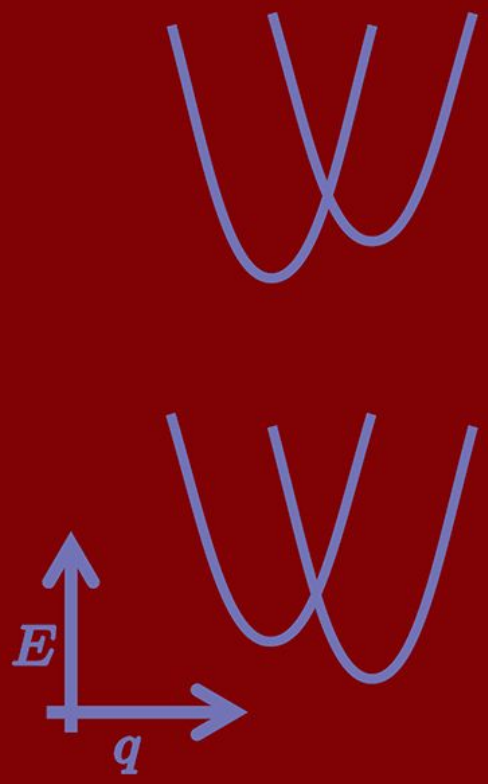
OXFORD



Semiconductor Physics

Principles, Theory and Nanoscale

Sandip Tiwari

ELECTROSCIENCE SERIES | **VOLUME 3**

SEMICONDUCTOR PHYSICS

This text is the third volume of the series titled Electroscience.

1. S. Tiwari, *Quantum, statistical and information mechanics: A unified introduction*, Electroscience 1, Oxford University Press, ISBN 978-0-198-75985-0,
2. S. Tiwari, *Device physics: Fundamentals of electronics and optoelectronics,* Electroscience 2, Oxford University Press, ISBN 978-0-198-75984-3,
3. S. Tiwari, *Semiconductor physics: Principles, theory and nanoscale,* Electroscience 3, Oxford University Press, ISBN 978-0-198-75986-7 (2020) and
4. S. Tiwari, *Nanoscale device physics: Science and engineering fundamentals,* Electroscience 4, Oxford University Press, ISBN 978-0-198-75987-4 (2017).

These volumes comprise a sequence of undergraduate to graduate textbooks of the underlying physical foundations leading up to advanced devices of nanometer scale. Teaching slides are available on the book's companion website at www.oup.co.uk/companion/semiconductorphysics2020, and the solutions manual may be requested at global.oup.com/uk/academic/physics/admin/solutions.

Semiconductor Physics

Principles, Theory and Nanoscale

ELECTROSCIENCE SERIES, VOLUME 3

Sandip Tiwari

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

First Edition published in 2020
Impression: 1

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data
Data available

Library of Congress Control Number: 2019941496

ISBN 978–0–19–875986–7

Printed and bound by
CPI Group (UK) Ltd, Croydon, CR0 4YY

To

Pampam bhai (Anupam Mishra),

my childhood hero and beacon

(those visits to kabadi bazaar mattered singularly in this life);

and to the Mishra family

for showing the way to

love, grace, dignity, conviction and patience by example.

॥ फर्क ॥

मेरा और तुम्हारा
सारा फर्क,
इतने में है
कि तुम लिखते हो
मैं बोलता हूं,
और कितना फर्क हो जाता है इससे
तुम ढांकते हो, मैं खोलता हूँ ।

भवानी प्रसाद मिश्र

The difference

The entire difference
between you and me
is that you write
and I talk,
and how entirely different they are.
You cover it and I open it.

Bhawani Prasad Mishra

Acknowledgments

One's approach to life and, more specifically, to the choices one makes in the course of it, the questions one asks and the irritants one learns to ignore have a cumulative outcome. Life is defined by beauty, love, taste and the passion that propels one in pursuit of these lambent aspirations. In this, one's family and friends, early and advanced education, the company one keeps and the approaches one learns from the especially gifted people that one has the good fortune of befriending all make a difference. So do people who mentor and support through the vicissitudes of life. I was lucky to have had a good share of exemplary teachers in early years, and during my undergraduate education at the Indian Institute of Technology, Kanpur, as well as particularly enlightened colleagues during my years at the research laboratory at the International Business Machines Corporation (*IBM*). The latter was a community with a preponderance of exceptional people. Every encounter had something to learn from it. Science ruled. Nonsense was unacceptable. All these benefactors have influenced choices I have made, especially the nature of questions I became interested in and asked and the multifaceted approaches of different perspectives I used in answering them. That some parts of businesses can run with academic ideals for the greater good and that academe is so much beholden to business norms has been a surprising observation of mid-life. The loss of the research institutions, particularly the *IBM* and Bell laboratories, and the reduction in the diversity of thoughts and approaches that has followed, has changed the American story. These institutions were exemplary for most of their lives in their approach to discovery, research and development, understood their interconnectedness and had an intellectual commitment that overrode palace intrigues. The test of good management is in how an institution responds and negotiates through hard times. *IBM* maneuvered remarkably well for a fair period.

Foremost among the people I wish to acknowledge are the colleagues I had at Yorktown during Research's golden

period. Among them, Paul Solomon, David Frank, Steve Laux, Massimo Fischetti, Wen Wang, Frank Stern, Rolf Landauer, Peter Price, Bob Dennard, Subu Iyer, Tom Jackson, Supratik Guha, Pat Mooney, Arvind Kumar, Chuck Black, Jeff Welser, Tom Theis, Doug Buchanan, Dan DiMaria, Jim Stathis, Emilio Mendez, Leo Easaki, Jim Misewich, Ravi Nair, Charles Bennett, Hans Rupprecht, Jerry Woodall, Peter Kirchner, Jeff Kash, Jimmy Tsang, Steve Wright, Dennis Rogers, Marshall Nathan, Frank Fang, Alan Fowler, Reuben Collins, Dick Rutz and many others shared technical and life-enriching wisdom.

This text being about semiconductors, I want to particularly acknowledge the long discussions with my contemporaries David Frank, Steve Laux and Max Fischetti, each with their own different interests but always ready to discuss the numerous interesting questions that kept unfolding as new technologies made new artificial creations in semiconductors with their own questions possible. Arvind Kumar later on joined this illustrious group. Paul Solomon straddled a sagacious understanding of devices and their materials' physics simultaneously. Among the senior colleagues, Frank Stern, with his Dirac-like devotion to choice of words and science, influenced all of us. His co-authored *Reviews of Modern Physics* book-sized article on two-dimensional systems continues to influence the community to this day. One had to be very careful in listening to Peter Price lest a gem of an insight—spoken in a few choice words—was missed. Soon after I left for academe, upon reading a letter of mine in the *Wall Street Journal* about the importance of bureaucracy in getting things done properly—I was responding to a published anti-government opinion that used developing nations as examples by pointing out that it was the select dedicated folks that partly saved the day when the politicians royally failed with the Indian partition and the accompanying migration—Peter wrote, "Sandip, the Op Ed page of *WSJ* is scarcely the place to look for rationality."

The conflict between a humane society and a living economy is a broader tension. It has always been. We fail at teaching our graduating students not to force fit "truth" to our biases. We succeed with too few. Part of the reason for this failure is that a business style—talking points, elevator pitch, slides to which one talks, spreadsheets, college bookstores as sweatshirt shops with little space for books, libraries as cafés with the loss of space and collection where one was exposed to printed thoughts that one was not actively searching for, and, underlying all this, finance over content—has broadly infected the educational institutions of this country. Science and technology are major social and economic

One notable aphorism from those days is as follows: "Beware the *PF*9s and *PF*10s." Before the internet, or the bitnet before that, there existed within *IBM* an internal network (*VNET*) for electronic mail, for accessing repositories of useful codes and technical documents that individuals had written, and for other network-wide computing tasks. *PF*9 and *PF*10 were programmable function keys for the "Receive" and the "Send" tasks in the mailing program. "*PF*9s and *PF*10s" was the euphemism for the doorkeepers, speed breakers and messengers—folks whose technical careers were short but who had the wherewithal to generate pointless work—from whom one needed to protect oneself to be a good scientist and engineer.

forces. By not finding a balance between being the source of technological and financial success, which is initially limited to only a few who are educated and have power, and is economic, and in developing citizenship, which is a broader goal and is social, academia seeds a future downfall if it is not fixed. The standards need to be high, and, for this social commitment theme, legal acceptability or checking boxes of ethics as dictated by accreditation boards is not that standard. This is particularly important in science and engineering, where the student usually has a very different mental focus than the student of humanities.

The early thoughts on how to organize this writing occurred during a sabbatical leave at Harvard in 2006–2007. I even wrote some preliminary notes. But work could not begin in earnest until I managed to relinquish several responsibilities that work life brings. The first drafts started in Ithaca around 2010, but serious work had to wait for my next sabbatical leave in 2012–2013, which also gave a chance to try the material—of the fourth and this third volume—with different student audience. At the Indian Institute of Science, my hosts included Professors Navakant Bhat, Rudra Pratap and S. Shivashankar; at Stanford University, Professor Roger Howe; and, at Technische Universität München, Professor Paolo Lugli. The environment at these institutions was ideal for what I had in mind. And, in addition, it provided an opportunity to be in the company of several other faculty with a joyful outlook to science and life: Professor Ambarish Ghosh, Srinivasan Raghavan, Philip Wong, Yoshio Nishi, Walter Harrison, Christian Jirauschek, Wolfgang Porod and Peter and Johannes Russer—my immense gratitude to them for a stimulating year. The students who participated in these courses around the world have provided invaluable feedback that is reflected in the writing and rewriting.

Many colleagues have read and commented on parts or the whole, and this has helped with the exposition. To Tom Theis, Wolfgang Porod, Max Fischetti, Jerry Tersoff, Supratik Guha, Ed Yu, Federico Capasso, Siegfried Selberherr, Srikrishnaa Vadivel and Kunal Tiwari, my thanks for sharing their time and suggestions. Jack and Mary East—my dear friends—have given crucial support through their constant interest in the progress of this work. My wife Mari has kept a careful eye on the goings-on, helped keep my centrifugal propensities in check and given invaluable advice on the presentation. The very constructive exchange with Sonke Adlung, Harriet Konishi and Elizabeth Farrell at Oxford University Press has been immensely valuable in the creation of the final form.

The LaTeX class for Edward Tufte's style suggestions has been largely followed in these texts. The authors of such open source

When I first arrived at IBM Research in the beginning of the 80s, *APL*—A Programming Language—as an abstract and compact computational and graphical language tool was an eye opener. It had beauty, and it let you tackle data manipulation, matrix calculations such as for differential equations, searches, et cetera, all quite compactly, together with publishable graphics. But, as used to be said about *APL* programs' readability, "It was tough to write. It should be tough to understand." I found deciphering even my own codes that were just a few years old hard going. Now, when using Python, where, as with *MATLAB*, I can see through where these systems and their syntax built itself from, I do appreciate the readability, reusability and clarity that underlie the evolution. This goes together with the difficulties I have certainly faced with undoing the programming relearning that came from having been exposed to *FORTRAN* first. Those indiscriminate *GO TO* statements make a complete tangle of the ball of string that has an $\mathcal{O}(n^2)$ complexity.

resources, which here also includes Python for the calculations in the exercises and figures, perform an immense service to the society.

Over the years, I have been fortunate to have had generous and understanding mentors and supporters who have made the research and academic pursuits fulfilling. The research environment of those early years was a reflection of the focus on research with a perspective. The management ably fostered this.

In days past, on the board outside Professor Les Eastman's office, there used to be a press clipping circa the late 1970s, from the Universal Press Syndicate, with the following quote: "Our futures almost certainly depend less on what Ronald Reagan and Walter Mondale say and do than on what is going on inside the head of some young Cornell graduate waiting for a plane in Pittsburgh." The routes are now through Philadelphia, but the thought is still right. To students who have interacted with me through the classes goes the ultimate tip of the hat.

Science touches us all through its beauty. My colleagues, teachers, family, students and others who have, through occasional talks, conversations, writings, the way an argument was framed, a clever twist of reasoning, or even plowing through when a situation demands, have enlightened this recognition.

Sandip Tiwari
Ithaca, Orsay, Southport and Bhopal

Contents

Introduction to the series

These books are a labor of love and love of labor. They reflect a personal philosophy of education and affection for this small but vital subject area of electroscience, one that has given me satisfaction. This subdiscipline is also a domain where knowledge has evolved rapidly, leaving in its wake an unsatisfactory state in the coherence of content tying mathematical and physical descriptions to its practice.

Engineering-oriented science education, even though not really that different from science education itself, is difficult for two reasons. It aims to provide strong scientific foundations and also to make the student capable of practicing it for society's benefit. Adequate knowledge of design and technology to invent and optimize within constraints demands a fundamental understanding of the natural and physical world we live in. Only then can we create usefully with these evolving tools and technology. Three hundred years ago, calculus and Kepler's and Newton's laws may have been adequate. Two hundred years ago, this basic foundation was expanded to include a broader understanding of thermodynamics; the Lagrangian approach to classical mechanics; probability; and the early curiosity about the compositional origins of lightening. A hundred years ago, the foundational knowledge had expanded again to include a fair understanding of the periodic table, Hamiltonians, electricity, magnetism, and statistical mechanics, and, yet again, it was incomplete, as the development of new, non-classical approaches, such as Planck's introduction of quantum of action, and relativity, showed. Our understanding still remains very incomplete today even as evolution is gathering pace via science and engineering with non-carbon forms of intelligent machines quite imaginable. Reductive and constructive approaches, as before, pervade the pursuit of science and engineering. Understanding singularities, whether in black holes, in phase transitions with their information mechanics implications, or for solving near-infinite differential equations with near-infinite variables and constraints

A slight aside. Behind this education is an urge to understand our universe, the nature in which we exist, and perhaps through our acts bringing about small changes around us to make life and the world a bit better. When my eldest son came back home after first semester in college, he said, "I now get it. Biology is the emergence of chemistry, chemistry is the emergence of physics, and physics of mathematics." This is a modern take of Galileo Galeilei's statement in *Sidereus Nuncius*, "Nature is written in that great book which ever lies before our eyes. · · · · The book is written in the mathematical language, and the symbols are triangles, circles and other geometrical figures without whose helps it is humanly impossible to comprehend a single word of it, and without which one wanders in vain through a dark labyrinth." The symbols and this language, especially via Leibniz's calculus, have expanded tremendously. This Cartesian world view of physical reality, though powerful, is reductionist and incomplete in the Gödelian sense. Mathematics, music, paintings, and good writing are universal languages that reach out from the natural world to our senses. Objectivity is not

and the networks they form, is central to science and engineering problems—connecting back the two ends of the string.

All this evolving knowledge and its usage would be deficient were we incapable of adequately using the tools available, which in their modern forms include software for the implementation of mathematics and their computational, observational and experiment-stimulating machinery and their operating software for designing and optimizing suitable answers to the questions posed. A physical understanding of the connections between different interactions, as well as the reasoning that leads one to identify the most primary of these interactions, is essential to utilizing them gainfully.

Another conundrum and this is particularly true for three volumes of this series is that much of this subject area is at the intersection of science and engineering with both important. A scientist is both an artist and a craftsman. The former in the sense of Edgar Degas who says "On voit comme on veut voir; c'est faux; et cette fausseté constitue l'art," Science in this art sense is a search for truth. Art is subjective. In the art of book writing, the choice of words and the exposition are our main tools for exploring the truth or maybe perhaps "what we want to see." The craft, on the other hand, has much objectivity to it. Objectivity can be tackled through the tools of mathematics. Both the art and the craft are important. The books are an attempt at finding that balance so that it appeals both to the foxes and the hedgehogs.

Engineering education, with these continuing changes in fundamental understanding and its practice, raises difficult questions of content and delivery too under the constraint of a fixed time period for education. It has also raised serious humanist questions of affordability, even as engineering education claims to aim at frugality through less expensive scaled delivery mechanisms. Engineering, more than science, is beholden to societal needs. In growing fields, particularly the ones that that have the most immediate societal relevance, this rapidly brings content and finances into conflict. As the amount of engineering and science knowledge required rapidly increases, with the rapidly evolving technology, training becomes obsolete just as rapidly. In technical areas, whose educational needs expand suddenly because of their societal use and consequent professional needs, specialized course offerings proliferate rapidly. This puts pressure on the teaching of the foundational knowledge of the disciplines, and the time available for it. The inclusion of broad skill sets into the core curriculum is threatened by the need to teach an expanding number of specialized topics in an ever-shrinking amount of time. The pace

just the Cartesian physical objectivity with an in-built bias in the exclusivity of certain properties but must also expand to domains of other experiences for such a coming together to explain the world. When science uses objective measures of information, it is just that, a measure of that specific information content. The jump to knowledge—a phenomenological objective view—is much more. I would also add that engineering is very much a Martin Heidegger's *Being and Time*, where being-in-the-world is central, and Sartre's progression to existentialism with *Being and Nothingness* as humanism.

Translated, Degas is saying "People see what they want to see; it is false; and this falseness constitutes art."

of and need for change through new offerings or modifications to courses risk introducing disjointedness and decreasing rigor, because modifying and harmonizing a curriculum is a difficult and time-consuming task.

This series of books is an experiment in attempting to answer today's needs in my areas of interest while preserving thoroughness and rigor. It is an attempt at coherent systematic education with discipline, while maintaining reverence and a healthy disrespect for received wisdom.

The books aim to be conceptual not mechanical. This series is aimed ultimately at the electroscience of the nanoscale—the current interest of the semiconductors and devices stream—but which is also far more interdisciplinary than the norms suggest. Its objective is to have students understand electronic devices, in the modern sense of that term, which includes magnetic, optical, mechanical, and informational devices, as well as the implications of the use of such devices. It aims, in four semester-scale courses, to introduce the underlying science, starting with the fundamentals of quantum, statistical and informational mechanics and connecting these to an exposition of classical device physics, then dive deeper into the condensed matter physics of semiconductors, and finally address advanced themes regarding devices of nanometer scale: so, starting with the basics and ending with the integration of electronics, optics, magnetics and mechanics at the nanoscale.

The first book[1] of the series explores the quantum, statistical and information mechanics foundations for understanding semiconductors and the solid state. The second[2] discusses microscale electronic, optical and optoelectronic devices, for which mostly classical interpretation and understanding suffice. The third[3] builds advanced foundations utilizing quantum and causal approaches to explore electrons, phonons and photons and their interaction in the solid state, particularly in semiconductors, as relevant to devices and to the properties of matter used in devices. The fourth book[4] is a treatment of the nanoscale-specific physics of electronic, optical, magnetic and mechanical devices of engineering interest. The second and the third volumes are for subjects that can be taught in parallel but are necessary for the fourth. The value of this approach is that this sequence can be completed by the first year of graduate school or even the senior year of undergraduate studies, for a good student, while leaving room for much else that the student must learn. For those interested in electrosciences, this still includes electromagnetics, deeper understanding of lasers, analog, digital and high frequency circuits, and other directions. The fourth book was the first to come out because of the urgency

A hallmark of the present times is introduction of new words when older ones lose their apparent luster or "branding." "Multidisciplinary" evolved to "interdisciplinary" with an expansion of indiscipline. "Transdisciplinary" must be trying to birth itself. Richard Feynman's statement, "In these days of specialization there are too few people who have such a deep understanding of two departments of our knowledge that they do not make a fools of themselves in one or the other" (from R. P. Feynman, "The meaning of it all: Thoughts of a citizen-scientist," Perseus ISBN 0-7382-0166-9 (1989), p. 9), is not inappropriate here.

[1] S. Tiwari, "Quantum, statistical and information mechanics: A unified introduction," Electroscience 1, Oxford University Press, ISBN 978-0-198-75985-0 (forthcoming).

[2] S. Tiwari, "Device physics: Fundamentals of electronics and optoelectronics," Electroscience 2, Oxford University Press, ISBN 978-0-198-75984-3 (forthcoming).

[3] S. Tiwari, "Semiconductor physics: Principles, theory and nanoscale," Electroscience 3, Oxford University Press, ISBN 978-0-198-75986-7 (2020).

[4] S. Tiwari, "Nanoscale device physics: Science and engineering fundamentals," Electroscience 4, Oxford University Press, ISBN 978-0-198-75987-4 (2017).

I felt. The third book puts together the foundational learning for the modern insights of semiconductors.

I have always admired simplicity of exposition with a thorough discussion that even if simplified, is devoid of propaganda or the much too common modern practice of using templates where depth and nuances are lost and doubts and questions are not addressed. Also consistency is easily lost when modern tools, instead of a pencil and paper, are employed. The style of these books follows these beliefs. Notations, figures, the occasional use of color and other stylistic choices are consistent across the book series.

From early years, I have been a devotee of marginalia—much of the learning and independent thought have come from doodling in the margins and the back pages of notebooks. These books are organized so that the reader will feel encouraged to do so.

A list of very readable, in-depth sources, with my perspectives serving as a trigger for different contents within the book, is to be found at the end of each chapter, in the section titled "Concluding remarks and bibliographic notes." No attempt has been made to credit original discoverers or authors. These remarks and notes ascribe them, or they are to be found by following the references in these notes to their origins.

The exercises are formulated for use in self-study and in the class-room. A subject cannot really be learned by simply reading. Problems requiring application of the information learned and encouraging further thinking and learning are necessary. When we discover for ourselves, we learn best. The exercises here are meant to inform and to be instructive. They are also ranked for difficulty—those that need only a short time but test fundamental understanding are marked as **(S)**, for simple; those requiring considerable effort, bordering on being research problems, are rated **(A)**, for advanced; and those that are intermediate are rated **(M)**, for moderate.

Teaching slides are available on the companion website recorded in the front. The solutions manual may also be requested by providing information at the second link furnished in the front. Slides, when in the modern template-based style, can seriously hinder teaching when they become a tool for filtering key information and explanation while emphasizing summary points. The available presentation material is a tool to avoiding mistakes in writing out equations and to carefully and graphically explain the relationships that science and engineering unfold. They do not substitute for the book and the instructor needs to be diligent in making sure that important themes of teaching—probing, questioning, reasoning, explaining, exploring evidence—come out credibly. I am also happy

The emphasis on probing, questioning, reasoning, explaining and exploring cannot be emphasized enough. I use paradoxes, puzzles, gedanken experiments and real world analogies as common tools. A simple capacitor switched abruptly connected to an ideal source lets one explore dissipation and energy conversion in its broader sense. Displacement currents are real currents, a capacitor as an antenna can radiate, that this radiation proceeding to infinity has a real characteristic impedance, that an infinite L and C transmission line network ends up as a line with real impedance even if made of reactive elements, and that dissipation may arise in the material too, can all be followed through from the poor lowly capacitor. How did the energy appear throughout the capacitor lets one probe Maxwell's equations and electromagnetic propagation even when this current is asymptotically vanishing in a slow charging process. And farther on a tying in of all these connections between fundamental laws and physical behavior entwining Maxwell's electromagnetism, the quantum-mechanical origins of the materials' properties and the diverse meanings of entropy from statistical mechanics. When I tried to introduce a question drawing on the basic understanding of capacitors in a qualifying examination, a fellow member—a *PF9*/PF10 with administration responsibilities—spoke up that we teach our students to never connect ideal voltage sources to capacitors. Science is not religion. Looking for contradictions helps one find the invariants—the physical principles—that stand tall. It is through such probings and mental experiments that one learns and understands. There is nothing more satisfying in education than this peeling of onion from a simple question to deeper and deeper insights. This is what education is about. Curiosity should never be discouraged.

to hear and discuss the subtleties and the different viewpoints of principles, approaches and the deeper meanings of a derived result.

Lots of people can grasp things remarkably quickly. But grasping is not the same as understanding. Understanding is a much deeper network in the brain. I hope students will find in this sequence of books the ballast to propel their own interests through the understanding.

The books could have been shorter and crisper had there been more time. But, what time there was has given enormous pleasure—a time out for integrity in the presence of the incessant pressure of existence, particularly of life in modern academe. For this escape, my gratitude to this world. For making possible the following of my wishes to produce these songs as the shadow of a life in research, teaching and writing, I thank the Hitkarini Foundation.

सर्वजन हिताय । सर्वजन सुखाय ॥

Introduction

Semiconductors, as crystalline, polycrystalline or amorphous inorganic solids, as ordered or disordered organic solids or even in glassy and liquid forms, form a large set of materials useful in active and passive devices. The control of their properties arising in an interaction of particles—atoms, electrons, photons, their elementary one- and many-body excitations, transport and the exchange between different energy forms—has been a fruitful human endeavor since the birth of the transistor, where they found their first large-scale use. Integrated electronics, through its social and commercial informational ubiquity; optoelectronics, through lasers and photovoltaics; and thermoelectronics and magneto-electronics, with their use in energy transformation and signal detection, are but a few of these gainful uses. Nanoscale, within this milieu, opens up a variety of perturbative and significantly more substantial and sensitive effects. Some are very useful, and some can be quite a bother.

Dating back to the 1950s, there exist numerous good textbooks for the solid state. From these early years, J. M. Ziman's *Electrons and phonons* for the details and *Principles of the theory of solids* for a thoughtful broad discussion are particularly of note. Another one is Rudolf Peierls' *Quantum theory of solids.* Among solid-state texts, these remain particularly alive because much of their content is appropriate to electronics of semiconductors. They certainly treat several of the semiconductor-specific scattering and transport topics rather well. As optics—later rebranded as photonics—became important, the divergence in texts increased. Later solid-state texts, with their emphasis on metals, ferroelectricity, ferromagnetism, superconductivity, et cetera, inevitably gave short shrift to semiconductors. That there is a quite informal completeness, consistency and unity to the diversity in the foundations is something that, except for the early books, few capture. Nanoscale makes matters even more divergent. I have felt that a book that

Semiconductor Physics: Principles, Theory and Nanoscale. Sandip Tiwari.
 DOI: 10.1093/oso/9780198759867.001.0001

brings together this unity and focuses on the foundations toward understanding why semiconductor matter behaves the way it does would be useful.

From an engineering perspective, and from that of science, information as the fountain from which much can be understood and explored, including the quantum-mechanical notions through the Bayesian interpretations, is a major change in our learning of recent times.

The Fermi surface of a metal, which gets much attention in a solid-state text, is of enormous import, but it is more of an anachronistic appendage to semiconductor matters. The Fermi surface in a semiconductor, while important, is not as complicated. But there are many static and dynamic interactions, transformations and fluctuation effects that have enormous import and need emphasis. Included within this group are the topics of noise and dissipation as consequences of fluctuations, linear response and causality appearing in Kramers-Kronig-type relationships in multitudes of places beyond just the dielectric function, collective effects and interactions such as those of plasmons or polaritons, strain, semiconductor alloys, the nature of heterostructures and their periodic structures, of defects and multiparticle Auger interactions and of nonlinearities in energy coupling, such as those embodied in Onsager relationships, and even transport from classical to mesoscopic in off-equilibrium conditions.

Add to this collection of topics the consequences of nanoscale from surface to bulk, dimensionality change, collective behavior and, together, their effect on various interactions and transformations as additional subjects of modern importance. In teaching these, with the implicit understanding of nanoscale devices as the ultimate goal, one has to resort to a fair collection of diverse classic resources and combine them with one's own thoughts. This makes the task of getting across to the student the necessary physics for understanding devices difficult, with styles, nomenclature, incompleteness and substantive jumps abounding.

In keeping with the spirit of this textbook series, this volume is devoted to semiconductor-specific solid-state physics aimed at students of engineering, particularly electronics and materials science, but also with utility, because of the exposition, for those from physics and chemistry.

It is organized to certainly include the classical underpinnings ranging from bandstructure approaches to phonon behavior, scattering, approximations, et cetera, but it particularly stresses topics that are modern and aimed toward nanoscale. All are presented with principles and theory as areas of emphasis,

Not to belabor the point, but the Copenhagen interpretation—a duality—and the emphasis on an observation and therefore, secondarily, an observer both are causes of why folks see a hint of spookiness in quantum mechanics. Philosophically, I subscribe to the notion of deep truths—truths where a statement and its opposite are both true such as with wave and particle, or insistence on one's privacy yet wanting governments to give one strong security in this internet and information-centric age, or carbon as both a source of nature's suffering and a source of joy by and for humans—as well as observation as the action that unveils information. This is consistent with the earlier Bohrian notions but also does away with the trust that it expects, and perhaps the wonder it raises, when first introduced in an undergraduate classroom. Bohr had this inclination toward the complementarity of truth and clarity. He is known to have used the following story often. A young person was sent to another village to listen to a great rabbi. Upon his return, he reported, "The rabbi spoke three times. The first talk was brilliant, clear and simple. I understood everything. The second was even better, deep and subtle. I didn't understand much, but the rabbi understood it all. The third was just superb and unforgettable. I understood nothing and the rabbi himself didn't understand much either." Keeping an information-centric perspective helps do away with quite a bit of the metaphysics that developed over the decades around quantum mechanics.

expecting that review papers and other narrower but deeper treatises will become analyzable, understandable and critiqueable to those prepared from this approach.

The book reviews the essential basics and the tools of the trade first, including the quantum methods for ensembles and their approximations, before moving on to the approaches of bandstructure calculation as well as their limitations, which help us describe the behavior of electrons and phonons in a semiconductor. This serves to then develop the treatment of transport, including within it semi-classical, quantum and mesoscopic approaches under scattering and the limit of no scattering under equilibrium and off-equilibrium conditions. For semiconductors, particularly in newer applications, spin-orbit coupling manifests itself in several places, so care is taken to bring the insights from bandstructures to the interactions for semiconductors at different dimensionality.

This sets the stage for the atypical topics of emphasis of the text. The first of these is the discussion of electrons and phonon behavior at surfaces. This is then reformulated for interfaces. Here, heterostructures—what really happens physically at the boundaries—also appear as an important subject for analysis and discussion. Zinc blende, diamond and wurtzite, encompassing elemental and compound semiconductors, including the nitrides, are explored together in emphasizing the principles. The text also discusses the newer and perhaps presently unconventional semiconductors, such as monolayers, in the final chapter, where we return to the themes of the initial chapters in light of all the learning in-between. All this discussion has electrons and phonons as its center, where defect-catalyzed interactions and their variety of behavior under compositional changes are also important.

Photons, electron-photon interactions and radiative and non-radiative phonon-assisted processes are tackled to bring about the interactions in a broadband of energies, so including Auger processes.

This sets the stage for discussion of the next order of complexity in ensemble interactions. We start with a discussion of causality and response theory, and within it the different places where fluctuation-dissipation and Kramers-Kronig forms appear. Ensemble interactions, also in their coupled forms, such as excitons, polaritons and plasmons, follow next. This discussion of higher order interactions is expanded to the variety of manifestations of dissipative transport. Particularly important here is noise, which is central to the use of semiconductor devices at nanoscale. Another of these next order effects is strain, whose use is now pervasive in semiconductors. Spin again becomes quite central to this discussion through bandstructure, as it does for topological reasons.

The spin-topological connections and their device implications are tackled separately in S. Tiwari, "Nanoscale device physics: Science and engineering fundamentals," Electroscience 4, Oxford University Press, ISBN 978-0-198-75987-4 (2017).

The high permittivity of gate dielectrics often used with semiconductors appears with soft phonons and is an essential part of the tool set of semiconductor devices. We look at their behavior and the local and remote coupling effects arising in them.

Energy couplings and their transfer between various forms—heat to electric, and stress to electric—their off-equilibrium behavior and the role of Onsager relationships in these energy transformations is an essential set of topics in important areas of use of semiconductors, from thermoelectrics to piezoelectrics. These are discussed in sufficient detail for the reader to get good insight into the operating principles and how many of the effects undergo some change—sometimes small, sometimes large—at nanoscale.

Quantum and statistical mechanics treatment at the level of S. Tiwari, "Quantum, statistical and information mechanics: A unified introduction," Electroscience 1, Oxford University Press, ISBN 978-0-198-75985-0 (forthcoming), is expected. The reader will find the integrated treatment of information mechanics within this description quite useful because of the common themes that tie energy, entropy and information together, as well as the dominant usage of semiconductors in the processing of information.

We follow this broad swath of physics discussion by looking at periodic structures and the nature of the various excitations of interest in them. So, we discuss superlattices for electrons, phonons, plasmons and plasmon-polaritons, as well as the role of dimensionality within them.

The text intends to provide the reader with an in-depth discussion of semiconductors, aiming toward the nanoscale through this range of development of the subject. Readers who have had an introductory course in quantum and statistical approaches and have a general understanding of the operation of electronic and optical devices will benefit.

An understanding of the operation of simple devices—p/n junctions, and unipolar and bipolar transistors—helps with understanding the relationship between the operational physics, such as that of high permittivity insulators, of noise, of strain or of heterostructures and the behavior of devices.

At the very least, readers must have internalized the meaning of equations at the end of the glossary and the principles of quantum and statistical mechanics and should be willing to pursue the appendices that sometimes serve as summary introductions of important ideas being employed in the main chapters.

The content here is quite comprehensive, as it tries to integrate a variety of ideas across the significant breadth of phenomena that one needs to understand in semiconductors. It is likely that, for some, it is more than can be tackled in a single semester, especially if the students have diverse educational backgrounds and disciplines. So, choices may need to be made. Mine have been to maintain balance between taste and the students' needs. And these have changed from year to year. This book, and this series, represent an attempt at a style where learning is also possible on one's own. The classroom is particularly useful in bringing about the connection of ideas, the emphasizing of principles, the creation of interesting segues where new thoughts can be explored, using the learning and the give and take that the classroom provides so well, and the stimulation of the students' spirit for adventure. The first two chapters of this text are an attempt at summarily introducing and reviewing major ideas, some part of the orthodoxy, but others,

such as Fisher entropy and information, not. This content is the part of the book that one does have the freedom of referring back to if one starts with Chapter 3. Good books and teaching are like music, where the beauty and joy comes from the constant returning back to powerful ideas with variations, each with a little different way of looking at the subject—a different rhythm, a different harmony, a different timbre. Each class is then a different piece of music. The first two chapters facilitate this for the rest of the text so that the learning can be fulfilling to both the student and the teacher. This comment holds just as well for the appendices, where important notions are summarily emphasized. The rest of the book can then be managed in a reasonably demanding course where the students are expected to stay abreast with their reading and thinking.

My favorites for this returning back again and again in music is Verdi's *La forza del destino* as a simultaneous multipath arrival, but many of Schubert's and Chopin's piano—single instrument—pieces bring the path to the heart and the mind perhaps even more convincingly. The Giuseppe Verdi creation is itself a variation on a play by the Spanish master Angel Perz de Saavedra. Verdi endowed his beautiful home—Casa Verdi—in Milan as a retirement home for musicians who need such support late in their life. Music lives there in love and in peace. Verdi's is a life whose variations continue to live a hundred-plus years later. Variations are the most powerful, whether in music or in books and teaching, when they play out forever in time and space.

1
Hamiltonians and solution techniques

Nature is composed of objects—particles, solids or other assemblages in various representations—whose behavior—properties, evolution in time, consequences of stimulation and others—we attempt to explore, understand, design and predict in science and engineering. A major success of classical mechanics from the mid-17th century on was the ability to mathematically describe the evolution in time of the objective values of properties of interest. For example, if a system—a bounded object—of known spatial coordinates and velocity (or momentum), that is, one whose "state" was known, was stimulated under the action of a force, one could predict the future values in Euclidean space. Take this same mathematical construction—usually a set of differential equations—and one could build bigger and smaller objects and predict their evolution by changing the parameters of this differential construction.

Post-Copernicus—the importance of observation and prediction, and mathematical tools such as calculus—is the age of modern science. Prior to that, Euclidean geometry had been the dogma since about 300 *BC*. Likewise, dating from the same time period, Aristotle's views that there are four elements, that heavier objects fall faster or that Earth is the center of universe made up the dogma that was not to be questioned. The former was overthrown by Nikolai Lobatchevski, and the latter needed the Renaissance. This is almost two millennia of scientific darkness!

A space—the state space—could be described with the object at some location in it for each moment of time. The dynamic system's change of state could be described through the equations of motion using either the Lagrangian or the Hamiltonian function once an initial state had been described through the complete specification of the initial dynamic variables together with that of the action on the system. The complete quantitative description of a set of simultaneously measurable parameters—position and velocity (or momentum)—is an essential requirement of this classical determinism. At the quantum scale, simultaneous precise measurement of parameters such as position and momentum is not possible. *The state of the system is not characterized by a set of dynamic variables with specific values. Instead, the state is characterized by a statefunction.* The statefunction is composed of a set of chosen variables—the canonic variables—and the time dependence of this statefunction describes the dynamics of the system. This function of

Semiconductor Physics: Principles, Theory and Nanoscale. Sandip Tiwari.
 DOI: 10.1093/oso/9780198759867.001.0001

time is our wavefunction of the quantum system. It has properties similar to those of waves but it also describes the state of the quantum object. The statistical/probabilistic nature of quantum is within this statefunction. Even for this quantum scale, one can write the Hamiltonian and the Lagrange functions by employing operators that correspond to an observable property—position, energy, momentum and others—and thus describe the evolution of the state.

What is extremely powerful in this approach of the Hamiltonian and Lagrangian functions is that these express the behavior of physical systems irrespective of whether they need to be treated classically or—in more depth—quantum-mechanically. They represent the principle of conservation of energy and the principle of least action as complementary articulations of nature's precept. Between the classical and quantum view, where the quantum view reduces to the classical in the limits, it is the quantum Heisenberg uncertainty, the quantum de Broglie wave-particle duality, the energetics of the interaction and the statistics of quantum to classical that make the enormous change we see in the real world happen. Atoms are stable—neutral—and an electric or magnetic field will have no effect were atoms just to be thought of as classical particles. Place them together in a solid and they form metals, semiconductors and insulators with a variety of seemingly magical properties that are quite different from those of the atom. Conduction and insulation both arise from the properties of the state. An electron in a propagating state leads to conduction. An electron in a bounded state leads to insulation. Largely unoccupied and largely occupied bands of states can both conduct, and conduction can be modulated! Tunneling can happen at microscopic scale. Control can be exercised at small energies—of the order of eVs of visible and infrared light, and therefore at useful bias voltages of a V in semiconductors—instead of Rydberg energies. And further modifications to properties become possible by making structures the size of an electron's wavelength. Quantum mechanics predicts this diverse complexity of solids, and specifically of semiconductors, and thereby makes it possible to use them judiciously.

For reasons of symmetry, explorations in quantum mechanics prefer the Hamiltonian methodology. If you can write the Hamiltonian, you have described the system and its evolution.

That writing the Hamiltonian suffices is written facetiously, of course. In theory, what one has to do is to use this Hamiltonian function in the classical approach, and the Hamiltonian operator in the quantum approach, to now solve the Hamiltonian equation describing the problem. In practice, only the simplest cases can be solved precisely. The rest need approximations.

A solid is a collection of particles—atoms and electrons as their simplest form in ordered or disordered arrangement—undergoing perturbations due to external stimulus because they exist at a temperature T connected to the rest of the universe as a reservoir with which it exchanges energy and particles. To understand

semiconductors, therefore, one needs to be able to describe the interactive evolution of the system. An atom is the simplest, a molecule another level up and then there is the larger ensemble of the semiconductor solid involving an Avogadro-scale number of these particles. So. how does one describe an assembly of particles is an important question to start with.

When making predictions of what a collection of particles will do, classical analysis will resort to either the Lagrangian or the Hamiltonian formalism. These approaches employ a space—the phase space—whose spatial points each represent a specific arrangement of the individual particles. The space is built of the two canonic coordinates over the N dimension of the N particles. The evolution of this point follows from a mathematical operation on a single function leading to a description of the dynamic behavior. This function is the Lagrangian or the Hamiltonian. The Lagrangian employs a general position coordinate $\{q\} = q_1, \ldots, q_N$—not necessarily a set of Cartesian positions, whose choice is determined by convenience, and bundles it with a generalized velocity $\{\dot{q}\} = \dot{q}_1, \ldots, \dot{q}_N$—again, not necessarily a Cartesian velocity. The equation of motion of the system then follows from the Euler-Lagrange equations:

$$\frac{d}{dt}\frac{\partial \mathscr{L}}{\partial \dot{q}_i} - \frac{\partial \mathscr{L}}{\partial q_i} = 0 \;\; \forall\, i = 1, \ldots, N. \tag{1.1}$$

The Lagrange function $\mathscr{L} = T - V$ as a difference of the kinetic (T) and the potential energy (V) due to all sources—internal and external—captures the entire behavior of the particle set. The equations state the principle of least or stationary action. In the configuration space, the evolution between two fixed end points occurs when the action, which is the integral of the Lagrange function along the line connecting the two end points, is a minimum. This is akin to the more easily visualized picture that a function $f(x)$ is a minimum, a maximum or a saddle point if $df/dx = 0$. It is a stationary point: a minimum, a maximum or a saddle point in higher dimensions.

What we wrote here is the Lagrangian for classical mechanics. Lagrangians can be written for all the variety of physical phenomena. As shown in Table 1.1, what they have in common is a square gradient term, of energy, although there may not be an explicit energy connection. In the following chapter (Chapter 2), one sees the informational link to the observation of the physical phenomena—the data that contains the information and therefore the entropy and energy connection—that this represents.

This nomenclature of configuration and phase space arose in explorations of approaches for dynamic systems, where mechanical systems were the first ones of interest from the 18th century onwards. In phase space, every possible state of the system defined by the values that the parameters take defines a point and its evolution a trajectory. The position and the momentum are the variables of the phase space. One could also describe this evolution through position and velocity. Multiple particles will have a multidimensional space. The configuration of the system is writable in generalized coordinates. The vector space defined by these coordinates is the configuration space of the physical system.

Throughout this text, in order to limit the unwieldiness of an equation, we may employ the prime mark to indicate a derivative with respect to spatial coordinates, and a dot above for a derivative with respect to time. So, $q' \equiv dq/dz$ when using a prime, and $\dot{q} \equiv dq/dt$ when using a dot.

Lagrangians have been employed fruitfully in genetics, macroeconomics, machine learning and various other places where new formulations of energy that are not kinetic or potential but where one can see an intuitive energy interpretation. Lagrangian's beauty includes that it teaches us symmetries, conservation laws and other properties in addition to the equation of motion of the dynamical system.

System	Lagrangian $\mathscr{L}$	Comments
Classical mechanics	$(1/2)m\left(\partial q/\partial t\right)^2 - V$	Kinetic minus potential energy
Compressible fluid	$(1/2)\rho\left[\left(\partial q/\partial t\right)^2 - v^2\nabla^2 q\right]$	ρ: density, v: flow velocity
Diffusion	$-\nabla^2_{\mathbf{r}}\psi - O$	ψ: a concentration, O: other terms
Schrödinger's equation	$-(\hbar^2/2m)\nabla^2_{\mathbf{r}}\psi - O$	ψ: statefunction
Elastic wave equation	$(1/2)\rho\left(\partial^2 q/\partial t^2\right) - O$	ρ: density
Helmholtz equation	$\nabla^2_{\mathbf{r}}\psi$	ψ: a field
Lorentz transformation	$\left(\partial_i q_n\right)^2$	Integral invariance

Table 1.1: Lagrangians of some common physical situations. These are all energies, and their integral is the "action."

The Hamiltonian picture is more symmetrical, represents an identical description but, being symmetric, provides a more convenient method for analysis, particularly in quantum mechanics. The generalized position coordinate set $\{q\}$ is taken together with generalized momentum $\{p\} = p_1, \ldots, p_N$. Again, these are not necessarily the Cartesian positions or linear momenta. $p_i = \partial\mathscr{L}/\partial\dot{q}_i$. The symmetry of the Hamiltonian evolution appears in the form

$$\dot{p}_i = \frac{dp_i}{dt} = -\frac{\partial\mathscr{H}}{\partial q_i}, \text{ and } \dot{q}_i = \frac{dq_i}{dt} = \frac{\partial\mathscr{H}}{\partial p_i} \quad \forall\, i = 1, \ldots, N. \tag{1.2}$$

From this "canonical" form, one can also read that

$$\frac{d\mathscr{H}}{dt} = 0, \tag{1.3}$$

which is the law of conservation of energy.

The equivalence of the two approaches can be noted through the ability to determine the canonical conjugates p_i or $\dot{q}_i$ of the two approaches equivalently. The Hamiltonian may be found from the Lagrangian using $\mathscr{H} = \sum_i \dot{q}_i(\partial\mathscr{L}/\partial\dot{q}_i) - \mathscr{L}$. The Hamiltonian function $\mathscr{H} = \mathscr{H}(\{q\};\{p\})$ describes the total energy of the system. Its quantum-mechanical operator is $\hat{\mathscr{H}}$.

1.1 *Hamiltonian*

BEING THE TOTAL ENERGY OF THE SYSTEM, in classical mechanics, the Hamiltonian $\mathscr{H}$ is the sum of the kinetic energy T and the potential energy V of all the particles of the system; that is, $\mathscr{H} = T + V$. In quantum mechanics, $\hat{\mathscr{H}}$ is an operator, corresponding to the observation of energy of the system; operating on the wavefunction describing the system results in the total energy $E = T + V$. For every observable, one can write an operator that, upon operating on the wavefunction describing the system, leads to the observable.

More subtle and therefore more consequential is that one may define the Hamiltonian more generally and hence more powerfully than as just the sum written here.

Operators may be developed from an observable's functional form but require some subtlety so that symmetry and antisymmetry consequences are properly accounted for.

See S. Tiwari, "Quantum, statistical and information mechanics: A unified introduction," Electroscience 1, Oxford University Press, ISBN 978-0-198-75985-0 (forthcoming) for remarks on finding operators corresponding to an observable.

The Schrödinger equation,

$$-\frac{\hbar}{i}\frac{\partial}{\partial t}\psi = \hat{\mathscr{H}}\psi = E\psi, \tag{1.4}$$

incorporating both the time-dependent and the time-independent parts, describes the wavefunction of the system. A system is stationary if $\langle\psi\left|\hat{\mathscr{H}}\right|\psi\rangle$ is invariant in time. Its expected energy and the probability distribution $\langle\psi|\psi\rangle$ are constants of time. For any expectation value of any observable, say A, the expectation value ($\langle A\rangle$) follows:

A statistical—probabilistic—interpretation of stationarity—of various orders—will appear in our discussion of noise in Chapter 16. See Appendix A also if you want to look ahead.

$$-\frac{\hbar}{i}\frac{d}{dt}\langle\hat{A}\rangle = \langle[\hat{A},\hat{\mathscr{H}}]\rangle - \frac{\hbar}{i}\langle\frac{\partial}{\partial t}\hat{A}\rangle\Big|_{\mathscr{H}}. \tag{1.5}$$

For a particle of mass m in a potential V, the Hamiltonian is

$$\hat{\mathscr{H}} = -\frac{\hbar^2}{2m}\nabla^2 + \hat{V}, \tag{1.6}$$

where the first term for kinetic energy follows from the operator for momentum,

$$\hat{\mathbf{p}} = \frac{\hbar}{i}\nabla. \tag{1.7}$$

because kinetic energy $T = \mathbf{p}^2/2m$. Table 1.2 summarizes some of the Hamiltonian operators of interest in common and simple systems.

Since our interest in this text is in understanding the various ways that interactions occur in semiconductors, and the manifestations of these interactions in properties, the important underlying theme is a reasonable understanding of the semiconductor solid itself and therefore the predictive edifice for the collection of electrons and atoms therein. Reasonably accurate solutions of the

System	Hamiltonian $\hat{\mathscr{H}}$	Comments
$1D$ harmonic oscillator	$-\frac{\hbar^2}{2m}\frac{d^2}{dx^2} + \frac{1}{2}kx^2$	k: force constant
Rotation in a plane	$-\frac{\hbar^2}{2I}\frac{d^2}{d\phi^2}$	I: moment of inertia
Rotation on a sphere	$-\frac{\hbar^2}{2I}\Lambda^2$	Λ: Legendrian operator $\Lambda^2 = \frac{\partial^2}{\partial\theta^2} + \frac{\cos\theta}{\sin\theta} + \frac{\partial}{\partial\theta} + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\phi^2}$
Hydrogenic atom	$-\frac{\hbar^2}{2\mu}\nabla^2 - \frac{1}{4\pi\epsilon_0}\frac{Ze^2}{r}$	Z: atomic number μ: reduced mass
Collection of charged particles	$-\sum_i \frac{\hbar^2}{2m_i}\nabla_i^2 + \sum_{i>j}\frac{1}{4\pi\epsilon_0}\frac{z_i z_j e^2}{\lvert\mathbf{r}_i-\mathbf{r}_j\rvert}$	$z_i e$: charge of ith particle
Electric dipole in a field	$-\mathbf{p}\cdot\boldsymbol{\mathcal{E}}$	$\mathbf{p} = ze\langle\mathbf{r}\rangle$: electric dipole moment
Magnetic dipole in a field	$-\mathbf{m}\cdot\mathbf{H}$	$\mathbf{m}$: magnetic moment

Table 1.2: Hamiltonians of some often–encountered situations in semiconductor physics.

Hamiltonian for this collection and its interactions are necessary. And all this will require approximations, for obvious reasons of complexity therein.

We will build this understanding by starting with a discussion of the formulation and then, following some comments on the approach, proceed to its usage in model problems that are instructive. For example, we first look at systems with few electrons and few atoms such as a molecule, or even an atom with its collection of electrons, and then let the number N of particles of this ensemble expand to larger numbers. As we explore, what is important is to understand the reasoning behind the approximations that we make, so that we also know the limits of their validity.

What we are doing is building models based upon our understanding and interpretation of what is most important. Their test of success comprises predictions that come about to be true. Not all of the possibilities can be tested. There will be a range of variations of parameters where our predictions may be trusted with a good model. But it is still a model, an approximation and not the complete reality. So, these are all just different levels of sophistication of "toy models." This is how we should always look at our analytic or algorithmic interpretations.

1.2 *Preliminaries*

CALCULATION OF THE DYNAMICS of a single particle (an electron, for us) requires us to write the Hamiltonian with the potential energy of its interaction and solve the single particle Schrödinger equation

$$\hat{\mathscr{H}}|\psi\rangle = \left(-\frac{\hbar^2}{2m}\nabla^2 + \hat{V}\right)|\psi\rangle = E|\psi\rangle, \tag{1.8}$$

under the constraints of the boundary. In principle, this is straightforward. It may be as simple as the wave solution for $V=0$, or it may be computationally demanding when $V(\mathbf{r})$ takes on odd complexities. When we make this a few particle system, with $V(\mathbf{r}_1, \mathbf{r}_2, \ldots)$ (a potential that is a function of the position of the individual particles), it immediately becomes unwieldy, the dominant reason being the two-body nature of Coulomb interaction in a multi-component many-body problem. Only when particles are non-interacting, that is, when $V(\mathbf{r}_1, \mathbf{r}_2, \ldots) = V(\mathbf{r}_1) + V(\mathbf{r}_2) + \cdots$, does this problem reduce to a straightforwardly solvable form with

$$\hat{\mathscr{H}}(\mathbf{r}_1, \mathbf{r}_2, \ldots) = \mathcal{H}(\mathbf{r}_1) + \mathcal{H}(\mathbf{r}_2) + \cdots = \sum_i \left[-\frac{\hbar^2}{2m_0}\nabla_i^2 + \hat{V}(\mathbf{r}_i)\right], \tag{1.9}$$

a set of independent single particle equations. The wavefunction solution then is $|\psi\rangle = \prod_i |\psi_i\rangle$, where $|\psi_i\rangle$ is the eigenfunction solution of the partitioned Hamiltonian. The energy of the system is just the sum of eigenenergies of these non-interacting particles. *But, this works only if these particles—electrons here—can be treated as being non-interacting, a very rare situation.* If these particles were interacting, each would influence the other, and this picture is invalid, since interactions, even if infinitesimally small, will modify properties.

We will use the hat symbol, $\hat{}$, to identify an operator. Any observable—a physical measurable quantity—is associated with a self-adjoint linear operator. Operators yield the physical value, which is an eigenvalue of the set of possibilities for the system. The wavefunction of the system provides the probability amplitude of finding the system in that state. Pure states have unit norm, so they can be represented by unit-norm vectors. The operators are Hermitian, since operators must yield real eigenvalues, whose probability is in the wavefunction through the square of the amplitude of the orthonormal eigenfunctions. We will be a little loose in writing. Sometimes, an operator hat should be there but may be missing. Sometimes a wavefunction maybe written without the ket symbol $|\rangle$ denoting its vector nature, and sometimes it will. It should be clear from the context.

As an analogy, non-interacting classical gas molecules have very well-defined macroscopic properties, but, microscopically, the motion of each is affected by the collision interactions, even the elastic ones.

The semiconductor solid—the item of interest to us—is a collection of atoms where the atom itself is a collection of particles as a nucleus surrounded by electrons, with the entire solid being charge neutral. With M_i, Ze and $\mathbf{R}_i$ as the nuclei's mass, charge (Z being the atomic number) and position, respectively, and m_0, e and $\mathbf{r}_i$ as the corresponding electron parameters, this solid's Hamiltonian is

$$\hat{\mathscr{H}}_{xtal} = \sum_i \left[-\frac{\hbar^2}{2M_i}\nabla_i^2 \right] + \frac{1}{2}\sum_{i\neq j} \frac{1}{4\pi\epsilon_0}\frac{(Ze)^2}{\mathbf{R}_i - \mathbf{R}_j} + \sum_l \left[-\frac{\hbar^2}{2m_0}\nabla_l^2 \right]$$
$$+ \frac{1}{2}\sum_{i\neq j} \frac{1}{4\pi\epsilon_0}\frac{e^2}{\mathbf{r}_i - \mathbf{r}_j} - \sum_{i\neq j} \frac{1}{4\pi\epsilon_0}\frac{Ze^2}{\mathbf{r}_i - \mathbf{R}_j}. \tag{1.10}$$

This Hamiltonian is made up of energy terms representing the nuclei's kinetic energy (the first term), the potential energy from internuclear Coulomb interactions (the second term), electrons' kinetic energy (the third term), the potential energy from inter-electron Coulomb interactions (the fourth term) and the potential energy contribution of electron-nuclear Coulomb interactions (the fifth term). There are two summations over the kinetic energy, and three summations over the electrostatic interaction.

Since electrons are fermions, the total electronic wavefunction must be antisymmetric whenever the coordinates of two electrons are exchanged (an exchange interaction). The nucleus may be of different species, in which case, they are distinguishable. If they are of the same species, then nuclear spin will also matter. To manage a solution, we have to make judgments on what is important and how to judiciously incorporate it in a manageable calculation, and what is irrelevant, peripheral or a perturbation to be tackled secondarily.

Just a few of the electrons—the valence electrons of the outermost shells—of the semiconductor solid will be important to specific properties of interest to us. The inner ones stay confined with the nucleus, and we may treat them as staying rigidly along with it. We pull these electrons together with the nucleus into an ion—an ion core—and modify the nuclear charge. These inner electrons have now been incorporated into an ion. These ions are massive compared to the electron. This means that the ion motion, and its Coulomb interaction with each other, may also be treated as being small and so can be accounted for as a secondary perturbation if our principle interest is in the electron motion; thus, these terms are eliminated, although the ion motion's perturbation consequence will be included as a later thought. Our problem has now been reduced to solving the electrons' Hamiltonian:

Of course, even the notions of Ze, M_i and m_0 have much complexity buried in them. A remark on this "simple equation" is in order. Paul Dirac, in a 1929 paper in the Proceedings of the Royal Society, says as an introduction, "The general theory of quantum mechanics is now almost complete, the imperfections that still remain being in connection with the exact fitting of the theory with relativity ideas. …The underlying physical laws necessary for the mathematical theory of a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble. It therefore becomes desirable that approximate practical methods should be developed, which can lead to an explanation of the main features of complex atomic systems without too much computation" (P. A. M. Dirac, "Quantum mechanics of many-electron systems," Proceedings of the Royal Society of London, **123**, 714–733 (1929)). Papers and pencils have now been replaced by electronic computers and algorithms. But, nearly a hundred years after this foresighted publication, and nearly forty years after high temperature superconductivity, we don't have an acceptable explanation for the latter crystal phenomena. *There is much subtlety buried away at the low energy end and spread out at the high energy end of the universe.*

In a molecule, the nuclear spin consideration can be important. H and ^3He, for example are fermions, due to their 1/2 nuclear spin, while D, ^4He and H_2 are bosons. Many different properties arise from this difference. This nuclear spin aspect is the least of our worries in discussing a crystal, where the electron and the electron-dressed nucleus considerations will dominate.

In combining inner atomic electrons and the nucleus, we have modified the meaning of Z. It is now the net charge number of the core, not the atomic number.

$$\hat{\mathscr{H}} = \sum_i \left[-\frac{\hbar^2}{2m_i} \nabla_l^2 \right] + \frac{1}{2} \sum_{i \neq j} \frac{1}{4\pi\epsilon_0} \frac{e^2}{\mathbf{r}_i - \mathbf{r}_j} - \frac{1}{2} \sum_{i \neq j} \frac{1}{4\pi\epsilon_0} \frac{Z^* e^2}{\mathbf{r}_i - \mathbf{R}_j}. \tag{1.11}$$

The ion motion can, of course, be very important. The ion motion will cause electrons to scatter—exchange energy and momentum with them—and thus affect transport properties. But we may bring this into our description as a perturbation.

We have now reduced Equation 1.10—a relatively complete description of atom assembly—to Equation 1.11, still a very accurate description where the inter-ion Coulomb interaction and the electron-nucleus Coulombic interaction have been approximated into the third term. The Z^*—ionic charge number—has a new meaning. It is the dressed charge of the nuclei with their surrounding core electrons. And the slow motion of the ion vis-à-vis the electron will let us tackle this term as a perturbation.

The extent of the role of the core of the atom here is as a source of positive charge. If we could further approximate this charge, instead of being localized at $\mathbf{R}_j$, as being uniformly spread out—a continuum—then we have the *jellium model*. This jellium solid, if one also ignores all the quantum-mechanical constraints on the electron, is now just a classical electron particle gas in the solid.

An extension of this jellium and classical discussion is understanding the consequences of dopants; for example, in devices. The classical treatment of dopants in devices is as a continuum; that is, a jellium treatment. A uniform distribution of charge is assumed to arise in them, which in the quasineutral material is balanced by the electron charge cloud. If one makes the device small and have only a few of these dopants, then many of the assumptions underlying the description break down. There are not enough of them to appear as a continuum, and since they do have individual perturbation effects locally, the consequences show up in a small device.

Equation 1.11's second term—a many-body term—is one that requires much attention. Electrons interact with other electrons and have a Coulomb energy associated with that interaction. An electron is also a fermion. And an electron does not interact with itself. The first reflects an electromagnetic force effect. The second is a quantum-mechanical constraint. And the third reflects something much deeper with possibly many interpretations, although it is certainly tied to the first also. *An electron's response arises through its interaction with its surroundings.*

That an electron does not interact with itself can be viewed at many levels. Electric fields are polar vectors. Fields must terminate. There must be a surrounding, and that is how lowering (or even raising, as in single electron effects at nanoscale) of energy happens. An electric field at this level of interpretation arises as $\boldsymbol{\mathcal{E}} = \lim_{q \to 0}(-\nabla_{\mathbf{r}} U_e/q)$. With no charge, there is no energy, and this equation now has a singularity. An electron needs the surrounding for it to be observable through its Coulomb interaction. Richard Feynman's discussion of the singularity and renormalization conundrum is especially powerful as enunciated in his Nobel lecture. It is a question he worried about, starting in his undergraduate years. The lecture can be found at https://www.nobelprize.org/nobel_prizes/physics/laureates/1965/feynman-lecture.html.

The three considerations are tackled by breaking this electron-electron term further. If we only consider classical electrostatic interaction energy, that is, the form $(1/2)(1/4\pi\epsilon_0) \int \rho(\mathbf{r})[\rho(\mathbf{r}')/|\mathbf{r} - \mathbf{r}'|] d\mathbf{r}\, d\mathbf{r}'$ in charge distribution, then we have used what is called a *Hartree energy term*, and a *Hartree equation* form will solve for it. The result has significant errors in it, since the second and third components have not been accounted for. So, we introduce corrections to the Hartree form.

Two electrons, when exchanged, being indistinguishable particles, and fermions, must have different wavefunctions, so the wavefunctions must be antisymmetric. We will see that a Slater determinant antisymmetrizes the many-body wavefunction. This takes care of this *exchange interaction* that has local and nonlocal contributions in it. The *Hartree-Fock equation* will apply this correction for us.

But the antisymmetrization process of the Slater determinant itself also employs one-electron wavefunctions. This requires us to have separability of Equation 1.11, that is, that an electron at any spatial coordinate in space is essentially independent of where the other electrons are. But, a repulsive electron-electron interaction prevents other electrons from approaching the electron at that spatial coordinate. *An electron in the jellium classical solid repels the other electrons electrostatically, thus exposing an equal and opposite positive background so that electric fields vanish far away.* The electrons correlate themselves in a way that screens the electric field. An electron here is surrounded by an equal and opposite charged hole in the electron density. So, there is an "exclusion" zone here. This is a *correlation interaction* representing the physical principle that an electron does not interact with itself and only with all else that surrounds it. The probability of an electron at this position depends very much on the location of the other electrons. Sometimes, this is also referred to as a *correlation hole*, which should not be confused with electron's quasiparticle "hole." This correlation hole is the Coulomb repulsion of other electrons from this electron's vicinity. It is a correlation charge hole. Figure 1.1 is a pictorial representation of such a correlation hole. In Sections 1.5 and 1.6, we will summarize the gradual increasing of accuracy in our search for the solution to this description of the solid. Since Equation 1.11 does not lend itself to easy decoupling, and this starting Equation 1.10 has 3× the total number of nuclei and electrons as its coupled degrees of freedom, we will first employ simple atom assemblies—molecules—to bring out the main physical features of the arguments that we will deploy.

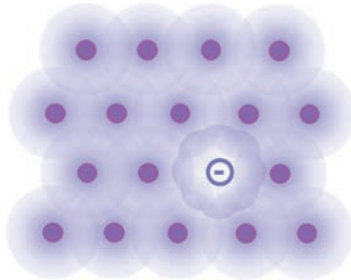


Figure 1.1: An electron in an ensemble of positive charge interacting with other electrons represented by variably filled areas, with atomic nuclei as the background. The lighter area surrounding the electron represents the exclusion zone of the Coulomb repulsive correlation. This is the correlation hole.

This outline of the problem of energetics in the solid shows us the incremental path that we have to take to find satisfactory solutions. If a solution to a problem that is close enough to a new problem is known, we employ perturbation techniques to find the solution to the new problem. This following section summarizes a few of the common perturbation techniques that will be utilized throughout this text, including for this semiconductor crystal outlined so far.

1.3 *Perturbation approaches*

In outlining some of the approaches to solving Hamiltonians, our interest here is in dispensing with how one satisfactorily arrives at the solution to problems that we will encounter in this text. We have to set the problem up right; only then can we

The Church thesis and its Turing machine form are examples of an algorithm for solving a problem by reducing it to a procedure that one steps through. Gamma functions—factorials for integer argument—can be solved for any n since $\Gamma(n+1) = n\Gamma(n)$, and $\Gamma(1) = 1$. This recursion approach is commonplace as a procedure for proof, although sometimes it is applied inappropriately heuristically. Gamma functions, as the integral analytic function, are

$$\Gamma(s) = \int_0^\infty \exp(-x)x^s \frac{dx}{x},$$

where $s = \sigma + it$ is complex, and have many quite amazing properties. They will appear for us during the use of Fermi integrals. Even more magical is the Riemann zeta function $\zeta(s)$, whose integral analytical form is

$$\zeta(s) = \frac{1}{\Gamma(s)} \int_0^\infty \frac{1}{\exp(x) - 1} x^s \frac{dx}{x}.$$

If $\sigma > 1$, this reduces to an infinite series, $\zeta(s) = \sum_{i=1}^{\infty} i^{-s}$. Riemann showed a relationship between the distribution of prime numbers and the non-trivial zeros of the zeta function. So, prime numbers are not randomly distributed, only pseudo-randomly. Other transcendent characteristics of this function include implications for Casimir forces, the cosmological constant, and the lack of Bose-Einstein condensation in two dimensions.

proceed to find the solution. Only under the most circumscribed of conditions in a many-body system may one find accurate direct solutions. Usually though, one has to transform a known related problem that has a known solution to the problem of interest as a perturbation. The Church and Turing forms provide important insight into information mechanics, particularly its deterministic form. The perturbation can be static, that is, time independent and steady state, or dynamic, where a time-dependent perturbation and a quantum system response unfold.

First, we take up the time-independent steady-state perturbation, and then we will take up the time-dependent perturbation. Later on, we will take on the adiabatic time-dependent perturbation, where a state evolves smoothly and continuously, maintaining its quantized identity. An example of time-independent perturbation is when two molecules come close enough, the energy changes. The properties of cohesion/adhesion and repulsion are also examples of the nature of this energetics, a quasi-steady state, where a perturbation causes changes in the energy landscape. Examples of a time-dependent perturbation are a photon exciting an electron, or an electron undergoing scattering during transport. An example of adiabatic perturbation is an electron in a confined quantized space as the size of the space is slowly changed electrically.

The prime number distribution ties it to the Lambert W function ($W(z)\exp[W(z)] = z$) and Ramanujan's series for $\zeta(3)$. Riemann's investigations in geometry inspired Einstein's relativity. A Riemann remark, paraphrased, that the geometry of physical space need not be a God-given Euclidean space but should be determined by experiment, not by hypothesis, stands as one of the most observant statements from a remarkable mathematician and a clergyman's son who, like Ramanujan, died too young. To balance these serious statements, here is a joke that punches at recursiveness. A Russian mathematician and a Russian engineer are visiting a European research institution. Smoking is still accepted. On the first day, when the engineer drops a lighted butt in the trash can sitting on the floor, the paper in it catches fire. So he uses the fire extinguisher to put it out. An identical sequence is repeated with the mathematician, who arrives later after a long night of work. The next day, upon cleaning, the janitor leaves the trash cans on their desks. Events repeat. The engineer uses the fire extinguisher again. The mathematician, however, places the trash can on the floor. *He had reduced it to a known problem.*

1.3.1 *Time-independent perturbation*

LET $\hat{\mathscr{H}}'$ BE A PERTURBATION; that is, let $\hat{\mathscr{H}} = \hat{\mathscr{H}}_0 + \hat{\mathscr{H}}'$ be the Hamiltonian of a quantum system whose statefunction is known in the perturbation's absence. The new statefunction is different. The new perturbation potential causes transitions between the states of the unperturbed Hamiltonian. Let $|u_i^0\rangle$ be the eigenfunctions for the Hamiltonian $\hat{\mathscr{H}}_0$ so that $\hat{\mathscr{H}}_0|u_i^0\rangle = E_i^0|u_i^0\rangle$. For the perturbation $\hat{\mathscr{H}}'$, the eigenenergy solutions exist if $\det|\mathscr{H} - E\mathscr{O}| = 0$. To determine the changes in energy and the eigenfunctions under perturbation, one uses the trick of separation of order where various combinations of terms leading to the same order of effect can be pooled together. Let

$$\hat{\mathscr{H}} = \hat{\mathscr{H}}_0 + \lambda\hat{\mathscr{H}}', \tag{1.12}$$

so that $\lambda = 0$ is the absence of perturbation, and $\lambda = 1$ is the complete turning on of perturbation. The eigenenergies and eigenfunctions of the perturbed system have changed and we will determine these as perturbational changes, by corrections of increasing order, using this λ. Let the eigenenergies and eigenfunctions for the Hamiltonian of Equation 1.12 be

The Hamiltonian equation, even if a simple equation to write summarily, can be quite difficult to solve. We will see this particularly when we dwell on bandstructure calculations where enormous-size matrices are encountered. Sometimes, it is prudent to solve a problem by writing its physical basis in its entirety, as in this equation. Sometimes, it is prudent to start from a known solution and find a solution with it perturbed. The latter is a linear view. The former is a nonlinear view. Each has its successes and failures. I have immense respect for the German organization of education: enough resources at every level, and all the education accessible to everybody. Early schools are equivalent. One doesn't choose a locality to live based on the local schools. Universities are essentially free. Technical professions are respected. A machinist is a precision worker—not a "blue collar"

$$E_n = |E_n^0\rangle + \lambda|E_n^1\rangle + \lambda^2|E_n^2\rangle + \cdots, \text{ and}$$

$$|u_n\rangle = |u_n^0\rangle + \lambda|u_n^1\rangle + \lambda^2|u_n^2\rangle + \cdots. \tag{1.13}$$

The problem posed to us is

$$(\hat{\mathscr{H}}_0 + \lambda\hat{\mathscr{H}}')\left(|u_n^0\rangle + \lambda|u_n^1\rangle + \lambda^2|u_n^2\rangle + \cdots\right)$$
$$= \left(E_n^0 + \lambda E_n^1 + \lambda^2 E_n^2 + \cdots\right)\left(|u_n^0\rangle + \lambda|u_n^1\rangle + \lambda^2|u_n^2\rangle + \cdots\right), \tag{1.14}$$

which can be partitioned in the powers of order λ as

$$\lambda^0 : \hat{\mathscr{H}}_0|u_n^0\rangle = E_n^0|u_n^0\rangle,$$
$$\lambda^1 : \hat{\mathscr{H}}_0|u_n^1\rangle + \hat{\mathscr{H}}'|u_n^0\rangle = E_n^0|u_n^1\rangle + E_n^1|u_n^0\rangle,$$
$$\lambda^2 : \hat{\mathscr{H}}_0|u_n^2\rangle + \hat{\mathscr{H}}'|u_n^1\rangle = E_n^0|u_n^2\rangle + E_n^1|u_n^1\rangle + E_n^2|u_n^0\rangle, \tag{1.15}$$

and so on. With $\lambda = 1$, we now have the posed problem but partitioned into different order corrections. The 0th order equation defines and describes the unperturbed system. The 1st order equation, with the lowest-order correction due to perturbation, has a term due to unperturbed Hamiltonian operating on the 1st order correction to the eigenfunction, and the perturbation Hamiltonian operating on the unperturbed eigenfunction. Both of these terms are of similar order correction. The 2nd order equation has three such terms pooling the same order of correction. The use of λ has let us achieve this deconvolving.

To obtain the first order correction, we use orthonormality by taking the inner product with the bra $\langle u_n^0|$ in Equation 1.15 of the λ^1 power:

$$\langle u_n^0|\hat{\mathscr{H}}_0|u_n^1\rangle + \langle u_n^0|\hat{\mathscr{H}}'|u_n^0\rangle = E_n^0\langle u_n^0|u_n^1\rangle + E_n^1|\langle u_n^0|u_n^0\rangle, \tag{1.16}$$

where $\langle u_n^0|u_n^0\rangle = 1$, $\langle u_n^0|u_n^1\rangle$ is finite and non-zero, and $\langle u_n^0|\hat{\mathscr{H}}_0|u_n^1\rangle = E_n^0\langle u_n^0|u_n^1\rangle$, since $\hat{\mathscr{H}}_0$ is Hermitian. Therefore,

$$E_n^1 = \langle u_n^0|\hat{\mathscr{H}}'|u_n^0\rangle, \text{ and}$$
$$(\hat{\mathscr{H}}_0 - E_n^0)|u_n^1\rangle = -(\hat{\mathscr{H}}' - E_n^1)|u_n^0\rangle. \tag{1.17}$$

The 1st order correction in the eigenenergy arose from the perturbation Hamiltonian and the unperturbed eigenstate. The 1st order correction to the eigenfunction needs a little reworking to write it in terms of the unperturbed orthonormal basis set of $|u_n^0\rangle$, which is a complete orthonormal basis set. We may write

$$|u_n^1\rangle = \sum_{i\neq n} c_i^1|u_i^0\rangle. \tag{1.18}$$

It is useful to exclude the $i = n$ term in the expansion. As $(\hat{\mathscr{H}}_0 - E_n^0)|u_n^0\rangle = 0$, the existence of a term based on $|u_n^0\rangle$ in Equation 1.18 is

with all its subtle prejorativeness—just as the academic is, with the two following different paths of education, one for practice, and one for research and teaching. Both can raise a family with equal opportunity. This becomes possible by the system assuring proper investments across all the stages of education so that enough qualified folks appear at the end for the technical and scientific needs. University may be free, but the student must show that he/she belongs, through effort and examinations. It is not a surprise that there are more car manufacturers in Germany than in the *USA*. And science in Germany has revived exactingly well since the events and tragedy of the Second World War. This setting up the problem right and solving it all together works quite often. But there are situations where it is a luxury and the complexity is such that a local perturbation is desired to get a quick and good-enough answer.

In many situations, the different terms in any order of perturbation are comparable as a product of a large and a small entity. If the perturbation is very small, then one can see it as an operation of a small energy on the starting eigenfunction, and of a large energy operator operating on a small disturbance in the eigenfunction.

dispensable. So, we expand $|u_n^1\rangle$ in the 0th order orthonormal basis set and, to find the lth term of the correction, we take the inner product with the bra $\langle u_l^0|$:

$$\sum_{i\neq n}(E_i^0 - E_n^0)c_i^1\langle u_l^0|u_i^0\rangle = -\langle u_l^0|\hat{\mathscr{H}}'|u_n^0\rangle + E_n^1\langle u_l^0|u_n^0\rangle$$

$$\therefore\ \ (E_l^0 - E_n^0)c_l^1 = -\langle u_l^0|\hat{\mathscr{H}}'|u_n^0\rangle \ \text{ for } \ l \neq n,$$

$$\therefore\ \ c_i^1 = \frac{\langle u_i^0|\hat{\mathscr{H}}'|u_n^0\rangle}{(E_n^0 - E_i^0)} \ \text{ with } \ i = l, \ \text{ and}$$

$$|u_n^1\rangle = \sum_{i\neq n}\frac{\langle u_i^0|\hat{\mathscr{H}}'|u_n^0\rangle}{E_n^0 - E_i^0}|u_i^0\rangle. \tag{1.19}$$

Both the eigenenergy correction (Equation 1.17) and the eigenfunction correction terms (Equation 1.19) are now known for the first order.

This formal approach is obviously extendable to higher orders recursively—more and more terms—and the writing of such an algorithm is quite straightforward. For the λ^2 set of terms,

$$\hat{\mathscr{H}}_0|u_n^2\rangle + \hat{\mathscr{H}}'|u_n^1\rangle = E_n^0|u_n^2\rangle + E_n^1|u_n^1\rangle + E_n^2|u_n^0\rangle; \tag{1.20}$$

therefore,

$$\langle u_n^0|\hat{\mathscr{H}}_0|u_n^2\rangle + \langle u_n^0|\hat{\mathscr{H}}'|u_n^1\rangle = E_n^0\langle u_n^0|u_n^2\rangle + E_n^1\langle u_n^0|u_n^1\rangle + E_n^2\langle u_n^0|u_n^0\rangle. \tag{1.21}$$

Again, since $\hat{\mathscr{H}}_0$ is Hermitian, $\langle u_n^0|\hat{\mathscr{H}}_0|u_n^2\rangle = E_n^0\langle u_n^0|u_n^2\rangle$, we have

$$E_n^2 = \langle u_n^0|\hat{\mathscr{H}}'|u_n^1\rangle - E_n^1\langle u_n^0|u_n^1\rangle, \tag{1.22}$$

and because $\langle u_n^0|u_n^1\rangle = \sum_{i\neq n} c_i^1\langle u_n^0|u_i^0\rangle = 0$, the 2nd order energy correction term is

$$E_n^2 = \langle u_n^0|\hat{\mathscr{H}}'|u_n^1\rangle = \sum_{i\neq n} c_i^1\langle u_n^0|\hat{\mathscr{H}}'|u_i^0\rangle$$

$$= \sum_{i\neq n}\frac{\langle u_i^0|\hat{\mathscr{H}}'|u_n^0\rangle\langle u_n^0|\hat{\mathscr{H}}'|u_i^0\rangle}{(E_n^0 - E_i^0)} = \sum_{i\neq n}\frac{\langle u_i^0|\hat{\mathscr{H}}'|u_n^0\rangle^2}{(E_n^0 - E_i^0)}. \tag{1.23}$$

The 1st order energy correction was the expectation of perturbation on the unperturbed state. The 2nd order energy correction is the perturbation on the 1st order corrected eigenfunction. One can proceed from this and now also build the 2nd order corrected eigenfunction by finding c_i^2.

A very simple consequence of these perturbational relationships is worth thinking through. In a semiconductor, one may represent the states as Bloch states that are propagating states spread out over

The last part of Equation 1.19 is a quantum-mechanical reflection on our classical intuition. Any classical state, under an energy perturbation, that is, the exercise of a force, changes. A mass moves. For any function $y = f(x)$ describing this classical picture, df is the marginal consequence, and df/dx the marginal rate. The lowest order correction for the change is the marginal efficiency of this perturbation. Force and acceleration—the rate change of velocity—are related, with a marginal efficiency determined by the inverse mass. This is an inertial mass. Mass is an emergent property, which superficially can be seen in the bundling of energy. Quantum-mechanically, how well two eigenstates will couple due to perturbation is again this derivative-like ratio of perturbational coupling energy and separation of the unperturbed states. And the statefunction reflects the statistical consequence of these couplings.

the entire semiconductor, and evanescent states that are localized, as at surfaces, defects, impurities, et cetera. The following chapters will expend considerable effort toward this analysis. An $E(n, \mathbf{k}) \equiv E_{n,\mathbf{k}}$ relationship, where n is a quantum number that identifies a band, and $\mathbf{k}$—another quantum number—which is the wavevector, describe the energies of the allowed eigenfunctions. Now suppose we apply a bias voltage V_{dc} (see Figure 1.2) to this semiconductor; that is, there is a net spatially invariant potential energy rise in the system. Do I now have to recalculate the $E(n, \mathbf{k})$? No. Why not? Because

$$E^1_{n,\mathbf{k}} = \langle u^0_{n,\mathbf{k}} | \hat{\mathscr{H}}' | u^0_{n,\mathbf{k}} \rangle = -eV_{dc} \langle u^0_{n,\mathbf{k}} | u^0_{n,\mathbf{k}} \rangle = -eV_{dc} \tag{1.24}$$

since $\hat{\mathscr{H}}_0$ and $-e\hat{V}_{dc}$ commute, that is, $[\hat{\mathscr{H}}_0, -e\hat{V}_{dc}] = 0$, so the same orthonormal basis set may be employed. What about the eigenfunction? Again,

$$|u^1_{n,\mathbf{k}}\rangle = \sum_{i \neq j} \frac{\langle u^0_{n,\mathbf{k}_i} | \hat{\mathscr{H}}' | u^0_{n,\mathbf{k}_j} \rangle}{(E^0_{n,\mathbf{k}_j} - E^0_{n,\mathbf{k}_i})} |u^0_{\mathbf{k}_i}\rangle = 0$$

$$\because \quad \langle u^0_{n,\mathbf{k}_i} | \hat{\mathscr{H}}' | u^0_{n,\mathbf{k}_j} \rangle = -eV_{dc} \langle u^0_{n,\mathbf{k}_i} | u^0_{n,\mathbf{k}_j} \rangle = 0 \;\; \forall\, i \neq j. \tag{1.25}$$

Here, we considered only the interactions within the band n, since the nearest states provide the strongest contribution, but the result is more general because of orthogonality. The implication is that the entire energy bandstructure may be shifted by this potential change and the eigenfunctions do not change. The device analysis is also in dynamic conditions where the time scale of electrodynamic response is much slower—adiabatic, which we discuss a little later—and hence one may again neglect any bandstructure consequence of the electrically applied stimuli's consequences. This underlies all the drawing of the band diagrams, where the conduction and valence bandedge lines represent the extrema of the bands.

There is one complication in the previous argument that we sidestepped but should address. The electron's state, absent magnetic interaction, with $s = 1/2$, and therefore the secondary spin quantum number of $\pm 1/2$, is degenerate in energy. The equation forms in Equations 1.17 and 1.23 actually blow up if one considers the interaction between these degenerate states. This, of course, is unphysical. The series does not converge. The perturbation expansion has this problem with states very close in energy. This can be effectively and efficiently tackled by treating the nearly degenerate states in the same way as we treated $|u_n\rangle$ in the perturbation expansion. This means that the 0th order state is being allowed to be an arbitrary linear combination of the degenerate states.

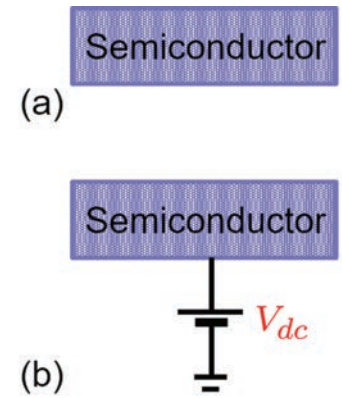


Figure 1.2: (a) A semiconductor whose bandstructure describing electron states has been determined to be $E(n, \mathbf{k}) \equiv E_{n,\mathbf{k}}$ with the semiconductor grounded (thermal equilibrium). (b) This semiconductor shown under the static bias voltage V_{dc}. Perturbation tells us that this just shifts the bandstructure—the eigenenergy solution—by $-eV_{dc}$.

Good thing too. Imagine having to calculate this bandstructure under all potentials and their distribution in real space in a device. There is more to this that we will see in our discussion in Chapter 4 of the parameter we call the effective mass, and its effectiveness.

Let there be N nearly degenerate states. We choose an orthonormal set of basis state in N; that is,

$$\langle\phi_j|\phi_i\rangle = \delta_{ji} \;\; \forall \;\; i,j \in N. \tag{1.26}$$

So, these $|\phi_i\rangle$'s together with their coefficients α_i let us expand the degenerate states into a non-degenerate basis. Now take the 0th and 1st order terms of the Schrödinger/Hamiltonian equation of the problem. Orthonormality lets us write

$$\sum_{i\in N} \langle\phi_j|\hat{\mathscr{H}_0} + \hat{\mathscr{H}'}|\phi_i\rangle\alpha_i = E\alpha_j. \tag{1.27}$$

The number of degenerate states determines the number of solutions to this eigenvalue equation set, writable as

$$\begin{bmatrix} \mathscr{H}_{11} & \cdots & \mathscr{H}_{1N} \\ \vdots & \vdots & \vdots \\ \mathscr{H}_{N1} & \cdots & \mathscr{H}_{NN} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix} = E \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_N \end{bmatrix}, \tag{1.28}$$

where $\mathscr{H}_{ji} = \langle\phi_j|\hat{\mathscr{H}_0} + \hat{\mathscr{H}'}|\phi_i\rangle$ is the matrix element for the complete Hamiltonian. The solution follows from the condition $\det|\mathscr{H}_{ji} - E\mathscr{O}| = 0$. α_i are now known, and using this orthonormal set to replace the degenerate set in the Equation 1.14 tackles the problem in first order perturbation.

In S. Tiwari, "Quantum, statistical and information mechanics: A unified introduction," Electroscience 1, Oxford University Press, ISBN 978-0-198-75985-0 (forthcoming), we have tackled these various perturbation situations: 1st, 2nd, degenerate and more; for example, Rabi oscillations and the evolution of two-level systems when one applies a static perturbation at an instant in time. This discussion is very germane here, and a reading of it is strongly recommended for insight.

1.3.2 *Time-dependent perturbation*

In semiconductor problems, non-steady-state interactions—a shining of light, an electron scattering due to a Coulomb impurity, interface roughness, phonon excitation or even bandedge fluctuations arising from atomic motion due to thermal energy—are ubiquitous. Perturbation with time dependence allows us to view the quantum evolution under these conditions.

We employ a 2-level system (see Figure 1.3), but in a way that is extendable to the full basis set because the time dependence of the evolution still conforms to the same picture. The basis eigenfunction set consists of kets $|m\rangle$ and $|k\rangle$ with eigenenergies E_m and E_k, respectively. The statefunction solution for the unperturbed state is $|\psi\rangle = c_m|m\rangle + c_k|k\rangle$, with $\langle m|k\rangle = 0$, $\hat{\mathscr{H}_0}|m\rangle = E_m|m\rangle$ and $\hat{\mathscr{H}_0}\langle k\rangle = E_k|k\rangle$. With perturbation,

$$\hat{\mathscr{H}} = \hat{\mathscr{H}_0} + \hat{\mathscr{H}'}, \;\; \text{and}$$
$$-\frac{\hbar}{i}\frac{\partial}{\partial t}|\psi\rangle = \hat{\mathscr{H}}|\psi\rangle, \tag{1.29}$$

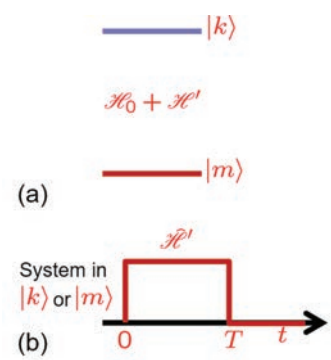


Figure 1.3: (a) A two-level quantum system. (b) This system initially in one or the other eigenstate is subjected to a perturbation for a time duration T. Consequently, it evolves.

describe the evolution of the statefunction. The state function evolves as

$$|\psi(t)\rangle = c_m(t)|m\rangle + c_k(t)|k\rangle, \tag{1.30}$$

which is prescribed by our restriction of the two-level orthonormal basis. If $\hat{\mathscr{H}}' = 0$—an unperturbed stationary state—$c_l(t) = c_l(0)\exp(-iE_l t/\hbar)$, where $l = k, m$ holds for the stationary state.

If a perturbation is applied, using the condition of orthonormality one can view the evolution for the coefficients through

$$-\frac{\hbar}{i}\frac{d}{dt}\begin{bmatrix} c_m(t) \\ c_k(t) \end{bmatrix} = \begin{bmatrix} \mathscr{H}_{mm} & \mathscr{H}'_{mk} \\ \mathscr{H}'_{km} & \mathscr{H}_{kk} \end{bmatrix}\begin{bmatrix} c_m(t) \\ c_k(t) \end{bmatrix}, \tag{1.31}$$

where $\mathscr{H}'_{mk} = \mathscr{H}'_{km}$ are real values of the energy observable. The statefunction $|\psi\rangle$ follows from

$$-\frac{\hbar}{i}\frac{\partial|\psi\rangle}{\partial t} = \hat{\mathscr{H}}|\psi\rangle = E|\psi\rangle, \tag{1.32}$$

that is,

$$\begin{bmatrix} c_m(t) \\ c_k(t) \end{bmatrix} = \exp\left(-i\frac{Et}{\hbar}\right)\begin{bmatrix} c_m(0) \\ c_k(0) \end{bmatrix}, \tag{1.33}$$

where

$$E\begin{bmatrix} c_m(t) \\ c_k(t) \end{bmatrix} = \begin{bmatrix} \mathscr{H}_{mm} & \mathscr{H}'_{mk} \\ \mathscr{H}'_{km} & \mathscr{H}_{kk} \end{bmatrix}\begin{bmatrix} c_m(0) \\ c_k(0) \end{bmatrix}. \tag{1.34}$$

Let $E = E_-, E_+$ be the perturbed eigenenergies; then, this two-level system, under this static perturbation turned on at $t = 0$, has the eigenfunction solutions

$$\begin{aligned} |\psi_-(t)\rangle &= \left[c_m^-(0)|m\rangle + c_k^-(0)|k\rangle\right]\exp\left(-i\frac{E_-t}{\hbar}\right) \\ &= (\cos\theta|m\rangle + \sin\theta|k\rangle)\exp\left(-i\frac{E_-t}{\hbar}\right), \text{ and} \\ |\psi_+(t)\rangle &= \left[c_m^+(0)|m\rangle + c_k^+(0)|k\rangle\right]\exp\left(-i\frac{E_+t}{\hbar}\right) \\ &= (-\sin\theta|m\rangle + \cos\theta|k\rangle)\exp\left(-i\frac{E_+t}{\hbar}\right). \end{aligned} \tag{1.35}$$

Here, the alternative set of equations are written with $c_m^-(0) = \cos\theta$, $c_k^-(0) = \sin\theta$, $c_m^+(0) = -\sin\theta$ and $c_k^+(0) = \cos\theta$. This maintains orthonormality and establishes a starting phase. If $\theta = 0$, then the system has been prepared in $|m\rangle$ before the turning on of the perturbation. The statefunction solution with perturbation is

$$\begin{aligned} |\psi(t)\rangle &= d_-|\psi_-(t)\rangle + d_+|\psi_+(t)\rangle \\ &= d_-|\psi_-(0)\rangle\exp\left(-i\frac{E_-t}{\hbar}\right) + d_+|\psi_+(0)\rangle\exp\left(-i\frac{E_+t}{\hbar}\right). \end{aligned} \tag{1.36}$$

Choosing the amplitudes $d_- = \cos\theta$ and $d_+ = \sin\theta$ leads to

$$\begin{aligned}|\psi(t)\rangle &= \left[\cos^2\theta \exp\left(-i\frac{E_-t}{\hbar}\right) + \sin^2\theta \exp\left(+i\frac{E_+t}{\hbar}\right)\right]|m\rangle \\ &\quad + \sin\theta\cos\theta\left[\exp\left(-i\frac{E_-t}{\hbar}\right) - \exp\left(+i\frac{E_+t}{\hbar}\right)\right]|l\rangle \\ &= c_m(t)|m\rangle + c_l(t)|l\rangle. \end{aligned} \tag{1.37}$$

This is an oscillatory, not stationary, solution. If the system were initialized in $|m\rangle$, the probability of the system being found in $|k\rangle$ at time t in the presence of this steady-state perturbation would be

$$\begin{aligned}|c_k(t)|^2 &= \sin^2(2\theta)\sin^2\left[\frac{(E_+ - E_-)t}{2\hbar}\right] \\ &= \frac{4\left|\mathscr{H}'_{mk}\right|^2}{(\mathscr{H}_{kk} - \mathscr{H}_{mm})^2 + 4\left|\mathscr{H}'_{mk}\right|^2}\sin^2\left(\frac{\Omega t}{2}\right), \end{aligned} \tag{1.38}$$

with

$$\Omega = \left[\left(\frac{\mathscr{H}_{kk} - \mathscr{H}_{mm}}{\hbar}\right)^2 + \frac{4\left|\mathscr{H}'_{mk}\right|^2}{\hbar^2}\right]^{1/2} \tag{1.39}$$

as the oscillation frequency—the Rabi frequency (Figure 1.4). The two-level system would have stayed in the prepared state absent perturbation. With static perturbation, it now oscillates at the slower frequency of $\Omega = (E_+ - E_-)/2\hbar$ determined by the eigenenergies of the statefunction under perturbation. The perturbation energy determines the cycling depending on the magnitude of coupling.

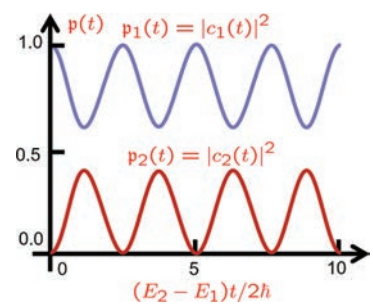


Figure 1.4: Rabi oscillation in a 2-level system under a perturbation.

What if the perturbation was for a short time duration T—a scattering event—as in Figure 1.3? We make this a harmonic perturbation. Our stimulus to this 2-level system is

$$\begin{aligned}\mathscr{H}'_{mk}(t) &= 0 \;\text{ for }\; t \le 0, \\ \mathscr{H}'_{mk}(t) &= 2\mathscr{H}'_{mk}\sin(\omega t), \;\text{ that is,} \\ &= i\mathscr{H}'_{mk}\left[\exp(-i\omega t) - \exp(i\omega t)\right] \;\text{ for }\; t > 0, \end{aligned} \tag{1.40}$$

where we will make this duration finite while looking at the solution of the evolution. The time dependence follows from Equation 1.31.

Take the case of a system starting in the eigenfunction state $|m\rangle$. So, where $c_m(0) = 1$ and $c_k(0) = 0$, Equation 1.31 states (note $\mathscr{H}'_{km} = \mathscr{H}'_{mk}$ because the operator is Hermitian)

$$-\frac{\hbar}{i}\frac{d}{dt}c_k(t) = \mathscr{H}'_{km}c_m(0) + \mathscr{H}_{kk}c_k(0) = \mathscr{H}'_{mk}c_m(0) + \mathscr{H}_{kk}c_k(0), \tag{1.41}$$

leading to

$$-\frac{\hbar}{i}c_k(t) = \int_0^t \mathscr{H}'_{mk}(\tau)\exp(i\omega_{km}\tau)\,d\tau, \tag{1.42}$$

with the oscillation time dependence of the perturbation explicitly included through the time variable τ in the integral spanning $t = 0$ to $t = t$. Also, remember the phase terms in the basis set. The solution is

$$\begin{aligned} c_k(t) &= \frac{\mathscr{H}'_{mk}}{\hbar}\int_0^t \left\{\exp\left[i(\omega_{km}-\omega)\tau\right] - \exp\left[i(\omega_{km}+\omega)\tau\right]\right\} d\tau \\ &= \frac{\mathscr{H}'_{mk}}{\hbar}\left\{\frac{\exp\left[i(\omega_{km}-\omega)t\right]-1}{i(\omega_{km}-\omega)}\right. \\ &\quad \left. -\frac{\exp\left[i(\omega_{km}+\omega)t\right]-1}{i(\omega_{km}+\omega)}\right\}. \end{aligned} \tag{1.43}$$

Here, $\omega_{km} = (E_k - E_m)/\hbar$. The largest contribution to amplitude evolution in time comes from where the denominator is the smallest. The perturbation frequency closest to the interlevel frequency has the largest effect. If $\omega_{km} \approx \omega$, the first term will dominate. Since we started with the lower level filled, the amplitude of the higher level is

$$\begin{aligned} c_k(t) &= \frac{2\mathscr{H}'_{mk}}{\hbar}\frac{\sin\left[(\omega_{km}-\omega)t/2\right]}{(\omega_{km}-\omega)}\exp\left[i(\omega_{km}-\omega)\frac{t}{2}\right] \\ \therefore\quad |c_k(t)|^2 &= \frac{4\left|\mathscr{H}'_{mk}\right|^2}{\hbar^2}\frac{\sin^2\left[(\omega_{km}-\omega)t/2\right]}{(\omega_{km}-\omega)^2}. \end{aligned} \tag{1.44}$$

This is the lowest-order correction for the probability of finding the system in $|k\rangle$ at time $t = T$, and let us say that we remove the perturbation at that point, leaving the system in that stationary state, is

$$|c_k(t)|^2 = \frac{4\left|\mathscr{H}'_{mk}\right|^2}{\hbar^2}\frac{\sin^2\left[(\omega_{km}-\omega)T/2\right]}{(\omega_{km}-\omega)^2} \quad \text{for} \quad t \geq T. \tag{1.45}$$

This response has a form that is the square of a sinc function. Figure 1.5 shows the normalized response. It has a peak when the frequencies/energies are precisely matched; that is, $\omega_{km} = \omega$. The peak of the normalized fraction is of magnitude $T^2/4$. The half width of the main peak is $\sim 5.6/T$. The transition probability per unit time $|c_{km}|^2/T$—a scattering rate—which we denote by S, is

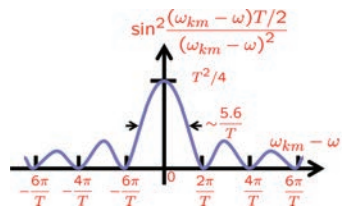


Figure 1.5: A plot of the term $\sin^2\left[(\omega_{km}-\omega)(T/2)\right]/(\omega_{km}-\omega)^2$, which is proportional to the probability of finding the 2-level system in the eigenfunction state $|k\rangle$ following an application of perturbation for a time duration T.

$$S_{mk} = \frac{4\left|\mathscr{H}'_{mk}(0)\right|^2}{\hbar^2}\frac{\sin^2\left[(\omega_{km}-\omega)\frac{T}{2}\right]}{(\omega_{km}-\omega)^2 T}. \tag{1.46}$$

If T is large enough, the sinc function asymptotes to the Dirac delta function, that is, for large-enough T over which the perturbation appears,

$$S_{mk} = \frac{2\pi}{\hbar}\left|\mathscr{H}'_{mk}(0)\right|^2\delta(E_k - E_m) \tag{1.47}$$

for $t \geq T$. The entire complexity of transitions under time-dependent finite time perturbation, under certain constraints, can be reduced

Not necessarily because of Matthew's principle, this relationship of Equation 1.47 is often referred to as Fermi's golden rule. It appeared in Fermi's quantum mechanics lectures. The compact and to-the-point lecture notes from his University of Chicago days—*Notes on quantum mechanics* from the University of Chicago Press is the 1954 version—are very worthwhile reading. Dirac had gotten there twenty years earlier. But the first order and second order perturbations

to a time-normalized transition probability; that is, transition probability per unit time, which has a very simple form. If one knows the perturbation Hamilitonian, and two states between which this interaction's transition rate is to be ascertained, Equation 1.47 ascertains it. This is the Golden rule.

There are a number of interesting features embedded here. The peak is proportional to $T^2/4$, and the full width at half maximum is $\sim 5.6/T$. The area is proportional to T, and the width to time duration's inverse ($1/T$). The uncertainty principle is reflected in this spread. The central peak contains about 90 % of the area. When time is large enough, this function narrows further to the Dirac delta. It has a peak that corresponds to $T^2/4$, so, for matched conditions, only short times are needed. But, uncertainty relationships—embedded in our calculation of this relationship—must still hold and do. In addition, $\Delta E \Delta t \geq \hbar/2$ or $\Delta\omega\Delta T \approx 1$ still applies, as reflected in the central peak's areal argument. So, very short times, for example as in semiconductor-specific energy transition problems, will reduce the probability in state $|k\rangle$. But, as a first order term, this transition probability through the Dirac delta relationship will still be useful. If matching is poor, it may even vanish under certain conditions. If $\omega_{km} - \omega = 2\pi/T$, the transition vanishes, since a full cycle of interaction brings back the system to its original state. For all the problems that we are interested in, the time scales of interactions are large enough that this relationship written in Dirac delta form suffices. The equation can also be extended when there are spreads in frequencies of excitation, or spreads—as in bandstructure—of states of transition. For these,

$$S = \frac{1}{T}\sum_k |c_k(t \geq T)|^2 = \frac{1}{T}\int |c_k(t \geq T)|^2 \mathscr{G}(k)\, dE_k, \tag{1.48}$$

which reduces to

$$S = \frac{2\pi}{\hbar}\left|\mathscr{H}'_{if}(0)\right|^2 \mathscr{G}(E_f)\delta(E_f - E_i). \tag{1.49}$$

From two levels to nearly continuous distribution simply follows as an extension through density of states. And if it is between two different distributions of density of states, then joint density of states will appear.

referred to as Golden rule 1 and Golden rule 2 in Fermi's notes struck a cord, and somewhere along the way his name got associated. We will call it just the Golden rule. It is of the first kind. We extracted it in the lowest order term of change. Enrico Fermi was an exceptional scientist, equally adept at experiment and theory. Fermi had a good sense of humor. In Rome, his Physics Institute was in the same compound as other senior government offices, where nobody worked on Sundays. Fermi was known to drive in wearing a hat and declaring himself a driver of dignitaries so that he could continue his experiments. His Nobel prize for discovering transuranic elements, which he named Ausonium ($Z = 93$) and Hesperium ($Z = 94$), is one of those for wrong reasons. The elements resulting from his slow neutron bombardment experiments were fission products, not heavier. But, like Bethe, who worked with him in Rome, there was plenty of other work for which the Nobel was deserved.

See Appendix B where a number of computationally useful functions that one encounters in a variety of forms are summarized. The Dirac function can be written in a variety of ways and is often a very convenient manipulation tool, as is his bra and ket notation for vectors.

1.3.3 *Scattering by the perturbation*

THE UTILITY OF THIS GOLDEN RULE can be illustrated through Coulomb scattering, as shown in Figure 1.6. Take the electron as a plane wave encountering the Coulomb attraction from a positive charge. This could be in free space, but, for us, this is particularly

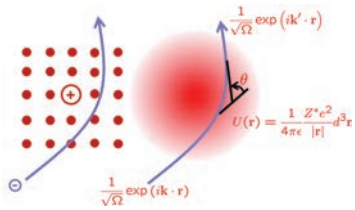


Figure 1.6: Coulomb scattering causing an electron wave to scatter—change momentum (magnitude and direction)—as it travels in a crystal. Coulomb energy is gained and then lost as the scattering takes place with a small net change.

germane due to its correspondence to impurity, which has a charge in the semiconductor, causing a strong scattering for an electron. This electron is a nearly free electron traveling around, and the wavefunction of $\psi = (1/\sqrt{\Omega}) \exp(i\mathbf{k}\cdot\mathbf{r})$ for the spatial component is an adequate representation. Here, Ω represents the volume for normalization of the probability. Due to the charge, the perturbation potential is $U(\mathbf{r}) = (1/4\pi\epsilon)Z^* e^2/|\mathbf{r}|$. $Z^* e$ here is the dressed static charge of the impurity. The electron's wavevector changes to $\mathbf{k}'$ as a result of this scattering. Under this perturbation, we have

We will generally use Ω for volume, and V for potential, except in rare situations, to avoid confusion. Ω_0 will be the volume of a unit cell.

$$\begin{aligned}\mathscr{H}'_{\mathbf{kk}'} = \langle \mathbf{k}'|\hat{\mathscr{H}}'|\mathbf{k}\rangle &= \frac{1}{\Omega}\int U(\mathbf{r}) \exp\left[i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}\right] d^3\mathbf{r} \\ &= \frac{1}{\Omega} U_{\mathbf{kk}'},\end{aligned} \tag{1.50}$$

where this last simple notation tells us that it is the Fourier component—in reciprocal space—for the continuous electrostatic potential. Any scattering into a solid angle $d\theta$ depends on the density of states of the available states there. Classically, this is a continuous distribution, as all positions and momenta are possible. Quantum-mechanically, it will depend on the availability of states. Connecting quantum to classical with isotropicity, we may write $dE = v d(\hbar k)$, so the density of states for scattering is

Later on, in Chapter 10, we will see anisotropic consequences as a result of this state argument, since semiconductors in general are not isotropic, neither do they necessarily have a continuous state distribution.

$$d\mathscr{G} = \frac{\Omega}{(2\pi)^3}\, d\theta\, \frac{k^2\, dk}{v d(\hbar k)} = \frac{k^2 \Omega}{8\pi^3 \hbar v}\, d\theta. \tag{1.51}$$

The interaction happens during transit near the impurity, and the scattering or transition rate, writing it both classically through scattering cross-section and quantum-mechanically through the Golden rule, is

$$S_\theta = \frac{d\sigma}{\Omega} v = \frac{v}{\Omega}\frac{d\sigma}{d\theta}\, d\theta = \frac{2\pi}{\hbar}\left|\frac{1}{\Omega} U_{\mathbf{kk}'}\right|^2 \frac{k^2\Omega}{8\pi^3\hbar v}\, d\theta. \tag{1.52}$$

Therefore, with $d\sigma/d\theta$ from this relationship and using the momentum correspondence of $\hbar k = m^* v$, where m^* is an effective mass, we have

$$\frac{d\sigma}{d\theta} = \frac{m^{*2}}{4\pi^2\hbar^4}\frac{k^2}{v^2}\left|U_{\mathbf{kk}'}\right|^2. \tag{1.53}$$

The Fourier component of the Coulomb perturbation is

$$\begin{aligned}\left|U_{\mathbf{kk}'}\right|^2 &= \int \frac{1}{4\pi\epsilon}\frac{Z^* e^2}{|\mathbf{r}|} \exp\left[(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}\right] d^3\mathbf{r} \\ &= \frac{1}{4\pi\epsilon}\frac{Z^* e^2}{|\mathbf{k}-\mathbf{k}'|^2} = \frac{Z^* e^2}{4k^2\sin^2(\theta/2)}.\end{aligned} \tag{1.54}$$

This last equation now gives a direct correspondence between the classical and quantum-mechanical pictures of a nearly free electron

scattering due to a charge in an isotropic condition. We could use the Golden rule to determine what this scattering rate will be in the angle $d\theta$. A classical scattering cross-section can be fitted to it. And, by correspondence, the quantum scattering will gracefully transform into a classical scattering relationship.

This discussion of states and their transformation due to interactions sets up a reasonable starting point for discussing the nature of the electrons and the atom systems and their analysis.

Many of the parameters—nearly constant—that one often utilizes in the physical world are the result of fast and slow at work. The friction coefficient—static and dynamic—arises in the electromagnetic and quantum-constrained interactions in the interface region between two objects. The couplings under static and dynamic conditions are different, but both are to a broadband of vibrational losses for the atoms of the objects. An electron undergoes fast scattering events, many of which are random, as it moves through the matter. This too is a broadband event in the frequency domain, with energy loss to the environment. Thermal equilibrium or steady state comes about because of the accumulation of the fast-and-slow events.

1.4 *Fast and slow, and the Born-Oppenheimer/adiabatic approximation*

SIMULTANEOUS PRESENCE OF FAST AND SLOW interactions is quite commonplace. A simple example of this is the variety of circumstances where one introduces a frictional damping. Brownian motion, conductance, et cetera, are illustrations of fast events—in these cases, random scattering encounters of a particle or an electron in a solid—in the presence of a slow external stimulus—the flow of the liquid and the particles, or of electrons—under an external cause of potential or kinetic energy change. For semiconductors, an illustration of this complexity is fast-moving electrons in the midst of the vibration of atoms around their equilibrium positions. The atomic motion—dressed nuclei, that is, nuclei and the core electrons vibrating around their equilibrium—is slow, since the mass is large. An electron transits a few atom distance ($\sim nm$) at a speed of $\sim 10^7$ cm/s in ~ 10 fs. Atomic vibration—a deformation, where the frequencies are of the order of a few THz—must lead to a change in the allowed states of the electron. So, if there is a scattering interaction between the electron and the perturbation due to this deformation, there will be the fast scattering event coupling to a slower deformation. The Golden rule lets us tackle the fast through the transition/scattering rate. The slow will follow for us from adiabatic approximation, which has its origins in the Born-Oppenheimer approximation in quantum mechanics' earliest application to the study of molecules. We will be particularly interested in this simultaneous presence of fast and slow processes because of its importance to transport and transitions. And we can then suitably put the two approaches together to understand fast and slow.

Adiabatic, a word of Greek origin, translates as "not to be passed through."

The *adiabatic process*, or *adiabatic approximation*, is an important analysis tool in quantum conditions. Both of these terms are also used in classical conditions but with subtle differences in meaning that need some elaboration. In classical mechanics, the adiabatic

In electronics, adiabatic circuits, by suppressing entropy production and by recovering energy, can consume vanishingly low energy. But they are slow, so this quantum-classical difference can cause plenty of tangle.

process is a process in which no heat is exchanged. The system can be viewed as one that is thermally isolated or in which change is taking place rapidly enough that transfer of energy as heat is absent. In the thermodynamic view, this lets one analyze conditions of rapid change—such as compression or expansion in a mechanical engine—and determine limits such as the Carnot efficiency. Absence of heat exchange means no entropy change. In the quantum view, adiabatic approximation implies a change that is sufficiently slow so that the eigenfunction evolves slowly from one to another, slowly enough that it is a tight coupling between one state to another state; that is, it remains reversible throughout the process. The state of the system remains the eigenstate of the instantaneous Hamiltonian. It is in this sense of reversibility that it corresponds to the classical use. However, in classical adiabatic conditions, the process needs to be rapid to eliminate entropy change; in quantum conditions, it needs to be quasistatic to allow state-to-state coupling. If the latter were not, a starting state would couple to a band of states—the final new state being a superposition in the new system—destroying reversibility.

In S. Tiwari, "Quantum, statistical and information mechanics: A unified introduction," Electroscience 1, Oxford University Press, ISBN 978-0-198-75985-0 (forthcoming), we looked at the reflection and transmission of an electron plane wave whose energy is larger than the barrier energy. An abrupt barrier—sudden change—causes reflection and transmission. A gradually changing barrier, where $\ell \gg 2\pi/k$, with ℓ as a length scale of change of the barrier's changing energy, and k as the wave vector of the incident wave, has the reflection suppressed. The incident wave of wavevector $\mathbf{k}$ will adiabatically adjust to a wavevector $\mathbf{k}'$ in the barrier region.

So, in quantum mechanics, an adiabatic process is a process where a system undergoing change—with modification to its energy levels—continues to remain in a single definite state. It maintains, for example, its quantum numbers. If a square well has a particle in the ith level, it remains there as the square well shape is changed adiabatically. The wavefunction adapts to the slowly changing parameters that define and mold the system.

The Born-Oppenheimer approximation is an example of resorting to adiabatic process in molecules and solids. When we reduced Equation 1.10 to Equation 1.11 to describe the atomic solid by decoupling the ionic motion—the slow process—from the electrons'—the fast process—Hamiltonian, we employed the adiabatic approximation. The deformation-induced scattering can now be added on as a perturbation to the solution. Another example is the simple system that we will take up next for analysis—a 2-electron and 2-atom system (Section 1.5)—where both an electronic and a nuclear part appear. Both have motional components. For any change from the initial state to a final state, one must consider both the electron and the nuclear part. The mass of the electron is significantly smaller than that of the nuclei, so nuclei can be viewed as moving so slowly that the electron distribution adjusts to them instantaneously responding to the changing potential. The consequence—a Born-Oppenheimer approximation—is that a fixed electronic wavefunction is calculable for any fixed nuclear locale. This is to say that, in the absence of

An N-particle system has $3N$ degrees of motional freedom; 3 will be translational, leaving $3N-3$ for assorted other possibilities. If we look at just what 2 atoms can do, of the 6 motional degrees of freedom, 3 are for translational movement of molecules in real space. This leaves 3. A complete and independent set of these is vibration along the axis, and 2 more for the rotational freedom in two orthogonal planes that intersect along the molecular axis. Water with 3 atoms has a larger collection of such modes. In a kitchen microwave, the 2.45 GHz frequency $\equiv 12$ cm of free space wavelength or a 10 μeV energy photon is absorbed by the water molecule in this motional freedom.

degeneracy, one may split the eigenfunction describing both the electrons and the nuclei as separable products; that is,

$$|\psi\rangle = |\phi(\{\mathbf{r}\},\{\mathbf{R}\})\rangle|\chi(\{\mathbf{R}\})\rangle, \tag{1.55}$$

where $\{\mathbf{R}\} = (\mathbf{R}_1, \ldots, \mathbf{R}_N)$ are all the nuclear displacements, and $\{\mathbf{r}\} = \mathbf{r}_1, \ldots, \mathbf{r}_N$ all the electron coordinates, which may be generalized to include spins if appropriate. This separability establishes that the electronic motion instantaneously responds to the atomic and nuclear motion—that is, it is a function of these coordinates—and is separable from the nuclear response, which can be written as a function of its own coordinates alone. The electron charge cloud follows rapidly the slow response of the electron and nuclear particles of the system. The energy of all possible arrangements is calculable in principle within the constraints of the accuracy of the method for the Hamiltonian's solution—Hartree-Fock—there. We will discuss Hartree and Hartree-Fock approaches to solving the Hamiltonian shortly. The equilibrium position is relatively accurate through the distribution of separations, even if the derivative around this equilibrium is not.

As an example of separation of the nuclear part, the Born-Oppenheimer approximation's use and its analysis, take the example of electromagnetic absorption by a molecule, which can also be extended to the crystal and is particularly useful for the nanoscale. Figure 1.7 shows the underlying process of an interaction of a molecule absorbing a radiation photon. This is an illustration of the Franck-Condon shift where fast and slow again makes its appearance. The Franck-Condon principle states that since the electron mass is much smaller than the nuclear mass, electronic transitions can be treated with a stationary nuclear framework. This is pretty much the Born-Oppenheimer approximation. The Franck-Condon approach is used for molecules, but it is also pertinent to the electron in the crystal with atoms attached to each other.

A *configuration coordinate* diagram shows the energetic changes as multiple coordinates; for example, the geometric spacing of the center of motion, and the relative spacings of components, together with the momenta of a system, also undergo a change. Figure 1.7 shows the energetics as a function of nuclear position for a system undergoing transition by interaction with a photon: the initial and final state electronic energy as a function of the displacement $\mathbf{R}_j$ of the jth nuclei. The electron energies $E(\mathbf{R})$ as a function of the atomic/nuclear displacement within the harmonic approximation for two different nuclear positions of the jth nuclei are shown here for both the initial state and the final state. The vibrational part, due to the nuclear motion, occurs slowly. The electron state change

The Born-Oppenheimer approximation is very useful but, nevertheless, imperfect. There is inevitably some mixing between states. An electron falling behind nuclear motion may be closer to another state—a mixing has happened—and therefore transitions that are not allowed may become possible, small shifts in energies may come about, or degeneracies may be removed.

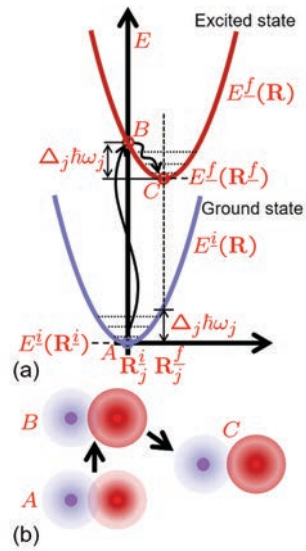


Figure 1.7: Franck-Condon shift, an example of fast-and-slow change, shown in configuration coordinate diagram. (a) A sequence of processes. An optical transition from A to B—a resonant absorption—is fast and accompanied by an ionic distortion since electrons move rapidly but the nucleus does not. This metastable excited state slowly relaxes to the excited state C with a Franck-Condon shift of $\Delta_j\hbar\omega_j$. (b) A physical interpretation at the atomic level. In the state C, the nuclei of the excited state are farther away. Eventually, another optical emission takes place, leaving the nuclei still far apart. And eventually they slowly come back to the starting state A. This picture of configuration coordinates is useful in understanding some of the deep levels in semiconductors— vacancies, atoms displaced, et cetera—where lattice distortion accompanies the electron capture or emission process.

occurs rapidly. The electronic and vibrational parts of the response can be separated. The vibration part, in the linear limit, is a simple harmonic oscillator, whose center will change as time proceeds, and this process will involve phonon emission in the solid. The different excitation states of the electron are the different dotted lines in the figure in the harmonic well.

These energies can be written as

$$E^i(\mathbf{R}) = E^i(\mathbf{R}^i) + \sum_{j=1}^{3N} \frac{1}{2} m\hbar\omega_j^2 (\mathbf{R}_j - \mathbf{R}_j^i)^2, \text{ and}$$

$$E^f_-(\mathbf{R}) = E^f_-(\mathbf{R}^f_-) + \sum_{j=1}^{3N} \frac{1}{2} m\hbar\omega_j^2 (\mathbf{R}_j - \mathbf{R}_j^f)^2. \tag{1.56}$$

The electronic energies differ in energy linearly with the displacement $\mathbf{R}_j$ if the coupling between electronic and nuclear motion is also linear under small displacement. The nuclear contribution oscillates around mean displacement position ($\mathbf{R}_j^i$ and $\mathbf{R}_j^f$). The change in mean position, representative of the electron-nuclei interaction, can be related through a parameter written here without proof as

$$\Delta_j = m\left(\frac{\omega_j}{2\hbar}\right)|\mathbf{R}_j^i - \mathbf{R}_j^f|^2. \tag{1.57}$$

Δ_j—the Huang-Rhys factor—is a dimensionless factor tying the strength of the coupling of the electronic states to the nuclear motional freedom.

The photon absorption induces an electronic transition from point A, with the system evolving, without nuclear positional change, to point B. Nuclear position now changes slowly to $\mathbf{R}_j^f$ (point C) via phonon emission, with the system still staying excited. The net difference between the final and initial state energies is

$$E^f_-(\mathbf{R}) - E^i(\mathbf{R}) = E^f_-(\mathbf{R}^f_-) - E^i(\mathbf{R}^i) + \sum_{j=1}^{3N} \Delta_j \hbar\omega_j + \sum_{j=1}^{3N} \sqrt{2m\hbar\omega_j^3 \Delta_j}|\mathbf{R}_j|. \tag{1.58}$$

Here, the first summation is the Franck-Condon energy corresponding to the net relaxation of the molecule or crystal, and the last term is the result of electron and nuclei motion interaction, which is the electron-phonon coupling.

The absorption of the photon—a fast process—causes a change in energy without a change in nuclear configuration, which thereon relaxes through the transfer of energy to the vibrations. We could separate the terms, since the nuclear part of the wavefunction could

be written as a product term solely in terms of nuclear position in the wavefunction of Equation 1.55. We could analytically resolve this because of the separability of the fast-and-slow process that adiabatic approximation could be applied to. This example also introduces us to the use of configuration diagrams when local positional changes have an effect on energy. We will encounter this a few times, particularly so in understanding defects, where the local crystal environment will have a consequence for electronic energy interactions, just as it did in this molecular example.

The heating of a water molecule in the microwave depends on an electronic transition that then shakes the water molecules, which is an effective way to heat food.

1.5 *A 2-electron and 2-atom system*

We start our exploration of solids with the simplest of cases of matter assembly: a 2-electron, 2-atom system—the hydrogen molecule—to understand the energetics and the different possibilities of the eigenstates as starting points. There is much here from which we can draw implications for the solid forms of interest to us. In this 2-electron, 2-atom example, the Hamiltonian operator, with the electron's kinetic and Coulomb potential energies included, is

$$\hat{\mathscr{H}} = -\frac{\hbar^2}{2m_0}\nabla_1^2 - \frac{\hbar^2}{2m_0}\nabla_2^2 - \frac{1}{4\pi\epsilon_0}\frac{e^2}{r_{1A}} - \frac{1}{4\pi\epsilon_0}\frac{e^2}{r_{2A}} - \frac{1}{4\pi\epsilon_0}\frac{e^2}{r_{1B}} - \frac{1}{4\pi\epsilon_0}\frac{e^2}{r_{2B}} + \frac{1}{4\pi\epsilon_0}\frac{e^2}{r_{12}} + \frac{1}{4\pi\epsilon_0}\frac{e^2}{R}, \tag{1.59}$$

where A and B denote the sites of the two atoms that are R apart and 1 and 2 represent the two electrons of mass m_0, so that r_{1A} means the separation of electron 1 of charge $-e$ from a residual core of charge $+e$ ($Z = 1$ in Equation 1.11) and r_{12} is the separation between the two electrons. We wish to find the lowest energies. It stands to reason that these will be reconstituted from the lowest energy states of the atoms from which the molecule is formed as the interaction evolves. Let $|u_{A\uparrow}\rangle$, $|u_{A\downarrow}\rangle$, $|u_{B\uparrow}\rangle$ and $|u_{B\downarrow}\rangle$ represent the 1s orbital wavefunction of the two atoms A and B of this hydrogen molecule, with spin up $\uparrow \equiv m_s = +1/2$ and spin down $\downarrow \equiv m_s = -1/2$ possibilities. These are the four possible spin orbitals.

In following through by the writing of Coulomb energy terms in this form, energy terms will appear as a Coulomb integral. An electron in an atomic orbital $|u_1\rangle$ has a charge density of $-e\langle u_1|u_1\rangle$ or a charge of $-e\langle u_1|u_1\rangle d^3r_1$ in volume d^3r_1. A second electron has a charge $-e\langle u_2|u_2\rangle d^3r_2$. For a separation r_{12}, the potential energy of the Coulomb interaction—the total electrostatic interaction between the two elemental charges—is the integral over the entire space of each of these volume elements. This is the Coulomb integral. It is a net increase in energy because charges are of the same sign.

Let the two electrons be represented by 1 and 2, respectively. So, $|u_{A\uparrow}(1)u_{B\uparrow}(2)\rangle$ speaks to electron 1 on atom A, and electron 2 on atom B, where both electrons have $m_s = +1/2$. We may exchange the electrons between the two atoms, in which case wavefunction $|u_{A\uparrow}(2)u_{B\uparrow}(1)\rangle$ is also a possibility. There are many more such possibilities—a total of ${}^4C_2 = 6$—that represent choosing any 2 out of 4 available spin orbitals. Two of these—$|u_{A\uparrow}u_{A\downarrow}\rangle$ and $|u_{B\uparrow}u_{B\downarrow}\rangle$—

This is the Pauli exclusion principle telling us that no two eigenfunction solutions for the fermions can be the same.

represent the 2 electrons on one atom (A or B). The other is bereft. So, we have a combination where one is now a negative ion, and the other a positive ion. We exclude these from our discussion. They are certainly possible under energetic circumstances, but we are interested in solutions where both centers are still neutral.

So, we need determinental functions that exclude the degenerate possibilities, for example, under exchange of the two electrons on the two centers ($|u_{A\uparrow}(1)u_{B\uparrow}(2)\rangle$ and $|u_{A\uparrow}(2)u_{B\uparrow}(1)\rangle$), and form a non-degenerate set for a stationary solution.

Let $\{|u_i^0\rangle\}$ be the set of orthonormal basis functions. We build a linear combination $|\psi\rangle = \sum_i c_i|u_i^0\rangle$. Let $\mathcal{O}_{jk} = \langle u_j^0|u_k^0\rangle$. This is an overlap matrix element. If $\mathscr{H}_{jk} = \langle u_k^0|\hat{\mathscr{H}}|u_j^0\rangle$, then stationarity requires

$$\sum_i c_i(\mathscr{H}_{jk} - E\mathcal{O}_{ji}) = 0, \tag{1.60}$$

by variational principle. For a unique stationary solution to exist for this set of equations, the secular equation

$$\det|\mathscr{H}_{jk} - E\mathcal{O}_{ji}| = 0 \tag{1.61}$$

must be satisfied. For N basis functions, there are N roots that are the eigenvalues. Each of these eigenvalues is associated with a combination of c_i's of Equation 1.60—a linear combination. For us, here these are to be built from $|u_{A\uparrow}u_{B\uparrow}\rangle$, $|u_{A\uparrow}u_{B\downarrow}\rangle$, $|u_{A\downarrow}u_{B\uparrow}\rangle$ and $|u_{A\downarrow}u_{B\downarrow}\rangle$. We have now discarded the identification of each electron, as it is implicit in this choice set, where M_s—the sum of the secondary spin number along the axis of quantization—changes from 1, to two with 0, and the last one with -1. This middle set of $M_s = 0$ leads to the linear combination through sum and difference, which are distinguishable, but which will also lead to a degenerate energy. Our four solutions, unnormalized, but explicitly including the spin and the electron and atom identity, are

$$\begin{aligned}
|u_{A\uparrow}u_{B\uparrow}\rangle &= [u_A(1)u_B(2) - u_A(2)u_B(1)]|\uparrow(1)\rangle|\uparrow(2)\rangle,\\
|u_{A\downarrow}u_{B\downarrow}\rangle &= [u_A(1)u_B(2) - u_A(2)u_B(1)]|\downarrow(1)\rangle|\downarrow(2)\rangle,\\
|u_{A\uparrow}u_{B\downarrow}\rangle + |u_{A\downarrow}u_{B\uparrow}\rangle &= [u_A(1)u_B(2) - u_A(2)u_B(1)]\\
&\quad\times[|\uparrow(1)\rangle|\downarrow(2)\rangle + |\downarrow(1)\rangle|\uparrow(2)\rangle], \text{ and}\\
|u_{A\uparrow}u_{B\downarrow}\rangle - |u_{A\downarrow}u_{B\uparrow}\rangle &= [u_A(1)u_B(2) + u_A(2)u_B(1)]\\
&\quad\times[|\uparrow(1)\rangle|\downarrow(2)\rangle - |\downarrow(1)\rangle|\uparrow(2)\rangle],
\end{aligned} \tag{1.62}$$

showing the separation of spatial and spin coordinates with their changing symmetries that make these combinations different from each other. The top three of these have antisymmetric spatial coordinates and spin coordinates are $M_s = 1, -1$ and 0.

A stationary state is one whose probability density ($\langle\psi|\psi\rangle$) is invariant in time. See the probability discussion of random processes in Appendix C.

See Appendix D for a discussion of variational principle and its usage. The reader will also find Appendix E and Appendix F as a short encapsulation of the important notions of thermodynamics and important distribution functions. Spin is discussed in Appendix G.

The simplest illustration of the removal of degeneracy in the presence of interaction is that of bonding and antibonding states. Let two systems share identical Hamiltonians. Independent, they have identical energy. Let $|l\rangle$ and $|r\rangle$ be the eigenfunctions that have this same energy E as their solution. Bring the two closer to cause each to perturb the other. Equation 1.61 gives the solution for the energy. Absent perturbation, so $\mathcal{O}_{ji}$s vanishing, the energy E is the solution for the two non-interacting systems in the $|l\rangle$ and $|r\rangle$ state. With perturbation, the energies change, and we have the solutions

$$\begin{aligned}
E^+ &= E + \Delta, \text{ with}\\
|+\rangle &= \frac{1}{\sqrt{2}}(|r\rangle - |l\rangle), \text{ and}\\
E^- &= E - \Delta, \text{ with}\\
|-\rangle &= \frac{1}{\sqrt{2}}(|r\rangle + |l\rangle),
\end{aligned}$$

where Δ is the energy change arising due to the perturbation. This perturbed system does not have $|l\rangle$ or $|r\rangle$ as its eigenfunction or E as its eigenenergy. The energies have changed to E^+ and E^-, and the new eigenfunctions are linear combinations of unperturbed eigenfunctions. $|l\rangle$ and $|r\rangle$ have hybridized. Degeneracy has been lifted. the higher energy state is the antibonding state. It is spatially antisymmetric. The lower energy state is the bonding state. It is spatially symmetric. This is a "molecular" description. States were localized. Such a "molecular" model is very useful in understanding many defects in semiconductors.

The eigenvalue of these three is identical, arising in the spatial dependences. Spatially antisymmetric functions form the triplet. The last wavefunction, also an eigenfunction and spatially symmetric, forms a singlet state.

With these functions known, one may calculate the energies following normalization, since the Hamiltonian of this problem only involves spatial dependence. This is shown in Figure 1.8 as a function of changing R. The singlet spatially symmetric solution has a lower energy, while the triplet spatially antisymmetric solution has higher energy. The former is our bonding state, and the latter the anti-bonding state. This approach is an illustration of the tight binding approach. We built a tight molecular construction using it. It is also known as linear combination of atomic orbitals (*LCAO*), since it built the hybrids, or the evolved eigenfunctions starting from the original atomic orbitals as the orthonormal set.

Bringing two atoms together here has led to a lowering of energy in the bonding state and has resulted in a stable molecule with one bonded spatially symmetric solution. It has illustrated to us a methodology that will be a stepping stone to more complicated constructs. For us, a very instructive one is of N electrons together with the nuclei.

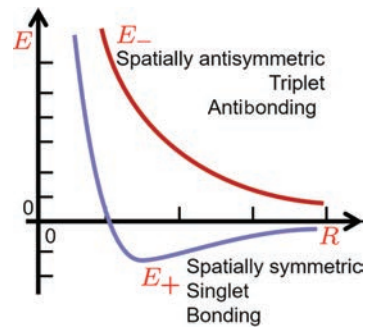


Figure 1.8: The energy, as a function of interatomic spacing, of a model hydrogen molecule for the four wavefunction solutions. The antisymmetric solution has a higher energy than the symmetric singlet solution.

1.6 *N non-interacting electrons in the presence of nuclei*

IF ONE HAS N NON-INTERACTING ELECTRONS, and they are also independent of each other, that is, no Pauli-exclusion and fermionic constraints, the Hamiltonian may be written as $\hat{\mathscr{H}} = \sum_1^N \hat{\mathcal{H}}(\mathbf{z}_i)$, where the Hamiltonian of the ith electron—a sum of the kinetic $((-\hbar^2/2m_0)\boldsymbol{\nabla}_i^2)$ and the potential energy (V) form—is a function of both the position ($\mathbf{r}_i$) and the spin ζ_i coordinates; that is, $\mathbf{z}_i = \mathbf{r}_i, \zeta_i$. These N Hamiltonians are identical, and their solutions degenerate, that is,

Spin often will become important because both Pauli exclusion and magnetic energetics relate to the spin's important role. See Appendix G for a discussion of the spin and spin matrices to represent spin coordinates.

$$\hat{\mathcal{H}}(\mathbf{z})|u_i(\mathbf{z}_i)\rangle = E_i|u_i(\mathbf{z}_i)\rangle \;\; \forall\; i = 1, \ldots, N, \tag{1.63}$$

so that the net energy of the N electrons is $E = \sum_{i=1}^N E_i$ and the wavefunction is also a product of the one-electron states, that is, $|\psi\rangle = \prod_{i=1}^N u_i(\mathbf{z}_i)$. The problem with this solution is that the wavefunction $|\psi\rangle$ is not antisymmetric, and it does not represent a collective ensemble of N non-interacting electrons that are not independent.

Another way of saying this is that we have found a solution for N one-electron one-nucleus systems that are all very far apart from each other so that they do not interact and are also independent of each other.

The wavefunction $|\psi\rangle$ for the ensemble Hamiltonian composed of these one-electron states must be an antisymmetric function for the fermion electron, since Pauli exclusion applies. This is the *Slater determinant*

$$|\psi\rangle = \frac{1}{\sqrt{N!}} \begin{vmatrix} |u_1(\mathbf{z}_1)\rangle & \cdots & |u_1(\mathbf{z}_N)\rangle \\ |u_2(\mathbf{z}_1)\rangle & \cdots & |u_2(\mathbf{z}_N)\rangle \\ \vdots & \vdots & \vdots \\ |u_N(\mathbf{z}_1)\rangle & \cdots & |u_N(\mathbf{z}_N)\rangle \end{vmatrix}, \tag{1.64}$$

A Slater determinant obeys antisymmetry by construction. Determinants change sign when rows or columns are interchanged. The Slater determinant also keeps the electrons indistinguishable.

with spin and which is orthonormal. The eigenenergy is $E = \sum_i^N E_i$. In this form, when two electrons are taken in identical wavefunction form, the determinant vanishes, as expected from Pauli exclusion, and the electrons are also indistinguishable. If the energy, and so also the Hamiltonian $\mathcal{H}(\mathbf{z})$, are independent of the spin, then one may separate the spin part from the spatial dependence; that is,

Indistinguishability here means that each electron is associated with each electron wavefunction.

$$|u_{i\sigma}(\mathbf{z})\rangle = |u_i(\mathbf{r})\rangle|v_\sigma(\zeta)\rangle. \tag{1.65}$$

Electrons of opposite spin may coexist in the same orbital state, so the ground state of this N non-interacting electron system is one where the entire gamut of lowest one-electron states is filled with two electrons, each of opposite spin per state.

This multiple electron question shows up as a problem of lowest energy in bonding in chemistry. Spin up and spin down are valid in a bonding state since the resulting eigenfunction is antisymmetric.

An important point of note here is that one could follow through this way and obtain the wavefunction for the system as a whole using the Slater determinant, because the electrons were non-interacting and because the total Hamiltonian could be written as a sum of each electron's Hamiltonian. And each electron's Hamiltonian was a function of its coordinate interacting only with nuclei. All other energetics, such as electron-electron interaction energies, for example, were, by fiat, zero. The independence let us write the total Hamiltonian as a sum of each electron's, and hence the wavefunction solutions found from Equation 1.63 served to build the system's wavefunction. The hydrogen molecule model in Section 1.5 is an example of this approach applied to a 2-electron, 2-nuclei problem.

1.7 *N interacting electrons in the presence of nuclei*

In an N-interacting-electron system, together with the atoms from which the electrons arose, the wavefunction of the system is a function of the coordinates of all electrons and of other particles, such as the cores, with which they meaningfully interact. Even with the simplification of these N electrons in a continuum approximation averaging the positively charged ion background, because of the interaction between all the electrons, the Hamiltonian is not separable as it was in previous case. We need to approximate the effect of interaction of an electron with all others by a potential

that is only a function of the electron coordinate. Only then is it possible to treat the electrons one at a time separately. This is an example of a *self-consistent field approximation*—finding a mean field in which interaction may be tackled for specific particle with attention to only its coordinates—in numerous places and is useful for our N-interacting-electron system. Since the self-consistent approach requires the interaction potential and a knowledge of the states, with each dependent on the other, one must compute the solution self-consistently, iterating until one finds a solution that is satisfactorily accurate.

Here, this self-consistent field approximation implies that, for the specific electron for which one is computing, the potential in which it is present can be calculated by freezing all the other electrons and taking their averaged distribution by a centrosymmetric potential source. Solving the Schrödinger equation gives an updated state description. And this becomes a starting point for improving the accuracy of description of other electrons, again using the frozen representation for the rest. One cycle of this procedure updates all the electrons' orbital descriptions, and the procedure may be repeated. The procedure has limitations but is quite useful. As the number of particles increases, it becomes increasingly more accurate. It is therefore quite useful for the determination of many of the properties that are of interest to us.

1.7.1 *Hartree approximation*

WITH N ELECTRONS AND THEIR NUCLEI as an ensemble, the Hamiltonian may be written as

$$\hat{\mathscr{H}} = \sum_{i=1}^{N} \hat{\mathcal{H}}(\mathbf{z}_i) + \frac{1}{2} \sum_{i \neq j} \frac{1}{4\pi\epsilon_0} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} + \hat{V}_{NN}. \tag{1.66}$$

The first term is as before (a sum of kinetic energy and the Coulomb interaction with the nuclei/ion), the second is the electron-electron Coulomb interaction while compensating for duplication and only including separate electrons, and the last is the Coulomb interaction between the nuclei $\hat{V}_{NN} = \sum_{i=1}^{N} \hat{V}(\mathbf{R}_i)$. It is the second term that mixes up the electron coordinates because of the electron-electron Coulomb interaction that makes the solution not decomposable to the form discussed in Section 1.6, where this interaction was excluded by making the independent electron approximation. So, this wavefunction solution is a function in the coordinates $(\mathbf{r}_1, \ldots, \mathbf{r}_N)$ has now become non-trivial.

Hartree's insight is that one may tackle each electron separately by viewing it as moving in the field of the nuclei (or ions, in the simplification for solids) and in the average field due to the other electrons. This self-consistent field approximation—the *Hartree approximation*—corresponds to finding the solution of

$$\left[\hat{\mathcal{H}}(\mathbf{z}_i) + \sum_{i \neq j} \int \frac{1}{4\pi\epsilon_0} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} \langle u_j(\mathbf{z}_j) | u_i(\mathbf{z}_i) \rangle \langle u_j((\mathbf{z}_j) | u_j((\mathbf{z}_j) \rangle d\mathbf{z}_j \right] |u_i(\mathbf{z}_i)\rangle$$
$$= E_i |u_i(\mathbf{z}_i)\rangle \;\; \forall\, i = 1, \ldots, N. \tag{1.67}$$

The problem has again been reduced to N equations, one for each electron, where each is in a field due to the other $N-1$ electrons.

On the left-hand side of Equation 1.67, we add and subtract an unphysical *self-interaction Coulomb term* for $i=j$, and one obtains the form

$$\left[\hat{\mathcal{H}}(\mathbf{z}) + \hat{V}_H(\mathbf{z}) - \hat{V}_{si}(\mathbf{z})\right] |u_i(\mathbf{z})\rangle = E_i |u_i(\mathbf{z})\rangle, \tag{1.68}$$

an equation that is now identical for all the electrons $i=1,2,\ldots,N$, where

$$V_H(\mathbf{z}) = \int \frac{1}{4\pi\epsilon_0} \frac{e^2}{|\mathbf{r}-\mathbf{r}'|} n(\mathbf{z}')\, d\mathbf{z}',$$

$$n(\mathbf{z}) = \sum_i o_i \langle |u_i(\mathbf{z})|u_i(\mathbf{z})\rangle, \quad \text{and}$$

$$V_{si}(\mathbf{z}) = \int \frac{1}{4\pi\epsilon_0} \frac{e^2}{|\mathbf{r}-\mathbf{r}'|} \langle |u_i(\mathbf{z}')|u_i(\mathbf{z}')\rangle\, d\mathbf{z}'. \tag{1.69}$$

This self-interaction term is introduced only for convenience of calculation. An electron, of course, does not interact with itself. But, through this subterfuge, one can write a Hartree potential as an integration over the ensemble.

Here, V_H is the *Hartree potential*—a Coulomb energy term arising in the electron interactions that includes the unphysical $i=j$ contribution and written in terms of electron density, which is a summation over the product of probability density of the ith state and its occupation factor (o_i, where o_i is 1 if an electron is present, and 0 if it is absent). The last term of V_{si} is a self-interaction term that is now being subtracted to compensate for what was artificially introduced in the Hartree potential.

This Hartree potential approach is quite a good starting point for describing an atom's electronic picture, or a many-electron picture of a solid, since the set of equations represented by Equation 1.67 are now solvable with the Hartree potential calculable through simple averaging in Equation 1.69. V_H represents a mean field effect, but one that includes the unphysical self-interaction term.

In the Hartree approximation, one starts with a trial wavefunction $|\psi\rangle$ composed of independent electrons; that is,

$$|\psi\rangle = \prod_{i=1}^{N} |u_i(\mathbf{z}_i)\rangle, \tag{1.70}$$

ignoring the antisymmetry. The desired solution is the one that has the minimum for $\langle\psi|\hat{\mathscr{H}}|\psi\rangle$. This requires the variational expectation to vanish; that is,

$$\langle\delta\psi|\hat{\mathscr{H}}|\psi\rangle = 0. \tag{1.71}$$

See Appendix D, where the approach of using the variational principle to minimize energy by varying from a good guess of a starting trial solution is discussed.

So, the variational treatment of $|u_i\rangle$ leads to the solution of Equations 1.66 and 1.67, our equations of the Hartree formulation. We have reduced the problem to solving N single particle equations with unknown $n(\mathbf{z}') - \langle u_i(\mathbf{z}')|u_i(\mathbf{z}')\rangle$. An iterative approach that brings about self-consistency between the density's implication

for potential (Equations 1.69) and the governing Hamiltonian (Equation 1.68) with the use of equilibrium statistics leading back to the density will tackle it.

But there are serious shortcomings. The most important is that these equations have no constraint that reflects Pauli exclusion (Equation 1.70 is not antisymmetric); that is, exchange.

1.7.2 *Hartree-Fock approximation*

In the Hartree approximation, we started with the trial $|\psi\rangle$ as the product of independent electron wavefunctions. The Slater determinant gives us the coupled solution that adheres to all the quantum-mechanical constraints. This antisymmetrization is the Hartree-Fock approximation. Both Hartree and Hartree-Fock are, however, still approximate with Hartree-Fock, an improvement toward accuracy. Multiple electron assemblies such as atoms and molecules—systems with restricted numbers—show this.

THE HARTREE-FOCK APPROXIMATION ANTISYMMETRIZES the wavefunction. It starts with the Slater determinant, with its orthonormality for spin orbitals of independent electrons as the initial trial wavefunction. For the Hamiltonian of the Hartree equation (Equation 1.66) for this N-interacting-electron system, the energy solution is

$$E = \sum_i o_i \langle u_i|\hat{\mathcal{H}}|u_i\rangle + \frac{1}{2}\sum_{i,j} o_i o_j \left(\langle u_i u_j \left| \frac{1}{4\pi\epsilon_0}\frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} \right| u_i u_j \rangle \right.$$
$$\left. - \langle u_i u_j \left| \frac{1}{4\pi\epsilon_0}\frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} \right| u_j u_i \rangle \right) + V_{NN}. \tag{1.72}$$

Note the antisymmetrization in the second term. Here, we have used the generalized notation

$$\langle u_k u_l \left| \frac{1}{4\pi\epsilon_0}\frac{e^2}{|\mathbf{r} - \mathbf{r}'|} \right| u_i u_j \rangle$$
$$= \int u_k^*(\mathbf{z}) u_l^*(\mathbf{z}') \frac{1}{4\pi\epsilon_0}\frac{e^2}{|\mathbf{r} - \mathbf{r}'|} u_i(\mathbf{z}) u_j(\mathbf{z}')\, d\mathbf{z}\, d\mathbf{z}' \tag{1.73}$$

for brevity.

Minimization of E for all $|u_i\rangle$ under the constraint of their orthonormality may be accomplished using Lagrangian multipliers; that is, we require

A short summary of the method of Lagrangian multipliers can be found in Appendix D. This approach to finding solutions under constraints is an essential instrument from the tool set of mathematics that we employ throughout.

$$\delta E - \sum_{i,j} \lambda_{ij} \int \delta u_j^*(\mathbf{z}) u_i(\mathbf{z})\, d\mathbf{z} = 0 \quad \forall\ \delta u_j^*. \tag{1.74}$$

The one-particle equation set that this corresponds to is

$$\left[\hat{\mathcal{H}}(\mathbf{z}) + \sum_j o_j \langle u_j(\mathbf{z}') \left| \frac{1}{4\pi\epsilon_0}\frac{e^2}{|\mathbf{r} - \mathbf{r}'|} \right| u_j(\mathbf{z}') \rangle \right] |u_i(\mathbf{z})\rangle$$
$$- \sum_j o_j \langle u_j(\mathbf{z}') \left| \frac{1}{4\pi\epsilon_0}\frac{e^2}{|\mathbf{r} - \mathbf{r}'|} \right| u_i(\mathbf{z}) \rangle |u_j(\mathbf{z})\rangle$$
$$= \sum_j \lambda_{ij} |u_j(\mathbf{z})\rangle. \tag{1.75}$$

The first of the summation interaction terms on the left is just the Hartree potential V_H that we encountered before. The second summation term on the left is an *exchange term*. Since the Slater determinant undergoes only a phase factor change under unitary transformation, the equations remain the same structurally under diagonalization of this equation set. So, one may reform Equation 1.75 to a diagonal form where $\lambda_{ij} = E_i \delta_{ij}$. We now have the equation set

$$\begin{aligned}&\left[\hat{\mathcal{H}}(\mathbf{z}) + V_H(\mathbf{z})\right] |u_i(\mathbf{z})\rangle \\ &\quad + \int \left[-\frac{1}{4\pi\epsilon_0} \frac{e^2}{|\mathbf{r}-\mathbf{r}'|} \sum_j o_j \langle u_j(\mathbf{z}')|u_j(\mathbf{z})\rangle \right] |u_i(\mathbf{z}')\rangle \, d\mathbf{z}' \\ &= E_i |u_i(\mathbf{z})\rangle \;\; \forall\; i = 1, \ldots, N. \end{aligned} \tag{1.76}$$

The last term on the left in this equation is a nonlocal exchange interaction. The equation form can be written more meaningfully—and simply—as

$$\left[\hat{\mathcal{H}}(\mathbf{z}) + V_H(\mathbf{z}) + V_x(\mathbf{z})\right] |u_i(\mathbf{z})\rangle = E_i |u_i(\mathbf{z})\rangle \;\; \forall\; i = 1, \ldots, N, \tag{1.77}$$

where

$$V_x(\mathbf{z}) = \frac{1}{u_i(\mathbf{z})} \int \left[-\frac{1}{4\pi\epsilon_0} \frac{e^2}{|\mathbf{r}-\mathbf{r}'|} \sum_j o_j \langle u_j(\mathbf{z}')|u_j(\mathbf{z})\rangle \right] |u_i(\mathbf{z}')\rangle \, d\mathbf{z}' \tag{1.78}$$

In Equation 1.76, the $i = j$ contribution arising in the last term on the left is precisely V_{si}. But this Hartree-Fock approximation has reformed it into a correction term V_x arising in exchange for $i \neq j$. It is more accurate, even if less intuitive, and it has pulled in the nonlocal exchange's energetic consequence. This Hartree-Fock equation set can tackle the spin orbital as factorized by Equation 1.65. It accomplishes this by transforming the integration over $\mathbf{z}'$ to $\mathbf{r}'$, while V_H gets doubled for spin degeneracy and the exchange term is unchanged since their contribution to Equation 1.75 vanishes. In the Hartree approximation (Equation 1.68), we had to explicitly exclude the $i = j$ term in the summation. In the Hartree-Fock approximation, we do not have to exclude this, since the exchange term sums cancel with the $i = j$ term. When N is large, with electron contributions scaling as $1/N$, the distinction between Hartree and Hartree-Fock rapidly vanishes.

But, at small N, so few electron and few atom systems, such as nanostructures, these approaches and their judicious correction for $i = j$—within the self-consistent field approximation—will have noticeable consequences. Even though the Hartree-Fock approach

should be expected to be more accurate by accounting for exchange, there will still be major shortcomings.

The next major shortcoming to consider is that we have not accounted for *correlation*. Correlation here is the notion that the mere presence of an electron causes a redistribution of the other electrons, due to electron-electron repulsion creating a "Coulomb hole." This issue of correlations and exchange-correlation holes is tackled in the next subsection.

We will encounter the word "correlation" often. Even Pauli exclusion is a form of correlation. Two electrons in a non-quantum mechanical view move independently of each other. But there is a low probability of them being in the same location, due to the Coulomb interaction. This is a charge correlation. The change in localization of electrons, such as in the hydrogen molecule, through a wavefunction that is a superposition of different configurations is a configurational correlation. Fermions avoiding each other for the same spin and clustering together when in opposite spin is a spin correlation. Any interaction between an electron and another electron that is not due to the Coulomb interaction is a quantum-mechanical consequence that is a correlation effect in this view. It appears in numerous forms, and we will see these throughout our discussion.

1.7.3 *Correlations*

DISCREPANCIES ARISE in a number of considerations neglected up to this point. In an atom or molecule, a principal one is from the relativistic effect in the core electrons with their large kinetic energy. Another one is due to *correlation energy*, which is important to atoms, molecules and atomic assemblies where many electrons will exist. Hartree-Fock ignores any local changes in the distribution of an electron, since it force fits a mean effect arising in the others.

Take the case of a molecule. When an electron is in the vicinity of another electron, Hartree-Fock accounts for it for the whole orbital as an average. This neglects any local electron-electron effect. This neglecting of electron correlations in its configuration form due to the mean field formalism makes energy calculation inaccurate at long separations and in the curvature at equilibrium. So, as seen in Figure 1.9, the poor representation of local distortion results in accurate representation of local equilibrium geometry but poor calculation of properties such as force constants, vibration frequencies, et cetera. Note that subsumed in the Hartree-Fock approximation also is the Born-Oppenheimer adiabatic approximation. The molecular potential energy is a function of relative nuclei locale. And any calculation where electrons follow any nuclear movement instantaneously will have increasing errors in the calculation of dynamic parameters. In a solid, this same correlation will cause similar inaccuracies when the number of interacting electrons is small and local electron-electron interaction important.

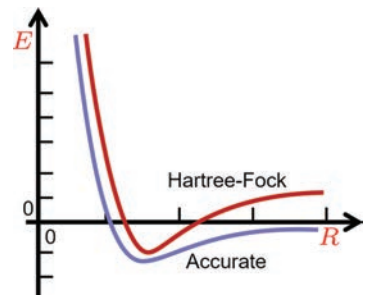


Figure 1.9: Molecular potential energy in Hartree-Frock approximation, where correlation is not accounted for. The minimum is geometrically close, but the derivatives are not, and the dissociation limit of $R \to \infty$ inaccurate.

We will modify the Hartree-Fock approach by accounting for this configuration interaction of correlation.

But, we also note that the Hartree-Fock equation is actually quite accurate and successful in a number of situations of interest to us. It is, for example, solvable and accurate for a free electron gas with a uniform compensating background. With one-electron wavefunction as a plane wave, the exchange interaction is calculable. The $1/r$ dependence then leads to a total energy that is proportional to

$n^{1/3}$, where $n = N/\Omega$ is the electron density for N electrons in volume Ω. The mean separation between the particles varies as $n^{-1/3}$ and the plane wave gives an equal probability everywhere, so any distortion effects naturally disappear with increasing particle population. We return to this problem shortly because of this construct's peculiarity.

The correlation effects have another important consequence: the existence of holes and mounds. A simple example is from spin correlation. Since two electrons of the same spin may not be found at the same point, if one looked at the probability of finding a second electron of identical spin, it will vanish at the point of no separation. The wavefunction of the second electron asymptotically vanishes in the vicinity of the first electron's locale. This is a *Fermi hole*. Likewise, for opposite spins, the second electron's probability is enhanced. Pauli exclusion, or exchange, has resulted in this spin correlation effect. In a configuration of atoms, the Hund rule's of maximum multiplicity is indirectly due to this spin correlation's consequence.

Hund's rule of multiplicity—an observation—is that when an atom or molecule has open electronic shells, then, in any given electronic configuration, the lowest energy form corresponds to the largest spin multiplicity. That is, the state with the largest total spin ($\sum_{m_s}$) is the one with the most stable configuration. So, with multiple orbitals of equal energy, electrons prefer to occupy them singly before pairing. Pairing reduces total spin. In silicon, which has $3s^2 3p^2$ as the outer orbital configuration, the $3p^2$ may exist as ^1D, ^3P and ^1S—multiplicity being $2S+1$ indicated in the superscript of $J = L + S$, with L the orbital quantum number—^3P, with $M_L = 1$ and $M_S = 1$—is the favored ground state. This is Hund's rule 1; rule 2 is an observation on the consequences of electron repulsion, for which the silicon triplet state has no choice. Hund's rule 2 states that, in as much as it is possible to be consistent with rule 1, the configuration that maximizes orbital angular momentum is favored. Hund's rule 3 minimizes spin-orbit coupling and so appears for larger atoms with d and f electrons. The first two rules really are because of the Coulomb energy implication of the arrangements. Hund's rule can be seen classically in orbital motion. Two pendulums interacting with each other lock orbital and spinning motion.

Since the Hartree-Fock approximation does not include correlation effects, *configuration interaction* needs to be computed as a modification. The eigenstates of the system of N interacting electrons can be built by expansion of the Slater determinant constructed off an infinite set of orthonormal one-particle spin orbitals whose starting point is the Hartree-Fock ground state determinant. Let $|u_{Sn}\rangle$ represent the basis of the Slater determinant from the infinite orthonormal single particle spin orbitals. We have the wavefunction

$$|\psi\rangle = \sum_{n=1}^{N} c_n |u_{Sn}\rangle. \tag{1.79}$$

If we think of a free electron gas, with the nuclear charges smeared as a positive background—a jellium as a continuum—balanced by an opposite and uniform electron density, then the mean effect is a zero potential. The one-particle Hartree equation (Equation 1.67) is then just the free electron wave

See Appendix H for a discussion of allowed **k** and the distribution of states in the reciprocal and real space.

$$u_k(\mathbf{r}) = \frac{1}{\sqrt{\Omega}} \exp(i\mathbf{k}\cdot\mathbf{r}), \quad \text{with} \quad E_k = \frac{\hbar^2 \mathbf{k}^2}{2m_0}, \tag{1.80}$$

where Ω is the free space volume, and the time dependence is implicit. A ground state has two electrons of opposite spin in these one-particle states. A Fermi level with energy E_F and wavevector $\mathbf{k}_F$ defines the highest state up to which these are filled at absolute zero temperature. Since the volume of **k**-space is $(4/3)\pi k_F^3$, with $2\pi/\Omega^{1/3}$ as the separation between the points of **k**-space that electrons with opposite spins may occupy,

$$N = 2\frac{(4/3)\pi k_F^3}{(2\pi)^3/\Omega} = \frac{k_F^3}{3\pi^2}\Omega \quad \therefore \quad n = \frac{N}{\Omega} = \frac{k_F^3}{3\pi^2}. \tag{1.81}$$

An electron has an effective space of a length scale $r_e = (3/4\pi n)^{-1/3}$, where n is the electron density, and its correspondence with the Fermi wavevector follows as

$$k_F = \frac{1}{\alpha r_e}, \text{ with } \alpha = \left(\frac{4}{9\pi}\right)^{1/3} \approx 0.521, \tag{1.82}$$

which is an indicator of the high level of filling and which is the reason why a jellium description becomes quite valid. The plane waves are eigenstates of the Hartree-Fock operator that had led to the exchange correction term of V_x (Equation 1.78) operating on the state $|u_i(\mathbf{z})\rangle$. This can be viewed as a one-body effective exchange potential operating on the one-particle eigenstate. For the plane waves, this energy function for a wavevector $\mathbf{k}$ is

$$V_{x,\mathbf{k}} = -\frac{4\pi e^2}{\Omega}\sum_{\mathbf{k}'}\frac{1}{4\pi\epsilon_0}o_{\mathbf{k}'}\frac{1}{|\mathbf{k}-\mathbf{k}'|}. \tag{1.83}$$

This term gives the exchange energy by summing over all the other states that are also occupied by electrons. Normalizing by the number of electrons N and avoiding double counting of the interaction by dividing by 2 gives the averaged exchange energy per electron of

$$E_x = -\frac{3}{4\pi}\frac{1}{4\pi\epsilon_0}\frac{e^2}{\alpha r_e} \approx -\frac{1}{4\pi\epsilon_0}\frac{0.458}{r_e} \tag{1.84}$$

for the free electron gas in a plane wave approximation for a metal.

The calculation that this all represents is that the electron doesn't interact with itself but with all others, and we have to self-consistently determine energy under this situation. The electron at $\mathbf{r}$ feels the field from other electrons, but, due to electron-electron repulsion, its presence in our calculation at $\mathbf{r}$ is also repelling these other electrons. So, it has a created a hole in the electron distribution around itself. This is a Coulomb hole due to exchange correlation—an exchange hole. Its presence is also changing the screening of the electron-electron interactions. Figure 1.1, in Section 1.2, is not an unreasonable representation. The exchange hole lowers the net energy. Charge neutrality also means that the electron and the Coulomb hole compensate each other locally. So, in this volume region, net charge still vanished, and the system is neutral. For free metal conducting systems, it has a fair and well-formed description.

V_x of Equation 1.78 is a one-body exchange potential on the one-body eigenstate $|u_i(\mathbf{z})\rangle$. One may view it as an electrostatic potential that arose due to the occupation density

$$o_{HF}(\mathbf{z},\mathbf{z}') = \sum_j o_j \frac{u_j^*(\mathbf{z}')u_j(\mathbf{z})u_i(\mathbf{z}')}{u_i(\mathbf{z})}. \tag{1.85}$$

Integrated over $\mathbf{z}'$, this must be unity, since it represents the existence of this occupied state for which the calculation is being performed. In any N-electron system, this electron at $\mathbf{r}$ is interacting with $N-1$ other electrons. The Hartree potential V_H of Equation 1.69 contains N electrons and one exchange hole. Equation 1.85 says that if $o_j = 1 \ \forall\, j$, then the hole is localized on the electron—a delta function ($\delta(\mathbf{z}-\mathbf{z}')$). But, our previous paragraph argues that this cannot be the case. So, there exists a *broadening*. It is this broadening that is of the order of $\lambda_F = 2\pi/k_F = 2\pi\alpha r_e$ for the free electron gas. The exchange hole in "free" electron metals—alkali being the closest approximation—spreads out a bit beyond the nearest neighbor.

The spreading just beyond makes sense and should be general. Two electrons of same spin cannot be in the same position. The configuration interaction correction to Hartree-Fock used the Slater determinant, which takes this exclusion to heart.

This exchange-correlation hole can now be easily interpreted and understood. The electron density is the probability of finding the electron per unit volume. It is the number of occupied states, and if we normalized it to the states, it is the fraction. With $\mathbf{r}_i$ as the electron positions,

$$n(\mathbf{r}) = \langle\psi \sum_{i=1}^{N} \delta(\mathbf{r}-\mathbf{r}_i)|\psi\rangle, \tag{1.86}$$

where $|\psi\rangle$ is the N-particle wavefunction. Let $n(\mathbf{r},\mathbf{r}')$ be a pair correlation of the squared probability of finding two electrons, one at $\mathbf{r}$ and another at $\mathbf{r}'$, that is,

$$n(\mathbf{r},\mathbf{r}') = \langle\psi \sum_{i\neq j} \delta(\mathbf{r}-\mathbf{r}_i)\delta(\mathbf{r}'-\mathbf{r}_j)|\psi\rangle. \tag{1.87}$$

The term includes any contributions of correlations between electrons. The system Coulomb energy is

$$V_{Coul} = \langle\psi|\frac{1}{2}\sum_{i\neq j}\frac{1}{4\pi\epsilon_0}\frac{e^2}{|\mathbf{r}_i-\mathbf{r}_j|}|\psi\rangle = \frac{e^2}{2}\frac{1}{4\pi\epsilon_0}\int_\Omega \frac{n(\mathbf{r},\mathbf{r}')}{|\mathbf{r}_i-\mathbf{r}_j|}\,d\mathbf{r}\,d\mathbf{r}'. \tag{1.88}$$

The Hartree and Hartree-Fock approach didn't account for correlation. This means that

$$n(\mathbf{r},\mathbf{r}') = n(\mathbf{r})n(\mathbf{r}') \tag{1.89}$$

for the Hartree and Hartree-Fock treatments.

When configuration interaction—correlation—is included, one may rewrite the pair correlation function in a first order expansion as

$$n(\mathbf{r},\mathbf{r}') = n(\mathbf{r})n(\mathbf{r}')\left[1+\alpha(\mathbf{r},\mathbf{r}')\right], \tag{1.90}$$

where $\alpha(\mathbf{r},\mathbf{r}')$ is a correlation parameter that contains correlation's consequences. Since an electron at $\mathbf{r}$ interacts with $N-1$ other electrons in the $\mathbf{r}'$-space, $\int n(\mathbf{r},\mathbf{r}')d\mathbf{r}' = N-1$. Therefore, because of Equation 1.90,

$$\int \alpha(\mathbf{r},\mathbf{r}')n(\mathbf{r}')\,d\mathbf{r}' = -1. \tag{1.91}$$

This is the mathematical expression for stating that the electron at $\mathbf{r}$ has an exchange-correlation hole enveloping it. It arises both due to exchange and due to correlation, and it gives us an intuitive way of looking at exchange and correlation.

To distinguish exchange's and correlation's consequences in the creation of the hole, consider spin and we can look at what happens with aligned and anti-aligned spins. These are both conditions where Equation 1.91 is still valid. If a system had N electrons composed of $N_\uparrow$ (spin up) and $N_\downarrow$ (spin down) electrons, the electron at $\mathbf{r}$ with up spin will interact with $N_\uparrow - 1$ of up-spin electrons and $N_\downarrow$ of down-spin electrons. Therefore, the integral of Equation 1.91 split up is

$$\int \alpha_{\uparrow\uparrow}(\mathbf{r},\mathbf{r}')n_\uparrow(\mathbf{r}')\,d\mathbf{r}' = -1, \quad \text{as before, and}$$
$$\int \alpha_{\uparrow\downarrow}(\mathbf{r},\mathbf{r}')n_\uparrow(\mathbf{r}')\,d\mathbf{r}' = 0. \tag{1.92}$$

An exchange hole exists (with this integral of -1) for the up-spin electron at $\mathbf{r}$, even with the correlation effect present. And, for interaction with electrons of opposite spin—no correlation—the local screening hole will have to be compensated for by the charge on the surface of the system so that the second part of Equation 1.92 is satisfied. At nanoscale, this effect will be of significance.

This discussion suffices for now to indicate that predictive description of large-N systems, such as solids, will require care. The energy state, the transitions under perturbations, will relate to the Hamiltonian description and its solution under the constraints of the circumstances. Bandstructure—energy states of the electrons—calculation will require related care. We will return to this calculation to summarize the different approaches—their salient points and applicability and limitations—in Chapter 4. Here, we continue with our discussion of approximation methods and now look at screening by the mobile charge.

1.8 Screening

How do we tackle spatial variation in electronic charge? It exists, since perturbations exist. The simplest type of perturbation may be a space charge, where electrons will locally rearrange themselves to minimize the energy of interaction by attempting to screen the perturbation. A positive space charge will attract, and a negative space charge will repel. The simple charge-induced perturbation may be static, but it could also be dynamic if it arises in an oscillatory phenomenon interacting with the screening electrons. If the electron states are filled up to some states up in energy—the Fermi energy of $\sim E_F$ with a Fermi wavevector of $\mathbf{k}_F$—then it is the electrons around the Fermi energy that are most likely to respond, since both filled and empty states are available around it. We have now created not only a dynamic condition—depending on the frequency scale—but also one where the consequences will be felt nonlocally, as the electrons at Fermi energy provide an oscillatory response. All these interactions will be mediated by permittivity, and which kind—static, intermediate or high frequency—will depend on the conditions of perturbation. A few remarks on the screening are therefore in order to understand how the particles moving around in the crystal respond.

The permittivity reflects the medium's ability to withstand the applied electric field as represented by displacement, or equivalently, polarization. In vacuum, this is quite clear. $\epsilon = \epsilon_0$, which does not depend on the frequency of the applied electromagnetic stimulus. In an atom, isolated in vacuum, when determining the orbitals, et cetera, without stimulus, it is again this $\epsilon = \epsilon_0$. Apply an electric field, and the atom responds by polarizing—slightly or significantly—and the response is concentrated in the outer orbitals, with electrons in the core orbitals shielded by the valence. Now, the permittivity needs some care and thought. In a semiconductor, the electrons or their anti-quasiparticle hole, sample the environment of the crystal. The binding energy, that is, the ionization energy of donors and acceptors (shallow hydrogenic) now must be the permittivity of the crystal. And it is the static permittivity, since this particle's binding exists in an unstimulated environment. Place the donor under very confined conditions in the semiconductor, and the permittivity must account for the change of the environment. In an unconfined crystal, the permittivity will change as a function of frequency, since the medium's response is changing. If the nearly free electron and an electromagnetic stimulus are interacting in the crystal environment, then this interaction will need to account for the frequency dependences and the time extent of the interaction, where phonons may also be important. We tackle this later. If an electromagnetic signal causes an electron transition from within the core, then, due to where it is from and the rapidity with which the change takes place up in the atom's higher orbitals—still localized—the permittivity is still free space permittivity. But, an excited electron localized at an atom relaxing into a delocalized state in the crystal environment will now need a more complicated permittivity analysis. So, use permittivity with care. The solution is generalizable. In electrical engineering texts, it appears as a question of how electrons screen a potential disturbance, that is, a field, such as when the jellium of

1.8.1 Debye-Hückel and Thomas-Fermi screening

The static screening problem—largely an electronic many-body problem with electrons interacting with fields arising from other charges—goes back to Debye and Hückel, who explored it for the case of electrons interacting with other electrons. Figure 1.1 was an example showing an electron with the exclusion zone due to correlation around it arising from Coulomb repulsion. This picture can be seen—within the jellium approximation—as an illustration of Debye-Hückel screening.

The Poisson equation, with an electron located at $\mathbf{r}_0$, a charge distribution due to electrons of $-en(\mathbf{r}, \mathbf{r}_0)$, and a uniform positive neutralizing background of concentration en (the ionic jellium), is

$$\nabla^2 V(\mathbf{r}) = -\frac{1}{\epsilon}\left[-e^2\delta(\mathbf{r}-\mathbf{r}_0) - e^2 n(\mathbf{r}-\mathbf{r}_0) + e^2 n\right], \qquad (1.93)$$

where V is the electrostatic potential. To include the correlation effect, we write the pair correlation function $g(\mathbf{r}|\mathbf{r}_0) = n(\mathbf{r}|\mathbf{r}_0)/n_0$. This function gives the probability of finding an electron at $\mathbf{r}$, given that there is another electron at $\mathbf{r}_0$. This pair distribution function

vanishes at $\mathbf{r} = \mathbf{r}_0$, that is, the probability of an electron vanishes, and it asymptotes to 1 at infinity. We now rewrite this equation as

$$\nabla^2 V(\mathbf{r}) = -\frac{e^2}{\epsilon}\left\{\delta(\mathbf{r}-\mathbf{r}_0) + n\left[g(\mathbf{r}|\mathbf{r}_0) - 1\right]\right\}. \tag{1.94}$$

We need to find the pair correlation function that solves this many-body simplified problem. This is possible at many different levels of accuracy. The equation as written holds true whether we need to include quantum constraints or not for physical charged particles. The quantum character of electrons introduces just additional non-triviality.

First, consider non-quantum classical conditions. The Boltzmann distribution applies. At very small $\mathbf{r}$ referenced to $\mathbf{r}_0$, this approximation will fail, but, at far enough distances where one may linearize the correlation function (a Poisson-Boltzmann function) of

$$g(\mathbf{r}) = g_{PB}(\mathbf{r}) = \exp\left[-\frac{V(\mathbf{r})}{k_B T}\right], \tag{1.95}$$

the solution will be quite accurate. Here, the position of perturbation at $\mathbf{r}_0$ is implicitly understood. With the linearization, the Poisson equation reduces to

$$\nabla^2 V(\mathbf{r}) = -\frac{e^2}{\epsilon}\delta(\mathbf{r}) + \frac{ne^2}{\epsilon k_B T}V(\mathbf{r}), \tag{1.96}$$

whose solution is

$$V(\mathbf{r}) = \frac{e^2}{4\pi\epsilon r}\exp\left(-\frac{r}{\lambda_{DH}}\right), \tag{1.97}$$

a form similar to that of the Yukawa potential encountered with massive bosons and is the static and spherically symmetric solution of the Klein-Gordon equation.

$$\lambda_{DH} = \left(\frac{\epsilon k_B T}{ne^2}\right)^{1/2} = \lambda_D \tag{1.98}$$

is the Debye-Hückel or just plain Debye screening length. This approximation is a linear screening approximation from that Boltzmann expression.

Now assume that the linearization is acceptable, but pair correlation as employed is not. We should still be able to use the thermal equilibrium condition, which brings about the equilibration of electrochemical potential. Electrostatic potential and chemical potential compensate each other. So, now, we have a screening length that is

$$\lambda_{scr} = \left[\frac{\epsilon\,(\partial E_F/\partial n)\,|_T}{e^2}\right]^{1/2}. \tag{1.99}$$

positive charge is uncovered. Edges of transition region are an example. So is the case where suddenly one has a sudden change in doping.

Appendix E and F have a short primer on thermodynamics and the statistical implications reflected in the distributions functions. A Boltzmann distribution will be a reasonable approximation where classical conditions are a good description. Electron description in non-degenerate semiconductor conditions is one example, even if electrons are quantum particles, and this quantum aspect is very necessary in describing the nature of their states in the semiconductor. Boltzmann distribution works pretty well here, since a deep potential also increases the probability of finding a particle there. The exponential arrives from the phase space.

Debye (Petrus Josephys Wilhemlmus Debije) is being given top billing. Electrical engineering's semiconductor device literature, with atoms as an afterthought except in reliability or processing discussions, largely ignores Hückel. Hückel, of course, finds a pride of place in chemistry. So does Debye, whom we first encounter through the Debye model for the low-frequency phonon contribution to specific heat, but which is only one of many significant contributions. He was Sommerfeld's student before Sommerfeld's Munich period. As with many of the European scientists who came of age in the pre-war years, there exists considerable tension and ambivalence in matters of life where science and society intersect. Debye became the head of the Kaiser-Wilhelm Institute in Berlin when Einstein left for the *USA* in the 1930s, and Debye himself moved to the *USA* just before the Second World War. He was the head of the Deutsche Physikalische Gesellschaft (the German physical society) from 1937 to 1939 and was among those who helped Lise Meitner escape, but one can also find letters that pay obeisance to powers that be, which in this case was Adolf Hitler, and an untenable situation brought on by his daughter's decision to stay back in Germany. Even Fermi was a member of the Fascist party.

If the conditions are degenerate, and a Fermi gas description is more appropriate, then we obtain the Thomas-Fermi screening length of

$$\lambda_{TF} = \left(\frac{2\epsilon E_F}{3ne^2}\right)^{1/2}, \tag{1.100}$$

for three-dimensional conditions, since, in degenerate conditions, $E_F \propto n^{2/3}$.

To paraphrase the great English mathematician, George Hardy, from his essay "A mathematician's apology," pure mathematics is the most beautiful mathematics, since it has no usefulness. And it is because of this very uselessness that the pursuit of pure mathematics cannot be misused to cause harm.

Figure 1.10 shows the magnitude of the screening length scale in *Si* as a function of the carrier concentration. At small carrier concentrations, the electrons are tens of *nm* or even more distant from the potential disturbance. At 10^{17} cm^{-3} carrier concentration, this screening length is of the order of 10 *nm*, and it is in the Debye-Hückel limit. As the semiconductor becomes degenerate (the effective density of states $\mathcal{N}_c \approx 2.8 \times 10^{19}$ cm^{-3} at room temperature), gradually, with degeneracy, the screening length scale bends over to the Thomas-Fermi limit. At the highest concentrations possible in *Si*, this screening length scale is a fair fraction of a *nm*, so spread over several atom spacings.

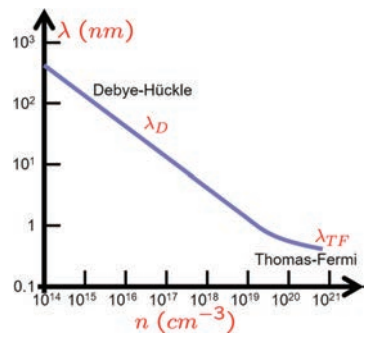


Figure 1.10: Electron screening length in *Si* at 300 K as a function of carrier concentration. The screening length scale can span over 100s of *nm* in the Debye-Hückel limit to sub-*nm* in the Thomas-Fermi limit.

1.8.2 *Static versus dynamic screening, and a note on permittivity*

INTERACTIONS CAN BE SLOW AND FAST. How screening will happen will depend very much on the pace of this interaction. A spatially fixed charge with electrons around it screening the perturbation is a static perturbation. The permittivity mediates it, and the electrons screen, present in these surroundings, through the static dielectric response. *The Debye and Thomas-Fermi screening are very applicable to such static circumstances, and it is the static dielectric constant that is applicable.* There are, however, circumstances where this will need modification. An electron in a confined condition, that is, with surrounding potential barriers that keep it in narrow atomic-scale regions, also feels the barrier and its behavior is not that of the electron in a crystalline surrounding of long-range periodicity. Its probability densities have changed, as have the state description and the energy and the wavevectors. The permittivity will change and needs to account for the change of the surroundings at such small dimensions, just as the eigenfuction description changes, leading to changes in energies and even the applicability of the mass assigned to the state of the electron.

Interactions can also be fast. Consider the absorption of phonons representing the quantization of crystal vibrations. A phonon energy of 50 *meV* is a 10 *THz* oscillating quantum. Since the time lengths of interactions at any energy change of ΔE also has time interaction

of $\hbar/\Delta E$, that optical phonon interaction time is 100 *fs*. These are dynamic conditions, and the permittivity, which reflects the medium's permission to allow electrical changes to take place, will also have a frequency dependence. It is clear from this argument that screening's mediation in the interaction will be influenced by permittivity, which depends on the polarization response of the medium and the extent of the medium interacting, and all these changes occurring simultaneously will need to be reconciled. Polarization arises in many sources, electrons are moving around and respond to the fields, and an oscillating field causes the charge cloud to respond too by oscillating under forced conditions. These are plasmons affecting the permittivity locally. Under certain conditions, there will also exist long-range consequences spatially. Vibrating atoms—ion and charge with their different inertia—also respond, and this leads to an ionic response. So, the general nature of the permittivity is to drop with increasing frequency. There are, however, regions in between where resonances in the forcing function's frequency and the polarizing response may cause large-scale changes of increase or decrease either side of the resonance. In these situations, we will have to modify substantially the nature of screening—as in the Debye or Thomas-Fermi static interpretation—and incorporate the permittivity's behavior at high frequency, even as there is a background asymptotic behavior that is incorporated in the static response.

The reader might wish to ponder why this term is called permittivity, or, for that matter, the origin of the word "displacement," in these field relationships. Permittivity, tied to the nature of the polar character of electrical fields, is a negpermittivity. A higher permittivity, arising with higher polarization, reduces electric field.

There is one other aspect of permittivity and electrons that needs emphasis in a Hamiltonian and perturbation discussion related to the conditions in time and environment that the mathematical description must account for. It is related to the time scale of rapidity of the interaction. An example is the absorption of light in a semiconductor with an electron moving from the valence band—a bonding-based quasi-continuum—to the conduction band—an antibonding quasi-continuum. The electron occupies states of these quasi-continuums that are defined by the electron's environment of being in the midst of this bonded collective of atoms. This forces its $E(\mathbf{k})$ dispersion, but the electron that transitions has changed states defined by the crystal because of the interaction with the photon, by absorbing the energy and the vanishingly small momentum. The crystal here is only a phase space locale for the states. The photon-electron interaction happens with the electron's free space mass, while the states reflect the effective mass that reflects the $E(\mathbf{k})$ dispersion. This argument also holds for the photon processes where phonons—the quantized atomic vibrations—are also involved as in indirect bandgap materials.

A different—and yet analogous—circumstance is when an electron confined very locally in an impurity that is not shallow—a deep level—has an electron move from one energy state to another energy state, while still remaining confined. An example is a $Au^{2+} \mapsto Au^{3+} + e^{-}$ transition. The electron was in a local state, not spread out over the crystal. It was confined to the atom. The electron mass relevant in the description here is the free space mass, and the permittivity is free space permittivity. On the other hand, if the impurity is a shallow hydrogenic state, that is, a few *meV* separated from other conducting states, with properties very similar to that of the host crystal atoms that gave rise to those conducting states, then the electron really feels its presence in the crystal. It is not that localized, and an effective mass of the crystal, and a permittivity of the crystal, will be a more apt set of parameters. Now, one can imagine potential impurities and defects, where the behavior may very well be in between—neither entirely localized nor entirely delocalized—and one may have to use either an interpolation or a more rigorous description.

This dynamic behavior of permittivity and of electrons, in a large-gap material, is more easily deconvolved. If permittivity changes are arising due to plasmonic response—the response of the charge cloud—then the plasmons can be incorporated into the electron-phonon scattering through an effective treatment of the dynamic screening. In large-gap materials, there is just one of the bands contributing to the conducting carriers that one needs to worry about. This treatment will also have to change when one confines carriers to a plane or to a quantum wire, because carriers are not free to move in all the directions for screening. Graphene is a zero bandgap material. Now this dynamic screening for electron-phonon scattering will become considerably more complicated.

1.9 *Summary*

THIS CHAPTER WAS AN INTRODUCTION to several of the common principles, techniques and approximations that will be employed throughout the text, with an emphasis on their implications, context and physical meaning so that we may employ them with due care and restraint. In this approach, our quantum view, its emergence into the classical view under many of the natural world's sizes—in dimensions as well as the number of participating entities and their interactions—and others such as the statistical and informational views, all have an important role. This last theme—of information—will be deployed in Chapter 2, where we also bring in thoughts

from statistical mechanics, which itself is spread throughout the text, toward understanding two of the most important pervasive themes in any study of natural phenomena: those of entropy and energy. The present chapter particularly stressed the quantum underpinnings toward understanding statics and dynamics at the quantum scale, with emphasis on many particles coming together in an ensemble, be it an atom, a molecule or an atomic assembly, and the methods for solving the Hamiltonians; that is, the energetics.

Understanding the energetics of multiparticle systems is essential to developing an understanding of the properties during interactions that define the internal properties and response of a semiconductor. Much skepticism, care and *understanding*—not just grasping the ideas—is important for prudent use, treatment and reaching a result that holds validity over a range of conditions. As an introductory chapter integrating partly the material that students will need to learn and the outlines of the underlying physical principles and techniques that will be employed throughout this text, we started with a discussion of Hamiltonians and Lagrangians as functional tools for unraveling the energetics. We sketched the broader nature of the Hamiltonian description of electrons and atoms in an assembly such as a metal, a semiconductor or a molecule, and then reduced it to the problem of understanding the electrons' interactions. Before embarking on the approximate solution techniques of this problem, we segued into the different perturbation approaches that are sprinkled throughout the text, since problems in general, and certainly the many-body problems, cannot usually be exactly solved. But the solutions can be approached via perturbation techniques. We illustrated the first order perturbation approach and applied it to both a time-dependent and a space-dependent perturbation. The former was useful in showing the Golden rule and the limits of its applicability. The latter was useful in illustrating scattering's quantum-mechanical origin and its classical fitting. Another important approximation technique was the use of the adiabatic or Born-Oppenheimer approximation, and, through it, the approach to situations where fast and slow phenomena interact. The adiabatic approximation is very important to calculating phonon-based; that is, atomic movement-based interactions with those of electrons. The former are slow, and the latter are fast. Energies may exist in multiple modalities—atomic bonding and vibrational and electronic kinetics, for example—and we outlined how a configuration coordinate diagram lets us see the slow and fast together in this energy transformation.

At this point, we returned back to solving the multiple electron problem to bring out the nature of many-body interactions. The

Understanding is an internalization—deeper than *knowing, being aware of* or other, similar terms. Approximation and model usage requires understanding and skepticism. In this, it is no different than in the populated world around us, where much marketing, manipulation and myth building abounds. Take history. The New Testament is so different from the Old Testament. In India, a different narration around Rama, a revered godly king, brings out violence and McCarthyism. Passage of time or distant lands is not necessary for creating mythologies. Take a statistic from the Second World War of the 20th century.

	Major event	Innocent deaths
Stalin	Russian famine	$6–7 \times 10^6$
Hitler	Death camps	2.8×10^6
Tôjô	East Indies	$2.4–4.0 \times 10^6$
Churchill	Bengal famine	$2.5–5.5 \times 10^6$
Truman	Atom bomb	$0.13–0.22 \times 10^6$

Stalin cannot be criticized in Russia, neither can Churchill in the West. The Soviet state's industrial and poor folks' transformation, as well as the war's transformational fight, is Stalin's contribution. Churchill's is his steadfastness in the war. Stalin's was the brutal killing of innocents in his home country. Churchill's was a tribal and racist view of freedom—Woodrow Wilson-like—as a white European prerogative. Even de Gaulle of France marched right back into Indochina, culminating in the Vietnam War. Narratives should always be looked at with caution. Stalin and Churchill stood up for their lands, and for that both should be lauded, but not worshipped. Indians still remember that, for the false promise of freedom, nearly 75,000 young men fell even in the First World War and have been forgotten, several during Churchill's Gallipoli folly. No site marks the forced fighters—the unknown soldiers—of the third world, even as a famous Western journalist declares himself and his kin the greatest generation. It turns out the

Hartree approximation is one where only Coulomb interaction is tackled and the electron as a point particle is a secondary thought in what is essentially a classical calculation. That the electron is a quantum particle—a fermion—and therefore requires a different wavefunction under exchange was brought in through the use of the Slater determinant and the evolution to a Hartree-Fock approximation. Exchange interaction is the result of Coulomb interactions between the electrons under the quantum constraints from spin. The Coulomb interaction becomes spin dependent under the constraint that the wavefunction of any pair of electrons must be antisymmetric with respect to any interchange of spatial coordinates and spins. When spins are parallel, the coordinate part must be antisymmetric. So, parallel spin pairing reduces the probability of two electrons of being spatially close, compared to the probability when possessing antiparallel spin. Parallel spin electrons, when more separated in space, have less repulsion and this lowers the energy of electrostatic interaction. Spin and orbit also interact, and this we will look at carefully in our discussion of valence bandstructure as well as defect-mediated point perturbations. In situations where the spin-orbit energetics is important, the velocity, as well as the structure of the wavefunction solution in a crystal assembly, which leads to the description of the motion of electrons on the atomic scale, affects the interaction and the electron g factor. We ignore nuclear spins, since nuclear magnetic moment is small ($\sim$2000$\times$ smaller) than that of an electron, and its consequences are through perturbations in semiconductors where spin-dependent transport and other phenomena are important.

man was also a sexual predator. This is the difference between "getting it" and "understanding it," where using the learning to solving a general—and not special—problem matters. Use of approximations and models requires tremendous care.

The g factor should be distinguished from the gyromagnetic ratio, which is the ratio of the magnetic moment to the angular momentum. g is dimensionless. The electron has charge and spin, but it is not quite appropriate to view it as an object with literal rotation about an axis. The g factor is the dimensionless number that modifies the gyromagnetic ratio as determined by the classical definition.

The final and very important, particularly so for nanoscale, interaction is that of correlation. An electron does not interact with itself. It only interacts with others. So, accounting for an electron in a Hamiltonian in the middle of other electrons is complex. If we take the electron away, it is a different problem. If we place the electron in, then it is also a different problem, since now the other electrons are responding to the presence of this electron. Mean field, as in the first case, is not representative completely, since the presence of the electron matters in the arrangements of others. Neither is the latter, since what the true energy picture needs is the arrangement of electrons where this electron takes into account exchange and correlation. Two spin-up electrons cannot be present simultaneously in identical space, but electrons with opposite spins can. The spin-up second electron has vanishing presence, while the spin-down electron's presence has been accentuated. The first has a hole, while the second formed a

Spin-dependent transport is an important subject area for devices and is discussed in depth in S. Tiwari, "Nanoscale device physics: Science and engineering fundamentals," Electroscience 4, Oxford University Press, ISBN 978-0-198-75987-4 (2017).

mound. This configuration also needs to be accounted for, and we outlined how one may do this approximately.

Another important analytic theme related to how these electrons are behaving in the solid is how they self-consistently respond to the created conditions. Electrons screen because of the Coulomb interaction, but under all the rest of the constraints that we just discussed. If the perturbation is static, and the electron population small—non-degenerate—then the Debye-Hückel, or Debye, for short, length scale suffices in how the potential perturbation is screened. If it is large and degeneracy prevails, we observed Thomas-Fermi screening. These will all be mediated by static permittivity. And if it is a rapid perturbation, we must also bring dynamic permittivity: the high-frequency aspects of the electronic or ionic or other polarization responses. And, into this, one must also take into account the nature of the behavior of the electron. Is it feeling this polarization environment or not? So, both the permittivity and the mass must reflect the realities of the dynamic perturbation.

1.10 *Concluding remarks and bibliographic notes*

This chapter was an introduction to several of the common techniques, principles and approximations that will be employed throughout the text, with an emphasis on their implications, context and physical meaning, so that we may employ them with due care and restraint. Solid state has a longer history and wider context than the subject of semiconductors, and the objective here was to introduce a few of the main techniques and the scope of the nature of the techniques that are particularly important for semiconductors.

Solid state has been the subject of numerous texts. A number of books have been standard bearers; historically, first and before all, are the conceptual and analytic discussions by Ziman. The first[1] is a very readable discussion at the senior undergraduate level, with an emphasis on scattering and transport as well as a semiconductor bent. The second[2], although it has much in common with the first book, has a more diverse treatment toward solid state, with magnetism, ferroelectricity and superconductivity as the ending points. This book is now in its second edition, having been revised in the early 1970s. Both of these books are worth reading so many decades after their writing. Another text, from the same time period—well, a little earlier—is the text by Peierls[3], which too has a treatment of phenomena from electrical and thermal conductivity, working from the behavior of electrons and phonons and ending in

[1] J. M. Ziman, "Electrons and phonons," Oxford (1960)

[2] J. M. Ziman, "Principles of the theory of solids," Cambridge, ISBN 0-521-29733-8 (1964)

[3] R. E. Peierls, "Quantum theory of solids," Oxford, ISBN 19-850781-X (1955)

broader solid-state topics as they were understood during that time period. A more advanced treatment from this period is the book by Pines[4]. This is a mathematically detailed text. One additional book very worthy of note, similar to Pines in its advanced treatment, is by Kittel[5]. Both Pines' and Kittel's first editions appeared in 1963. I reference these books since they have stood the test of time and are worth reading to get a perspective, from quite different ways of looking, by many of the luminaries of the early days of the marriage of quantum and the solid state.

A set of solid-state texts that have established themselves as standard texts in the undergraduate and early graduate curricula around the world, from this side of the Atlantic, are those of Kittel[6], whose first edition appeared in 1963, and Ashcroft and Mermin[7], whose first edition appeared in 1976. They are different in style from each other, but both have a very carefully and clearly written exposition.

Quantum mechanics is a subject with an even vaster collection of texts. Two that have become standards are one at the introductory level, by Griffiths[8], and one that is a little more advanced (intermediate), by Sakurai[9], both of which have gone through several incarnations. These texts are quite lucid in their exposition of the perturbation theory, the Golden rule, and the Golden rule's limitations. A mathematically sophisticated treatment is in the series of books by Landau and Lifshitz, which all physics students have since they encompass much of the formalism of physics through the 1960s. The volume devoted to quantum mechanics[10] is a translation by J. B. Sykes and J. S. Bell. Any book that Bell spent time translating has to stand head and shoulders above the rest. Elsewhere, Bell also likes the text by Gottfried[11], whose first edition is from 1966 and whose copy at *CERN* Bell found very well worn, and worth discussing in a work entitled *Speakables and unspeakables in quantum mechanics*, a subject that Bell had much to contribute to through his Bell inequalities that are so illuminating.

A book from the early times with an excellent discussion of the finer points embedded within the formulation of quantum mechanics and its application to the description of solids is the book by Slater[12]. Another book mixing solid-state and quantum matters and which is a favorite of mine for its lucidity, a stronger bending toward semiconductors, and restrained and yet thorough discussions is the one by Harrison[13].

There are a few additional books that are quite representative of the physical intuition necessary in this transition from our observational classical thinking to the reality of the quantum-mechanical.

[4] D. Pines, "Elementary excitations in solids," Perseus, ISBN 0-7382-0115-4 (1999)

[5] C. Kittel, "Quantum theory of solids," John Wiley, ISBN 0-471-62412-8 (1987)

[6] C. Kittel, "Introduction to solid state physics ," Wiley, ISBN 13 978-0471415268 (2004)

[7] N. Ashcroft and D. Mermin, "Solid state physics," Saunders, ISBN 13 978-0030839931 (2003)

[8] D. J. Griffiths, "Introduction to quantum mechanics," Pearson, ISBN 0-13-191175-9 (2005)

[9] J. J. Sakurai, "Modern quantum mechanics," Addison-Wesley, ISBN 0-201-53929-2 (1967)

[10] L. D. Landau and E. M. Lifshitz, "Quantum mechanics," Butterworth-Heinemann, ISBN 13 978-0750635394 (2003)

[11] K. S. Gottfried, "Quantum mechanics," ISBN 0-387- 95576-3, Springer (2003)

[12] J. C. Slater, "Quantum theory of atomic structure," **1**, McGraw-Hill (1960)

[13] W. Harrison, "Sold state theory," Dover, ISBN 0-486-63948-7 (1979)

Dyakanov[14] discusses the different spin-based issues that appear in semiconductors. Spin has consequences through the Pauli principle and through exchange interactions. In semiconductors, the major manifestations include the spin-orbit interaction and the role it plays in optical transitions.

[14] M. I. Dyakanov, "Spin physics in semiconductors," Springer, ISBN 978-3-540-78819-5 (2008)

Hartree, Hartree-Fock and correlations have occupied considerable space in the discussion of this chapter, since these really represent the major collection of ways that we treat a multi-electron assembly, and, as the ensembles get smaller, from nanoscale to a molecule, the consequences as the quantum nature manifests itself in different pronounced ways. A good book for this discussion is the one by Delerue and Lannoo[15]. It discusses the general modeling techniques, their usage in quantum-confined systems and the variety of properties that result. Its early exposition is quite close to the several points we have emphasized, but it goes quite a bit beyond. Another good text for understanding the Hartree-correlation spectrum of subjects is the text by Kohanoff[16]. It develops the subject all the way through to density functional theory and Car-Parrinello techniques that we did not dwell on. Density functional theory will appear in a minor form in the discussion of bandstructures (Chapter 4).

[15] C. Delerue and M. Lannoo, "Nanostructures," Springer, ISBN 3-540-20694-9 (2004)

[16] J. J. Kohanoff, "Electronic structure calculations for solids and molecules," ISBN 13 978-0521815918, Cambridge (2006)

1.11 *Exercises*

1. The Maxwell's equations can be transformed into a simpler group under source-free free space conditions; that is, with $\mathbf{J} = 0$, $\rho = 0$, $\mathbf{D} = \epsilon_0 \boldsymbol{\mathcal{E}}$, $\mathbf{B} = \mu_0 \mathbf{H}$ and $1/c^2 = \mu_0 \epsilon_0$. For this simplified free space source-free form,

$$\text{General form} \mapsto \text{Source free and free space form,}$$

$$\nabla \cdot \mathbf{D} = \rho \mapsto \nabla \cdot \boldsymbol{\mathcal{E}} = 0,$$

$$\nabla \cdot \mathbf{B} = 0 \mapsto \nabla \cdot \mathbf{B} = 0,$$

$$\nabla \times \boldsymbol{\mathcal{E}} = -\frac{\partial \mathbf{B}}{\partial t} \mapsto \nabla \times \boldsymbol{\mathcal{E}} = -\frac{\partial \mathbf{B}}{\partial t}, \text{ and}$$

$$\nabla \times \mathbf{H} = \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t} \mapsto \nabla \times \mathbf{B} = \frac{1}{c^2}\frac{\partial \boldsymbol{\mathcal{E}}}{\partial t}.$$

The Lagrangian function for the free space problem is

$$\mathscr{L} = \frac{1}{2}\epsilon_0 \mathcal{E}^2 - \frac{1}{2}\mu_0 H^2 = \frac{1}{2}\epsilon_0 \mathcal{E}^2 - \frac{1}{2\mu_0}B^2.$$

Show that, in the presence of external sources ρ and $\mathbf{J}$, and a generalized medium, the Lagrangian function has the form

$$\mathscr{L} = \frac{1}{2}\epsilon\mathcal{E}^2 - \frac{1}{2}\mu H^2 + \alpha\rho\phi + \beta\mathbf{J}\cdot\mathbf{A}.$$

Find α and β so that Euler-Lagrange equations reproduce the Maxwell's equations in the presence of sources. **[M]**

See Appendix D for a broader discussion of the Lagrangians and variational methods.

2. Consider a Hamiltonian operator $\hat{\mathscr{H}}$ that has discrete eigenvalues. It is also Hermitian, so

$$\langle\psi|\hat{\mathscr{H}}|\phi\rangle = \langle\phi|\hat{\mathscr{H}}|\psi\rangle^*$$

by definition. Show that
- the eigenvalues for this Hamiltonian are real, and that
- the eigenfunctions of $\hat{\mathscr{H}}$ that correspond to different eigenvalues must be orthogonal to each other. **[S]**

3. This problem is to emphasize the power of the Golden rule, and a view of scattering that we mentioned but did not discuss in much detail—one that is particularly apropos of nanoscale devices with a finite and low number of scattering events, and also in mesoscopic transport. We will look at transmission and reflection at a barrier, the working example of which is shown in Figure 1.11, by two methods. We consider just a one-dimensional structure where waves transmit or reflect back, here due to an incident wave, $\exp(ik_iz)$. One can look at the net effect of the transmitted wave $t_B\exp(ik_iz)$ and the reflected wave $r_B\exp(ik_iz)$ as arising from multiple transmissions and reflections as the wave rattles back and forth between the two non-adiabatic discontinuities—the net effect being a convergent series, as shown in the figure. Show that the transmission coefficient $\mathscr{T}_B$ and the reflection coefficient $\mathscr{R}_B$ arising from the barrier can be written as

$$\mathscr{T}_B = |t_B|^2 = \frac{\mathscr{T}^2}{1 + R^2 - 2R\cos(2k_2d)},$$

$$\mathscr{R}_B = |r_B|^2 = \frac{2\mathscr{R} - 2\mathscr{R}\cos(2k_2d)}{1 + \mathscr{R}^2 - 2\mathscr{R}\cos(2k_2d)},$$

which add to unity and where $\mathscr{T}$ and $\mathscr{R}$ are the transmission and reflection coefficients for individual step. Now use the Golden rule to calculate $\mathscr{R}_B$ and compare with this result, remarking on the conditions under which the two are in accord. **[S]**

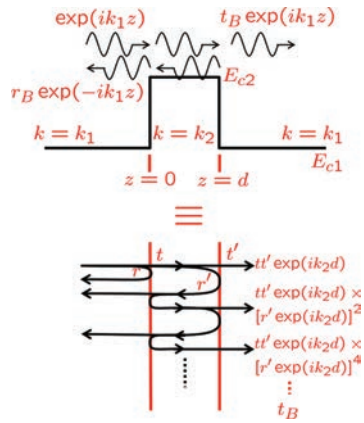


Figure 1.11: Multiple transmissions and reflections at a potential barrier, leading to net transmission and reflection.

The barrier is a perturbation!

4. For a system of particles of mass m in state ψ, the particle flux (number per unit time per unit perpendicular-to-motion area) is given by

$$\mathbf{S} = \frac{\hbar}{2im}\left(\psi^*\nabla\psi - \psi\nabla\psi^*\right).$$

- Show the validity of this expression, starting from the momentum operator, and
- show that, in the classical limit—a beam of free particles at velocity v—the one-dimensional expression gives

$$S = v \times \text{particle density.}$$ **[S]**

This is how current (both quantum-mechanical tunneling and normal) can be calculated across a potential barrier. This problem shows that the correspondence principle holds here (quantum-mechanical expressions reducing to classical ones in the asymptotic limit).

5. Following Appendix E, show that, for bosons with $E_k = \hbar^2 k^2/2m$—the classical limit from quantum with $\hbar \to 1$—that the chemical potential μ is always negative. **[S]**

6. Again, following Appendix E, find the asymptotic behavior of the chemical potential as a function of temperature ($\mu(T)$) and show that the distribution function reduces to the Boltzmann distribution as $T \to \infty$. **[S]**

7. If the energy of all configurations is the same, show that the entropy as arrived at in Appendix E is equal to the logarithm of the number of configurations. **[S]**

8. Again, following Appendix E, show that a boson gas with n particles in state $|k\rangle$ will have an entropy of

$$S = -k_B \sum_k [n_k \ln n_k - (1 + n_k)\ln(1 + n_k)],$$

and that, for fermion gas, it will be

$$S = -k_B \sum_k [n_k \ln n_k - (1 - n_k)\ln(1 - n_k)].$$ **[S]**

9. The notions of Appendix E can also be applied to photons as bosons. Does a collection of photons—as a gas—have entropy? In thermodynamic equilibrium, as in blackbody radiation, what is the chemical potential of a photon? **[M]**

10. Let us make some order of magnitude energy estimates based on quasi-classical-quantum fitting to see the adiabatic approximation's use. We explore the situation of the applicability of electrons not undergoing transitions between stationary states. Take a molecular system. Such a system will have three different types of motion: electronic, nuclear vibration and rotation. If a is an interatomic distance, it is a length scale for electron movement, so $E_e \approx \hbar^2/2ma^2$ is an electronic energy scale. Estimate this for $a = 0.1 \quad nm$. The nuclear motion, vibrational, has an energy estimate of $E_{vib} = \hbar\omega_{\mathbf{q}}$. A mass M, moving a distance a at a frequency $\omega_{\mathbf{q}}$, has an energy of $\sim M\omega_{\mathbf{q}}^2 a^2$. Such a motion—of distance a—would remove the atom from the molecule. It is

bond-breaking energy, since chemical bonding is the binding of atoms as molecules through shared electrons. So, it is of the same order as E_e, that is, $E_e \approx M\omega_{\mathbf{q}}^2 a^2$. Hence, $E_{vib} = \hbar\omega_{\mathbf{q}} \approx \sqrt{m/M}E_e$. Rotation, being related to angular momentum, is quantized by the action $\hbar$. So, if I is the inertia, and L the angular momentum, then the rotational energy is $E_r = L^2/I \approx \hbar^2/Ma^2 = (m/M)E_e$. The energies are related as $E_r \approx \kappa^2 E_v \approx \kappa^4 E_e$. Take the example of an N_2 molecule and estimate these energies, that is, the binding energy of electron in the atom, the vibrational excitation energy and the rotational energy separation. **[S]**

11. This problem is an exploration of the Slater determinant in search of its hidden secrets. We have seen that an n-electron Schrödinger equation became separable in position $\mathbf{r}_i$ if the potential energies could also be separated in the spatial coordinates, so, with

$$V(\mathbf{r}_1, \ldots, \mathbf{r}_n) = \sum_{i=1}^{n} V(\mathbf{r}_i),$$

the eigenfunction of Equation 1.8 becomes

$$|\psi(\mathbf{r}_1, \ldots, \mathbf{r}_n)\rangle = \prod_{i=1}^{n} |u_i(\mathbf{r}_i)\rangle, \quad \text{and} \quad E = \sum_{i=1}^{n} E_i,$$

where $|u_i\rangle$ satisfies the one-independent-electron Schrödinger equation

$$\left[-\frac{\hbar^2}{2m_0}\nabla_{\mathbf{r}_i} + V(\mathbf{r}_i)\right]|u_i\rangle = E_i|u_i\rangle.$$

The Slater determinant incorporates the antisymmetry dictated by Pauli exclusion. This Hamiltonian does not operate on the spin coordinate σ of the electron. So, when we write

$$|u_i(\mathbf{r}_i, \sigma_i)\rangle = |u_{i\sigma}(\boldsymbol{\tau}_i)\rangle = u_i(\mathbf{r}_i)\xi_i(\boldsymbol{\sigma}_i),$$

this Pauli-conditioned form also satisfies the Hamiltonian. This form, where we now have

$$|\psi_\sigma(\boldsymbol{\tau})\rangle = \prod_{i=1}^{N} |u_{i\sigma}(\boldsymbol{\tau}_i)\rangle,$$

is this solution, using the same argument of product in the independent electron approximation. These can be written as determinants of all possible configurations

$$|\Psi\rangle = \mathscr{A}\begin{vmatrix} |u_{1\sigma}(\boldsymbol{\tau}_1)\rangle & \cdots & |u_{1\sigma}(\boldsymbol{\tau}_N)\rangle \\ |u_{2\sigma}(\boldsymbol{\tau}_1)\rangle & \cdots & |u_{2\sigma}(\boldsymbol{\tau}_N)\rangle \\ \vdots & \vdots & \vdots \\ |u_{N\sigma}(\boldsymbol{\tau}_1)\rangle & \cdots & |u_{N\sigma}(\boldsymbol{\tau}_N)\rangle \end{vmatrix}.$$

Show that, with $|u_i\rangle$ orthogonal, a normalized $|\psi\rangle$ requires that $\mathcal{A} = 1/\sqrt{N!}$. **[M]**

12. To understand the implication of spin and Pauli exclusion, another exercise of interest is reworking our interpretation of the singlet and triplet construction for the hydrogen molecule. Let A and B be the atoms, and 1 and 2 the coordinates of the electrons. Using $|u_A\rangle$ and $|u_B\rangle$ as the basis orbital functions for trial functions,

$$|\psi_s\rangle = |u_A(1)\rangle|u_B(2)\rangle + |u_B(1)\rangle|u_A(2)\rangle, \quad \text{and}$$
$$|\psi_a\rangle = |u_A(1)\rangle|u_B(2)\rangle - |u_B(1)\rangle|u_A(2)\rangle,$$

where the former is symmetric, and the latter, asymmetric. Now consider new wavefunctions incorporating spin, $|u_\alpha\rangle$ and $|u_\beta\rangle$, and show that these may be written in the form

$$|\psi\rangle = [u_A(1)u_B(2) \pm u_B(1)u_A(2)]\,\xi(1,2),$$

with $\xi(1,2)$ as the spin functions of the two electrons. If $\hat{s}_z$ is an operator for the z component of the spin of the electron, then

$$\hat{s}_z|u_\alpha\rangle = \frac{1}{2}|u_\alpha\rangle, \quad \text{and} \quad \hat{s}_z|u_\beta\rangle = -\frac{1}{2}|u_\beta\rangle;$$

lets us then construct the two-electron spin function $|\xi(1,2)\rangle$. Show that this $|\xi\rangle$ is the eigenfunction of the square of the total spin ($|S|^2$) with the eigenvalues of 0 or 2. These lead to the Heitler-London functions $|\psi_s\rangle$ and $|\psi_a\rangle$ corresponding to the singlet and the triplet states. **[M]**

13. Consider an assembly of atoms, subscripted as A and B, so of only two types. The internuclear repulsion is $\sum_{A<B} Z_A Z_B e^2/|\mathbf{R}_B - \mathbf{R}_A|$. When solving for the eigenenergy E of the electron, we include this repulsion energy even if it has internuclear spatial parameters. It sets the potential energy in the Schrödinger equation for nuclear motion. Take a center-of-mass form Schrödinger formulation for a diatomic molecule at rest. The reduced mass is $\mu = M_A M_B/(M_A + M_B)$, and the equation is

$$\left[-\frac{\hbar^2}{2\mu}\nabla_{\mathbf{r}}^2 + V(R)\right]|\psi\rangle = E|\psi\rangle,$$

where $R = |\mathbf{R}_B - \mathbf{R}_A|$, and V is the potential ("mechanical") from the interatomic forces. Let $V = (1/2)k_s(R - R_0)^2$, where R_0 is the equilibrium value of internuclear distance, and k_s is a force constant. Find the eigenfunctions $|\psi^{\underline{n}}\rangle$ and the eigenenergies E. **[M]**