Advanced Theory of Semiconductor Devices
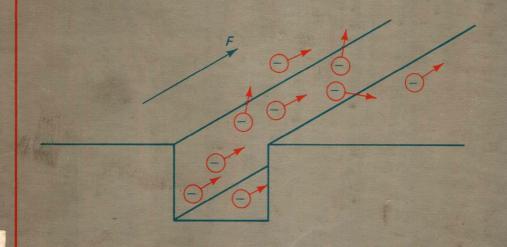

Karl Hess

ADVANCED THEORY OF SEMICONDUCTOR DEVICES

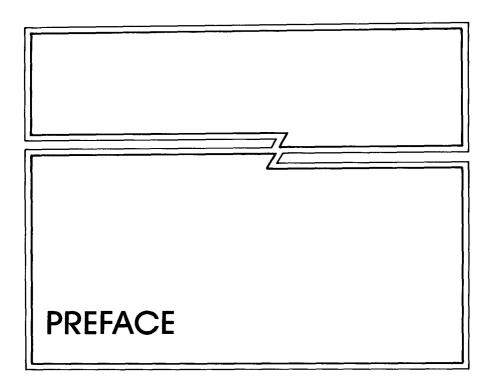**KARL HESS**

University of Illinois

To the memory of my father

PREFACE

Since the invention of the bipolar transistor by Bardeen and Brattain in late 1947, semiconductor devices have developed at an astonishing pace. A large variety of single device components for numerous uses evolved over the first two decades of the "golden semiconductor age," while in the third decade (starting about 1970) integrated circuits revolutionized semiconductor electronics. Semiconductor memories have replaced other components and have brought us not only the video game but also the supercomputer. Current devices approach submicrometer dimensions corresponding to 10^6 elements on a chip of centimeter size. At the same time, two newer technologies of crystal growth have evolved from the vapor phase epitaxy introduced in 1960, molecular beam epitaxy (MBE) and metal organic chemical vapor deposition (MOCVD), which make it possible to grow lattice-matched semiconductor structures having a characteristic dimension of 10^{-6} cm (and less). These heterostructures (superlattices) have already shown interesting effects with high device potential caused by quantum effects.

Astonishingly enough, the theory of semiconductor devices as a branch of solid state theory has received relatively little attention. Any theory of devices must contain a careful account of (1) electron (hole) drift and diffusion, including the effect of high electric fields and high energies; (2) generation recombination (which is usually treated very cursorily in solid state texts); and (3) the self-consistent fields of variable densities of electrons, holes

xiii

as well as donors and acceptors. For a proper treatment of (1)–(3) some basic knowledge of energy bands, electron impurity (lattice vibration) interaction, and basic quantum theory (tunneling, size quantization) is necessary. This background per se can be found in many solid state texts; unfortunately, however, little effort has been made to link it to electronic devices. Although much of the material and especially the depth of treatment of these texts are directly influenced by device applications, this fact is hardly ever mentioned. Often the device application is considered as something "dirty," and any remarks about it are more or less shamefully avoided. The physics student therefore usually does not understand why Gunn devices and field effect transistors are not made out of InSb or PbTe (which have higher minima and high electronic mobilities) but out of GaAs instead. I have also not found a single text on solid state that explains well why silicon is so special. On the other hand, there exist enormous amounts of information on devices with little link to the basics. In these texts, equations that describe the device operation are typically "introduced without much justification" and then are integrated to arrive at the final result. Nobody really knows just why this or that equation has been used and what has to be modified when the device operation is slightly changed or when the dimensions are shrunk, for example. Even in excellent texts on devices, principles are consistently used that are obsolete and have nothing to do with reality. For example, the electrical conductivity is always assumed to be proportional to $T^{3/2}$ (T is the absolute temperature) when ionized impurity scattering is dominant. In fact, in metal-oxide-silicon transistors this is never true and T^0 law is much better over most of the temperature range. This is only one of many examples.

To understand devices, and especially very small (submicrometer) devices, we need to establish our approach on more basic principles. We need to know about the effective mass theorem, Fermi's Golden Rule, the scattering probabilities, velocity overshoot, and the generation-recombination mechanism, to give a few examples. A book that treats all the necessary principles, including new developments and the resulting devices, would be much too long and probably unnecessary. Streetman's *Solid State Electronic Devices* provides a working knowledge of most of the important devices. It is the purpose of the present book to outline the basic unifying principles that are necessary to understand these devices in greater detail and to enable generalizations for future development. Consequently, I will discuss only certain ideal models of devices and refer to sources such as Streetman for more details on specific devices. However, much detail will be given with respect to the basic principles. Therefore, this book addresses graduate students and researchers in fields connected with semiconductor devices who wish to derive and understand the basic equations for a novel or current device, equations that are general enough to contain all the necessary basic physical effects and specific enough to make an effective solution possible. It is this quality that distinguishes a device theory from mere device modeling or from a general theory of the solid state.
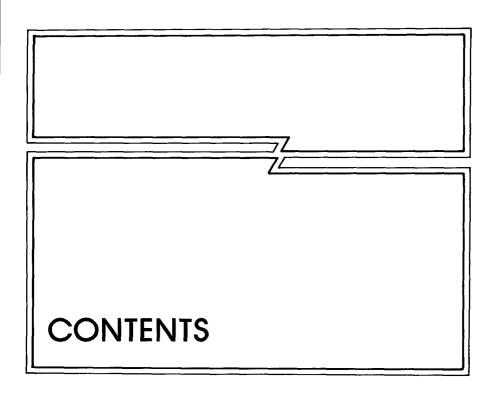
In addition, I desire to show that semiconductor device theory can be a discipline by itself that contains the necessary richness and complications to attract scientific interest and provides the possibility of new developments in future decades. This richness is caused, in my opinion, by the existence and importance of generation-recombination, statistical effects (Boltzmann equation), and the possibility to control the boundary conditions on a quantum level (submicron devices, quantum well superlattice structures). These complicated effects have not only become important for small devices but can in turn be included in modern device theory and simulation, because of the enormous progress of computers. Large-scale computation will be a major tool to understand small devices, and this book is also intended to give the physical basis for large-scale computational models.

As a text for graduate courses, this book can be used in various ways. I have used it for a one-semester course for first- and second-year graduate students of physics and electrical engineering. Not all of the material can be taught in one semester. Depending on student interest, one can emphasize the solid state aspects and leave out much of the last chapters on devices, or one can emphasize the device aspects and skip much of Chaps. 3 and 9 and all of Chaps. 2 and 7. I have also used the contents of this book for a two-semester course and have supplemented the text in the second semester by a description of numerical methods in device simulations (MINIMOS, PISCES) and on another occasion by developing in class a band structure and Monte Carlo simulation. The required background for the book is an introduction to the principles of quantum mechanics and, even more so, a working knowledge of advanced calculus. The material is presented in a dense form, and, in order to really understand it, one needs to "go over it with a pencil" and in some instances consult the referenced literature.

My thinking and approach in writing this book have been much influenced by my colleagues B. G. Streetman and N. Holonyak, Jr., and by the inspiring environment at the University of Illinois.

Thanks go to E. Kesler, C. Willms, and to my wife Sylvia for their help in preparing the manuscript.

Karl Hess

CONTENTS

BRIEF REVIEW
OF THE RELEVANT
BASIC EQUATIONS
OF PHYSICS

It is clear that from a mathematical viewpoint all equations of physics (microscopic and macroscopic) are relevant for semiconductor devices. In an absolutely strict theoretical way, we therefore would have to proceed from the fundamentals of quantum field theory and write down the $\sim 10^{23}$ coupled equations for all the atoms in the semiconductor device. Then we would have to solve these equations, including the complicated geometrical boundary conditions. However, the outcome of such an attempt is clear to everyone who has tried to solve only one of the 10^{23} equations.

Any realistic approach has to proceed differently. Based on the experience and investigations of many excellent scientists in this field, we neglect effects that would influence the results only slightly. In this way many relativistic effects become irrelevant. In my experience the spin of electrons plays a minor role in the theory of semiconductor devices and can be accounted for in a simple way (the correct inclusion of a factor of 2 in some equations).

Most effects of statistics can be understood classically, and we will need only a very limited amount of quantum statistical mechanics. This leaves us essentially with the Hamiltonian equations (classical mechanics), the Schrödinger equation (quantum effects), the Boltzmann equation (statistics), and the Maxwell equations (electromagnetics).

It is clear that the atoms that constitute a solid are coupled, and therefore the equations for the movement of atoms and electrons in a solid are

coupled. This still presents a major problem. We will see, however, that there are powerful methods to decouple the equations and therefore make explicit solutions possible. My presentation attempts a delicate balance between rigor and intuitive concepts. In this way, the fundamental laws of physics are finally reduced to laws of semiconductor devices that are tractable and whose limitations are clearly stated.

Many body effects such as superconductivity are excluded from our treatment. Because of the low density of mobile particles in semiconductors, many body effects are rare, except for effects connected to screening, which will be treated in detail in Chap. 7. Effects of high magnetic fields are also excluded since they are unimportant for most device applications.

1.1 THE EQUATIONS OF CLASSICAL MECHANICS

Some time after the work of Galileo, Kepler, Copernicus, and Newton, Hamilton was able to give the laws of mechanics a very elegant and powerful form. He found that the laws of mechanics can be closely linked to the sum of kinetic and potential energy written as a function of momentumlike (p_i) and spacelike (x_i) coordinates. This function is now called the Hamiltonian function $H(p_i, x_i)$. The laws of mechanics are

$$\frac{dp_i}{dt} = -\frac{\partial H(p_i, x_i)}{\partial x_i} \tag{1.1}$$

and

$$\frac{dx_i}{dt} = \frac{\partial H(p_i, x_i)}{\partial p_i} \tag{1.2}$$

where t is time and $i = 1,2,3$. Instead of x_i, we sometimes denote the space coordinates by x, y, z.

Some simple special cases can be solved immediately. The free particle (potential energy = constant) moves according to

$$H = \sum_i p_i^2 / 2m$$

and we have from Eq. (1.1)

$$\frac{dp_i}{dt} = 0; \qquad p_i = \text{constant}$$

which is Newton's first law of steady motion without forces.

If we have a potential energy $V(x_1)$ that varies in the x_1 direction, we obtain from Eq. (1.1)

$$\frac{dp_1}{dt} = -\frac{\partial V(x_1)}{\partial x_1} \equiv F \tag{1.3}$$

The quantity defined as F is the force, and Eq. (1.3) is Newton's second law of mechanics.

These examples are enough for our purpose, since we will make less use of classical mechanics than we do of quantum mechanics, which is discussed in the next section.

1.2 THE EQUATIONS OF QUANTUM MECHANICS

At the beginning of the twentieth century, A. Einstein, M. Planck, L. de Broglie, E. Schrödinger, W. Heisenberg, and M. Born (to name a few) realized that nature cannot be strictly divided into waves and particles. They found that light has definite particlelike properties and cannot always be viewed as a wave, and particles such as electrons revealed definite wavelike behavior under certain circumstances. They are, for example, diffracted by gratings as if they had a wavelength

$$\lambda = h/|\mathbf{p}| \tag{1.4}$$

where $\hbar \equiv h/2\pi \simeq 6.58 \cdot 10^{-16}$ eVs and $\mathbf{p}$ is the electron momentum.

Schrödinger showed that the mechanics of atoms can be understood as boundary value problems. In his theory, electrons are represented by a wave function $\psi(\mathbf{r})$, which can have real and imaginary parts, and follows an eigenvalue differential equation:

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r})\right)\psi(\mathbf{r}) = E\psi(\mathbf{r}) \tag{1.5}$$

The part of the left-side of Eq. (1.5) that operates on ψ is now called the Hamiltonian operator H.

Formally this operator is obtained from the classical Hamiltonian by replacing momentum with the operator $\hbar/i \nabla (i = $ imaginary unit), where

$$\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right).$$

The meaning of the wave function $\psi(\mathbf{r})$ was not clearly understood at the time of Schrödinger; in fact, $\psi(\mathbf{r})$ was misinterpreted by Schrödinger himself. It is now agreed that $|\psi(\mathbf{r})|^2$ is the probability of finding an electron in a volume element $d\mathbf{r}$ at $\mathbf{r}$. In other words, we have to think of the electron as a point charge with a statistical interpretation of its whereabouts (the wavelike nature!) It is usually difficult to get a deeper understanding of this viewpoint of nature; even Einstein had trouble with it. It is, however, a very successful viewpoint that describes exactly all phenomena we are interested in. To obtain a better feeling for the significance of $\psi(\mathbf{r})$, we will solve Eq. (1.5) for several special cases. As in the classical case, the simplest solution is obtained for constant potential. Choosing an appropriate energy scale, we put $V(\mathbf{r}) = 0$ everywhere.

By inspection we can see that the function

$$C \exp (i\, \mathbf{k} \cdot \mathbf{r}) = C \,(\cos \mathbf{k} \cdot \mathbf{r} \, + \, i \sin \mathbf{k} \cdot \mathbf{r}) \qquad (1.6)$$

is a solution of Eq. (1.5) with

$$\hbar^2 k^2 / 2m = E \qquad (1.7)$$

and C a constant.

The significance of the vector $\mathbf{k}$ can be understood from analogies to well-known wave phenomena in optics and from the classical equations. Since E is the kinetic energy, $\hbar \mathbf{k}$ has to be equal to the classical momentum $\mathbf{p}$ to satisfy $E = p^2 / 2m$. On the other hand, in optics

$$|\mathbf{k}| = 2\pi / \lambda \qquad (1.8)$$

which gives, together with Eq. (1.4),

$$\hbar \mathbf{k} = \mathbf{p}$$

which is consistent with the mechanical result.

How can the result of Eq. (1.6) be understood in terms of the statistical interpretation of $\psi(\mathbf{r})$? Apparently

$$|\psi(\mathbf{r})|^2 = |C|^2 (\cos^2 \mathbf{k} \cdot \mathbf{r} + \sin^2 \mathbf{k} \cdot \mathbf{r}) = |C|^2$$

This means that the probability of finding the electron at any place is equal to C^2. If we know that the electron has to be in a certain volume V_{ol} (e.g., of a crystal), then the probability of finding the electron in the crystal must be one. Therefore

$$\int_{V_{ol}} |C|^2 \, d\mathbf{r} = V_{ol} |C|^2 = 1$$

and

$$|C| = 1 / \sqrt{V_{ol}} \qquad (1.9)$$

In other words, the probability of finding an electron with momentum $\hbar \mathbf{k}$ at a certain point $\mathbf{r}$ is the same in the whole volume and equals $1/V_{ol}$.

This is a peculiar result that can only be understood if one is either very familiar with optics (coherence conditions in interference problems) or if one understands in detail the uncertainty principle. The unfamiliar reader is referred to the literature given at the end of this section (Feynman).

By confining the electron to a volume, we have already contradicted our assumption of constant potential $V(\mathbf{r}) = 0$. (Electrons can only be confined in potential wells.) If, however, the volume is large, our mistake is insignificant for many purposes.

Let us now consider the confinement of an electron in a one-dimensional potential well (although such a thing does not exist in nature). We assume that

the potential energy $V(\mathbf{r})$ is zero over the distance $(0, L)$ on the x axis and infinite at the boundaries 0 and L.

The Schrödinger equation, Eq. (1.5), reads in one dimension (x direction, $V(x) = 0$)

$$-\frac{\hbar^2}{2m}\frac{\partial^2 \psi(x)}{\partial x^2} = E\psi(x) \tag{1.10}$$

Inspection shows that the function

$$\psi(x) = \sqrt{\frac{2}{L}} \sin\frac{n\pi}{L}x \qquad \text{with } n = 1,2,3,\ldots \tag{1.11}$$

satisfies Eq. (1.10) as well as the boundary conditions. The boundary conditions are, of course, that ψ vanishes outside the walls, since we assumed an infinite impenetrable potential barrier. In the case of a finite potential well, the wave function penetrates into the boundary and the solution is more complicated. If the barrier has a finite width, the electron can even leak out of the well (tunnel). This is a very important quantum phenomenon the reader should be familiar with. We will return to the tunneling effect below.

The wave function, Eq. (1.11), corresponds to energies E

$$E = \frac{n^2\pi^2\hbar^2}{2mL^2} \tag{1.12}$$

Since n is an integer, the electron can assume only certain discrete energies while other energies are not allowed. These discrete energies that can be assumed are called *quantum states* and are characterized by the *quantum number n*. The wave function and corresponding energy are therefore also denoted by ψ_n, E_n.

Think of a violin string vibrating in various modes at higher and lower tones (frequency ν), depending on the length L, and consider Einstein's law:

$$E = h\nu \tag{1.13}$$

If we compare the modes of vibration of the string with the form of the wave function for various n, then we can appreciate the title of Schrödinger's paper, "Quantization as a Boundary Value Problem."

From the above examples it is clear that the solution of the Schrödinger equation requires considerable effort and we would be in trouble if we had to find an exact solution for all kinds of potentials $V(\mathbf{r})$ that are of interest in the theory of semiconductor devices. Fortunately there is a powerful method of approximation, perturbation theory, that gives us the solution for arbitrary weak potentials. The method is very general and applies to any kind of equation.

Consider an equation of the form

$$(H_0 + \epsilon H_1)\psi = 0 \tag{1.14}$$

where H_0 and H_1 are differential operators of arbitrary complication and ϵ is a small positive number.

If we know the solution ψ_0 of the equation

$$H_0\psi_0 = 0$$

then we can assume that the solution of Eq. (1.14) has the form $\psi_0 + \epsilon\psi_1$. Inserting this form into Eq. (1.14), we obtain

$$(H_0 + \epsilon H_1)(\psi_0 + \epsilon\psi_1) = H_0\psi_0 + \epsilon H_1\psi_0 + H_0\epsilon\psi_1 + \epsilon^2 H_1\psi_1$$

We now can neglect the term proportional to ϵ^2 (because ϵ is small), and since $H_0\psi_0 = 0$, we have

$$H_1\psi_0 + H_0\psi_1 = 0 \qquad (1.15)$$

This equation is now considerably simpler than Eq. (1.14) since ψ_0 is known. Therefore, ψ_1 can be determined easily if H_0 has a simple form no matter how complicated H_1 is. The form of the Schrödinger equation allows further simplification, and it is easy to show that any problem involving a small given perturbation H_1 reduces to solving a three-dimensional integral.

Assume that we know the solutions of a Schrödinger equation:

$$H_0\psi_n = E_n\psi_n \qquad n = 1,2,3,\ldots \qquad (1.16)$$

and we would like to know the solutions of

$$(H_0 + H_1)\phi_m = W_m\phi_m \qquad \text{with } H_1 \ll H_0 \qquad (1.17)$$

Then it is shown in elementary texts on quantum mechanics (Baym) that

$$W_m = E_m + M_{mm} + \sum_{n \neq m} \frac{|M_{nm}|^2}{E_m - E_n} \qquad (1.18)$$

with

$$\phi_m = \psi_m + \sum_{n \neq m} \frac{M_{nm}}{E_m - E_n} \psi_n \qquad (1.19)$$

and

$$M_{mn} = \int_{V_{ol}} \psi_n^* H_1\psi_m \, d\mathbf{r} \qquad (1.20)$$

where $d\mathbf{r}$ stands for $dxdydz$ (integration over the volume V_{ol}) and ψ_n^* is the complex conjugate of ψ_n.

We have, therefore, reduced the solution of the Schrödinger equation of the new problem to the integration of Eq. (1.20). In addition, one needs to know the solution of Eq. (1.16). Depending on the nature of the problem, this could be the solution for the free electron, the one-dimensional well, the hydrogen atom, the harmonic oscillator, or other well-known solutions.

The formalism outlined above and the examples given are independent