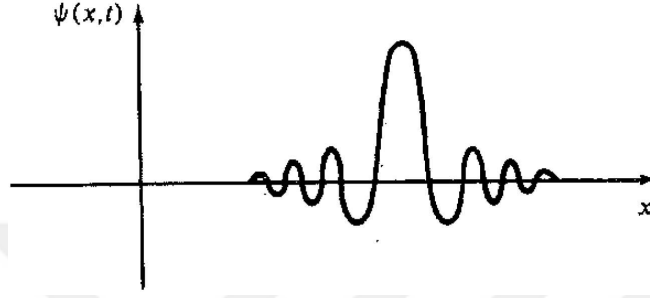


Figure 3.1: A wave packet at some time t ; taken from Ref. [25].

For the free particle problem at hand, the group velocity formula (3.42) leads to the correct classical velocity, as it should be:

$$v_g = \frac{\partial \omega}{\partial k} = \frac{\partial \hbar \omega}{\hbar \partial k} = \frac{1}{\hbar} \frac{\partial E}{\partial k} = \frac{1}{\hbar} \frac{\partial (\hbar^2 k^2 / 2m)}{\partial k} = \frac{\hbar k}{m} = \frac{p}{m} = v_{\text{CL}}$$

3.2 Quantum Dynamics

Suppose that we know $\psi(\mathbf{r}, 0)$, the initial value of a state function at $t = 0$, and we want to determine $\psi(\mathbf{r}, t)$, its value at a later time $t > 0$. Using the time-dependent Schrödinger equation (3.28) we can find the desired solution. To this end we first define the so-called *time-evolution operator* $\hat{U}$ as [26]

$$\hat{U} = \exp\left(-\frac{it\hat{H}}{\hbar}\right) \quad (3.43)$$

with its inverse

$$\hat{U}^{-1} = \exp\left(\frac{it\hat{H}}{\hbar}\right) \quad (3.44)$$

It is clear that

$$\hat{U}\hat{U}^{-1} = \exp\left(-\frac{it\hat{H}}{\hbar}\right)\exp\left(\frac{it\hat{H}}{\hbar}\right) = \hat{I} \quad (3.45)$$

where $\hat{I}$ is the *identity operator*.

The time-evolution operator $\hat{U}$ and its inverse $\hat{U}^{-1}$ are the function of another operator, the Hamiltonian operator $\hat{H}$. We define such operators in terms of a *Taylor series expansion*. As an example, the full form of (3.44) is

$$\hat{U}^{-1} = \exp\left(\frac{it\hat{H}}{\hbar}\right) = 1 + \frac{it\hat{H}}{\hbar} + \frac{1}{2!}\left(\frac{it\hat{H}}{\hbar}\right)^2 + \dots \quad (3.46)$$

Now we can go back to the solution of our initial value problem at hand. We first rewrite the time-dependent Schrödinger equation (3.28) as

$$\frac{\partial}{\partial t} \psi(\mathbf{r}, t) + \frac{i\hat{H}}{\hbar} \psi(\mathbf{r}, t) = 0 \quad (3.47)$$

An attentive eye can readily see that the inverse time-evolution operator (3.44) is also the *integrating factor* for the above differential equation. We then multiply (3.47) by (3.44) on the left and obtain

$$\frac{\partial}{\partial t} \left[\exp\left(\frac{it\hat{H}}{\hbar}\right) \psi(\mathbf{r}, t) \right] = 0$$

Time-integration from 0 to t yields

$$\exp\left(\frac{it\hat{H}}{\hbar}\right) \psi(\mathbf{r}, t) - \psi(\mathbf{r}, 0) = 0$$

As the last step, we multiply this on the left by (3.43) and use (3.45). We have

$$\psi(\mathbf{r}, t) = \exp\left(\frac{-it\hat{H}}{\hbar}\right)\psi(\mathbf{r}, 0) \quad (3.48)$$

This is what we have set out to do.

As an example we suppose that the initial state is a specific eigenstate of the Hamiltonian operator $\hat{H}$. Let it be φ_n . That is,

$$\hat{H}\varphi_n = E_n\varphi_n \quad \text{and} \quad \psi(\mathbf{r}, 0) = \varphi_n(\mathbf{r})$$

For this the recipe (3.48) yields for a later time that

$$\psi(\mathbf{r}, t) = \exp\left(\frac{-it\hat{H}}{\hbar}\right)\varphi_n(\mathbf{r}) = \exp\left(\frac{-itE_n}{\hbar}\right)\varphi_n(\mathbf{r}) = e^{-i\omega_n t}\varphi_n(\mathbf{r})$$

where we have let in the last part

$$E_n = \hbar\omega_n$$

This result is the solution to the time-dependent Schrödinger equation for an initial state. Previously we had obtained the same solution, (3.36), using the separation of variables technique. Comparison of (3.36) with (3.48) reveals that the latter is the most general.

CHAPTER 4

SCATTERING PROBLEMS IN ONE DIMENSION

In this chapter we shall deal with 1D scattering problems. We will first discuss properties of the 1D Schrödinger equation in general terms. Then there will be a discussion about *bound* and *unbound* states. This chapter will continue with the *continuity equation* and with its meaning. A detailed section for the *transmission* and *reflection coefficients* will precede two handy worked examples to demonstrate how quantum mechanics finds exquisitely its way in 1D scattering problems.

4.1 General Features of the 1D Schrödinger Equation

We consider a particle of mass m that is confined to move in one dimension, say, in the x -axis, under the effect of a position-dependent potential (energy) field $V(x)$. The time-independent Schrödinger equation (3.12) reads

$$\left[-\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + V(x) \right] \varphi(x) = E \varphi(x) \quad (4.1)$$

where E is the total energy of the particle,

$$E = T + V \quad (4.2)$$

We rewrite (4.1) in the form

$$\frac{\partial^2 \varphi}{\partial x^2} = -\frac{2m}{\hbar^2} (E - V) \varphi \quad (4.3)$$

Here there are two cases that must be treated separately. If $E > V$, as in domain I of Fig. 4.1, the particle is said to be in the classically *allowed region*; its kinetic energy T is positive and we can identify it as $T = E - V = \hbar^2 k^2 / 2m$. With this identification,

the differential equation (4.3) and its most basic solution form, apart from a constant, are written as

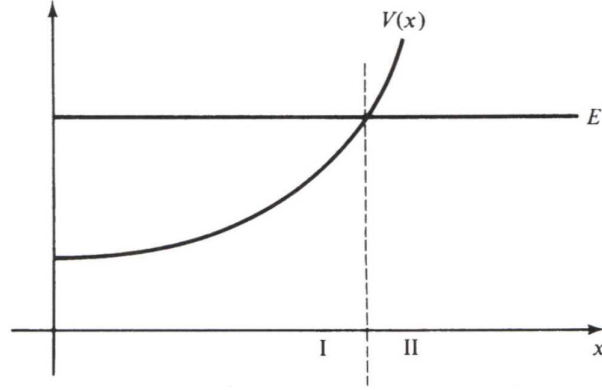


Figure 4.1: Two distinct domains of a particle that is in motion in a 1D potential (energy) field $V(x)$. Domain I is classically *allowed region*, while domain II is classically *forbidden region*. The dashed border passes through the so-called classical *turning point*, where $T = V$; taken from Ref. [27].

$$\begin{aligned}\frac{\hbar^2 k^2}{2m} &= E - V > 0 \\ \frac{\partial^2 \varphi}{\partial x^2} &= -k^2(x) \varphi \\ \varphi(x) &\sim e^{\pm i k x}\end{aligned}\tag{4.4}$$

If $V > E$, as in domain II of Fig. 4.1, the particle is now in the classically *forbidden region*; its kinetic energy T is negative: $T = E - V = -\hbar^2 \kappa^2 / 2m$. For this case, the differential equation (4.3) and its basic solution might be written in the form

$$\begin{aligned}\frac{\hbar^2 \kappa^2}{2m} &= V - E > 0 \\ \frac{\partial^2 \varphi}{\partial x^2} &= \kappa^2(x) \varphi \\ \varphi(x) &\sim e^{\pm \kappa x}\end{aligned}\tag{4.5}$$

It is clear that when the particle is in the classically allowed region, where $E > V$, we have an *oscillatory* solution ($\sim e^{\pm i k x}$), whereas when it is in the classically forbidden region, where $V > E$, we have a *decaying* ($\sim e^{-\kappa x}$), or *growing* ($\sim e^{+\kappa x}$) solution. It is also seen that when $E = V$, the particle is at a classical *turning point*, for which (4.3)

reduces to

$$E = V$$

$$\frac{\partial^2 \varphi}{\partial x^2} = 0 \quad (4.6)$$

which corresponds to $\partial\varphi/\partial x = 0$, that means, in turn, that we have a solution $\varphi(x)$ with a constant slope at that turning point. With all these, the (eigen)solution to the Hamiltonian for the energy profile given in Fig. 4.1 can be depicted as the one in Fig. 4.2.

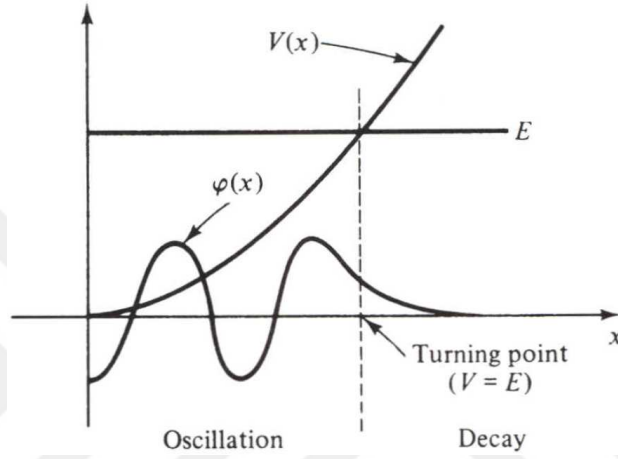


Figure 4.2: A representative wave function for the energy profile shown in Fig. 4.1; taken from Ref. [28].

4.2 Bound and Unbound States

If a wave function ψ of, say, a 1D system is *normalizable*, we say that it represents a *bound state*. In mathematical terms, a bound state ψ obeys the condition

$$|\psi|^2 \rightarrow 0 \quad \text{for} \quad |x| \rightarrow \infty \quad (\text{for all time } t) \quad (4.7)$$

Conversely, if the wave function does not fulfil this requirement, so that we cannot normalize it, it is said to represent an *unbound state*.

The integral of the *norm* of a bound state over the infinite interval results, therefore, in a finite number

$$\int_{-\infty}^{\infty} |\psi|^2 dx < \infty \quad (4.8)$$

and, on the other hand, that of an unbound state gives a finite number *only* over a finite interval

$$\int_a^b |\psi|^2 dx < \infty \quad \text{for} \quad |b - a| < \infty \quad (4.9)$$

The eigenstates (3.10) of the 1D momentum operator $\hat{p}$,

$$\varphi_k(x) = A e^{ikx} = \frac{1}{\sqrt{2\pi}} e^{ikx}, \quad (4.10)$$

which are closely related to the eigenstates of the free-electron problem, are all unbound states.¹

As we shall see also in the remaining parts of the present manuscript, unbound states are indispensable in the proper treatment of scattering problems. Problems of this kind usually contain a beam (or, bunch) of particles, say, neutrons or electrons of similar energies or linear momenta, that is incident on a central potential (energy) barrier, as the one depicted in Fig.4.3.

Because $\int_{-\infty}^{\infty} |\psi|^2 dx$ does not converge for any unbound state, it is better in scattering problems to normalize a wave function ψ with respect to the *particle density* ρ , which is defined as the (differential) number of particles per differential length. We have for 1D problems

$$\rho = \frac{dN}{dx} \quad (4.11)$$

But we know that the norm $|\psi|^2$ of the wave function at a point x is directly related to the “presence” the particle there. This means that $|\psi|^2 dx$ must be proportional to ρdx at any point x . Taking the (arbitrary) proportionality constant equal to unity, we may write

$$|\psi|^2 dx = \rho dx = dN \quad (4.12)$$

¹ The coefficient $1/\sqrt{2\pi}$ in (4.10) does not originate from any normalization procedure; in fact, it comes from the orthonormality condition of the eigenstates. We say that the eigenstates of any observable must be *orthonormal* (or, in loose speaking, *perpendicular*) to each other. This also means that we cannot express any eigenstate of an observable in terms of the other eigenstates of that same observable. This corresponds mathematically to

$$\int_{-\infty}^{\infty} \varphi_k^*(x) \varphi_{k'}(x) dx = \delta_{kk'} = \begin{cases} 1 & \text{if } k = k', \\ 0 & \text{if } k \neq k' \end{cases}$$

where $\delta_{kk'}$ is called the *Kronecker delta function* and it is actually a *discrete* version of the *continuous* Dirac delta function (3.22). We have for the eigenstate (4.10) of the linear momentum operator at hand that

$$\int_{-\infty}^{\infty} \varphi_k^*(x) \varphi_{k'}(x) dx = \int_{-\infty}^{\infty} \left(\frac{1}{\sqrt{2\pi}} e^{-ikx} \right) \left(\frac{1}{\sqrt{2\pi}} e^{ik'x} \right) dx = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{ix(k' - k)} dx = \delta(k' - k) = \begin{cases} 1 & \text{if } k = k', \\ 0 & \text{if } k \neq k' \end{cases}$$

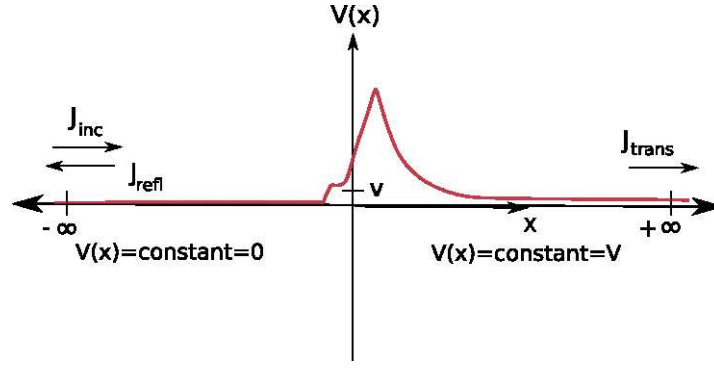


Figure 4.3: A representative potential (energy) profile of a 1D scattering problem. An incident beam with current density $\mathbb{J}_{\text{inc}}$ comes from $x = -\infty$. It encounters the central potential barrier $V(x)$; some of it is transmitted towards $x = +\infty$ with $\mathbb{J}_{\text{trans}}$ and the remaining is reflected back again towards $x = -\infty$ with $\mathbb{J}_{\text{refl}}$. Before and after the central scattering region, the potential $V(x)$ is constant at some specific values. Adapted from Ref. [29].

so that the total number of particle in any finite interval $(b - a)$ is equated (or, loosely speaking, *normalized*) to N through

$$\int_a^b |\psi|^2 dx = N \quad (4.13)$$

As an illustrating example [29], we may consider a 1D beam of 10^3 neutrons/cm, which are altogether in motion with the same specific linear momentum $p = \hbar k_0$. We may write the relevant wave function as

$$\psi(x, t) = 10^{3/2} e^{i(k_0 x - \omega t)}$$

whose norm gives exactly the total number of momentum per length as

$$|\psi|^2 = 10^3 \text{ cm}^{-1}$$

with

$$\frac{\hbar^2 k_0^2}{2m} = \hbar \omega$$

as being their common energy.

As we have seen in this example, for a uniform beam of particles, $|\psi|^2$ is constant. This means at the same time that the probability of finding particles *anywhere* (not at a specific x -position!) within the beam is exactly unity. Put in another way, the position of a particle with the certainly known linear momentum $p = \hbar k_0$ is absolutely uncertain, obeying thus Heisenberg's uncertainty principle.

4.2.1 Continuity Equation

The relationship between the current density $\mathbf{J}$ and particle density ρ in the most general 3D case is established by is the *continuity equation*

$$\frac{\partial \rho}{\partial t} + \nabla \cdot \mathbf{J} = 0 \quad (4.14)$$

In order to interpret this equation, we integrate the two terms over a closed region of volume V and surface area S :

$$\int_V \frac{\partial \rho}{\partial t} d\mathbf{r} + \int_V \nabla \cdot \mathbf{J} d\mathbf{r} = 0$$

or, changing the order of derivation and integration in the first term,

$$\frac{\partial}{\partial t} \int_V \rho d\mathbf{r} + \int_V \nabla \cdot \mathbf{J} d\mathbf{r} = 0 \quad (4.15)$$

The volume integral in the first term gives just the total number N of the particles enclosed within the volume V ,

$$N = \int_V \rho d\mathbf{r} \quad (4.16)$$

For the second term in (4.15), we make use of *Gauss's theorem*:

$$\int_V \nabla \cdot \mathbf{J} d\mathbf{r} = \int_S \mathbf{J} \cdot d\mathbf{S} \quad (4.17)$$

Feeding (4.16) and (4.17) into (4.15), we obtain

$$\frac{\partial N}{\partial t} + \int_S \mathbf{J} \cdot d\mathbf{S} = 0 \quad (4.18)$$

Since the current density $\mathbf{J}$ is directly proportional to the drift velocity $\mathbf{v}_d$ of the particles, as we have seen before in (2.20),

$$\mathbf{J} = nq\mathbf{v}_d$$

the surface integral in (4.18) is analogue to the elementary *mass flux* concept. In the present context, if $\int_S \mathbf{J} \cdot d\mathbf{S} > 0$, this surface integral gives the total number particles passing through the surface S , out of the volume V , from inside to outside S . The requirement (4.18) says then that $\partial N / \partial t < 0$ and that the total number particles out of the volume V is exactly equal to the *decrease* of the particles in V . In other words, the continuity equation (4.14), via (4.18), is just a concise mathematical statement for the fundamental concept of *conservation of matter*.

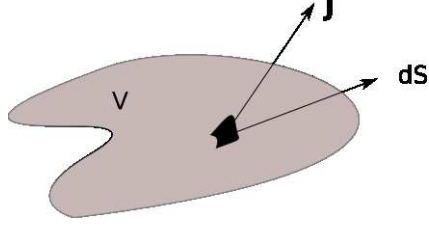


Figure 4.4: The geometry for the continuity equation (4.14); adapted from Ref. [30].

For a beam of particles confined to move in 1D, say, in the x -direction, the current density vector reduces to

$$\mathbf{J} = \mathbb{J}_x \hat{\mathbf{x}} \quad (4.19)$$

for which the continuity equation (4.14) will be

$$\frac{\partial \rho}{\partial t} + \frac{\partial \mathbb{J}_x}{\partial x} = 0 \quad (4.20)$$

Through (4.12), we have already associated the particle density with the square of the norm of the wave function.

$$\rho = |\psi|^2$$

We must similarly associate $\mathbb{J}_x$ with ψ . To do this, we should establish an equation identical to (4.20) in which $|\psi|^2$ is to be replaced with ρ . The best way is to make use of the time-dependent Schrödinger equation (3.28) that governs our 1D particles. We have

$$\frac{\partial \psi}{\partial t} = -\frac{i}{\hbar} \hat{H} \psi \quad \text{and} \quad \frac{\partial \psi^*}{\partial t} = +\frac{i}{\hbar} \hat{H} \psi^* \quad (4.21)$$

with $\hat{H}$ being the typical 1D Hamiltonian²

$$\hat{H} = \frac{\hat{p}_x^2}{2m} + V(x) = -\frac{\hbar^2}{2m} \frac{\partial^2}{\partial x^2} + V(x) \quad (4.22)$$

Our derivation should contain the time derivative of $|\psi|^2 = \psi^* \psi$, which is the particle density ρ . With (4.21), we obtain

$$\frac{\partial \psi^* \psi}{\partial t} = \psi^* \frac{\partial \psi}{\partial t} + \psi \frac{\partial \psi^*}{\partial t} = \psi^* \left(\frac{-i\hat{H}}{\hbar} \psi \right) + \psi \left(\frac{+i\hat{H}}{\hbar} \psi^* \right)$$

Insertion of the second part of (4.22) yields

$$\frac{\partial \psi^* \psi}{\partial t} = \frac{i\hbar}{2m} \left(\psi^* \frac{\partial^2 \psi}{\partial x^2} - \psi \frac{\partial^2 \psi^*}{\partial x^2} \right)$$

² The potential (energy) function $V(x)$ in (4.22) must be *real*. It can be shown that a complex potential (energy) function does not comply with the continuity equation (4.14).

or

$$\frac{\partial \psi^* \psi}{\partial t} + \frac{\partial}{\partial x} \left[\frac{\hbar}{2mi} \left(\psi^* \frac{\partial \psi}{\partial x} - \psi \frac{\partial \psi^*}{\partial x} \right) \right] = 0 \quad (4.23)$$

A closer look at (4.20) and (4.23) reveals that the 1D current density that we are after must be identified as³

$$\mathbb{J}_x = \frac{\hbar}{2mi} \left(\psi^* \frac{\partial \psi}{\partial x} - \psi \frac{\partial \psi^*}{\partial x} \right) \quad (4.24)$$

We should notice here that the dimensions of $\mathbb{J}_x$ is “particles per time,” which can be readily seen from (4.20).

A straightforward generalization of (4.24) yields the most general 3D current density as

$$\mathbf{J} = \frac{\hbar}{2mi} (\psi^* \nabla \psi - \psi \nabla \psi^*) \quad (4.25)$$

or, in terms of the 3D momentum operator $\hat{\mathbf{p}} = -i\hbar \nabla$,

$$\mathbf{J} = \frac{1}{2m} [\psi^* \hat{\mathbf{p}} \psi - (\psi^* \hat{\mathbf{p}} \psi)^*] \quad (4.26)$$

with dimensions “particles per area per time,” that follow most easily from (4.14).

4.2.2 Transmission and Reflection Coefficients

Next we turn to the discussion of concepts of *transmission* and *reflection* in 1D scattering problems, where the particles have definite momenta in the so-called *plane-wave states*. We assume that the wave functions pertinent to incident, transmitted, and reflected beams of particles are known, and that we have already obtained the related current densities by using (4.24). We *define* the *transmission* and *reflection coefficients*, respectively, as

$$T \equiv \left| \frac{\mathbb{J}_{\text{trans}}}{\mathbb{J}_{\text{inc}}} \right| \quad \text{and} \quad R \equiv \left| \frac{\mathbb{J}_{\text{refl}}}{\mathbb{J}_{\text{inc}}} \right| \quad (4.27)$$

We assume that the incident particles all have a definite momentum

$$p_{\text{inc}} = \hbar k_1 \quad (4.28)$$

so that the reflected particles all have exactly the *opposite* momentum

$$p_{\text{refl}} = -\hbar k_1 \quad (4.29)$$

³ The structure of $\mathbb{J}_x$ in (4.24) offers a good plausibility argument that if we are to have a nontrivial current density, the solution $\psi(x, t)$ of the relevant time-dependent Schrödinger equation must be a *complex* function: if $\psi(x, t)$ happens to be real, we obtain $\mathbb{J}_x = 0$ for all x .

(Remember that in 1D problems it is the sign of the momentum, or speed, of the particle that indicates the direction of its motion.) If we consider a general situation in which the potential (energy) profiles before and after the scattering region, just like the case depicted in Fig. 4.3, are different from each other, then the transmitted particles will have a particular momentum $\hbar k_2$ which will be different from that of the incident particles, $\hbar k_2 \neq \hbar k_1$:

$$p_{\text{trans}} = \hbar k_2 \quad (4.30)$$

In writing the three momenta above, we have assumed that the potential energies of the particles before and after the central scattering region are constant to some specific values. (In Figure 4.3, for example, the potential energies before and after the scattering region are zero and $V > 0$, respectively.) This means that we have *free-particle* wave functions in both of these incoming and outgoing domains. We might thus write

$$\begin{aligned} \psi_{\text{inc}} &= A e^{i(k_1 x - \omega_1 t)}, & \hbar \omega_1 &= E_{\text{inc}} = \frac{\hbar^2 k_1^2}{2m} \\ \psi_{\text{refl}} &= B e^{-i(k_1 x + \omega_1 t)}, & \hbar \omega_1 &= E_{\text{refl}} = \frac{\hbar^2 k_1^2}{2m} = E_{\text{inc}} \\ \psi_{\text{trans}} &= C e^{i(k_2 x - \omega_2 t)}, & \hbar \omega_2 &= E_{\text{trans}} = \frac{\hbar^2 k_2^2}{2m} + V = E_{\text{inc}} = \hbar \omega_1 \end{aligned} \quad (4.31)$$

where in the last line we have stated the fact that the (total) energy E of a particle in the beam must be conserved and that the frequency must be kept intact ($\omega_1 = \omega_2$) in passing through the central scattering region. As to the change in wave number k , it results from the obvious change in the kinetic energy (and momentum) of the particle.

Now using the 1D current density formula (4.24) with the wave functions in (4.31), we obtain

$$\begin{aligned} \mathbb{J}_{\text{inc}} &= \frac{\hbar}{m} k_1 |A|^2 \\ \mathbb{J}_{\text{trans}} &= \frac{\hbar}{m} k_2 |C|^2 \\ \mathbb{J}_{\text{refl}} &= \frac{\hbar}{m} k_1 |B|^2 \end{aligned} \quad (4.32)$$

We here notice immediately that the above results are the quantum mechanical counterparts of a classical particle current; the interconnection is made via $\mathbb{J} = \rho v$, $\rho = |\psi|^2$, and $v = \hbar k/m$.

Putting the relations in (4.32) into those in (4.27), we obtain the desired formulas for the transmission and reflection coefficients as

$$T = \left| \frac{C}{A} \right|^2 \frac{k_2}{k_1} \quad \text{and} \quad R = \left| \frac{B}{A} \right|^2 \quad (4.33)$$

It is seen that if the potential energies before and after the scattering region happen to be the same, so are then the corresponding wave numbers $k_1 = k_2$, and, in turn, the transmission coefficient formula takes the simple form $T = |C/A|^2$.

As we shall see later, out of computational easiness, most of the time the coefficient A of the incident wave is taken to be unity. Nevertheless, in order to obtain C/A and B/A in (4.33), we must first solve the related Schrödinger equation across the central potential (energy) barrier.

4.2.3 The Simple Step Problem: A Worked Example

In order to put what we have seen so far into practice, so that we may grasp firmly the way quantum mechanics works, we shall here work out a simple 1D scattering problem. It will contain a beam of particles, of (total) energy E , incident on a central, simple step, potential (energy) barrier, like the one depicted in Fig. 4.5. Although the energy of the incident particles is greater than the “hight” V of the barrier, $E > V$, in this specific problem at hand, all the other cases can be treated readily by applying the same lines of reasoning that will be discussed here.

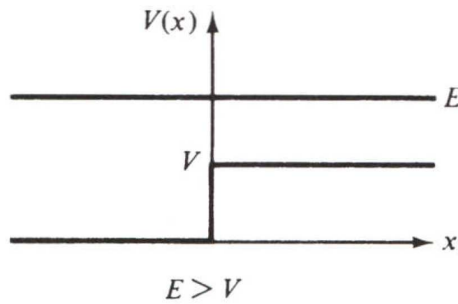


Figure 4.5: The simple step potential (energy) barrier treated in the worked example.

Our aim is to obtain the wave function $\varphi(x)$ everywhere as a function of x as the solution of the time-independent Schrödinger equation; we already know the time-dependence of the wave functions. The potential (energy) before the step (call it region I) is zero; it is constant V after the step (call it region II):

$$V(x) = \begin{cases} 0 & \text{for } x < 0, \\ V = \text{const.} & \text{for } x \geq 0 \end{cases}$$

In region I, with $V(x) = 0$, the total energy E is seen to be absolutely kinetic; the time-independent Schrödinger equation (4.1) for this region reads

$$-\frac{\hbar^2}{2m} \frac{\partial^2 \varphi}{\partial x^2} = E \varphi \quad (4.34)$$

If we let

$$\frac{\hbar^2 k_1^2}{2m} = E \quad (4.35)$$

then (4.34) becomes

$$\frac{\partial^2 \varphi}{\partial x^2} = -k_1^2 \varphi \quad (\text{region I}) \quad (4.36)$$

In region II, we have $V(x) = V = \text{const.}$; the kinetic energy T of the particles is seen to decrease by an amount V . The time-independent Schrödinger equation for this region becomes

$$-\frac{\hbar^2}{2m} \frac{\partial^2 \varphi}{\partial x^2} = (E - V) \varphi \quad (4.37)$$

Letting

$$\frac{\hbar^2 k_2^2}{2m} = E - V = T \quad (4.38)$$

we rewrite (4.37) as

$$\frac{\partial^2 \varphi}{\partial x^2} = -k_2^2 \varphi \quad (\text{region II}) \quad (4.39)$$

We have already encountered before this kind of differential equations given in (4.36) and (4.39); their solutions are elementary and they are respectively written in their most general forms as

$$\begin{aligned} \varphi_I &= A e^{ik_1 x} + B e^{-ik_1 x} \\ \varphi_{II} &= C e^{ik_2 x} + D e^{-ik_2 x} \end{aligned} \quad (4.40)$$

Actually here the terms on the right all have an unwritten time-dependency factor of the form $e^{-i\omega t}$. If we put these factors, we can interpret the very nature of individual terms as

$$\begin{aligned}
Ae^{ik_1x} &\rightarrow Ae^{ik_1x} e^{-i\omega_1 t} = Ae^{i(k_1x - \omega_1 t)} && \text{(incident wave moving to right in region I)} \\
Be^{-ik_1x} &\rightarrow Be^{-ik_1x} e^{-i\omega_1 t} = Be^{-i(k_1x + \omega_1 t)} && \text{(reflected wave moving to left in region I)} \\
Ce^{ik_2x} &\rightarrow Ce^{ik_2x} e^{-i\omega_2 t} = Ce^{i(k_2x - \omega_2 t)} && \text{(transmitted wave moving to right in region II)} \\
De^{-ik_2x} &\rightarrow De^{-ik_2x} e^{-i\omega_2 t} = De^{-i(k_2x + \omega_2 t)} && \text{(a wave moving to left in region II)}
\end{aligned}$$

Because of the wording of our original problem at hand, there must be no beam of particles originating from $x = +\infty$ and moving to the left in region II. Consequently we must set $D = 0$ in (4.40). We are then left with

$$\begin{aligned}
\varphi_I &= Ae^{ik_1x} + Be^{-ik_1x} \\
\varphi_{II} &= Ce^{ik_2x}
\end{aligned} \tag{4.41}$$

Figure 4.6 shows the potential (energy) barrier, again, with the solutions in both regions.

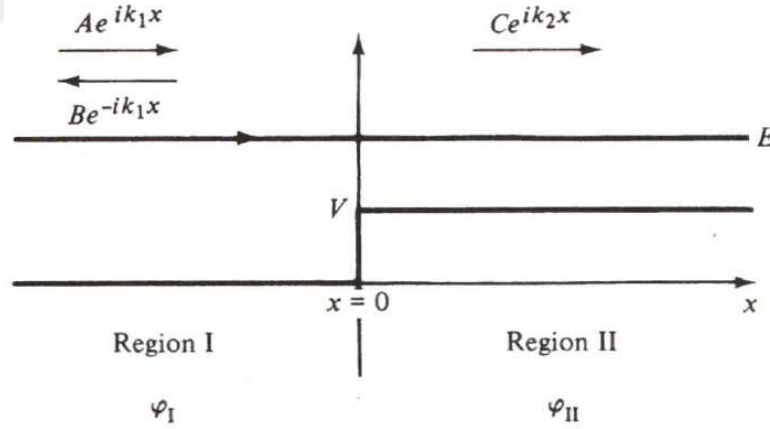


Figure 4.6: The simple step barrier with representative solutions in both regions; taken from Ref. [31].

Now, we know from our previous discussions that, as being the *single* solution of the time-independent Schrödinger equation, the wave function in (4.41) and its first space derivative must be continuous at $x = 0$:

$$\begin{aligned}
\varphi_I(0) &= \varphi_{II}(0) \\
\frac{\partial}{\partial x} \varphi_I(0) &= \frac{\partial}{\partial x} \varphi_{II}(0)
\end{aligned} \tag{4.42}$$

which yield

$$A + B = C$$

$$A - B = \frac{k_2}{k_1} C$$

Solving these two for the desired unknowns, we obtain

$$\frac{C}{A} = \frac{2}{1 + k_2/k_1} \quad \text{and} \quad \frac{B}{A} = \frac{1 - k_2/k_1}{1 + k_2/k_1} \quad (4.43)$$

Finally we insert these into (4.33) and end up with the transmission and reflection coefficients that we are after:

$$T = \frac{4k_2/k_1}{[1 + (k_2/k_1)]^2} \quad \text{and} \quad R = \left| \frac{1 - k_2/k_1}{1 + k_2/k_1} \right|^2 \quad (4.44)$$

It follows from these two coefficients that their sum is always unity, as might be intuitively expected:

$$T + R = 1 \quad (4.45)$$

Although it has been drawn from the simple step problem at hand, the conclusion (4.45) is valid for all kinds of 1D barrier problems.

Solving (4.35) and (4.38) for k_2/k_1 gives

$$\left(\frac{k_2}{k_1} \right)^2 = 1 - \frac{V}{E} \quad (4.46)$$

We confer that:

1. if $E \geq V$, as in the present situation, then it is always true that $0 \leq k_2/k_1 \leq 1$;
2. if $E \gg V$, then $k_2/k_1 \rightarrow 1$, which means $T \rightarrow 1$ and $R \rightarrow 0$; we say that we witness a phenomenon called *total transmission*; and
3. if $E = V$, then $k_2/k_1 = 0$, which corresponds to $T = 0$ and $R = 1$; we now speak of the concept of *total reflection* (or, *zero transmission*).

CHAPTER 5

DELTA-FUNCTION POTENTIALS

5.1 Delta-Function Well [32]

In this section we shall solve another 1D problem that is simple yet beautiful with its results applicable to real practical situations. It includes only a single *delta-function (potential energy) well* positioned at the origin, as shown in Fig. 5.1

$$V(x) = -\alpha\delta(x) \quad (5.1)$$

where $\alpha > 0$ is some constant for the “strength” of the well.

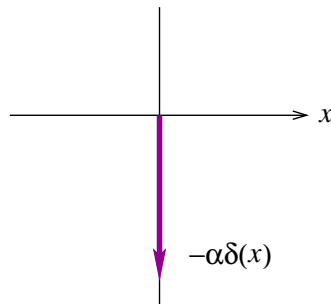


Figure 5.1: The (Dirac) delta-function (potential energy) well.

Though it is not absolutely realistic, the delta-function well is an excellent approximation for problems that naturally include an *extremely* narrow, *extremely* strong attractive potential (energy) profile. Such a potential might be a characteristic representative for an electron under the effect of a nucleus, or for a proton under the effect of another proton within the nucleus.

With the potential (5.1), the time-independent Schrödinger equation (3.12) takes the usual form

$$-\frac{\hbar^2}{2m} \frac{d^2\varphi}{dx^2} - \alpha\delta(x)\varphi(x) = E\varphi(x) \quad (5.2)$$

which has solutions for both *bound* states for which $E < 0$ and *unbound* (sometimes *scattering*) states with $E > 0$. We shall look first at the latter.

5.1.1 Bound States ($E < 0$)

We consider a particle, of mass m , confined within an infinitely narrow and deep (so that infinitely strong) potential well so that its energy is nowhere kinetic. It can be spotted only in 1D, say, again in the x -axis. The delta-well is located at the origin, as given by (5.2) and depicted in Fig. (5.1).

On the left we have $x < 0$ and $V = 0$, the Schrödinger equation (5.2) reduces to

$$\frac{d^2\varphi}{dx^2} = -\frac{2mE}{\hbar^2} \varphi(x) = \kappa^2 \varphi(x) \quad (5.3)$$

with

$$\kappa \equiv \frac{\sqrt{-2mE}}{\hbar} \quad (5.4)$$

(Note here that κ is a real and positive since $E < 0$.) The solution to (5.3) elementary; it is written in general terms as

$$\varphi(x) = Ae^{-\kappa x} + Be^{\kappa x}$$

where only the second term is admissible because the first one blows up as $x \rightarrow -\infty$; we must take $A = 0$. (Note that the second term goes to zero as $x \rightarrow -\infty$ so that it is perfectly normalizable.) Thus we have for the left part

$$\varphi(x) = Be^{\kappa x} \quad (x < 0) \quad (5.5)$$

Similar arguments are valid on the right of the well, where $x > 0$. We just write

$$\varphi(x) = Ce^{-\kappa x} + De^{\kappa x}$$

This time, as $x \rightarrow \infty$, the first term goes to zero and the second one blows up, so the latter is not acceptable, we choose $D = 0$:

$$\varphi(x) = De^{-\kappa x} \quad (x > 0) \quad (5.6)$$

The continuity condition for bound states dictates that these two solutions (5.5) and (5.6) must match at the origin where the delta-well is located. This means that $D = B$, so

$$\varphi(x) = \begin{cases} Be^{\kappa x} & (x \leq 0) \\ Be^{-\kappa x} & (x \geq 0) \end{cases} \quad (5.7)$$

Actually we have another condition that $d\varphi/dx$ be continuous everywhere except at the origin where the potential (energy) is infinitely deep. But this condition gives us nothing new, at least for the time being.

We plot the wave function $\varphi(x)$ in Fig. 5.2 (the parameters seen on this figure will be calculated shortly). We see that as $x \rightarrow \pm\infty$, $\varphi(x) \rightarrow 0$, as it must be; and that it has a kink at the origin which is the joint point for the left and right solutions.

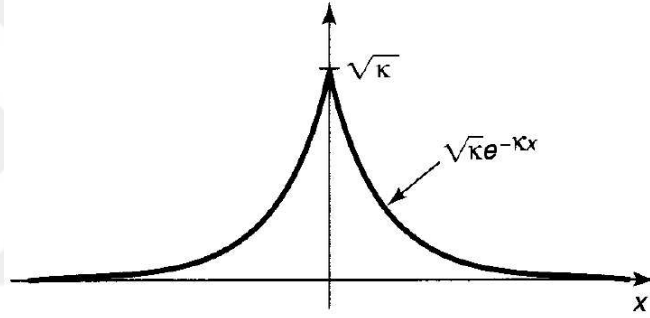


Figure 5.2: The wave function $\varphi(x)$ of the only bound state of the delta-function well; taken from Ref. [33].

Until this point we have not yet made use of the delta function so much. We shall see in the following that it is the delta function that settles down the discontinuity of $d\varphi/dx$ at the origin. The fact that $d\varphi/dx$ is only piecewise continuous will be a by-product of our mathematical trick.

We will first integrate each term in the Schrödinger equation (5.2) from $-\epsilon$ to $+\epsilon$, where $0 < \epsilon \ll 1$, and then allow $\epsilon \rightarrow 0$ in a usual limiting process:

$$-\frac{\hbar^2}{2m} \int_{-\epsilon}^{+\epsilon} \frac{d^2\varphi}{dx^2} dx - \alpha \int_{-\epsilon}^{+\epsilon} \delta(x)\varphi(x) dx = E \int_{-\epsilon}^{+\epsilon} \varphi(x) dx \quad (5.8)$$

It is clear that $\epsilon \rightarrow 0$, the last integral goes also to zero, because the area under the $\varphi(x)$ graph with constant height B and vanishing width $\Delta\epsilon$ is zero. The middle

integral contains the Dirac delta function, and its gives

$$-\alpha \int_{-\epsilon}^{+\epsilon} \delta(x) \varphi(x) dx = -\alpha \varphi(0)$$

The first integral is just $d\varphi/dx$ evaluated at the integral boundaries. With these and with the use of (5.7), (5.8) becomes

$$\begin{aligned} \left. \frac{d\varphi}{dx} \right|_{+\epsilon} - \left. \frac{d\varphi}{dx} \right|_{-\epsilon} &= -\frac{2m\alpha}{\hbar^2} \varphi(0) \\ \left. \frac{d}{dx} (Be^{-\kappa x}) \right|_{+\epsilon} - \left. \frac{d}{dx} (Be^{+\kappa x}) \right|_{-\epsilon} &= -\frac{2m\alpha}{\hbar^2} B \\ (-\kappa e^{-\kappa\epsilon}) - (+\kappa e^{+\kappa\epsilon}) &= -\frac{2m\alpha}{\hbar^2} \\ \kappa (e^{-\kappa\epsilon} + e^{+\kappa\epsilon}) &= \frac{2m\alpha}{\hbar^2} \end{aligned} \quad (5.9)$$

where we see that $d\varphi/dx$ is *not* continuous and that $\varphi(x)$ has a kink at the origin. If we take the limit $\epsilon \rightarrow 0$ in the last line, we obtain $2\kappa = 2m\alpha/\hbar^2$, or

$$\kappa = \frac{m\alpha}{\hbar^2} \quad (5.10)$$

With this the *only* allowed energy is calculated from (5.4) as

$$E = -\frac{\hbar^2 \kappa^2}{2m} = -\frac{m\alpha^2}{2\hbar^2} \quad (5.11)$$

Now we can normalize the *only* wave function φ of the system. With the fact that $\varphi(x)$ is an even function, we have

$$\int_{-\infty}^{+\infty} |\varphi(x)|^2 dx = 2|B|^2 \int_0^{\infty} e^{-2\kappa x} dx = 2|B|^2 \left[-\frac{1}{2\kappa} e^{-2\kappa x} \right]_0^{\infty} = \frac{|B|^2}{\kappa} = 1$$

We here, for convenience, choose the real positive root, and, with (5.10), we have

$$B = \sqrt{\kappa} = \frac{\sqrt{m\alpha}}{\hbar} \quad (5.12)$$

To recapitulate, the delta-function well problem has one, and only one, bound state, that is written in its final form as

$$\varphi(x) = \frac{\sqrt{m\alpha}}{\hbar} e^{-m\alpha|x|/\hbar^2} \quad \text{with} \quad E = -\frac{m\alpha^2}{2\hbar^2} \quad (5.13)$$

What we have learnt from this problem? We have a closer look again at Fig. 5.2 that shows the wave function $\varphi(x)$ for the only bound state of the delta-function well problem at hand. We first recall that $|\varphi(x)|^2 dx$ is directly proportional to probability for the particle to be found within the infinitesimal length dx which is a distance x from the origin. Since $|\varphi(x)|^2$ has its maximum at the origin, we conclude that it is highly likely for the particle to be found near the origin; this is what we expect *classically*: the particle can never escape from the infinitely deep potential well. On the other hand, the same figure suggests that $|\varphi(x)|^2$ has *non-zero* values near the origin, though it goes rapidly to zero on the left and right of the origin. This is an absolutely *quantum mechanical* outcome. Classically these regions are absolutely *forbidden*, but quantum mechanics says that there is a definite non-zero probability for the particle to be detected in two these regions!

This unprecedented result implies also another possibility. As a simple example, consider a single particle which is within an infinitely deep and strong potential well. Further, let we have another identical delta-function well next to the first one. (In reality, this might correspond to an electron under the effect of two nuclei.) Provided that the distance between these two delta-function wells is sufficiently small, there is a certain non-zero probability that the particle detected at some time in one of the potential wells can readily be detected at a later time in the other potential well! To compound the situation further, the particle can be spotted again in the first potential well at another time! The particle will simply “tunnel” the wall of the potential wells from one to the other at will. This and similar phenomena are really something that are completely impossible from the classical point of view.

5.1.2 Unbound States ($E > 0$)

It might be better to describe the problem for this case in order for the reader to grasp firmly what is going on from the very outset. We consider again a particle, of mass m , confined to *move* in an 1D path, along the x -axis. It *does* come (so that it has a definite kinetic energy) from $-\infty$ and encounters at the origin an infinitely narrow and infinitely deep delta-function potential well (Fig. 5.1). After it is *scattered* by the potential well, it *does* go towards $+\infty$ with a definite kinetic energy again. The

delta-function well is, to repeat, given by (5.2). In order to keep the formulation as general as possible, we will not exclude the probable opposite case in which the particle comes from $+\infty$, is scattered at the origin, and then goes away to $-\infty$.

At regions away from delta-function well, i.e., for $x \neq 0$, $V = 0$ and the particle is completely free; its energy is purely kinetic. In other words, we will be dealing with traveling waves in these regions, as we will see shortly.

On the left and on the right of the potential well, the Schrödinger equation (5.2), with $V = 0$, reduces to

$$\frac{d^2\varphi}{dx^2} = -\frac{2mE}{\hbar^2} \varphi(x) = -k^2\varphi(x)$$

where

$$k \equiv \frac{\sqrt{2mE}}{\hbar} \quad (5.14)$$

which is a positive and real number; it will be seen to be the *wave number* of the incident, reflected, and transmitted traveling waves. The most general solution to the above differential equation for the two regions can be written as

$$\begin{aligned} \varphi(x) &= Ae^{ikx} + Be^{-ikx} & (x < 0) \\ \varphi(x) &= Ce^{ikx} + De^{-ikx} & (x > 0) \end{aligned} \quad (5.15)$$

Note that this time we cannot rule out any term because they are all of sinusoidal nature; there is no blowing-up term. Figure 5.3 illustrates all the possible components considered in (5.15). We have a prior experience to interpret them: with $e^{-i\omega t}$ coupled to them, Ae^{ikx} and Ce^{ikx} (Be^{-ikx} and De^{-ikx}) represent waves traveling to the right (left). Therefore, if particle comes from $-\infty$ and scattered toward $+\infty$, we should take $D = 0$ in (5.15); if it originates from $+\infty$ and continues toward $-\infty$, we should then take $A = 0$.

The two wave functions in (5.15) must be the continuation of each other at the origin. In other words, we must have only one *single* solution and this solution must be *continuous* everywhere, in the present case, at the origin. We then have

$$C + D = A + B \quad (5.16)$$

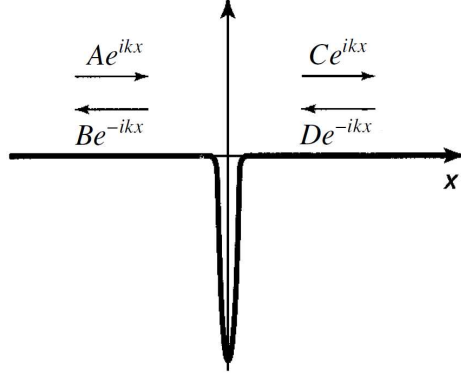


Figure 5.3: All the possible components of the wave function $\varphi(x)$ for unbound states of the delta-function well; adapted from Ref. [34].

For the derivative $d\varphi/dx$ of the wave function, we just reiterate our “small- ϵ ” trick. Repeating the steps from (5.8) through (5.9), we proceed this time as

$$\begin{aligned} \left. \frac{d\varphi}{dx} \right|_{+\epsilon} - \left. \frac{d\varphi}{dx} \right|_{-\epsilon} &= -\frac{2m\alpha}{\hbar^2} \varphi(0) \\ \left. \frac{d}{dx} (Ce^{ikx} + De^{-ikx}) \right|_{+\epsilon} - \left. \frac{d}{dx} (Ae^{ikx} + Be^{-ikx}) \right|_{-\epsilon} &= -\frac{2m\alpha}{\hbar^2} (A + B) \\ ik(Ce^{ik\epsilon} - De^{-ik\epsilon}) - ik(Ae^{ik\epsilon} - Be^{-ik\epsilon}) &= -\frac{2m\alpha}{\hbar^2} (A + B) \end{aligned}$$

Taking the limit $\epsilon \rightarrow 0$ in the last line, we obtain

$$ik(C - D - A + B) = -\frac{2m\alpha}{\hbar^2} (A + B) \quad (5.17)$$

or, in terms of a new parameter

$$\beta = \frac{m\alpha}{\hbar^2 k}, \quad (5.18)$$

we have a more compact expression for (5.17):

$$C - D = A(1 + 2i\beta) - B(1 - 2i\beta) \quad (5.19)$$

Now for a typical scattering experiment in which the particle comes from the *left*, from $-\infty$, we take $D = 0$ and rewrite (5.16) and (5.19) as

$$C = A + B$$

$$C = A(1 + 2i\beta) - B(1 - 2i\beta)$$

The solutions for B and C , the respective amplitudes of the reflected and transmitted waves, in terms of the amplitude A of the incident wave are found as

$$B = \frac{i\beta}{1 - i\beta} A \quad \text{and} \quad C = \frac{1}{1 - i\beta} A \quad (5.20)$$

With these we can now calculate two important quantities of interest: the *transmission* and *reflection* coefficients defined previously (4.33). Noting that the wave numbers of the waves before and after scattering are the same, we have

$$T = \frac{|C|^2}{|A|^2} = \left| \frac{1}{1 - i\beta} \right|^2 = \frac{1}{|1 - i\beta|^2} \quad (5.21)$$

and

$$T = \frac{1}{1 + \beta^2} \quad (5.22)$$

and similarly

$$R = \frac{|B|^2}{|A|^2} = \left| \frac{i\beta}{1 - i\beta} \right|^2 = \frac{|\beta|^2}{|1 - i\beta|^2} \quad (5.23)$$

and

$$R = \frac{\beta^2}{1 + \beta^2} \quad (5.24)$$

We immediately note that the sum of transmission and reflection coefficients is unity, as it must be:

$$T + R = 1 \quad (5.25)$$

Using (5.14) and (5.18), we can express these coefficients, (5.22) and (5.24), in terms of the original parameters used in the wording of the problem:

$$T = \frac{1}{1 + m\alpha^2/2\hbar^2 E} \quad \text{and} \quad R = \frac{1}{1 + 2\hbar^2 E/m\alpha^2} \quad (5.26)$$

It follows from these that as the energy of the incoming particle, which is purely kinetic, the probability of transmission increases and that of reflection decreases. These are all reasonably expected.

Now let us consider the other scattering experiment in which the particle comes from the *right*, from $+\infty$. This time we take $A = 0$ so that (5.16) and (5.19) take the forms

$$C + D = B$$

$$C - D = -B(1 - 2i\beta)$$

Solving these for B and C , the respective amplitudes, this time, of the *transmitted* and *reflected* waves, in terms of the amplitude D of the *incident* wave, we obtain

$$B = \frac{1}{1 - i\beta} D \quad \text{and} \quad C = \frac{i\beta}{1 - i\beta} D \quad (5.27)$$

Then, the transmission and reflection coefficients are found as

$$\begin{aligned} T &= \frac{|B|^2}{|D|^2} = \frac{1}{1 + \beta^2} \\ R &= \frac{|C|^2}{|D|^2} = \frac{\beta^2}{1 + \beta^2} \end{aligned} \tag{5.28}$$

as before.

The attentive reader perhaps have already noticed that the results of the problem at hand must not be dependent on the direction of the incoming particle; it might come from the left or right and that does not matter at all.¹ This means that if we have a closer look at Fig. 5.3, and detect the incoming and outgoing waves for the left and right scattering experiments, it is sufficient for us to make changes $A \rightarrow D$, $B \rightarrow C$, and $C \rightarrow B$ in (5.20), (5.21), and (5.23) to reach the results in (5.27) and (5.28), as the reader can readily verify.

5.2 Delta-Function Barrier

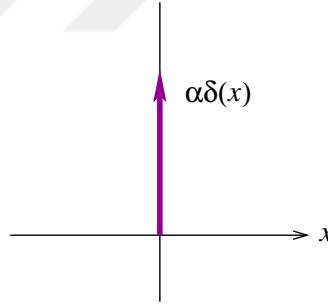


Figure 5.4: The (Dirac) delta-function barrier.

Let us now touch on the problem of a *delta-function barrier* (Fig. 5.4), but only briefly. We have again a particle of mass m that is confined to move along the x -axis. In this case there is *not*, of course, any bound state ($E < 0$) because the particle comes with a definite kinetic energy, say, from the left and encounters at the origin an infinitely narrow and infinitely high delta-function potential barrier, given this time by

$$V(x) = \alpha\delta(x) \tag{5.29}$$

¹ “In space there is no preferred direction,” *n’est-ce pas?*

with $\alpha > 0$ being, again, the “strength” of the barrier. After it is *scattered* by this barrier, the particle finally goes to the right with the same kinetic energy again. Therefore, the problem of delta-function barrier includes only *unbound* states ($E > 0$).

We notice at once that the way we should follow to solve the present scattering from a delta-function barrier problem is exactly identical to that we followed to solve the previous scattering from a delta-function well problem. All the details are the same; all we must do is to change the sign of the “strength” α . This sign change, however, will have no effect on the transmission and reflection coefficients (5.26), for they depend not on the sign of α but on its square, α^2 . It might appear peculiar at first sight that the likelihood of the passing of the particle *through* the delta-function barrier amounts conclusively to its passing *over* the delta-function well! This is again a notably bizarre dictation of quantum mechanics. In classical point of view, the particle can never pass through an extremely high potential energy barrier, no matter how large its energy is. Actually the scattering phenomenon in classical mechanics is rather dreary: When $E > V$, then $T = 1$ and $R = 0$ so that the particle passes over the barrier for certain. In the reverse case, when $E < V$, this time $T = 0$ and $R = 1$; the particle goes up the potential barrier until it exhausts its kinetic energy, and stops, and then returns back through the same route that it came.

The quantum mechanical version of scattering problem is far more richer. There is always a definite non-zero probability for the particle to “tunnels” through a potential energy barrier even if $E < V$. (This situation is similar to the case of a single particle within a delta-function well that we discussed above, where we have seen that the particle can “tunnel” through the walls of the well at will and can find itself in a neighboring well.) In the reverse case, even if $E > V$, there is always a certain probability that the particle will be reflected back, as we have just seen in the delta-function well where we have witnessed a non-zero reflection coefficient, written in (5.26).

5.3 Scattering Matrix [35]

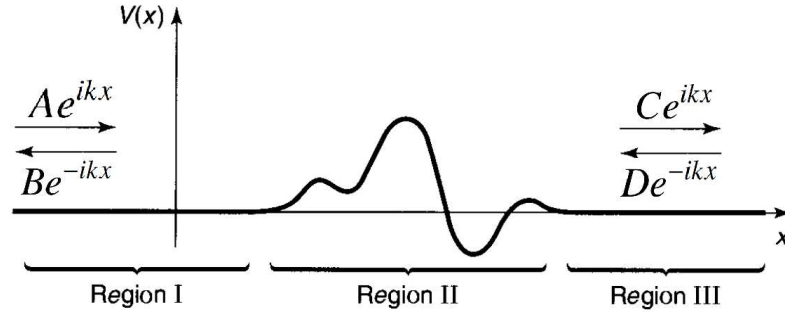


Figure 5.5: A depiction for a generalized scattering problem. Note that $V(x) = 0$ in Region I and II. Adapted from Ref. [35].

We can generalize scattering problems ($E > 0$), and their solutions, for any arbitrarily shaped central potential(s). Figure 5.5 below depicts the most general situation, where $V(x) = 0$ everywhere except possibly in Region II. Then, in Region I and III we have

$$\begin{aligned}\varphi_I(x) &= Ae^{ikx} + Be^{-ikx} \\ \varphi_{III}(x) &= Ce^{ikx} + De^{-ikx}\end{aligned}\tag{5.30}$$

with

$$k \equiv \frac{\sqrt{2mE}}{\hbar}\tag{5.31}$$

being the wave number of the incoming and outgoing wave functions. We cannot say anything about the wave function in Region II until we specify the central potential profile. Nonetheless, since the time-independent Schrödinger equation is a 2nd order linear differential equation, we know in advance that its solution everywhere must be composed of two *linearly independent* particular solutions [36]. Therefore, the wave function in Region II can be written as

$$\varphi_{II}(x) = \alpha f(x) + \beta g(x)\tag{5.32}$$

There are totally four boundary conditions: two for the border between Regions I and II, and two for that between Regions II and III. Then we can use two of four in eliminating α and β and the remaining two to determine B and C in terms of A and D . As a result, we end up with

$$B = S_{11}A + S_{12}D \quad \text{and} \quad C = S_{21}A + S_{22}D\tag{5.33}$$

where coefficients S_{ij} depend on the energy E via the wave number k given in (5.31). The structure of these equations is reminiscent of a 2×2 matrix whose constituents are S_{ij} :

$$\mathbf{S} = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \quad (5.34)$$

This is the so-called *scattering matrix* (or, *S-matrix*) giving us the connection between the outgoing amplitudes (B and C) and the incoming amplitudes (A and D) through the matrix equation

$$\begin{pmatrix} B \\ C \end{pmatrix} = \mathbf{S} \begin{pmatrix} A \\ D \end{pmatrix} \quad (5.35)$$

or, in full form,

$$\begin{pmatrix} B \\ C \end{pmatrix} = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \begin{pmatrix} A \\ D \end{pmatrix} \quad (5.36)$$

In the problem of scattering from the left, we take $D = 0$ and the reflection and transmission coefficients follow from (5.33) as

$$R_\ell = \frac{|B|^2}{|A|^2} \Big|_{D=0} = |S_{11}|^2 \quad \text{and} \quad T_\ell = \frac{|C|^2}{|A|^2} \Big|_{D=0} = |S_{21}|^2 \quad (5.37)$$

and in the problem of scattering from the right, we obtain with $A = 0$ that

$$R_r = \frac{|C|^2}{|D|^2} \Big|_{A=0} = |S_{22}|^2 \quad \text{and} \quad T_r = \frac{|B|^2}{|D|^2} \Big|_{A=0} = |S_{12}|^2 \quad (5.38)$$

Everything about a scattering problem is embedded within the S-matrix. Even so, information about bound states, if any, is contained in the S-matrix, though in a hidden form. For instance, for $E < 0$, the wave function $\varphi(x)$ is of the form

$$\varphi(x) = \begin{cases} Be^{\kappa x} & \text{(Region I)} \\ \alpha f(x) + \beta g(x) & \text{(Region II)} \\ Ce^{-\kappa x} & \text{(Region III)} \end{cases}$$

where

$$\kappa \equiv \frac{\sqrt{-2mE}}{\hbar}$$

Are these recoverable from the S-matrix? Answer is yes. To see this we note that all the boundary conditions for equations in (5.30) and (5.32) are the same as before.

Therefore, the S-matrix must have the same structure, except that the energy E is *negative* now, so must take $k \rightarrow i\kappa$ in these equations: we obtain from (5.30) that

$$\varphi_I(x) \rightarrow Ae^{-\kappa x} + Be^{\kappa x}$$

$$\varphi_{III}(x) \rightarrow Ce^{-\kappa x} + De^{\kappa x}$$

But then, here we must take $A = 0$ and $D = 0$ because the exponential terms next to them blow up at $-\infty$ and $+\infty$, respectively. This means, in turn, from (5.33) that at least two constituents of the S-matrix will be infinite. In other words, if we happen to obtain the S-matrix (necessarily for $E > 0$), and if we are to pin down the bound states, if any, we just make the change $k \rightarrow i\kappa$, and check for the energies at which the elements of the S-matrix blow up.

5.3.1 The S-Matrix of a Delta-function Well: A Worked Example

As a concrete example, let us form the S-matrix for the problem in which a particle is scattered from a delta-function well, which we studied in detail in Section 5.1.2. We have had two equations, (5.16) and (5.19), for the amplitudes of the incoming and outgoing waves. We repeat again them here:

$$C + D = A + B \tag{5.39}$$

$$C - D = (1 + 2i\beta)A - (1 - 2i\beta)B \tag{5.40}$$

where

$$\beta = \frac{m\alpha}{\hbar^2 k} \quad \text{and} \quad k = \frac{\sqrt{2mE}}{\hbar} \tag{5.41}$$

In order to construct the relevant S-matrix, we must obtain two equations in the form of (5.33):

$$B = S_{11}A + S_{12}D \quad \text{and} \quad C = S_{21}A + S_{22}D \tag{5.42}$$

If we subtract (5.40) from (5.39), we get

$$2D = -2i\beta A + 2(1 - i\beta)B$$

whence

$$B = \frac{1}{1 - i\beta} (i\beta A + D) \tag{5.43}$$

Next, multiplying (5.39) by $(1 - 2i\beta)$ and add the resulting expression to (5.40), we obtain

$$2(1 - i\beta)C - 2i\beta D = 2A$$

whence

$$C = \frac{1}{1 - i\beta} (A + i\beta D) \quad (5.44)$$

Comparing the elements of (5.43) and (5.44) with those in (5.42), we construct the desired S-matrix as

$$S = \frac{1}{1 - i\beta} \begin{pmatrix} i\beta & 1 \\ 1 & i\beta \end{pmatrix} \quad (5.45)$$

Now we can use this S-matrix to extract the energy of the bound state. The ratio $1/(1 - i\beta)$ is common in all the elements of this matrix. We first write this ratio in its explicit form and then make in it the change $k \rightarrow i\kappa$:

$$\frac{1}{1 - i\beta} = \frac{1}{1 - im\alpha/\hbar^2 k} \xrightarrow{k \rightarrow i\kappa} \frac{1}{1 - im\alpha/\hbar^2 i\kappa} = \frac{1}{1 - m\alpha/\hbar^2 \kappa}$$

We see that this ratio blows up whenever

$$\frac{m\alpha}{\hbar^2 \kappa} = 1 \quad \text{or} \quad \kappa = \frac{m\alpha}{\hbar^2}$$

With this, the energy $E < 0$ of the bound state is found from $\kappa = \sqrt{-2mE}/\hbar$ as

$$E = -\frac{\hbar^2 \kappa^2}{2m} = -\frac{\hbar^2}{2m} \left(\frac{m\alpha}{\hbar^2} \right)^2$$

whence

$$E = -\frac{m\alpha^2}{2\hbar^2} \quad (5.46)$$

If we compare this result with that in (5.11), we see that this energy amounts exactly to the one we have found previously, for which we were required considerable time and effort!

CHAPTER 6

DIRAC NOTATION FOR QUANTUM STATES

6.1 Dirac Notation for Continuous States [37]

In this section we shall introduce the so-called *Dirac notation*, a very powerful mathematical tool in quantum mechanical calculations. Here we will confine the discussion to continuous states only. Since the nature of the subject of this work is in general *discrete*, we keep the discussion for this section rather brief. In the next one, we will expand the use of Dirac notation, with its all details, to discrete states. Henceforth, it will be one of our essential tools that we will use in the remainder of the manuscript.

The Dirac notation is a short-hand monogram to the integral of type

$$\langle \psi | \varphi \rangle = \int_{-\infty}^{\infty} \psi^*(x) \varphi(x) dx \quad (6.1)$$

where $\psi(x)$ and $\varphi(x)$ are two state functions. As it is clearly seen, $\langle \psi | \varphi \rangle$ is a mathematical recipe, instructing: (1) take the complex conjugate of the item, *whatever it is*, in the first slot, $\psi \rightarrow \psi^*$, and then, (2) integrate the product $\psi^* \varphi$.

For two state functions $\psi(x)$ and $\varphi(x)$ satisfying

$$\int_{-\infty}^{\infty} \psi^*(x) \varphi(x) dx < \infty \quad (6.2)$$

and for a complex number a , the following properties always hold:

$$\langle \psi | a \varphi \rangle = a \langle \psi | \varphi \rangle \quad (6.3)$$

$$\langle a \psi | \varphi \rangle = a^* \langle \psi | \varphi \rangle \quad (6.4)$$

$$\langle \psi | \varphi \rangle^* = \langle \varphi | \psi \rangle \quad (6.5)$$

$$\langle \psi_1 + \psi_2 | \varphi_1 + \varphi_2 \rangle = \langle \psi_1 | \varphi_1 \rangle + \langle \psi_1 | \varphi_2 \rangle + \langle \psi_2 | \varphi_1 \rangle + \langle \psi_2 | \varphi_2 \rangle \quad (6.6)$$

In the Dirac notation $\langle\psi|\varphi\rangle$, the first part $\langle\psi|$ is said to be a “bra vector,” the second part $|\varphi\rangle$ a “ket vector.” Whenever they combine, we obtain the “bra-ket,” $\langle\psi|\varphi\rangle$.

6.2 Dirac Notation for Discrete, Path-Like, States [38]

For this version of the Dirac notation, we will make use of the experimental features of the famous double-slit experiment (Fig. 6.1). We have an electron gun as the source s of electrons, then a wall containing two slits, and then a detector which is a distance x to the center of a second wall.

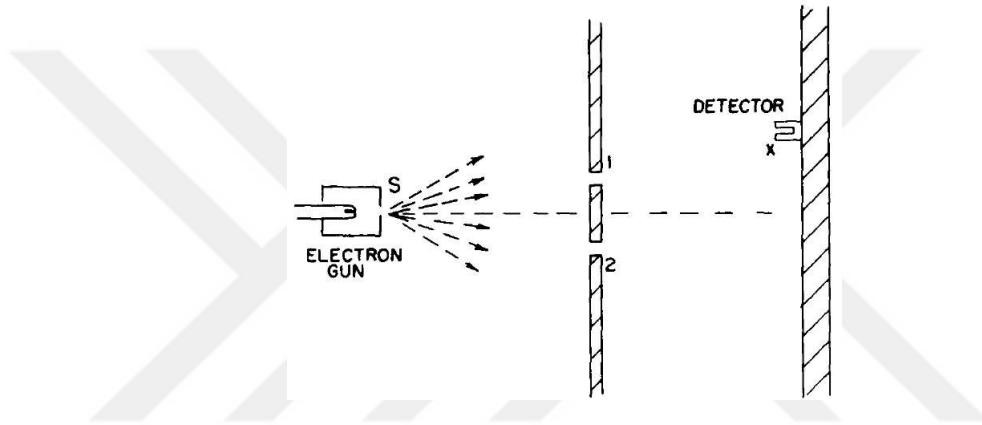


Figure 6.1: The double-slit experiment; taken from Ref. [27].

We may ask: what is the probability that an electron will be detected at x ? The *first general principle* of quantum mechanics says that the *probability* that an electron fired from the source s will reach x can be quantitatively described by the absolute square of, in general, a complex number, referred to as a (*probability*) *amplitude*.¹ In the present case, we speak of “the amplitude that an electron liberated from the source s will be detected at x , by the detector.” We can write this amplitude using a short-hand notation, introduced first by Dirac, as

$$\langle \text{electron reaches } x \mid \text{electron is liberated from } s \rangle \quad (6.7)$$

¹ It is very important to distinguish *probability* from *probability amplitude* (or, simply *amplitude*, from now on). A wordy relationship between them can be written as

$$\text{probability} \sim |\text{amplitude}|^2$$

or, more compactly,

$$\langle x | s \rangle \quad (6.8)$$

Thus, in the monogram $\langle x | s \rangle$, the two brackets $\langle \rangle$ stands for “the amplitude that,” the ket part $| \rangle$ contains the *starting* point (or, condition), and the bra part $\langle |$ contains the *final* point (or, condition).

The *second general principle* of quantum mechanics states that if an electron can reach a final state by two distinct ways, the total amplitude for the whole event is equal to the *sum* of separate amplitudes for the two distinct ways followed separately. We may then write

$$\langle x | s \rangle_{\text{both slits open}} = \langle x | s \rangle_{\text{via slit 1}} + \langle x | s \rangle_{\text{via slit 2}} \quad (6.9)$$

Now we come to the *third general principle* of quantum mechanics, saying that if an electron follows a *specific* way, the amplitude for that way might be written as the *product* of the *amplitude* for some initial part of the way with the *amplitude* for the rest of the way. For example, if an electron in the system depicted in Fig. 6.1 starts from source s and arrives the detector at x going through slit 1, we write the amplitude for this event as the amplitude for starting from s and arriving at 1, multiplied by the amplitude for starting from 1 and arriving at x :

$$\langle x | s \rangle_{\text{via slit 1}} = \langle x | 1 \rangle \langle 1 | s \rangle \quad (6.10)$$

We should notice here that the both sides of this equation are to be read from left to right! Similarly, we can write the amplitude for the event in which an electron starts from s and arrives at x going through slit 2 as

$$\langle x | s \rangle_{\text{via slit 2}} = \langle x | 2 \rangle \langle 2 | s \rangle \quad (6.11)$$

Using (6.10) and (6.11), we can rewrite (6.9) in a more compact way as

$$\langle x | s \rangle_{\text{both slits open}} = \langle x | 1 \rangle \langle 1 | s \rangle + \langle x | 2 \rangle \langle 2 | s \rangle \quad (6.12)$$

We can make use of the three general principles of quantum mechanics stated above in considering a more involved multi-slit problem, like the one depicted in Fig. 6.2. We have again the same electron source s and the same wall with two slits. But this time we have another wall with three slits behind the first wall, and again the same detector at x .

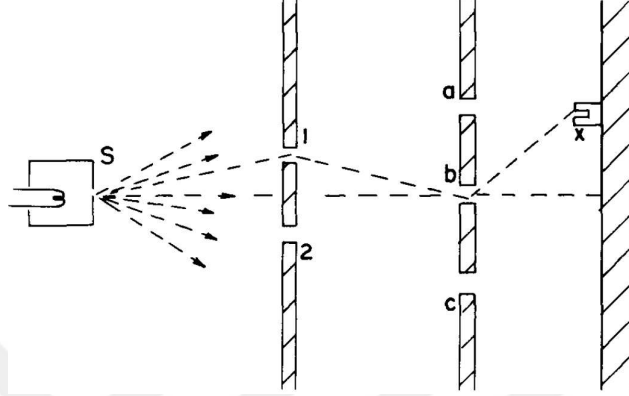


Figure 6.2: A multi-slit experiment; taken from Ref. [27].

An electron starting from s can arrive at x using six different alternatives: we have two initial alternatives from s to 1 or 2, and after the first wall we have another three alternatives for *each* particular way through 1 and 2. According to the above recipe, we can write the total amplitude for the event in which starts from the source s and reacts the detector at x as

$$\begin{aligned} \langle x | s \rangle = & \langle x | a \rangle \langle a | 1 \rangle \langle 1 | s \rangle + \langle x | b \rangle \langle b | 1 \rangle \langle 1 | s \rangle + \langle x | c \rangle \langle c | 1 \rangle \langle 1 | s \rangle + \\ & \langle x | a \rangle \langle a | 2 \rangle \langle 2 | s \rangle + \langle x | b \rangle \langle b | 2 \rangle \langle 2 | s \rangle + \langle x | c \rangle \langle c | 2 \rangle \langle 2 | s \rangle \end{aligned}$$

or, in a more concise way,

$$\langle x | s \rangle = \sum_{\substack{i=1,2 \\ \alpha=a,b,c}} \langle x | \alpha \rangle \langle \alpha | i \rangle \langle i | s \rangle \quad (6.13)$$

If we are to make calculations using the above principles, the first thing we should know is the amplitude for an electron (or, more generally, for a particle) from one point to another. Actually we already know of such an amplitude, from our previous discussions: the amplitude for a free particle. Let us suppose that a (nonrelativistic) free particle of mass m moves from a position $\mathbf{r}_1$ to another $\mathbf{r}_2$ in vacuum. no force

acts on it on its way. Its (linear) momentum is $\mathbf{p}$ so that its energy, that is is purely kinetic, is

$$E = \frac{p^2}{2m}$$

Apart from a numerical constant, the relevant amplitude for this motion is given as

$$\langle \mathbf{r}_2 | \mathbf{r}_1 \rangle = \frac{e^{i\mathbf{p} \cdot (\mathbf{r}_2 - \mathbf{r}_1)}}{\mathbf{r}_2 - \mathbf{r}_1} \quad (6.14)$$

What this equation states in effect is that the free particle inherits, from nature, wave-like properties (not a surprising fact at this very moment): it is the amplitude that propagates as a way whose wave number $\mathbf{k}$ is the (linear) momentum $\mathbf{p}$ of the particle divided by $\hbar$, or, in our previous notation, $\mathbf{p} = \hbar\mathbf{k}$.

6.2.1 Time Evolution of Amplitudes in Time [39]

Let us digress a little from the development of Dirac notation to touch, again, on the time evolution of probability amplitudes. A particle of rest mass m_0 in vacuum might have some specific energy. When it is at rest, for example, it will have only its rest energy: $E_0 = m_0c^2$, with c being the speed of light. For such a particle at rest, the (quantum mechanical) *amplitude* to find it at any point does *not* depend on position; it is the *same everywhere*. (This means, in turn, that the *probability* to find the particle anywhere is also the same everywhere.) Nevertheless, this amplitude *does* depend on the time *as its phase*. According to one of our previous discussions, for a particle in a state of certain energy E_0 , the amplitude to find it at any point (x, y, z) in space at time t is of the form

$$ae^{-i(E_0/\hbar)t} = ae^{-i\omega_0 t} \quad (6.15)$$

where a is a constant and

$$E_0 = \hbar\omega_0 = m_0c^2 \quad (6.16)$$

When we take the absolute square of the amplitude (6.15), we obtain the probability of finding the particle at point (x, y, z) , which is going to be *time-independent*. In other words, for a particle of certain specific energy, the amplitudes change with time only as an imaginary exponential and their absolute squares will remain always the same.²

² Actually this is the very reason why a particle in a state of certain definite energy is said to be in a *stationary state*.

It might appear strange to talk about a “particle” having the same amplitudes to be detected anywhere in space, because we usually think of a “particle” as a compact object that is pin down “somewhere.” As we have discussed previously, if a particle has a certain energy, its (linear) momentum is also certain, so there is no uncertainty in its momentum, $\Delta p = 0$. Then, according to the Heisenberg uncertainty principle, $\Delta p \Delta x \approx \hbar$, the uncertainty in the position of the particle must be infinite, $\Delta x \sim \infty$. This amounts to saying that the amplitude to find the particle at any point in space is the same everywhere.

If we are to speak of probabilities that change in time, we must have the interference of at least two amplitudes at two distinct frequencies, corresponding to two distinct energies, say, E_1 and E_2 . In such a case, we lose the information about energy; we cannot know it for certain. The particle will possess one amplitude to be found in a state with energy E_1 and another amplitude to be found in another state with energy E_2 .

As we shall see later in a concrete example, if the particle is somehow in a “condition” that consists of two distinct states with two distinct energies, then the corresponding amplitudes will change in time according to (6.15), as

$$e^{-i(E_1/\hbar)t} = e^{-i\omega_1 t} \quad \text{and} \quad e^{-i(E_2/\hbar)t} = e^{-i\omega_2 t} \quad (6.17)$$

and whenever we are to combine the relevant amplitudes, we will witness an interference; the probabilities will then *be* time-dependent.

Before passing, we would like to touch on another point which will be important when we discuss the propagation of an electron through a crystalline lattice structure. We notice that we are allowed to multiply all the amplitudes by the same exponential factor $e^{-i(A/\hbar)t}$, as:

$$e^{-i(E_1+A/\hbar)t} \quad \text{and} \quad e^{-i(E_2+A/\hbar)t} \quad (6.18)$$

This means that all possible linear combinations taking part in interferences will have this very same factor $e^{-i(A/\hbar)t}$. Whenever we take the absolute squares, the resultant probabilities will remain the same. This actually amounts in effect to the adding the same amount of energy A to the state energies E_1 and E_2 , and this addition will make no difference. We are now just using a different energy scale in which all the energy values—they are just numbers—are increased or decreased by the same constant

amount A . In other words, we are absolutely free to choose any origin for our energy scale, and our choice will not cause any difference at all on the final probability values.

So much for this little digression. Now we can return back to the Dirac notation used for discrete states, but this time in a much broader manner.

6.3 Dirac Notation for Discrete States (Continued) [40]

The following mathematical tools not only will be used in the rest of the manuscript, but they are also used widely in the literature. The serious reader is strongly recommended excel himself/herself in grasping and keeping them at hand just in case they will be needed in reading papers and books on the present subject.

We first remember from the above discussion that the probability amplitude to start from state ϕ and arrive at state χ might be written as a sum of over a set middle states i , of the same kind, of the amplitude to start from ϕ , next into a particular middle state i , and then to the final state χ , just as we did before for the double-slit experiment, as expressed in (6.12):

$$\langle \chi | \phi \rangle = \sum_{\text{all } i} \langle \chi | i \rangle \langle i | \phi \rangle \quad (6.19)$$

(From now on we call the middle states as the *base states* because of the reason that will be apparent shortly.) But the structure of this equation is similar to that of the equation for the dot product of two ordinary Cartesian vectors $\mathbf{B}$ and $\mathbf{A}$, given, though a little bit peculiar, as

$$\mathbf{B} \cdot \mathbf{A} = \sum_{\text{all } i} (\mathbf{B} \cdot \hat{\mathbf{e}}_i) (\hat{\mathbf{e}}_i \cdot \mathbf{A}) \quad (6.20)$$

where $\mathbf{e}_i$ represent the three usual unit (or *base*) vectors in the rectangular Cartesian coordinates,

$$\hat{\mathbf{e}}_1 \equiv \hat{\mathbf{x}}, \quad \hat{\mathbf{e}}_2 \equiv \hat{\mathbf{y}}, \quad \text{and} \quad \hat{\mathbf{e}}_3 \equiv \hat{\mathbf{z}}$$

and $\mathbf{B} \cdot \hat{\mathbf{e}}_i$ are seen to be

$$\mathbf{B} \cdot \hat{\mathbf{e}}_1 = \hat{\mathbf{e}}_1 \cdot \mathbf{B} = B_x, \quad \mathbf{B} \cdot \hat{\mathbf{e}}_2 = \hat{\mathbf{e}}_2 \cdot \mathbf{B} = B_y, \quad \text{and} \quad \mathbf{B} \cdot \hat{\mathbf{e}}_3 = \hat{\mathbf{e}}_3 \cdot \mathbf{B} = B_z$$

so that (6.20) is written in the more familiar form as

$$\mathbf{B} \cdot \mathbf{A} = B_x A_x + B_y A_y + B_z A_z$$

Now let us compare (6.19) and (6.20). We see that the two states χ and ϕ are counterparts to the two usual vectors $\mathbf{A}$ and $\mathbf{B}$, and that the *base states* i are counterparts to the rectangular unit (or *base*) vectors, $\hat{\mathbf{e}}_i = \{\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}\}$.

The analogy between (6.19) and (6.20) is obvious and complete. We know that we can express any vector $\mathbf{A}$ as a linear combination of the unit vectors, or, “base vectors” $\hat{\mathbf{e}}_i$. Whenever we specify the coefficients A_i , which are the projections of $\mathbf{A}$ onto the “base vectors” $\hat{\mathbf{e}}_i$, we already know everything that can be known about this vector. For instance, a vector can be specified uniquely as

$$\mathbf{A} = 3.4 \hat{\mathbf{x}} - 2.1 \hat{\mathbf{y}} + 0.9 \hat{\mathbf{z}}$$

with

$$A_x = 3.4, \quad A_y = -2.1, \quad \text{and} \quad A_z = 0.9$$

Similarly, speaking of quantum mechanics, any state ϕ can be represented as a linear combination of the base states i . Whenever the “coefficients” $\langle i | \phi \rangle$, which are the “projections” of the state ϕ onto the base states i , the state ϕ is completely and uniquely specified. Due to this close and strong resemblance, a quantum mechanical *state* is often referred to as a “state vector” and a *base state* as a “base vector.”

Let us summarize the analogy between the two pictures:

$$\begin{aligned} \text{vector } \mathbf{A} &\rightarrow \text{state } \phi, \text{ or } |\phi\rangle, \text{ or } \langle\phi| \\ \text{base vector } \hat{\mathbf{e}}_i &\rightarrow \text{base state } i, \text{ or } |i\rangle, \text{ or } \langle i| \\ \text{dot product } \mathbf{A} \cdot \mathbf{B} &\rightarrow \text{amplitude } \langle\phi|\chi\rangle \\ \text{coefficient } A_i = \hat{\mathbf{e}}_i \cdot \mathbf{A} &\rightarrow \text{coefficient } C_i = \langle i|\phi\rangle \end{aligned}$$

Because the base vectors are perpendicular to each other, we have

$$\hat{\mathbf{e}}_i \cdot \hat{\mathbf{e}}_j = \delta_{ij} \tag{6.21}$$

The corresponding relation for the quantum mechanical base states is

$$\langle i | j \rangle = \delta_{ij} \tag{6.22}$$

Because of this, we say that the base states are all *orthonormal*.

There is also a small difference between the dot product and (6.19): the dot product is *commutative*

$$\mathbf{A} \cdot \mathbf{B} = \mathbf{B} \cdot \mathbf{A}$$

but the corresponding quantum mechanical amplitude is not [see also (6.5)]:

$$\langle \phi | \chi \rangle = \langle \chi | \phi \rangle^* \quad (6.23)$$

The lesson to be learned here that in quantum mechanical calculations do *not* change the order of the terms, for most of the time they are crucially important.

Let us consider another oddly written but correct vector expression:

$$\mathbf{A} = \sum_i \hat{\mathbf{e}}_i (\hat{\mathbf{e}}_i \cdot \mathbf{A}) \quad (6.24)$$

Since $\hat{\mathbf{e}}_i \cdot \mathbf{A} = A_i$, the above expression is equivalent to

$$\mathbf{A} = \sum_i A_i \hat{\mathbf{e}}_i = A_x \hat{\mathbf{e}}_x + A_y \hat{\mathbf{e}}_y + A_z \hat{\mathbf{e}}_z \quad (6.25)$$

Comparing (6.24) with (6.20) we are tempted to say that someone has obtained (6.24) just by removing the “ $\mathbf{B} \cdot$ ” part from the both sides of (6.19) and that we can perform the same operation in (6.19) by removing the bra $\langle \chi |$ from the both sides of it. We end up with

$$| \phi \rangle = \sum_i | i \rangle \langle i | \phi \rangle \quad (6.26)$$

We have thus just divided the amplitude $\langle \chi | \phi \rangle$ into two and obtained the bra $\langle \chi |$ and the ket $| \phi \rangle$. It is because of the very structure of (6.26), which is identical to that of (6.25), that we call $\langle \chi |$ and $| \phi \rangle$ also as *state vectors*.

Most of the time we are asked to find the results of our quantum mechanical calculations as “numbers.” However, the result of (6.26) is *not* a number, it looks like a “vector equation.” With its present structure, it can be said to be an “unfinished” equation. If we want a number from this equation, we must multiply its both sides by a *bra* from the left, then it will have been “finished.”

Have a closer look again at (6.26). This equation is true for *any* state ϕ , so it also seems redundant, just like the state χ in (6.20). Why do we have to keep it? We can also abstract it away from (6.26) and we are left with

$$| = \sum_i | i \rangle \langle i | \quad (6.27)$$

Believe it or not, this is one of the greatest laws of quantum mechanics (no kidding, really it *is*). Besides, it has no counterpart in ordinary vector analysis. It is sometimes called the *identity operator*, implying that you can insert it wherever and whenever you like into any quantum mechanical equations. Although it is innocent looking, its meaning is much deeper: if the base states $|i\rangle$ are associated, via the Schrödinger equation, with a physical observable (with the linear momentum operator $\hat{p}$, for example), we say that the space (called the *Hilbert space*) relevant to this observable is “spanned” by these base states $|i\rangle$, just like the ordinary rectangular Cartesian space is spanned by the three base (or, unit) vectors $\hat{\mathbf{e}}_i = \{\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}\}$. Consequently, whenever (6.27) holds for an observable, we say that the base states $|i\rangle$ span the relevant (Hilbert) space *completely*; it is owing to this fact that (6.27) is sometimes referred to as the *completeness relation*.

Now let us go back to (6.26). We pointed out above that we can express any state vector $|\phi\rangle$ as a linear combinations of the base vectors $|i\rangle$ with the corresponding “coefficients” given by $\langle i | \phi \rangle$, which are most of the time *complex* numbers. We can write

$$\langle i | \phi \rangle = C_i$$

where we have deliberately used the letter C as a reminder that $\langle i | \phi \rangle$ are in general “complex” numbers. With this abbreviation, we rewrite (6.26) in a more meaningful form as

$$|\phi\rangle = \sum_i |i\rangle C_i \quad (6.28)$$

We may readily write a similar equation for another state vector $|\chi\rangle$ using different coefficients D_i as

$$|\chi\rangle = \sum_i |i\rangle D_i \quad (6.29)$$

where $D_i = \langle i | \chi \rangle$ are just the relevant amplitudes.

If we had removed first $|\phi\rangle$ from (6.19), we would have had

$$\langle \chi | = \sum_i \langle \chi | i \rangle \langle i | \quad (6.30)$$

But we know from (6.23) that $\langle \chi | i \rangle = \langle i | \chi \rangle^*$, so we can re-express the above equation as

$$\langle \chi | = \sum_i D_i^* \langle i | \quad (6.31)$$

Now if we multiply (6.31) and (6.28), we obtain back $\langle \chi | \phi \rangle$:

$$\langle \chi | \phi \rangle = \sum_{ij} D_j^* \langle j | i \rangle C_i \quad (6.32)$$

(It is very important to note that we have changed the dummy index i of (6.31) to j before multiplying. A golden rule says that never use the same indices in a multiplication.) According to the orthonormality of the base states, (6.22), we have $\langle i | j \rangle = \delta_{ij}$, so that the above equation becomes

$$\langle \chi | \phi \rangle = \sum_{ij} D_j^* \delta_{ij} C_i$$

Running first the index j we see that only the terms with $j = i$ survive:

$$\langle \chi | \phi \rangle = \sum_{ij} D_i^* C_i \quad (6.33)$$

Again there is a very close correspondence with the dot product

$$\mathbf{A} \cdot \mathbf{B} = \sum_i A_i B_i$$

with one minor difference that we have this time the complex conjugate D_i^* . This result is of course just our starting equation (6.19) expressed with new symbols.

Before passing to the next section it seems in order to stress that a vector $\mathbf{A}$ in the usual rectangular Cartesian coordinate system is expressed in terms of only *three* orthonormal unit vectors $\hat{\mathbf{e}}_i = \{\hat{\mathbf{x}}, \hat{\mathbf{y}}, \hat{\mathbf{z}}\}$, while a state vector $|\phi\rangle$ in its Hilbert space might be described by two, or four, or ten, or even an infinite number of base states $|i\rangle$.

6.3.1 Measurements in Dirac Notation [41]

We have already talked about measurements in quantum mechanics in Section 3.1.2 when as we have stated Postulate II. We repeat it here: if we measure a physical observable A on a system and if we obtain as a result the specific value a , this measurement process leaves the system in the definite state φ_a , which is the eigenfunction of $\hat{A}$ for the specific eigenvalue a . The underlying fact in this postulate is that the measurement process *changes* (actually *collapses*) the initial state (which is most of the time is uncertain, naturally, for us) into a (definite) final state (and leaves it there).

Suppose a system is initially in a state ϕ and we measure a physical observable A . What is the (probability) amplitude that after the measurement we find the system in a state χ . We can write the answer in the Dirac notation as

$$\langle \chi | \hat{A} | \phi \rangle \quad (6.34)$$

Such a use is somewhat special to quantum mechanical calculations, it does not have any analog in the usual vector algebra. Using twice the identity operator (6.27) in this expression, we can rewrite it in the form

$$\langle \chi | \hat{A} | \phi \rangle = \sum_{ij} \langle \chi | i \rangle \langle i | \hat{A} | j \rangle \langle j | \phi \rangle \quad (6.35)$$

If we make two successive measurements on the state ϕ , the first one for the observable A and the second one for another observable B , then the amplitude that the system is found after the measurement in a state χ is written as

$$\langle \chi | \hat{B} \hat{A} | \phi \rangle$$

which can be expressed, with the triple uses of the identity operator (6.27), as

$$\langle \chi | \hat{B} \hat{A} | \phi \rangle = \sum_{ijk} \langle \chi | i \rangle \langle i | \hat{B} | j \rangle \langle j | \hat{A} | k \rangle \langle k | \phi \rangle \quad (6.36)$$

Let us now consider (6.35) in a different way. Suppose, again, a system is initially in a state ϕ and we measure a physical observable A and the system is left in another state ψ . We can then ask: Is there a state ψ such that the amplitude to start from state ψ

and to arrive at state χ , i.e., the amplitude $\langle \chi | \psi \rangle$, is always the same as the amplitude $\langle \chi | \hat{A} | \phi \rangle$? Our answer is yes. What we want is to replace (6.35) with

$$\langle \chi | \psi \rangle = \sum_i \langle \chi | i \rangle \langle i | \psi \rangle \quad (6.37)$$

This can be obviously done if

$$\langle i | \psi \rangle = \sum_j \langle i | \hat{A} | j \rangle \langle j | \phi \rangle = \langle i | \hat{A} | \phi \rangle \quad (6.38)$$

which definitely determines ψ in components-wise manner, $\langle i | \psi \rangle = C_i$. (Whenever you determine the components A_x , A_y , and A_z of a Cartesian vector $\mathbf{A}$, you determine it uniquely.) Let us play a little with the last part of this equation:

$$\langle i | \psi \rangle = \langle i | \hat{A} | \phi \rangle = \sum_j \langle i | j \rangle \langle j | \hat{A} | \phi \rangle \quad (6.39)$$

Because this equation holds for all i , the bras $\langle i |$ on the leftmost part on its both sides might be taken off:

$$| \psi \rangle = \sum_j | j \rangle \langle j | \hat{A} | \phi \rangle \quad (6.40)$$

which simply says that state ψ is what we obtain if make a measurement for the observable A on the initial state ϕ .

Of course we can abstract away the identity operator from this equation and end up with

$$| \psi \rangle = \hat{A} | \phi \rangle \quad (6.41)$$

Actually this expression solidifies the meaning the word “operator”: whenever an “operator” $\hat{A}$ “operates on” a state $| \phi \rangle$ “from its left” as $\hat{A} | \phi \rangle$, we get a *new* state $| \psi \rangle$.

We can describe any operator $\hat{A}$ completely by giving the elements of the amplitude matrix

$$A_{ij} = \langle i | \hat{A} | j \rangle$$

in terms of any base states.

6.4 The Time Evolution of States in the Dirac Notation [42]

Suppose we have a system in a state ϕ at a time t_1 and let it sit until a later time t_2 . Because of the particular physical conditions in which the system sits (the system is perhaps in an electric or magnetic, or gravitational field) the initial state ϕ will change in time and will evolve into another state χ at t_2 . Because “sitting” can be considered to be a kind of measurement of an “observable” U (wait and *measure* to *see* that the initial state ϕ changed into a final state χ), we may describe what is happening, as it should be usual now, by an amplitude

$$\langle \chi | \hat{U}(t_2, t_1) | \phi \rangle \quad (6.42)$$

where $\hat{U}(t_2, t_1)$ is called the *time-evolution operator* [43]. Using the identity operator (6.27) twice, we represent this amplitude in terms of any basis states as

$$\langle \chi | \hat{U}(t_2, t_1) | \phi \rangle = \sum_{ij} \langle \chi | i \rangle \langle i | \hat{U}(t_2, t_1) | j \rangle \langle j | \phi \rangle \quad (6.43)$$

Thus we say that $\hat{U}$ is described completely by specifying the elements of the amplitude its matrix, the amplitude matrix,³

$$U_{ij} = \langle i | \hat{U}(t_2, t_1) | j \rangle \quad (6.44)$$

Let us think about the possible properties of the time-evolution operator $\hat{U}$. If we make two successive “measurements” for the observable U , the first from t_1 to t_2 , and the second from t_2 to t_3 , where $t_1 < t_2 < t_3$, these measurements should be equivalent to a single measurement from t_1 to t_3 . This means that

$$\hat{U}(t_3, t_1) = \hat{U}(t_3, t_2) \hat{U}(t_2, t_1) \quad (6.45)$$

In other words, if we are asked to analyze a given time interval, we can do this by analyzing shorter time intervals sequentially in between.

³ In the limiting case in which $t_1 \rightarrow -\infty$ and $t_2 \rightarrow +\infty$, the operator $\hat{U}(t_2, t_1)$ said to become the $\hat{S}$ operator and the amplitude matrix

$$S_{ij} = \langle i | \hat{S} | j \rangle$$

is called $\hat{S}$ -matrix.

We can now consider the time evolution for an infinitesimal time interval Δt . The question is: given an initial state $\phi(t)$ at time $t_1 = t$, what will it look like after an infinitesimal time Δt , at $t_2 = t + \Delta t$? The answer to this question is given in terms of the time-evolution operator as

$$|\phi(t + \Delta t)\rangle = \hat{U}(t + \Delta t, t) |\phi(t)\rangle \quad (6.46)$$

The amplitude of finding χ as the final state at $t + \Delta t$ is then

$$\langle \chi | \phi(t + \Delta t) \rangle = \langle \chi | \hat{U}(t + \Delta t, t) | \phi(t) \rangle \quad (6.47)$$

It seems better to work with a set of definite base states. Multiplying both sides of (6.46) by $\langle i |$, we obtain

$$\langle i | \phi(t + \Delta t) \rangle = \langle i | \hat{U}(t + \Delta t, t) | \phi(t) \rangle \quad (6.48)$$

and then decompose $|\phi(t)\rangle$ into the same base states as

$$\langle i | \phi(t + \Delta t) \rangle = \sum_j \langle i | \hat{U}(t + \Delta t, t) | j \rangle \langle j | \phi(t) \rangle \quad (6.49)$$

Now let

$$C_i(t) = \langle i | \phi(t) \rangle,$$

they are just (complex) numbers, be the amplitude that the initial state $\phi(t)$ is measured to be in the base state i . It varies with time, and it becomes after an infinitesimal time interval Δt that

$$C_i(t + \Delta t) = \langle i | \phi(t + \Delta t) \rangle$$

If we write the time-evolution operator also in matrix form,

$$U_{ij}(t + \Delta t, t) = \langle i | \hat{U}(t + \Delta t, t) | j \rangle,$$

these are also just (complex) numbers, we then cast (6.49) of the form

$$C_i(t + \Delta t) = \sum_j U_{ij}(t + \Delta t, t) C_j(t) \quad (6.50)$$

This is a great equation describing *how* a particular amplitude $C_i(t)$ changes with time; it states that a particular amplitude at a later time, $C_i(t + \Delta t, t)$, is directly proportional to the *all of the other* amplitudes $C_j(t)$, including itself, at the initial time t , multiplied by some particular coefficients $U_{ij}(t + \Delta t, t)$.

Until now we do not know the form of U_{ij} . But we are sure about one thing: in the limit $\Delta t \rightarrow 0$, we must just obtain the original state. This means that we must have

$$\lim_{\Delta t \rightarrow 0} U_{ij}(t + \Delta t, t) = \delta_{ij}$$

so that

$$C_i(t + \Delta t) = \sum_j U_{ij}(t + \Delta t, t) C_j(t) \rightarrow \sum_j \delta_{ij} C_j(t) = C_i(t)$$

Also, it is reasonable that for an infinitesimally small Δt , the individual coefficients U_{ij} each should deviate, from δ_{ij} , only a little amount, and that this deviation should be proportional to Δt . We these properties in our mind, we can write the form of U_{ij} as

$$U_{ij} = \delta_{ij} + K_{ij} \Delta t$$

However, it is an historical adoption to abstract the factor $(-i/\hbar)$ out of the (complex) coefficient K_{ij} to its front and to write this equation as⁴

$$U_{ij}(t + \Delta t, t) = \delta_{ij} - \frac{i}{\hbar} H_{ij}(t) \Delta t \quad (6.51)$$

A close look at this expression reveals that the terms H_{ij} are just equal to the factor $(-i/\hbar)$ times the time derivatives of U_{ij} at time t .

Feeding this form into (6.50) gives

$$\begin{aligned} C_i(t + \Delta t) &= \sum_j \left[\delta_{ij} - \frac{i}{\hbar} H_{ij}(t) \Delta t \right] C_j(t) \\ &= \sum_j \delta_{ij} C_j(t) - \frac{i}{\hbar} \sum_j H_{ij}(t) C_j(t) \Delta t \\ &= C_i(t) - \frac{i}{\hbar} \sum_j H_{ij}(t) C_j(t) \Delta t \end{aligned}$$

or

$$\frac{C_i(t + \Delta t) - C_i(t)}{\Delta t} = -\frac{i}{\hbar} \sum_j H_{ij}(t) C_j(t)$$

In the limit $\Delta t \rightarrow 0$, the left-hand side of this equation is recognized as the time-derivative of the amplitudes $C_i(t)$. Thus,

$$i\hbar \frac{dC_i(t)}{dt} = \sum_j H_{ij}(t) C_j(t) \quad (6.52)$$

⁴ Yes, you are right in your suspicion: the terms H_{ij} in (6.51) here are a herald of the Hamiltonian operator $\hat{H}$.

This is our target dynamical equation that governs how the amplitude $C_i(t)$ changes in time, which is, to repeat, the sum of all the other possible amplitudes $C_j(t)$, including itself, multiplied by some particular (complex) numbers $H_{ij}(t)$. But we remember that $C_i(t) = \langle i | \phi(t) \rangle$ is the amplitude of finding the *any* state ϕ , at time t , in a base state i . Consequently, (6.52) describes, equivalently, how *any* state ϕ changes in time.

6.5 The Hamiltonian (or, Energy) Matrix

It is also stated in (6.52) that it is $H_{ij}(t)$ that *dictates* the way $C_i(t)$ changes in time, for it includes the physical circumstances—in which the system under question resides—causing the system under question evolve in time. In other words, whenever we happen to know the elements H_{ij} , we have automatically a description, which is absolutely complete in every respect, for the time evolution of the system. We then have *everything* about the system, we can readily answer to every question about what will occur next in the system. Therefore, we have to know how to obtain the matrix elements H_{ij} for any system and for any circumstances. And this is the most difficult part. It is enough for the time being to acknowledge the fact that the full true $\hat{H}$ is not known for the whole cosmos.

As you have perhaps already anticipated, the coefficients H_{ij} are referred to as *the Hamiltonian matrix* or the *energy matrix*, or just *the Hamiltonian*. It has one special property that

$$H_{ij}^* = H_{ji} \quad (6.53)$$

This feature follows from the requirement that the *total* probability

$$\sum_i |C_i(t)|^2$$

for the system to be in *some* state does not vary with time. If this condition is to hold for any initial state ϕ , then (6.53) must also hold.⁵

⁵ We can prove (6.53) following the way the great master Feynman naively indicates. We have from (6.52) that

$$i\hbar \frac{dC_i(t)}{dt} = \sum_j H_{ij}(t)C_j(t) \quad \text{and} \quad -i\hbar \frac{dC_i^*(t)}{dt} = \sum_j H_{ij}^*(t)C_j^*(t)$$

The *total* probability for the system to be in *some* state does not vary with time:

$$\sum_i |C_i(t)|^2 = \text{const.}$$

Although it was somewhat difficult to reach (6.52), its application to real problems is not difficult at all. The good thing is for simple problems we do not have to the Hamiltonian $\hat{H}$ of the system—provided that it is independent of time—because the H_{ij} are just (complex) numbers! Let us consider a plain example in which the *external* physical conditions do not vary with time, so the Hamiltonian $\hat{H}$ is not dependent on time. We also chose a very simple system, so simple that we require only one single base state (say, “state 1”) for its description. This single state can readily be a first-order approximation to the hydrogen atom at rest. Equation (6.52) reduces to one single equation

$$i\hbar \frac{dC_1(t)}{dt} = H_{11} C_1(t) \quad (6.54)$$

so that

$$\begin{aligned} \frac{d}{dt} \sum_i |C_i(t)|^2 &= 0 \\ \sum_i \frac{d}{dt} |C_i(t)|^2 &= 0 \\ \sum_i \frac{d}{dt} C_i(t) C_i^*(t) &= 0 \\ \sum_i \left[\frac{dC_i(t)}{dt} C_i^*(t) + C_i(t) \frac{dC_i^*(t)}{dt} \right] &= 0 \\ \sum_i \left[i\hbar \frac{dC_i(t)}{dt} \cdot C_i^*(t) + C_i(t) \cdot i\hbar \frac{dC_i^*(t)}{dt} \right] &= 0 \\ \sum_i \left[\sum_j H_{ij}(t) C_j(t) C_i^*(t) - C_i(t) \sum_j H_{ji}^*(t) C_j^*(t) \right] &= 0 \\ \sum_i \sum_j H_{ij}(t) C_j(t) C_i^*(t) - \sum_i \sum_j C_i(t) H_{ji}^*(t) C_j^*(t) &= 0 \end{aligned}$$

In the second term, we change the roles of the indices i and j to get

$$\begin{aligned} \sum_i \sum_j H_{ij}(t) C_j(t) C_i^*(t) - \sum_j \sum_i C_j(t) H_{ji}^*(t) C_i^*(t) &= 0 \\ \sum_{ij} [H_{ij}(t) C_j(t) C_i^*(t) - C_j(t) H_{ji}^*(t) C_i^*(t)] &= 0 \\ \sum_{ij} \{ C_j(t) C_i^*(t) [H_{ij}(t) - H_{ji}^*(t)] \} &= 0 \\ H_{ij}(t) - H_{ji}^*(t) &= 0 \\ \therefore H_{ij}(t) &= H_{ji}^*(t) \quad (\text{QED}) \end{aligned}$$

Because H_{11} is just a constant, the solution is

$$C_1(t) = (\text{const.}) e^{-(i/\hbar)H_{11}t} \quad (6.55)$$

This gives the time variation of the single state whose energy is $E = H_{11}$. Although the amplitude $C_1(t) = \langle 1 | \phi \rangle$ for, say, the hydrogen atom to be in state 1 has a certain time-dependence, its absolute square

$$|C_1(t)|^2 = C_1^*(t) C_1(t) = |\text{const.}|^2$$

is not time-dependent; it is always constant. This is expected because the hydrogen atom is within an environment for which the *external* physical conditions do not change in time; there is no external agent on the hydrogen atom to cause it to change its state. (Actually we knew this in advance. If, at the outset, there were a possibility for the system to be in more than one base states, we must have included them into calculation at the very beginning.)

What if we have a system with two base states? Equation (6.52) now yields two separate equations, as

$$\begin{aligned} i\hbar \frac{dC_1(t)}{dt} &= H_{11} C_1(t) + H_{12} C_2(t) \\ i\hbar \frac{dC_2(t)}{dt} &= H_{21} C_1(t) + H_{22} C_2(t) \end{aligned} \quad (6.56)$$

As we remarked above we can readily solve these two equations if the Hamiltonian $\hat{H}$ of the system is independent of time again without knowing its explicit form! In the next chapter we will solve them in considering real problems.

CHAPTER 7

TWO-STATE PROBLEMS

7.1 A Fully Worked Example: The Ammonia Molecule [44]

As depicted in Fig. 7.1, an ammonia molecule consists of three hydrogen atoms that are placed in a triangular plane and one nitrogen atom that is below or above the plane of hydrogens.

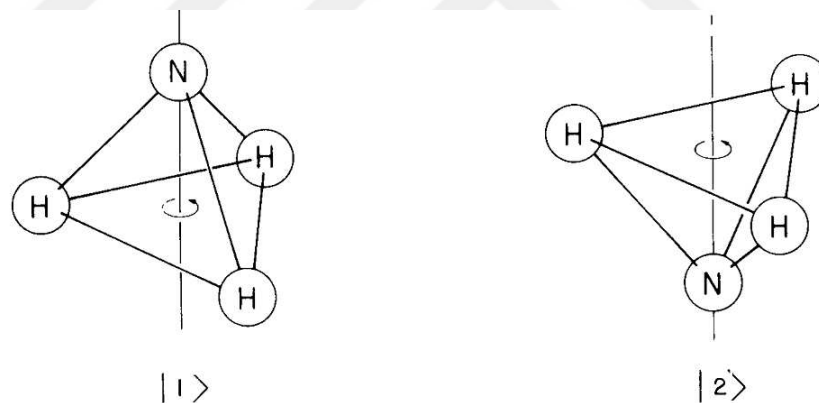


Figure 7.1: The two states of the ammonia molecule; taken from Ref. [45].

The ammonia molecule might have infinitely many states: it might be spinning around some axes; it might be moving along some directions; it might be vibrating; the possibilities go on and on. Consequently, it is absolutely not in any way a two-state system. Nonetheless, we are going to make an approximation in which we shall assume that

1. the molecule is spinning around its symmetry axis (passing through the nitrogen atom), as indicated in the figure;

2. it is not undergoing any translational motion;
3. it is not vibrating (or, barely vibrating); and
4. all the other unmentioned possible states are kept fixed (because they are not relevant to what we are considering).

All these settle down all but one circumstances about the molecule: *the nitrogen atom still has two probable positions*—it might be above or below the triangular plane containing hydrogen atoms. So we can consider the molecule to be a two-state system. (Even if the molecule is spinning with a particular momentum around the symmetry axis *and* it has a non-zero translational momentum, *and* it is vibrating in a certain way, it still has these two probable states. We have only two states to contemplate; all the other states are neglected without any risk.)

Let the molecule be in the state $|1\rangle$ when the nitrogen atom “up,” and in the state $|2\rangle$ when the nitrogen atom is “down,” as shown in Fig. 7.1. These two states, $|1\rangle$ and $|2\rangle$, are our base states. We also let $|\psi\rangle$ denote the *actual* state of the ammonia molecule, so the (complex) coefficients $C_1(t) = \langle 1 | \psi \rangle$ and $C_2(t) = \langle 2 | \psi \rangle$ give the amplitudes that the actual state $|\psi\rangle$ is in state $|1\rangle$ and in state $|2\rangle$, respectively. Then, using the identity operator (6.27) once, the state vector $|\psi\rangle$ is written as

$$|\psi\rangle = \sum_{i=1}^2 |i\rangle \langle i | \psi \rangle = |1\rangle \langle 1 | \psi \rangle + |2\rangle \langle 2 | \psi \rangle$$

or

$$|\psi\rangle = |1\rangle C_1 + |2\rangle C_2 \quad (7.1)$$

Now the crux of the problem is that if the molecule is definitely in some state, say $|1\rangle$, at some time, it will *not* stay at the same state at a later time; there is a certain non-zero, though small, probability that it will be observed in state $|2\rangle$.

The two coefficients C_1 and C_2 will change with time as the equations (6.56) dictate. We repeat them here:

$$\begin{aligned} i\hbar \frac{dC_1}{dt} &= H_{11} C_1 + H_{12} C_2 \\ i\hbar \frac{dC_2}{dt} &= H_{21} C_1 + H_{22} C_2 \end{aligned} \quad (7.2)$$

The solutions of these equations will give us how these amplitudes evolve in time. But we have a “small” problem here: we are not given the elements H_{ij} of the Hamiltonian matrix. Nevertheless, we *do* have some arguments to follow. Consider a case in which the molecule is in state $|1\rangle$ and there is no chance for it to reach state $|2\rangle$, and vice versa. In these cases the matrix elements H_{12} and H_{21} will be zero and equations (7.2)

$$i\hbar \frac{dC_1}{dt} = H_{11} C_1 \quad \text{and} \quad i\hbar \frac{dC_2}{dt} = H_{22} C_2$$

Their solutions are

$$C_1(t) = (\text{const.}) e^{-(i/\hbar)H_{11}t} \quad \text{and} \quad C_2(t) = (\text{const.}) e^{-(i/\hbar)H_{22}t} \quad (7.3)$$

These are merely the time evolutions of the amplitudes for the two *stationary* states whose energies are $E_1 = H_{11}$ and $E_2 = H_{22}$. There is however one more thing about these energies: as Fig. 7.1 also suggests, the two states $|1\rangle$ and $|2\rangle$ of the ammonia molecule possess a well-defined symmetry, and it follows directly from this definite symmetry that their energies H_{11} and H_{22} must be equal. Let E_0 denote them both, since they are the energies of the two states for the case in which the matrix elements H_{12} and H_{21} are zero:

$$H_{11} = H_{22} = E_0$$

so that

$$C_1(t) = (\text{const.}) e^{-(i/\hbar)E_0t} \quad \text{and} \quad C_2(t) = (\text{const.}) e^{-(i/\hbar)E_0t}$$

The above treatment, with $H_{12} = H_{21} = 0$, does not account for what the ammonia molecule actually does. It turns out that the “up” nitrogen atom in state $|1\rangle$ happens to move somehow through the planes of hydrogen atoms and appears at the other opposite side to become the “down” nitrogen atom in state $|2\rangle$. But this process must be very difficult, for even reaching the middle hydrogen-plane demands a very high energy; the situation gets worse if the nitrogen atom does not have sufficient energy to do this. But the nitrogen atom behaves in spite of, almost in defiance of, all these considerably serious hardships. As we have seen before in the discussion of the bound states of the delta-potential well in Section 5.1, quantum mechanics allows a particle to *sneak* into an energetically forbidden region; it is kind of illicite region from the classical point of view. The conclusion to be drawn is that there exist some non-zero amplitude that an ammonia molecule initially in state $|1\rangle$ will be found at a later time

in state $|2\rangle$. Therefore, the matrix elements H_{12} and H_{21} must not be zero. Also, because of the symmetry of the two states, again, these elements must be equal to each other, in magnitude at least:¹

$$H_{12} = H_{21} = -A$$

where we have equated them to a negative number for later convenience. We all these, the complete temporal description of the ammonia molecule is given, via (7.2), by the two differential equations

$$\begin{aligned} i\hbar \frac{dC_1}{dt} &= E_0 C_1 - A C_2 \\ i\hbar \frac{dC_2}{dt} &= -A C_1 + E_0 C_2 \end{aligned} \quad (7.4)$$

To solve these two *coupled* equations, we first take their sum and then their difference, and obtain

$$\begin{aligned} i\hbar \frac{d}{dt} (C_1 + C_2) &= (E_0 - A) (C_1 + C_2) \\ i\hbar \frac{d}{dt} (C_1 - C_2) &= (E_0 + A) (C_1 - C_2) \end{aligned}$$

Now these two are elementary differential equations; their respective solutions are

$$\begin{aligned} C_1 + C_2 &= a e^{-(i/\hbar)(E_0 - A)t} \\ C_1 - C_2 &= b e^{-(i/\hbar)(E_0 + A)t} \end{aligned} \quad (7.5)$$

where a and b are integration constants (which are determined by the initial conditions according to the wording of the problem). By first adding and then subtracting these, we obtain the desired expressions for C_1 and C_2 as

$$\begin{aligned} C_1(t) &= \frac{a}{2} e^{-(i/\hbar)(E_0 - A)t} + \frac{b}{2} e^{-(i/\hbar)(E_0 + A)t} \\ C_2(t) &= \frac{a}{2} e^{-(i/\hbar)(E_0 - A)t} - \frac{b}{2} e^{-(i/\hbar)(E_0 + A)t} \end{aligned} \quad (7.6)$$

They are the same expression apart from the negative sign in front of the second term.

¹ Actually this equality can be seen to follow also directly from (6.53). If the property $H_{ij}^* = H_{ji}$ is to hold, the elements H_{ij} and H_{ji} are allowed to differ only by a phase.

We have two solutions at hand. What about their meanings? We first note that when $b = 0$, both C_1 and C_2 will “oscillate” with the same frequency $\omega = (E_0 - A) / \hbar$. This means that everything in the ammonia molecule evolve at one single frequency, and that it is in a state of definite energy, given by $(E_0 - A)$. Thus, this is a *stationary state* in which C_1 and C_2 are equal. If we remember the meaning of the amplitudes C_1 and C_2 , we conclude that the ammonia molecule possesses a well-defined energy $(E_0 - A)$ if the amplitudes that the nitrogen atom is found to be in the “up” and “down” positions are equal.

When $a = 0$ in (7.6), both amplitudes C_1 and C_2 again oscillate with the same frequency $\omega' = (E_0 + A) / \hbar$, but this time $C_2 = -C_1$. Therefore, we have another stationary state with a definite energy $(E_0 + A)$.

We draw the conclusion that (1) there is some non-zero probability that the nitrogen atom is able to flip from one side to the other opposite site at will; (2) the energy of the ammonia molecule is not just E_0 , as might be expected; on the contrary (3) there are now two *stationary states* with definite energies $(E_0 - A)$ and $(E_0 + A)$. We say that the particular state of rotation, with energy E_0 , has been *splitted* into two states, one is greater than E_0 , the other one is less than E_0 .

Actually, every possible state that might be considered about the ammonia molecule (like translation, vibration, etc.) *splits* into two states. We say that whenever the flip-flop of the nitrogen atom is taken into account, there is a *doublet* of energy levels for every possible condition.

Suppose now that we know *for certain* that the ammonia molecule is initially in state $|1\rangle$ at $t = 0$, so that we have $C_1(0) = 1$ and $C_2(0) = 0$. With these, we can determine the constants a and b in (7.6) at $t = 0$ as

$$C_1(0) = \frac{a+b}{2} = 1 \quad \text{and} \quad C_2(0) = \frac{a-b}{2} = 0$$

whence $a = b = 1$. Insertion of these into (7.6) and some arrangements lead to

$$C_1(t) = e^{-(i/\hbar)E_0 t} \left(\frac{e^{(i/\hbar)At} + e^{-(i/\hbar)At}}{2} \right)$$

$$C_2(t) = e^{-(i/\hbar)E_0 t} \left(\frac{e^{(i/\hbar)At} - e^{-(i/\hbar)At}}{2} \right)$$

which can be written in more compact forms as

$$C_1(t) = e^{-(i/\hbar)E_0 t} \cos \frac{At}{\hbar} \quad (7.7)$$

$$C_2(t) = e^{-(i/\hbar)E_0 t} \sin \frac{At}{\hbar} \quad (7.8)$$

The absolute square of the amplitude $C_2(t)$ gives the probability for the ammonia molecule to be found in state $|2\rangle$:

$$|C_2(t)|^2 = \sin^2 \frac{At}{\hbar} \quad (7.9)$$

In Figure 7.2 we plot the probability curves, with $P_1(t) = |C_1(t)|^2$ and $P_2(t) = |C_2(t)|^2$ being the respective probabilities that the molecule is found in state $|1\rangle$ or $|2\rangle$. We see that $P_2(t)$ at $t = 0$ is zero, as it should be, then increases and becomes unity at $t = (\pi/2)\hbar/A$, then decreases, and becomes again zero at $t = \pi\hbar/A$; then repeats this behaviour, oscillates back and forth between zero and unity. The same oscillating feature is seen for $P_1(t)$, but exactly in the opposite manner: whenever $P_2(t)$ is zero, $P_1(t)$ is unity, and vice versa; whenever $P_2(t)$ increases, $P_1(t)$ decreases simultaneously, and vice versa.

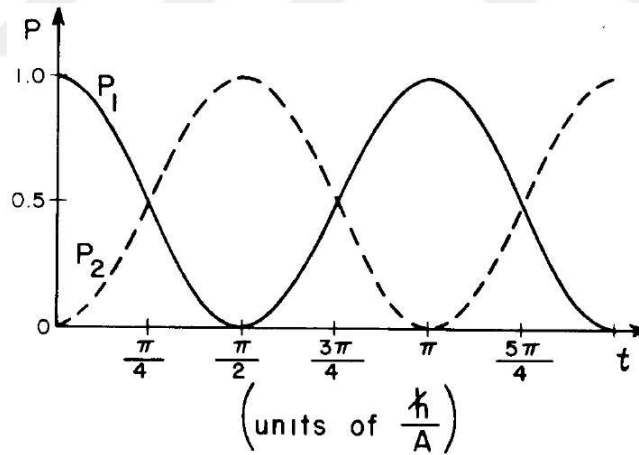


Figure 7.2: Probability curves for the ammonia molecule. P_1 (P_2) is the probability that the molecule is found in state $|1\rangle$ ($|2\rangle$). Taken from Ref. [46].

The behaviour of the probability curves in Fig. 7.2 reminds us the behaviour of two identical pendula coupled slightly [47]. When one of them is lifted back and set free, it swings. The other one is initially at rest, but it little by little starts to swing also. After a while, the second pendulum accumulates all the energy and swings largely, at

the same time the first one has been lost all of its energy and comes to a halt. Then the whole process is reversed and the first pendulum accumulates all the energy again. The energy is swapped back and forth at a rate that is dependent on the strength of the coupling between the two pendula—a measure of “oscillation leaking” between them. We remember one more thing about the two coupled pendula experiment that they have two especially particular motions, each have a well-defined frequency: if both pendula are lifted to the left, for example, and then set free, they will swing together at one special frequency; if one of them is lifted to the left and the other to the right, and then set free, they will swing one toward the other, and then away from each other, again simultaneously at another special frequency. These two particular motions are the “stationary states” of the two coupled pendula, and are referred to as the *fundamental notes*.

We observe an analogous situation in the ammonia molecule. Speaking mathematically, the molecule behaves like a pair of coupled identical pendula whose “fundamental note” frequencies are $(E_0 \pm A) / \hbar$. If the nitrogen atom is pulled to one side, we then witness a *superposition* of these two fundamental notes. (All the possible motions of the coupled pendula system can be expressed as a linear combinations of its fundamental notes.)

The similarity between the ammonia molecule and a pair of couple pendula is not deeper than that explained above. The phenomenon of the splitting of energy levels into two in the ammonia molecule is purely a quantum mechanical result.

At long last, we have seen an example of real physical problem that we were able to understand only with the power of quantum mechanics!

7.2 The Hydrogen Molecular Ion [48]

The hydrogen molecular ion, H_2^+ , consists of two protons and one electron wandering about them. If the distance between the two protons is very large, we can speak of two states for H_2^+ : in one of these states, the electron will linger near one proton, say, proton 1, to constitute an isolated hydrogen atom² and the other proton, proton 2, will stay as an isolated positive ion; in the other *symmetric* state, the electron will be near proton 2 and proton 1 will stay as a positive ion. We take these two as the base states for the H_2^+ molecule, and designate them as $|1\rangle$ and $|2\rangle$. They are depicted in Fig. 7.3.

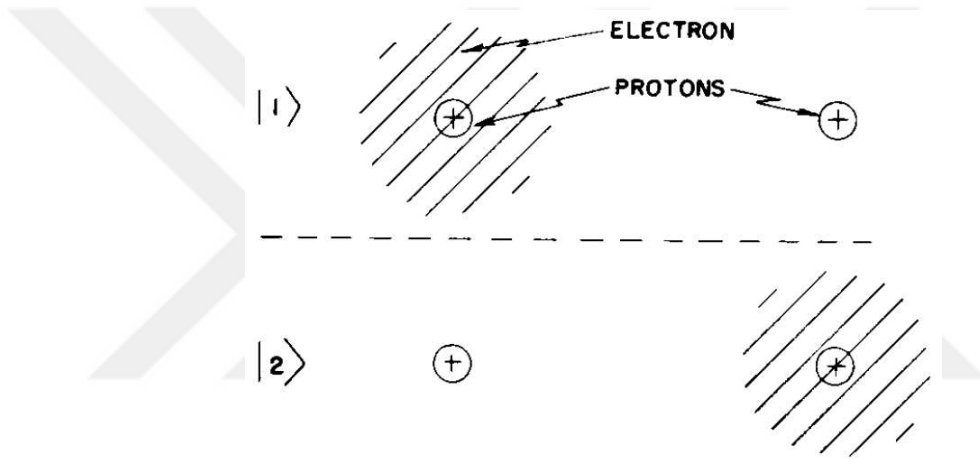


Figure 7.3: The base states for the H_2^+ molecule; taken from Ref. [49].

Actually we can speak of infinitely many number of states for the H_2^+ molecule. But except for $|1\rangle$ and $|2\rangle$, all the others are for the excited states of the molecule. We here confine the discussion to the hydrogen atom in its ground state. Also, for the time being, the spin of the electron will be neglected because it has no effect for the present problem (provided that the H_2^+ molecule is not within a magnetic field).

² We can say that the isolated hydrogen atom in this case is found in its lowest 1s-state because the other proton is too far apart from it to affect it.

We know that in order to take the single electron off a hydrogen atom, a 13.6-eV energy is required. In our case in which the two protons are very apart from each other, to bring the electron to the middle point of the imaginary line connecting the two protons still requires about this amount of energy. In other words, just like the nitrogen atom in the ammonia molecule, it is (classically) a very demanding task for the electron to get away from one proton and to reach and “attached” to the other proton. But quantum mechanics says that this is perfectly possible, although the relevant probability is very small. We have a small, but non-zero, amplitude that the electron starts from one proton and get to the other. Therefore, as a first approximation, both of the base states $|1\rangle$ and $|2\rangle$ will have the same energy E_0 , which is the energy of the hydrogen atom in the presence of the isolated positive ion. This means that we can take the Hamiltonian (energy) matrix elements H_{11} and H_{22} to amounts approximately to E_0 .

$$H_{11} = H_{22} = E_0$$

We take the other elements H_{12} and H_{21} , the amplitudes that the electron flip-flops between the two protons, to be equal again to $-A$:

$$H_{12} = H_{21} = -A$$

If we did not allow, somehow, the electron to flip-flop between the protons, we would have the two states $|1\rangle$ and $|2\rangle$ with the same energy E_0 . But whenever the electron is allowed to go back and forth, this energy level will be splitted into two as $E_0 \pm A$. We see that the greater the amplitude A that the electron flip-flops, the greater the separation in the splitting. We write these two relevant states with definite energies as [see also (7.5)]

$$\begin{aligned} |I\rangle &= \frac{1}{\sqrt{2}}(|1\rangle - |2\rangle) \quad \text{with} \quad E_I = E_0 + A \\ |II\rangle &= \frac{1}{\sqrt{2}}(|1\rangle + |2\rangle) \quad \text{with} \quad E_{II} = E_0 - A \end{aligned} \tag{7.10}$$

where we have now *normalized* them.

It follows from our solution that the electron in a H_2^+ molecule does not stay around one proton only, but will move back and forth between the protons. If the molecule is initially in state $|1\rangle$, for example, it will “swing” between the two states $|1\rangle$ and $|2\rangle$, and the relevant solution, like the one in (7.9), will *be* time-varying. If we are to have the solution with the lowest energy $E_{II} = E_0 - A$, the molecular must be in the *stationary state* $|II\rangle = (|1\rangle + |2\rangle) / \sqrt{2}$, which does *not* change with time. This means, in turn, that the *only* electron in the molecule must spend equal times, with the probability amplitudes $C_1 = C_2 = 1 / \sqrt{2}$, or with the probabilities $|C_1|^2 = |C_2|^2 = 1/2$, near each proton.³

Now the amplitude A that the electron moves back and forth between the protons must be dependent on the distance between them: the smaller this distance, the greater the amplitude A . This is something that is naturally expected because it is easier for the electron to move through a shorter distance. As we have remarked above, if an electron is to reach another proton, it must “penetrate” the classically forbidden energy barrier. Although quantum mechanics allows this, it is still a very demanding task for the electron. And this difficulty escalates with the increasing separation of the protons. In other words, the amplitude A that the electron “penetrates” through the barrier towards the other proton decreases, somehow exponentially, with the increasing distance. In this case the separation of the energy levels E_I and E_{II} will diminish gradually to zero. In the reverse case, when the separation between the protons is smaller, the amplitude A will be larger and the amount of splitting between the energy levels E_I and E_{II} , which is equal to $2A$, will be also larger.

Let D be the distance between the two protons in the H_2^+ molecule. According to the preceding discussion, if the molecule is in state $|I\rangle$, its energy $E_I = E_0 + A$ will increase as the distance D decreases. The net effect, which is a purely quantum mechanical effect, will be a *repulsive* force which causes the two protons to stay *apart*. In the reverse case, if the molecule is in state $|II\rangle$, its energy $E_{II} = E_0 - A$ will decrease as the distance D decreases. This time, there will be an *attractive* force which causes the two protons to stay *closer*. In Fig. 7.4 we give the rough plots of

³ You may oppose here that the other stationary state $|I\rangle = (|1\rangle - |2\rangle) / \sqrt{2}$ with a higher energy $E_I = E_0 + A$ describes the same situation in which the electron spends equal time, with the probability amplitudes $C_1 = -C_2 = 1 / \sqrt{2}$, or again with the probabilities $|C_1|^2 = |C_2|^2 = 1/2$, near each proton. Just wait a little for an additional explanation that you will see shortly.

these two stationary-state energies E_I and E_{II} of the H_2^+ molecule as a function of the separation of the protons.

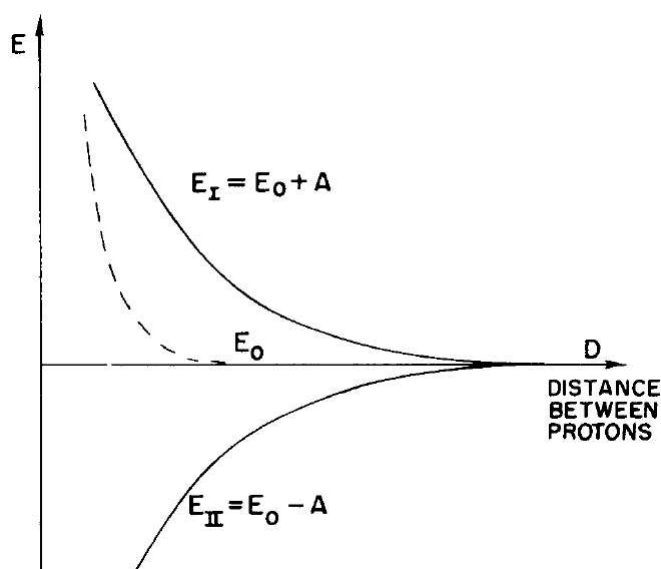


Figure 7.4: The stationary-state energies E_I and E_{II} of the hydrogen molecular ion, plotted against the distance D between its two protons, for the case in which the repulsive proton-proton electrostatic energy is *not* included. Taken from Ref. [50].

But there is one more force we must take into consideration as well as the (*attractive* or *repulsive*) force we have just talked about. It is the electrostatic Coulomb force between the protons, which is also *repulsive*. When the separation between the two protons is very large, the negatively charged electron shields the positively charged proton so efficiently that the “isolated” proton beholds only an electrically neutral hydrogen atom. Therefore, at large proton distances, this proton-proton electrostatic force is not significant, so negligible. On the other hand, at very small proton distances (so small that the “isolated” proton is, most of the time, much more closer to the other proton than it is to the electron), the no longer “isolated” proton is now within the electron cloud distribution. Since at this distances the negatively charged electron cannot effectively shield the positively charged proton, the proton-proton electrostatic force *is* now profoundly significant, so cannot be neglected at all. As a result, there appears some *extra* electrostatic energy that is positive. This energy is, of course, also changes with the proton separation D , and must be incorporated into E_0 . To this end, we describe E_0 like to one represented by the broken-line curve

in Fig. 7.4 which increases quickly at distances less than “effective radius” of the hydrogen atom. After the transition energy A is added to and subtracted from this newly considered E_0 , the stationary energies E_I and E_{II} takes the new forms as depicted in Fig. 7.5, which is highly detailed. In this figure the distance between the two protons is given in $\text{\AA}$ ($= 10^{-10} \text{ m}$) and the *excess* energy ΔE of the molecule (the repulsive proton-proton electrostatic energy plus the energy of a single hydrogen atom) is given in units of “Rydbergs,” $E_H = 13.6 \text{ eV}$, which is the *binding energy* (or, *bond energy*) of the hydrogen atom.

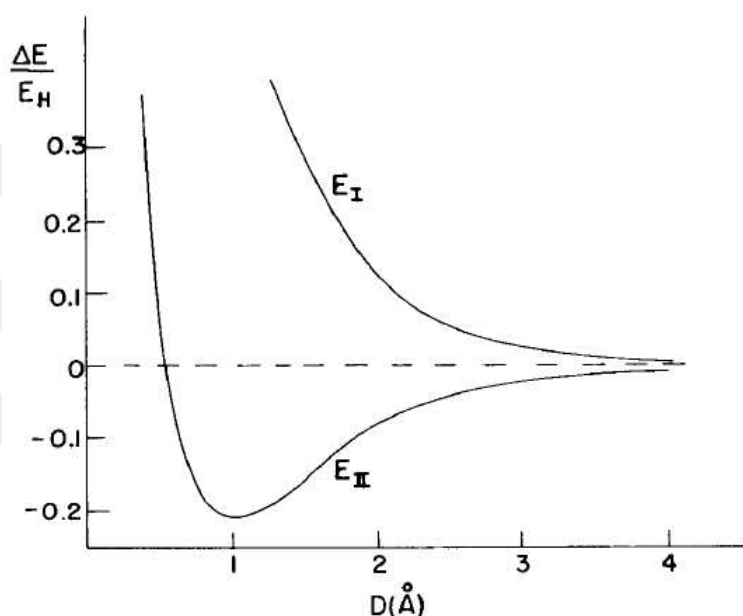


Figure 7.5: The same as Fig. 7.4, but this time for the case in which the repulsive proton-proton electrostatic energy *is* included. Taken from Ref. [50].

It is seen from Fig. 7.5 that state $|II\rangle$ possesses an energetically minimum point, which corresponds to the equilibrium configuration of the H_2^+ molecule. The energy of the molecule at this special point is less than that of an isolated proton and a hydrogen atom, so we say that “the system is bound.” The single electron here acts as a “mediator” in holding the two protons together. This type of bond would be called a “one-electron bond” by a chemist.

Until so far, we have seen a simple and equally beautiful quantum mechanical description for the concept of binding in the hydrogen molecule ion. The golden rule crowning our efforts is: “every system in the universe always seeks to minimize its energy.”

Now we have come to the crux of the matter at hand. Why does state $|II\rangle$ have a lower energy than that of an isolated proton and a hydrogen atom? To resolve this point, at least *plausibly*, we first remember that an electron near a proton is “spread out.” This is because of the uncertainty principle, $\Delta p \Delta x \approx \hbar$: the electron looks for a balance between reaching a low (Coulombic) *potential* energy and not getting restricted into a too little space ($\Delta x \sim 0$), which results in a large *kinetic* energy ($\Delta p \sim \infty$ so that $T = p^2/2m$ is large). This is for a single electron around a single proton. We next consider a situation in which there are two fixed protons (the distance between them is not too large) and there is an electron near them. Now the single electron has more space to lower its potential energy: it can readily “spread out” to reduce its kinetic energy without any danger of increasing its potential energy. The end result is that we have a hydrogen molecular ion H_2^+ that has a less energy than a hydrogen atom H does.

Next comes the second seemingly subtle issue that awaiting an explanation for long: why does the other state $|I\rangle$ possesses a higher energy? This state is given as, to repeat, $|I\rangle = (|1\rangle - |2\rangle) / \sqrt{2}$ so it is the *difference* of the two base states $|1\rangle$ and $|2\rangle$, each of which has the 1s *spherical* symmetry around the protons. Because of this symmetry and the difference of states, the amplitude that the electron is found at the middle point of the (imaginary) line connecting the two protons must be zero. This means, in turn, that the single electron is much more *restricted* (it might be near one proton with a probability 1/2, or near the other proton with the same probability, but never between them), which results in a larger energy, again due to the uncertainty principle.

7.3 The Hydrogen Molecule [51]

The neutral hydrogen molecule, H_2 , contains two protons and two electrons. Assuming for the time being that they are *distinguishable*, we label the two electrons as “electron a ” and “electron b .” We again think of two possible base states to work with. If the distance between two two protons is very large, like two separate hydrogen atoms, “electron a ” might be found near the first proton and “electron b ” near the second. Let the ket $|1\rangle$ denote this state. In the other possible state, “electron b ” might be near the first proton and “electron a ” near the second; we let the ket $|2\rangle$ denote this state. Figure 7.6 shows these two base states. Of course, there are infinitely many other states describing the other possible situations. For example, “electron a ” and “electron b ” might be near the *same* proton. But all these belong to energetically higher “excited” states. (If two electrons are close to the same proton, there will be a very large, positive, repulsive Coulomb potential energy.) In order to attain a greater accuracy, we should obtain all of these excited states, but the essential results of the problem will not change. Therefore, we will be content with the use of only two states $|1\rangle$ and $|2\rangle$.

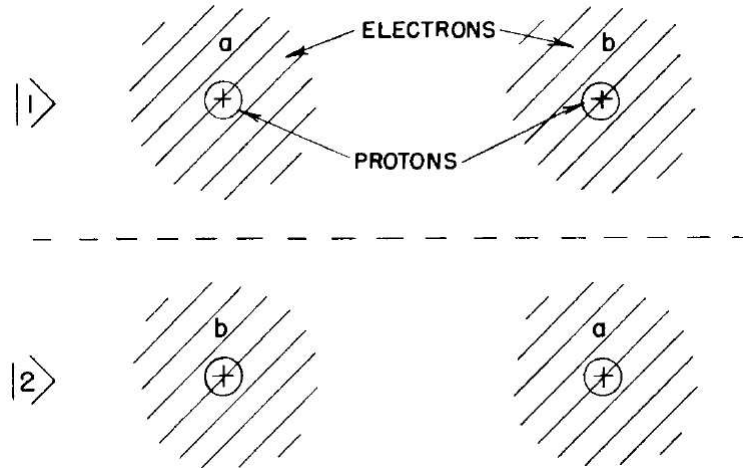


Figure 7.6: The two bases states for the hydrogen molecule; taken from Ref. [52].

Because of the apparent symmetry of the system, these two base states has identical energies, but, as will be seen soon, the (total) energy of the hydrogen molecule is *not* the sum of that of two hydrogen atoms.

In our two-state approximation, we can write any state vector $|\phi\rangle$ in the usual linear combination form as

$$|\phi\rangle = \sum_i |i\rangle \langle i|\phi\rangle$$

with, again, the usual interpretation that the amplitudes $\langle 1|\phi\rangle$ and $\langle 2|\phi\rangle$ give the amplitude that the system is state $|1\rangle$ and $|2\rangle$, respectively. Also, there will be an amplitude A for the electrons to exchange their places by going through all the intervening space. As we have seen in the hydrogen molecular ion case above, this likelihood of exchange will cause the energy of the system to be split. If the two protons of the system are well apart, this splitting is only a little. As the separation between the protons increases, the amplitude that the electrons move back and forth will increase, so will the splitting. There will be an attractive force holding the atoms together; this will correspond to the decrease seen in the *lower* energy E_{II} (not the upper energy!). Owing to the repulsive Coulombic repulsion between the protons, the energy levels, *both* the lower and upper energies, will increase as the separation between the protons decreases to very small values. The final result is that the two *stationary* states will have energies that changes with the distance D between the protons as depicted in Fig. 7.7 (where $E_H = 13.6$ eV). It is seen that at $D \approx 0.74$ Å (this is the real proton-proton separation in the real hydrogen molecule), the *lower* energy level assumes its minimum value.

At this point we should return back to one of our initial assumptions. We assumed that the two electrons in the hydrogen atom were distinguishable. But this is *wrong*: we cannot simply label one of the electrons as “a” and the other “b”. We cannot paint them as the one being “red” and the other “blue.” We have absolutely now way to distinguish which is which. We must resolve this subtle issue in our treatment of the hydrogen molecule.

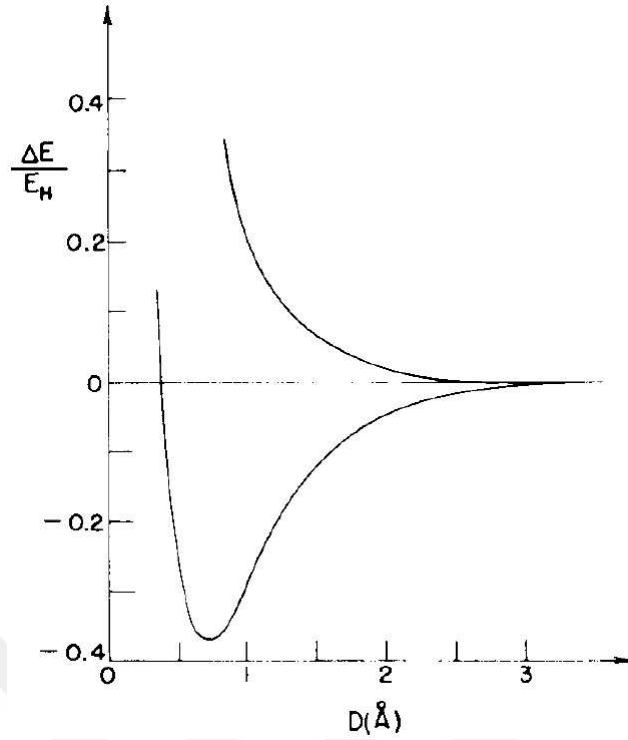


Figure 7.7: The stationary-state energies of the hydrogen molecular, plotted against the distance D between its two protons, where the repulsive proton-proton electrostatic energy is included. Taken from Ref. [52].

Electrons are referred to as *Fermi particles*. If there exist two alternative ways for a physical phenomenon can take place by exchanging Fermi particles (or electrons as in the H_2 molecule), the state of the system must then be *antisymmetric*: if we exchange the labels of the electrons in a system, the sign of its state must reverse with a negative sign [53].

But we have just conjectured in the above discussion that the bound state of the H_2 molecule would be, at $t = 0$,

$$|II\rangle = \frac{1}{\sqrt{2}}(|1\rangle + |2\rangle)$$

However, due to the preceding antisymmetry discussion, this state cannot describe the H_2 molecule. When we exchange the labels of the electrons, we obtain

$$\frac{1}{\sqrt{2}}(|2\rangle + |1\rangle)$$

which is again the original state $|II\rangle$ with *no* sign change.

These are all correct if the two electrons happen to have the same spin (both spin-up or spin-down). It is also correct that for such two electrons with the same spin, the only permissible state is

$$|I\rangle = \frac{1}{\sqrt{2}}(|1\rangle - |2\rangle)$$

for which the exchange of two electrons leads to

$$\frac{1}{\sqrt{2}}(|2\rangle - |1\rangle)$$

which equals $-|I\rangle$, as the antisymmetry condition requires. Thus, if two H atoms with their electrons having the same spin are brought close to each other to form an H_2 molecule, they can enter *only* into state $|I\rangle$ and *not* into state $|II\rangle$. But we have seen that state $|I\rangle$ corresponds energetically to the *upper* state and that its energy-separation graph does not have any minimum at all. The two H atoms will continuously repel each other and they will not produce a molecule. Therefore, we draw the conclusion that it is simply impossible to obtain an H_2 molecule with the same electron spins. And this is correct.

On the other hand, if we are to bring two H atoms to get a H_2 molecule, we must then insist on state $|II\rangle$, which is, however, completely symmetric for the exchange of two electrons.⁴ The remedy is simple: the spins of the electrons *must* be opposite, one spin-up and the other is spin-down⁵, and that is all!⁶

Let us recapitulate the results of above discussions: (1) Energetically the lowest-lying state—which is the only bound state—of the hydrogen molecule has two electrons whose spins are directed in opposite directions; (2) therefore, the net spin angular momentum (of the two electrons) is zero.⁷ (3) In the case of two nearby hydrogen atoms with electrons having the same spins, the system must be in the energetically higher state—which is the unbound state in which the two atoms repel each other; and (4) for this case (the magnitude of) the net spin angular momentum is $\hbar$.⁸

⁴ We say then that the *spacious part* of the molecular state is *symmetric*.

⁵ We say then that *spin part* of the molecular state is *antisymmetric*.

⁶ The total molecular state is the product of the spacious and the spin parts. The so-obtained total molecular state is now *antisymmetric*, as it must be.

⁷ $\left(+\frac{\hbar}{2}\right) + \left(-\frac{\hbar}{2}\right) = 0$

⁸ $\left(+\frac{\hbar}{2}\right) + \left(+\frac{\hbar}{2}\right) = +\hbar$, or $\left(-\frac{\hbar}{2}\right) + \left(-\frac{\hbar}{2}\right) = -\hbar$

There is an interesting yet thought-provoking connection between spins and energies: as we have just witnessed in the H_2 molecule problem, there appears that two spins “interacts” somehow, since a system of two opposite spins has a lower energy than that of two parallel spins. We may say that two parallel spins will always try to reach the antiparallel-spins condition; when they accomplish this, there will be an automatic energy liberation so that the total energy of the system will have been reduced. The intriguing thing is that these all result from the *exclusion principle*, not from any large magnetic force between the two spins!



CHAPTER 8

PROPAGATION OF AN ELECTRON IN A CRYSTAL LATTICE¹

It might seem at first a very demanding task for a (low-energy) electron to pass through a 3D crystal lattice. The atoms in such a structure are packed together rather tightly, the nearest-neighbor distance between them is only a few angstroms. The effective diameter of an atom taking part in electron scattering is only around one angstrom. In other words, relative to their interatomic nearest-neighbor distances, atoms are big, so big indeed that the average *mean free path* of an electron between two successive collisions with atoms should be at most several angstroms, which means almost nothing. According to this classical point of view, an electron will hit a “large” atom just in front of it on its way, perhaps change direction, and then almost immediately hit another “large” atom, and this will go on and on like this; at the end, its displacement will be only a little. Nonetheless, nature, in the language of quantum mechanics, says that if the structure of crystal is perfectly flawless (so that it is referred to as a *lattice*), an electron can move readily through this lattice structure as if it were a free particle in vacuum; its travel through the lattice will be easy and effortless, and amazingly smooth and plain. It is exactly because of this peculiar fact that electrical and thermal conduction in metals (in which “free” electrons are abundant) are incomparably higher than those in insulators (they have only “bound” electrons that *cannot* freely wander through the lattice structure). This very fact has also played the number one role in paving the way for the development of today’s many indispensable industrial devices. An exquisite example is transistors (in which electrons travel freely through the lattice structure) that have been created to imitate the conventional radio tubes (also in which electrons move absolutely freely in vacuum).

¹ The material here is based on the thirteenth chapter of one of Feynman’s colossal masterpieces [54].

The notion of the conduction of electrons in a crystalline lattice structure is actually a ubiquitous natural phenomenon. It has been verified that many other “things” like some kind of *disorder*, *disturbance*, or (atomic) *excitations* can also propagate through the lattice, just like a free electron does. Consequently, the phenomenon we are about to study is encountered in many branches of (the solid state) physics, so the serious reader should pay sufficient attention to the fine details that we will expound shortly.

In the previous chapter we have studied three examples of systems with two states. The electron in the molecular hydrogen ion, for instance, can be each one of two positions; each position has exactly identical environments for the single electron. There is some amplitude $H_{12} = -A$ that it can move from one proton (position) to the other and the same amplitude $H_{21} = -A$ to move back. Quantum mechanics then dictates that: (1) There are two probable base states $|1\rangle$ and $|2\rangle$ of particular energy $H_{11} = H_{22} = E_0$ for the electron; we describe each of these states by the amplitudes C_1 and C_2 that the electron wanders close to each of the two protons (positions). In either of these two base states, the amplitudes C_1 and C_2 evolve in time as

$$C_1(t) = (\text{const.}) e^{-(i/\hbar)E_0 t} \quad \text{and} \quad C_2(t) = (\text{const.}) e^{-(i/\hbar)E_0 t}$$

where we see that their magnitudes are constant and that their phases change in time with the same frequency. (2) If the electron starts around one of the protons (positions), at a later time it will move to the other proton (position), and at another later time it will come back again near to first proton (position). We said that the amplitude is similar to two coupled pendula.

We now think of an absolutely perfect crystalline lattice in which we envisage that a single electron can be positioned in a “pit” at one specific atom with some specific energy. We also assume that there is some amplitude that the electron go into another pit at one of the nearest-neighbor atoms. The situation is analogous to our previous two-state system—but it has now been compounded by the fact that when the electron gets to a nearby atom, it can then go on to another nearest-neighbor atom or it may return back to the initial atom. This new case is now similar to innumerable many pendula which are coupled in, again, nearest-neighbor manner.

If we have a simple harmonic oscillator (SHO) that is coupled to another SHO, and, in turn, that one is coupled to another SHO, and so on ..., and if we create an *irregularity* somewhere, we observe then that that irregularity shall propagate through the 1D line of SHOs as a wave. The same phenomenon is seen if an electron is put at any one atom of a long 1D chain of atoms.

When we want to solve a mechanical problem, most of the time we do not solve it by considering its temporal initial point, in which a pulse starts, or final points, where the pulse diminishes. The initial and final solutions contain the so-referred *transient* temporal parts that last usually for a very short time interval which is, in general, of no importance. We start to deal with the problem by taking any intermediate temporal point and solve it in *steady-wave* terms that give the steady part of the solution which last for a long time interval and is independent of time.

8.1 The Base States for an Electron in a 1D Chain of Atoms

Now we are ready to analyze the problem at hand quantitatively in view of quantum mechanics. We restrict our analysis to a 1D (lattice) system which is composed of a (very) long chain of atoms, as depicted in Fig. 8.1(a). A crystalline (lattice) system is of course 3D but its physics is a very straightforward generalization of the relevant 1D system: whenever one grasps the details of the 1D version, they will easily and readily extend their accumulated acquaintance to the 3D version. The question to ask is: what happens when a single *extra* electron is placed on this chain of atoms? Of course, there are already billions and billions of electrons in an actual crystalline structure. But a large majority of them (in fact nearly all of them in a nonconducting crystal) are already situated near their own atoms undergoing a somehow well-defined pattern of motion. In other words, the whole lot *was* rather stationary before the introduction of our additional *extra* electron; it is this extra electron that starts an “irregularity” and causes it to propagate through the 1D chain of atoms. We shall not contemplate on what the other, already available, electrons in the lattice structure are doing by supposing that it requires a large amount of excitation energy to alter their on-going motions. We shall add an *extra* electron as though to obtain a single negative ion which is bound somewhat slightly. We will just keep a careful eye on what this single

extra electron does, and will ignore all the other on-going happenings within the 1D chain.

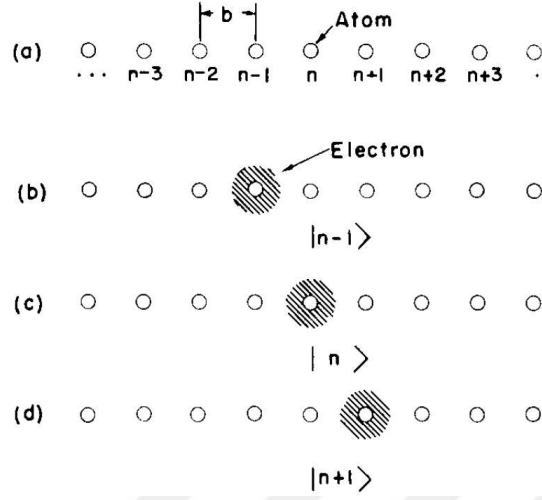


Figure 8.1: (a) A 1D (lattice) composed of a very long chain of atoms. (b)-(d) The base states of the extra electron. Taken from Ref. [55].

Because of the nearest-neighbor scheme we follow, we suppose that the extra single electron (*the electron*, from now on) might be at one particular atom with some amplitude, or it may move from this particular atom to its neighboring-atoms with some other amplitudes. We also suppose that the separations between any two neighboring atoms are all the same. (We say that the distribution of the atoms on the chain are *uniform*.) It seems better to label the atoms on the chain with integer numbers n , as we did Fig. 8.1(a).

In order to describe mathematically this 1D chain of atoms with one electron, we need some base states. One of them is that the electron is at the 6th atom, another is that it is at the 7th atom, or at the 8th atom, and so on. More generally, the n th state is that the electron is at the n th atom, which we shall designate as $|n\rangle$. Figures 8.1(b)-(d) depict clearly the meanings of the base states $|n-1\rangle$, $|n\rangle$, and $|n+1\rangle$. Using these, we can describe any *molecular* state $|\phi\rangle$ of our 1D chain in terms of

$$\langle n | \phi \rangle \equiv C_n(t)$$

that is the amplitude that the molecular state $|\phi\rangle$ is in base state $|n\rangle$ in which the electron is at n th atom. Thus, the molecular state $|\phi\rangle$ can be written just like as a

linear combination of atomic orbitals (LCAO)

$$|\phi\rangle = \sum_n |n\rangle \langle n|\phi\rangle = \sum_n |n\rangle C_n(t) \quad (8.1)$$

If we happen to know the amplitudes $C_n(t)$ at any time t , we take their absolute squares $|C_n(t)|^2$ and thus obtain the probability that the electron is found at the n th atom at time t .

What will happen at a later time? At this point the analogy between the present problem and the two-state problems we have studied previously should be obvious. Taking the amplitudes that the electron moves from the n th atom to the $(n \pm 1)$ th atoms as $iA/\hbar$ (per unit time), as we did before, we may suggest that the main dynamical equation that governs how the amplitude $C_n(t)$ changes with time should be written in the form

$$i\hbar \frac{dC_n(t)}{dt} = E_0 C_n(t) - AC_{n+1}(t) - AC_{n-1}(t) \quad (8.2)$$

where E_0 is the energy of the electron for the case in which it cannot move away its position from one of the atoms in the chain. (Actually E_0 is nothing but the *zero-energy choice* in our energy scale, as we shall see shortly.) The second and third terms in (8.2) is the amplitudes, per unit time, for the electron to move into the n th atom from the $(n + 1)$ th and $(n - 1)$ th atoms, respectively. Here A is again taken as a time independent constant.

Our 1D chain extends from $-\infty$ to $+\infty$ so that it contains an infinitely many number of atoms. This means that the number of the (time-dependent) Hamiltonian equations of the form (8.2) is also indefinitely large:

$$\begin{aligned} & \vdots \\ i\hbar \frac{dC_{n-1}}{dt} &= E_0 C_{n-1} - AC_{n-2} - AC_n \\ i\hbar \frac{dC_n}{dt} &= E_0 C_n - AC_{n-1} - AC_{n+1} \\ i\hbar \frac{dC_{n+1}}{dt} &= E_0 C_{n+1} - AC_n - AC_{n+2} \\ & \vdots \end{aligned} \quad (8.3)$$

8.2 The States Possessing Definite Energy

As we have pointed out before, when we talk about a state of definite energy, we mean a case in which *all* the amplitudes $C_n(t)$ evolve in time at the same frequency. We might then try a solution of form

$$C_n(t) = a_n e^{-iEt/\hbar} \quad (8.4)$$

where the (complex) constant a_n is for the time-independent part of the amplitude $C_n(t)$, whose absolute square gives the probability that the electron is found at atom number n . Feeding this solution form into (8.2) we end up with

$$Ea_n = E_0 a_n - Aa_{n+1} - Aa_{n-1} \quad (8.5)$$

In order to solve these equations (the number of such equations are infinite!), we first mark the atoms in the chain according to their positions. Let x_n be the position of the n th atom, x_{n+1} be that of the $(n + 1)$ th atom, and so on. If the distance between two nearest-neighbor atoms is b , as shown in Fig. 8.1, we see that the relationship among $\{x_n\}$ should be

$$x_{n+1} = x_n + b \quad (8.6)$$

By choosing the origin of the 1D chain at the position of the 0th atom, we can even write

$$x_n = nb \quad (8.7)$$

Now we can recast (8.4) as

$$C_n(t) = a(x_n) e^{-iEt/\hbar} \quad (8.8)$$

with the understanding that $a(x_n)$ is a function of position. With this (8.5) takes the alternative form

$$E a(x_n) = E_0 a(x_n) - A a(x_{n+1}) - A a(x_{n-1}) \quad (8.9)$$

Or, using (8.6), we can have another form

$$E a(x_n) = E_0 a(x_n) - A a(x_n + b) - A a(x_n - b) \quad (8.10)$$

This is a *difference equation*. It says that the function $a(x)$ at a point $x = x_n$ is interconnected to the itself at nearest-neighbor points $x = x_n \pm b$.

We can try to solve (8.10) by an educated guess (or, *ansatz*)

$$a(x_n) = e^{ikx_n}, \quad (8.11)$$

where k is a *real* number, so that (8.10) becomes

$$Ee^{ikx_n} = E_0e^{ikx_n} - Ae^{ik(x_n+b)} - Ae^{ik(x_n-b)}$$

or, eliminating the common factor e^{ikx_n} ,

$$E = E_0 - Ae^{ikb} - Ae^{-ikb} \quad (8.12)$$

Since

$$-Ae^{ikb} - Ae^{-ikb} = -2A \cos kb$$

we have finally

$$E = E_0 - 2A \cos kb \quad (8.13)$$

This important result tells us that for *any* value of k , there corresponds a solution with energy given by (8.13). Since there is no restriction on k , the number of such solutions is infinite—this should be unsurprising, for the number of base states is also infinite.

For each value of k , the spatial functions $a(x_n)$ are given by (8.11). Then, the full form of amplitudes $C_n(t)$ in (8.8) is now

$$C_n(t) = e^{ikx_n} e^{-(i/\hbar)Et} \quad \text{with} \quad E = E_0 - 2A \cos kb \quad (8.14)$$

The amplitudes $C_n(t)$ are seen to contain a time-dependent part, $e^{-(i/\hbar)Et}$, and a space-dependent part, e^{ikx_n} ; the latter says that the amplitudes $C_n(t)$ oscillate from one atom to the neighboring one as we go along the 1D chain.

What we mean is that, for *any* fixed time t , the space part of the amplitude $C_n(t)$, given by

$$e^{ikx_n} = \cos kx_n + i \sin kx_n, \quad (8.15)$$

exhibits a *complex* oscillation. Its magnitude is unity for *any* atom because

$$|e^{ikx_n}|^2 = e^{ikx_n} e^{-ikx_n} = 1,$$

or because

$$|e^{ikx_n}|^2 = \cos^2 kx_n + \sin^2 kx_n = 1,$$

but its phase differs by an amount ikb as we move from one atom to the neighboring one. In Figure 8.2 we plot the graph of $\text{Re}[C_n(t)] = \cos kx_n$ with x_n , which is the real part of the amplitude $C_n(t)$ at a fixed time t . (Note that this plot is discrete, *not* continuous. Only vertical lines at each atom are meaningful. The broken line that envelopes these vertical lines is then the continuous graph of $\cos kx$.) The graph of the imaginary part, $\text{Im}[C_n(t)] = \sin kx_n$, of the amplitude $C_n(t)$ at a fixed time t also oscillates, but it is shifted in phase by 90° with respect to the real part.

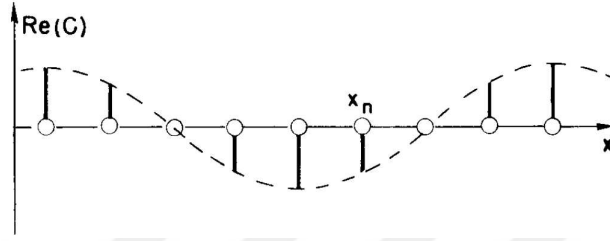


Figure 8.2: The graph of the real part of the amplitude $C_n(t)$ at a fixed time t . Taken from Ref. [56].

Let us recapitulate our findings as: (1) For any value of k we obtain a *stationary* molecular state with a specific energy E . (2) For any such stationary molecular state, the probability that the electron is found at *any* atom at any time is unity, thus the same. This means that there is no preference for the electron to be positioned at one atom or another atom. (3) At a fixed time t , only the phase of the space part varies. (4) As time evolves, we see another phase variation because of the time part. (5) It follows from (8.14) that the real and imaginary parts of the amplitude the amplitude $C_n(t)$, rewritten as

$$C_n(t) = e^{i[kx_n - (E/\hbar)t]}, \quad (8.16)$$

propagate as *traveling waves* along the 1D chain of atoms. The direction of this traveling wave, toward the positive or negative x -direction, depends on the sign of the *wave number* k . We shall never forget that these are *not* real waves, they are merely the so-called *probability waves*.

Note that the wave number k in the trial solution (8.11) was assumed to be a real number. It must be so especially if we are to consider a chain of atoms of infinite length. To see this, we let k be an imaginary number, as $k \rightarrow ik'$. Then the space

part $a(x_n)$ of the amplitude $C_n(t)$ in (8.10) becomes $e^{k'x_n}$. But this gets larger and larger for large positive (negative) values of x_n if k' is a positive (negative) number. This sort of solution is acceptable only when we consider a chain of atoms of *finite* length; it is not a physical solution for the present problem of infinitely long chain because it will lead to infinite probabilities. Later of we will see a special situation in which an imaginary wave number k will be meaningful.

We plot in Fig. 8.3 the graph of the energy E as a function of the wave number k , as given in (8.13). It is seen that E varies between $(E_0 - 2A)$ at $k = 0$ and $(E_0 + 2A)$ at $k = \pm\pi/b$. In this graph A is positive; if it were negative, the range would stay the same, but we would just invert the curve with respect to the line $E = E_0$. Here the important result is that the energy E is allowed to be only within a specific *energy band*; it cannot assume any value outside of this band. In other words, as a result of the assumptions made in the solution of this problem, if the electron in the 1D chain of atoms is in a *stationary* state, it must have an energy value *within* this band.

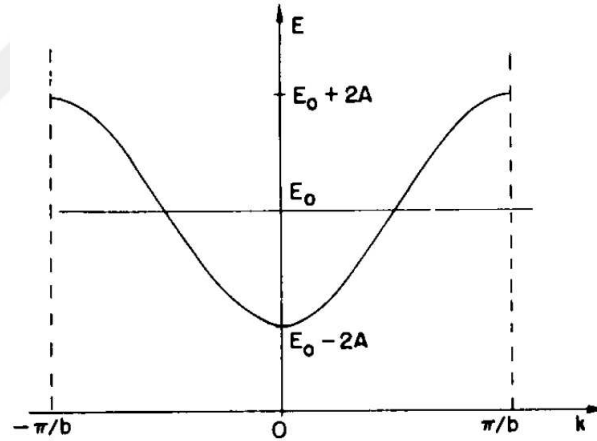


Figure 8.3: The graph of $E = E_0 - 2A \cos kb$, $A > 0$, as a function of the wave number k ; taken from Ref. [57].

It follows from (8.13) that the smallest values of k in magnitude belong to low-energy states, with $E \approx (E_0 - 2A)$. As is seen also from Fig. 8.3, if the value of k increases in magnitude from zero to positive or negative values, the energy E increases, at first slowly, then steeply, after then slowly again, and finally it reaches the maximum value $E = (E_0 + 2A)$ at $k = \pm\pi/b$.

What happens when the values k is beyond $\pm\pi/b$ towards both directions? The energy E will just begin to decrease again. Actually we do not need to consider such values of k beyond $\pm\pi/b$, for they will not result in new states, instead, they will lead to the states that we already have for smaller $|k|$ values. In order to see this more clearly, we consider the lowest energy state corresponding to $k = 0$. Equation (8.11) says that the space-part amplitude $a(x_n)$ is equal to unity for *all* x_n :

$$a(x_n) = e^{ikx_n} = e^{i(0)x_n} = e^0 = 1$$

with the energy $E(0) = E_0 - 2A$. If we calculate the same quantities for $k = 2\pi/b$ we see that this state has the same energy, $E(2\pi/b) = E_0 - 2A$. Furthermore,

$$a(x_n) = e^{ikx_n} = e^{i(2\pi/b)x_n}$$

or, taking the 0th atom at the origin with $x_0 = 0$ and setting $x_n = nb$ [see (8.7)], we have

$$a(x_n) = e^{i(2\pi/b)x_n} = e^{i(2\pi/b)(nb)} = e^{i2\pi n} = 1$$

Therefore, because the two states for $k = 0$ and $k = 2\pi/b$ have the same space-part amplitude and the same energies, we draw the conclusion that these two states represents physically the same state.

For another example, consider the states with $k = \pi/4b$ and $k = 7\pi/4b$. First notice from (8.14) that these two states have the same energy $E = E_0 - \sqrt{2}A$. Observe the two curves plotted in Fig. 8.4; curve 1 is the real part of the space-part amplitude $a(x_n)$ for $k = \pi/4b$ and curve 2 is that for $k = 7\pi/4b$. They are $\text{Re}[a(x_n)] = \cos[(\pi/4b)x_n]$ and $\text{Re}[a(x_n)] = \cos[(7\pi/4b)x_n]$, respectively. (As we have pointed out before, *only* vertical lines at each atom, i.e., *at points* x_n , are meaningful; the full cosine curves are of *no* significance.) It is seen that the two values of k leads to the same $\text{Re}[a(x_n)]$ values at all points x_n .

At this point, it seems in order to touch on the subtle subject of the nearest-neighbor approximation that we have employed. Let us consider a situation in which the electron cannot only move to the nearest-neighbor atom with the usual amplitude $iA/\hbar$, but it can also move, in one direct jump, to the *next-nearest neighbor* atom with another amplitude $iB/\hbar$. (Here perhaps $|B| < |A|$.) We will find that the space-part amplitude is again in the form $a(x_n) = e^{ikx_n}$; in other words, this kind of solution is somehow

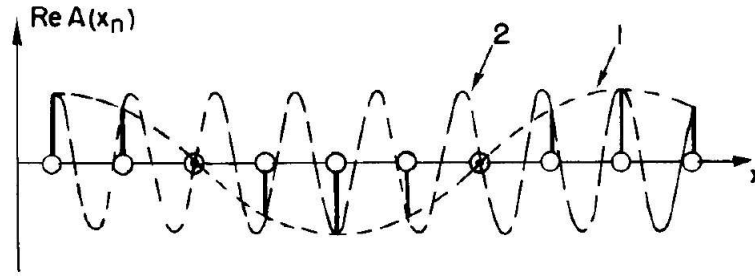


Figure 8.4: The graphs of the real part of the space-part amplitude $a(x_n)$ for $k = \pi/4b$ (curve 1) and $k = 7\pi/4b$ (curve 2); taken from Ref. [57].

universal. We will find also that the energy E of a *stationary* state corresponding to a certain wave number k is

$$E = E_0 - 2A \cos kb - 2B \cos 2kb$$

This means that the graphs of the energy E versus the wave number k are *not* universal at all, but they depend on the specific assumptions made in the treatment of the problem—it may not be a cosine curve, or even, it may not be symmetrical with respect to some (horizontal) line. It is always correct, however, that the $E(k)$ curve renews itself outside of the interval $[-\pi/b, +\pi/b]$, so we shall never need to take the other values of k into consideration.

Now let us see what happens when the wave number k is small, which corresponds to the states for which the variations of the space-part amplitude $a(x_n) = e^{ikx_n}$ from one atom to the next one are rather slow. [See also, for example, the real part of (8.15).] If we make the *zero-energy choice* by setting $E_0 = 2A$, then the minimum energy value of the graph in Fig. 8.3 will be at $E = 0$. For small values of k , we have from the small-angle approximation that

$$\cos kb \cong 1 - k^2 b^2 / 2$$

so the energy in (8.13) reduces, in the new energy scale, to

$$E = Ak^2 b^2 \tag{8.17}$$

We see that the energy of the stationary state is directly proportional to the square of the wave number k , which governs the variations of the space-part amplitude $a(x_n)$, and, in turn, that of the spatial part of the amplitudes $C_n(t)$.

8.3 The Time-Dependent States

We have learned that if $C_n(t)$ is the amplitude that the electron is at an atom at x_n , then $|C_n(t)|^2$ gives the probability that it is found there. If we talk about the *stationary* states represented by (8.14), the probabilities that the electron is found at *any* atoms in the chain are all the same; they do *not* change in time. Now the question is: how can we describe a situation if the electron is *localized* in a particular region with a particular energy—thus we can say it is more probable to spot it at an atom than at some other atom? We can achieve this mathematically by superposing several states whose wave numbers k , so that their energies E , differ slightly. Then at least at $t = 0$, the amplitudes C_n will differ with position x_n because the various terms will certainly interfere, just as the “beats” obtained whenever there is an assortment of waves possessing slightly different wavelengths. (See Fig. 8.5 for a beautiful example.) So we can readily form a *wave packet* whose primary wave number is k_0 ; the others are just distributed around this k_0 .

A “beat” is formed in the places where the *total* probability $|C(x_n, t)|^2$ is large; it is because of exactly this reason that such a beat is said to represent (the state of) an electron or a particle in the 1D lattice. It is a well known phenomenon that such a beat travel along the 1D lattice with a *group velocity* that is closely related to the variation of the wave number k via

$$v_{\text{group}} = \frac{d\omega}{dk} \quad (8.18)$$

Thus (the state of) an electron that is a “cluster”—the total amplitude $C(x_n, t)$ varies *in space* just like the wave packet shown in Fig. (8.6) will travel along the 1D lattice with a speed (8.18), where the relationship between the frequency ω and the energy E is given by $E = \hbar\omega$. For the case in which the primary wave number k_0 of the wave packet is small, we can use (8.17) for E and obtain

$$v_{\text{group}} = \frac{2Ab^2}{\hbar} k_0 \quad (8.19)$$

This is a very important result. Remembering that the (linear) momentum of a (free) particle is given by $p = \hbar k$, the above equation tells us that the electron travel along the 1D lattice with a speed proportional to the classical momentum $p = mv$. Furthermore, it then follows from (8.17) that the energy of the electron is directly proportional

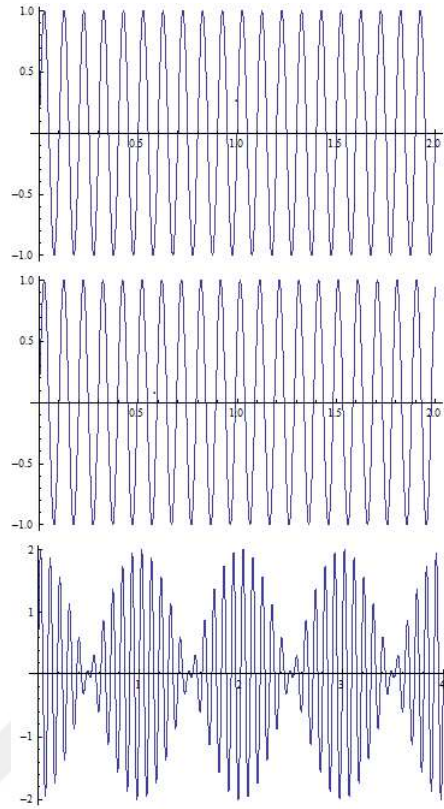


Figure 8.5: Formation of beats as a superposition of a two waves with *slightly* different wavelengths. The upper plot is for $f_1(x) = \sin[2\pi(10)x]$, the middle plot is for $f_2(x) = \sin[2\pi(10.1)x]$, and the lower plot is their sum, $f_1(x) + f_2(x)$.

to the square of its speed (or its momentum).² In other words, *the electron behaves just like a classical point particle!*

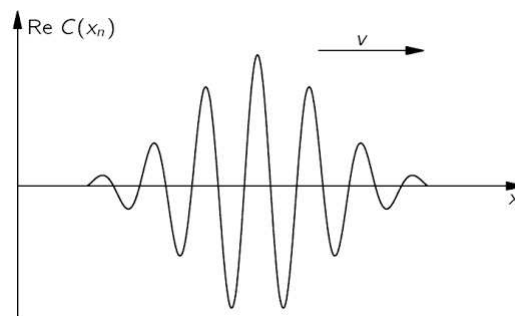


Figure 8.6: The real part of the total amplitude $C(x_n, t)$ versus the atomic position x . The plot is said to be in the shape of a wave packet because it is the superposition of various states whose wave numbers are only slightly different. Taken from Ref. [58].

² Remember that the (kinetic) energy of a *classical* point particle is $T = mv^2/2 = p^2/2m$.

The general conclusion we can draw from these results is that whenever we look at any entity on a big scale so that we cannot distinguish any of its fine features, the quantum mechanical results starts to yield well-known classical results. (This is actually the so-called *Bohr's correspondence principle*.)

Solving (8.19) for the primary wave number k_0 of the wave packet, we get

$$k_0 = \frac{\hbar}{2Ab^2} v_{\text{group}}$$

We feed this into (8.17) and obtain

$$E = Ak_0^2 b^2 = \frac{\hbar^2}{4Ab^2} v_{\text{group}}^2$$

We can make this energy expression resemble the classical (kinetic) energy counterpart as by writing it in the form

$$E = \frac{1}{2} m_{\text{eff}} v_{\text{group}}^2 \quad (8.20)$$

where m_{eff} is referred to as the *effective mass of the electron*, which is seen to be

$$m_{\text{eff}} = \frac{\hbar^2}{2Ab^2} \quad (8.21)$$

To underline the legitimacy of “assigning” $\hbar^2/2Ab^2$ to the idea of “mass,” we just check also the idea of “momentum.” It follows from (8.19) and (8.21) that

$$m_{\text{eff}} v_{\text{group}} = \hbar k_0 \quad (8.22)$$

But this is exactly equal to the (linear) momentum of a free particle from the quantum mechanical point of view! Hence, there is a perfect concordance between the classical and quantum mechanical pictures, demonstrating the validity and reliability of our simple (quantum mechanical) treatment of a 1D (perfect) lattice system.

We shall never mistake the effective mass m_{eff} of the electron (within a crystalline lattice structure) for its real mass m_e ; they are *not* equal to each other. But, curiously enough, these two masses of the electron happen to be of the same order of magnitude in real crystalline structures where m_{eff} ranges usually from $2m_e$ to $20m_e$.

We have at last wonderfully succeeded in explaining an equally wonderful mystery: how on earth an electron can easily, readily, and smoothly, and absolutely freely, travel through a crystalline lattice structure although it is supposed (in classical view) to collide with all the atoms on its way. Quantum mechanics has told us that the electron does so by somehow being blessed with its amplitudes allowing it to jump from one atom to the neighboring one through the lattice structure. If it moves within a “perfect” crystalline lattice structure, there will be nothing to oppose its motion, and we say that that crystal has absolutely no resistance to the flow of current. But whenever the structure of crystal is *not* perfect (there may be some *interstitial* atoms of some other type; there happens to be some missing atoms with their completely empty places; or there may be some (local) distortion in the lattice structure) so that it has some “irregularity”, then the electron will not be able to move so freely, and we will end up with a crystal having a particular resistance to the flow of current. The larger the degree of this “irregularity” a crystal structure has, the bigger a resistance it has.

In this section, we have not only learned the phenomenon of the conduction of electricity in a crystalline structure, but have also learned how a normally insulating crystalline structure (like germanium) can be transformed into a conducting one (by just inserting, by some means, some *extra* electrons in it).

8.4 An Electron in a 3D (Perfect) Crystalline Lattice

Let us now generalize the above treatment to the case of a 3D (perfect) crystalline lattice structure with a single *extra* electron. We shall see that the results will come out to be resembling very much to those of the 1D case. We consider a rectangular lattice structure formed by atoms of the same kind. Let the distances between nearest-neighbor atoms be a , b , and c in the respective x , y , and z directions. We suppose also that the amplitudes that the electron moves from one atom to the nearest-neighbor atom in the x , y , and z directions are $iA_x/\hbar$, $iA_y/\hbar$, and $iA_z/\hbar$, respectively. We specify any *lattice point* with (x, y, z) . If we choose the location of an atom to be at the origin, we can define all the other lattice points with respect to this origin as

$$x = n_x a, \quad y = n_y b, \quad \text{and} \quad z = n_z c \quad \text{with} \quad n_i = 0, \pm 1, \pm 2, \pm 3, \dots$$

As in the 1D situation, one obvious *base state* is that the electron is situated at an atom located at the lattice point (x, y, z) ; we represent this base state with the ket $|\text{electron at } x, y, z\rangle$. Furthermore, we designate the amplitude that the electron in some particular state $|\psi\rangle$ is found in any one of these base states $|\text{electron at } x, y, z\rangle$ as $C(x, y, z) = \langle \text{electron at } x, y, z | \psi \rangle$.

Although do not write it explicitly, the amplitudes $C(x, y, z)$ will understood to change in time: $C(x, y, z) \equiv C(x, y, z; t)$. Since our 3D lattice contains an infinitely many number of atoms, the number of the (time-dependent) Hamiltonian equations is also indefinitely large, which are in the form

$$\begin{aligned} i\hbar \frac{dC(x, y, z)}{dt} = & E_0 C(x, y, z) - A_x C(x + a, y, z) - A_x C(x - a, y, z) \\ & - A_y C(x, y + b, z) - A_y C(x, y - b, z) \\ & - A_z C(x, y, z + c) - A_z C(x, y, z - c) \end{aligned} \quad (8.23)$$

There corresponds such an equation for *each* atom that are at the lattice point (x, y, z) .

We expect again a *stationary state* in which all the amplitudes C 's changes in time in the same identical manner,³ so that our trial solution will be again exponential in the form

$$C(x, y, z) = e^{-iEt/\hbar} e^{i(k_x x + k_y y + k_z z)} \quad (8.24)$$

Upon feeding this into (8.23) we see that this is indeed the solution, and that the energy E of the system in this state is obtained as

$$E = E_0 - 2A_x \cos k_x a - 2A_y \cos k_y b - 2A_z \cos k_z c \quad (8.25)$$

We have now three wave numbers k_x , k_y , and k_z , one for each direction; they are, incidentally, the components of a 3D vector $\mathbf{k}$. With this, we can rewrite (8.24) in the more compact vector notation as

$$C(x, y, z) = e^{-iEt/\hbar} e^{-i\mathbf{k} \cdot \mathbf{r}} \quad (8.26)$$

where $\mathbf{r} = x \hat{\mathbf{x}} + y \hat{\mathbf{y}} + z \hat{\mathbf{z}}$ is the position vector of a lattice point (x, y, z) at which an atom is situated. It follows again that the amplitudes C 's behaves as a 3D complex *plane wave* whose wave number is

$$k = (k_x^2 + k_y^2 + k_z^2)^{1/2} \quad (8.27)$$

³ Here we remember again our old motto: "There is no preferred direction in the space."

The direction of this plane “wave” is that of the vector $\mathbf{k}$.

As before, if the signs of A_x , A_y , and A_z are all positive, and if the variation of the wave number k is very small from one atom to a next one in 3D, we obtain from the small-angle approximation that

$$E = E_{\min} + A_x a^2 k_x^2 + A_y b^2 k_y^2 + A_z c^2 k_z^2 \quad (8.28)$$

which is the 3D version of (8.17).

In the special case of a cubic lattice in which the distances between any two nearest-neighbor atoms in the usual three directions are all a , it is natural to expect that A_x , A_y , and A_z should be equal to the same constant, say, A . The above expression then reduces to

$$E = E_{\min} + Aa^2 (k_x^2 + k_y^2 + k_z^2)$$

or, with (8.27,)

$$E = E_{\min} + Aa^2 k_0^2 \quad (8.29)$$

which is exactly in the same form as (8.17). Therefore, we again draw the conclusion that an electron represented by a 3D traveling wave packet (containing many states with energies, or equivalently, wave number differing only slightly around a primary wave number k_0), travels just like a classical point particle with a particular effective mass m_{eff} .

In a non-cubic crystalline lattice structure (or in a cubic structure in which the electron’s state has no atomic symmetry), it is highly likely that the three constants A_x , A_y , and A_z all will be different. This means that the “effective mass” m_{eff} of a localized electron will be utterly dependent on the direction of its motion: it might have a particular effective mass m_x if it moves in the x -direction and a completely different one m_y if it moves in the y -direction.⁴

⁴ It seems to us that the word “mass” in the expression “effective mass” is a poor choice. It is perhaps more appropriate to use, in the present context, the word “inertia” or, “translational inertia”, which convey properly the original meaning of “mass.”

8.5 Other Probable States in a Crystalline Lattice Structure

It follows from (13.24) that the electron states can assume energy values only in a particular “energy band” whose respective bottom and top values are

$$E_{\min} = E_0 - 2(A_x + A_y + A_z) \quad \text{and} \quad E_{\max} = E_0 + 2(A_x + A_y + A_z)$$

Of course other energy values are also probable, but they are pertinent to other kinds of electron states. In the present analysis, we have chosen our base states in which the (single) electron is situated at an atom of the (perfect) crystalline lattice structure in some definite state with, say, the lowest energy E_0 or $E_{\min}$.

Suppose that there is an atom in vacuum, and that an electron is added to it to form an ion. The electron might have an infinite number of states in this system. It may be in the minimum-energy state, with a definite energy; or it may be in one of other *excited states* with an energy higher than the minimum. The same situation takes place in a crystalline lattice structure. Let E_0 be the energy of a base state in which the electron assumes its lowest energy, as we did before. We might also consider a new set of *base states* in which the electron does *not* assume the lowest possible energy, instead it is situated around an atom in a somewhat different manner so that its energy E' is certainly higher than the lowest energy E_0 . As we did previously, we assign some amplitude A' , which is not equal to the previous one A , that the electron will move from its present excited state at an atom to another excited state at a nearest-neighbor atom. Everything will be the same: we will end up with an energy band whose central energy is the new larger energy E' . The general conclusion to draw from this discussion is that we might have many (actually infinitely many) possible energy bands, each of them corresponds to a different excitation energy.

There might also be many other cases that we can consider. There can be a some cartain amplitude A'' for the electron to move from an excited state at an atom to some ground state at a nearest-neighbor atom. (Such an event is referred to as an interaction between two energy bands.) As we take into consideration more and more energy bands and put into the analysis more and more amplitudes A , A' , A'' for the electron between states, the mathematics we have to cope with will get more and more involved. The good thing is, however, that there will be *no* new ideas that we have to contemplate; the final equations will turn out to be of the same kind and of the same

level of difficulty.

What about the coefficient A that we have used throughout the whole theory? Actually there is not a lot to say anything about it, for it is a very demanding task to determine it theoretically. For this reason, in real practical calculations one usually uses coefficients which are determined by some empirical means.

We can also talk about other cases in which the physics, mathematics, and even wording of the problem is exceptionally similar to the problem of a single electron which travels through a crystalline lattice structure (that we have analysed above), but with a very small difference: the thing (or “the object”) that travels is *not* an electron. Consider as an example again our original 1D chain of *neutral* atoms, each of them has a loosely bound outer electron, just like the atoms in the group 1A of the periodic table. (As usual, the tightly bound inner electrons are not taken into account.) Then we assume that we have taken away one *single* electron from this chain. Which atom out of some infinite number of atoms has been stripped of its single outer electron? Let $C_n(x_n)$ denote the amplitude for the electron to be *missing* from the atom at position x_n . What may then happen? The outer electron of a nearest-neighbor atom, say atom number $(n - 1)$, just move to atom number n (which already has room for the newcomer electron), leaving atom number $(n - 1)$ stripped of its single electron. There will be some amplitude $iA/\hbar$ to describe this event. But this situation is exactly the same as *reverse* situation for which we say that there is some amplitude $iA/\hbar$ that the “missing electron” moves from atom number n to its nearest-neighbor atom number $(n - 1)$. We shall see again that the mathematics and its results will be exactly the same as before: we will again end up with the same formulas for the energy E and the probability “waves” $C_n(t)$, given in (8.14), which travels through the 1D chain with a group velocity given by (8.19). We will even see for sure that this single “missing” electron will have some definite “effective mass” (or, translational inertia) given by (8.21), and so on. The only difference now is that the probability waves, which are mathematically 1D traveling waves, describe how the single *missing electron* behaves, instead of a real electron. In the literature, such a missing electron is called as a “hole.” To sum up, such a “hole” behaves just like a classical charged point particle with an effective mass m_{eff} and a *positive* charge $+e$, which is equal in amount to the charge of an electron.

As another illustrative example, we consider again the 1D chain of neutral atoms. We suppose that one of the atoms is an excited state so that its energy is greater than its (unexcited) ground state energy. Let $C_n(x_n)$ denote the amplitude for atom number n at position x_n to have this excitation. This atom might interact with a nearest-neighbor atom by giving it its excess energy and thus return back to its ground state. We describe this event by some amplitude $iA/\hbar$. It should be obvious at this point to the diligent reader is that the formulation of the problem, its mathematics, and its analysis are all the same as before. This time the moving “object” through the 1D chain is referred to as an *exciton*, which behaves as if it were a neutral classical point “particle.” It carries with itself the excess excitation energy. Such a phenomenon is said to be an important mechanism in some biological events such as photosynthesis or our vision ability to see with our eyes. It has been hypothesized that whenever the retina absorbs a light, an “exciton” is formed that travels within some periodic construction. The journey of such excitons ends at some special accumulation spots where their excess energy is consumed to start a chemical reaction.

8.6 Scattering from Irregularities (or Imperfections) in a 1D Lattice

In this section we investigate the problem of again a single electron in a 1D atomic chain that is *not* perfect. Our previous analysis says that perfect crystalline lattice structures have absolutely no *resistance* to the motion of the electron—it goes readily slipping through the lattice, as if it were a free particle traveling in a vacuum. One of the most important reasons that can prevent the electron from traveling forever is an irregularity (or an imperfection) in the lattice structure. For instance, there might be a missing atom somewhere in the lattice, or there happen to be a different kind of atom at any lattice point so that the conditions there are not the same as those at other lattice points. Or, the energy E_0 at a lattice point or the amplitude A between nearest-neighbor lattice points might be different. How can we mathematically depict what happens in such cases?

Now let us be specific. We assume that the 0^{th} lattice point has a single “impurity” atom which has a different energy value of E_0 than other atoms at the other lattice points. Let $(E_0 + F)$ denote the energy of this impurity atom. What is next then?

When an electron happens to come the 0th atom, there will be some probability that is scattered back, or, equivalently, there will be some other probability that it will be transmitted forward. (These two probabilities are interrelated, as we will see shortly.) If we represent an electron with a wave packet, and such a wave packet arrives at the impurity atom whose conditions are a little different, *some* of it will continue and transmitted forward, and the remaining part will be *reflected* backward. From mathematical point of view it is a formidable task to analyze this problem working with a wave packet, since the time dependencies of the individual waves forming the wave packet are all different (because of their different wave numbers.) On the other hand, working with steady-state solutions is incomparably easier, where there will be a single traveling wave incident on to the 0th atom, another one reflected backward, and a final one transmitted forward.

We begin with equations like (8.5); we write a similar equation for each atom n . We obtain

$$\begin{aligned}
 & \vdots \\
 Ea_{-2} &= E_0 a_{-2} - Aa_{-1} - Aa_{-3} \\
 Ea_{-1} &= E_0 a_{-1} - Aa_0 - Aa_{-2} \\
 Ea_0 &= (E_0 + F) a_0 - Aa_1 - Aa_{-1} \\
 Ea_1 &= E_0 a_1 - Aa_2 - Aa_0 \\
 Ea_2 &= E_0 a_2 - Aa_3 - Aa_1 \\
 & \vdots
 \end{aligned} \tag{8.30}$$

Notice here that the energy of the 0th atom, $(E_0 + F)$, has been written different than those, E_0 , of the other atoms. Actually, for the most general case we should have used a different coefficient A for the amplitude for the electron to move from the 0th atom to its nearest-neighbors the -1^{th} and $+1^{\text{th}}$ atoms, but main characteristics of the final solution will remain the same. (Later on, within the subject of this thesis, we will allow it to be different.)

Solution (8.11) is still valid for all the equations in (8.30) except the middle one for the 0th atom. We repeat it here

$$a_n \equiv a(x_n) = e^{ikx_n}$$

which describes a wave moving in the $+x$ -direction. A similar wave traveling in the $-x$ -direction is also an equally valid solution, which is written as

$$a_n = e^{-ikx_n}$$

With these two, we can write the most general solution to the equations in (8.30), except the middle one, as a combination of them:

$$a_n = \alpha e^{ikx_n} + \beta e^{-ikx_n} \quad (8.31)$$

This solution is a blend of two complex waves, one with amplitude⁵ α traveling in the positive x -direction and the other with amplitude β traveling in the negative x -direction.

The general solution (8.31) is valid for every atoms but the 0th atom. We might make it belong only to the atoms with $n \leq -1$. Then we say for its physical meaning that there is an *incident* wave with amplitude α traveling in the positive x -direction towards the 0th atom, called *scatterer*. There is also a *reflected* or *scattered* wave with amplitude β traveling in the negative x -direction away from the 0th atom. There will not be any loss of generality if we take the amplitude α of the incident wave as being unity. Then, the amplitude β of the reflected wave will be, speaking generally, a complex number.

By the same line of reasoning, we can write the solution which belongs only to the atoms with $n \geq +1$; because the amplitudes of the incoming and the outgoing waves might be different, we can write this solution as

$$a_n = \gamma e^{ikx_n} + \delta e^{-ikx_n} \quad \text{for } n \geq +1 \quad (8.32)$$

where the first (second) term on the right depicts a complex wave of amplitude γ (δ), traveling in the positive (negative) x -direction. If we consider a specific case in which the incident wave originates from $-\infty$, approaches the central scatterer atom; there will be a reflected wave moving back to $-\infty$. Then there will be only a single *transmitted*

⁵ There is a funny incidence here. Actually, the word “amplitude”, meaning the *strength* of the (complex) wave, is not to be confused with our old friend “(probability) amplitude.” The good thing is here that since $a_n \equiv a(x_n)$ itself denotes the amplitude that the electron is situated at atom number n , the phrase “amplitude α ” literally and logically means “the amplitude of the amplitude”; the latter is for the same meaning, the amplitude that the electron is situated at atom number n , and the former is for the amplitude that the electron at atom number n is moving from left to right. What we want to say in short is that the usage of the word “amplitude” seems perfectly correct in our present context.

wave moving towards $+\infty$, away from the central scatterer atom. Then we can set the amplitude δ in (8.32) to zero. To sum up, then, we write the solution satisfying all of the equations, except the middle one, in (8.30) as

$$\begin{aligned} a_n (n < 0) &= e^{ikx_n} + \beta e^{-ikx_n} \\ a_n (n > 0) &= \gamma e^{ikx_n} \end{aligned} \quad (8.33)$$

with the additional requirement that the wave number k is relevant to the energy E and the distance b between two nearest-neighbor atom by

$$E = E_0 - 2A \cos kb \quad (8.34)$$

Figure 8.7 depicts clearly the situation that we are trying to outline.

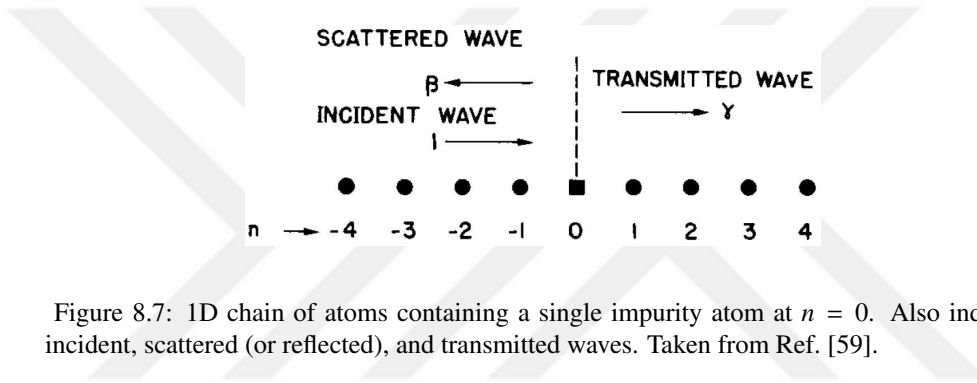


Figure 8.7: 1D chain of atoms containing a single impurity atom at $n = 0$. Also indicated are the incident, scattered (or reflected), and transmitted waves. Taken from Ref. [59].

Now we use the formulas (8.33) for the two solutions a_1 and a_1 ; we then insert them into the three middle equations in (8.30). Taking $x_n = nb$, we obtain

$$\begin{aligned} (E - E_0) [e^{ik(-b)} + \beta e^{-ik(-b)}] &= -A [a_0 + e^{ik(-2b)} + \beta e^{-ik(-2b)}] \\ (E - E_0 - F) a_0 &= -A [\gamma e^{ikb} + e^{ik(-b)} + \beta e^{-ik(-b)}] \\ (E - E_0) \gamma e^{ikb} &= -A [\gamma e^{ik(2b)} + a_0] \end{aligned} \quad (8.35)$$

We can solve these three equations for a_0 and for the two amplitudes β and γ , and thus we will have had the complete solution.

Remembering that $\cos x = \frac{1}{2} (e^{ix} + e^{-ix})$, we can rewrite (8.34) as

$$(E - E_0) = -A (e^{ikb} + e^{-ikb})$$

Inserting this into the first and last equations in (8.35) we obtain in succession

$$a_0 = 1 + \beta \quad (8.36)$$

and

$$a_0 = \gamma \quad (8.37)$$

so that

$$\gamma = 1 + \beta \quad (8.38)$$

This last result tell us that the amplitude γ of the transmitted wave is equal to sum of the amplitudes 1 and β of the incident and reflected waves, respectively. This result is *not* always correct; it is just an ordinary outcome special only to the present scattering problem with one single impurity atom.

What remains is the amplitude β of the reflected wave; we find it from the second equation of (8.35) as

$$\beta = \frac{-F}{F - 2iA \sin kb} \quad (8.39)$$

With this, we are done. We have managed to the obtain the complete solution for the 1D chain of atoms with a single impurity atom.

Before passing, the expression (8.38) might be misleading. It might deceptively implies that the transmitted outgoing wave is “more” than the incoming wave. This is wrong. The two amplitudes β and γ are not of some significance on their own because they are, in general, complex numbers. We remember that the probability (or the number of particle) notion is related to a wave through the absolute square of the amplitude, not the amplitude itself. As we have discussed before, if and only if the absolute square of the amplitude of the incident wave is equal to the sum of those of the reflected and transmitted waves, that is,

$$1 = |\beta|^2 + |\gamma|^2 \quad (8.40)$$

we can then speak of “conservation of particles.” This equation is *always* correct, as it readily verified for the present case at hand.

8.7 An Electron Trapped by an Impurity Atom⁶

Let us consider the situation in which F is negative. This means the energy ($E_0 + F$) of the electron at the 0th atom, is less than its energy E_0 at any other atom in the chain. It is then highly likely that the electron will be caught by the impurity atom. Let's be precise. If the energy ($E_0 + F$) is lower than ($E_0 - 2A$), which is the bottom of the energy band, then the electron will be trapped in a state with an energy $E < E_0 - 2A$. We cannot obtain such a solution from what we have done above. We can attain *this* solution, nevertheless, if we let the wave number k in the trial solution (8.16) be an imaginary number. To this end, let us take $k = i\kappa$ with κ being a real number. We can write a probable solution for the left part of the system, where $n < 0$, as

$$a_n (n < 0) = ce^{+\kappa x_n} \quad (8.41)$$

where we have taken the sign of κ positive in order for the exponent not to blow up as n approaches to $-\infty$. In a similar manner, we write a probable solution for the right part of the system, where $n > 0$, as

$$a_n (n > 0) = c'e^{-\kappa x_n} \quad (8.42)$$

These two trial solutions satisfy all the equations in (8.30) except the middle three. We obtain

$$E = E_0 - A(e^{kb} + e^{-kb}) \quad (8.43)$$

Here we notice that the inequality $(e^{kb} + e^{-kb}) > 2$ is always true; which means, in turn, that the energy given in (8.43) is always below (the bottom of) the usual energy band. We can satisfy the remaining three middle equations in (8.30) if we choose the amplitudes as $c = c'$, and κ as

$$A(e^{kb} - e^{-kb}) = -F \quad (8.44)$$

⁶This problem is *not* continuation of the preceding problem; it is a distinct problem on its own. Here there is *no* incoming or reflected or outgoing waves. The single electron can only be found situated at any atoms of the chain.

If we combine this result with (13.41), we obtain⁷

$$E = E_0 - \sqrt{4A^2 + F^2} \quad (8.45)$$

which the energy of the electron which got caught on the central impurity atom. It is readily that this energy is unique, and that it is below (the bottom of) the energy band (or, equivalently, the *conduction* band).

We notice that the amplitudes (8.41) and (8.42) tell us *nothing* about the place of the trapped electron; it might be sitting just on the central impurity atom, or on its neighboring atoms on the right or on the left. We calculate the probability that the electron is found at any atoms in the 1D chain by taking the square of these amplitudes. If we plot this probability against the positions x_n of the atoms in the chain, we obtain a curve just like the one depicted in Fig. 8.8. We see that the probability that the electron is sitting on the central impurity atom is the largest. The probability for the neighboring atoms drops off drastically in an exponential manner as the distance x_n from the central atom increases. This is a good demonstration for the phenomenon of *tunneling* (or *barrier penetration*). Classical mechanics says that the electron does not have sufficient energy to jump from the central impurity atom (which is referred to as the *energy hole*) to the neighboring ones. But quantum mechanics says it can, though it can get away only a little.

⁷ For inexperienced eyes, it might be a little subtle to obtain (8.45) by combining (8.43) and (8.44). We just remember the definitions of the hyperbolic functions

$$\cosh x = \frac{1}{2}(e^x + e^{-x}) \quad \text{and} \quad \sinh x = \frac{1}{2}(e^x - e^{-x}) \quad \text{with} \quad \cosh^2 x - \sinh^2 x = 1$$

We then note from (8.43) and (8.44) that

$$2A \cosh kb = E_0 - E \quad \text{and} \quad 2A \sinh kb = -F$$

or

$$4A^2 \cosh^2 kb = (E_0 - E)^2 \quad \text{and} \quad 4A^2 \sinh^2 kb = F^2$$

Subtracting the latter from the former, we get the desired result as

$$4A^2 = (E_0 - E)^2 - F^2 \quad \text{or} \quad 4A^2 + F^2 = (E_0 - E)^2 \quad \text{or} \quad \sqrt{4A^2 + F^2} = E_0 - E$$

and finally

$$E = E_0 - \sqrt{4A^2 + F^2} \quad (\text{QED})$$

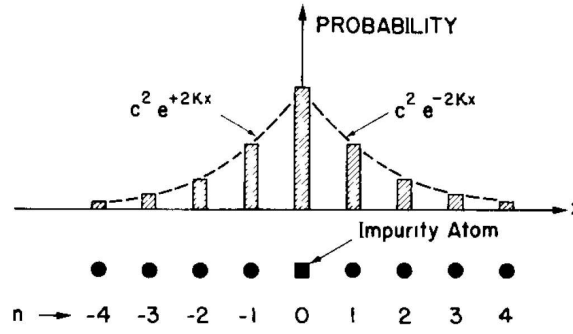


Figure 8.8: The curves for the probabilities that the single “trapped” electron is found at any atoms near the central impurity atom; taken from Ref. [60].

8.8 Bound States, Again, and Their Relationship to Scattering Amplitudes

Remember the way a proton and an electron come together in a bound state to form a hydrogen atom. [In a *bound state*, two initially infinitely separate (or free) particles come together to make a combination whose energy is lower than the energy of the two separate (free) particles.]

A theory says that a bound state will be present at energy values for which the scattering (or reflection) amplitude β is infinite if it is algebraically extrapolated (or it is continued analytically) to energy values outside of the regular energy band.

We explain the physical logic behind this as follows. Loosely speaking, in a bound state we have only waves attached to a point; there is *no* “incoming” wave to start the bound state, it is just there on its own. The proportion of the “scattered” (or reflected or created) wave to “outgoing” wave is infinite. Let us test this idea.

We first rewrite the amplitude of the scattered wave given in (8.39) in terms of E , the energy of the scattered (or reflected) particle, instead of k . To do so, we first need $2A \sin kb$ in terms of E . To this end, we rewrite (8.34) as⁸

$$2A \sin kb = \sqrt{4A^2 - (E - E_0)^2} \quad (8.46)$$

⁸ Notice that because of (8.34), we have

$$-2A \leq E - E_0 < 2A \quad \text{or} \quad |E - E_0| < 2A \quad \text{or} \quad (E - E_0)^2 < 4A^2$$

so the expression within the square root sign in (8.46) is always equal or greater than zero.

and using this in (8.39) we find the amplitude of the scattered wave

$$\beta = \frac{-F}{F - i\sqrt{4A^2 - (E - E_0)^2}} \quad (8.47)$$

Our original derivation dictates that the above equation holds only for *real* (physical) states, for which the energy E scattered of the particle is within the energy band, $|E - E_0| < 2A$. But let us assume that we were not aware of this fact and broaden the use of (8.47) into the *virtual* (unphysical) state, for which the energy E is *outside* the energy band, $|E - E_0| > 2A$. For these virtual states we may write⁹

$$\sqrt{4A^2 - (E - E_0)^2} = i\sqrt{(E - E_0)^2 - 4A^2}$$

Inserting this into (8.47), we get the “scattering amplitude,” who knows what it might mean, as

$$\beta = \frac{-F}{F + \sqrt{(E - E_0)^2 - 4A^2}} \quad (8.48)$$

Now we seek for any energy value E which makes β infinite [i.e., which makes the expression (8.48) for β have a “pole.”] Provided that F is negative, the denominator of (8.48) will become zero whenever

$$(E - E_0)^2 - 4A^2 = F^2$$

or

$$E = E_0 \pm \sqrt{4A^2 + F^2} \quad (8.49)$$

Here the energy with the negative sign is the energy in (8.45) that we have found for the trapped particle. What about the one with the positive sign? It is an energy *above* the top of the allowed regular energy band. The amazing thing is that there exists indeed another bound state which were at large when the equations (8.30) were solved in our previous analysis.

⁹ Here we do not bother ourselves with the sign of the root, a subtle point. It is relevant to the permitted signs of κ in (8.41) and (8.42).

CHAPTER 9

1D FINITE ATOMIC-CHAIN and ATOMIC-RING STRUCTURES

9.1 Introduction

At last, after an unusually lengthy theoretical foundation, we have come to our own subject. Now that the (diligent) reader has become a learned one at the subject of the propagation of an electron¹ through a 1D crystalline lattice structure, we have a sincere humble hope that s/he will easily and readily follow our present subject for which we shall not offer so much detail. Because the present manuscript has already become rather voluminous, we will henceforth focus more on the problems and their results, but we will not go into fine or obvious details so much, except the chapter devoted to the Aharonov-Bohm rings.

In the below derivations, since those used by Feynman, the Great Master, are now a little “old”, we will develop our own notation. We will also provide the reader with the nomenclature widely used in the literature. Our new notation and nomenclature will be highly commensurate with those of today’s literature, so the reader will possibly feel more comfortable in following them.

There will be two major changes in the formulation. Firstly, we will use the symbol t for the amplitude that the particle can move from one atom to one of its nearest-neighbor atoms instead of A that we have been using so far. Secondly, we will set the separation b between two nearest-neighbor atoms to unity: $b = 1$. This means that

¹ As the reader comfortably appreciate, the use of an electron here is not mandatory, instead of it we can use a particle or an “object” or an imperfection, or an irregularity, and so on. The only this we need is something that causes the previously perfectly ordered structure to “disordered” or “disturbed,” “perturbed,” but only *a little bit*.

all of the 1D and seemingly 1D (crystalline) lattice structures will be topologically perfect. The “disorder” on which we study will always be introduced to the system by the presence of our (extra) single electron.

In this chapter we will first consider 1D finite (linear) perfect chains of a *finite* number of atoms (of the same kind). The atomic chains that we have studied above were all containing an infinite number of atoms. We will then study perfect rings of, again, a finite number of atoms. Later on, these ring structures will imitate the so-called Aharonov-Bohm (AB) rings, perhaps threaded by some magnetic flux. The theory of the AB rings will be developed in a separate chapter in depth.

What will do with these 1D finite linear atomic chains and atomic rings? They will be our *central structures* connected from the left and right to semi-infinite, again, 1D linear atomic chains. Then an electron will come from the left towards this central systems and use them as a “mediator” in being transmitted to the right. Our main purpose will be to explore the strict conditions under which the electron will be transmitted to the right, a phenomenon called *resonance transmission*.

9.2 A 1D (Linear) Chain of N Atoms

We consider a 1D (linear) chain of N atoms, as shown in Fig. 9.1. We may think of the atoms being hydrogen atoms so that the system might be seen as a *hypothetical* H_N molecule, as a natural continuation of the H_2 molecule that we have studied previously. We said “hypothetical” because there is actually no such an H_N molecule; nevertheless its results will be illustrative, giving valuable insights into how a (3D) real crystalline solid material might be considered as a very large “molecule.”

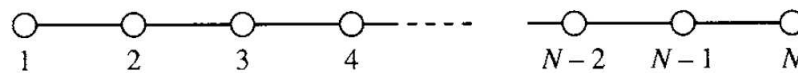


Figure 9.1: A 1D (linear) chain composed of N hydrogen-like atoms; taken from Ref. [61].

Because we are working with H atoms, each of them will possess only one bound *ground* state, namely, the 1s state, when they are kept isolated from the others. This means that there are totally N electrons in the whole molecule. We also keep in mind that each atom can possess two electrons, provided that their spins are opposite, but we shall not go into this kind of details. Also, for any possible two-electron case on a single atom we will ignore the Coulomb interaction between the two electrons, a phenomenon called as the *Coulomb blockage*. In other words, we will assume that there is no electron-electron interaction

We let $|j\rangle$ denote the 1s base state that is localized on the j th atom (or *atomic site*, or simply a *site*). The set of base states, $\{|j\rangle\} = \{|1\rangle, |2\rangle, |3\rangle, \dots, |N\rangle\}$, will be postulated to be *complete*

$$\sum_{j=1}^N |j\rangle\langle j| = \hat{1} \quad (9.1)$$

and *orthonormal*

$$\langle j|\ell\rangle = \delta_{j\ell} \quad (9.2)$$

so that the total *ground state* wave function $|\Psi\rangle$ of the molecule can be *expanded* onto the base states $\{|j\rangle\}$.² This means in mathematical terms that

$$|\Psi\rangle = \sum_{j=1}^N \phi_j |j\rangle \quad (9.3)$$

where ϕ_j are (complex) expansion coefficients that are to be determined.

The expansion postulate (9.3) allows us to solve the time-independent Schrödinger equation

$$\hat{H}|\Psi\rangle = E|\Psi\rangle \quad (9.4)$$

where E is the total molecular energy of the chain. To this end, we insert (9.3) into (9.4) and obtain

$$\hat{H} \sum_{j=1}^N \phi_j |j\rangle = E \sum_{j=1}^N \phi_j |j\rangle$$

or

$$\sum_{j=1}^N \phi_j \hat{H} |j\rangle = E \sum_{j=1}^N \phi_j |j\rangle$$

² How on earth do we know that these base states are adequate for such an expansion? The answer is simple: because we do *not* have any other simple option; we just make use of what we have at hand!

Multiplication from the left by the bra $\langle \ell |$ yields

$$\sum_{j=1}^N \phi_j \langle \ell | \hat{H} | j \rangle = E \sum_{j=1}^N \phi_j \underbrace{\langle \ell | j \rangle}_{\delta_{\ell j}}$$

whence

$$\sum_{j=1}^N \phi_j \langle \ell | \hat{H} | j \rangle = E \phi_\ell \quad (9.5)$$

As we did many times before, we restrict the matrix elements $\langle \ell | \hat{H} | j \rangle$ to being non-zero only for adjacent atoms, within the usual *nearest-neighbor approximation*. We make two definitions:

$$\langle j | \hat{H} | j \rangle = v \quad (9.6)$$

$$\langle j | \hat{H} | j \pm 1 \rangle = -t \quad (9.7)$$

where v is usually referred to as the *on-site energy* and t as the *hopping energy*.³

Now we use (9.5), together with (9.6) and (9.7), for different ℓ values ($1 \leq \ell \leq N$). We obtain

$$v\phi_1 - t\phi_2 = E\phi_1 \quad (\text{for } \ell = 1) \quad (9.8)$$

$$-t\phi_1 + v\phi_2 - t\phi_3 = E\phi_2 \quad (\text{for } \ell = 2)$$

$$-t\phi_2 + v\phi_3 - t\phi_4 = E\phi_3 \quad (\text{for } \ell = 3)$$

$$\vdots$$

$$-t\phi_{s-1} + v\phi_s - t\phi_{s+1} = E\phi_s \quad (\text{for } \ell = s) \quad (9.9)$$

$$\vdots$$

$$-t\phi_{N-2} + v\phi_{N-1} - t\phi_N = E\phi_{N-1} \quad (\text{for } \ell = N - 1)$$

$$-t\phi_{N-1} + v\phi_N = E\phi_N \quad (\text{for } \ell = N) \quad (9.10)$$

³ The attentive reader will notice that these two parameters v and t can be interpreted in two different ways. For the case in which the 1D chain represents an H_N molecule, v will be the energy of the molecule in which all the constituent protons are well separated and each of them has their “own” electrons wandering close to themselves; and t will then be the (probability) amplitude that two any nearest-neighbor proton exchange their electrons. In the other case in which the 1D chain represents a system of any atomic chain, and a *single* electron moves through the chain; for such a system v will be the energy of the single electron for the case in which it cannot move away its position from one of the atoms in the chain, and t will be again the amplitude that the electron move from one atom to any of its nearest-neighbor atom.

We can write the general equation (9.9) as

$$\phi_{s-1} + \phi_{s+1} = E' \phi_s \quad (9.11)$$

with the *dimensionless* energy

$$E' = (v - E)/t \quad (9.12)$$

The now-well-known ansatz for the general solution to this equation is

$$\phi_s = Ae^{iks} + Be^{-iks} \quad (9.13)$$

where

$$E' = e^{ik} + e^{-ik} = 2 \cos k \quad (9.14)$$

We here note that we have taken the distance between two nearest-neighbor atoms as being unity.

Equations (9.13) and (9.14) contain totally three parameters, A , B , and k . Since we are left with only two equations, (9.8) and (9.10), we can determine only two of them. From (9.8) we have

$$\phi_2 = E' \phi_1$$

Using the ansatz (9.13) and the first part of (9.14), we get

$$A + B = 0$$

or

$$B = -A \quad (9.15)$$

This result should have been expected if we reflect on the physical symmetry of the system. Similarly we have from (9.10) that

$$\phi_{N-1} = E' \phi_N$$

Again with (9.13) and the first part of (9.14), we obtain

$$Ae^{ik(N+1)} + Be^{-ik(N+1)} = 0$$

Using (9.15) we rewrite this as

$$A \underbrace{\left[e^{ik(N+1)} - e^{-ik(N+1)} \right]}_{\sim \sin[k(N+1)]} = 0$$

so

$$\sin[k(N+1)] = 0$$

and finally

$$k = \frac{n\pi}{N+1}, \quad (n = 1, 2, 3, \dots, N) \quad (9.16)$$

These are the allowed values of k . Here we exclude the $n = 0$ case because, with (9.15), it leads, due to (9.13), to $\phi_s = 0$, which is an unwanted trivial solution.

Recapitulating our findings, the general solution, from (9.9)–(9.16) is

$$\phi_j^n = A \left(e^{ikj} - e^{-ikj} \right) = A' \sin(kj) = A' \sin\left(\frac{n\pi}{N+1} j\right)$$

where constants A or A' can be determined from a normalization procedure. The result for the latter is

$$A' = \sqrt{\frac{2}{N+1}}$$

Thus the complete solution is

$$\phi_j^n = \sqrt{\frac{2}{N+1}} \sin\left(\frac{n\pi}{N+1} j\right) \quad (9.17)$$

with the allowed *discrete* eigenenergies

$$E = v - 2t \cos k = v - 2t \cos\left(\frac{n\pi}{N+1}\right) \quad (9.18)$$

with

$$n = 1, 2, 3, \dots, N$$

We note that there are totally N different solutions in the form

$$|\Psi^n\rangle = \sum_{j=1}^N \phi_j^n |j\rangle \quad (9.19)$$

where $|\Psi^n\rangle$ is usually called a *molecular orbital*.

If we look closely at (9.17), we see that they are standing waves on a grid $x_j = jb$, where we have taken $b = 1$. This behaviour is very similar to the problem of one particle in an infinitely high quantum well. We recall from an introductory quantum mechanics course that if the length of a such well is L , the eigenfunctions of the single particle is given by [62]

$$\psi_n(x) = \sqrt{\frac{2}{L}} \sin \frac{n\pi x}{L}, \quad (n = 1, 2, 3, \dots)$$

But this result is exactly the same as (9.17) if we observe that the length of the atomic chain is equal to $L = (N + 1)b = N + 1$.

Using (9.18), which we repeat here

$$E = v - 2t \cos k = v - 2t \cos \left(\frac{n\pi}{N + 1} \right),$$

we can obtain the eigenenergy spectrum of any 1D chain with a specific number of atoms N . Figure 9.2 shows examples of such spectra for the smallest five chains and for the chain with an infinite number of atoms. The first thing to be noticed is that all the eigenenergy values are evenly *discrete*, there is no *degenerate* eigenvalue at all. We see that as the number of atoms N increases, the gap between the consecutive eigenenergy values diminishes to zero, the width of the energy spectrum approaches to $4t$, and the variable k becomes continuous between 0 and π :

$$0 \leq k \leq \pi \quad (\text{for a 1D linear chain}) \quad (9.20)$$

Thus we confer that the spectrum of a 1D chain with an infinite number of atoms is continuous. It is called an *energy band*. Its *bandwidth* is $4t$, being proportional to the hopping constant t . Of course, this treatment assumes the use of nearest-neighbor approximation.

We can interpret these eigenenergy values again for two different cases. If the 1D chain represents an H_N molecule, then E will be the total energy of the molecule. The states with $E > v$ will be the *antibonding* states, whereas those with $E < v$ will be the *bonding states*. On the other hand, if the 1D chain represents, say, a system of protons through which a *single electron* can wander freely, then E will be the only allowed energy values of the electron. What does this mean? It means that if in some means a single electron will use this 1D chain as a “mediator,” or bridge, in passing from

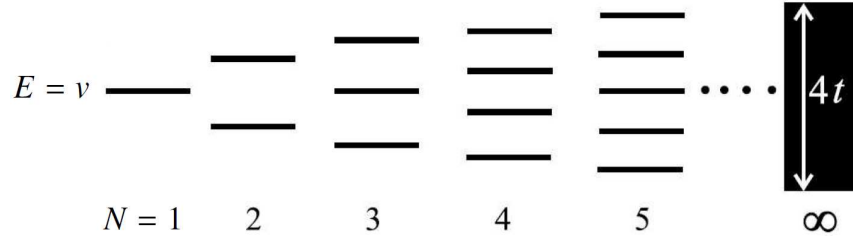


Figure 9.2: The eigenenergy spectra for 1D (linear) chains containing $N = 1-5$, and ∞ atoms, depicted using (9.18); adapted from Ref. [63].

some source to some sink, then that electron must have one of the eigenenergies of the chain. Otherwise it will “scattered” back to the source! We shall say more on this point later on when we discuss about the transport properties of such systems.

9.2.1 An Easy Solution Leading to (9.17) and (9.18)

We just add two artificial end sites $|0\rangle$ and $|N+1\rangle$ to the original chain. We assume the same solution as (9.3) which now includes these two end sites, as

$$|\Psi\rangle = \sum_{j=0}^{N+1} \phi_j |j\rangle$$

with two boundary conditions

$$\phi_0 = \phi_{N+1} = 0$$

Thus for all values of j from $j = 1$ to $j = N$, we have the same solution of the form

$$-t\phi_{j-1} + v\phi_j - t\phi_{j+1} = E\phi_j$$

The general solution is again of the form

$$\phi_j = Ae^{ikj} + Be^{-ikj}$$

Then $\phi_0 = 0$ gives $A = -B$, and $\phi_{N+1} = 0$ gives $\sin[k(N+1)] = 0$ so that we recover $k = n\pi/(N+1)$ ✓.

Now let us go back to equations (9.8)–(9.10). We can write them concisely in a three-diagonal (*sparse*) matrix form as

$$\begin{pmatrix} v & -t & 0 & \cdots & 0 & 0 \\ -t & v & -t & \cdots & 0 & 0 \\ 0 & -t & v & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & v & -t \\ 0 & 0 & 0 & \cdots & -t & v \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \vdots \\ \phi_{N-1} \\ \phi_N \end{pmatrix} = E \begin{pmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \vdots \\ \phi_{N-1} \\ \phi_N \end{pmatrix} \quad (9.21)$$

This structure of this matrix implies that that, with (9.6) and (9.7), we can write the Hamiltonian $\hat{H}$ of the system as

$$\hat{H} = \sum_{j=1}^N \left(v |j\rangle\langle j| - t |j\rangle\langle j+1| - t |j-1\rangle\langle j| \right) \quad (9.22)$$

or, allowing the hopping constant t to be complex, which is necessary when we deal with Aharonov-Bohm ring structures,

$$\hat{H} = \sum_{j=1}^N \left(v |j\rangle\langle j| + t |j\rangle\langle j+1| + t^* |j\rangle\langle j-1| \right) \quad (9.23)$$

or, in the *second-quantization notation*,⁴

$$\hat{H} = \sum_{j=1}^N \left(v c_j^\dagger c_j + t c_j^\dagger c_{j+1} + t^* c_j^\dagger c_{j-1} \right) \quad (9.24)$$

⁴ We give (9.24) only for the sake of completeness; we are not exploiting this notation in this work.

9.3 A Ring-Like Chain of N Atoms

We can now consider a ring-like chain (or a *ring*) of N atoms, as the one shown in Fig. 9.3. To be specific, we may let the atoms be hydrogen atoms so that the system might be seen as, again, a *hypotetical* H_N ring molecule.

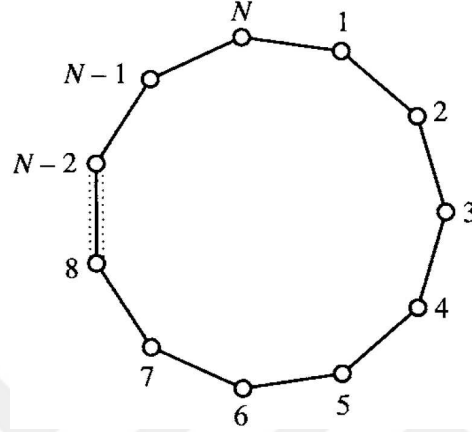


Figure 9.3: A ring-like chain composed of N hydrogen-like atoms; taken from Ref. [64].

The Hamiltonian of such a ring is the same as that for the 1D linear chain with one exception that the 1st and N th atoms are now nearest neighbors. Therefore, we have still the same general equation (9.11) to be solved. Since everything goes on in the same manner, the solution ansatz (9.13) and the eigenenergy spectrum (9.14) are all valid also for the ring structure. What remains is only the determination of the wave number k for the system at hand.

Having been experienced enough with the 1D chain case, we may also follow a straight approach in studying the ring problem. We can directly assume a *normalized* solution of the form

$$|\Psi\rangle = \frac{1}{\sqrt{N}} \sum_{j=1}^N e^{ikj} |j\rangle \quad (9.25)$$

(This above form is actually a *Fourier transform*.) In other words, we have just taken

$$\phi_j \sim e^{ikj} \quad (9.26)$$

Because of the physical periodicity of the ring, we have only one single boundary

condition

$$\phi_{j+N} = \phi_j \quad (9.27)$$

which leads to

$$e^{ik(j+N)} = e^{ikj}$$

whence

$$e^{ikN} = 1$$

or

$$\cos kN + i \sin kN = 1$$

This equation is satisfied if and only if

$$kN = 2\pi n$$

or

$$k = \frac{2\pi n}{N}, \quad n = 0, 1, 2, \dots, (N-1) \quad (9.28)$$

Thus, to recapitulate, the total molecular energy of the H_N ring (or the energy of a *single* electron wandering freely around an atomic ring) is given again by

$$E = v - 2t \cos k \quad (9.29)$$

or, in its explicit form,

$$E = v - 2t \cos \frac{2\pi n}{N}, \quad n = 0, 1, 2, \dots, (N-1) \quad (9.30)$$

Note that contrary to the 1D linear chain case, this time the $n = 0$ case *is* included because it does not give any trivial zero solution. Also, we have taken the maximum value of n as $N - 1$, since after this value, the $\cos k$ function in (9.30) repeats itself. Therefore, the total number of eigenenergy values is N , being equal to the number of atoms.

There is a nice connection between the 1D linear chain and the ring structure in the limit of infinite number of atoms. Within the nearest-neighbor approximation that we are following, when $N \rightarrow \infty$, the atoms in the linear chain does not feel what is going on at the end atoms of the chain. This means simply that an infinite linear atomic chain and an infinite atomic ring can be safely treated as the same system because both of them have exactly the same eigenenergy values in the limit of $N \rightarrow \infty$.

If we write the Schrödinger equation for the ring system in matrix form, we obtain

$$\begin{pmatrix} v & -t & 0 & \cdots & 0 & -t \\ -t & v & -t & \cdots & 0 & 0 \\ 0 & -t & v & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & v & -t \\ -t & 0 & 0 & \cdots & -t & v \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \vdots \\ \phi_{N-1} \\ \phi_N \end{pmatrix} = E \begin{pmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \vdots \\ \phi_{N-1} \\ \phi_N \end{pmatrix} \quad (9.31)$$

If we compare this with the one for the 1D linear chain given in (9.21), we see that the Hamiltonian matrix for the ring structure is still *sparse*, but no longer three-diagonal, because it contains two extra $-t$ elements in the upper-right and the lower-left corners. This small detail might be useful especially in writing simulation codes.

Making use of (9.30), we can gain insights into the eigenenergy spectrum of any atomic ring with a specific number of atoms N . Figure 9.4 gives examples of such spectra for chains with $N = 1, 3-5$, and ∞ . We notice immediately that, contrary to the case of linear chain depicted in Fig. 9.2, some of the eigenenergy values are *degenerate*. For example, for $N = 3$ the eigenvalue $v+t$ and for $N = 4$ the eigenvalue v are both *doubly* degenerate. We again also see that as the number of atoms N in the ring increases, the gap between the successive eigenenergy values diminishes to zero, the width of the energy spectrum approaches to $4t$, and, in turn, the variable k becomes continuous between 0 and 2π :

$$0 \leq k \leq 2\pi \quad (\text{for an atomic ring structure}) \quad (9.32)$$

We again conclude one more time that the spectrum of an atomic ring with an infinite number of atoms is continuous. The bandwidth of the relevant energy band is $4t$, which is directly proportional to the hopping constant t .

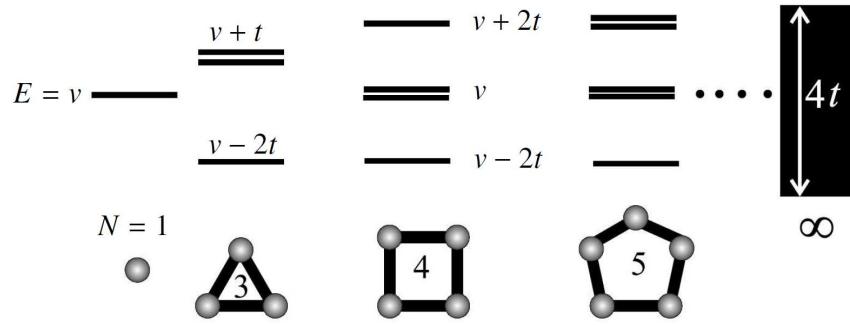


Figure 9.4: The eigenenergy spectra for atomic ring structures containing $N = 1, 3-5$, and ∞ atoms, depicted using (9.30); adapted from Ref. [65].

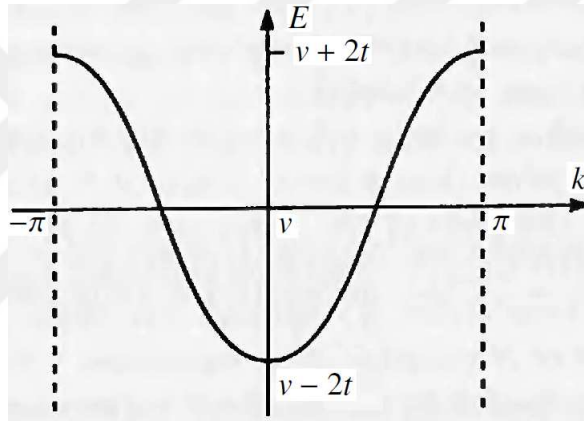


Figure 9.5: The band structure of the ring containing an infinite number of atoms; taken and modified from Ref. [65].

Thanks to the periodic cosine function, it is a common practice to redefine the range of the wave number k in the interval $-\pi < k \leq +\pi$ (or, actually $-\pi/b < k \leq +\pi/b$, where the interatomic distance between two nearest-neighbor atoms), as depicted in Fig. 9.5. The aim of such a practice is to accentuate the degeneracy of eigenenergy values between $-k$ and $+k$.

CHAPTER 10

RESONANT TRANSPORT: A COLLECTION OF PROBLEMS

10.1 Transport Through a Single Central Site

Here we set the rules of the game which will be played in the rest of the manuscript. Starting from this chapter, we will study the transport properties of a *single* particle, say an electron, through a 1D atomic chain which contains some “irregularity” or “disorder,” or “impurity” to some extent. Observe Fig. (10.1) that might be our prototype for such 1D systems. In terms of the nomenclature of today’s literature, such a 1D system under question is said to be composed of a *left lead* for $j \leq -1$, a *right lead* for $j \geq 1$, and a *central part* for $j = 0$, where j represents the j^{th} (atomic) site.

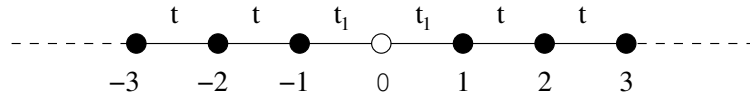


Figure 10.1: A 1D (linear) atomic chain containing a single “impurity” at site $j = 0$.

The left and right leads are semi-infinite atomic chains, which are perfectly structured 1D crystalline structures without any flaw, not containing any impurity. Thanks to their physical perfectness, the single electron can move through these two leads without encountering any hindrance or resistance, so they are said to be *ideal* conductors. Their sole aim is to provide the single electron with a “path” from a *source*, where the electron starts to its journey, to a *sink*, where the electron’s is disposed of and its journey terminates. In other words, the electron is ready to move through these

two leads, from left to right or vice versa, so a “current” is obtained. To this end we connect an external voltage source to the leads and then there said to be a *voltage bias*, or voltage difference between them, causing the electron to move.

The (probability) amplitude that the single electron moves from any atomic site j , in the left and right leads, to its two next-neighboring atomic sites $j \pm 1$ will be designated as t , whose other names are, as you well remember now, the *hopping constant* or *hopping energy*. We remember that the energy E of the electron in such a 1D atomic chain of infinite length is given by

$$E = v - 2t \cos k \quad (10.1)$$

where we have taken the distance between two neighboring atomic sites as being unity, $b = 1$. In this relation v is the now-usual *on-site energy*, that is, the energy of the electron when it would wander close to an atomic site in the case in which it cannot move to the neighboring sites, *in the presence of all the other atomic sites*. In practice, the on-site energy v is nothing but our zero-energy choice in the energy scale we work with, and it is supplied by the external voltage source connected to the lead under question.

We may again stress that the energy E of the single electron in the left and right leads are continuous, *not* discrete.¹ It follows from (10.1) that the electron can have any energy in the interval

$$v - 2t \leq E \leq v + 2t$$

If we think, from the point of view of band theory of solids, that the left and right leads are composed of hydrogen-like atoms each containing one s-type electron only, *at zero temperature* these leads are said to have a *half-filled* energy band, because each s-type atomic orbital has a capacity of housing two electrons with opposite spins. Because the leads are infinitely long, they will contain infinitely many, on the order of Avogadro’s number, hydrogen-like atoms and electrons.

¹ Actually, the energy E of the electron in an infinitely long atomic chain *is* always discrete; but the difference between any two successive energy levels are so small that we can safely assume that the energy E is continuous. It is said in such cases that as the number of atoms in a chain increases, the energy E becomes a *continuum*.

We said that at zero temperature, $T = 0$, the left and right leads are half-filled, up to the *Fermi energy* level, E_F . Thus, there is seemingly infinitely many electrons which are completely free to move through the leads. Let us say more on this point, but only briefly.

According to the free electron theory for the conduction phenomenon in metals, the free valence (or conduction) electrons (the 1s-electrons in our leads) must fill the quantized energy levels obeying *Pauli's exclusion principle*. The number of energy states, per unit volume of a 3D metallic crystalline structure, accessible for free valence electrons having energies between E and $E + dE$ is given by [66]

$$N(E) dE = C \frac{\sqrt{E} dE}{e^{(E-E_F)/k_B T} + 1}$$

where C is any constant,

$$k_B = 1.38 \times 10^{-23} \text{ J/K}$$

is *Boltzmann's constant*, and E_F is the so called *Fermi energy*. At zero temperature, $T = 0$, all the energy levels below the Fermi energy level E_F are completely filled, whereas all the levels above E_F are entirely empty. Now is the crux of the matter at hand: *only the electrons which possess energies near the Fermi energy level E_F can take a role for the conductivity of metals.*

Figure 10.2 shows a plot of the number of energy states $N(E)$ in a 3D metal versus energy E at zero temperature, $T = 0$, and at room temperature, $T = 300 \text{ K}$. We see that at 300 K only a tiny percentage of the free conduction electrons, usually much less than 1 %, are available above the Fermi energy E_F . This fact leads to a very important conclusion that any mathematical derivation carried out at $T = 0$ is nearly perfectly valid for all real-life room temperatures!

Having digressed a little, we can return back to our problem at hand. What about the central part? The first obvious fact about this part is that it is the primary source of “disorder” introduced to the system. In other words, if the central atomic site were identical with any other atomic site of the leads, our 1D system would be perfectly “ordered,” so that the single electron coming from any lead would travel through the central site to the other lead without encountering any difficulty, or resistance, to its travel. But we have now a “disorder” in the system at the central site and this very site

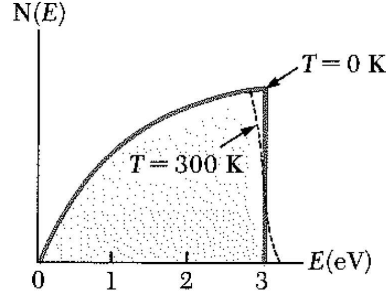


Figure 10.2: A plot for the number of energy states $N(E)$, per unit volume of a 3D metallic crystalline structure, versus energy E at $T = 0$ and $T = 300$ K. The Fermi energy is taken as $E_F = 3$ eV. Taken from Ref. [67].

will oppose to the motion of the electron from one lead to the other in such a way that part of the wave function associated with the electron will be *transmitted*, say from the left lead to the right lead hopping through the central site, and the remaining part will be *reflected* back to the left lead off, again, the central site. It may follow also from Fig. 10.1 that the “disorder” can be introduced to the system in two different ways: the on-site energy v_0 of the central site might be different from those of the atomic sites on the left and right leads, or the hopping constant t_1 from the leads to the central site might be, again, different from the hopping constant t between any next-neighboring atomic sites in the leads.

In our calculations, we will always take the hopping constant t for the two leads to be unity,

$$t = 1,$$

signifying a *perfect hopping* for the single electron from one site to the next-neighboring one. On the other hand, the hopping constant t_1 from the central site to the two leads, or for the reverse case of hopping from leads to the central site, will be taken between zero and unity,

$$0 \leq t_1 \leq 1,$$

where 0 means absolutely *no-hopping* and 1 means, again, perfect hopping.

The main equation that we have derived, in the previous chapter, from the time-independent Schrödinger equation is

$$E\phi_j = v_j \phi_j - \sum_{\langle j,k \rangle} t_{jk} \phi_k \quad (10.2)$$

where ϕ_j are the (complex) expansion coefficients to be determined, and $\langle j, k \rangle$ means that only the *nearest-neighbor hopping* will be taken into consideration. This is our fundamental energy equation from which everything will follow.

The following derivation will be as general as possible in that it will contain the cases in which the incoming electron might loose some of its energy as it passes through the central site² (or, for the most general cases, the central *region*), so that its wave numbers in the two leads *might* be different, that is, $k \neq k'$, where k (k') is the wave number of the single electron at the left (right) lead.

The ansatz for the left and right leads are already known. They are

$$\phi_{j<0} = 1 \cdot e^{+ikj} + R e^{-ikj} \quad (\text{for the left lead}) \quad (10.3)$$

$$\phi_{j>0} = T e^{+ik'j} \quad (\text{for the right lead}) \quad (10.4)$$

where 1 in the upper equation means that there is exactly “one electron” coming from the right lead. The coefficient R is the *reflection amplitude* giving the (probability) amplitude for the portion of (the wave function of) the electron which goes back through the left lead. And the remaining coefficient T is the *transmission amplitude* giving information about how much of (the wave function of) the electron goes on its way through the right lead.

The *transmission coefficient* T and the *reflection coefficient* R are generally given as

$$\mathsf{T} = |T|^2 \frac{k'}{k} \quad (10.5)$$

$$\mathsf{R} = |R|^2 \quad (10.6)$$

For our current problem, the transmission coefficient (10.5) becomes³

$$\mathsf{T} = |T|^2 \frac{\sin k'}{\sin k} \quad (10.7)$$

² Such a case is indeed possible if, for example, the central region contains a vibrational atomic system; the single electron would loose (or gain) energy when it excites the vibrational system to a higher (lower) energy level.

³ We have left the verification of the passage from (10.5) to (10.7) to the reader. There is no need to check the internet for the answer, for it is within the manuscript!

Our ultimate aim is to determine the relevant coefficient T (and/or R) for the problem under study. We should keep in mind, for the purpose of checking the consistency of any analytical derivation with the corresponding simulation results, that the sum of these two coefficients is always unity for all kinds of one-dimensional barrier problems:

$$T + R = 1 \quad (10.8)$$

Later on in this chapter, and in the following chapters, we will numerically verify this statement for several systems under study.

The total energy of the electron is conserved from the left to the right lead. We write

$$E = v - 2t \cos k = v' - 2t \cos k' \quad (10.9)$$

where v and v' are the *bias potentials* of the left and right leads. We shall later on refer to the second terms of the above equations as the “energy” of incoming and outgoing electrons because v and v' are just two constants.

We note that

$$e^{ik} = \cos k + i \sin k \quad \text{and} \quad e^{-ik} = \cos k - i \sin k$$

whence

$$e^{ik} + e^{-ik} = 2 \cos k$$

so that we can manipulate (10.9) for the left lead as

$$E - v = -2t \cos k = -t(2 \cos k)$$

and

$$E - v = -t(e^{ik} + e^{-ik}) \quad (10.10)$$

Similarly, for the right lead we have

$$E - v' = -t(e^{ik'} + e^{-ik'}) \quad (10.11)$$

We *do* know the solution for $\phi_{j<0}$ and $\phi_{j>0}$ but we *do not* know the solution for ϕ_0 . To this end, we will first determine ϕ_1 , ϕ_2 , ϕ_{-1} , and ϕ_{-2} .

Using (10.4) we have for ϕ_1 that

$$\phi_1 = T e^{ik'} \quad (10.12)$$

Using again (10.4) we have for ϕ_2 that

$$\phi_2 = T e^{2ik'} = T e^{ik' + ik'} = T e^{ik'} e^{ik'} = e^{ik'} \cdot T e^{ik'}$$

and

$$\phi_2 = e^{ik'} \phi_1 \quad (10.13)$$

where in the last line we have used (10.12).

Similarly, coefficients ϕ_{-1} and ϕ_{-2} follow from (10.3) as

$$\phi_{-1} = e^{-ik} + R e^{ik} \quad (10.14)$$

and

$$\begin{aligned} \phi_{-2} &= e^{-2ik} + R e^{2ik} \\ &= e^{-2ik} + R e^{ik} e^{ik} \\ &= e^{-2ik} + e^{ik} \cdot R e^{ik} \\ &= e^{-2ik} + e^{ik} (\phi_{-1} - e^{-ik}) \end{aligned}$$

and finally

$$\phi_{-2} = e^{ik} \phi_{-1} + e^{-2ik} - 1 \quad (10.15)$$

where we have used (10.14) for $R e^{ik}$.

Now using (10.2), we obtain the connection formulas between the central site ϕ_0 and the two edge sites ϕ_{-1} and ϕ_1 . For the edge site $j = -1$, we obtain

$$E \phi_{-1} = v \phi_{-1} - t \phi_{-2} - t_1 \phi_0$$

$$(E - v) \phi_{-1} = -t \phi_{-2} - t_1 \phi_0$$

$$-t (e^{ik} + e^{-ik}) \phi_{-1} = -t (e^{ik} \phi_{-1} + e^{-2ik} - 1) - t_1 \phi_0$$

where we have used (10.10) for $(E - v)$ and (10.15) for ϕ_{-2} . To proceed, we first divide every term by $-t$,

$$(e^{ik} + e^{-ik}) \phi_{-1} = e^{ik} \phi_{-1} + e^{-2ik} - 1 + \frac{t_1}{t} \phi_0$$

$$e^{ik} \phi_{-1} + e^{-ik} \phi_{-1} = e^{ik} \phi_{-1} + e^{-2ik} - 1 + \frac{t_1}{t} \phi_0$$

$$e^{-ik} \phi_{-1} = e^{-2ik} - 1 + \frac{t_1}{t} \phi_0$$

and finally multiply each term by e^{ik} to obtain ϕ_{-1} in terms of ϕ_0 as

$$\phi_{-1} = \frac{t_1}{t} e^{ik} \phi_0 + e^{-ik} - e^{ik} \quad (10.16)$$

In the same manner, we put $j = 1$ in (10.2) for the edge site ϕ_1 :

$$E\phi_1 = v'\phi_1 - t\phi_2 - t_1\phi_0$$

$$(E - v')\phi_1 = -t\phi_2 - t_1\phi_0$$

$$-t(e^{ik'} + e^{-ik'})\phi_1 = -t \cdot e^{ik'}\phi_1 - t_1\phi_0$$

where we have used (10.11) for $(E - v')$ and (10.13) for ϕ_2 . We divide both sides by $-t$ and obtain

$$\begin{aligned} (e^{ik'} + e^{-ik'})\phi_1 &= e^{ik'}\phi_1 + \frac{t_1}{t}\phi_0 \\ e^{ik'}\phi_1 + e^{-ik'}\phi_1 &= e^{ik'}\phi_1 + \frac{t_1}{t}\phi_0 \\ e^{-ik'}\phi_1 &= \frac{t_1}{t}\phi_0 \end{aligned}$$

Finally we multiply each side by e^{ik_R} , we obtain ϕ_1 in terms of ϕ_0 as

$$\phi_1 = \frac{t_1}{t} e^{ik'} \phi_0 \quad (10.17)$$

Now we use (10.2) for $j = 0$ and determine ϕ_0 . We have

$$E\phi_0 = v_0\phi_0 - t_1\phi_{-1} - t_1\phi_1$$

$$(E - v_0)\phi_0 = -t_1\phi_{-1} - t_1\phi_1$$

$$= -t_1 \left(\frac{t_1}{t} e^{ik} \phi_0 + e^{-ik} - e^{ik} \right) - t_1 \left(\frac{t_1}{t} e^{ik'} \phi_0 \right)$$

$$= -\frac{t_1^2}{t} e^{ik} \phi_0 - \frac{t_1^2}{t} e^{ik'} \phi_0 + t_1 (e^{ik} - e^{-ik})$$

$$\therefore (E - v_0)\phi_0 = -\frac{t_1^2}{t} (e^{ik} + e^{ik'}) \phi_0 + t_1 (e^{ik} - e^{-ik}) \quad (10.18)$$

where we have used Eqs.(10.16) and (10.17) for ϕ_{-1} and ϕ_1 , respectively.

Now we just think that our single electron comes from the left lead with an “energy” ε and going on with another “energy” ε' through the right lead:

$$\varepsilon = -2t \cos k \quad (10.19)$$

$$\varepsilon' = -2t \cos k'$$

The condition of energy conservation requires that

$$E = v + \varepsilon = v' + \varepsilon' \quad (10.20)$$

With the left part of this in (10.18), ϕ_0 finally takes the form

$$\left[(\varepsilon + v - v_0) + \frac{t_1^2}{t} (e^{ik} + e^{ik'}) \right] \phi_0 = t_1 (e^{ik} - e^{-ik})$$

and whence

$$\phi_0 = \frac{t_1 (e^{ik} - e^{-ik})}{(\varepsilon + v - v_0) + \frac{t_1^2}{t} (e^{ik} + e^{ik'})} \quad (10.21)$$

It is finally time to determine the transmission and reflection coefficients. With the help of (10.12), we first determine the transmission *amplitude* T . It follows from

$$\phi_1 = T e^{ik'}$$

that

$$T = e^{-ik'} \phi_1$$

Inserting (10.17) for ϕ_1 , we obtain

$$\begin{aligned} T &= e^{-ik'} \cdot \frac{t_1}{t} e^{ik'} \phi_0 \\ \therefore \quad T &= \frac{t_1}{t} \phi_0 \end{aligned} \quad (10.22)$$

or, with (10.21),

$$T = \frac{t_1^2 (e^{ik} - e^{-ik})}{t (\varepsilon + v - v_0) + t_1^2 (e^{ik} + e^{ik'})} \quad (10.23)$$

Here in this last result, the attentive reader will note, for dimensional analysis purposes to avoid any possible error, that each term in the numerator and in the denominator has a dimension of $[\text{energy}]^2$, so the transmission amplitude T is just a (complex) number, as it must be.

Using (10.22), because of its simplicity, the desired transmission *coefficient* $\mathbb{T}$ in (10.7) can be expressed as

$$\mathbb{T}(\varepsilon, \varepsilon') = \left(\frac{t_1}{t} \right)^2 |\phi_0|^2 \frac{\sin k'}{\sin k} \quad (10.24)$$

For the reflection *amplitude* R , we use (10.14) in reverse order:

$$\begin{aligned}
e^{-ik} + Re^{ik} &= \phi_{-1} \\
e^{-ik} + Re^{ik} &= \frac{t_1}{t} e^{ik} \phi_0 + e^{-ik} - e^{ik} \\
Re^{ik} &= \frac{t_1}{t} e^{ik} \phi_0 - e^{ik} \\
\therefore R &= \frac{t_1}{t} \phi_0 - 1
\end{aligned} \tag{10.25}$$

With this, the desired reflection *coefficient* R in (10.6) is expressed as

$$R = \left| \frac{t_1}{t} \phi_0 - 1 \right|^2 \tag{10.26}$$

We can now perform our first simulation. To this end, we have written a FORTRAN 77 code for trial and testing purposes and a series of MATHEMATICA[®] programs. Referring to Fig. 10.1, the system parameters that we use are $t = 1.0$, $t_1 = 0.2$, $v_c \equiv v_0 = 0.0$ (no potential is applied to the single central site), and $v = v' = 0$ (no external bias is applied).

As you see, we do not apply any external bias to the system, so will we in the rest of the manuscript: within the scope of this work, we will *not* manipulate our systems by applying any external bias. The reason for this is twofolds. Firstly, because they play the primary role in determining which electron with a specific energy will be transmitted to the right lead, we just want to obtain accurately the eigenenergies of the central part. (This somewhat vague sentence will assume its very meaning shortly.) Secondly, as usual, there will be always one single electron coming from the left lead with an energy ε in the interval between $-2t$ and $2t$ (or between -2 and 2 , since we take $t = 1.0$), it will hop to the central site, and then it will be transmitted to the right lead with an energy in the interval between $-2t$ and $2t$. We will always want the (gate) potential of the central site also within the same interval between $-2t$ and $2t$. Now, we will assume implicitly that the left lead always houses such an incoming electron and the right lead will always welcome the electron coming toward itself. Technically speaking, the single electron will come from the *occupied* part of the the left lead's energy band, and it will be accepted, after passing the central site, by the *empty* part of the right lead's energy band, and we take this fact as granted. Doing so, we will

have avoided the highly likely case of *complex* wave number k , thus we will guarantee that, if the other circumstances allow, the single electron will definitely move through and pass the central site (or region); it will not be *trapped* there.

Figure 10.3 (a) shows, as the result of numerical simulation, the plot of the transmission coefficient $T(\varepsilon, \varepsilon')$ versus the incoming electron energy ε for the system under consideration. We cannot learn a lot from this plot. We can only say, or conjecture, that whenever the energy ε of the incoming electron is around about (*not* necessarily exactly equal to) the gate potential $v_c = v_0$ of the cenral site, it can hop from the left lead to the central site and from there it can hop to the right lead. In all other cases it will be reflected back through the left lead. Let us write this as a hypothesis for comparison or resemblance purposes with our later systems including two or more central sites, and see that whether it is correct or not.

$$\begin{aligned} \text{HYPOTHESIS: if } \varepsilon \sim v_c, \text{ the electron can hop to the right lead;} \\ \text{otherwise, it is reflected back.} \end{aligned} \quad (10.27)$$

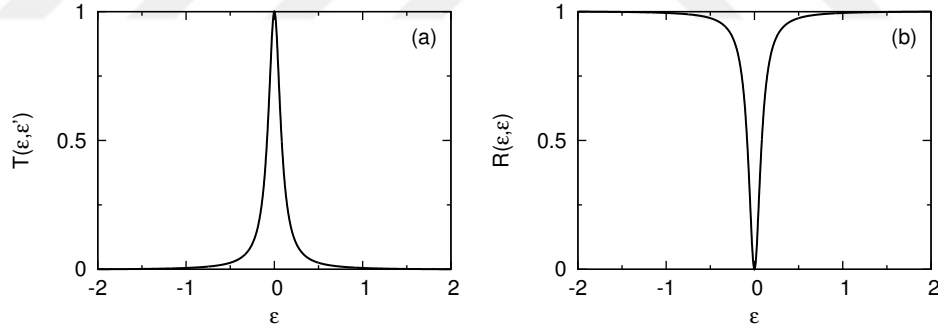


Figure 10.3: (a) The transmission coefficient $T(\varepsilon, \varepsilon')$ and (b) the reflection coefficient $R(\varepsilon, \varepsilon)$ versus the incoming electron energy ε for the 1D linear chain with single impurity. The common system parameters are $t = 1.0$, $t_1 = 0.2$, $v_c \equiv v_0 = 0$, and $v = v' = 0$.

As we pointed out before, in performing simulations we should also check the reflection coefficient R . Figure 10.3 (b) shows the plot of R versus the incoming electron energy ε for our present system. It is seen that the sum of transmission and reflection coefficients is indeed unity, $T + R = 1$, as it must be.

Now we want to check the hypothesis (10.27) for different values of the central gate potential. The simulation results are plotted in Fig. 10.4, which verify the correctness of the hypothesis: indeed, whenever the energy ε of the incoming electron is around the central gate potential $v_c = v_0$, say $\varepsilon \sim v_c = 0.6$, it will be transmitted to the right lead, as it is demonstrated clearly in Fig. 10.4 (c); otherwise it will be reflected back through the left lead. The reader has to wait for the later chapters to see whether the hypothesis (10.27) is universally correct or just a conjecture, or a special case of a more general truth; for the time being it is correct for our simple 1D chain with single impurity.

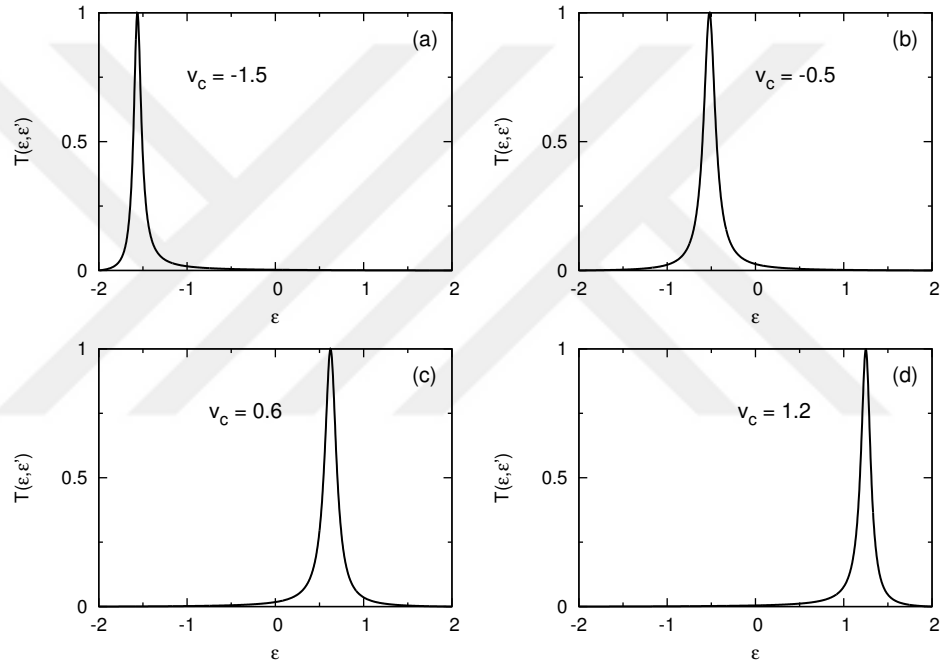


Figure 10.4: The transmission coefficient $T(\varepsilon, \varepsilon')$ versus the incoming electron energy ε for the 1D linear chain with single impurity for different values of the central gate potential $v_c \equiv v_0$ (see Fig. 10.1). The other common system parameters are $t = 1.0$, $t_1 = 0.2$, and $v = v' = 0$.

Making use of the present system, we can learn the role of the hopping constant t_1 that mediates the transmission between the central site and the left and right leads (see Fig. 10.1). To do this, we calculated the transmission coefficient $T(\varepsilon, \varepsilon')$ for different values of t_1 . Figure 10.5 shows the plots of the simulation results. We see that the smaller (the larger) the value of t_1 , the less (the more) likely that the single electron is transmitted to the right lead. In other point of view, Fig. 10.5 (a) with

$t_1 = 0.10$ being a rather small value, for example, says that if there are many electrons incoming *one-by-one* through the left lead, only a few of them might pass through the central site and continue their travel through the right lead; a great majority of them might not overcome the central impurity site and will be reflected back through the left lead. In the opposite case of Fig. 10.5 in which $t_1 = 0.75$, a highly large value, we can safely say that a great deal of the incoming electrons will be transmitted to the right lead; the reflected electrons will be only a small minority.

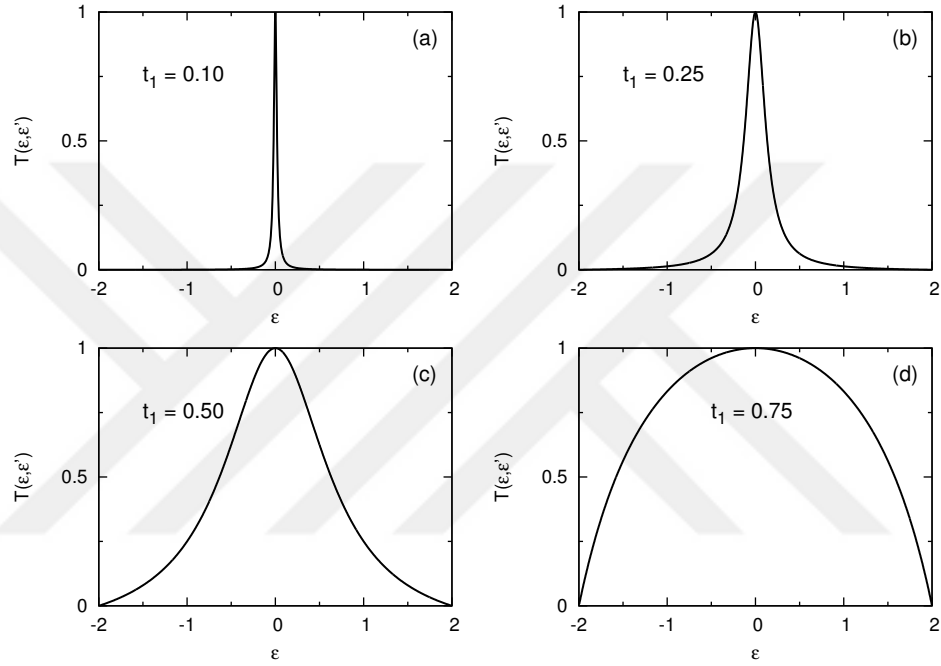


Figure 10.5: The transmission coefficient $T(\varepsilon, \varepsilon')$ versus the incoming electron energy ε for the 1D linear chain with single impurity for different values the hopping constant t_1 mediating the transmission between the central site and the left and right leads (see Fig. 10.1). The other common system parameters are $t = 1.0$, $v_c \equiv v_0 = 0$, and $v = v' = 0$.

It should now be obvious to the reader that if t_1 is infinitesimally small, but not equal to zero, only the very electron whose energy ε is *exactly* equal to the central gate potential v_c will be transmitted to the right lead. In the most extreme case in which $t_1 = 0$, all of the electrons will be reflected back, and none will be transmitted.

So much is enough for the present so-simple system with one single impurity. In the next chapter we will study another 1D system containing two impurities.

10.2 Transport Through Two Horizontal Sites

We now consider the transport properties of one single electron through two central impurities, connected to each other and to the leads in succession, as the one depicted in Fig. 10.6.

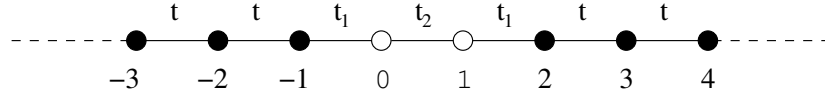


Figure 10.6: A 1D (linear) atomic chain containing two “impurities” at sites $j = 0$ and $j = 1$.

The parameters of this system are the same as those used for the system with one-site impurity that we have considered above (Fig. 10.1). The only additional parameter for the present system is the hopping constant t_2 that mediates the transport of the single electron between the two central (impurity) sites:

$$t_2 \equiv t_{01} = t_{10}$$

We will suppose also that the same gate potential v_c is applied too these two central sites:

$$v_c \equiv v_0 = v_1$$

We are again after the transmission, T , and reflection, R , amplitudes. It seems sufficient for us just to practice the same lines of reasoning that we have followed for the previous one-site problem. The transmission amplitude T for the present system follows from (10.22) with one small modification as

$$T = \frac{t_1}{t} \phi_1 \quad (10.28)$$

whereas the reflection amplitude (10.25) does not need any change, given again as

$$R = \frac{t_1}{t} \phi_0 - 1 \quad (10.29)$$

In other words, if we are to determine T and R , we must find out first ϕ_0 and ϕ_1 .

We already know the “edge” solutions ϕ_{-1} and ϕ_2 of the right and left leads which are the nearest neighbors of the two central sites. With one necessary modification in the latter, they follow from (10.16) and (10.17) respectively as

$$\begin{aligned}\phi_{-1} &= \frac{t_1}{t} e^{ik} \phi_0 + e^{-ik} - e^{ik} \\ &= \frac{t_1}{t} e^{ik} \phi_0 + \beta\end{aligned}\tag{10.30}$$

$$\phi_2 = \frac{t_1}{t} e^{ik'} \phi_1\tag{10.31}$$

where in the former we have let, for the sake of simplicity,

$$\beta = e^{-ik} - e^{ik}\tag{10.32}$$

To write down the equations for ϕ_0 and ϕ_1 , we use our main transport equation (10.2), which we repeat it here:

$$E\phi_j = v_j \phi_j - \sum_{\langle j,k \rangle} t_{jk} \phi_k\tag{10.33}$$

We then have for site $j = 0$ that

$$\begin{aligned}E\phi_0 &= v_0\phi_0 - t_1\phi_{-1} - t_2\phi_1 \\ (E - v_0)\phi_0 &= -t_1\phi_{-1} - t_2\phi_1 \\ \therefore \quad \alpha\phi_0 &= -t_1\phi_{-1} - t_2\phi_1\end{aligned}\tag{10.34}$$

where we have let

$$\alpha = E - v_0\tag{10.35}$$

Similarly, the relation for site $j = 1$ results from (10.33) as

$$\begin{aligned}E\phi_1 &= v_0\phi_1 - t_2\phi_0 - t_1\phi_2 \\ (E - v_0)\phi_1 &= -t_2\phi_0 - t_1\phi_2 \\ \therefore \quad \alpha\phi_1 &= -t_2\phi_0 - t_1\phi_2\end{aligned}\tag{10.36}$$

Putting (10.30) into (10.34), we advance as

$$\begin{aligned}
\alpha\phi_0 &= -t_1\phi_{-1} - t_2\phi_1 \\
&= -t_1\left(\frac{t_1}{t}e^{ik}\phi_0 + \beta\right) - t_2\phi_1 \\
&= -\frac{t_1^2}{t}e^{ik}\phi_0 - t_1\beta - t_2\phi_1 \\
\therefore \quad \left(\alpha + \frac{t_1^2}{t}e^{ik}\right)\phi_0 &= -t_2\phi_1 - t_1\beta
\end{aligned}$$

Multiplying each side by t , we obtain

$$(t\alpha + t_1^2e^{ik})\phi_0 = -tt_2\phi_1 - tt_1\beta \quad (10.37)$$

Similarly, putting (10.31) into (10.36) gives

$$\begin{aligned}
\alpha\phi_1 &= -t_2\phi_0 - t_1\phi_2 \\
&= -t_2\phi_0 - t_1 \cdot \frac{t_1}{t}e^{ik'}\phi_1 \\
&= -t_2\phi_0 - \frac{t_1^2}{t}e^{ik'}\phi_1
\end{aligned}$$

Multiply, again, each side by t to get

$$\begin{aligned}
t\alpha\phi_1 &= -tt_2\phi_0 - t_1^2e^{ik'}\phi_1 \\
\therefore \quad (t\alpha + t_1^2e^{ik'})\phi_1 &= -tt_2\phi_0 \quad (10.38)
\end{aligned}$$

Now we first multiply (10.37) by $-tt_2$ and use (10.38) for $-tt_2\phi_0$. We have

$$\begin{aligned}
(t\alpha + t_1^2e^{ik})\phi_0 &= -tt_2\phi_1 - tt_1\beta \\
(t\alpha + t_1^2e^{ik})(-tt_2\phi_0) &= t^2t_2^2\phi_1 + t^2t_1t_2\beta \\
(t\alpha + t_1^2e^{ik}) \cdot (t\alpha + t_1^2e^{ik'})\phi_1 &= t^2t_2^2\phi_1 + t^2t_1t_2\beta \\
\left[(t\alpha + t_1^2e^{ik})(t\alpha + t_1^2e^{ik'}) - t^2t_2^2\right]\phi_1 &= t^2t_1t_2\beta
\end{aligned}$$

At last we obtain ϕ_1 as

$$\phi_1 = \frac{t^2t_1t_2\beta}{(t\alpha + t_1^2e^{ik})(t\alpha + t_1^2e^{ik'}) - t^2t_2^2} \quad (10.39)$$

Similarly, we multiply (10.38) by $-tt_2$ and use (10.37) for $-tt_2\phi_1$. We proceed as

$$\begin{aligned}
(t\alpha + t_1^2 e^{ik'})\phi_1 &= -tt_2\phi_0 \\
(t\alpha + t_1^2 e^{ik'})(-tt_2\phi_1) &= t^2 t_2^2 \phi_0 \\
(t\alpha + t_1^2 e^{ik'}) \cdot \left[(t\alpha + t_1^2 e^{ik})\phi_0 + tt_1\beta \right] &= t^2 t_2^2 \phi_0 \\
(t\alpha + t_1^2 e^{ik'}) (t\alpha + t_1^2 e^{ik})\phi_0 + (t\alpha + t_1^2 e^{ik'}) tt_1\beta &= t^2 t_2^2 \phi_0 \\
\left[t^2 t_2^2 - (t\alpha + t_1^2 e^{ik'}) (t\alpha + t_1^2 e^{ik}) \right] \phi_0 &= tt_1 (t\alpha + t_1^2 e^{ik'})\beta
\end{aligned}$$

And we obtain ϕ_0 as

$$\phi_0 = \frac{tt_1 (t\alpha + t_1^2 e^{ik'})\beta}{t^2 t_2^2 - (t\alpha + t_1^2 e^{ik'}) (t\alpha + t_1^2 e^{ik})} \quad (10.40)$$

Finally we can obtain the transmission and reflection amplitudes by feeding (10.39) into (10.28) and (10.40) into (10.29), respectively, as

$$T = \frac{tt_1^2 t_2 \beta}{(t\alpha + t_1^2 e^{ik}) (t\alpha + t_1^2 e^{ik'}) - t^2 t_2^2} \quad (10.41)$$

and

$$R = \frac{t_1^2 (t\alpha + t_1^2 e^{ik'})\beta}{t^2 t_2^2 - (t\alpha + t_1^2 e^{ik'}) (t\alpha + t_1^2 e^{ik})} - 1 \quad (10.42)$$

with

$$\alpha = E - v_0 \quad \text{and} \quad \beta = e^{-ik} - e^{ik} \quad (10.43)$$

As usual, the transmission and reflection coefficients are calculated via (10.7) and (10.6), to repeat,

$$\mathsf{T} = |T|^2 \frac{\sin k'}{\sin k} \quad (10.44)$$

$$\mathsf{R} = |R|^2 \quad (10.45)$$

As we have done before for the previous problem with central impurity site, we again think that the single electron comes from the left lead with an “energy” ε and continue with another “energy” ε' through the right lead:

$$\varepsilon = -2t \cos k \quad \text{and} \quad \varepsilon' = -2t \cos k' \quad (10.46)$$

with the energy conservation

$$E = v + \varepsilon = v' + \varepsilon' \quad (10.47)$$

We can now pass to the numerical results. We again wrote a series of MATHEMATICA[®] codes. Referring to Fig. 10.6, the system parameters that we use are $t = 1.0$, $t_1 = 0.2$, $t_2 = 1.0$, $v_c \equiv v_0 = v_1 = 0.0$, and $v = v' = 0$.

In Fig. 10.7 is the first result of numerical simulation: the plots of the transmission coefficient $T(\varepsilon, \varepsilon')$ and the reflection coefficient $R(\varepsilon, \varepsilon)$ drawn against the incoming electron energy ε for the system. We immediately notice that the sum of the two coefficients is always unity, $T + R = 1$, as it must be if our derivation is to be correct.

Using the plot for the transmission coefficient T in Fig. 10.7 (a), we can check the validity of our previous hypothesis (10.27), that was saying that if the energy ε of the incoming electron is around about the gate potential v_c of the central site (or region), it can hop from the left lead to the central site and from there it can hop to the right lead, otherwise it will be reflected back. If this hypothesis is correct for the system with two central sites, we might expect the transmission plot to have only one peak around the central gate potential $v_c = 0.0$. But Fig. 10.7 (a) exhibits something very different: we see two peaks around $\varepsilon = \pm 1$ and do not see any peak around $\varepsilon = 0$ at all! Is hypothesis (10.27) not correct?

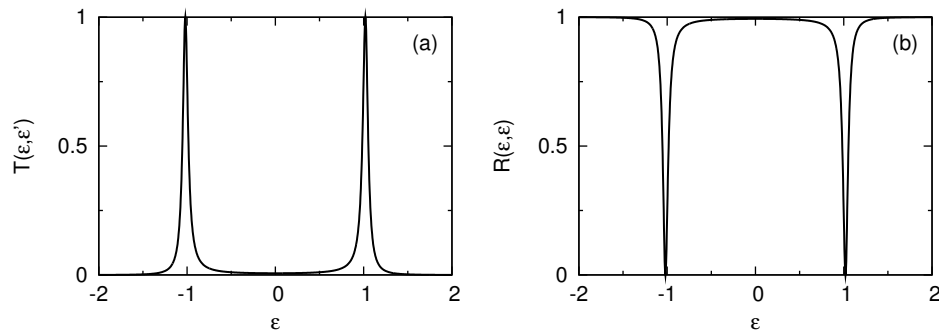


Figure 10.7: (a) The transmission coefficient $T(\varepsilon, \varepsilon')$ and (b) the reflection coefficient $R(\varepsilon, \varepsilon)$ versus the incoming electron energy ε for the 1D linear chain with two impurities. The common system parameters are $t = 1.0$, $t_1 = 0.2$, $t_2 = 1.0$, $v_c \equiv v_0 = v_1 = 0$, and $v = v' = 0$.

In order to resolve this difficulty, let us have a closer look at Fig. 10.7 (a). We see *two* peaks that are *symmetric* with respect to the origin, $\varepsilon = 0$, of the graph. Actually we have witnessed, though not explicitly, such a situation before in discussing the 1D linear chain problem. We observe for the present system that *the two peaks correspond to the **eigenenergies** E_c , given in (9.18), of the central two-site linear chain!*. These are given by

$$\begin{aligned} E_c &= v_c + \varepsilon \\ &= v_c - 2t_2 \cos k_c \\ &= v_c - 2t_2 \cos \frac{n\pi}{N+1} \quad (n = 1, 2, 3, \dots, N) \end{aligned} \quad (10.48)$$

which becomes for the case at hand with $v_c = 0.0$ and $N = 2$ that

$$E_c = -2t_2 \cos \frac{n\pi}{3} \quad (n = 1, 2) \quad (10.49)$$

or, explicitly,

$$E_c = \begin{cases} -2t_2 \cos \frac{\pi}{3} \\ -2t_2 \cos \frac{2\pi}{3} \end{cases} \quad (10.50)$$

and, with $t_2 = 1.0$, we have $E_c = \pm 1.0$ and these are just the locations of the two peaks that we have observed in Fig. 10.7 (a). It seems that we have resolved the trouble.

What about hypothesis (10.27)? Actually it was also *correct* if we happen to see that the eigenenergy of a 1D “chain” with single site is just equal to the gate potential applied to it:

$$\begin{aligned} E_c &= v_c - 2t_2 \cos \frac{n\pi}{N+1} \quad (n = 1) \\ &= v_c - 2t_2 \cos \frac{\pi}{2} \\ &= v_c \end{aligned}$$

Then, we can correct hypothesis (10.27) as a now-well-established corollary as the following:

$$\begin{aligned} \text{COROLLARY:} \quad &\text{if } \varepsilon \sim E_c, \text{ the electron can hop to the right lead;} \\ &\text{otherwise, it is reflected back.} \end{aligned} \quad (10.51)$$

In other words, whenever the energy ε of the single electron is around (*not necessarily equal to*) the eigenenergies E_c of the central region, it is probable that it hops from the left lead to the central region and from there to the right lead. Otherwise it will be reflected back through the left lead.

What is wonderful about this corollary, which is highly promising in designing any nanoelectronic device, is that the single electron has now N different *energy channels*—corresponding to the eigenenergies (10.48)—for hopping from the left lead to the right lead if they are connected to a 1D linear chain with N sites. And, moreover, what is best is that the numerical values of these energy channels can be easily modulated, via (10.48), by changing the gate potential v_c (an easy job) or the hopping amplitude t_2 between any two next-neighboring central sites (a not-so-easy job). These two features are beautifully demonstrated in Fig. 10.8 for the present system with two central impurities. For the numerical values given in Fig. 10.8 (a)–(d), the eigenenergies follow from (10.48), with $n = 1$ and 2 , as

$$\begin{aligned}
 E_c &= 0 - (2)(0.25) \cos \frac{n\pi}{3} = \pm 0.25 && \text{[for (a)]} \\
 &= 0 - (2)(0.75) \cos \frac{n\pi}{3} = \pm 0.25 && \text{[for (b)]} \\
 &= 1 - (2)(0.25) \cos \frac{n\pi}{3} = 0.75, 1.25 && \text{[for (c)]} \\
 &= 1 - (2)(0.75) \cos \frac{n\pi}{3} = 0.25, 1.75 && \text{[for (d)]}
 \end{aligned}$$

and these are all the energies ε of the single electron spotted in Fig. 10.8 (a)–(d); whenever it has an energy around about these values, it can be transported from the left lead to the right lead.

The gate potential v_c and the hopping amplitude t_2 are not the only means to play with the transport properties of the electron. As you remember, we have another mean to increase or decrease the probability that the electron can be transported from left to the right over the central impurity region; it is the hopping constant t_1 modulating the transport between the leads and the central region. Figure 10.9 shows the plots of the transmission coefficient $T(\varepsilon, \varepsilon')$ versus the incoming electron energy ε for the present system under consideration for different values t_1 . As we have seen previously for the system with single central site (see Fig. 10.5), when t_1 is so small (but *not* equal to zero), if and only if the energy ε of the incoming electron is *very* close to one of

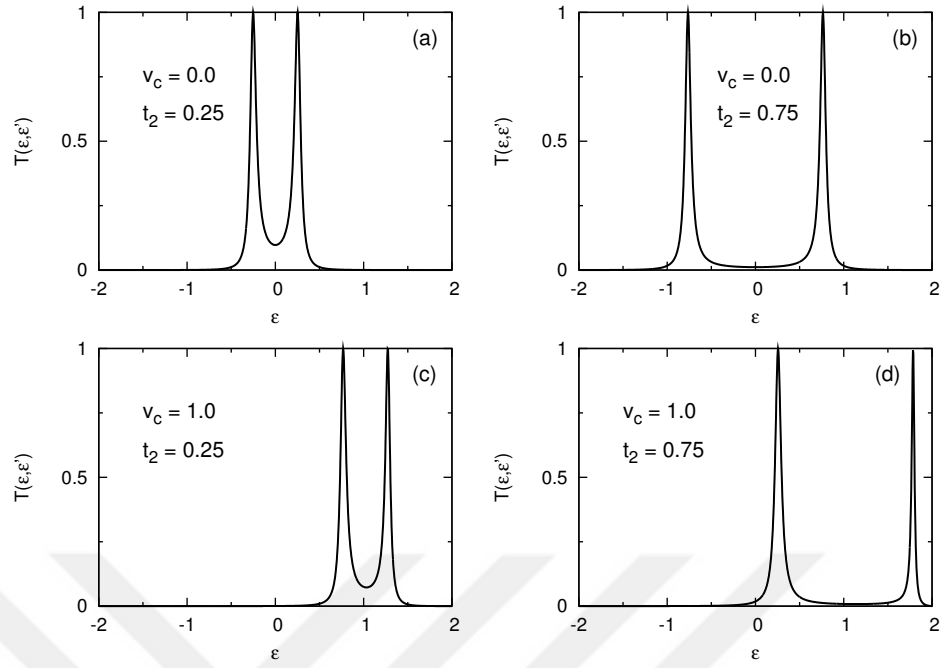


Figure 10.8: The transmission coefficient $T(\varepsilon, \varepsilon')$ versus the incoming electron energy ε for the 1D linear chain with two central impurities for different values the hopping constant t_2 mediating the transmission between the two central sites (see Fig. 10.6) and the gate potentials $v_c \equiv v_0 = v_1$. The other common system parameters are $t = 1.0$, $t_1 = 0.2$ and $v = v' = 0$.

the eigenenergies of the central part, it can be transported to the right. If t_1 is large, and if ε is close (*not necessarily very close*) enough to one of the eigenenergies of the central part, the electron can be readily transported. If we speak of a beam of electrons incident through the left lead towards the central impurity region, the last two facts can be formulated differently: if the hopping constant t_1 is small (large), the fraction of electrons transported from left to right is also small (large); this simply means, in turn, that the amount of “current” flowing through the right lead is also small (large).

We have now come to understand that the present simple system with two impurities has already opened an auspicious gate for the entrance to the (quantum mechanical) nanoelectronic devices. The reader should have noticed that the passage from the system with single central site to the other with two central sites is accomplished merely by an additional extra site. And it is this additional extra site that has caused the system to have an incomparably high flexibility to be manipulated according to the some pre-desired end results.

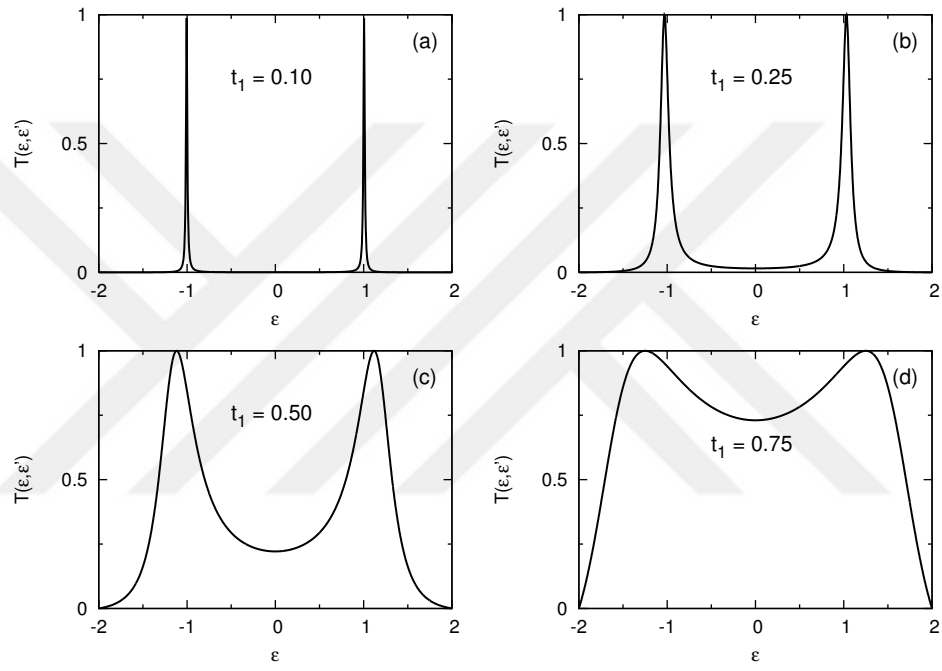


Figure 10.9: The transmission coefficient $T(\varepsilon, \varepsilon')$ versus the incoming electron energy ε for the 1D linear chain with two central impurities for different values the hopping constant t_1 mediating the transmission between the central impurity region and the left and right leads (see Fig. 10.6). The other common system parameters are $t = 1.0$, $t_2 = 1.0$, $v_c \equiv v_0 = v_1 = 0$, and $v = v' = 0$.

10.3 Transport Through Three Horizontal Sites

We now study the transport problem of one single electron through three central impurities, connected to the two leads and to each other in series, as shown in Fig. 10.10. Since the reader is supposed to have already been used to the way we derive the equations, thanks especially to the last chapter, our derivations and the relevant outcomes will not be discussed in detail.⁴

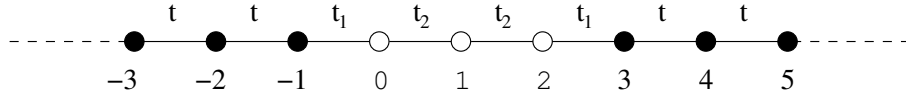


Figure 10.10: A 1D (linear) atomic chain containing three “impurities” at sites $j = 0$, $j = 1$, and $j = 2$.

The parameters of this system are the same as those used for the system with two-site impurity that we have studied thoroughly in the preceding chapter (Fig. 10.6). As is customary we will apply the same gate potential v_c to these three central sites:

$$v_c \equiv v_0 = v_1 = v_2$$

Our aim is to determine analytically the transmission, T , and reflection, R , amplitudes. The former for the present system is modified from (10.22) as

$$T = \frac{t_1}{t} \phi_2 \quad (10.52)$$

and the latter is written from (10.25) without any change as

$$R = \frac{t_1}{t} \phi_0 - 1 \quad (10.53)$$

We thus conclude that we must work out the solutions ϕ_0 and ϕ_2 (and ϕ_1), belonging to the central impurity sites, in order to specify T and R completely.

⁴ We are of course aware of the fact that, except the first one for the system with single central impurity, the second derivation carried out in the preceding chapter and the third one we are about to perform now are somehow arduous, and may seem redundant, too. These types of derivations are referred to as “must-be-done-once-in-a-lifetime—not twice.” We hope the reader will feel assured shortly that s/he will be rewarded with some invaluable insights into the transport problem by practising these “annoying” derivations by hand. Besides, *à force de forger, on devient forgeron*, n’est-ce pas?

The “edge” solutions ϕ_{-1} and ϕ_3 are already known. They are modified from (10.16) and (10.17) as

$$\begin{aligned}\phi_{-1} &= \frac{t_1}{t} e^{ik} \phi_0 + e^{-ik} - e^{ik} \\ &= \frac{t_1}{t} e^{ik} \phi_0 + \beta\end{aligned}\tag{10.54}$$

$$\phi_3 = \frac{t_1}{t} e^{ik'} \phi_2\tag{10.55}$$

where in the former we have used the abbreviation

$$\beta = e^{-ik} - e^{ik}\tag{10.56}$$

We make use the main transport equation

$$E\phi_j = v_j \phi_j - \sum_{\langle j,k \rangle} t_{jk} \phi_k\tag{10.57}$$

to write down the equations for ϕ_0 , ϕ_1 , and ϕ_2 .

The equation for site $j = 0$ is

$$\begin{aligned}E\phi_0 &= v_0\phi_0 - t_1\phi_{-1} - t_2\phi_1 \\ (E - v_0)\phi_0 &= -t_1\phi_{-1} - t_2\phi_1 \\ \therefore \quad \alpha\phi_0 &= -t_1\phi_{-1} - t_2\phi_1\end{aligned}\tag{10.58}$$

where

$$\alpha = E - v_0\tag{10.59}$$

Similarly, the equations for sites $j = 1$ and $j = 2$ are

$$\begin{aligned}E\phi_1 &= v_0\phi_1 - t_2\phi_0 - t_2\phi_2 \\ (E - v_0)\phi_1 &= -t_2\phi_0 - t_2\phi_2 \\ \therefore \quad \alpha\phi_1 &= -t_2\phi_0 - t_2\phi_2\end{aligned}\tag{10.60}$$

and

$$\begin{aligned}E\phi_2 &= v_0\phi_2 - t_2\phi_1 - t_1\phi_3 \\ (E - v_0)\phi_2 &= -t_2\phi_1 - t_1\phi_3 \\ \therefore \quad \alpha\phi_2 &= -t_2\phi_1 - t_1\phi_3\end{aligned}\tag{10.61}$$

Feeding (10.54) into (10.58), we proceed as

$$\begin{aligned}
\alpha\phi_0 &= -t_1\phi_{-1} - t_2\phi_1 \\
&= -t_1\left(\frac{t_1}{t}e^{ik}\phi_0 + \beta\right) - t_2\phi_1 \\
&= -\frac{t_1^2}{t}e^{ik}\phi_0 - t_1\beta - t_2\phi_1 \\
\therefore \quad \left(\alpha + \frac{t_1^2}{t}e^{ik}\right)\phi_0 &= -t_2\phi_1 - t_1\beta
\end{aligned}$$

Multiplication of each side by t yields

$$(t\alpha + t_1^2e^{ik})\phi_0 = -tt_2\phi_1 - tt_1\beta \quad (10.62)$$

Similarly, feeding (10.55) into (10.61) gives

$$\begin{aligned}
\alpha\phi_2 &= -t_2\phi_1 - t_1\phi_3 \\
&= -t_2\phi_1 - t_1 \cdot \frac{t_1}{t}e^{ik'}\phi_2 \\
&= -t_2\phi_1 - \frac{t_1^2}{t}e^{ik'}\phi_2
\end{aligned}$$

Again, multiplying each side by t we get

$$\begin{aligned}
t\alpha\phi_2 &= -tt_2\phi_1 - t_1^2e^{ik'}\phi_2 \\
\therefore \quad (t\alpha + t_1^2e^{ik'})\phi_2 &= -tt_2\phi_1 \quad (10.63)
\end{aligned}$$

We have three equations for ϕ_0 , ϕ_1 , and ϕ_2 expressed in (10.62), (10.60), and (10.63), respectively. Let us write them again altogether:

$$(t\alpha + t_1^2e^{ik})\phi_0 = -tt_2\phi_1 - tt_1\beta \quad (10.64)$$

$$\alpha\phi_1 = -t_2\phi_0 - t_2\phi_2 \quad (10.65)$$

$$(t\alpha + t_1^2e^{ik'})\phi_2 = -tt_2\phi_1 \quad (10.66)$$

Let us eliminate ϕ_1 . To do this, we first multiply (10.64) by α and then use (10.65) for $\alpha\phi_1$. We proceed as

$$\begin{aligned}
(t\alpha + t_1^2 e^{ik})\phi_0 &= -tt_2\phi_1 - tt_1\beta \\
(t\alpha^2 + t_1^2 \alpha e^{ik})\phi_0 &= -tt_2 \cdot \alpha\phi_1 - tt_1\alpha\beta \\
(t\alpha^2 + t_1^2 \alpha e^{ik})\phi_0 &= -tt_2(-t_2\phi_0 - t_2\phi_2) - tt_1\alpha\beta \\
(t\alpha^2 + t_1^2 \alpha e^{ik})\phi_0 &= tt_2^2\phi_0 + tt_2^2\phi_2 - tt_1\alpha\beta \\
(t\alpha^2 + t_1^2 \alpha e^{ik} - tt_2^2)\phi_0 &= tt_2^2\phi_2 - tt_1\alpha\beta
\end{aligned} \tag{10.67}$$

Now in similar fashion, we multiply (10.66) by α and then use again (10.65) for $\alpha\phi_1$. We move on as

$$\begin{aligned}
(t\alpha + t_1^2 e^{ik'})\phi_2 &= -tt_2\phi_1 \\
(t\alpha^2 + t_1^2 \alpha e^{ik'})\phi_2 &= -tt_2 \cdot \alpha\phi_1 \\
(t\alpha^2 + t_1^2 \alpha e^{ik'})\phi_2 &= -tt_2(-t_2\phi_0 - t_2\phi_2) \\
(t\alpha^2 + t_1^2 \alpha e^{ik'})\phi_2 &= tt_2^2\phi_0 + tt_2^2\phi_2 \\
(t\alpha^2 + t_1^2 \alpha e^{ik'} - tt_2^2)\phi_2 &= tt_2^2\phi_0
\end{aligned} \tag{10.68}$$

To obtain ϕ_2 , we multiply (10.67) by tt_2^2 and use (10.68) for $tt_2^2\phi_0$. We have

$$\begin{aligned}
(t\alpha^2 + t_1^2 \alpha e^{ik} - tt_2^2)\phi_0 &= tt_2^2\phi_2 - tt_1\alpha\beta \\
(t\alpha^2 + t_1^2 \alpha e^{ik} - tt_2^2) \cdot tt_2^2\phi_0 &= t^2 t_2^4 \phi_2 - t^2 t_1 t_2^2 \alpha\beta \\
(t\alpha^2 + t_1^2 \alpha e^{ik} - tt_2^2)(t\alpha^2 + t_1^2 \alpha e^{ik'} - tt_2^2)\phi_2 &= t^2 t_2^4 \phi_2 - t^2 t_1 t_2^2 \alpha\beta \\
\left[(t\alpha^2 + t_1^2 \alpha e^{ik} - tt_2^2)(t\alpha^2 + t_1^2 \alpha e^{ik'} - tt_2^2) - t^2 t_2^4\right]\phi_2 &= -t^2 t_1 t_2^2 \alpha\beta
\end{aligned}$$

Out of space concerns, we reverse the last line and carry out the multiplications to open the parantheses:

$$\begin{aligned}
-t^2 t_1 t_2^2 \alpha \beta &= \left[(t\alpha^2 + t_1^2 \alpha e^{ik} - tt_2^2) (t\alpha^2 + t_1^2 \alpha e^{ik'} - tt_2^2) - t^2 t_2^4 \right] \phi_2 \\
&= \left[t^2 \alpha^4 + tt_1^2 \alpha^3 e^{ik'} - t^2 t_2^2 \alpha^2 + tt_1^2 \alpha^3 e^{ik} + t_1^4 \alpha^2 e^{i(k+k')} - tt_1^2 t_2^2 \alpha e^{ik} \right. \\
&\quad \left. - t^2 t_2^2 \alpha^2 - tt_1^2 t_2^2 \alpha e^{ik'} + t^2 t_2^4 - t^2 t_2^4 \right] \phi_2 \\
&= \left[t^2 \alpha^4 - 2t^2 t_2^2 \alpha^2 + tt_1^2 \alpha (\alpha^2 - t_2^2) e^{ik} + tt_1^2 \alpha (\alpha^2 - t_2^2) e^{ik'} + t_1^4 \alpha^2 e^{i(k+k')} \right] \phi_2 \\
&= \alpha \left[t^2 \alpha^3 - 2t^2 t_2^2 \alpha + tt_1^2 (\alpha^2 - t_2^2) (e^{ik} + e^{ik'}) + t_1^4 \alpha e^{i(k+k')} \right] \phi_2
\end{aligned}$$

We first reverse the last line and then divide each side by α and obtain

$$\begin{aligned}
\left[t^2 \alpha^3 - 2t^2 t_2^2 \alpha + tt_1^2 (\alpha^2 - t_2^2) (e^{ik} + e^{ik'}) + t_1^4 \alpha e^{i(k+k')} \right] \phi_2 &= -t^2 t_1 t_2^2 \beta \\
\left[t^2 \alpha (\alpha^2 - 2t_2^2) + tt_1^2 (\alpha^2 - t_2^2) (e^{ik} + e^{ik'}) + t_1^4 \alpha e^{i(k+k')} \right] \phi_2 &= -t^2 t_1 t_2^2 \beta
\end{aligned}$$

And we finally obtain ϕ_2 as

$$\phi_2 = \frac{-t^2 t_1 t_2^2 \beta}{t^2 \alpha (\alpha^2 - 2t_2^2) + tt_1^2 (\alpha^2 - t_2^2) (e^{ik} + e^{ik'}) + t_1^4 \alpha e^{i(k+k')}} \quad (10.69)$$

In order to obtain ϕ_0 , we multiply (10.68) by tt_2^2 and use (10.67) for $tt_2^2 \phi_2$. We have

$$\begin{aligned}
(t\alpha^2 + t_1^2 \alpha e^{ik'} - tt_2^2) \phi_2 &= tt_2^2 \phi_0 \\
(t\alpha^2 + t_1^2 \alpha e^{ik'} - tt_2^2) \cdot tt_2^2 \phi_2 &= t^2 t_2^4 \phi_0 \\
(t\alpha^2 + t_1^2 \alpha e^{ik'} - tt_2^2) \left[(t\alpha^2 + t_1^2 \alpha e^{ik} - tt_2^2) \phi_0 + tt_1 \alpha \beta \right] &= t^2 t_2^4 \phi_0 \\
(t\alpha^2 + t_1^2 \alpha e^{ik'} - tt_2^2) (t\alpha^2 + t_1^2 \alpha e^{ik} - tt_2^2) \phi_0 + tt_1 \alpha \beta (t\alpha^2 + t_1^2 \alpha e^{ik'} - tt_2^2) &= t^2 t_2^4 \phi_0 \\
(t\alpha^2 + t_1^2 \alpha e^{ik'} - tt_2^2) (t\alpha^2 + t_1^2 \alpha e^{ik} - tt_2^2) \phi_0 + tt_1 \alpha \beta \left[t_1^2 \alpha e^{ik'} + t (\alpha^2 - t_2^2) \right] &= t^2 t_2^4 \phi_0
\end{aligned}$$

Adopting momentarily an abbreviation as

$$\gamma = tt_1\alpha\beta \left[t_1^2\alpha e^{ik'} + t(\alpha^2 - t_2^2) \right] \quad (10.70)$$

and arranging the terms further, and reversing the order, we have

$$\begin{aligned} -\gamma &= \left[(t\alpha^2 + t_1^2\alpha e^{ik'} - tt_2^2)(t\alpha^2 + t_1^2\alpha e^{ik} - tt_2^2) - t^2t_2^4 \right] \phi_0 \\ &= \left[t^2\alpha^4 + tt_1^2\alpha^3 e^{ik} - t^2t_2^2\alpha^2 + tt_1^2\alpha^3 e^{ik'} + t_1^4\alpha^2 e^{i(k+k')} - tt_1^2t_2^2\alpha e^{ik'} \right. \\ &\quad \left. - t^2t_2^2\alpha^2 - tt_1^2t_2^2\alpha e^{ik} + t^2t_2^4 - t^2t_2^4 \right] \phi_0 \\ &= \left[t^2\alpha^4 - 2t^2t_2^2\alpha^2 + (tt_1^2\alpha^3 - tt_1^2t_2^2\alpha) e^{ik} + (tt_1^2\alpha^3 - tt_1^2t_2^2\alpha) e^{ik'} + t_1^4\alpha^2 e^{i(k+k')} \right] \phi_0 \\ &= \left[t^2\alpha^2(\alpha^2 - 2t_2^2) + (tt_1^2\alpha^3 - tt_1^2t_2^2\alpha)(e^{ik} + e^{ik'}) + t_1^4\alpha^2 e^{i(k+k')} \right] \phi_0 \\ &= \left[t^2\alpha^2(\alpha^2 - 2t_2^2) + tt_1^2\alpha(\alpha^2 - t_2^2)(e^{ik} + e^{ik'}) + t_1^4\alpha^2 e^{i(k+k')} \right] \phi_0 \\ &= \alpha \left[t^2\alpha(\alpha^2 - 2t_2^2) + tt_1^2(\alpha^2 - t_2^2)(e^{ik} + e^{ik'}) + t_1^4\alpha e^{i(k+k')} \right] \phi_0 \end{aligned}$$

Reversing the order and inserting (10.70) back, we obtain

$$\begin{aligned} \alpha \left[t^2\alpha(\alpha^2 - 2t_2^2) + tt_1^2(\alpha^2 - t_2^2)(e^{ik} + e^{ik'}) + t_1^4\alpha e^{i(k+k')} \right] \phi_0 &= -tt_1\alpha\beta \left[t_1^2\alpha e^{ik'} + t(\alpha^2 - t_2^2) \right] \phi_0 \\ \left[t^2\alpha(\alpha^2 - 2t_2^2) + tt_1^2(\alpha^2 - t_2^2)(e^{ik} + e^{ik'}) + t_1^4\alpha e^{i(k+k')} \right] \phi_0 &= -tt_1\beta \left[t_1^2\alpha e^{ik'} + t(\alpha^2 - t_2^2) \right] \phi_0 \end{aligned}$$

At last we obtain ϕ_0 as

$$\phi_0 = \frac{-tt_1\beta \left[t_1^2\alpha e^{ik'} + t(\alpha^2 - t_2^2) \right]}{t^2\alpha(\alpha^2 - 2t_2^2) + tt_1^2(\alpha^2 - t_2^2)(e^{ik} + e^{ik'}) + t_1^4\alpha e^{i(k+k')}} \quad (10.71)$$

Finally the transmission and reflection amplitudes are reached by inserting (10.69) into (10.52) and (10.71) into (10.53), respectively, as

$$T = \frac{-tt_1^2t_2^2\beta}{t^2\alpha(\alpha^2 - 2t_2^2) + tt_1^2(\alpha^2 - t_2^2)(e^{ik} + e^{ik'}) + t_1^4\alpha e^{i(k+k')}} \quad (10.72)$$

and

$$R = \frac{-t_1^2\beta \left[t_1^2\alpha e^{ik'} + t(\alpha^2 - t_2^2) \right]}{t^2\alpha(\alpha^2 - 2t_2^2) + tt_1^2(\alpha^2 - t_2^2)(e^{ik} + e^{ik'}) + t_1^4\alpha e^{i(k+k')}} - 1 \quad (10.73)$$

where

$$\alpha = E - v_0 \quad \text{and} \quad \beta = e^{-ik} - e^{ik} \quad (10.74)$$

Again, the transmission and reflection coefficients are to be calculated through (10.7) and (10.6):

$$T = |T|^2 \frac{\sin k'}{\sin k} \quad (10.75)$$

$$R = |R|^2 \quad (10.76)$$

Now we repeat the lines of the same old story: the single electron originates from the external source, moves through the left lead with an “energy” ε , if circumstances allow, it hops to the central impurity region and hops from there to the right lead, and goes on towards to the external sink with another “energy” ε' :

$$\varepsilon = -2t \cos k \quad \text{and} \quad \varepsilon' = -2t \cos k' \quad (10.77)$$

where the total energy of the system is conserved via

$$E = v + \varepsilon = v' + \varepsilon' \quad (10.78)$$

The last stage for this system is to embed these results into a series of MATHEMATICA[®] codes. The system parameters used are again the same; referring to Fig. 10.10, they are $t = 1.0$, $t_1 = 0.2$, $t_2 = 1.0$, $v_c \equiv v_0 = v_1 = v_2 = 0.0$, and $v = v' = 0$. Figure 10.11 gives the plots of $T(\varepsilon, \varepsilon')$ and $R(\varepsilon, \varepsilon)$, the respective transmission and reflection coefficients, drawn against the energy ε of the incoming electron. As being the first and immediate check, we observe and become pleased with the fact that, for each value of ε , the sum of the two coefficients is consistently unity, $T + R = 1$, which is an acutely strong indication that the above derivation is precise.⁵

⁵ This is not just a glossy conclusion that might appear to the reader. S/he should behold the complexity of (10.72) and (10.73) and appreciate the fact that, with the nudge of a highly possible tiniest flaw, all the derivation, together with the numerical results presented in Fig. 10.10, will right away collapse.

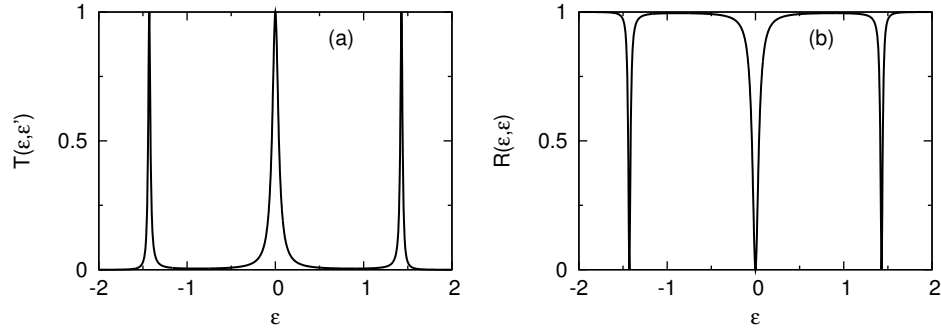


Figure 10.11: (a) The transmission coefficient $T(\varepsilon, \varepsilon')$ and (b) the reflection coefficient $R(\varepsilon, \varepsilon)$ versus the incoming electron energy ε for the 1D linear chain with three impurities. The common system parameters are $t = 1.0$, $t_1 = 0.2$, $t_2 = 1.0$, $v_c \equiv v_0 = v_1 = v_2 = 0.0$, and $v = v' = 0$.

As we expect, we see in Fig. 10.11 (a) *three* peaks which are *symmetric* with respect to the origin. They correspond to the eigenenergies E_c of the central three-site linear chain, and are given generally by

$$\begin{aligned} E_c &= v_c - 2t_2 \cos k_c \\ &= v_c - 2t_2 \cos \frac{n\pi}{N+1} \quad (n = 1, 2, 3, \dots, N) \end{aligned} \quad (10.79)$$

which, for the present system with $v_c = 0.0$ and $N = 3$, assumes the form

$$E_c = -2t_2 \cos \frac{n\pi}{4} \quad (n = 1, 2, 3)$$

or, with $t_2 = 1.0$,

$$E_c = \begin{cases} -2 \cos \frac{\pi}{4} = -\sqrt{2} \approx -1.41 \\ -2 \cos \frac{2\pi}{4} = 0 \\ -2 \cos \frac{3\pi}{4} = +\sqrt{2} \approx +1.41 \end{cases}$$

and these are exactly the positions of the three peaks discerned in Fig. 10.11 (a). All in this plot goes as anticipated: if the energy ε of the single incoming electron is about any of the eigenenergies E_c of the central impurity region, it will be transmitted to right lead; if not, it will be reflected back through the left lead.

In Fig. 10.12 are shown the plots of the transmission coefficient $T(\varepsilon, \varepsilon')$ versus the energy ε incoming electron for the present system for different values the hopping constant t_2 , between any of the next-neighboring three central sites and the gate potentials $v_c \equiv v_0 = v_1 = v_2$. Using the numerical values given in these plots, we predict the eigenenergies of the central region from (10.79), with $n = 1, 2$, and 3 , to three-digit accuracy as

$$E_c = 0 - (2)(0.25) \cos \frac{n\pi}{4} = -0.354, 0, 0.354 \quad [\text{for (a)}]$$

$$= 0 - (2)(0.75) \cos \frac{n\pi}{4} = -1.06, 0, 1.06 \quad [\text{for (b)}]$$

$$= 1 - (2)(0.25) \cos \frac{n\pi}{4} = 0.646, 1, 1.35 \quad [\text{for (c)}]$$

$$= 1 - (2)(0.75) \cos \frac{n\pi}{4} = -0.061, 1, 2.06 \quad [\text{for (d)}]$$

These are all but one the energies ε of the single electron that can be readily discerned the plots of Fig. 10.12. Whenever the incoming electron has an energy around about these values, its transmission to the right lead will take place. The reader is to notice that the plot in Fig. 10.12 (d) has only two peaks, instead of three. Is something wrong? No. The missing eigenenergy in that plot has the numerical value 2.06 and the electron is not allowed to possess this energy; it is outside the permitted energy band between $-2t$ and $+2t$ (or, with $t = 1$, between -2 and $+2$), therefore we cannot see it in that plot.

In Fig. 10.13 we present the plots of the transmission coefficient $T(\varepsilon, \varepsilon')$ drawn against the incoming electron energy ε for the present system for different values t_1 , which mediates the transport of the single electron between the leads and the central impurity region. Everything is just we expect. When $t_1 \ll 1$ (but *not* equal to zero), only when energy ε of the incoming electron is nearly to equal to one of the eigenenergies of the central region, it can pass to the right lead. When t_1 is large, on the other hand, and when the numerical value of ε is sufficiently near one of the eigenenergies of the central region, the electron can readily pass to the right lead with ease.

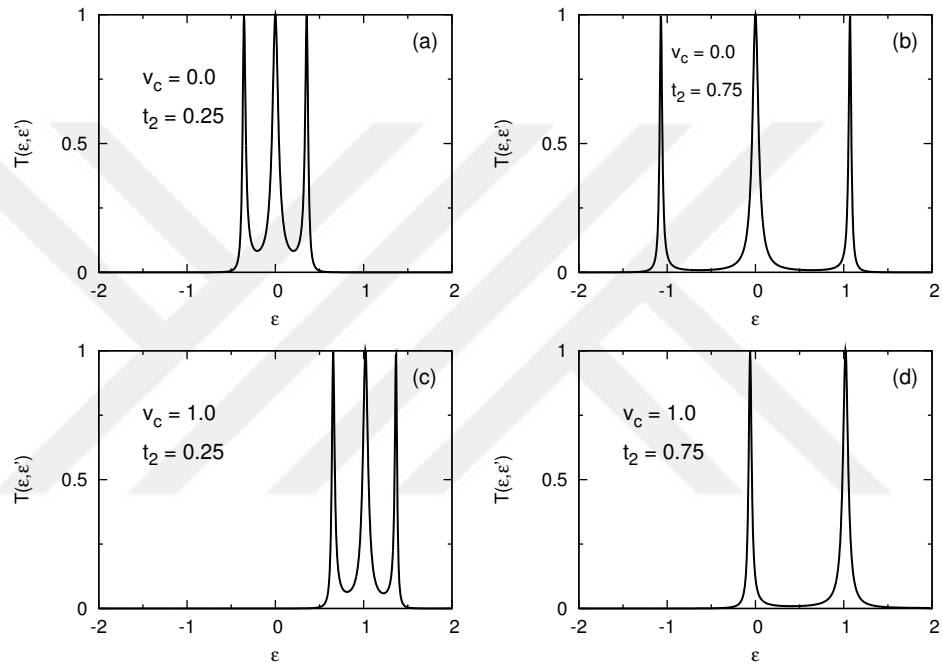


Figure 10.12: The transmission coefficient $T(\varepsilon, \varepsilon')$ versus the incoming electron energy ε for the 1D linear chain with three central impurities for different values the hopping constant t_2 mediating the transmission between any of the next-neighboring three central sites (see Fig. 10.10) and the gate potentials $v_c \equiv v_0 = v_1 = v_2$. The other common system parameters are $t = 1.0$, $t_1 = 0.2$ and $v = v' = 0$.

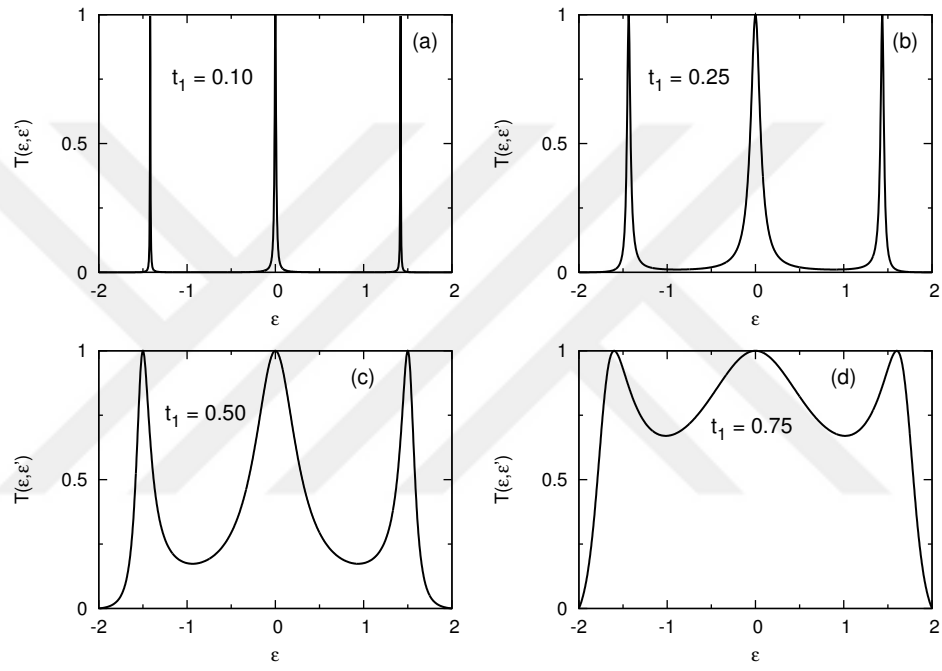


Figure 10.13: The transmission coefficient $T(\varepsilon, \varepsilon')$ versus the incoming electron energy ε for the 1D linear chain with three impurities for different values the hopping constant t_1 mediating the transmission between the central impurity region and the left and right leads (see Fig. 10.10). The other common system parameters are $t = 1.0$, $t_2 = 1.0$, $v_c \equiv v_0 = v_1 = v_2 = 0$, and $v = v' = 0$.

10.4 Transport Through Four Horizontal Sites

The next study that we deal with is the transport problem through four central impurities, as is depicted in Fig. 10.14. Because the reader has already been acquainted with all the derivations and all the important results, we here will not expound anymore on them; instead, we provide only the relevant plots in Figs. 10.15–10.17.

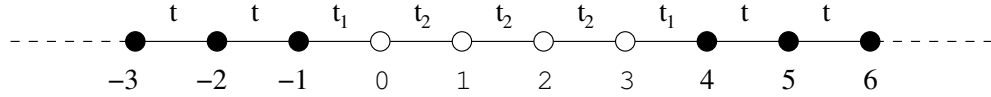


Figure 10.14: A 1D (linear) atomic chain containing four “impurities” at sites $j = 0$ –4.

The transport amplitude T of this system is calculated to be

$$T = \frac{t_1^2 t_2^3 \beta}{t^2 (t_2^4 - 3t_2^2 \alpha^2 + \alpha^4) + t t_1^2 \alpha (\alpha^2 - 2t_2^2) (e^{ik} + e^{ik'}) + t_1^4 (\alpha^2 - t_2^2) e^{i(k+k')}} \quad (10.80)$$

which we give only for the reader’s ease of reference.

As usual Fig. 10.15 indicates that the compulsory condition $T + R = 1$ always holds.

The eigenenergies in Fig. 10.15 (a) are to be evaluated from

$$\begin{aligned} E_c &= v_c - 2t_2 \cos k_c \\ &= v_c - 2t_2 \cos \frac{n\pi}{N+1} \quad (n = 1, 2, 3, \dots, N) \end{aligned} \quad (10.81)$$

which, for the present system with $v_c = 0.0$ and $N = 4$, is seen to take the form

$$E_c = -2t_2 \cos \frac{n\pi}{4} \quad (n = 1, 2, 3, 4)$$

and, with $t_2 = 1.0$,

$$E_c = \begin{cases} -2 \cos \frac{\pi}{4} \approx -1.62 \\ -2 \cos \frac{2\pi}{4} \approx -0.618 \\ -2 \cos \frac{3\pi}{4} \approx +0.618 \\ -2 \cos \frac{4\pi}{4} \approx +1.62 \end{cases}$$

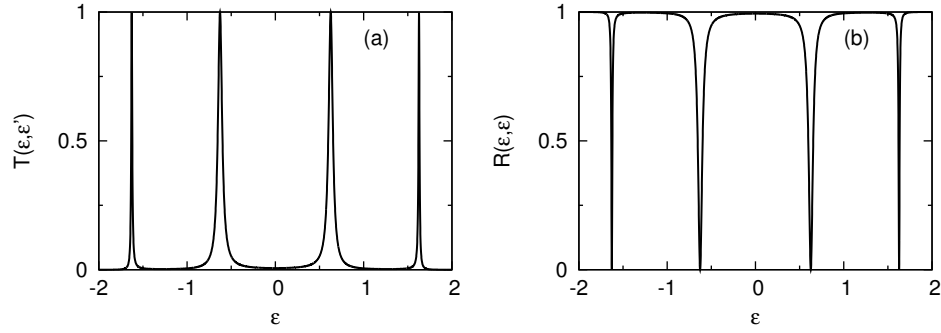


Figure 10.15: (a) The transmission coefficient $T(\epsilon, \epsilon')$ and (b) the reflection coefficient $R(\epsilon, \epsilon)$ versus the incoming electron energy ϵ for the 1D linear chain with four impurities. The common system parameters are $t = 1.0$, $t_1 = 0.2$, $t_2 = 1.0$, $v_c \equiv v_0 = v_1 = v_2 = v_3 = 0.0$, and $v = v' = 0$.

We see again in Fig. 10.16 the “sifting” properties of the central gate potential v_c and the hopping amplitude t_2 for the energies of the incoming electron. With $n = 1, \dots, 4$, the eigenenergies spotted in these plots follow from (10.81), to three-digit accuracy, as

$$\begin{aligned}
 E_c &= 0 - (2)(0.25) \cos \frac{n\pi}{5} = \pm 0.405, \pm 0.155 && \text{[for (a)]} \\
 &= 0 - (2)(0.75) \cos \frac{n\pi}{5} = \pm 1.21, \pm 0.464 && \text{[for (b)]} \\
 &= 1 - (2)(0.25) \cos \frac{n\pi}{5} = 0.595, 0.845, 1.15, 1.40 && \text{[for (c)]} \\
 &= 1 - (2)(0.75) \cos \frac{n\pi}{5} = -0.214, 0.536, 1.46, 2.21 && \text{[for (d)]}
 \end{aligned}$$

It is to be noticed that the last eigenvalue, 2.21, is not seen in Fig. 10.16 (d) because it is not included in the permitted energy band, between -2 and $+2$, of the incoming electron.

And finally, Fig. 10.17 illustrates once more the “sifting” property of the hopping constant t_1 for the number of electron(s) hopping from left lead to the right lead through the central impurity region.

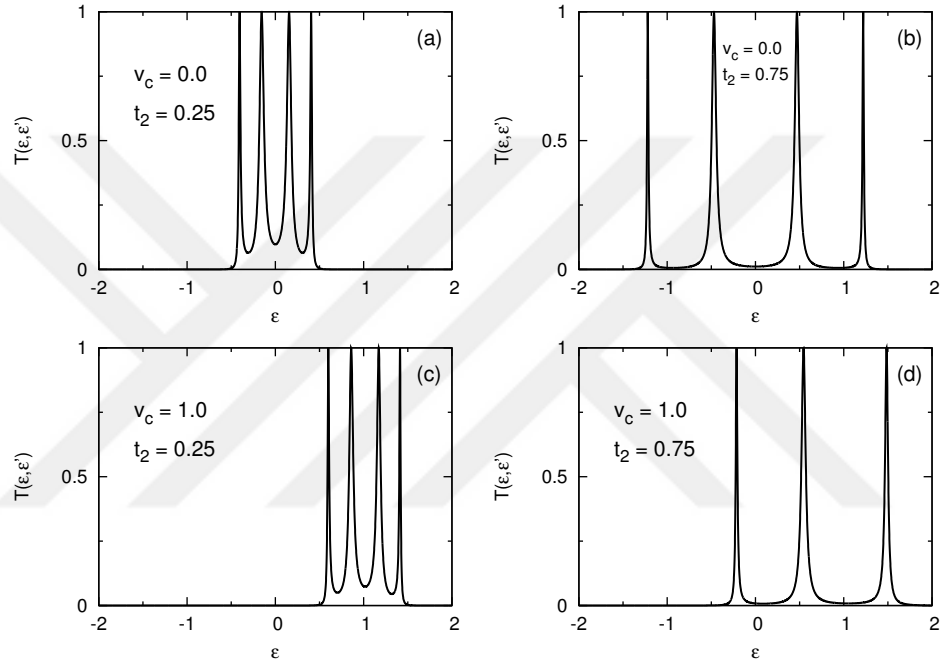


Figure 10.16: The transmission coefficient $T(\varepsilon, \varepsilon')$ versus the incoming electron energy ε for the 1D linear chain with four central impurities for different values the hopping constant t_2 mediating the transmission between any of the next-neighboring three central sites (see Fig. 10.14) and the gate potentials $v_c \equiv v_0 = v_1 = v_2 = v_3$. The other common system parameters are $t = 1.0$, $t_1 = 0.2$ and $v = v' = 0$.

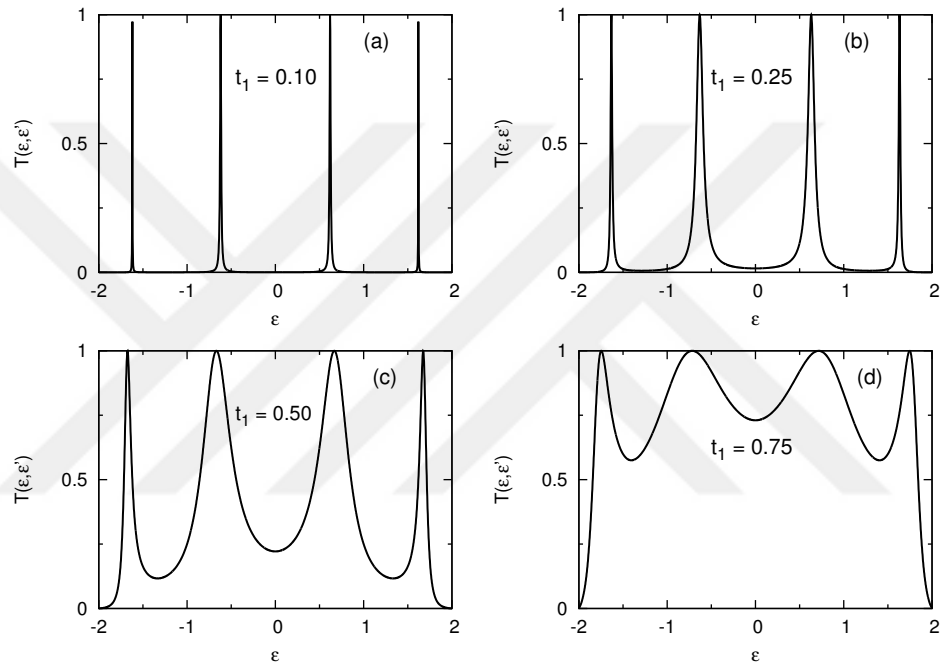


Figure 10.17: The transmission coefficient $T(\varepsilon, \varepsilon')$ versus the incoming electron energy ε for the 1D linear chain with four impurities for different values the hopping constant t_1 mediating the transmission between the central impurity region and the left and right leads (see Fig. 10.14). The other common system parameters are $t = 1.0$, $t_2 = 1.0$, $v_c \equiv v_0 = v_1 = v_2 = v_3 = 0$, and $v = v' = 0$.

10.5 Transport Through a Four-Site Ring Structure

The last study in this chapter is about the transport problem through four ring-like central impurities, like the one shown in Fig. 10.18. This simple system will be a *discrete* version of the so-called Aharonov-Bohm ring structures that we will investigate later on both theoretically and numerically in great detail. Therefore, for the time being, within the scope of this chapter, it will suffice for us to present only the general results, just as we have done routinely for the previous 1D linear chains with one- to four-site central impurities.

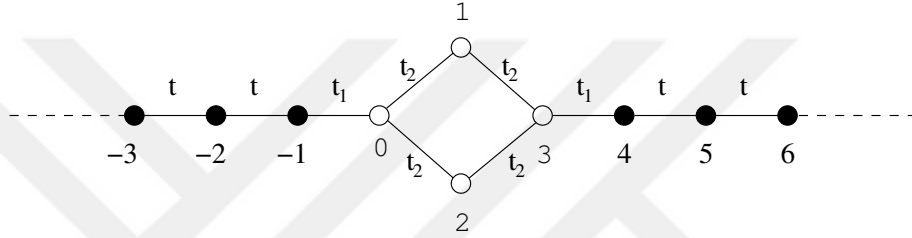


Figure 10.18: Two 1D (linear) atomic chains connected to a central ring-like structure containing four “impurities” at sites $j = 0-4$.

Again, we give only the transport amplitude T of this system for ease of reference:

$$T = \frac{-2tt_1^2t_2^2\beta}{t^2\alpha(\alpha^2 - 4t_2^2) + tt_1^2(\alpha^2 - 2t_2^2)(e^{ik} + e^{ik'}) + t_1^4\alpha e^{i(k+k')}} \quad (10.82)$$

The alert reader will immediately notice that the above expression is decidedly, somehow, similar to the transport amplitude of the system with three horizontal central impurity sites [see (10.72)].

Figure 10.19 (where, in order to observe more clearly the two edge eigenenergy values around ± 2.0 , we have taken $t_2 = 0.9$ instead of 1.0 as we have done so far) manifests again, one more time, that the unavoidable requirement $T + R = 1$ is satisfied, indeed, for every energy value, ε , of the incoming electron.

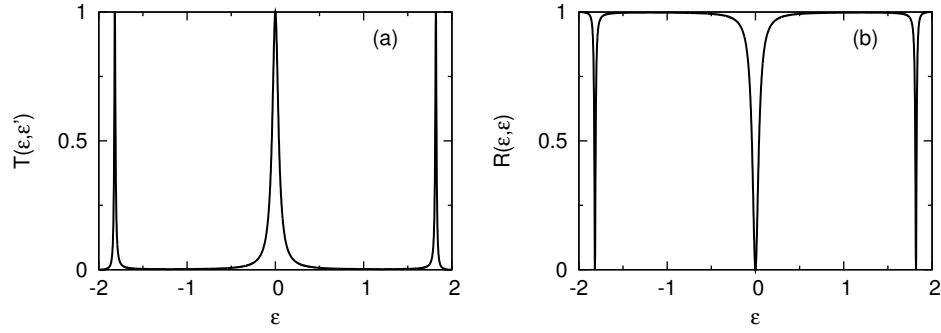


Figure 10.19: (a) The transmission coefficient $T(\varepsilon, \varepsilon')$ and (b) the reflection coefficient $R(\varepsilon, \varepsilon)$ versus the incoming electron energy ε for two 1D (linear) atomic chains connected to a central four-site ring-like “impurity” structure. The common system parameters are $t = 1.0$, $t_1 = 0.2$, $t_2 = 0.9$, $v_c \equiv v_0 = v_1 = v_2 = v_3 = 0.0$, and $v = v' = 0$.

The reader, perhaps having seen only *three* transmission peaks in Fig. 10.19 (a) [or three reflection minima in Fig. 10.19 (b)] might have been led to believe that there should be something wrong with these plots: except this present ring system, s/he has always witnessed that the number of peaks (or minima) should be equal to that of the central impurity sites; the s/he might expect to see *four* peaks also for the system at hand. We can resolve this issue by calculating the eigenenergy values E_c of the central ring system which we should be able to spot in Fig. 10.19 (a). It follows from (9.29) and (9.30) that

$$\begin{aligned} E_c &= v_c - 2t_2 \cos k_c \\ &= v_c - 2t_2 \cos \frac{2\pi n}{N} \quad (n = 0, 1, 2, \dots, N-1) \end{aligned} \quad (10.83)$$

which, for the system at hand with $v_c = 0.0$ and $N = 4$, takes the form

$$\begin{aligned} E_c &= -2t_2 \cos \frac{2\pi n}{4} \\ &= -2t_2 \cos \frac{n\pi}{2} \quad (n = 0, 1, 2, 3) \end{aligned}$$

and, with $t_2 = 0.9$,

$$E_c = \begin{cases} -1.8 \cos 0 = -1.8 \\ -1.8 \cos \frac{\pi}{2} = 0 \\ -1.8 \cos \pi = +1.8 \\ -1.8 \cos \frac{3\pi}{2} = 0 \end{cases}$$

and these are really what we observe in Fig. 10.19 (a). Thus the above difficulty has been resolved: there are certainly *four* eigenenergies for the central ring-like structure, but not all of them are *discrete*, for the $E_c = 0$ value is *doubly degenerate*; therefore, we mark only three peaks, *not* four in Fig. 10.19 (a).

Is there any way to “decouple” this doubly degenerate eigenenergy value? Or, can we remove this degeneracy? The answer is yes. In the last chapter we shall exemplify a beautiful means, with the help of a magnetic flux allowed to thread the central ring-like structure, to expel such a degeneracy.

It seems in order before passing to draw the attention of the reader to the fact that we have for the first time proved numerically with this present problem the validity of the formula (10.83), which gives the eigenenergies of a ring-like structure.

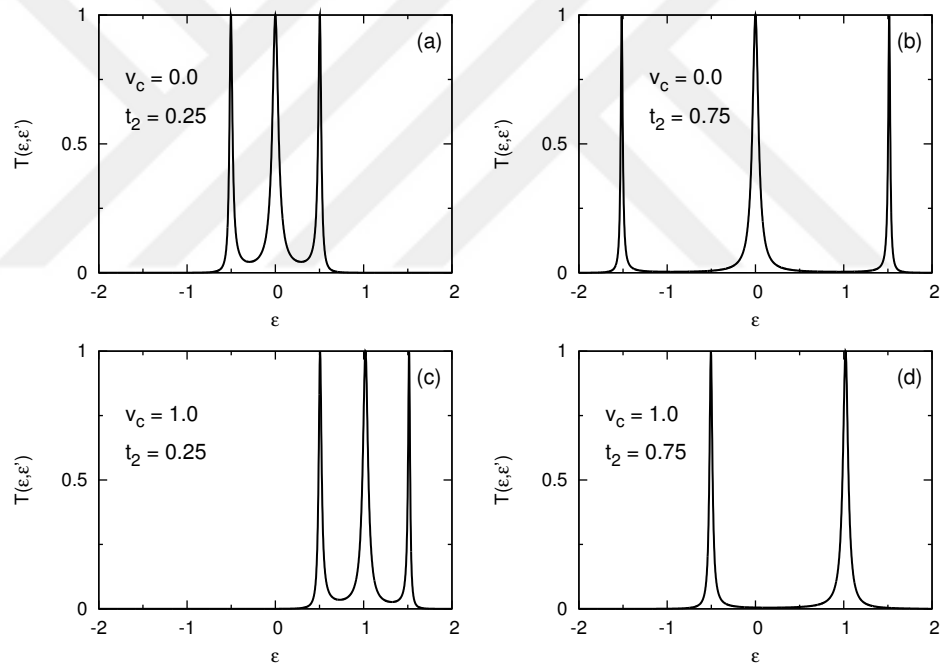


Figure 10.20: The transmission coefficient $T(\epsilon, \epsilon')$ versus the incoming electron energy ϵ for two 1D (linear) atomic chains connected to a central four-site ring-like “impurity” structure for different values the hopping constant t_2 mediating the transmission between any of the next-neighboring three central sites (see Fig. 10.18) and the gate potentials $v_c \equiv v_0 = v_1 = v_2 = v_3$. The other common system parameters are $t = 1.0$, $t_1 = 0.2$ and $v = v' = 0$.

In Fig. 10.20 above are the plots for demonstrating clearly, again, the “sifting” effects of the central gate potential v_c and the hopping amplitude t_2 on the energies of the incoming electron. With $n = 0, \dots, 3$, the eigenenergy values seen in these plots are evaluated from (10.83) as

$$\begin{aligned}
 E_c &= 0 - (2)(0.25) \cos \frac{n\pi}{2} = -0.5, 0, 0.5, 0 && \text{[for (a)]} \\
 &= 0 - (2)(0.75) \cos \frac{n\pi}{2} = -1.5, 0, 1.5, 0 && \text{[for (b)]} \\
 &= 1 - (2)(0.25) \cos \frac{n\pi}{2} = 0.5, 1.0, 1.5, 1.0 && \text{[for (c)]} \\
 &= 1 - (2)(0.75) \cos \frac{n\pi}{2} = -0.5, 1.0, 2.5, 1.0 && \text{[for (d)]}
 \end{aligned}$$

We note that the 2.5-unit eigenenergy is not seen in Fig. 10.20 (d) because the permitted energy band of the incoming electron, between -2.0 and $+2.0$, excludes this value.

And finally, at least for this chapter, Fig. 10.21 gives the usual “sifting” effect of the hopping constant t_1 on the number of electron(s) being able to hop from the left lead to the right lead via the central ring-like impurity region. There is no need to give any further explanation, for every feature in these plots is already obvious.

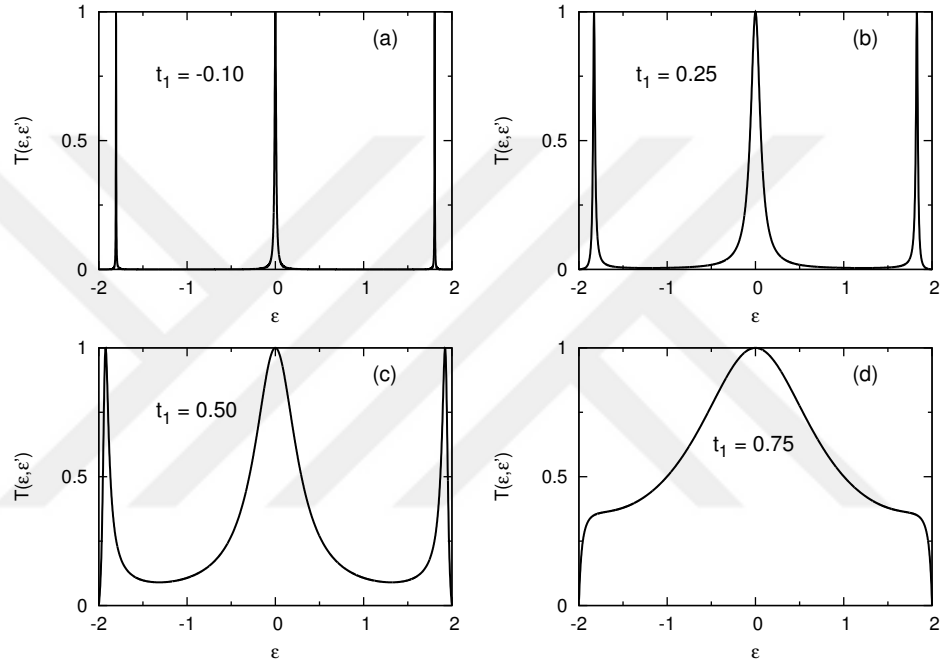


Figure 10.21: The transmission coefficient $T(\varepsilon, \varepsilon')$ versus the incoming electron energy ε for two 1D (linear) atomic chains connected to a central four-site ring-like “impurity” structure for different values the hopping constant t_1 mediating the transmission between the central impurity region and the left and right leads (see Fig. 10.18). The other common system parameters are $t = 1.0$, $t_2 = 0.9$, $\nu_c \equiv \nu_0 = \nu_1 = \nu_2 = \nu_3 = 0$, and $\nu = \nu' = 0$.

CHAPTER 11

AHARONOV–BOHM RINGS: A DETAILED EXPOSITION¹

In this chapter the mathematics and physics of the Aharonov–Bohm (AB) rings is given at some length. Firstly Lagrangian and Hamiltonian dynamics are briefly given and then the Hamiltonian of a charged particle moving in an electromagnetic field is obtained. Secondly the gauge transformations are discussed and their implications are exemplified with the aid of the double–slit experiment and the AB effect. Next the isolated AB ring problem without and with a constant magnetic flux is studied. This discussion is then extended to the AB ring which is now part of an external circuit. At last the problem of the AB ring exposed to a time–dependent magnetic flux is discussed. In this final part the main emphasis is placed on the case where the magnetic flux is time–periodic which allows the application of the so–called Floquet theory to the solution of the problem.

Although the treatment in this chapter is not so rigorous, most of the formulas and statements are proved or verified. The proofs or verifications are usually given at footnotes for the sake of clarity and continuity of the discussion.

11.1 Introduction

In order to build a complete picture of the Aharonov–Bohm rings, the first and the most critical task is to construct a reliable Hamiltonian for a charged particle moving in an electromagnetic field. In the following treatment, the discussion will be confined to a *nonrelativistic view*, which is adequate for our purposes.

¹ This chapter was originally written in a full-article format, so its style might be a little different from that of the rest of the manuscript.

We remember from one of our previous discussions that, in classical electrodynamics, the force on a particle of charge q moving with velocity $\mathbf{v}$ through electric and magnetic fields $\mathbf{E}$ and $\mathbf{B}$ is given by the *Lorentz force law*:

$$\mathbf{F} = q \left(\mathbf{E} + \frac{\mathbf{v}}{c} \times \mathbf{B} \right) \quad (11.1)$$

[Here we have written this law in the conventional CGS unit system; see also (2.38) which was written in the SI unit system.] Unfortunately this force cannot be expressed as the gradient of a scalar potential U , and therefore the Schrödinger equation in the usual form

$$i\hbar \frac{\partial \Psi}{\partial t} = -\frac{\hbar^2}{2m} \nabla^2 \Psi + U\Psi \quad (11.2)$$

cannot accommodate the Lorentz force law. What we should do is to use the more sophisticated form

$$i\hbar \frac{\partial \Psi}{\partial t} = H\Psi \quad (11.3)$$

and resort to a method to find out the correct Hamiltonian H for the system. The path we shall follow is first to determine the Lagrange equation of motion using the Lorentz force law (11.1), second to identify the Lagrangian $\mathcal{L}$ of the system from this equation of motion, third to obtain the *classical* Hamiltonian $\mathcal{H}$ from the Lagrangian, and finally to arrive at the *quantum mechanical* Hamiltonian H through the canonical substitution $\mathbf{p} \rightarrow -i\hbar\nabla$. Therefore, it would be appropriate to review briefly the Lagrangian and Hamiltonian dynamics.

11.2 A Review of Classical Mechanics

11.2.1 The Lagrangian Dynamics

Everything is based on the celebrated *Hamilton's least action principle* [68]: Let a dynamical system start from one specific point and terminate at another specific point in the so-called *configuration space* within a definite time interval. There are infinitely many paths the system may follow from the initial point to the final point. The real path followed by the system is the one for which the time integral of $(T - U)$, the difference between the kinetic and potential energies of the system, is minimum. In

the language of the calculus of variations, this amounts to

$$\delta \int_{t_1}^{t_2} (T - U) dt = 0 \quad (11.4)$$

This energy difference, called the *Lagrangian* of the system, is designated as

$$\mathcal{L} \equiv T - U \quad (11.5)$$

which is a function of *generalized coordinates* q_n and *generalized velocities* $\dot{q}_n$:

$$\mathcal{L} = \mathcal{L}(q_n, \dot{q}_n; t)$$

With this definition, we write (11.4) as

$$\delta \int_{t_1}^{t_2} \mathcal{L}(q_n, \dot{q}_n; t) dt = 0 \quad (11.6)$$

which leads to the *Lagrange equations of motion* for the system:²

$$\frac{\partial \mathcal{L}}{\partial q_n} - \frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{q}_n} = 0, \quad n = 1, 2, 3, \dots \quad (11.7)$$

11.2.2 The Hamiltonian Dynamics

The passage from Lagrangian dynamics to Hamiltonian dynamics is achieved via the following transformation:

$$\mathcal{H}(q_n, p_n; t) = \sum_m p_m \dot{q}_m - \mathcal{L}(q_n, \dot{q}_n; t) \quad (11.8)$$

where p_n are called the *generalized momenta* and given by

$$p_n \equiv \frac{\partial \mathcal{L}}{\partial \dot{q}_n} \quad (11.9)$$

² **Proof for (11.7):** We assume that there is no variation at the end points, $\delta q(t_1) = \delta q(t_2) = 0$. Then

$$\begin{aligned} 0 &= \delta \int_{t_1}^{t_2} \mathcal{L}(q, \dot{q}; t) dt = \int_{t_1}^{t_2} \delta \mathcal{L}(q, \dot{q}; t) dt = \int_{t_1}^{t_2} \left(\frac{\partial \mathcal{L}}{\partial q} \delta q + \frac{\partial \mathcal{L}}{\partial \dot{q}} \delta \dot{q} \right) dt \\ &= \int_{t_1}^{t_2} \frac{\partial \mathcal{L}}{\partial q} \delta q dt + \int_{t_1}^{t_2} \frac{\partial \mathcal{L}}{\partial \dot{q}} \delta \dot{q} dt = \int_{t_1}^{t_2} \frac{\partial \mathcal{L}}{\partial q} \delta q dt + \underbrace{\frac{\partial \mathcal{L}}{\partial \dot{q}} \delta q}_{=0} \Big|_{t_1}^{t_2} - \int_{t_1}^{t_2} \left(\frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{q}} \right) \delta q dt \\ \therefore 0 &= \int_{t_1}^{t_2} \left(\frac{\partial \mathcal{L}}{\partial q} - \frac{d}{dt} \frac{\partial \mathcal{L}}{\partial \dot{q}} \right) \delta q dt \end{aligned}$$

Here the integrand must vanish in order for this result to be valid for every variation δq ; with this, we are led to (11.7). $\square$

We immediately note that with this definition the Lagrange equations of motion (11.7) are expressed by

$$\dot{p}_n = \frac{\partial \mathcal{L}}{\partial q_n} \quad (11.10)$$

The quantity $\mathcal{H}$ is called the *Hamiltonian* of the system; it is a function of generalized coordinates q_n and generalized momenta p_n . The *Hamilton equations of motion* are given by³

$$\dot{q}_n = \frac{\partial \mathcal{H}}{\partial p_n} \quad \text{and} \quad \dot{p}_n = -\frac{\partial \mathcal{H}}{\partial q_n} \quad (11.16)$$

11.2.3 Hamiltonian of a Charged Particle in an Electromagnetic Field

We recall from the electromagnetic theory that the electric field $\mathbf{E}$ and magnetic field $\mathbf{B}$ are expressed by the scalar potential ϕ and the vector potential $\mathbf{A}$ through the relations

$$\mathbf{E} = -\nabla\phi - \frac{1}{c} \frac{\partial \mathbf{A}}{\partial t} \quad \text{and} \quad \mathbf{B} = \nabla \times \mathbf{A} \quad (11.17)$$

With these the Lorentz force law (11.1) becomes

$$\mathbf{F} = q \left\{ -\nabla\phi - \frac{1}{c} \left[\frac{\partial \mathbf{A}}{\partial t} - \mathbf{v} \times (\nabla \times \mathbf{A}) \right] \right\} \quad (11.18)$$

³ **Proof for (11.16):** Since $\mathcal{H} = \mathcal{H}(q_n, p_n; t)$, the total differential of $\mathcal{H}$ is

$$d\mathcal{H} = \sum_n \left(\frac{\partial \mathcal{H}}{\partial q_n} dq_n + \frac{\partial \mathcal{H}}{\partial p_n} dp_n \right) + \frac{\partial \mathcal{H}}{\partial t} dt \quad (11.11)$$

We can also write, using (11.8),

$$d\mathcal{H} = \sum_n \left(\dot{q}_n dp_n + p_n d\dot{q}_n - \frac{\partial \mathcal{L}}{\partial q_n} dq_n - \frac{\partial \mathcal{L}}{\partial \dot{q}_n} d\dot{q}_n \right) - \frac{\partial \mathcal{L}}{\partial t} dt \quad (11.12)$$

Using (11.9) and (11.10) for the expressions $\partial \mathcal{L}/\partial q_n$ and $\partial \mathcal{L}/\partial \dot{q}_n$, the second and forth terms in the parentheses in (11.12) cancel, and we are left with

$$d\mathcal{H} = \sum_n (\dot{q}_n dp_n - \dot{p}_n dq_n) - \frac{\partial \mathcal{L}}{\partial t} dt \quad (11.13)$$

If we identify the coefficients of dq_n and dp_n between (11.11) and (11.13), we get the Hamilton equations of motion, (11.16). $\square$

Another byproduct comes from the identification of the coefficients of dt between the same equations:

$$-\frac{\partial \mathcal{L}}{\partial t} = \frac{\partial \mathcal{H}}{\partial t} \quad (11.14)$$

Furthermore, using (11.16) in (11.11), each term in the parentheses vanishes, and it follows that

$$\frac{d\mathcal{H}}{dt} = \frac{\partial \mathcal{H}}{\partial t} \quad (11.15)$$

This is an important result; it expresses the fact that if $\mathcal{H}$ does not explicitly contain the time, then the Hamiltonian is a conserved quantity.

It is an easy but tedious and tricky task to obtain the Lagrangian of the system from this force. The result is⁴

$$\mathcal{L}(\mathbf{r}, \mathbf{v}; t) = \frac{1}{2} m v^2 - q\phi + \frac{q}{c} \mathbf{v} \cdot \mathbf{A} \quad (11.22)$$

For obtaining the Hamiltonian of the system, we need the canonical momentum $\mathbf{p}$ from (11.9):

$$\frac{\partial \mathcal{L}}{\partial \mathbf{v}} = \mathbf{p} = m\mathbf{v} + \frac{q}{c} \mathbf{A} \quad (11.23)$$

⁴ **Proof for (11.22):** In most general circumstances we have $\mathbf{A} = \mathbf{A}(\mathbf{r}, t)$ so that the total time derivative of $\mathbf{A}$ can be written as

$$\frac{d\mathbf{A}}{dt} \equiv \frac{dA_n}{dt} = \frac{\partial A_n}{\partial r_m} \frac{dr_m}{dt} + \frac{\partial A_n}{\partial t} \equiv (\mathbf{v} \cdot \nabla) \mathbf{A} + \frac{\partial \mathbf{A}}{\partial t}$$

whence

$$\frac{\partial \mathbf{A}}{\partial t} = \frac{d\mathbf{A}}{dt} - (\mathbf{v} \cdot \nabla) \mathbf{A} \quad (11.19)$$

For the term with crosses in (11.18) we make use of a vector identity as follows:

$$\nabla(\mathbf{v} \cdot \mathbf{A}) = (\mathbf{v} \cdot \nabla) \mathbf{A} + \underbrace{(\mathbf{A} \cdot \nabla) \mathbf{v}}_{=0} + \mathbf{v} \times (\nabla \times \mathbf{A}) + \underbrace{\mathbf{A} \times (\nabla \times \mathbf{v})}_{=0}$$

whence

$$\mathbf{v} \times (\nabla \times \mathbf{A}) = \nabla(\mathbf{v} \cdot \mathbf{A}) - (\mathbf{v} \cdot \nabla) \mathbf{A} \quad (11.20)$$

Inserting (11.19) and (11.20) into (11.18), we obtain

$$\mathbf{F} = q \left[-\nabla \left(\phi - \frac{1}{c} \mathbf{v} \cdot \mathbf{A} \right) - \frac{1}{c} \frac{d\mathbf{A}}{dt} \right] \quad (11.21)$$

Note that we may also write

$$-\frac{\mathbf{A}}{c} = \frac{\partial}{\partial \mathbf{v}} \left(\phi - \frac{1}{c} \mathbf{v} \cdot \mathbf{A} \right)$$

so that (11.21) can be written as

$$\mathbf{F} = q \left[-\nabla \left(\phi - \frac{1}{c} \mathbf{v} \cdot \mathbf{A} \right) + \frac{d}{dt} \frac{\partial}{\partial \mathbf{v}} \left(\phi - \frac{1}{c} \mathbf{v} \cdot \mathbf{A} \right) \right]$$

On the other hand

$$\mathbf{F} = m \frac{d\mathbf{v}}{dt} = m \frac{d}{dt} \frac{\partial}{\partial \mathbf{v}} \left(\frac{1}{2} \mathbf{v} \cdot \mathbf{v} \right) = \frac{d}{dt} \frac{\partial}{\partial \mathbf{v}} \left(\frac{1}{2} m v^2 \right)$$

Combining these last two equations we get

$$\nabla \left(-q\phi + \frac{q}{c} \mathbf{v} \cdot \mathbf{A} \right) - \frac{d}{dt} \frac{\partial}{\partial \mathbf{v}} \left(\frac{1}{2} m v^2 - q\phi + \frac{q}{c} \mathbf{v} \cdot \mathbf{A} \right) = 0$$

If we add a $mv^2/2$ term into the first parenthesis, its effect will be nothing but we obtain a more symmetrical result:

$$\nabla \left(\frac{1}{2} m v^2 - q\phi + \frac{q}{c} \mathbf{v} \cdot \mathbf{A} \right) - \frac{d}{dt} \frac{\partial}{\partial \mathbf{v}} \left(\frac{1}{2} m v^2 - q\phi + \frac{q}{c} \mathbf{v} \cdot \mathbf{A} \right) = 0$$

Now we can compare this result with the Lagrange equations of motion, (11.7), and identify the Lagrangian for a single charged particle as given in (11.22). $\square$

Finally using (11.8) we cast the desired classical Hamiltonian for a single particle of mass m and charge q moving in an electromagnetic field with velocity $\mathbf{v}$ as

$$\begin{aligned}\mathcal{H}(\mathbf{r}, \mathbf{p}; t) &= \mathbf{p} \cdot \mathbf{v} - \mathcal{L}(\mathbf{r}, \mathbf{v}; t) \\ &= \frac{1}{2} m v^2 + q\phi \\ &= \frac{1}{2m} \left(\mathbf{p} - \frac{q}{c} \mathbf{A} \right)^2 + q\phi\end{aligned}\quad (11.24)$$

where we have used (11.23) for $\mathbf{v}$ in passing to the last line.

It is now time to get the quantum mechanical Hamiltonian H which will be obtained by making the canonical substitution $\mathbf{p} \rightarrow -i\hbar\nabla$ in the classical Hamiltonian, as we advertised before:

$$H(\mathbf{r}, \mathbf{p}; t) = \frac{1}{2m} \left(-i\hbar\nabla - \frac{q}{c} \mathbf{A} \right)^2 + q\phi \quad (11.25)$$

11.3 Gauge Transformations

The electromagnetic fields $\mathbf{E}$ and $\mathbf{B}$, which are expressed by the scalar potential ϕ and the vector potential $\mathbf{A}$ as given in (11.17), are *not* uniquely described by ϕ and $\mathbf{A}$ since both $\mathbf{E}$ and $\mathbf{B}$ are left unchanged by gauge transformations. An example of such a transformation is by the following set:⁵

$$\mathbf{A} \rightarrow \mathbf{A}' = \mathbf{A} + \nabla\Lambda \quad \text{and} \quad \phi \rightarrow \phi' = \phi - \frac{1}{c} \frac{\partial\Lambda}{\partial t} \quad (11.26)$$

where $\Lambda = \Lambda(\mathbf{r}, t)$ is an arbitrary scalar field.

Now consider the Schrödinger equation (11.3) with the Hamiltonian (11.25):

$$i\hbar \frac{\partial\Psi}{\partial t} = \frac{1}{2m} \left(-i\hbar\nabla - \frac{q}{c} \mathbf{A} \right)^2 \Psi + q\phi \Psi \quad (11.27)$$

⁵ We can easily verify that the set (11.26) leaves both $\mathbf{E}$ and $\mathbf{B}$ unchanged as follows:

$$\mathbf{B} \rightarrow \mathbf{B}' = \nabla \times \mathbf{A}' = \nabla \times (\mathbf{A} + \nabla\Lambda) = \nabla \times \mathbf{A} + \underbrace{\nabla \times \nabla\Lambda}_{=0} = \nabla \times \mathbf{A} = \mathbf{B} \quad \square$$

$$\begin{aligned}\mathbf{E} \rightarrow \mathbf{E}' &= -\nabla\phi' - \frac{1}{c} \frac{\partial\mathbf{A}'}{\partial t} = -\nabla\left(\phi - \frac{1}{c} \frac{\partial\Lambda}{\partial t}\right) - \frac{1}{c} \frac{\partial}{\partial t}(\mathbf{A} + \nabla\Lambda) \\ &= -\nabla\phi + \frac{1}{c} \frac{\partial}{\partial t}(\nabla\Lambda) - \frac{1}{c} \frac{\partial\mathbf{A}}{\partial t} - \frac{1}{c} \frac{\partial}{\partial t}(\nabla\Lambda) = -\nabla\phi - \frac{1}{c} \frac{\partial\mathbf{A}}{\partial t} = \mathbf{E} \quad \square\end{aligned}$$

If this equation is to remain invariant under the gauge transformation (11.26), the phase of the wave function Ψ must also change. A straightforward calculation shows that the change in the wave function is like [69]

$$\Psi \rightarrow \Psi' = e^{iq\Lambda/\hbar c} \Psi \quad (11.28)$$

To recapitulate things, the set of transformations

$$\begin{aligned} \mathbf{A} &\rightarrow \mathbf{A}' = \mathbf{A} + \nabla \Lambda \\ \phi &\rightarrow \phi' = \phi - \frac{1}{c} \frac{\partial \Lambda}{\partial t} \\ \Psi &\rightarrow \Psi' = e^{iq\Lambda/\hbar c} \Psi \end{aligned} \quad (11.29)$$

leaves the Schrödinger equation (11.27) unchanged or invariant.⁶ This result has an important physical interpretation: whatever set of potentials ϕ and $\mathbf{A}$ is chosen, the resultant Schrödinger equation will not depend on the choice as long as these potentials are connected by gauge transformations or, equivalently, describe the same electromagnetic fields $\mathbf{E}$ and $\mathbf{B}$.

⁶ **Verification of (11.29):**

$$(\text{Sch. Eq.}) \rightarrow (\text{Sch. Eq.})' : \quad \underbrace{i\hbar \frac{\partial \Psi'}{\partial t}}_{\text{I}} = \underbrace{\frac{1}{2m} \left(-i\hbar \nabla - \frac{q}{c} \mathbf{A}' \right)^2 \Psi'}_{\text{II}} + \underbrace{q\phi' \Psi'}_{\text{III}}$$

$$\text{I} = i\hbar \frac{\partial}{\partial t} \left(e^{iq\Lambda/\hbar c} \Psi \right) = e^{iq\Lambda/\hbar c} \left(-\frac{q}{c} \frac{\partial \Lambda}{\partial t} \Psi + i\hbar \frac{\partial \Psi}{\partial t} \right)$$

$$\begin{aligned} \text{II} &= \frac{1}{2m} \left[-i\hbar \nabla - \frac{q}{c} (\mathbf{A} + \nabla \Lambda) \right]^2 e^{iq\Lambda/\hbar c} \Psi \\ &= \frac{1}{2m} \left[-i\hbar \nabla - \frac{q}{c} (\mathbf{A} + \nabla \Lambda) \right] e^{iq\Lambda/\hbar c} \left(-i\hbar \nabla - \frac{q}{c} \mathbf{A} \right) \Psi \\ &= \frac{1}{2m} e^{iq\Lambda/\hbar c} \left(-i\hbar \nabla - \frac{q}{c} \mathbf{A} \right) \left(-i\hbar \nabla - \frac{q}{c} \mathbf{A} \right) \Psi \\ &= e^{iq\Lambda/\hbar c} \left[\frac{1}{2m} \left(-i\hbar \nabla - \frac{q}{c} \mathbf{A} \right)^2 \right] \Psi \end{aligned}$$

$$\text{III} = q \left(\phi - \frac{1}{c} \frac{\partial \Lambda}{\partial t} \right) e^{iq\Lambda/\hbar c} \Psi = e^{iq\Lambda/\hbar c} \left(q\phi \Psi - \frac{q}{c} \frac{\partial \Lambda}{\partial t} \Psi \right)$$

If we put the pieces I, II, and III altogether and cancel the exponential factors we get

$$(\text{Sch. Eq.})' : \quad \text{I} = \text{II} + \text{III} : \quad i\hbar \frac{\partial \Psi}{\partial t} = \frac{1}{2m} \left(-i\hbar \nabla - \frac{q}{c} \mathbf{A} \right)^2 \Psi + q\phi \Psi$$

so that $(\text{Sch. Eq.})' = (\text{Sch. Eq.})$. $\square$

11.3.1 Double–Slit Diffraction

Gauge invariance implies that the phase changes introduced into the wave function by a gauge transformation do not cause any change in any diffraction pattern. Here we will test this claim by comparing the phase shifts introduced along different paths by a gauge transformation.

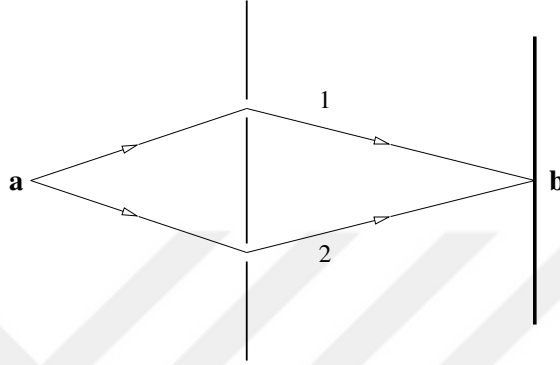


Figure 11.1: Double–slit diffraction experiment.

Consider a simple double–slit diffraction experiment, as in Fig. 11.1, and assume that there is no magnetic field present *anywhere* so that we take $\mathbf{A} = 0$. There will be then a certain diffraction pattern on the screen. Suppose next that we transform to a nonzero, but still curl–free, vector potential

$$\mathbf{A} \rightarrow \mathbf{A}' = \nabla \Lambda(\mathbf{r})$$

According to (11.28), this transformation introduces a local *phase change*

$$\varphi(\mathbf{r}) = \frac{q}{\hbar c} \Lambda(\mathbf{r})$$

into the transformed wave function Ψ' , relative to original wave function Ψ . Because this phase shift is not constant, but depends on $\mathbf{r}$, there will in general be a change in the phase *difference* between any points $\mathbf{a}$ and $\mathbf{b}$, such as the point of origin of an electron stream and a point of observation on the screen,

$$\Delta\varphi_{\mathbf{ab}} \equiv \varphi(\mathbf{b}) - \varphi(\mathbf{a}) = \frac{q}{\hbar c} [\Lambda(\mathbf{b}) - \Lambda(\mathbf{a})]$$

We write this change as a path integral

$$\Delta\varphi_{\mathbf{ab}} = \frac{q}{\hbar c} \int_{\mathbf{a}}^{\mathbf{b}} (\nabla \Lambda) \cdot d\mathbf{l} = \frac{q}{\hbar c} \int_{\mathbf{a}}^{\mathbf{b}} \mathbf{A}' \cdot d\mathbf{l}$$

A nonzero value of $\Delta\varphi_{ab}$ means that the total number of wavelengths contained in the path **ab** has been changed by the gauge transformation. However, what matters is not the phase change along path 1 or path 2 alone, but the *difference*—if any—between the two paths:

$$\Delta\varphi = \Delta\varphi_1 - \Delta\varphi_2 = \frac{q}{\hbar c} \left(\int_1 \mathbf{A}' \cdot d\mathbf{l} - \int_2 \mathbf{A}' \cdot d\mathbf{l} \right) = \frac{q}{\hbar c} \oint \mathbf{A}' \cdot d\mathbf{l}$$

The (counterclockwise) loop integral is evaluated by using Stokes' theorem:

$$\Delta\varphi = \frac{q}{\hbar c} \oint \mathbf{A}' \cdot d\mathbf{l} = \frac{q}{\hbar c} \int_S (\nabla \times \mathbf{A}') \cdot d\mathbf{S} = \frac{q}{\hbar c} \int_S \mathbf{B}' \cdot d\mathbf{S} = \frac{q}{\hbar c} \Phi \quad (11.30)$$

where the surface integrals are over the area enclosed by the two paths and Φ is the magnetic flux—if any—enclosed by the paths. Because of our assumption of no magnetic field anywhere, the enclosed magnetic flux is zero, and we obtain a zero phase difference between the two interference paths, thus proving our claim.

11.3.2 Aharonov–Bohm Effect

The above discussion suggests that there *should* be observable interference effects if the two path enclose a region containing a nonzero magnetic flux, even if the magnetic field along the electron paths themselves is zero. This is the celebrated *Aharonov–Bohm effect* [70], which has been confirmed experimentally [71].

Consider a specific example of a longitudinally magnetized iron wire, as shown in Fig. 11.2. Such a wire may contain an *internal* magnetic flux with no *external* magnetic field. According to (11.30) an external vector potential *must* then be present in the field-free space outside the wire. Suppose next that such a wire is placed into the wave shadow between the two slits in a double-slit diffraction screen (Fig. 11.3). The magnetic field inside the wire points into the plane of the drawing, there will then be a clockwise circular vector potential present.

The vector potential outside the wire is curl-free and may therefore still be written as the gradient of a scalar function. We may still treat external vector potential *as if* it had been generated from the zero vector potential case by a gauge transformation. However, compared to the true zero-field case without any enclosed flux, the gauge function $\Lambda(\mathbf{r})$ is no longer single-valued. Suppose we start at some point outside the

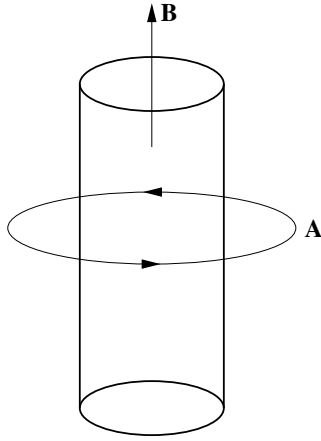


Figure 11.2: A longitudinally magnetized iron.

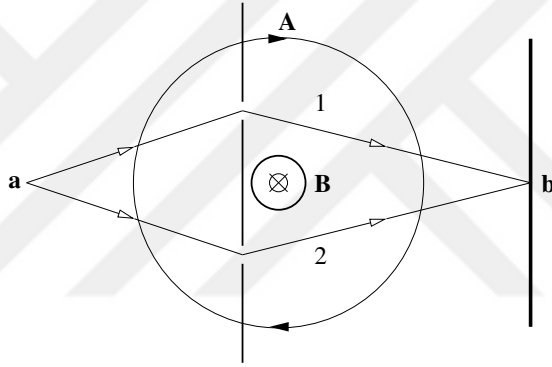


Figure 11.3: Double-slit diffraction around a magnetic flux tube.

wire, and follow $\Lambda(\mathbf{r})$ as we round the wire, and return to the starting point. The gauge function $\Lambda(\mathbf{r})$ then does not return to its original value, but will be incremented by

$$\Delta\Lambda = \oint \mathbf{A}' \cdot d\mathbf{l} = \Phi \quad (11.31)$$

Associated with this change in Λ is a phase difference between the two interfering wave:

$$\Delta\varphi = \frac{q}{\hbar c} \Phi \quad (11.32)$$

Our earlier proof of gauge invariance is not applicable to such multivalued gauge functions.

In classical mechanics, what is important is only the magnetic field along the particle trajectory, and as long as the particle stays outside the magnetic field itself, a multi-valued curl-free vector potential has no observable consequences. In quantum mechanics, this is no longer true: the propagation of particles is now wave-like rather than trajectory-like. But waves have a phase, and this phase is changed by the presence of a vector potential, even a curl-free one. The phase along the upper path is retarded (for electrons!) by the vector potential, while that along the lower part is advanced. Equation (11.32) is the resulting overall phase difference.

What happens if an electron moves in a field-free region which encloses a “hole” containing flux Φ ? Upon completing a circuit around the hole, the electron wave function acquires an additional phase of magnitude $|e|\Phi/\hbar c$. The requirement that the electron wave function be single-valued, so that the phase factor is unity, implies that *the magnetic flux is quantized*:

$$\Phi = n \frac{2\pi\hbar c}{|e|} = n\Phi_0 \quad n = 0, \pm 1, \pm 2, \dots \quad (11.33)$$

where

$$\Phi_0 = \frac{2\pi\hbar c}{|e|} = \frac{hc}{|e|} = 4.135 \times 10^{-7} \text{ Gauss-cm}^2 \quad (11.34)$$

is the *fundamental unit of magnetic flux*.

11.4 An Isolated Aharonov–Bohm Ring

Although it is of no interest, we study the isolated Aharonov–Bohm (AB) ring problem in order to appreciate the results of the next section where we shall study an AB ring which will be part of an external circuit.

11.4.1 Field–Free Case

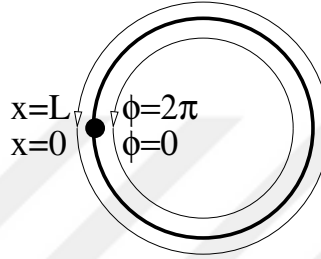


Figure 11.4: An Aharonov–Bohm ring with no magnetic field.

Consider an electron confined to move around a circle of radius R (Fig. 11.4), with the circumference $L = 2\pi R$. In the absence of an electromagnetic field, the (time–independent) Schrödinger equation of the system is

$$-\frac{\hbar^2}{2m}\nabla^2\Psi = E\Psi$$

where m is the mass of the electron.

Working in polar coordinates unnecessarily complicates the problem. The best way to follow is to think of the electron confined to move freely on a line of length L lying, say, on *some* x –axis, with the restriction that at the end points of the line the electron wave function be single–valued, $\Psi(0) = \Psi(L)$. According to Fig. 11.4 we take $dx \hat{x} = d\phi \hat{\phi}$ for the counterclockwise motion of the electron. With this convention the above Schrödinger equation takes the form

$$-\frac{\hbar^2}{2m}\frac{d^2}{dx^2}\Psi(x) = E\Psi(x) \tag{11.35}$$

which has the solution of the form

$$\Psi(x) = e^{ikx}, \quad k = \frac{1}{\hbar}\sqrt{2mE}$$

where we did not bother with normalizing the wave function. Applying the single-valuedness condition we get

$$\begin{aligned}\Psi(0) &= \Psi(L) \\ 1 &= e^{ikL} \\ kL &= 2\pi n \\ \therefore k &= \frac{2\pi}{L} n, \quad n = 0, \pm 1, \pm 2, \dots\end{aligned}$$

so that the complete solution is

$$\Psi_n(x) = e^{i(2\pi n/L)x} \quad \text{with} \quad E_n = \frac{\hbar^2}{2m} \left(\frac{2\pi}{L} \right)^2 n^2 \quad (11.36)$$

Here positive (negative) n means counterclockwise (clockwise) motion of the electron. It should be noted that there are *two* independent solutions for each energy E_n , corresponding to counterclockwise and clockwise motion. That is, each energy E_n has twofold *degeneracy*.

11.4.2 Constant Magnetic Field Case

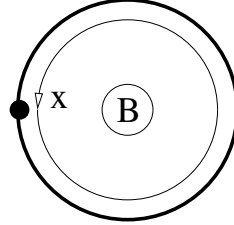


Figure 11.5: An Aharonov–Bohm ring exposed to a magnetic field $\mathbf{B} = B_0 \hat{z}$.

We now apply a *constant* magnetic field perpendicular to and inside the ring, $\mathbf{B} = B_0 \hat{z}$, as shown in Fig. 11.5. We may achieve this by placing a solenoid of radius $a < R$ along the ring axis z , carrying a magnetic field $\mathbf{B}$. If the solenoid is long enough, the field inside the solenoid will be uniform and constant, and that outside the solenoid will be zero. But the vector potential outside the solenoid will not be zero, instead

$$\mathbf{A} = \frac{\Phi}{2\pi r} \hat{\phi}, \quad r > a$$

where $\Phi = \pi a^2 B_0$ is the magnetic flux through the solenoid. On the ring we have

$$\mathbf{A} = \frac{\Phi}{2\pi R} \hat{\phi} = \frac{\Phi}{L} \hat{\phi}$$

and adopting the same convention used in the zero-field case above, we can finally write

$$\mathbf{A} = A \hat{x} = \frac{\Phi}{L} \hat{x} \quad (11.37)$$

With $q = e$, where $e < 0$, and $\phi = 0$, the Hamiltonian of the system is written from (11.25) as

$$\begin{aligned} H &= \frac{1}{2m} \left(-i\hbar \frac{d}{dx} - \frac{e}{c} A \right)^2 \\ &= \frac{1}{2m} \left(-i\hbar \frac{d}{dx} - \frac{e}{c} \frac{\Phi}{L} \right)^2 \\ &= -\frac{\hbar^2}{2m} \left(\frac{d}{dx} - i \frac{e}{\hbar c} \frac{\Phi}{L} \right)^2 \end{aligned}$$

Here it is convenient to work with the fundamental unit of magnetic flux given in (11.34).

We re-define the flux Φ in the fundamental unit via $f = \Phi/\Phi_0$, so that the Hamiltonian takes the final form as

$$H(x) = -\frac{\hbar^2}{2m} \left(\frac{d}{dx} + i \frac{2\pi f}{L} \right)^2 \quad (11.38)$$

Using this Hamiltonian, the time-independent Schrödinger equation is written as

$$-\frac{\hbar^2}{2m} \left(\frac{d}{dx} + i \frac{2\pi f}{L} \right)^2 \Psi = E \Psi$$

or

$$\frac{d^2 \Psi}{dx^2} + 2i\alpha \frac{d\Psi}{dx} + \beta \Psi = 0$$

where

$$\alpha = \frac{2\pi f}{L} \quad \text{and} \quad \beta = \frac{2mE}{\hbar^2} - \alpha^2$$

Solutions are of the form

$$\Psi = e^{i\kappa x} \quad \text{with} \quad \kappa = -\alpha \pm \frac{1}{\hbar} \sqrt{2mE}$$

Applying the single-valuedness condition

$$\Psi(0) = \Psi(L)$$

we get, as before,

$$\kappa = \frac{2\pi}{L} n', \quad n' = 0, \pm 1, \pm 2, \dots$$

so that the complete solution is

$$\Psi_{n'}(x) = e^{i(2\pi n'/L)x} \quad \text{with} \quad E_{n'} = \frac{\hbar^2}{2m} \left(\frac{2\pi}{L} \right)^2 (n' + f)^2 \quad (11.39)$$

Comparison of this solution with that for the field-free case, (11.36), reveals that the twofold degeneracy in the former has now been lifted in the latter by the magnetic field. According to above result negative n' case, which corresponds to clockwise motion of the electron (which in turn corresponds to the motion in the opposite direction as the current in the solenoid), has a somewhat *lower* energy than positive n' case, which describes the counterclockwise motion of the electron. What is more important is the fact that the allowed energies strictly depend on the magnetic flux, even though the magnetic field on the ring is zero.

11.4.3 Gauge-Invariance Treatment

Making use of the invariance property of the system under gauge transformations, we can deal with the problem in an easier and more elegant manner. This time we will be interested in the transformed wave function of the electron in lieu of (the transformed) Hamiltonian. If we choose naturally the vector potential for the field-free case as being zero, $\mathbf{A} = 0$, we have the free-electron Schrödinger equation given in (11.35) and its solution given in (11.36), we rewrite them:

$$-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} \Psi(x) = E \Psi(x) \quad (11.40)$$

$$\Psi_n(x) = e^{i(2\pi n/L)x}, \quad E_n = \frac{\hbar^2}{2m} \left(\frac{2\pi}{L} \right)^2 n^2 \quad (11.41)$$

In the presence of magnetic field it suffices for us to make a gauge transformation as just $\mathbf{A} \rightarrow \mathbf{A}' = \nabla \Lambda$, or, commensurate with our convention,

$$\mathbf{A}' = \frac{d\Lambda}{dx} \hat{x}$$

Here the choice of gauge function Λ is to be such that the path integral of the resulting $\mathbf{A}'$ from point $x = 0$ to $x = L$ should give the magnetic flux (or, equivalently, the

difference in the values of the gauge function at those points should equal the magnetic flux) just as in (11.31). For computational purposes we should attempt to find a gauge function that simplifies the computations as much as possible. In our case the simplest gauge function is

$$\Lambda(x) = \frac{\Phi}{L} x \quad (11.42)$$

Of course, we could have written down this gauge function directly from (11.37). Under the effect of this gauge function the Schrödinger equation of the system will remain exactly in the same as before (i.e., we still have, of course, the free-electron Schrödinger equation) but the electron wave function will change according to the recipe (11.28):

$$\begin{aligned} \Psi_n(x) &\rightarrow \Psi'_n(x) = e^{ie\Lambda/\hbar c} \Psi_n(x) \\ &= e^{-i|e|\Lambda/\hbar c} \Psi_n(x) \\ &= e^{-i(|e|\Phi/\hbar c L)x} e^{i(2\pi n/L)x} \end{aligned}$$

whence

$$\Psi'_n(x) = \exp \left[i \frac{2\pi}{L} (n - f) x \right] \quad (11.43)$$

It is important to note that the energy of the system is still given by

$$E'_n = \frac{\hbar^2}{2m} \left(\frac{2\pi}{L} \right)^2 n^2 \quad (11.44)$$

because we loaded all the changes on to the wave function. We note that the solution given in (11.39) is harmonious with those found above.

It would be better to emphasize here that all the above derivations have been made according to a magnetic field pointed in the z -direction. (Or, more generally, we have taken the increasing direction the coordinate x in such a way that is consistent with the familiar right-hand rule: if the right thumb points in the direction of magnetic field, the fingers of the right-hand curl around the wire in the increasing direction of x .) If we change the direction of magnetic field (or the increasing direction of the coordinate x), this means simply a change in the sign of the flux f so that it is sufficient to make a change in f in the above formulas as $f \rightarrow -f$.

Now let us have a closer look at the wave functions given in Eqs.(11.41) and (11.43) at points $x = 0$ and $x = L$:

$$\begin{aligned}\Psi_n(0) &= 1, & \Psi_n(L) &= e^{i2\pi n} & (\text{without magnetic flux}) \\ \Psi'_n(0) &= 1, & \Psi'_n(L) &= e^{i2\pi n - i2\pi f} & (\text{with magnetic flux})\end{aligned}$$

It follows from these that an electron moving *counterclockwise* from point $x = 0$ to $x = L$, thus returning to its departing point, under the influence of the magnetic field, acquires a *extra* phase factor $e^{-i2\pi f}$ compared with that in zero-field case. If the motion of the electron were *clockwise* from $x = L$ to $x = 0$, then the extra phase acquisition would be $e^{i2\pi f}$.

11.5 The Aharonov–Bohm Ring as a Part of an External Circuit

11.5.1 Field-Free Case

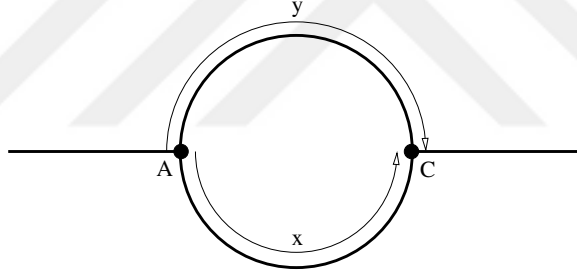


Figure 11.6: An isolated Aharonov–Bohm ring which is now part of an external circuit.

Consider an AB ring connected to an external circuit at the junctions A and C, as shown in Fig. 11.6. The electron is no longer confined to move on the circular part so that the energy of the electron is now continuous. For the sake of clarity we separate the ring into two parts as the lower and the upper ones. Choosing point A as the origin of both lower and upper part (although this bears no importance), the Schrödinger equation and its solution are now

$$\begin{aligned}-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} \Psi(x) &= E\Psi(x), & \Psi(x) &= e^{\pm ikx} & (\text{lower part}) \\ -\frac{\hbar^2}{2m} \frac{d^2}{dy^2} \Psi(y) &= E\Psi(y), & \Psi(y) &= e^{\pm iky} & (\text{upper part})\end{aligned}$$

where for both parts

$$E = \frac{\hbar^2 k^2}{2m}, \quad k = \frac{1}{\hbar} \sqrt{2mE}$$

11.5.2 Constant Magnetic Field Case

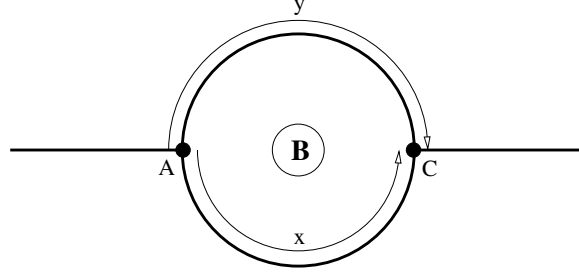


Figure 11.7: An Aharonov–Bohm ring, which is now part of an external circuit, exposed to a constant magnetic field $\mathbf{B} = B_0 \hat{z}$.

We again apply a constant magnetic field perpendicular to and inside the ring, as being $\mathbf{B} = B_0 \hat{z}$, and take point A again as the origin of both lower and upper part, as shown in Fig. 11.7. The Schrödinger equations, with $L = 2\ell$, are

$$-\frac{\hbar^2}{2m} \left(\frac{d}{dx} + i \frac{\pi f}{\ell} \right)^2 \Psi(x) = E \Psi(x) \quad (\text{lower part})$$

$$-\frac{\hbar^2}{2m} \left(\frac{d}{dy} - i \frac{\pi f}{\ell} \right)^2 \Psi(y) = E \Psi(y) \quad (\text{upper part})$$

Exploiting the gauge invariance property, we modify the wave function (11.43) as

$$\Psi(x) = \exp \left[i \left(k - \frac{\pi f}{\ell} \right) x \right] \quad (\text{lower part})$$

$$\Psi(y) = \exp \left[i \left(k + \frac{\pi f}{\ell} \right) y \right] \quad (\text{upper part})$$

Actually these forms of solutions are a little bit cumbersome for later calculations because they make implicit dependence to the field-free case. We can make things better as the following. It follows from the above results that, under the influence of the magnetic field, an electron, for example, going from junction A to junction C using the lower (upper) part acquires an *extra* phase factor $e^{-i\pi f}$ ($e^{i\pi f}$) compared with that in zero-field case. A more general form of this statement is: if the electron moves

counterclockwise (clockwise), then upon completing a way of half circumference ℓ on the ring it acquires an extra phase factor $e^{-i\pi f}$ ($e^{i\pi f}$) compared with that in zero-field case. (Note that we implicitly assume $\mathbf{B} = B_0 \hat{z}$.) With the generalizations, we no longer need to use two coordinates, x and y , instead we can use only x for both of them (or for every part of a given circuit) provided that the origin of each part be explicitly stated. Then with the understanding that the point A is the origin for both lower and upper part, we may write the most general solution for both lower and upper part in the same form as, referring to Fig. 11.8,

$$\begin{aligned}\Psi_1(x) &= a_1 e^{ikx} + b_1 e^{-ikx} & (\text{upper part}) \\ \Psi_2(x) &= a_2 e^{ikx} + b_2 e^{-ikx} & (\text{lower part})\end{aligned}\tag{11.45}$$

and we add manually the effect of the magnetic field as phase factors at the junctions according to the above recipe. To repeat, if an electron moves counterclockwise (clockwise), then upon completing a way of half circumference ℓ on the ring it acquires an extra phase factor $e^{-i\pi f}$ ($e^{i\pi f}$). This is what has been done in the work of Yi *et al.* [72].

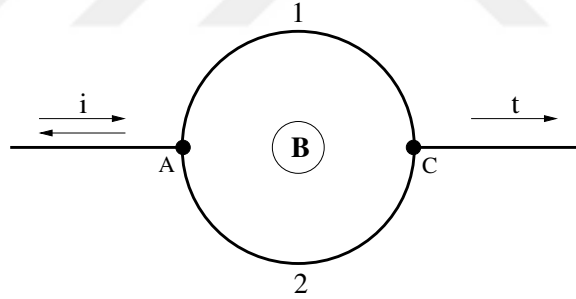


Figure 11.8: The same system as that in Fig. 11.7.

Now we write the so far neglected incident and transmitted wave functions as

$$\Psi_i(x) = e^{-ikx} + r e^{ikx} \quad (\text{incident wave function}) \tag{11.46}$$

$$\Psi_t(x) = t e^{ikx} \quad (\text{transmitted wave function}) \tag{11.47}$$

where we have taken junction A as the origin of the input circuit and junction C as that of the output circuit (beware of the directions!). Now referring to Fig. 11.9, we

observe that the wave functions at junctions assume the values

$$\left. \begin{aligned} \Psi_i(0) &= 1 + r \\ \Psi_1(0) &= a_1 + b_1 e^{-i\pi f} \\ \Psi_2(0) &= a_2 + b_2 e^{i\pi f} \end{aligned} \right\} \text{ at junction A}$$

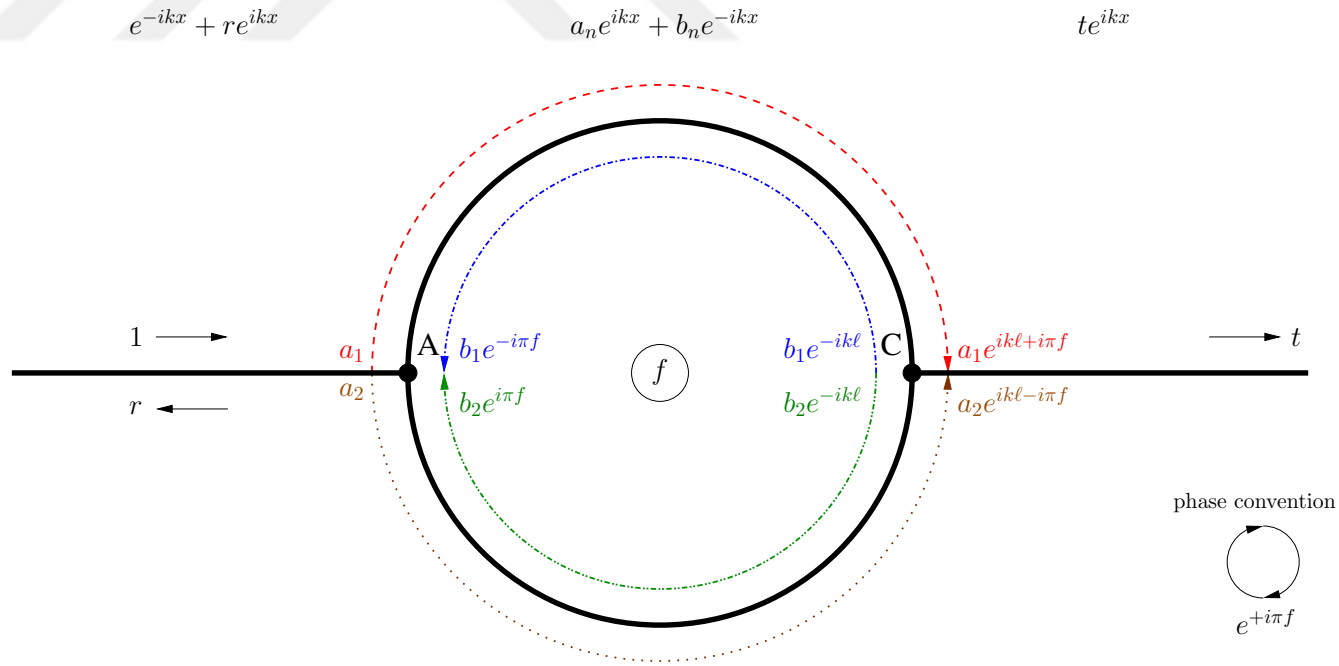
$$\left. \begin{aligned} \Psi_i(\ell) &= t \\ \Psi_1(\ell) &= a_1 e^{ik\ell + i\pi f} + b_1 e^{-ik\ell} \\ \Psi_2(\ell) &= a_2 e^{ik\ell - i\pi f} + b_2 e^{-ik\ell} \end{aligned} \right\} \text{ at junction C}$$

Similarly their first derivatives, in multiples of ik ,

$$\left. \begin{aligned} \Psi'_i(0) &= -1 + r \\ \Psi'_1(0) &= a_1 - b_1 e^{-i\pi f} \\ \Psi'_2(0) &= a_2 - b_2 e^{i\pi f} \end{aligned} \right\} \text{ at junction A}$$

$$\left. \begin{aligned} \Psi'_i(\ell) &= t \\ \Psi'_1(\ell) &= a_1 e^{ik\ell + i\pi f} - b_1 e^{-ik\ell} \\ \Psi'_2(\ell) &= a_2 e^{ik\ell - i\pi f} - b_2 e^{-ik\ell} \end{aligned} \right\} \text{ at junction C}$$

VALUES OF WAVE FUNCTIONS AT THE JUNCTIONS



Point A is the origin of the incident and middle circuits and point B is that for the outgoing surface.

Figure 11.9: The same system as that in Fig. 11.7 with the values of the wave functions at the two junctions.

At the junctions the continuity condition of the wave functions then demand that

$$1 + r = a_1 + b_1 e^{-i\pi f} = a_2 + b_2 e^{i\pi f} \quad (\text{at junction A}) \quad (11.48)$$

$$t = a_1 e^{ik\ell + i\pi f} + b_1 e^{-ik\ell} = a_2 e^{ik\ell - i\pi f} + b_2 e^{-ik\ell} \quad (\text{at junction C}) \quad (11.49)$$

In addition, from the conservation of current density

$$\sum_i \Psi'_i(x) = 0 \quad (11.50)$$

we must have

$$(-1 + r) + (a_1 - b_1 e^{-i\pi f}) + (a_2 - b_2 e^{i\pi f}) = 0 \quad (\text{at junction A}) \quad (11.51)$$

$$t + (a_1 e^{ik\ell + i\pi f} - b_1 e^{-ik\ell}) + (a_2 e^{ik\ell - i\pi f} - b_2 e^{-ik\ell}) = 0 \quad (\text{at junction C}) \quad (11.52)$$

With the help of an appropriate software, such as MATHEMATICA[®], these sets of equations can be readily solved for the six unknowns r, t, a_1, a_2, b_1, b_2 . What is important for us is the transmission coefficient $T = tt^*$. The result is, as given in the above mentioned work of Yi *et al.* [72],

$$T_s(k\ell, f) = \frac{16 \sin^2 k\ell \cos^2 \pi f}{9 \sin^4 k\ell + (16 - 40 \sin^2 \pi f) \sin^2 k\ell + 16 \sin^4 \pi f} \quad (11.53)$$

The method outlined here can be applied to any number of AB rings connected in series in chain-like manner. Yi *et al.* [72] also gave the result for the transmission coefficient T of a double identical AB ring shown in Fig. 11.10 as

$$T_d(k\ell, f) = \frac{4 \cos^4 \pi f \sin^2 k\ell}{(4 - 3 \sin^2 k\ell) \left[3 \sin^4 k\ell + (1 - 8 \sin^2 \pi f) \sin^2 k\ell + 4 \sin^4 \pi f \right]} \quad (11.54)$$

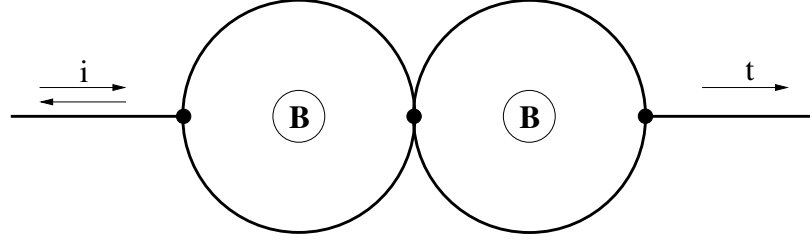


Figure 11.10: A double Aharonov–Bohm ring system exposed to a magnetic field $\mathbf{B} = B_0 \hat{z}$ and connected to two external leads.

The reader will readily notice that both T_s and T_d are periodic in f with period 1 as well as in $k\ell$ with period π . It is therefore sufficient to consider only the regions $0 \leq f < 1$ and $0 \leq k\ell < \pi$. It follows from (11.54) that there exist two kinds of *flux-independent* transmission line. They are

The reader will notice readily that both T_s and T_d are periodic in f with period 1 as well as in $k\ell$ with period π . It is therefore sufficient to consider only the regions $0 \leq f < 1$ and $0 \leq k\ell < \pi$. It follows from (11.54) that there exist two kinds of *flux-independent* transmission line. They are

$$T_d(k\ell = \frac{\pi}{2}, f) = 1 \quad \text{except when} \quad f = \frac{1}{2}$$

$$T_d(k\ell = 0, f) = 0 \quad \text{except when} \quad f = 0$$

In addition we also note that

$$\lim_{k\ell \rightarrow \pi/2} T_d(k\ell, f = \frac{1}{2}) = 0 \quad \text{and} \quad \lim_{k\ell \rightarrow 0} T_d(k\ell, f = 0) = 1$$

Therefore we conclude that the points $(0, 0)$ and $(\frac{\pi}{2}, \frac{1}{2})$ in $(k\ell, f)$ space are singular, and yield, respectively, *resonance* and *antiresonance* near these points.

A closer look at (11.53) reveals that the single ring system has only the resonance structure near point $(0, 0)$ and has no antiresonance.

Yi *et al.* [72] stated that the resonance feature appears both in the odd-numbered and the even-numbered ring chains, whereas the antiresonance feature appears only in even-numbered ring structure. This is clearly seen in the plots of (11.53) and (11.54) in Fig. 11.11; we drew them as they are exceptionally illustrating examples. (Writing

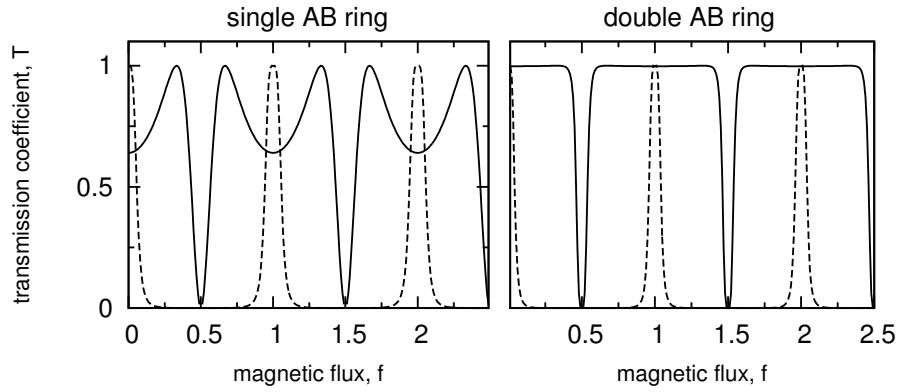


Figure 11.11: Transmission coefficient T versus magnetic flux f for $k\ell = 0.01\pi$ (solid curve) and $k\ell = 0.49\pi$ (dotted curve) for the single and double AB ring structures. It is seen that both resonance and antiresonance phenomena exist in the double AB rings case, whereas only the resonance phenomenon exists in the single AB ring case. These are the plots of (11.53) and (11.54) that we have produced with the help of a MATHEMATICA code). The reader should compare these plots with the original ones reported in the work of Yi *et al.* [72].

some MATHEMATICA[®] codes, we have checked the above statement of Yi *et al.* and have seen that it is correct indeed.)

11.6 An Aharonov–Bohm Ring Exposed to a Time–Dependent Magnetic Field⁷

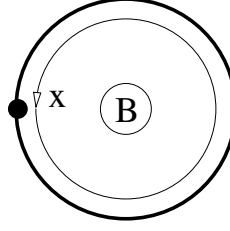


Figure 11.12: An Aharonov–Bohm ring exposed to a time–dependent magnetic field $\mathbf{B}(t) = B_0(t) \hat{z}$.

Consider first an isolated AB ring under the influence of a time–dependent magnetic field $\mathbf{B}(t)$. This means we have now a time dependent flux $f(t)$. Let us use the same system used in the constant magnetic field case above (Fig. 11.12). The Hamiltonian of the system remains almost in the same form as that in (11.38),

$$H(x, t) = -\frac{\hbar^2}{2m} \left[\frac{\partial}{\partial x} + i \frac{2\pi}{L} f(t) \right]^2 \quad (11.55)$$

and this problem requires to solve the time–dependent Schrödinger equation. For an arbitrary time dependence of f , the solution is usually intractable; but for an oscillating flux, the *Floquet theorem* provides us with a good method of handling the time–dependent Schrödinger equation.

⁷This section is not directly related to the present work. As we have indicated before, we want to include the time-dependent formalism into the manuscript just for the sake of the completeness and integrity of the general treatment; what we had in our mind was that in the future, in a Ph.D. study we hope, we desire to evolve the present formalism into more complicated, but potentially fruitful, problems that might contain time-dependent excitations, phonon-electron excitations, etc. Anyway, this and the following sub-section might be skipped in first reading.

11.6.1 Oscillating Magnetic Flux⁸

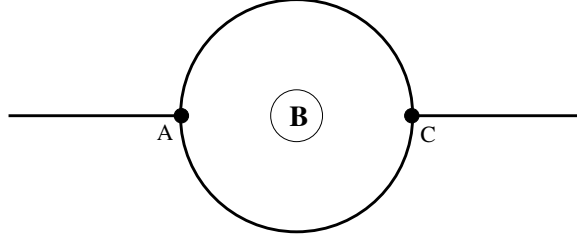


Figure 11.13: An Aharonov–Bohm ring exposed to an oscillating magnetic field $\mathbf{B}(t) = B_0(t) \hat{z}$.

We here consider an AB ring, which is a part of an external circuit, threaded by an oscillating magnetic flux, as shown in the below figure. The ring is made of thin metal wire, in which the transverse modes are well separated so that only a single transverse mode is treated. The oscillating magnetic flux is chosen to be

$$f(t) = f_s + f_d \cos(\omega t + \phi) \quad (11.56)$$

where f_s and f_d are constants which determine the strength of *static* flux and *dynamic* flux, respectively. Here the phase constant ϕ has been inserted for a later possible use. From (11.55) we write the Hamiltonian of the system *for the upper part of the ring* as

$$H(x, t) = -\frac{\hbar^2}{2m_e} \left[\frac{\partial}{\partial x} - i \frac{\pi}{\ell} f(t) \right]^2 \quad (11.57)$$

where we took $L = 2\ell$, with ℓ being half of the circumference of the ring. Since this Hamiltonian is periodic with period $2\pi/\omega$, the Floquet theorem is used in solving the time–dependent Schrödinger equation

$$i\hbar \frac{\partial}{\partial t} \Psi(x, t) = H(x, t) \Psi(x, t) \quad (11.58)$$

The Floquet theorem says that the wave function $\Psi(x, t)$ can be written as

$$\Psi(x, t) = e^{-iE_F t/\hbar} \varphi(x, t) \quad (11.59)$$

where $0 \leq E_F \leq \hbar\omega$ is the Floquet energy and $\varphi(x, t)$ is a periodic function which has the same period, $2\pi/\omega$, as the Hamiltonian:

$$\varphi(x, t + 2\pi/\omega) = \varphi(x, t) \quad (11.60)$$

⁸The treatment in this section is all based on the work *Electron transport in the Aharonov–Bohm pump* by Shin and Hong [73].

Substitution of (11.59) into (11.58) results in the following eigenvalue equation:

$$E_F \varphi(x, t) = \left[H(x, t) - i\hbar \frac{\partial}{\partial t} \right] \varphi(x, t) \quad (11.61)$$

The function $\varphi(x, t)$ is written as a sum of wave vector components k :

$$\varphi(x, t) = \sum_k a_k \xi_k(t) \chi_k(x) \quad (11.62)$$

with

$$\chi_k(x) = \exp \left[i \left(k + \frac{\pi}{\ell} f_s \right) x \right] \quad (11.63)$$

and

$$\xi_k(t) = \exp \left\{ -\frac{i}{\hbar} \left[\varepsilon t - \alpha_k \sin(\omega t + \phi) + \frac{\gamma}{2\omega} \sin 2(\omega t + \phi) \right] \right\} \quad (11.64)$$

where

$$\varepsilon = \frac{\hbar^2 k^2}{2m_e} + \gamma - E_F, \quad \gamma = \frac{\pi^2 \hbar^2 f_d^2}{4m_e \ell^2}, \quad \text{and} \quad \alpha_k = k \frac{\pi \hbar^2 f_d}{m_e \ell \omega}$$

Here it should be noted that the product $\xi_k(t) \chi_k(x)$ satisfies (11.61).

The periodic property (11.60) imposes the following constraint on the energy ε :

$$\varepsilon_m = m\hbar\omega \quad (11.65)$$

where m , being an integer, indicates the *Floquet channel*. Therefore, the wave vector k is also restricted:

$$\hbar k_m = \sqrt{2m_e(E_F + m\hbar\omega - \gamma)} \quad (11.66)$$

After some easy but somewhat cumbersome algebra, the wave function (11.59) is obtained as⁹

$$\begin{aligned} \Psi(x, t) = & \sum_{m,n,l=-\infty}^{\infty} a_m J_{m-n+2l} \left(\frac{\alpha_{k_m}}{\hbar} \right) J_l \left(\frac{\gamma}{2\hbar\omega} \right) \\ & \times \exp \left\{ i \left[\left(k_m + \frac{\pi}{\ell} f_s \right) x - (E_F + n\hbar\omega) t / \hbar + (m - n)\phi \right] \right\} \end{aligned} \quad (11.69)$$

It is very important to note that this formula depends both on the choice of the origin and on the direction of the electron motion; in particular, the above result takes the junction A as the origin and the clockwise for the direction of the electron motion. (Note that we defined k_m to be positive.) Because of the other junction C, there are actually four different version of the above result. The first thing to note that an electron moving counterclockwise direction sees the opposite flux of that seen by an electron moving clockwise; i.e. $f(t) \rightarrow -f_s - f_d \cos(\omega t + \phi)$. This change makes the argument of the first Bessel function in (11.69) negative, and using the second identity given in (11.68), the overall effect for the clockwise motion is $J_{m-n+2l} \rightarrow (-1)^{m-n} J_{m-n+2l}$. Secondly, the first term in the square parentheses must be modified as

⁹ **Verification of (11.69):** With the constraint (11.65) we write the time-dependent part (11.64) in full form as

$$\xi_{k_m}(t) = \exp \left[-im\omega t + i \frac{\alpha_{k_m}}{\hbar} \sin(\omega t + \phi) - i \frac{\gamma}{2\hbar\omega} \sin 2(\omega t + \phi) \right] \quad (11.67)$$

Using the identities

$$e^{-iz \sin \theta} = \sum_{n=-\infty}^{\infty} J_n(z) e^{-in\theta} \quad \text{and} \quad J_n(-z) = J_{-n}(z) = (-1)^n J_n(z) \quad (11.68)$$

we can write

$$\begin{aligned} \exp \left[i \frac{\alpha_{k_m}}{\hbar} \sin(\omega t + \phi) \right] &= \sum_{n=-\infty}^{\infty} J_n \left(\frac{\alpha_{k_m}}{\hbar} \right) e^{in(\omega t + \phi)} \\ \exp \left[-i \frac{\gamma}{2\hbar\omega} \sin 2(\omega t + \phi) \right] &= \sum_{l=-\infty}^{\infty} J_l \left(\frac{\gamma}{2\hbar\omega} \right) e^{-i2l(\omega t + \phi)} \end{aligned}$$

With these, (11.67) becomes

$$\xi_{k_m}(t) = \sum_{n,l} J_n \left(\frac{\alpha_{k_m}}{\hbar} \right) J_l \left(\frac{\gamma}{2\hbar\omega} \right) \exp \{ i [(-m + n - 2l)\omega t + (n - 2l)\phi] \}$$

The aim of all this algebra is to obtain the quantized form of (11.62). Noting that the sum k is now over m , we have

$$\varphi(x, t) = \sum_{m,n,l=-\infty}^{\infty} a_m J_n \left(\frac{\alpha_{k_m}}{\hbar} \right) J_l \left(\frac{\gamma}{2\hbar\omega} \right) \exp \left\{ i \left[\left(k_m + \frac{\pi}{\ell} f_s \right) x + (-m + n - 2l)\omega t + (n - 2l)\phi \right] \right\}$$

With the change of the indices as $-m + n - 2l \rightarrow -n'$, we get

$$\varphi(x, t) = \sum_{m,n,l=-\infty}^{\infty} a_m J_{m-n+2l} \left(\frac{\alpha_{k_m}}{\hbar} \right) J_l \left(\frac{\gamma}{2\hbar\omega} \right) \exp \left\{ i \left[\left(k_m + \frac{\pi}{\ell} f_s \right) x - n\omega t + (m - n)\phi \right] \right\}$$

Feeding this into (11.59) we obtain (11.69). $\square$

the following:

$$\begin{aligned}
& \left(k_m + \frac{\pi}{\ell} f_s\right) x && \text{(upper arm, clockwise)} \\
& \left(k_m - \frac{\pi}{\ell} f_s\right) x && \text{(lower arm, counterclockwise)} \\
& -\left(k_m - \frac{\pi}{\ell} f_s\right) (x - \ell) && \text{(upper arm, counterclockwise)} \\
& -\left(k_m + \frac{\pi}{\ell} f_s\right) (x - \ell) && \text{(lower arm, clockwise)}
\end{aligned}$$

Making the following definitions

$$K_m^\pm = k_m \pm \frac{\pi}{\ell} f_s \quad (11.70)$$

we write the final expression for the full wave function on the ring as

$$\begin{aligned}
\Psi^\pm(x, t) = & \sum_{m,n,l=-\infty}^{\infty} \left[a_m^\pm e^{iK_m^\pm x} + b_m^\pm (-1)^{(m-n)} e^{-iK_m^\mp (x-\ell)} \right] \\
& \times J_{m-n+2l} \left(\frac{\alpha_{k_m}}{\hbar} \right) J_l \left(\frac{\gamma}{2\hbar\omega} \right) e^{iF(t,\phi)}
\end{aligned} \quad (11.71)$$

with

$$F(t, \phi) = -(E_F + n\hbar\omega) t/\hbar + (m - n)\phi \quad (11.72)$$

The upper (lower) *sign* in (11.71) denotes the upper (lower) arm of the ring.¹⁰

Now we have come to the Floquet expressions of the wave function on both leads:

$$\Psi^{(i)}(x, t) = \sum_{n=-\infty}^{\infty} \left[a_n^{(i)} e^{iq_n x} + b_n^{(i)} e^{-iq_n x} \right] e^{-i(E_F + n\hbar\omega)t/\hbar} \quad (11.73)$$

where $(i) = (L), (R)$ denote the left and right leads, respectively. The boundary conditions are the wave function matching and the conservation of current density, as we did before for the constant magnetic field case. What is obtained is the system's scattering matrix of $\infty \times \infty$ dimension:

$$S = \begin{pmatrix} R & T' \\ T & R' \end{pmatrix} \quad (11.74)$$

where R and T (R' and T') denote matrices whose elements are probability amplitudes of reflection and transmission, respectively, for the propagating channel incident from

¹⁰ There is no mistake in writing $K_m^\mp$ in the second term inside the square brackets!

the left (right) lead. [See also (5.34) of our previous discussion about scattering matrices.] For example, t_{nm} , the matrix element of T , gives the probability amplitude of transmission for an electron incident from channel m on the left lead to appear at a channel n on the right lead.

We here put an end to this chapter. We tried to give a general formalism about AB rings exposed to a (possibly, time-periodic) magnetic flux. The remaining details can be found in the original papers of Yi *et al.* [72] and Shin & Hong [73], and the references cited therein.



CHAPTER 12

THE RICATTI-RATIOS METHOD IN DISCRETE TRANSPORT PROBLEMS¹

12.1 The Ricatti System

We consider a 1D mesoscopic system represented by a number of central sites which are coupled to the left and right leads. The model system under investigation is shown in Fig. 12.1; henceforth we refer to this model as the *Ricatti system* and designate it by R_s , where s is the number of central sites. Let j be the site index which will later designate the position of an electron. The sites for the left ($-\infty < j \leq -1$) and the right leads ($s \leq j < \infty$) are represented by filled circles, and the central sites ($0 \leq j \leq s-1$) by open circles. We investigate in the Ricatti system the resonant transmission of an electron tunneling through the central region, from the left to the right lead.

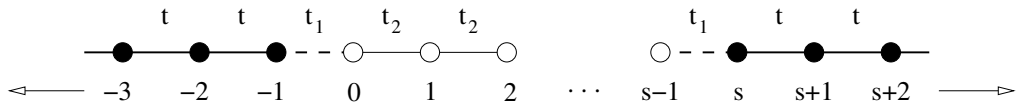


Figure 12.1: The Riccati system, R_s , under study.

¹ This chapter was originally written in a full-article format, so the reader might anticipate that its style might be a little different from that of the rest of the manuscript and that there might be some repetitions or overlapping of the previously elucidated details.

Since there is no any prescribed interaction, the main equation describing the motion of the single electron throughout all sites is just

$$E\phi_j = v_j \phi_j - \sum_{\langle j,k \rangle} t_{jk} \phi_k \quad (12.1)$$

which we have derived previously from the time-independent Schrödinger equation. In this working equation $\langle j, k \rangle$ means that only the nearest-neighbor hoppings are taken into consideration and that the case $j = \ell$ is to be excluded. We assign an onsite energy v_j , to site j , which is either a constant voltage bias with $v_{j \leq -1} = v$ and $v_{j \geq s} = v'$, or a constant (gate) voltage on the central sites with $v_{0 \leq j \leq s-1} = v_c$. Finally, $t_{j\ell}$ are the hopping amplitudes which mediate transport between two possible next-neighboring sites j and ℓ .

We can re-write (12.1) in the more clear form

$$E\phi_j = v_j \phi_j - t_{j,j+1} \phi_{j+1} - t_{j,j-1} \phi_{j-1} \quad (12.2)$$

As usual, we set the hopping energies $t_{j,j\pm 1}$ to t within leads sites, t_1 between the end sites of leads and the central region, and t_2 within central sites, as also indicated in Fig. 12.1 above.

The solution to (12.2) for an electron incident from the left lead is again given by

$$\phi_{j \leq -1} = e^{i(j+1)k} + R e^{-i(j+1)k} \quad (12.3)$$

$$\phi_{j \geq s} = T e^{i(j-s)k'} \quad (12.4)$$

where R and T are respectively the reflection and transmission amplitudes. (We take the amplitude of the incoming wave as being unity.) Also, k and k' in above solutions are the wave vectors of incoming and outgoing electron, respectively, which are to be determined from the condition of energy conservation,

$$E = v - 2t \cos k = v' - 2t \cos k' \quad (12.5)$$

where we take the distance between two neighboring sites as unity. It is now customary to use ε (ε') for the incoming (outgoing) energy of the electron,

$$\varepsilon = -2t \cos k \quad \text{and} \quad \varepsilon' = -2t \cos k' \quad (12.6)$$

so that

$$\varepsilon' = (v - v') + \varepsilon \quad (12.7)$$

which will be used during the numerical calculation.

The reader should notice one more time the mere role of leads in the system: their infinite length provides us with the electron with a continuous k -vector, or equivalently, with continuous energies, ε and ε' , between $-2t$ and $2t$ (or between -2 and 2).

Finally, the transmission coefficient T , from left to right, is to be calculated via

$$\mathsf{T}(\varepsilon, \varepsilon') = \frac{\sin k'}{\sin k} |T|^2 \quad (12.8)$$

We now demonstrate that it is amusingly straightforward to obtain the transmission amplitude T by defining the so-called *Ricatti-ratio* Y_j [8, 9]:

$$\phi_{j+1} = Y_j \phi_j \quad (12.9)$$

The use of this definition in (12.2) results in a two-point recursion relation for the Ricatti-ratios as

$$Y_{j-1} = t_{j,j-1} / (v_j - E - t_{j,j+1} Y_j) \quad (12.10)$$

From the solution (12.3) we have $\phi_s = T$ and $\phi_{-1} = 1 + R$. We relate T and R to each other by using (12.9) recursively:

$$\begin{aligned} \phi_s &= Y_{s-1} \phi_{s-1} = Y_{s-1} Y_{s-2} \phi_{s-2} \\ &= Y_{s-1} Y_{s-2} \cdots Y_{-1} \phi_{-1} \end{aligned} \quad (12.11)$$

whence

$$T = Y_{s-1} Y_{s-2} \cdots Y_{-1} (1 + R) \quad (12.12)$$

Thus, in order to obtain T explicitly in terms of Y_j , it will suffice an expression for R , which is to be determined again by using (12.9) for $j = -2$:

$$\phi_{-1} = Y_{-2} \phi_{-2}$$

But since, from the solution (12.3),

$$\begin{aligned} \phi_{-1} &= 1 + R \\ \phi_{-2} &= e^{-ik} + R e^{ik} \end{aligned}$$

we have, after a little algebra,

$$R = \frac{2i \sin k}{e^{ik} - 1/Y_{-2}} - 1 \quad (12.13)$$

We put this result into (12.12) and get the desired expression for T as

$$T = Y_{s-1} Y_{s-2} \cdots Y_{-1} \frac{2i \sin k}{e^{ik} - 1/Y_{-2}} \quad (12.14)$$

We also need to know Y_s , the initial datum; to obtain it we first note from the solution (12.4) that

$$\begin{aligned} \phi_s &= T \\ \phi_{s+1} &= T e^{ik'} \end{aligned}$$

Then Y_s follows from the use of (12.9) for $j = s$ as

$$Y_s = e^{ik'} \quad (12.15)$$

At this point the scenario is almost obvious: knowing the initial datum Y_s , we use the recursion relation (12.10) to obtain Y_{s-1} ; using Y_{s-1} , in turn, we get Y_{s-2} ; and so on until we determine Y_{-2} . We finally employ (12.14) and attain the desired transition amplitude T .

Equation (12.14), along with (12.10) is the central result of this chapter, which is, apart from the nearest-neighbor scheme used, exact. Because it directly follows from the Schrödinger equation, the transmission amplitude T embodies every bit of information about the system under question. All quantities of interest such as transmission, current, conductivity, etc. can be obtained through appropriate uses of T . The form of (12.14) reveals the reason of associating the Ricatti-ratios method with the adjective “pruning”: the effect of the semi-infinite right lead enter the formulation indirectly via the initial datum (12.15) as if we cut, or “prune,” the right lead and deposited all the information about it to site s . The reader is referred to the work of Haule & Bonca [10] to see another example of pruning.

We will now test the above recipe for the Ricatti systems with one- to four-site central parts, i.e., for the R_1 - R_4 systems. Referring to Fig. 12.1, the common system parameters that we choose are $v = v' = 0.0$ (no bias applied), $v_c \equiv v_0 = v_1 = v_2 = v_3 = 0.0$ (no gate potential applied to central sites), $t = 1.0$, $t_1 = 0.2$, and $t_2 = 1.0$. The reason for applying no bias and no gate potential is that we desire, as before, to see whether we can obtain correctly the eigenenergies of the central part. The plots in Fig. 12.2 show the transmission probability $T(\varepsilon, \varepsilon')$ versus the incoming electron energy ε for the systems under consideration. If we considered the central part as isolated, i.e., there were no left and right leads, equation (12.2) would then lead to the following now-well-learned expression for the eigenenergies ε_c of the central part:

$$\varepsilon_c(s) = -2t_2 \cos k_c \quad (12.16)$$

with

$$k_c = \frac{\pi n}{s+1} \quad (n = 1, 2, 3, \dots, s) \quad (12.17)$$

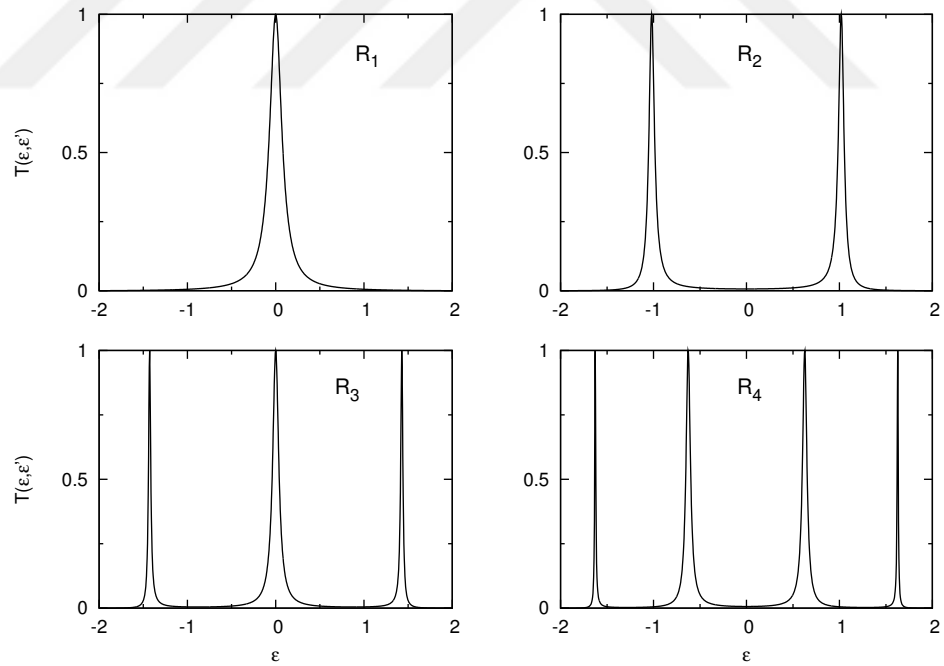


Figure 12.2: The transmission probability $T(\varepsilon, \varepsilon')$ versus the incoming electron energy ε for the R_1 - R_4 systems

Since $t_2 = 1.0$, the eigenenergies for isolated one- to four-site systems would be

$$\varepsilon_c(1) = 0.0$$

$$\varepsilon_c(2) = \pm 1.0$$

$$\varepsilon_c(3) = 0.0, \pm \sqrt{2}$$

$$\varepsilon_c(4) = \pm(\sqrt{5} + 1)/2, \pm(\sqrt{5} - 1)/2$$

Since there is no interaction, we expect the existence of leads to introduce only a barely discernible shift in the eigenenergies ε_c . Therefore, whenever the energy ε of incoming electron is around the above values, we expect a *total transmission*. What we read from Fig. 12.2 for the $\varepsilon(s)$ -values at which total transmission occurs, in the systems under study, are:

$$\varepsilon(1) \approx 0.0$$

$$\varepsilon(2) \approx \pm 1.02$$

$$\varepsilon(3) \approx 0.0, \pm 1.43$$

$$\varepsilon(4) \approx \pm 1.63, \pm 0.63$$

These values are all in very close proximity to the above values, a vindication of the recipe just presented.

12.2 Models Which Are Practically Reducible to Ricatti Systems: the Decimation Technique

12.2.1 Single Aharonov-Bohm Ring

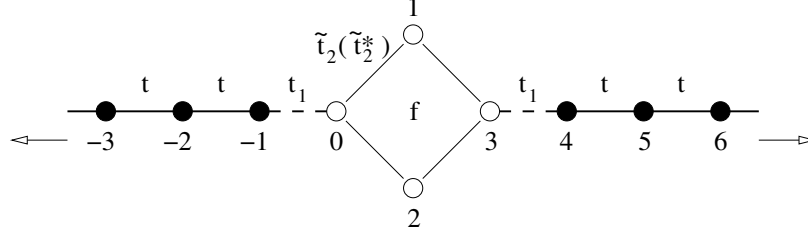


Figure 12.3: An imitation of a single AB ring.

What happens if any site in the central part (or anywhere in the infinite chain) happens to have more than two nearest neighbors? Figure 12.3 shows such a system in which sites $j = 0$ and $j = 3$ have three nearest neighbors. To make the matter at hand compounded, we allow a *constant* magnetic flux f to thread the “ring” formed by four central sites so that we practically deal with a simple single Aharonov-Bohm (AB) ring. In order to keep the discussion simple, we choose the onsite energies of central part zero, $v_c = 0.0$, and we apply some constant potential v (v') to the left (right) lead. Since it is developed for the Ricatti systems with sites which have only two nearest neighbors, the above formalism seems at first glance not practical for this case. We shall below demonstrate that, by making use of the decimation technique [7], such a system like in Fig. 12.3 can be transformed into a Ricatti system, after some simple algebra. First, we re-write (12.1), the equation for expansion coefficients ϕ_j ,

$$E\phi_j = v_j\phi_j - \sum_{\langle \ell, j \rangle} t_{j\ell}\phi_\ell \quad (12.18)$$

and we remember again that $\langle \ell, j \rangle$ indicates that only the nearest-neighbor of site j will be taken into account. The reader is now to notice that the hopping amplitude $t_{j\ell}$ might be taken, in general, to be a complex number because the effect of the magnetic flux f is to be included into the formalism by attaching a phase difference to the hopping constants within the ring. In other words, following the ideas presented in the preceding chapter, a gauge is chosen for the vector potential in which the effect

of magnetic fields appears as a change in the phase. Our convention is, again, that an electron traveling in clockwise manner, say, from site 0 to site 1 acquires a phase factor $e^{i\pi f/2}$ and from site 0 to site 1 in counterclockwise manner acquires another phase factor $e^{-i\pi f/2}$, which is the complex conjugate of the former. Then, in terms of hopping amplitudes, for example,

$$\begin{aligned} t_{01} &= \tilde{t}_2 = t_2 e^{i\pi f/2} \\ t_{10} &= \tilde{t}_2^* = t_2 e^{-i\pi f/2} \end{aligned}$$

with t_2 being now real. With the aid of Fig. 12.3, we can write from (12.18) the following set of equations:

$$\begin{aligned} &\vdots \\ E\phi_0 &= -t_1\phi_{-1} - \tilde{t}_2\phi_1 - \tilde{t}_2^*\phi_2 \\ E\phi_1 &= -\tilde{t}_2^*\phi_0 - \tilde{t}_2\phi_3 \\ E\phi_2 &= -\tilde{t}_2\phi_0 - \tilde{t}_2^*\phi_3 \\ E\phi_3 &= -\tilde{t}_2^*\phi_1 - \tilde{t}_2\phi_2 - t_1\phi_4 \\ &\vdots \end{aligned} \tag{12.19}$$

For the case for which $E \neq 0$, we can eliminate ϕ_1 and ϕ_2 in favor of ϕ_0 and ϕ_3 , and obtain the set

$$\begin{aligned} &\vdots \\ E\phi_0 &= v'_0\phi_0 - t_1\phi_{-1} - t_3\phi_3 \\ E\phi_3 &= v'_3\phi_3 - t_3\phi_0 - t_1\phi_4 \\ &\vdots \end{aligned} \tag{12.20}$$

with

$$v'_0 = v'_3 = \frac{2|\tilde{t}_2|^2}{E} = \frac{2t_2^2}{E} \tag{12.21}$$

and

$$t_3 = -\frac{1}{E} \left[(\tilde{t}_2)^2 + (\tilde{t}_2^*)^2 \right] = -\frac{2t_2^2}{E} \cos \pi f \tag{12.22}$$

Since the dots in both sets (12.19) and (12.20) stand for identical equations, a closer look at the latter set (12.20) reveals that it represents an R_2 system, shown in Fig. 12.4, in which every site has again only two nearest neighbors. What we did is that we transformed the single AB ring system with four-site central part (Fig. 12.3) into the R_2 system. We can now use the recipe of the previous section to investigate the transport properties of the AB ring under question. [For a comfortable use of (12.12), a re-enumeration of the sites in Fig. 12.4 is recommended.] The decimation technique [7] demonstrated straightforwardly here is again a different kind of pruning technique: the effect of sites 1 and 2 which have been “pruned” returns to the formalism as new gate potentials v'_0 and v'_3 to sites 0 and 3, respectively, together with a new hopping amplitude $t_{03} = t_3$. (Notice that we had $v_0 = v_3 = 0.0 = t_{03}$ before the pruning.)

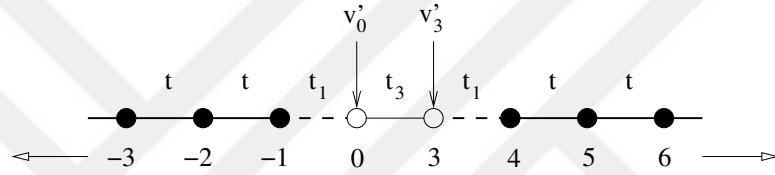


Figure 12.4: The transformed form of the single AB ring in Fig. 12.3.

In order to handle the seemingly problematical case in which $E = 0$, we resort to a limiting process. We first evaluate every quantity in our recipe in the previous section (i.e., k , k' , Y_s , T , and $\mathcal{T}$) as functions of E . Afterwards, to obtain the value of the transmission coefficient $\mathcal{T}$ at $E = 0$, we simply take its the upper and lower limits around $E = 0$ and assign

$$\mathcal{T}(0) = \lim_{E \rightarrow 0^+} \mathcal{T}(E) = \lim_{E \rightarrow 0^-} \mathcal{T}(E) \quad (12.23)$$

provided that both upper and lower limits are the same. This limiting process should be reasonable because, inasmuch as the both limits are equal, we do not expect, as a general rule, a one-point deviation at $E = 0$ from the result of these limits.

We show in Fig. 12.5 the graphs of the transmission probability $T(\varepsilon, \varepsilon')$ versus the incoming electron energy ε for the single AB ring shown in Fig. 12.3 (or, equivalently, for the R_2 system in Fig. 12.4) for different flux values in an consecutive manner. The common system parameters are $v = v' = 0.0$, $v_c = 0.0$, $t = 1.0$, $t_1 = 0.3$, and $t_2 = 0.5$. We see that when $f = 0.0$, the ε -values around which total transmission occurs are $\varepsilon = 0, \pm 1$. These are just what we expect from the formula

$$\varepsilon_c(s) = -2t_2 \cos k_c \quad (12.24)$$

with

$$k_c = \frac{2\pi n}{s} \quad (n = 0, 1, 2, \dots, s-1)$$

which gives the eigenenergies of an isolated ring consisting of s sites (with zero onsite potentials). These results demonstrate again that the existence of leads causes the eigenenergies given by (12.24) to shift only indiscernibly.

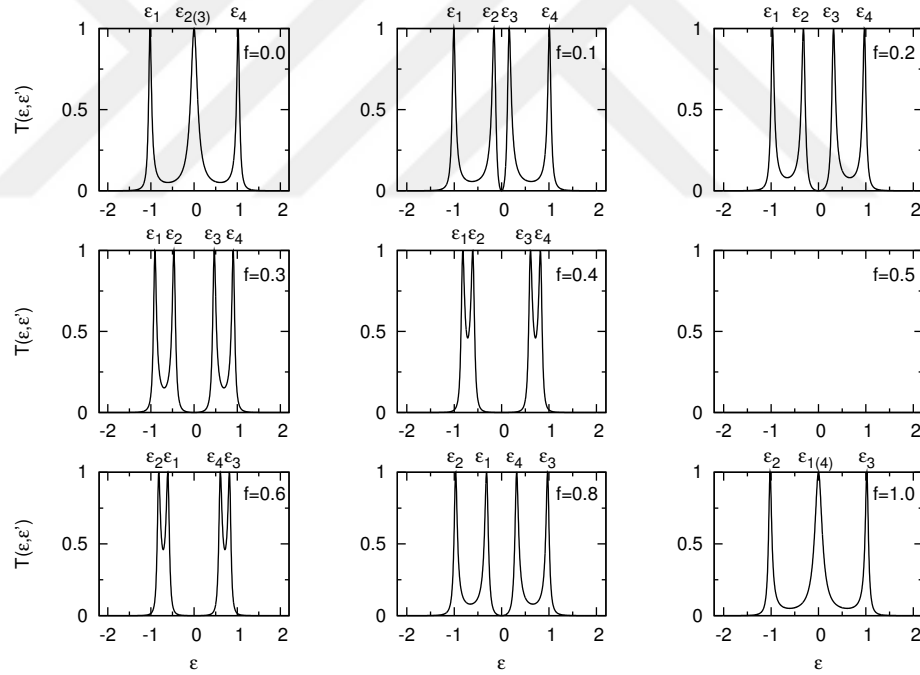


Figure 12.5: The transmission probability $T(\varepsilon, \varepsilon')$ versus the incoming electron energy ε for the single AB ring shown in Fig. 12.3.

We also added the labels of eigenvalues to the graphs in Fig. 12.5; for instance, in the graph for $f = 0.0$, they are $\varepsilon_1 = -1.0$, $\varepsilon_{2(3)} = 0.0$, and $\varepsilon_4 = 1.0$. When we apply a nonzero flux (observe, for instance, the graph for $f = 0.1$), the degeneracy of the eigenvalue zero is lifted, and there appear four distinct eigenvalues. Perhaps what is more interesting in those figures is that the eigenvalues seem to be in motion as f varies from 0.0 to 1.0: ε_1 and ε_3 move to the right, and ε_2 and ε_4 to the left. Around $f \approx 0.5$ (not shown), ε_1 and ε_2 (ε_3 and ε_4) will practically combine to one value around $\varepsilon \approx -\sqrt{2}$ ($\sqrt{2}$). Because of the perfect destructive interference, *exactly* at $f = 0.5$, as shown in Fig. 12.5, the transmission probability is thoroughly zero for every value of electron's incoming energy ε . In other words, at $f = 0.5$ we witness a *total reflection* (antiresonance) phenomenon irrespective of the ε -value [72, 74]. If f is continued to increase more from 0.5 to 1.0, we might figure out two scenarios. In the first scenario, as demonstrated in Fig. 12.5 we may say that eigenvalues ε_1 - ε_4 keep moving in the same directions as before, i.e., ε_1 and ε_3 move to the right, and ε_2 and ε_4 to the left, and they end their journey at the configuration shown in the figure for $f = 1.0$. In the second scenario (not shown), the eigenvalues stop their motion at $f = 0.5$, and, as f is increased, they start to go back, ε_1 and ε_3 move to the left, and ε_2 and ε_4 to the right, and they come to rest at $f = 1.0$ at the exactly same configuration they started from at $f = 0.0$.

The behavior of eigenvalues in Fig. 12.5 is actually predictable; they obey the formula, according to the gauge we chose,

$$\varepsilon_c(s) = -2t_2 \cos k_c \quad (12.25)$$

with

$$k_c = \frac{2\pi}{s}(n - f) \quad (n = 0, 1, 2, \dots, s - 1)$$

which gives the eigenenergies of an isolated ring, threaded by the flux f , consisting of s sites (which is 4 in our case), with no onsite potentials. We wish to again stress that the existence of leads causes the eigenenergies given by (12.25) to shift only indiscernibly although this time there is a magnetic flux threaded the ring.

Before passing, we want to look at the limiting process about the singular point $E = 0$ (or, equivalently, $\varepsilon = 0$ in our sample AB ring). We evaluated the upper and lower limits around $\varepsilon = 0$ and found out $\lim_{\varepsilon \rightarrow 0^+} T(\varepsilon) = 1.0$ for the cases in which $f = 0.0$ or $f = 1.0$, and $10^{-46} \leq \lim_{\varepsilon \rightarrow 0^+} T(\varepsilon) \leq 10^{-12}$ for all other cases shown in Fig. 12.5. Accepting the latter limit values practically zero, we see that these values are just what we expect, a good confirmation of the limiting process applied around singular point $\varepsilon = 0.0$. As a result, in practical numerical calculations the point $\varepsilon = 0.0$ can be omitted.

12.2.2 Double AB Ring

In the previous illustrations we studied some simple test systems whose transmission properties were known in advance. By using first the decimation method, if necessary as in the case of the single AB ring, we transformed them into a Ricatti system, and then applying the Ricatti-ratios method we were able to reproduce well-known features of the systems under question. In the following we study a more intriguing system whose transmission properties cannot be estimated immediately. As shown in Fig. 12.6, we add a second neighboring ring to our primitive model in Fig. 12.3 to obtain an imitation of a double AB ring system. The two rings are threaded by the same constant magnetic flux f . Unlike the models studied above, it is not easy to write down an expression giving the eigenenergies of the isolated central double-ring system with or without the magnetic flux f . By working in the reverse direction, we now demonstrate that the decimation and Ricatti-ratios methods can be readily used for determining the eigenenergies of the isolated central part (whose physical structure can be complicated to any desired extent) within nearest-neighbor level.

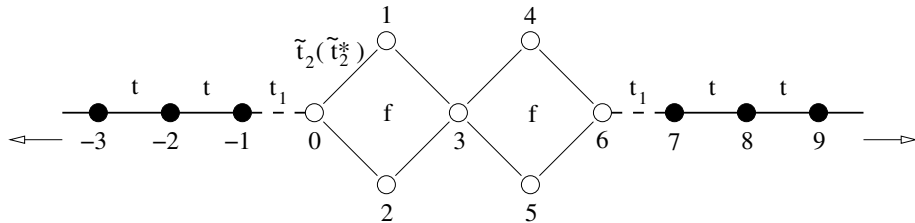


Figure 12.6: An imitation of a double AB ring.

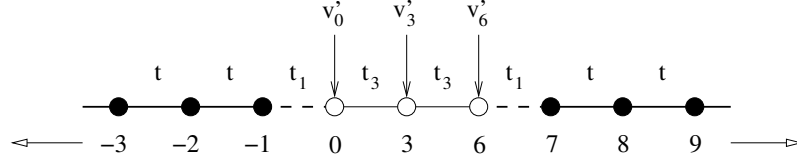


Figure 12.7: The transformed form of the double AB ring in Fig. 12.6.

As in the case of single AB ring, we again choose all the onsite energies of the central part as being zero, $v_c = 0.0$, and we apply some constant potential v (v') to the left (right) lead. Following the same procedure outlined for the single AB ring, the reader now should be able to verify that the double AB ring in Fig. 12.6 can be transformed into the one in Fig. 12.7 with the new parameters (where $E \neq 0$)

$$\begin{aligned} v'_0 &= v'_6 = \frac{2t_2^2}{E} \\ v'_3 &= \frac{4t_2^2}{E} \\ t_3 &= -\frac{2t_2^2}{E} \cos \pi f \end{aligned} \tag{12.26}$$

For the singular $E = 0$ case we again resort to the same limiting procedure as we did in the single AB ring case. This time we reduced the double AB ring to a R_3 system. (We again recommend to re-enumerate the sites in Fig. 12.7 before the use of the Ricatti-ratios method.)

Figure 12.8 demonstrates the graphs of the transmission probability $T(\varepsilon, \varepsilon')$ versus the incoming electron energy ε for the double AB ring in Fig. 12.6 (or, equivalently, for the R_3 system in Fig. 12.7) for different flux values. The common system parameters are the same as those used in the single AB ring which we studied above: $v = v' = 0.0$, $v_c = 0.0$, $t = 1.0$, $t_1 = 0.3$, and $t_2 = 0.5$. We now recall again that the leads cause only a barely discernible shift in the eigenenergies ε_c of the isolated central part and that whenever the energy ε of the incoming electron is around any of these ε_c values we expect the transmission probability to be unity. Then, we make use of the graph of the $f = 0.0$ case, and the data it is based on, to determine the eigenenergies of the isolated double-ring system, for the system parameters chosen, as $\varepsilon_c \approx 0.0, \pm 0.72, \pm 1.23$.

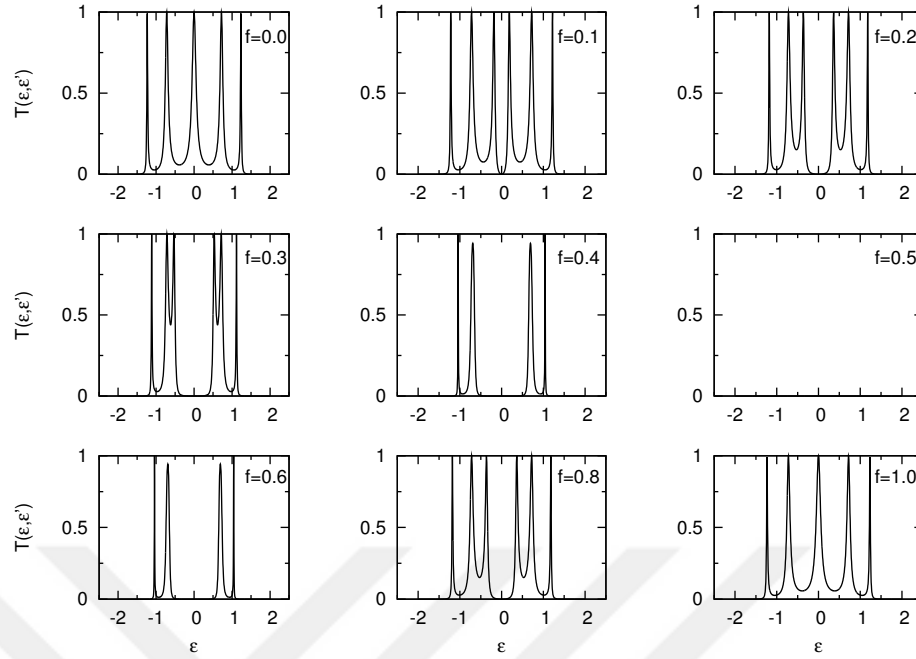


Figure 12.8: The transmission probability $T(\varepsilon, \varepsilon')$ versus the incoming electron energy ε for the double AB ring shown in Fig. 12.6.

Furthermore, the graph of the $f = 0.1$ case clearly shows that the zero eigenenergy is doubly degenerate. To recapitulate the main point, the reduction and Ricatti-ratios methods presented in this work can be successfully applied to determine the energetic properties of any system in an a priori manner just before a more complex analysis.

As we have witnessed in the single AB ring, the eigenenergies of the double AB ring are, too, in motion as flux changes, as is clearly seen in Fig. 12.8. As f increases, they move, combine, diminish, reappear, and so on. At $f = 0.5$ *exactly*, we again witness a total reflection (antiresonance) phenomenon irrespective of ε because of perfect destructive interference.

We clearly see in Figs. 12.5 and 12.8 that in both the single and double AB rings the antiresonance phenomenon comes into play at $f = 0.5$ in which there exists a total reflection irrespective of the incoming electron energy ε . According to the findings of Yi *et al.* [72] presented in the previous chapter, at some special values of $(k\ell, f)$, where k is the wave number of the incoming electron and ℓ is the half length of the single ring, the continuous versions of the single and double AB rings exhibit resonance

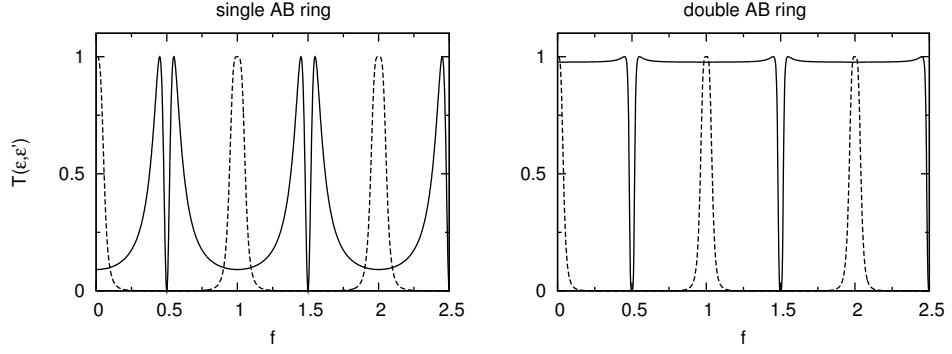


Figure 12.9: The transmission probability $T(\varepsilon, \varepsilon')$ versus the magnetic flux f for the single and double AB rings at $k = 0.05\pi$ (solid lines) and at $k = 0.495\pi$ (dashed lines).

or antiresonance phenomena. (Note that $\ell = 2$ units in our cases.) We might expect the results of the *discontinuous* single and double AB rings to be reminiscent of the continuous versions. In order to check this, we take the common system parameters in both AB rings as $\nu = \nu' = 0.0$, $\nu_c = 0.0$, $t = t_1 = t_2 = 1.0$ and evaluate the transmission probability at some specific k -values near zero and 0.5π while varying f . Figure 12.9 shows the results; we immediately note that while the double AB ring system exhibits both resonance and antiresonance, the single AB ring system exhibits only antiresonance. We see the exactly the same behavior in the continuous case: the graphs in Fig. 12.9 are the humble simulacra of those of Yi *et al.* [72], which we have shown in Fig. 11.11 of the previous chapter. What is different between the continuous case and the present discontinuous case of Yi *et al.* is that the product $k\ell$ of the former case corresponds to $2k$ of the latter case and the graphs near $k = 0$ ($k = 0.5\pi$) in the discontinuous AB rings correspond to those near $k\ell = 0.5\pi$ ($k\ell = 0$) in the continuous AB rings. All the remaining features of the graphs in both works are strikingly the same.

In this last chapter, we have had a closer look at the quantum transport of an electron through a 1D quantum-wire-like mesoscopic system coupled to two noninteracting reservoirs (i.e., left and right leads). We have based our formalism on the nearest-neighbor approximation. We have shown how to attain the transmission properties of the system under question by making use of the Ricatti-ratios method, a pruning technique transforming the Schrödinger equation into a two-point recursive equation. The Ricatti-ratios recipe has been then tested in a series of 1D systems and all the well-

known results have been recovered, showing that the presented method is numerically exact. The decimation method, a powerful reduction method, has been introduced in reducing the quasi 2D AB ring systems to one dimension. The results for the single and double AB rings further confirmed the power of two methods: we were able to show the resonance and antiresonance phenomena in these discontinuous ring systems, which were well-known in their continuous counterparts. We have seen in all the systems considered here that the existence of leads has introduced only a barely discernible shift in the eigenenergies of the central part, as we just expected (so that the name “resonant transport” finds its meaning). Using the double AB ring as an example, we have exhibited how to use the two method to determine the eigenenergies of a more complicated mesoscopic quantum dot system in an a priori manner.

The obtained results render the Ricatti-ratios method promising in the applications to more complex mesoscopic systems. It can be exploited while studying the resonant tunneling of an electron through a mesoscopic system in the presence of electron-phonon and/or electron-photon interaction, or with any type of similar inelastic interactions. It is perfectly possible to derive a matrix version of this method that merges the tight-binding and the Floquet theory into a two-point recursive matrix equation; such a version is particularly useful in considering time-periodic excitations. Along with the decimation technique, the matrix version may be then successfully utilized in investigating AB type quasi 2D electron pump systems with or without electron-phonon interactions [9, 3, 4]. The list of unexplored possibilities for the use of this method can go on and on.

REFERENCES

- [1] Kagoshima S., Nagasawa H., and Sambongi T., *One-Dimensional Conductors*, ISBN 0-387-18154-7, Springer-Verlag, Berlin, 1988.
- [2] Mattis D.C., in: *Physics in One Dimension*, edited by Bernasconi J. and Schneider T., ISBN 3-540-10586-7, Springer-Verlag, Berlin, 1981.
- [3] Oymak H., Shin D., and Hong J., *Europhysics Letters* **77**, 37005 (2007).
- [4] Shin D., H. Oymak H., Hong J., *J. Phys.: Condens. Matter* **19**, 226211 (2007).
- [5] Bloch F., *Z. Phys.* **52**, 555 (1928).
- [6] Goringe C.M., Bowler D.R., and Hernández E., *Rep. Prog. Phys.* **60**, 1447 (1997).
- [7] Pastawski H.M. and Medina E., *Rev. Mex. Fis.* **47**, 1 (2001).
- [8] Heinrichs J., *Phys. Rev. B* **65**, 075112 (2002).
- [9] Faizabadi E. and Ebrahimi F. *J. Phys.: Condens. Matter* **16**, 1789 (2004).
- [10] Haule K. and Bonča J., *Phys. Rev. B* **59**, 13087 (1999).
- [11] Ref. [23]; p. 743.
- [12] Halliday D., Resnick R., and Walker J., *Fundamentals of Physics*, Extended Eighth Edition. ISBN 978-0-471-75801-3, John Wiley and Sons, Inc., 2008; p. 686.
- [13] Ref. [12]; p. 690.
- [14] Ref. [12]; p. 692.
- [15] Ref. [23]; pp. 752-755.
- [16] Ref. [16]; p. 513.
- [17] Ref. [23]; p. 753.
- [18] Liboff R.L., *Introductory Quantum Mechanics*, Third Edition. ISBN 0-201-87879, Addison-Wesley Publishing Company, Inc., 1998.
- [19] Stapp H.P., *Am. J. Phys.* **40**, 1098 (1972).
- [20] Beiser A., *Concepts of Modern Physics*, Fifth Edition. ISBN 0-07-004814-2, McGraw-Hill, Inc., 1995; p. 161.
- [21] Tipler P.A. and Mosca G., *Physics for Scientists and Engineers*, Extended Fifth Edition. ISBN 0-7167-4389-2, W.H. Freeman and Company, New York, 2004; p. 474.

- [22] *Ibid*; p. 466.
- [23] Serway R.A., *Physics for Scientists and Engineers with Modern Physics*, Third Edition. ISBN 0-03-031353-8, Saunders College Publishing, 1990; p. 495.
- [24] Beiser A., *Concepts of Modern Physics*, Fifth Edition. ISBN 0-07-004814-2, McGraw-Hill, Inc., 1995; p. 97.
- [25] Ref. [18]; p. 85.
- [26] Sakurai J.J., *Modern Quantum Mechanics*, First Revised Edition. ISBN 0-201-53929-2, Addison-Wesley Publishing Company, Inc., 1994; p. 72.
- [27] Ref. [18]; p. 188.
- [28] Ref. [18]; p. 190.
- [29] Ref. [18]; p. 217.
- [30] Ref. [18]; p. 218.
- [31] Ref. [18]; p. 225.
- [32] Griffiths D.J., *Introduction to Quantum Mechanics*. ISBN 0-13-124405-1, Prentice-Hall, Inc., 1995; pp. 53-59.
- [33] Ref. [32]; p. 54.
- [34] Ref. [32]; p. 56.
- [35] Ref. [32]; p. 66.
- [36] Ross S.L., *Differential Equations*, Third Edition. ISBN 978-81-265-1537-0, John Wiley and Sons, Inc., 2004; p. 109.
- [37] Ref. [18]; p. 97.
- [38] Feynman R.P., *The Feynman Lectures on Physics. Volume 3: Quantum Mechanics*. ISBN 0-201-02010-6-H, Addison-Wesley Publishing Company, 1963; pp. 3-2–3-4.
- [39] Ref. [38]; pp. 7-1–7-3.
- [40] Ref. [38]; pp. 8-1–8-3.
- [41] Ref. [38]; pp. 8-3–8-5.
- [42] Ref. [38]; pp. 8-7–8-11.
- [43] Ref. [26]; p. 69.
- [44] Ref. [38]; pp. 8-11–8-14.
- [45] Ref. [38]; p. 8-11
- [46] Ref. [38]; p. 8-14

- [47] Russell D., *Coupled Oscillators*, a video demonstration at “<https://www.youtube.com/watch?v=YyOUJUOUvso>”.
- [48] Ref. [38]; pp. 10-2–10-3.
- [49] Ref. [38]; p. 10-2.
- [50] Ref. [38]; p. 10-3.
- [51] Ref. [38]; pp. 10-8–10-10.
- [52] Ref. [38]; p. 10-8.
- [53] Ref. [18]; p. 600.
- [54] Ref. [38]; pp. 13-1–13-13.
- [55] Ref. [38]; p. 13-2.
- [56] Ref. [38]; p. 13-4.
- [57] Ref. [38]; p. 13-5.
- [58] Ref. [38]; p. 13-7.
- [59] Ref. [38]; p. 13-11.
- [60] Ref. [38]; p. 13-12.
- [61] Sutton A.P., *Electronic Structure of Materials*. ISBN 0-19-851754-8, Clarendon Press, Oxford, 2004; p. 39.
- [62] Ref. [23]; p. 1190.
- [63] Ref. [61]; p. 42.
- [64] Ref. [61]; p. 44.
- [65] Ref. [61]; p. 45.
- [66] Ref. [16]; pp. 1268-1271.
- [67] Ref. [16]; pp. 1269.
- [68] Marion J.B. and Thornton S.T., *Classical Dynamics of Particles and Systems*. ISBN 0-03-097302-3, Saunders College Publishing, New York, 1995; pp. 234-235.
- [69] Ref. [32]; p. 175.
- [70] Aharonov Y., Bohm D., Phys. Rev. **115**, 485 (1959).
- [71] Chambers R.G., Phys. Rev. Lett. **5**, 3 (1960).
- [72] Yi J., Wei. J.H., Hong J., and Lee S.-I. Phys. Rev. B **65**, 033305 (2001).
- [73] Shin D. and Hong J., and Lee S.-I. Phys. Rev. B **70**, 073301 (2004).
- [74] Park W. and Hong J., Phys. Rev. B **69** 035319 (2004).