1
Introduction to Silicon Photonics

1.1 Introduction

Silicon is the material par excellence. It is the most widely studied material in the history of civilization. In fact, the present-day information age has dawned with an electronics revolution brought about by the maturity of silicon-based microelectronics.

The growth of the silicon industry follows the now-famous Moore’s law, which states that the number of transistors in an integrated circuit chip doubles every 12 months (since revised to every 18 months). However, during the last few years there has been indication of the decline of Moore’s law. There are doubts whether in future silicon-based integrated circuits (ICs) will deliver the same advantages and functionalities as shown today.

The weakest point of silicon is that proper light emitters and modulators cannot be realized by using it due to the indirect nature of its band gap. On the other hand, there is a steady increase in the area of photonics, in the form of optical communication and networking, optical information processing, and consumer electronics based on light. Present-day photonics relies on compound semiconductors and their alloys. Although discrete devices using these materials show very good performance, when it comes to integration of these devices, preferably on the same substrate, the levels of integration and performance are far below what has already been achieved in electronic integration. It is natural to expect that monolithic optoelectronic integrated circuits (OEICs) will provide the same advantages, that is, low cost due to batch fabrication, high functionality, scaling for denser integration, and so on, as provided by silicon ICs.

If, however, it is possible to grow OEICs on silicon and integrate with electronic ICs by using the same production facilities, the benefits to be accrued need no further elaboration. Si-based systems will then be used in all fields of electronics, computers, and communication. This is the dream cherished by many workers over the last few decades, though that dream is yet to materialize. In spite of this, Si-based photonics remained an active area of research and over the last 10–15 years some significant milestones have been achieved.
Another important area of application of silicon photonics is in the very large-scale integrated (VLSI) circuit itself. The complexity of present-day ICs has reached such a high level that the interconnects within it are formed on a number of levels. At present the number is six, but within a few years it will be doubled. The metallic interconnects, mainly Cu, provide delay due to resistor–capacitor (RC) time constants, which far exceed the transit time delay associated with the individual transistors. If the increase in speed is to be maintained at the same rate for the next-generation ICs, the interconnect bottleneck must be properly addressed. Optics is believed to be the right solution to the problem.

In the present chapter, we shall give an overview of the developments in silicon electronics, the present status, and the problems faced to achieve the goals discussed in this section. We first give a very short history of the development of silicon-based microelectronics, the present status, and the international roadmap for future development. The two most important areas in which presence of silicon is needed, that is, communication and interconnect in chips, will then be described, giving the reader an idea of the present scenario and the problems faced. In this connection, the alternative to monolithic integration, that is, the hybrid integration technology, followed at present and the related problems will be pointed out. Finally, the scope of the present book will be outlined.

A number of text, research monographs, and reviews have already appeared dealing with silicon photonics. The reader is referred to such sources, all of which contain a large number of useful references [1–13].

1.2 VLSI: Past, Present, and Future Roadmap

The announcement of the first point contact transistor on December 23, 1947, marks the birth of the electronics era. Although the material used in this device and its improved version, the junction transistor, was Ge, it was soon felt that single-crystal silicon would be a better alternative. The first silicon bipolar transistor came in 1954. The concept of the IC was first explored in 1958, and its working was demonstrated by using discrete components. A few months later, an IC using planar technology was developed. The bipolar transistor technology was developed earlier and was applied to the first IC memory in the 1960s. Although bipolar transistors are the fastest at the individual circuit level, their large power dissipation and very low integration level (~10^4 circuits per chip), compared to today’s VLSI standard, do not promote their use.

The control of conductivity in the surface of a semiconductor by an external electric field was proposed in the early 1930s. Attempts for conductivity modulation during the early part of the 1950s were not very successful. The first metal oxide semiconductor field effect transistor (MOSFET) using SiO2 as the gate insulator on silicon substrate was fabricated in the 1960s. The complementary MOS (CMOS) transistor was fabricated in 1963, and its advantage of lower power consumption was firmly established. The one-transistor dynamic random-access memory (DRAM) cell was announced in 1968, and the first microprocessor was marketed in 1971 [14].

The advantages of silicon as an electronic material are too many to recount here. Silicon is available in nature abundantly. It can be purified to a very high level. Native oxide silica is a very good insulator, is stable, and can withstand a large field across it; and above all, the interface charge between silica and silicon can be reduced to a minimal level. SiO2 can be easily patterned by photolithography [15].
Although bipolar transistors are faster than MOSFETs at the individual level, the low power consumption in CMOS, adaptability to planar processing, reduced size of the transistor, and larger packing density and ease of fabrication with reduced number of masks have made the CMOS technology on silicon substrate the sole technology followed by industries [16]. Figure 1.1 illustrates the developments of VLSIs over the last three decades and the technology roadmap for the coming decade [17].

At the rate shown in Figure 1.1, there were 1 billion transistors on a single die before 2007. With increasing numbers of transistors per die, the minimum feature size, or roughly the channel length in a single transistor, was around 70 nm in 2008. For almost the past 30 years, the feature size in IC lithography has been reduced at a rate of $0.7^{1/3}$ every three years. It is predicted that the feature size will reach 35 nm in 2014.

1.3 The Interconnect Problem in VLSI

On-chip interconnect is nothing but electrical wiring. According to the International Technology Roadmap for Semiconductors (ITRS), an interconnect is electrical wiring that distributes clock and other signals, and provides power and ground to and among the various circuits or systems functions on a chip.

The process devoted to metallization and interconnect involves deposition of metals, interlevel dielectrics deposition, and etching steps. A typical interconnect structure is shown in Figure 1.2.

The earlier approach was aluminum deposition and dry-etch definition. Currently, copper wires are introduced. The global wires connect different functional units, distributing clock signals and power among them. The length of the global wires scales down with chip size. On the other hand, local wires connect the gates, sources, and drains of close MOSFETs of the same functional units and their length scales with gate size.

Although the downscaling of transistor increases the speed, the same is not true for the downsizing of interconnect. In earlier ICs in use around the 1980s, the delay in interconnects was propagation limited, that is, it was limited by the time of propagation of the electromagnetic waves associated to SiO$_2$, rather than the RC time constant, which was $\sim 1$ ps. It is predicted that for 35 nm technology generation, the interconnect response time of 0.1 mm copper line with a low-k dielectric ($k = 2$) will be about 250 ps, about two orders of magnitude higher, and will account for delays related to the RC time constant of the wire.
The second issue with interconnect is power consumption. With scaling of transistors, the power consumption by the interconnect exceeds that by a transistor. For example, for 1.0\(\mu m\) devices, the switching energies in the transistor and 1 mm long interconnect were, respectively, 300 and 400 fJ. The predicted values for 35 nm technology are 0.1 and 3 fJ, respectively, indicating that the ratio between dissipation in the interconnect and in the transistor is about 30.

Figure 1.3 shows a trend of interconnect propagation delay for 1 cm length with feature size and year for aluminum metal and silica insulators, copper and low-k dielectrics, and projected optical waveguide technology. It indicates that Al-based technology has reached the performance limit at 0.55\(\mu m\); the reduced resistance due to Cu and reduced capacitance due to low-k dielectric ensure a performance improvement limit up to 0.18\(\mu m\) technology.

Figure 1.3  Variation of interconnects propagation delay with year and feature size for Al–SiO\(_2\), Cu–low-k dielectric, and projected optical interconnects technology.
The optical interconnect using silicon microphotonics technology offers a potential solution to the RC time delay associated with traditional metal interconnects. Using photons as bits of information, instead of electrons, a speedier performance of the devices is expected. Use of photons also solves the power dissipation problem. Photons propagate in transparent media with less heat dissipation and almost no cross-talk. Unlike electrical current beams, light beams can cross one another without using any insulator. The multilevel interconnection scheme shown in Figure 1.2 is not needed when light beams are used for interconnects within the chip.

Further discussion of the use of light waves for chip-to-chip or board-to-board connections will follow in Section 1.6.

1.4 The Long-Haul Optical Communication Link

Fiber-optic communication links have spread today over the whole globe like a spider’s net. A still larger number of links is being added. Today’s fiber-optic links employ dense wavelength division multiplexing (DWDM), in which huge amounts of data carried by hundreds of carrier wavelengths, each modulated at a high bit rate (~10 Gb/s or more), are transmitted by a single strand of a fiber. In Section 1.4.1, we shall first discuss the basic link, and the components used [18]. In Section 1.4.2, the materials used to grow the devices and the methods of integration of the devices will be pointed out.

1.4.1 Basic Link and Components

Figure 1.4 shows a block schematic of the WDM communication link. Voice, picture, or computer data, in digital format, are impressed on each laser emitting at a particular wavelength (e.g., $\lambda_1$). Either the laser may be directly modulated, or an external modulator may be used to impress the signal on the laser beam. A multiplexer combines the modulated signals coming from the bank of lasers, and the combination is transmitted by an optical fiber. After traversing a distance of a few hundred kilometers, the signal becomes attenuated.

![Figure 1.4 Schematic diagram of a WDM point-to-point communication link.](image-url)

and the digital pulses considerably spread due to material dispersion of the fiber. The combined signal is then regenerated by a regenerator (not shown in Figure 1.4). A photodetector first converts the weak and distorted optical signal into a stream of electrical pulses. These pulses are then reshaped and retimed by a decision circuit. The cleaned electrical pulses are then converted to optical pulses by a laser, and the stream of pulses propagates through another long section of the fiber. At present, a number of optical amplifiers are inserted at regular intervals in the link to boost up the intensity of optical signals, adding noise at the same time. A repeater or regenerator, which includes a detector, a laser, and different electronic circuits as described in this chapter, is then employed to reshape the pulses.

The transmitter and receiver units of the link need additional sub-units. The basic device in the transmitter, that is, the laser, is to be properly biased by a driver and the light output power should be accurately controlled by a monitor circuit. An optical amplifier, usually a semiconductor optical amplifier (SOA), may boost the laser power up. A variable optical attenuator (VOA) is sometimes necessary to reduce or control the intensity. In the receiver unit, the optical signal is detected by a photodetector. The weak electrical signal is then amplified first by a low-noise preamplifier and then by power amplifiers. Further processing systems are needed before these signals are converted back to the original format, at which time they are transmitted. The bottom part of Figure 1.5 indicates the occurrence of various devices in the order they appear in an optical link. Further discussion of Figure 1.5 will be made in connection with the integration of devices, which is discussed in Section 1.4.2.

The present-day communication links work around 1.55 μm, the wavelength at which optical fibers have minimum attenuation. The optical amplifiers, the Er-doped fiber amplifiers (EDFAs), also work around this wavelength. Earlier systems used a 1.3 μm
transmission window of the fiber, at which the material dispersion of the fiber is minimum. The DWDM system employs hundreds of wavelengths, separated approximately from each other by about 100 GHz (0.8 nm), covering both the 1.3 and 1.55 μm windows.

The requirement that all the optoelectronic and electronic components must work at the two wavelengths completely rules out the use of silicon-based active devices in the telecommunications network, since the cutoff wavelength of Si is about 1.1 μm. The materials of choice are quaternary alloy InGaAsP and ternary alloy InGaAs grown on InP substrate.

1.4.2 Materials and Integration

There are many differences between today’s microelectronics and photonics industries. While silicon is the only material in the former, a variety of materials are used in photonics. InP substrate is used for source and detector development, silica as the fiber material, different semiconductors and even insulators like LiNbO3 for modulators, an Si platform for passive lightwave circuits used in DWDM, and Si-based ICs for driver and controller circuits. No single material or single technology is leading the market. The production technology is still primitive, and the level of integration is far below the level achieved in microelectronics.

In a truly monolithic IC, all components, that is, electronic circuits, light sources, photodetectors, modulators, waveguides, multiplexers, and so on, are grown on the same piece of semiconductor substrate. Since good sources and modulators have not yet been realized using silicon, and since efficient silicon photodetectors do not exist at 1.55 μm, monolithic integration on a silicon platform is at present ruled out. It is possible, in principle, to use InP substrate for integration. However, the small wafer size, high cost, and other factors limiting the manufacturability have hindered progress in this direction.

Hybrid integration, in which disparate parts are assembled onto one common platform, has been pursued for quite some time. In common hybrid optical components, III–V compound light sources and detectors are attached onto silicon on insulator (SOI), silica, or polymer platforms. These components are commercially available. Figure 1.5 gives examples of discrete components in package, hybrid integration, and monolithic integration. The order of the devices on the axis agrees with the order of appearance of the devices in an optical link. As mentioned already, monolithic integration is almost exclusively on an InP platform; however, photodetectors and transimpedance amplifiers (TIAs) have been grown on silicon.

The main interest in hybrid technologies lies in the combination of III–V semiconductor laser diodes with Si integrated circuits for optical fiber communication or optical interconnects. For this purpose, GaAs or InP is grown on Si and then processed, or, alternatively, laser devices are detached from their substrate by an epitaxial liftoff process and then bonded to Si substrate.

For growth of GaAs and InP on Si, the large lattice mismatch (4% for GaAs and 8% for InP), different thermal expansion coefficients, and fast diffusion of Si as impurities all create difficulties in maintaining low defect densities in compounds for laser production. Although several new techniques have been developed to overcome these difficulties, it is too early to predict the long-term success of the techniques.

In the epitaxial liftoff technique, wet chemical etching is performed and then the III–V heterostructure is floated off and transferred to a planar Si substrate. The bonding occurs due
to the van der Waals force. In the wafer fusion process, the two materials (of high quality) are brought into intimate contact under hydrogen ambient at around 450 °C. Under uniform direct pressure, the substrates form robust chemical bonds. One of the substrates, for example compound semiconductor, is selectively etched and photonic devices grown onto it are integrated with silicon electronic circuits. Once again, it is difficult to predict how far these technologies will be successful in commercial production.

It appears, therefore, that the most satisfactory solution to the above-mentioned problems would be achieved when all the optoelectronic and photonic components could be grown on a single substrate, for which silicon seems to be the best choice. The extensive experience in Si fabrication and processing could then be put to maximum use. Unfortunately, however, the lack of suitable emitter and especially a laser based on Si, as well as of a fast modulator, stands in the way of achieving the coveted goal.

1.5 Data Network

While long-haul optical communication systems work at 1.55 μm, to exploit the minimum attenuation in the fiber, local area networks span smaller distance and area. In this case, working at other wavelengths, at the cost of higher attenuation of signals, may be of advantage due to the availability of cheap components. A possible system employs GaAs-based lasers at around 800 nm, at which wavelength Si photodetectors and other electronic circuits would offer a low-cost solution. If, in addition, Si emitters are available, further reduction in cost is highly expected.

A large part of data communication network is anchored to servers and desktop computers that utilize Si devices. The large potential volume of the market and the competition with copper cables will necessitate more use of inexpensive optical fibers. Si-based photonic components will offer the cheapest solution to the network.

1.6 Conclusions

From the discussions in the above sections, the following points emerge:

- The interconnect problem within a chip is taking an alarming shape. Optical interconnect based on silicon technology may offer a viable solution.
- The long-haul optical communication link employs at present a number of different devices (viz., lasers, modulators, power monitor and control, amplifiers, photodetectors, photoreceivers, multiplexers, demultiplexers, filters and other passive lightwave circuits, and active network components like wavelength converters, etc.). Apart from passive components, most of the active components are fabricated on the InP platform. A truly monolithic OEIC on silicon may offer all the advantages of integration including cost reduction.
- Si-based photonic devices may offer lower cost in the sector of data networks covering shorter distances.
- Si microphotonics seem to be an attractive solution for next-generation optical interconnects for chip-to-chip or board-to-board interconnects.
- Discrete silicon photonic devices like light-emitting diodes (LEDs) and lasers are in demand for consumer electronics, display, and mobile communication, and as mid-infrared or THz emitters.
Si-based passive lightwave circuits are well developed and find use in commercial optical communication and networking systems. Si-based photodetectors technology is also mature, and the devices are already in use. There are attempts to extend the operation to the important telecommunication window at 1.55 μm. On the other hand, the indirect band gap in silicon makes it difficult to realize efficient light emitter and to achieve laser action. This is also the reason why the intrinsic modulation bandwidth of silicon-based modulators is substantially lower than the compound semiconductor-based counterpart.

The challenge in the area of Si photonics is therefore to develop suitable emitter and modulator and then to integrate all the active and passive devices on a single silicon chip. The next challenge would obviously be to increase the level of integration, functionality, and yield coupled with reduction of cost.

1.7 Scope of the Book

In the earlier sections, an attempt has been made to identify the areas where Si photonics may be useful and advantageous. A few photonic devices are in the matured state of development, while a few others, notably emitters and modulators, are in the stage of early research or at most at the development stage in the laboratory. The book aims at providing the basic principles of operation of the devices, the structures of the devices, and an idea of the state-of-the-art developments. The following is a brief description of the chapter-wise coverage of different topics.

Chapter 2 describes the fundamental electronic properties of silicon and its alloys, and of heterostructures made with Si-based materials. The band structure, density-of-states in bulk silicon, and Ge are discussed first followed by similar discussions on Si and Ge-based alloys. A general introduction of heterostructures and band line-up then follows. The special features of Si-based heterostructures, the pseudomorphic growth, are then introduced. The band structure modifications arising out of strain and band offsets and band line-ups are then introduced. Recently, direct band gap has been achieved in tensile-strained Ge layers grown on Sn-based Si alloys. The theory and results obtained are presented.

Chapter 3 is devoted to quantum nanostructures. In the beginning, a simplified picture of quantum confinement and calculation of energy levels, density-of-states in quantum wells (QWs) using well-studied GaAs–GaAlAs material system is given, followed by a brief description of refined theories. Quantum wires (QWRs) and quantum dots (QDs) are then introduced. Similar discussions using Si–SiGe systems are then included, mentioning the recent results on direct band gap systems. Finally, the effect of electric field on the subband structures is discussed.

Chapter 4 gives an introduction to the optical processes in bulk semiconductors. The semiclassical theory of absorption is first introduced. Both direct gap and indirect gap materials are covered. The relationship between absorption and gain is then established. The chapter then discusses other types of absorption and different forms of radiative and nonradiative recombination. The basic idea of exciton formation and a simplified theory of excitonic absorption are then presented.

Chapter 5 is similar in structure as in Chapter 4; however, here the basic theory of optical processes as modified for QW, QWR, and QD is presented. The theory of absorption, gain, and recombination in direct and indirect gap materials and in different types of band
alignments is presented. Excitonic processes and the effect of electric fields on the properties of excitons, particularly in QWs, are discussed.

The remaining chapters are devoted to silicon photonic devices, covering description of the structures, operating principles, performance, and the application areas. These chapters may be grouped into two categories. Active devices including light emitters, modulators, photodetectors, and recently reported Raman lasers form the first group to which Chapters 5–9 belong. Passive lightwave circuits belong to the second group, and Chapters 10–12 discuss various aspects related to the circuits. The topics covered are the propagation of electromagnetic waves, waveguiding action, loss mechanisms, coupling between waveguides, various passive devices used in dense wavelength division multiplexed fiber-optic communication systems, and the application areas. The device fabrication processes are described in Chapter 13.

Chapter 6 discusses the phenomena of light emission in silicon, related materials, and nanostructures formed using Si, Ge, and their alloys. The devices reported are still at the primitive stage. The chapter makes a list of all the methods followed so far to achieve efficient light emission, the structures of the devices, the basic principle of light emission in each structure, and the performance achieved so far.

Chapter 7 deals with Si-based light modulators. This device also is in the early stage of development. The different approaches followed by different authors, the results obtained by them, and possible ways of improving the performance form the subject of discussion in this chapter. Some recent breakthroughs in the area are also reported.

Chapter 8 is devoted to silicon photodetectors. It starts with a general discussion of optical receiver systems, and it discusses optical detection from an engineering perspective and the most important performance measures of photodetectors. Then various examples of silicon-based–pn photodiodes, pin photodetectors, Schottky barrier photodiodes, avalanche photodiodes, and bipolar and MOS phototransistors are presented and discussed. Various examples of photodiodes and of phototransistors in standard silicon CMOS and BiCMOS technologies are described. Important emerging technologies such as silicon-on-insulator, and various types of photodetectors that could be fabricated using heteroepitaxial techniques such as silicon–silicon germanium multiple quantum wells and germanium-on-silicon, are presented and discussed. The chapter concludes with related theoretical discussion from a practical perspective so that time constants, operating frequencies, and signal-to-noise ratios can be calculated. Throughout the chapter, numerical examples are provided to add numerical details to the discussions.

Chapter 9 is devoted to stimulated Raman scattering and other nonlinear effects in silicon. After giving a brief idea of Raman scattering, recent results of stimulated Raman gain in silicon and the structure and characteristics of Raman lasers are described. The chapter ends with a brief discussion of other nonlinear effects in silicon that play important roles in silicon-based photonic devices.

Chapter 10 gives the introduction to the waveguides that form the basic unit of all passive lightwave circuits to be discussed in later chapters. This chapter explains the principle of waveguiding in dielectric waveguides from both the ray-optic theory and the electromagnetic theory. The simple forms of slab and three-dimensional waveguides are presented and the propagation characteristics in these structures are then discussed. This chapter then introduces various sources of loss in a waveguide and underlines the
principle of calculating the loss. It also discusses how waveguides are coupled to sources, detectors, and other elements.

Chapter 11 discusses the principle of operation of several planar waveguide devices used in actual optical communication systems and networking. Coupling between different waveguides occurs in many such devices. The general coupled mode theory is presented first, followed by the application of the theory to two important classes of devices: directional coupler and Bragg grating. The operating principles of splitters, directional couplers, Mach–Zehnder interferometers, Fabry–Perot resonators, Bragg gratings, mirrors, ring resonators, and resonant cavities are then presented.

Chapter 12 embodies the description of important devices used in DWDM systems. At first, the arrayed wave guides (AWGs), extensively used in present systems, are described in detail including the structure, principle of operation, performance, and so on. The different material systems on silicon platform are then discussed, and some comparative figures are quoted. The chapter includes two more devices, the Bragg grating in planar form and the Eshelle grating.

Chapter 13 is devoted to description of device fabrication processes. The different growth processes, like epitaxy and chemical vapor deposition (CVD), are first discussed. The planar technology, lithography, and etching processes are then described. Different methods used for the fabrication of waveguides are also presented in this chapter.

References


