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, it is the prowess and maturity of silicon-based technologies that spawned microelectronics, and now the nanoelectronics era, that impact almost all aspects of our lives.

The rapid and continued growth of the silicon industry has followed the now-famous Moore's law, which states that the number of transistors in an integrated circuit chip doubles every 12 months (initially, but was later revised to every 18 months). However, during the last few years, there has been an indication of the decline of Moore's law. In the semiconductor industry, there are challenges on how future silicon-based integrated circuits (ICs) will deliver the same advantages and functionalities as shown today, but research and technology development work continue. Accordingly, the International Roadmap for Semiconductors (ITRS) and its successor, the International Roadmap for Devices and Systems (IRDS - https://irds.ieee.org/editions/2023), continue to play a leading role in guiding the evolution of silicon-based semiconductor technologies. In fact, the 2023 IRDS edition provides key trends and a roadmap related to silicon-based devices and systems with a 15-year horizon.

On the one hand, with current silicon technologies, a prominent weak point is that proper light emitters and modulators cannot be easily realized by using it due to the indirect nature of silicon's 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. The 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 level of integration and performance are far below what has already been achieved in electronic integration. In fact, it is natural to expect that monolithic optoelectronic integrated circuits (OEICs) will provide the same advantages such as low cost due to batch fabrication, high functionality, and scaling for denser integration, the same advantages offered by silicon ICs.

In the past several years, there have been several research and technology development efforts to grow OEICs on silicon and to integrate them with electronic ICs, by using the same or compatible production facilities for silicon ICs. The goal is to leverage the mature manufacturing processes and material properties of silicon that have been pioneered and perfected over decades in the semiconductor industry to create photonic OEICs. This integration can revolutionize sectors ranging from telecommunications and data communications to sensing and computing, offering solutions to meet the ever-increasing demand for faster, more efficient, and scalable information processing, all at the low cost that is common in silicon ICs. This is the dream cherished by many workers over the last few decades, a dream in which there has been progress, but it is not yet fully materialized. In spite of this, Si-based photonics remains an active area of research, and over the last decade, some significant milestones have been achieved. For example, in a recent 2023 article discussing the goal of integrating lasers with silicon photonics, the merits and challenges of four strategies were proposed—flip-chip processing, microtransfer printing, wafer bonding, and monolithic integration [1]. In addition, there are other recent publications on photonic-electronic co-integration [2], "*roadmapping the next generation of silicon photonics*," [3] hybrid integrated silicon photonics [4], and "*perspective on the future of silicon photonics and electronics*." [5]

Another important area of application of silicon photonics is in very-large-scale integrated (VLSI) circuits themselves. The complexity of present-day ICs has reached such a high level that the interconnects within them are formed in a number of levels. At present, the number of metal layers used for interconnects is between 12 and 15 for decananometer silicon technology. The metallic interconnects, mainly copper (Cu), provide delay due to RC time constants, which far exceeds the transit time delay associated with the individual transistors. In addition, they also affect power consumption and chip design because of the complex trade-offs between performance, power consumption, and area utilization. If the increase in speed is to be maintained at the same rate for the next generation of ICs, then the interconnect bottleneck must be properly addressed. Optical interconnects are believed to be the right solution to the interconnect bottleneck problem.

In the present chapter, we shall give an overview of the developments in silicon electronics, the present status, and the problems faced in achieving the above-mentioned goals. 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 the presence of silicon photonics is needed, i.e., communication and interconnections in chips, will be described then, giving the reader an idea of the present scenario and the problems faced. In this connection, the alternative to monolithic integration, i.e., 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 texts, 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 [1–21].

1.2 VLSI: Past, Present, and Future Roadmap

The announcement of the first point-contact transistor on 23 December 1947 marked 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 was introduced in 1954. The concept of IC was proposed in 1958, and its working was demonstrated by using discrete components. A few months later, the 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 ($\sim 10^4$ circuits per chip) compared to today's VLSI standard have prevented their widespread use in silicon semiconductor technology.

The control of conductivity on the surface of a semiconductor by an external electric field was proposed in the 1920s. Attempts at conductivity modulation during the early part of the 1950s were not successful because of problems such as low material purity, poor device reliability, and manufacturing techniques still in their infancy. The first metal-oxide semiconductor field-effect transistor (MOSFET) using SiO_2 as the gate insulator on a 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, 22]. Since then, there have been remarkable advances in silicon CMOS technology that have enabled the exponential growth in computing power and miniaturization of electronic devices. In fact, the advances in silicon CMOS technology have enabled the creation of faster, smaller, and more powerful electronic devices, which in turn drove innovation in various fields, including computing, telecommunications, automotive, and consumer electronics.

The advantages of silicon as an electronic material are numerous. Silicon is available in nature abundantly – in fact, it is the second most abundant element in the Earth's crust (after oxygen). It can be purified to a very high level. The native oxide – silica – is a very good insulator, is stable, and can withstand a large field across it; above all, the interface charge between silica and silicon can be reduced to a minimal level. SiO_2 can easily be patterned by photolithography [23]. Its bandgap of $\sim$1.1 eV allows for effective switching between "*on*" and "*off*" states that are essential for transistor operation.

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 higher packing density, and ease of fabrication with a reduced number of masks have made the CMOS technology on a silicon substrate the standard technology followed by industries [24]. Figure 1.1 illustrates the developments of VLSIs over several decades [25].

At the rate shown in Figure 1.1, there are more than one billion transistors on a single die. With the increasing number of transistors per die, the minimum feature size, or roughly the channel length in a single transistor, is just a few nanometers in 2024. For almost 30 years, the feature size in IC lithography has been reduced at a rate of 0.7 × every 3 years, to just a few nanometers in 2024.

At present, VLSI technology has reached unprecedented levels of complexity and integration, driven largely by advancements in semiconductor processes such as lithography, materials science, and manufacturing processes. This has led to nanometer-scale transistors allowing for billions of transistors per chip; multicore processors and parallel architectures to optimize performance and manage energy efficiency; and specialized accelerators such as GPUs (graphics processing units), TPUs (tensor processing units), and AI accelerators,

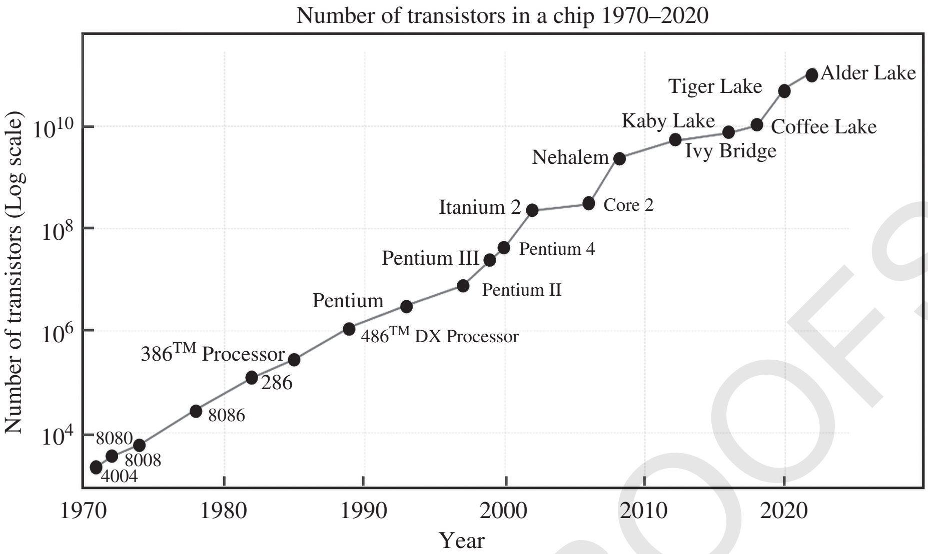

Figure 1.1 *Graph showing the number of transistors in a chip and the evolution of microprocessors since 1970 (See for example [25–28]).*

engineered for specific tasks such as deep learning inference and training, machine learning, and artificial intelligence.

In the future, we can expect that VLSI will face some key challenges. The trend of continued downscaling leads to physical limitations and costs associated with shrinking transistor sizes below a few nanometers, as well as the challenges of developing new lithography tools and new materials. Another trend is related to new system architectures, including 3D integration, in which chips are stacked vertically, brain-inspired architectures for neuromorphic computing, and leveraging quantum mechanics for computation in quantum computing, for which there are many design and development challenges to be overcome. A third trend is the growing use of VLSI chips in critical infrastructure and Internet of Things (IoT) devices for which the challenges are ensuring security against cyber threats and reliability through fault-tolerant design.

1.3 The Interconnect Problem in VLSI

On-chip interconnect is nothing but electrical wiring. According to ITRS and its successor, IRDS, an interconnect is electrical wiring that distributes clock and other signals among transistors, logic gates, memory cells, etc., and provides power/ground to and among the various circuits/systems functions on a chip.

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

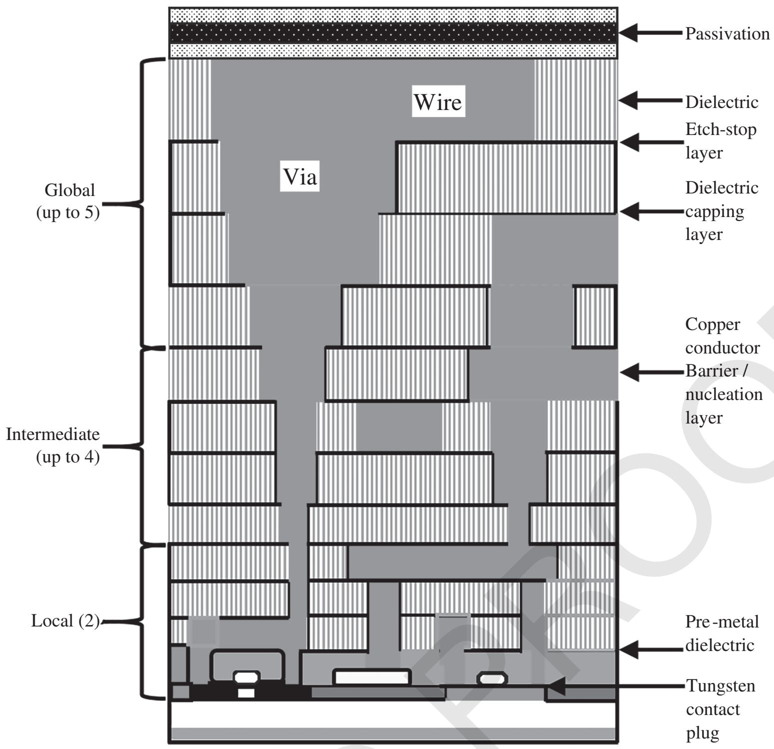

Figure 1.2 *Schematic view of electrical interconnect in VLSI.*

The earlier approach used aluminum deposition and dry etch definition. Currently, copper wires are used instead. 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 gates, sources, and drains of close MOSFETs of the same functional units, and their length scales with the gate size.

Although downscaling of transistors increases the speed, the same is not true for downsizing of interconnects. With the increasing density of components on a chip, the interconnects become thinner and longer, resulting in higher resistance and capacitance. In earlier ICs, used around the 1980s, the delay in interconnects was propagation-limited, i.e., limited by the time of propagation of the electromagnetic waves associated with SiO_2, rather than the *RC* time constant, which was ~1 ps. For example, for the 35-nm technology generation, the interconnect response time of a 0.1-mm copper line with a low-k dielectric ($k = 2$) is about 250 ps, about two orders of magnitude higher, and accounts for the delay related to the *RC* time constant of the wire. Since the signal propagation delay through interconnects becomes comparable to or higher than gate delays, the signal integrity can degrade, leading to problems such as noise, unwanted coupling between adjacent wires (crosstalk), and timing mismatches.

The second issue with interconnections is power consumption. The high resistance in interconnects leads to increased power dissipation because more power is required to drive signals through longer wires. This increases the overall power consumption and heat generation, which can in turn impact reliability and efficiency. As noted previously, with the scaling of transistors, the power consumption of the interconnect exceeds that of a transistor. For example, for 1.0-μm devices, the switching energies in the transistor and a 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 insulator, copper and low-k dielectric, and projected optical waveguide technology. It indicates that Al-based technology has reached the performance limit at 0.55 μm; the reduced resistance due to Cu and reduced capacitance due to low-k dielectric ensure a performance improvement limit up to 0.18-μm technology.

Future advancements in VLSI will require further reductions in interconnect delays and power consumption while enhancing signal integrity and reliability. To achieve this goal, several strategies are being pursued. One strategy is related to innovations in materials science, such as the development of new conductive and insulating materials. Another strategy is the integration of optical interconnects for high-speed data transmission. And a third

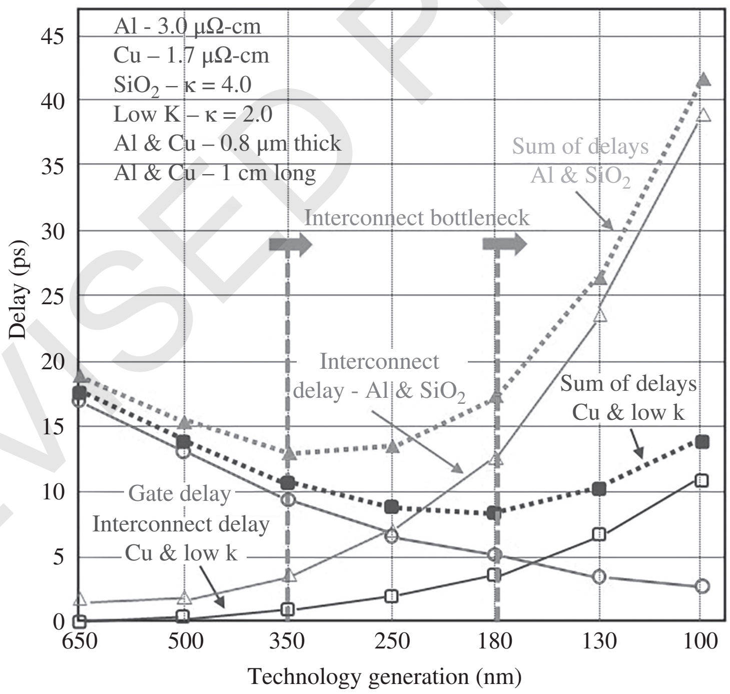

Figure 1.3 *Variation of interconnect propagation delay with feature size for Al/SiO_2, Cu/low-k dielectric and projected optical interconnect technology.*

strategy is with novel architectures that can minimize the impact of interconnect scaling on overall chip performance. Addressing the interconnect problem is crucial for the development of next-generation computing, communication, and sensory applications.

Here, we focus on optical interconnects using silicon microphotonics technology, as it offers a very promising solution to the *RC* time delay associated with traditional metal interconnects as well as interconnects at the inter-chip and inter-board levels. Using photons as bits of information, instead of electrons, a speedier performance of the devices is expected. The use of photons also solves the power dissipation problem. Photons propagate in transparent media with less heat dissipation and almost no crosstalk. 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 across the globe like a spider's web. And still, a larger number of links is being continuously added. These long-haul links are high-capacity telecommunications connections designed to transmit large volume at high data rates over distances ranging from hundreds to thousands of kilometers. Further, these links form the backbone of our global communication networks, providing the transmission of huge amounts of data between cities, countries, and continents at very high data rates with high signal fidelity and low latency.

Today's fiber-optic links employ dense wavelength division multiplexing (DWDM), in which a huge amount of data is carried by hundreds of carrier wavelengths, each modulated at a high bit rate ($\sim$10 Gb/s or more) and transmitted by a single strand of fiber. DWDM systems are used in long-haul telecommunications networks, metropolitan and access networks, as well as data centers, high-performance computing, and cloud computing platforms. In the following, first, we shall discuss the basic link, and second, the components used [29]. In the next subsection, 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

The key components of a basic link are the following:

Transmitters: This includes semiconductor diode lasers that convert electrical pulses into coherent light signals at specific wavelengths to carry data. Transmitters include laser diodes, modulators, and multiplexers in wavelength division multiplexing systems.

Wavelength division multiplexers and demultiplexers: Wavelength division multiplexing (WDM) is used to combine multiple wavelengths of light pulses to increase the data capacity of the optical fiber. Note that each wavelength acts as an independent channel. Demultiplexers are at the receiving end of the channel and are used to separate the multiple wavelengths back into their original individual wavelengths to be detected.

Optical Fiber: This is the transmission medium for optical communications and is usually made of plastic or glass to guide the data-carrying light pulses. Optical fibers typically have very low attenuation (signal loss) and dispersion over long distances, especially at 1.3 and 1.55 μm, allowing them to carry signals over long distances with minimal signal degradation.

Optical Amplifiers: Erbium-doped fiber amplifiers (EDFAs) are placed along the fiber cable route to amplify optical signals periodically without converting them back to electrical signals. With EDFAs, the optical signals can travel longer distances.

Receivers: At the receiving end of the fiber link, photodetectors convert the received optical signals back into electrical signals. Receivers include photodetectors, signal conditioning and error detection and correction ICs, and demultiplexers in WDM systems.

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 (λ_1 for example). The laser may either be directly modulated, or an external modulator may be used to impress the signal on the laser beam. A multiplexer (Mux) combines the modulated signals coming from the bank of lasers, and the combination is transmitted by an optical fiber. After traversing a long distance (usually several hundreds of kilometers) in the fiber, the signal becomes significantly attenuated and the digital pulses considerably spread due to material dispersion of the fiber; then the combined signal needs to be regenerated by a regenerator (this is not shown in Figure 1.4).

In the regenerator, a photodetector converts the weak and distorted optical signals 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. Sometimes, a number of optical amplifiers, such as EDFAs, are inserted at regular intervals in the link to boost the intensity of optical signals, but they add noise at the same time. In this case, a repeater or regenerator may not be needed to reshape the pulses. The received signal at the end of the fiber-optic link is then de-multiplexed (DeMux) into the individual data wavelengths, which are detected using photodetectors.

The transmitter and receiver units of the link need additional subunits. The basic device in the transmitter, i.e., the laser, is to be properly biased by a driver, and the light output

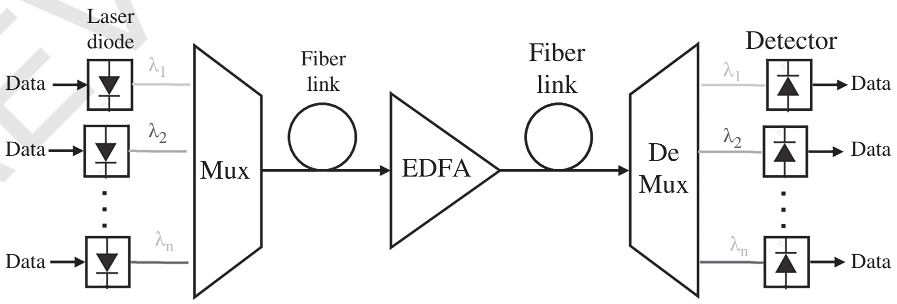

Figure 1.4 Schematic diagram of a WDM point-to-point communication link.

power should be accurately controlled by a monitor circuit. An optical amplifier, usually a semiconductor optical amplifier (SOA), may boost the laser power. A variable optical attenuator (VOA) is sometimes necessary to reduce or control the intensity. In the receiver unit, each demultiplexed 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 (e.g., signal conditioning and error correction) are needed before these signals are converted back to the original format in which they were 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 the figure will be made in connection with the integration of devices to be discussed shortly.

The present-day long-haul optical communication links work around 1.55 μm, the wavelength at which optical fibers have minimum attenuation. The 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. DWDM enables higher channel densities and greater overall data capacity.

The requirement that all the optoelectronic and electronic components must work at the two wavelengths rules out the use of silicon-based active devices in the telecommunication 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 an InP substrate.

In a fiber-optic communication link, there are many challenges to be addressed. They include dealing with attenuation and dispersion.

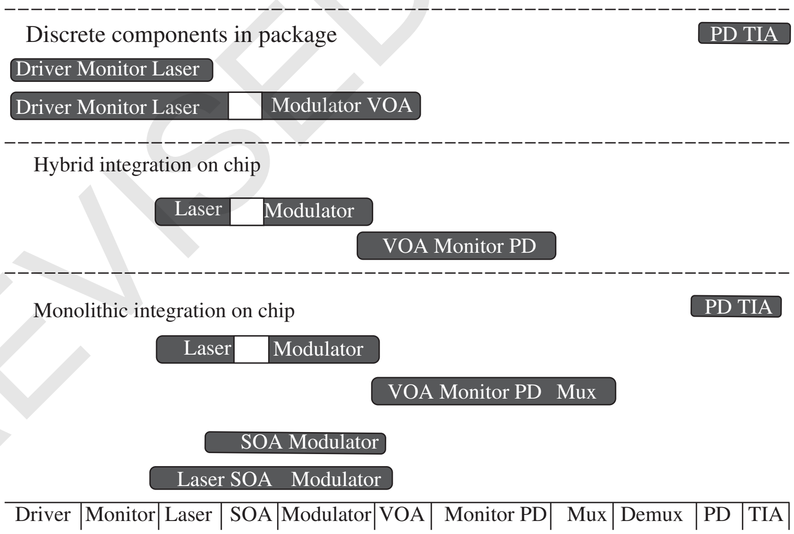

Figure 1.5 Level of device integration in commercially available products. Reproduced with permission from [7]. Copyright (2004) Springer Science + Business Media.

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 $LiNbO_3$ for modulators, a 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 evolving, and the level of integration is below the level achieved in microelectronics. In addition to silicon, silicon dioxide (SiO_2) and silicon nitride (Si_3N_4) are commonly used as insulating materials or as cladding layers in silicon photonics waveguides. SiO_2 and Si_3N_4 are used to guide or confine light within the silicon structure. Also, Si_3N_4 has low optical loss and broader transparency windows compared to SiO_2, making it very suitable for high-performance waveguides and resonators.

In a truly monolithic IC, all components, i.e., electronic circuits, light sources, photodetectors, modulators, waveguides, multiplexers, etc., 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 an InP substrate for integration. However, the small wafer size, high cost, and other factors limiting the manufacturability have hindered the 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 the InP platform; however, photodetectors and transimpedance amplifiers (TIAs) have been grown on silicon. For example, germanium grown on silicon is used as the photodetector material because of its high absorption coefficient for light at near-infrared wavelengths.

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 lift-off process and then bonded to a Si substrate.

For the 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 lift-off 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 a hydrogen ambient at around 450°C. Under uniform direct pressure, the substrates form robust chemical bonds. One of the substrates, e.g., a 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 therefore appears that the most satisfactory solution to the previously mentioned challenges 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 as current Si nanofabrication allows for precise control over dimensions and properties of photonic components at the nanoscale, which ensures high performance and excellent reliability. Unfortunately, however, the lack of a 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.

As a summary of this section, it is noted that materials and integration technologies will play a crucial role in advancing the development of silicon photonics. This would then allow for the development of compact, high-performance photonic devices integrated with silicon ICs. However, we require continued innovation in materials science and fabrication and integration technologies to drive further advances with the expected goal of broader adoption of silicon photonics in industries related to high-speed communications, high-performance computing platforms, biomedical sensing, and quantum photonics.

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 a smaller distance and area. In this case, working at other shorter 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 the 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.

Recent developments in the application of Si photonics in data centers, advances in different Si-based photonic devices, component integration in Si photonics chips, advances in electronics-photonics co-packaging, and integrated Si photonics examples in different interconnect industries are summarized in the paper by Margalit et al. [5]. A brief list of emerging applications of Si Photonics, e.g., in LIght Detection And Ranging (LIDAR), sensors, biosensors, gyroscopes, and spectrometers, is also given there. A very recent discussion of silicon photonics for LIDAR applications was given in [30]. Here, a discussion of how silicon photonics merges LIDAR's precision with Radio Detection And Ranging's (RADAR's) all-weather reliability, creating compact, high-performance sensors for autonomous vehicles, was presented. The synergy of LIDAR and RADAR enhances safety through superior object detection, overcoming individual limitations in challenging environments.

1.6 Conclusions

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

- The interconnection problem within a chip is taking an alarming shape. An 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, namely, lasers, modulators, power monitors and controls, 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 appears 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, 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 photodetector 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 an efficient light emitter and to achieve laser action. This is also the reason why the intrinsic modulation bandwidth of a silicon-based modulator is substantially lower than that of a compound semiconductor-based counterpart.

The challenge in the area of Si photonics is therefore to develop a 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 a 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 mature 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 to provide 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 is concerned with the basic properties of silicon, the foundation of modern electronics. Silicon possesses structural, electronic, and optical properties that define its dominance in microelectronics and its emerging role in photonics. Its diamond cubic crystal structure influences its electronic behavior, with an indirect bandgap of 1.12 eV at room temperature. Unlike direct-gap semiconductors such as gallium arsenide (GaAs), silicon requires phonon-assisted transitions for electron-hole recombination, making it inefficient for light emission. However, advances in strain engineering and heterostructures are transforming this limitation. The chapter explores silicon's band structure, comparing it with GaAs and germanium (Ge). Silicon's six Δ-valleys and Ge's L-valleys highlight the challenges of indirect transitions, while spin–orbit coupling splits the valence band into heavy-hole (HH), light-hole (LH), and split-off (SO) bands. Temperature and pressure shift band edges, while strain from heteroepitaxy significantly alters band alignment, enhancing carrier mobility and confinement. Heterojunctions, such as sicon–silicon germanium (SiGe) interfaces, leverage strain to modify electronic properties. Tensile strain in silicon improves electron mobility by lowering the Δ_2 valleys, whereas compressive strain in SiGe raises the valence band edge, benefiting hole transport. These effects enable high-performance devices like heterojunction bipolar transistors (HBTs) and quantum wells (QWs). Doping further refines silicon's properties, with donors and acceptors introducing shallow energy levels. SiGe and GeSn alloys enable bandgap tuning and quasi-direct transitions, crucial for silicon-compatible optoelectronics. Emerging techniques in strain engineering and nanostructuring are unlocking silicon's potential for photonic applications, paving the way for integrated photonic-electronic systems in data communication, sensing, and quantum computing.

Chapter 3 explores quantum structures—nanoscale semiconductor systems where quantum confinement tailors electronic and optical properties. Key systems include QWs, quantum wires (QWRs), and quantum dots (QDs), each reducing spatial dimensionality to create discrete energy levels and modified density of states (DOS). In QWs, electrons are confined in one dimension, forming a 2D electron gas with quantized subbands. The DOS becomes step-like, enabling applications like high-speed transistors and lasers. QWRs (1D confinement) and QDs (0D confinement) exhibit sharper DOS singularities and atom-like discrete states, ideal for single-photon sources. The chapter contrasts compound semiconductors (e.g., GaAs/AlGaAs) with silicon-based systems, where strain and anisotropic effective masses complicate subband engineering. Superlattices introduce periodic potentials, creating minibands for tailored carrier transport. External electric fields induce quantum-confined Stark effects (QCSEs), shifting energy levels for electro-optic modulation. By leveraging quantum confinement, these structures revolutionize optoelectronics.

Chapter 4 gives an introduction to the optical processes in bulk semiconductors. Different optical constants in the conductor and dielectric, and their interrelations, are obtained using Maxwell's equations, Poynting vector, etc. Three basic processes in the interaction of light with matter are absorption, spontaneous emission, and stimulated emission. The concept of defining absorption and emission rates for a two-level atomic system in terms of Einstein's coefficients can be extended to the optical processes in semiconductors. The origins and characteristics of absorption in semiconductors are discussed by semiclassical theory. Conservation laws for fundamental absorption in direct band-gap semiconductors show that the momentum vector for the transition remains unchanged. Absorption coefficient calculation stems from the Fermi's Golden rule, and the perturbation methods of quantum mechanics are employed. In a direct band-gap semiconductor, the first-order perturbation method is employed. In an indirect band-gap semiconductor, the calculation requires additional participation of phonons, and second-order perturbation theory is employed. If there is population inversion in a direct band-gap semiconductor, then stimulated emission can occur, and when it exceeds absorption, it gives rise to gain in the medium. Quasi–Fermi levels should satisfy certain conditions to achieve gain. Intervalence band absorption between the LH, HH, and SO bands is possible in a p-type semiconductor. Free carrier absorption (FCA) can occur for conduction band (CB) electrons and valence band (VB) holes, which may affect the refractive index as well as attenuation. Different types of recombination may occur in a semiconductor. Radiative recombination gives rise to luminescence. Nonradiative recombination includes recombination via traps: Shockley-Read-Hall (SRH) process, Auger recombination, surface recombination, and recombination via complexes. Excitonic absorption, particularly at low temperature, is shown as line spectra in the absorption spectra below the band-gap energy. Impurity states and isoelectronic states also play an important role in certain absorption/recombination processes.

Chapter 5 (similar in structure to Chapter 4) explores optical processes in quantum structures, focusing on QWs, QWRs, and QDs. Unlike bulk semiconductors, these nanostructures exhibit quantum confinement effects, leading to discrete energy levels that enhance optical properties such as absorption, recombination, and gain. In direct-bandgap QWs, absorption is step-like due to 2D DOS, differing from the continuous absorption in bulk materials. The overlap of electron-hole wavefunctions (governed by the envelope function overlap integral, C_{mn}) critically influences transition probabilities. Gain in QWs arises when carrier injection creates population inversion, with peak gain occurring near the effective bandgap. For indirect-bandgap materials (Si, Ge), phonon-assisted transitions dominate, producing distinct absorption spectra. Type-II QWs, where electrons and holes are spatially separated, exhibit weaker absorption due to reduced wavefunction overlap. Intersubband transitions (within the same band) enable applications like quantum cascade lasers (QCLs) and quantum well infrared photodetectors (QWIPs). These transitions are highly polarization-dependent, with TE (transverse electric) and TM (transverse magnetic) modes governed by selection rules. Excitonic effects are stronger in QWs than in bulk due to enhanced Coulomb interaction. Excitons in 2D systems have fourfold higher binding energies, making them observable even at room temperature. Electric fields (via the quantum-confined Stark effect, QCSE) shift exciton peaks, useful for electroabsorption modulators. QWRs and QDs exhibit sharper DOS, leading to higher gain and lower threshold currents. However, size fluctuations in QDs broaden absorption peaks, reducing performance. Recent advances in Ge/SiGe QWs demonstrate strong QCSE, enabling Si-based optoelectronic devices.

Chapters 6–12 are devoted to silicon photonic devices, covering descriptions of the structures, operating principles, performance, and 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 6–9 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. Chapter 13 is dedicated to nonlinear silicon photonics, where nonlinear optical processes, along with theoretical aspects, resulting phenomena, and applications are discussed. A detailed discussion on fabrication techniques and material systems is given in Chapter 14.

Chapter 6 discusses the phenomena of light emission in silicon and related materials. Light emission is due to the recombination of electrons and holes and to have an efficient light emitter nonradiative recombination lifetime should be much larger than the radiative recombination lifetime. Silicon, being an indirect band-gap semiconductor, defects arising out of its alloy disorders may provide the necessary momentum for recombination, such as heterojunction of Si and $Si_{1-x}Ge_x$ or $Si_{1-x-y}Ge_xC_y$. Quantum confined structures provide spatial confinement, and following Heisenbug's uncertainty relation, the spread of momentum may satisfy the momentum conservation requirement without the need for any phonon. Porous silicon, Si nanocrystals, Si-based QWs, wires, and dots have demonstrated light emission from silicon-based structures. Impurities, such as isoelectronic and rare-earth impurities, bind excitons in silicon and dissociate to give light emission. The prospect of stimulated emission in silicon or Si-based devices is indicated. QCL based on p-SiGe/Si for THz emission is an intersubband recombination process. Light emission is also reported using a Si–Ge superlattice structure, a tensile-strained Ge layer, and GeSn layers grown on a Si platform.

Silicon-based light modulators have emerged as indispensable components for optical communications and integrated photonic systems. Unlike conventional III–V compound modulators that exploit the linear Pockels effect, silicon's crystalline symmetry demands innovative approaches to light modulation. Chapter 7 examines three principal mechanisms that overcome this limitation: *free-carrier plasma dispersion*, where injected or depleted charge carriers dynamically alter both refractive index and absorption; the *thermo-optic effect*, offering precise refractive index tuning ($dn/dT \approx 1.86 \times 10^{-4}\,K^{-1}$ at room temperature) through thermal control; and the *quantum-confined Stark effect* (QCSE), particularly effective in strained Ge–SiGe QWs for electroabsorption modulation. The architectural landscape features Mach–Zehnder interferometers (MZIs) that deliver high-speed operation through phase modulation, albeit with millimeter-scale footprints. In contrast, resonant structures like ring and Fabry–Pérot cavities provide sub-micron compactness and high quality factors, though they require careful thermal stabilization. MOS capacitor-based designs stand out for their CMOS compatibility, achieving bandwidths exceeding tens of GHz through efficient carrier accumulation. Performance benchmarks reveal critical trade-offs: ring resonators achieve impressive 15 dB modulation depths, while reverse-biased p-n junctions push bandwidths to tens of GHz. Optimized designs demonstrate remarkably low power consumption down to 25 μW. Recent breakthroughs include Ge–SiGe QCSE modulators with high contrast ratios, hybrid MZI-ring configurations that significantly slash drive voltages, and monolithic integration techniques enabling dense on-chip photonic networks. Despite these advances, challenges persist in harmonizing speed, power efficiency, and miniaturization while addressing silicon's inherent indirect bandgap constraints. Emerging solutions incorporating strain engineering and III–V material integration herald a new era of performance enhancements, positioning silicon photonics at the forefront of optical innovation.

Chapter 8 is devoted to silicon photodetectors, which are solid-state devices that convert light into electrical signals, enabling applications in imaging, fiber-optic communications, and sensing. These detectors operate based on the internal photoemission effect, where photons with energy exceeding the semiconductor's bandgap generate electron-hole pairs, producing a photocurrent under an applied electric field. Silicon's indirect bandgap limits its absorption to wavelengths below $\sim$1.1 μm, but alloying with germanium extends detection into the infrared. In this chapter, key performance metrics that are discussed include quantum efficiency (QE)—fraction of incident photons converted to detectable electrons; responsivity—photocurrent per unit incident optical power (A/W); bandwidth—determined by carrier transit time and RC time constants; noise—includes shot noise, thermal noise, and dark current, impacting sensitivity; and gain—avalanche photodiodes (APDs) provide internal amplification via impact ionization. Different photodetector types, which include p-n and p-i-n diodes—simple structures with trade-offs between speed and sensitivity; Schottky and metal-semiconductor-metal (MSM) detectors—high-speed but lower responsivity; APDs—high gain but introduce excess noise; and single-photon avalanche diodes (SPADs)—operate above breakdown voltage for single-photon detection, are described. Also, challenges include integrating silicon photonics with CMOS, reducing noise, optimizing spectral response, and improving cost-effectiveness. Emerging solutions involving silicon–germanium heterostructures, resonant cavities, and advanced quenching circuits for SPADs are described. The chapter also includes 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 the Raman laser in silicon. It is based on the stimulated Raman scattering (SRS) in the medium, which is one of the third-order nonlinear processes discussed later in Chapter 13. Raman scattering is an incoherent scattering due to the interaction between the incident photon and optical phonons, resulting in Stokes and anti-Stokes lines in the scattered signal. A simplified theory shows that there exists a certain condition for SRS to occur. Differential equations representing the forward and backward propagation of pump and Stokes signals can be used to find the round-trip gain of the Stokes signal and, hence, the threshold pump power for the Raman laser. A CW Raman laser was reported later compared to the pulsed version of the laser because of the loss due to FCA arising from two-photon absorption (TPA). Ways to circumvent this problem, such as a reverse-biased p-i-n structure to wipe out free carriers, etc., led to the CW operation of the Si Raman laser. Further improvements in performance in the Si Raman laser are also being reported.

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 optics theory and the electromagnetic theory. An effective index can be defined for the propagating mode in a waveguide. In a slab waveguide, there is a cut-off condition for a propagating wave, and from the mode profiles, the confinement factor can be calculated. Practical waveguide (3D) geometries include ridge waveguide, strip-loaded waveguide, and buried channel waveguide. Though the mode patterns in a 3D waveguide are obtained numerically, there are approximate analytical methods, such as Marcatili's method and the effective index method, to obtain closed-form analytical solutions when the guided mode is far from cut-off and the aspect ratio is larger than unity. Various sources of loss in a waveguide are scattering loss, absorption loss, radiation loss, and coupling loss. Coupling of light into the waveguides can be done by using a grating and butt/end-fire coupling. Tapers are used for coupling different cross-sections of two 3D geometries. Other than waveguide structures, photonic crystal structures, and plasmonic structures may also be used for guiding light waves in an efficient and compact manner.

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. Coupling of one mode into another mode may be understood from the general coupled mode theory. This can be applied to study the directional couplers and distributed Bragg reflectors. The operating principles of splitters, dual-channel directional couplers, Mach–Zehnder Interferometer, Fabry–Perot resonators, Bragg gratings, dielectric mirrors, ring resonators, and VOA are then presented. A multimode interferometer is an important waveguide component of a photonic integrated circuit to elevate the data transmission capability.

Chapter 12 embodies the description of important devices used in dense wavelength division multiplexing (DWDM) systems. The arrayed wave guides (AWGs), extensively used in present systems, are described in detail, including the structure and principle of operation. Its performance characteristics include tuning, free spectral range, frequency response, channel crosstalk, and polarization dependence. There are methods to make the grating insensitive to changes in polarization and temperature, and to achieve a flat frequency response. AWGs have applications in multiplexers/demultiplexers, wavelength routers, multiwavelength receivers/transmitters, add-drop multiplexers, optical cross-connects, and dispersion equalizers. The different material systems on the 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 Echelle grating.

Chapter 13 deals with the optical nonlinearity in silicon. Nonlinearity gives rise to interesting properties not achievable by linear behavior. The enhanced nonlinearity due to high refractive index contrast of core-cladding in silicon-based waveguide devices, small cross-section in integrated devices, and CMOS compatibility are strong motivations for developing nonlinear silicon photonic devices. The polarization-to-field relationship shows that for low intensity of the optical field, it is a first-order or linear optical process, and with a large enough electric field, higher-order or nonlinear optical processes become significant. Second- and third-order nonlinear processes are usually discussed from an application point of view. Silicon shows third-order nonlinearity, and by ways of breaking inversion symmetry, silicon can also be used to show second-order nonlinearity. Such nonlinearities show various phenomena, such as sum frequency generation (SFG), difference frequency generation (DFG), second harmonic generation (SHG), third harmonic generation (THG), optical Kerr effect, self-phase modulation (SPM), cross-phase modulation (XPM), and four-wave mixing (FWM). Phase matching is a requirement to obtain efficient output from several nonlinear processes. The propagation of an optical pulse in a nonlinear and dispersive medium can be understood from the wave equation, which can be reduced to a nonlinear Schrödinger equation (NLSE) under the slowly varying amplitude assumption. There are some standard approaches for solving the NLSE. Typical silicon-based structures showing nonlinearity include strip waveguide, slot waveguide, and microcavity resonator structures. The various nonlinear processes in silicon are studied by single or coupled NLSEs, depending on the process. Second-order nonlinearity in silicon finds applications in the generation of green light, high-resolution spectroscopy, the generation of THz signals, optical rectification, Pockels effect modulator, etc. SPM, a third-order nonlinear effect, balanced by group velocity dispersion (GVD), may generate an optical soliton that can propagate in the medium maintaining its shape. The complex nature of permittivity also gives rise to TPA, so free-carrier absorption (FCA) due to third-order nonlinearity adds to overall absorption. FWM finds applications in parametric amplification, spectral translation, optical frequency comb generation, etc. Supercontinuum generation due to nonlinearity finds applications in accurate-frequency metrology, high-resolution optical coherence tomography, etc. SRS is an important third-order nonlinear process with involvement of phonons (molecular vibration), which finds applications in light amplification, etc. Stimulated Brillouin scattering (SBS) is another third-order nonlinear process that involves acoustic phonons. It can find applications in the conversion of optical waves to acoustic waves, etc.

Chapter 14 provides a comprehensive overview of the fabrication techniques and material systems used in silicon-based planar lightwave circuits (PLCs). Silicon photonics leverages well-established semiconductor manufacturing processes, enabling cost-effective, high-reliability, and scalable production of optical devices. The chapter discusses key fabrication methods, including deposition (chemical vapor deposition [CVD], physical vapor deposition [PVD], molecular beam epitaxy [MBE]), planarization (chemical-mechanical planarization [CMP]), doping (diffusion, ion implantation), oxidation, lithography (photolithography, e-beam, nanoimprint), etching (wet, dry, reactive ion), and bonding (wafer, adhesive). Two primary waveguide platforms are emphasized: silica-on-silicon and SOI. Silica-on-silicon waveguides, fabricated via flame hydrolysis deposition (FHD) or plasma-enhanced chemical vapor deposition (PECVD), offer low propagation loss (<0.1 dB/cm) and excellent fiber coupling but require large bend radii due to low refractive index contrast. SOI waveguides, formed using Separation by Implanted Oxygen (SIMOX), Bond and Etch-back Silicon-On-Insulator (BESOI), or smart-cut processes, provide strong optical confinement (high index contrast) and compact device footprints but suffer from higher coupling losses due to mode mismatch with optical fibers. Other material systems, such as SiGe–Si, SiON, and amorphous silicon, are also discussed, each offering unique advantages in refractive index tuning, strain engineering, and thermal management. Challenges such as lattice mismatch, defect density, and thermal budget constraints are addressed, along with solutions like graded buffer layers and optimized growth conditions. Finally, the chapter explores grating formation techniques (interferometric, phase mask) and compares different waveguide materials in terms of propagation loss, coupling efficiency, and polarization sensitivity. The chapter concludes by highlighting the trade-offs between various material platforms and their suitability for different integrated photonic applications.

References

1. Baets, R., Van Campenhout, J., Kunert, B., and Roelkens, G. (2023). 4 ways to put lasers on silicon: you can make many things with silicon photonics, but a laser is not one of them. *IEEE Spectr.* 60 (5): 32–37.
2. Ossieur, P., Zilkie, A., Menezo, S. et al. (2023). Editorial: introduction to JSTQE special issue on photonic electronic co-integration and advanced transfer printing. *IEEE J. Sel. Top. Quantum Electron.* 29 (3): 3–6.
3. Shekhar, S., Bogaerts, W., Chrostowski, L. et al. (2024). Roadmapping the next generation of silicon photonics. *Nat. Commun.* 15: 751. https://doi.org/10.1038/s41467-024-44750-0.
4. Prete, D., Amanti, F., Andrini, G. et al. (2024). Hybrid integrated silicon photonics based on nanomaterials. *Photonics* 11 (5): 418. https://doi.org/10.3390/photonics11050418.

5. Margalit, N., Xiang, C., Bowers, S.M. et al. (2021). Perspective on the future of silicon photonics and electronics. *Appl. Phys. Lett.* 118: 220501. https://doi.org/10.1063/5.0050117.
6. Reed, G.T. and Knights, A.P. (2004). *Silicon Photonics: An Introduction*. New York: Wiley Interscience.
7. Pavesi, L. and Lockwood, D.J. (ed.) (2004). *Silicon Photonics*, Topics in Applied Physics, vol. 94. Berlin: Springer-Verlag.
8. Soref, R.A. (2008). The advances and challenges of silicon photonics. *Adv. Opt. Technol.* 2008: 472305.
9. Fitzgerald, E.A. and Kimerling, L.C. (1998). Silicon-based microphotonics and integrated optoelectronics. *MRS Bull.* 23: 39–47.
10. Lockwod, D.J. (ed.) (1997). *Light Emission in Silicon: From Physics to Devices*, Semiconductors and Semimetals, vol. 49. San Diego, CA: Academic Press.
11. Pavesi, L., Gaponenko, S., and Dal Negro, L. (ed.) (2003). *Towards the First Silicon Laser*, NATO Science Series. Dordrecht: Kluwer Academic.
12. Bisi, O., Campisano, U., Pavesi, L., and Priolo, F. (ed.) (1999). *Silicon Based Microphotonics: From Basics to Applications*. Amsterdam: IOS Press.
13. Ossicini, S., Pavesi, L., and Priolo, F. (2003). *Light Emitting Silicon for Microphotonics*. Berlin: Springer-Verlag.
14. Pavesi, L. (2003). Will silicon be the photonic material of the third millenium? *J. Phys.: Condens. Matter* 15: R1169–R1196.
15. Salib, M., Liao, L., Jones, R., and Morse, M. (2004). Silicon photonics. *Intel Technol. J.* 8, (2): 143–160.
16. See papers in Special issue on Silicon Photonics, (2006). *IEEE J. Sel. Top. Quantum Electron.* 12 (6: Part 2): 1327–1755.
17. Jalali, B. (2008). Can silicon change photonics? *Phys. Status Solidi A* 205: 213–224.
18. Jalali, B., Paniccia, M., and Reed, G.T. (2006). Silicon photonics. *IEEE Microwave Mag.* 58–68.
19. Chrostowski, L. and Hochberg, M. (2015). *Silicon Photonics Design: From Devices to Systems*. Cambridge: Cambridge University Press.
20. Inniss, D. and Rubenstein, R. (2016). *Silicon Photonics*. Morgan Kaufmann.
21. Westerveld, W.J. and Urbach, H.P. (2017). *Silicon Photonics: Electromagnetic Theory*. Bristol: IOP Publishing.
22. For history of development of devices, the reader is referred to (1976). *IEEE Trans. Electron Dev.* ED 23 (7): 595–785.
23. Plummer, J.D., Deal, M.D., and Griffin, P.B. (2001). *Silicon VLSI Technology: Fundamentals, Practice and Modeling*. New York: Prentice Hall.
24. Taur, Y. and Ning, T.H. (1999). *Fundamentals of Modern VLSI Devices*. Cambridge: Cambridge University Press.
25. Moore, S.K. and Schneider, D. (2022). The state of the transistor in 3 charts. *IEEE Spectrum*. https://spectrum.ieee.org/magazine/2022/december/ (accessed 1 September 2025).
26. Cass, S. (2022). The ultimate transistor timeline. *IEEE Spectrum*. https://spectrum.ieee.org/magazine/2022/december/ (accessed 1 September 2025).
27. Moore, S.K. (2022). The transistor of 2047: expert predictions. *IEEE Spectrum*. https://spectrum.ieee.org/magazine/2022/december/ (accessed 1 September 2025).
28. Liu, M. and Wong, H.S.P. (2024). The path to a 1-trillion-transistor GPU: AI's Boom Demands New Chip Technology. *IEEE Spectr.* 61 (7): 22–27.
29. Kumar, S. and Deen, M.J. (2014). *Fiber Optic Communications: Fundamentals and Applications*. Chichester: Wiley.
30. Cassese, T. and Vornicu, I. (2025). Silicon photonics. *Photonics Spectra*, 30–34. https://www.photonics.com/Articles/Silicon_Photonics_Brings_a_Collaborative/p5/a70612 (accessed 1 September 2025).