1 Introduction

1.1 The early days of MOS technology

In the past couple of decades, the increasing influence of electronics on human life has promoted MOS technology to a role of similar significance for cultural change as, for example, electric power transmission and combustion engine transport. The basic device for this development, the metal–oxide–semiconductor field-effect transistor (MOSFET), was patented in 1928 by Lilienfeld. The invention had to wait for realization until 1961 when Khang at Bell Telephone Labs first demonstrated a working device. Until then, one of the main hurdles for implementing Lilienfeld's idea was finding a material combination such that a surface channel for charge carriers could be brought about by an external electric field. A charge-free surface or interface was needed, which required a structure free of charge carrier traps. Here, silicon technology opened new possibilities. By thermally oxidizing the surface of silicon crystals into SiO_2, an insulator was obtained with eminent properties and with a low concentration of traps at the SiO_2/Si interface and in its volume. At the beginning of the 1960s, a considerable amount of work was performed to optimize the properties of SiO_2 prepared this way and to understand the metal–oxide–semiconductor system. Important contributions to the understanding of the MOS system came from a group of William Shockley's former disciples at Fairchild Semiconductor in Palo Alto. In the same period, activities were also initiated at the IBM Thomas Watson Research Center, at the Bell labs and at some universities in the USA.

In parallel with this research on silicon dioxide, the MOSFET was developed. When a voltage is applied to the gate electrode of a MOS structure, a bending of the semiconductor energy bands occurs such that charge carriers can be injected from a source contact and collected by a drain contact on the opposite side of the gate. A channel is then opened at the Si/SiO_2 interface, and the current between source and drain can be regulated by the voltage on the gate, producing transistor action. After mastering the materials problems, this ingenious and simple geometrical design conquered other technologies for logic circuitry. The tremendous development potential of the device can be realized by considering that the channel lengths of the original MOSFETs were tens of microns while in today's transistor development, the corresponding distance is a couple of tens of nanometers. This was the ramp that launched the expansion of communications and computation present everywhere in global society today. At the time of writing this book, no serious competing device structure has been

proposed to replace the MOSFET for future applications. Observers of the field forecast that MOS technology will continue to dominate electronics applications for the foreseeable future.

1.2 Nature's freak of fate

As the MOS system operates by creating a channel between the insulator and the semiconductor for charge carrier transport, this sets high demands on the degree of perfection at this interface. Oxides used as the insulator material are generally amorphous, while the semiconductor material is crystalline. Therefore, the atomic structure at the interface between these two material types contains geometrical misfits, and thus imperfections due to open electron orbitals, which give rise to electron traps. Such "dangling bonds" are important sources of charge at insulator/semiconductor interfaces and influence the transport and stability properties of MOSFETs. At its beginning and for a long period of time, the whole MOS concept was leaning on the possibility to prepare the gate oxide by thermal oxidation of the silicon crystal surface. By a freak of nature, the SiO_2/Si interface built up that way. Not only existed a semiconductor material based on an element with a high abundance on earth, with a crystal structure of high mechanical strength and an energy bandgap well suited for electronic applications, but its natural oxide was almost perfect for use as an insulator! These materials properties together with the simple and flexible design of the MOSFET are the linchpins that have carried MOS technology to its present position.

1.3 Silicon dioxide becomes inadequate

In the downscaling of device dimensions, which in the past four or five decades has roughly followed Moore's law and resulted in a rapidly increasing number of transistors per chip, one of the main issues has been to design the properties of the transistor channel. Decreasing the distance between source and drain, thus making the depletion regions of these two p-n junctions approach each other, gives rise to increased current leakage and decreased threshold voltage. This problem, known as the "short channel effect" and occurring for each new technology generation, was solved by introducing sophisticated doping geometries for the source and drain contacts and by increasing the channel doping. However, the latter measure, which serves to decrease transistor leakage, has a detrimental influence on the capacitive coupling between the gate and the channel. This in turn influences the ability of the transistor to switch between its on- and off-states.

The switching capacity is one of the most important properties of the MOSFET. It depends on the characteristics of the MOS system constituting the gate/channel combination. For a given change in gate voltage, a high share needs to be supplied to the semiconductor in order to flip its energy bands efficiently and create or eliminate the channel. This voltage partition depends on the relation between the capacitances

of the gate oxide and of the channel region, respectively. A high ratio between these two quantities is desirable. However, increasing the channel doping in order to solve the problem of the short channel effect will increase the channel capacitance for a given voltage drop and thus impair the semiconductor's switching capability. The remedy is to increase the oxide capacitance. If SiO_2 is kept as the insulator material, this means that its thickness needs to be decreased. Such steps were part of transistor development until the SiO_2 layer thickness approached the lower limit of about 1.5 nm, where the leakage current between gate and channel could no longer be accepted. At this point, the way to proceed included the use of oxynitrides, thus increasing the dielectric constants a couple of units from the value of 3.9 for SiO_2. For a couple of technology generations, this allowed the use of a thicker insulator with high enough capacitance and low enough leakage. Finally, in order to be able to continue into channel lengths shorter than about 40 nm, SiO_2-based gate insulators could no longer be used. A drastic step had to be taken by changing to materials based on elements from other parts of the periodic table, the so-called "high-*k*" dielectrics, where *k* stands for the dielectric constant. The first development in this area was made for the 45 nm CMOS node, where a "hafnium-based" oxide was used. For the continued race of downscaling, therefore, the problem of finding new gate insulator materials will be one of the most important issues.

This necessary turn into novel materials for MOS development has been reflected in other parts of the transistor architecture. In the early period of this technology escalation, the atomic species used were mainly silicon, oxygen, aluminum, boron, arsenic and phosphorus. Prerequisites of more recent progress have been the use of, for example, copper and low-*k* materials for interconnects, and silicides based on various metals and different metal combinations for gate contacts; the search for novel metal oxides has included a large part of the transition and rare-earth metal range of the periodic system.

1.4 High-*k* dielectrics

The search for suitable materials to satisfy the gate functions of future MOS transistors has led the scientific semiconductor community on a quest through the periodic system to find the "Dielectric Grail." The goal is a material with acceptable energy offset values between the energy bands of the dielectric and the silicon crystal while, at the same time, having a high enough dielectric constant. The former quantity influences leakage for a given thickness, the latter secures channel coupling for a given leakage. So far for CMOS applications, most of the efforts have been limited to metal oxides. The change from the extremely well-mastered thermal SiO_2 material to an oxide based on metals among the transition or rare-earth series has revealed obstacles that were unnoticeable using traditional technology. As well as problems of chemical stability between these new "high-*k*" oxides and the silicon substrate, issues of crystallization, sensitivity to humid environment, higher concentrations of oxide traps and interface states have been encountered. Driven by technology, this has made it important to understand their microscopic properties from the chemical, physical and electrical points of view.

A common attribute of high-k oxide films deposited on silicon is the occurrence of an SiO_x interlayer between the high-k material and the silicon crystal. This evokes interface electron state properties similar to those at thermal SiO_2/Si interfaces. Even if the interlayer lowers the effective k-value of the film, it often gives better conditions for a transistor channel than those offered by a direct interface due to lower charge carrier scattering by the former. However, it must be paid for by an extra interface between the SiO_x and the high-k material. As the total physical thickness of the film is in the range of a couple of nanometers, on this length scale the transition from SiO_x to the high-k material can hardly be considered abrupt. It is found to include a transition region with undefined stoichiometry and thus with possible structural instabilities.

In order to use a convenient measure when comparing the thicknesses of insulators with different dielectric constants, the concept of "equivalent oxide thickness" (EOT) has been introduced. This is the thickness that a layer of SiO_2 would have for the same capacitance as a certain high-k layer. In order to fulfill the demands on transistors for the 22 nm node and beyond, the EOT values must be below 1 nm. The SiO_x interlayer has a k-value of roughly that same magnitude as SiO_2 and decreases the effective dielectric constant for the entire high-k stack. As it often appears with a thickness of about 1 nm, it must be eliminated in future applications. One possible solution to this problem has been the replacement of SiO_x by silicate.

1.5 Characterizing the MOS system

As the MOS system constitutes a capacitive device, a natural way to investigate its properties from an electrical point of view is by studying its admittance. From such measurements, capacitance and conductance as functions of voltage and frequency can be extracted and used as diagnostic data. Information about the properties of interface states and oxide charge can also be deduced from such results. Consequently, two traditional methods in wide use for MOS characterization are the capacitance versus voltage (C–V) and the "conductance method." In the pioneering work of the 1960s on understanding the physical properties of MOS structures, the C–V method was used to establish the main qualities of oxide/silicon interfaces. Based on this tool, the importance of interface state densities and oxide charge were revealed. Later, this technique was joined by the conductance method, where the real part of the MOS admittance was shown to complete data from the C–V method. From these initiatives, a number of variations and a huge amount of applications have blossomed for a continued detailed understanding of the technologically useful MOS structure.

The conductance method has lived through a discursive development after the first preliminary study, published in 1965 by Nicollian and Goetzberger. In that original work, the interpretation of measured data was oversimplified and a more realistic treatment followed by the same authors in a later published paper. The conductance method is based on resonance phenomena between the frequency of a probing voltage from the measurement set-up and the rate by which charge carriers are emitted from and captured to interface states. From such data, interface state densities and capture cross

sections of the charge carrier states are obtained. For more detailed diagnosis, the technique recently was developed into "multi-parameter admittance spectroscopy" (MPAS) for collection of capacitance and conductance data as functions of voltage, frequency and temperature, as discussed in detail in Chapter 6.

Properties of interface states at most oxide/silicon interfaces investigated so far, whether for SiO_2 or for high-k oxides with or without the SiO_x interlayer mentioned above, are surprisingly similar and seem to be dominated by dangling bonds. Also defects in the bulk of SiO_2 show a kinship with defects in high-k oxides. An important and widely investigated defect in SiO_2 is the E' center, which originates from an oxygen vacancy. For most high-k oxides it has been argued that the dominating defect has a similar origin. Here electron spin resonance has had a central role for elucidating the detailed physical properties of interface and bulk oxide states.

In the SiO_2 era, physical methods like electron spin resonance (ESR) and X-ray photon spectroscopy (XPS) were successfully used to investigate the oxide/silicon interface for electron state and bonding phenomena. As the novel high-k materials have brought more complicated physiochemical problems, additional methods have been introduced, such as X-ray diffraction (XRD) for investigating amorphicity and medium energy ion spectroscopy (MEIS) to map the distribution of elements.

Advances within two specific areas have had a most vital importance for the development of advanced device architectures: microscopy and computer development. The development of electron microscopy, by both scanned (SEM) and transmitted (TEM) electron beams, has reached a level where single atoms in a solid matrix can be detected. Together with additional characterization tools included in these instruments for identification of atomic species, and the family of atomic probe microscopy (ASM, AFM etc.), completely new insight has been achieved for nanoscale exploration over the past couple of decades. The feedback system for development of semiconductor devices may be considered starting with the theory for refinement and characterization of materials. Here, the incredible development of computers has had a decisive importance, presently making first-principle calculations of physical quantities possible with very high precision. Such input into the preparation of novel materials combinations has led to improved device performance to be fed back into ever more powerful computers.

1.6 Next episode

The outstanding rate of technical development in MOS technology and its influence on society has no comparison with any other technical evolution in human history. Attempts to forecast the consequences of its continued influence are fraught with difficulty. A look in the rear-view mirror reminds us of the problem: in 1970, mobile telephones of pocket size were science fiction; in 1980, it was believed that a couple of computers would serve the needs of any big nation; in 1990, the internet was an almost unknown concept.

For the technology itself somewhat more stringent predictions may be expressed. How far is it possible to pursue MOS technology? As believed by most observers at the time of printing of this book, future MOSFET fabrication will be dominated by silicon-on-insulator (SOI) technology. This means that an ultrathin intrinsic semiconductor material will be used for the transistor channel which will help in mastering the short channel effect, mentioned above. In a natural continued development, the channel will probably be shaped as a wire until its decreasing length will result in a quantum dot with 5 nm side lengths and height. A possible arrival at this destination will be preceded by an exciting voyage, demanding refined materials and ingenious design solutions by researchers and engineers. The following chapters are aimed as a contribution to the vehicle needed to carry on along that road.

1.7 Overview of the subject

The book is divided into three parts: Part I describes a basic physical background of MOS systems; Part II presents the most common characterization methods; and Part III treats the novel MOS systems from their actual perspective.

The ambition has been to create a text, not with the purpose to function as a dictionary for formulas to be used in technical development reports or in scientific paper production, but to provide a feeling of understanding to the reader. From the author's experience as an engineer in industry and as a researcher in the academic environment, metaphoric representations of physical conditions give a base for selecting the consecutive steps when carrying out an experimental sequence in the laboratory. For the subject of this book, this means to achieve a qualitative inside image of how a MOS system reacts when changing the conditions for perturbing it. Such basic understanding is served, for example, in Chapter 2, and signifies the methodology for the remainder of this survey: to accompany the derivation of physical relations with a discussion of calculated examples.

Chapter 3 similarly describes the conditions for the most important physical properties of MOS systems: the dielectric constant and the energy barriers separating the metal from the semiconductor. Chapters 4 and 5 start from statistical mechanics to pronounce the properties of electron states and the energy quantities involved in charge carrier traffic as a preparation for understanding the dynamic influence from interface states and bulk oxide traps. This is the base for understanding the most common characterization methods, which are covered in Chapters 6–9.

In Part III, Chapter 10 portrays the properties of silicon dioxide, mainly in its role connected with high-k oxides, followed in Chapter 11 by an overview of the status of these new dielectric gate materials and, in Chapter 12, a discussion of the delicate problems involved in selecting gate metals. Because of its sensitivity to sample specific quantities, the problem of gate current leakage is treated, in Chapter 13, merely from a simplified and basic point of view, to give the reader an understanding of its physical principles and consequences for device design. At the end, Chapter 14

describes the lively development activities for finding new transistor channel materials going on at the time of printing.

To a few specific parts the following text is critical to some earlier results published in literature. Such opposing views constitute the driving force of science. The reader should take these parts as an exercise given by the author: to find them and to critically create one's own view of where the truth lies.

Part I

Basic properties