Review Article

Recent Progress in Silicon Photonics: A Review

Zhou Fang and Ce Zhou Zhao

Department of Electrical and Electronic Engineering, Xi’an Jiaotong-Liverpool University, Suzhou 215123, China

Correspondence should be addressed to Ce Zhou Zhao, cezhou.zhao@xjtlu.edu.cn

Received 10 December 2011; Accepted 28 December 2011

Academic Editors: A. E. Miroshnichenko and S. F. Yu

Copyright © 2012 Z. Fang and C. Z. Zhao. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the increasing bandwidth requirement in computing and signal processing, the inherent limitations in metallic interconnection are seriously threatening the future of traditional IC industry. Silicon photonics can provide a low-cost approach to overcome the bottleneck of the high data rate transmission by replacing the original electronic integrated circuits with photonic integrated circuits. Although the commercial promise has not been realized, this perspective gives huge impetus to the development of silicon photonics these years. This paper provides an overview of the progress and the state of the art of each component in silicon photonics, including waveguides, filters, modulators, detectors, and lasers, mainly in the last five years.

1. Introduction

Silicon (Si) has been the mainstay of the electronics industry for more than 40 years and once revolutionized the way the world operates. By employing more precise lithography technology and multicore structures, the development of processors can still barely follow the Moore’s law. However, with increasing requirement of bandwidth, the parasitic effects in current metallic interconnection have gradually become a main obstacle for further improvements, since electric signal attenuation and power dissipation rise dramatically with higher data rate. To overcome the bottleneck of the high data rate transmission, one possible solution could be employing optical interconnect, in which the information signals are carried by photons instead. Compared with electrons, photons have zero rest mass and zero charge, which means that they can travel at velocity of light without the interference with electromagnetic field, so optical systems can theoretically achieve signal transmission with much higher data rate and better stability than electrical system. Therefore, it is highly desirable to replace traditional electrical circuits with optical circuits, and, under these motivations, a popular subject, called optoelectronic integrated circuit (OEIC), has been built since late 1980s.

Si has an apparent window from the wavelength of 1100 nm to 7000 nm approximately, which is far from being limited to the near-infrared (IR) communication band of 1300–1550 nm. Some excellent optical properties, like large optical damage threshold and thermal conductivity, also appear in Si. Furthermore, today’s mature complementary metal-oxide semiconductor (CMOS) techniques could also allow low-cost, large-scale manufacturing for Si photonic devices. All of these reasons select Si as a remarkable candidate for photonics. The development of Si photonics can be tracked back to the pioneering work done by Soref et al. in mid-1980s [1, 2], from which technology boom in this field mainly focused on Si-based waveguides, switches, and modulators ([3–5], e.g.). It was not until mid-2000s when significant momentum occurred in this field. Until now, fast Si modulators and high responsivity in epitaxial Germanium (Ge) detectors have both been achieved. Furthermore, great breakthroughs, such as Raman Si lasers, Mid-IR Si sources, epitaxial GaAs diodes grown on Si at room temperature, and even Ge-on-Si lasers, have been made as light sources, the main obstacle in traditional Si photonics. These technical advances surely boost the promise that Si can be considered as an optical material for OEIC system.

Instead of reviewing the history of this field, this paper aims to evaluate the potential impacts of silicon photonics by describing the recent breakthroughs, novel devices, and the existing challenges, mainly in the past five years. Some
current commercialized devices and long-term projects of several big corporations about this field will also be reviewed.

2. Passive Devices

2.1. Waveguides. One basic passive device in photonics is the waveguides, through which optical signals can be transferred from one point to another. In general, both size and propagation loss should be considered when evaluating the performance of a specific kind of waveguide. Propagation loss can be divided into two parts, which are intrinsic loss (such as carrier absorption) and extrinsic loss (such as sidewall scattering and radiation into the substrate). The former one is the main loss source for doping-based waveguides, while the later one becomes significant when the sizes of waveguides are relatively small (such as SOI waveguides), due to the field density at the silicon surface and the roughness at the interface.

There are many kinds of Si-related materials which can be applied to make the waveguides. One of the most commonly applied waveguide materials is silica. By doping III/V ions into the silica, the refractive index of the material will change (typically around 0.1% to 0.75%), resulting in a contrast in refractive index between the core and the cladding [6]. One advantage for silica waveguide is that the refractive index contrast can be adjusted flexibly by changing the doping density. However, the low refractive index means a weak confined structure for silica waveguides, which results in both thick cladding layer (typically ~50 μm) and wide spacing between waveguides [6, 7]. Therefore, it is not truly compatible with electronic IC technology. Fortunately, a fortuitous discovery of silicon-on-insulator (SOI) has offered an excellent platform for Si/SiO₂ waveguides. Due to the large refractive index contrast between Si (n = 3.45) and SiO₂ (n = 1.45), a strong confinement is formed within the waveguides, which makes it possible to scale down the size of the waveguides mode to approximately 0.1 μm². Such lateral and vertical dimensions agree with the requirement for economic compatibility with CMOS technology today (Figure 1). However, as devices are downsized, deviations from ideal profiles in fabrication will become prominent. In particular, sidewall roughness, introduced by imperfect etching, will result in scattering at the interface between waveguide core and cladding, which becomes a major source of propagation loss [8]. This scattering loss, typically around 0.2–3.0 dB/cm, is related to both sidewall roughness and the size of waveguide, which has been deeply analysed by Yap et al. in [9].

The surface roughness, however, can be reduced effectively by thermal oxidation [10, 11]. Higher temperature and longer oxidation time would improve the interface roughness more distinctly, and 1100 °C is enough to avoid the profile distortion within 1 hour [10, 11]. However, the defects of this method are that this process consumes more silicon, and it also induces residual stress in silicon [12]. The scattering loss can also be reduced by either surface encapsulation with SiNx [13], or high-temperature annealing with hydrogen ambient [13, 14], in which the surface mobility of silicon atoms is enhanced so that those migrating atoms can smooth out the surface roughness [12]. However, these methods involve harsh processed conditions, like high temperature, and, more importantly, they are not selective, which means that some other components that are not supposed to be processed at the same time on the same chip will be affected simultaneously, especially for chips with high density integration. One novel solution is to use self-perfection by liquefaction (SPEL). A XeCl excimer laser with 308 nm wavelength and 20 ns pulse duration is applied to melt the surface layer of the waveguide selectively, which results in smooth sidewall upon resolidification (Figure 2) [15]. Based on this method, reduction of roughness from 13 to 3 nm and roughness-induced propagation loss from 53 to 3 dB/cm have been reported. Another advantage for SPEL is that the process is ultrafast (~100 ns).

The problem of coupling light between the single-mode fibres (SMF) and the silicon waveguides is another challenge for passive devices. Since the waveguide dimension of silicon photonics is very small, a considerable mode size mismatch appears between nanoscale silicon waveguides and SMFs.
Figure 2: (a) 4 μm wide Si waveguide on SiO₂ with rough sidewalls was smoothed. (b) After exposure to 20 laser pulses of 900 mJ/cm² at a repetition rate of 1 Hz. (c) Roughness profile along the length direction of a 4 μm wide Si waveguide on SiO₂ before and after. (d) Self-perfection (SPEL). Figures are taken from [15].

Figure 3: (a) Grating coupler for coupling between photonic wire waveguide and fibre. (b) Oblique view of a holographic lens connected to silicon waveguides. The optical fibre illuminates the holographic lens from a position normal to the surface. The core of the fibre is shown by the dashed lines. Figure (a) is taken from [17, 18]. Figure (b) is taken from [22].

Thus, the coupling loss will be very large if directly coupling the waveguides with SMFs. One common solution is using surface grating (Figure 3(a)) to butt-couple light from a SMF, perpendicular to the surface, into planar waveguides [16–21]. Coupling efficiency around 65.6% at 1550 nm has been demonstrated in experiments [20] and is expected to be improved to 90% by optimizing grating design and layer stack [17, 18]. Since light can be injected or extracted at arbitrary location on the chip, conventional wafer scaling test can be applied without antireflection coating or facet polishing [20]. Furthermore, alignment tolerance is also relatively large for grating type coupler (typically >1 dB additional loss for alignment error of ±1 μm), hence reducing the packaging cost [17, 18, 20]. Although the bandwidth of surface grating couplers is relatively narrow, broad 3 dB bandwidth of 45 μm has been demonstrated, which is adequate for a wide range of applications [19]. In addition, vertical coupling can also be achieved using lens with a similar structure (Figure 3(b)) [22].
Another promising alternative is coupling light between fibres and waveguides by combining a Si inverse taper and a medium index contrast waveguide, made by polymer or SiN for example, (Figure 4) [23–28]. This technique is based on the gradual expansion from the core-guided mode to a much larger cladding-guided mode. Coupling loss as low as 0.36 dB/cm for TE modes and 0.66 dB/cm for TM modes has been demonstrated [26].

Most recently, a novel type of converter based on suspended structure is reported by Fang et al. (Figure 5) [28]. It consists of a suspended SiO\textsubscript{2} waveguide with overlapped Si nanotapers located in the centre of the waveguide, with SiO\textsubscript{2} beams laterally connected to provide structural support to the waveguide. In order to confine the optical mode within the SiO\textsubscript{2} waveguide, it is necessary to etch and isolate the SiO\textsubscript{2} waveguide from the Si substrate, so as to form a suspended SiO\textsubscript{2} waveguide. The optical signal will first be launched into the suspended SiO\textsubscript{2} waveguide and then be coupled into the silicon nanotapers. The advantages of this type of converters are that they are almost wavelength independent and they have a large alignment tolerance (1.7 μm for 1 dB excessive loss), hence overcoming the shortcoming of couplers in both grating type and taper type, respectively, though some improvements on its coupling loss are expected.

2.2. Filters. With the applications in wavelength multiplexing (WDM) and demultiplexing, optical filters have been a crucial participant in modern optical network. Early work can be tracked back to late 1990s and early 2000s using waveguide grating [29, 30], and, more recently, microdisk and microring have gained increasing interest for the same application due to their potential for high-density integration [31, 32]. For a common microring-based add-drop filter, light travels through a waveguide in close proximity to a ring, so that the evanescent fields of the optical modes overlap, and the optical energy can transfer to the ring [32]. The gap distance between the waveguide and ring will determine the strength of the coupling. In order to achieve resonance, the wavelength of signal must satisfy

$$\lambda = \frac{n_{\text{eff}} L}{m},$$

(1)

where $L$ is the ring diameter, $n_{\text{eff}}$ is the effective index of the mode propagating in the ring, and $m$ is the resonance order. The free spectra range (FSR), which is the wavelength separation between resonances, is given by

$$\text{FSR} = \frac{\lambda^2}{n_{g} L},$$

(2)

where $n_{g}$ is the group index of the propagating modes. Therefore, the resonance frequency will be blue-shifted if reducing the diameter of the ring $L$ or the refractive index. The resonance depth, which is also called the extinction ratio (ER), is another crucial parameter for ring filter, which is defined as

$$\text{ER} = 10 \log \frac{P_{\text{on}}}{P_{\text{off}}},$$

(3)

where $P_{\text{on}}$ and $P_{\text{off}}$ are the optical power on-resonance and off-resonance.

For general applications, filters with wide FSR and large ER are required, which are also the main challenges today. In addition, flat passband, shape roll-off and low drop loss are also important for WDM purpose. Recently, an impressively high ER at >50 dB with less than 2 dB drop loss has been achieved by both Dong et al. in Kotura [33, 34] (Figure 6(a)) and Smith’s group in MIT (Figures 6(b) and 6(c)) based on topology, while FSR over 30 nm is also demonstrated for add-drop filters [35]. However, since the ring and the busline waveguides are formed in the same layer for the planar structure of these filters, a very precise fabrication technique is required, which is a major challenge for size controlling and reproducibility [31]. In comparison, a group of filters with vertically coupled microrings have been demonstrated to simplify the fabrication [31, 36]. An example achieved by Kokubun using cross grid configuration is discussed here (Figure 7) [36]. Rather than etching of fine gap, vertical separation is achieved by well-controlled deposition, so the coupling strength can be controlled more precisely. Moreover, since the overlap of the field profile is much less sensitive to the offset than in lateral coupling, this type of filters also enjoys a relatively large fabrication tolerance of the lateral misalignment. Another popular group of filters is the lattice type (Figure 8). Using snakelike delay line, the channel dropping bandwidth can be controlled by the number of periods in this structure, and, more importantly, it can achieve a much wider FSR (2.75 times) than the ring filters with the same radius [25, 37].

All of those filters introduced above are relatively static, and any change in bandwidth, ER, FSR, and so on can be accomplished only by redesigning and fabricating it again. Therefore, for modern broadband applications, highly dynamic performance for one specific filter may be required [38], and several novel filters with tenability are briefly introduced here, though some of them become active devices
Figure 5: Schematic diagram of suspended converter inset: overlapped Si tapers. In order to confine the optical mode within the SiO$_2$ waveguide, it is necessary to etch and isolate the SiO$_2$ waveguide from the Si substrate. Figure is taken from [28].

Figure 6: (a) Optical images of fully fabricated 2nd-order (top) and 5th-order (bottom) ring filters by Kotura. (b) SEM of fabricated three-stage filter by MIT group. (c) Measured response spectra of MIT’s three-stage filter, showing high in-band extinction. Figure (a) is taken from [33]. Figures (b) and (c) are taken from [34].

instead. One common method is using microelectromechanical system (Figure 9). The waveguide is completely uncoupled at zero bias. When a voltage is applied on the electrodes, the waveguides are bent towards the microdisk, increasing the coupling strength between them, and the coupling is stronger with higher bias. Based on this mechanism, filters with adjustable attenuation have been achieved [39], and similarly a filter with tunable bandwidth from 2.8 GHz to 78.4 GHz has also been reported later [40]. Alternatively, reconfigurable add-drop filter can also be formed using arrays of micromirrors (Figure 10) [41]. About 750,000 mirrors can addressed in digital micromirror device (DMD)
chip, and each of them can be rotated by ±9.2° depending on the control signal applied to the memory cell. Therefore, dynamic configuration of the device parameters is obtained by controlling the tilt of the selected group of micromirrors, without any optical or mechanical adjustments, and extinction ratio at 35 dB with microseconds reconfiguration time is achieved.

3. Modulators

Optical modulators, the workhorse of interconnection technology, have experienced a dramatic development in recent years. In general, a modulator is a device which can vary the fundamental characteristics of a light beam with respect to an information signal. Several factors should be considered when evaluating the performance of a modulator. The first one is modulation depth, also called extinction ratio, referring to the ratio between the optical intensity at maximum transmission and minimum transmission. A large modulation depth can achieve good bit error rate and high maximum transmission and minimum transmission. The second one is modulation speed, which is the ability to carry data at a certain rate. The third one is bandwidth, and it usually represents the frequency at which the modulation is reduced to 50% of the maximum value. In addition, insertion loss, area efficiency (size or footprint), and power consumption are also important in modulator design.

Ideally, optical modulators need to have high modulation speed and large bandwidth, together with low losses, small size, and ultralow power consumption. However, these requirements often contradict to each other, and hence an optimal tradeoff is necessary to be achieved by the engineers for different purposes. There are two main branches for today’s research in this area. Some groups continue to focus on modulators in IV-group materials, while others are seeking for another alternative way of hybrid combination of different materials on the same silicon plate. Both methods will be discussed in detail in this part later, based on those capability metrics listed above.

3.1. IV-Group-Based Modulators

Based on the potential of integration with electronics in a cost-effective manner, silicon has been one main candidate for modulator fabrication. However, some primary electrical field effects, including the Pockels effect, the Kerr effect, and the Franz-Keldysh effect, are weak in pure silicon at telecommunication wavelengths of 1.3 μm and 1.55 μm, due to the centrosymmetric crystal structure of Si [2, 42, 43]. This currently leaves thermal-optic effect and plasma dispersion the only viable methods for modulation in silicon. Owing to the large thermal-optic coefficient of Si (1.86 × 10^-4/K) [43], thermal modulation, which is based on the temperature dependent of the material refractive, is realizable in Si. Recently, tunable thermal-optic switches and modulators have been demonstrated [44], and the power consumption of this type of modulators is reported being similar to that of plasma dispersion-based modulators [43]. The coefficient can be further increased by using nanostructure Si in silicon-rich silicon oxide (SRSO) thin film [45]. However, although Della Corte et al. has successfully decreased the response time of a thermal-optic modulator to around several microseconds by optimizing the driving signals, this speed is still too slow for the high frequencies required by modern telecommunication applications [44].

In comparison, plasma dispersion is the most common method of achieving modulation in Si, in which the refractive index, both the real part n (called electrorefraction) and the imaginary part α (called electroabsorption), changes with the concentration of free carriers in Si. Soref and Bennett evaluated the changes in the refractive index in experimentally produced absorption curves, focusing on the communication wavelengths, and they produced the following empirical expression to quantify the changes due to injection or depletion [2, 43]. At λ0 = 1300 nm

\[
\Delta n = \Delta n_e + \Delta n_h
\]

\[
= -\left[6.2 \times 10^{-22} \Delta N_e + 6.0 \times 10^{-18} (\Delta N_h)^{0.8}\right],
\]

\[
\Delta \alpha = \Delta \alpha_e + \Delta \alpha_h
\]

\[
= 6.0 \times 10^{-18} \Delta N_e + 4.0 \times 10^{-18} \Delta N_h,
\]

where \(\Delta n_e\) and \(\Delta n_h\) are changes in refractive index resulting from changes in the free-electron and free-hole carrier concentrations, respectively, and \(\Delta \alpha_e\) and \(\Delta \alpha_h\) are the changes in absorption resulting from changes in the free-electron...
Figure 8: Schematic diagram of Si-wire-based lattice filter. Figure is taken from [37].

Figure 9: Schematic diagrams of the microdisk resonator with deformable waveguides. (a) At zero bias, the microdisk is completely uncoupled since the suspended waveguides are far away from the microdisk. (b) With voltage applied on the electrodes, the waveguides are bent toward the microdisk, increasing the coupling between the waveguides and the disk. Figures are taken from [39].

and free-hole carrier concentrations, respectively. Similarly, at $\lambda_0 = 1550$ nm

$$\Delta n = \Delta n_e + \Delta n_h$$

$$\Delta n = -\left[8.8 \times 10^{-22} \Delta N_e + 8.5 \times 10^{-18}(\Delta N_h)^{0.8}\right],$$

$$\Delta n = 8.5 \times 10^{-18} \Delta N_e + 6.0 \times 10^{-18} \Delta N_h.$$
Compared with MZI modulators, ring modulators have smaller footprints and higher power efficiency, since MZI modulators require a long interaction length, resulting in greater insertion loss and power consumption. However, the drawback of ring modulators is their narrow bandwidth, which limits their application and, more importantly, causes a high sensitivity to both fabrication tolerance and temperature variation [42, 48]. Therefore, tradeoffs are involved when choosing the types of modulators.

Carrier injection is the most sophisticated technology to obtain carrier concentration variation, which can be tracked to mid-1980s when Soref and Bennett proposed [49] the first plasma dispersion modulator. This modulator has a p⁺-n-n⁺ structure, and the interaction length is less than 1 mm for 180° phase shift [49]. Later, intensive researches of injection-type modulators focused on p-i-n diode structure instead to avoid excessive optical loss. One advantage of current injection-based modulators is the relatively uniform index modulation cross the entail waveguide, so that the optical mode and modulation region are completely overlapped [50]. However, the major problem of this type of modulators is its slow operation speed originating from the long minority carrier life time in silicon. Performance can be enhanced by reducing device dimensions and optimizing the structure and doping density, which has been studied in details by Png et al. [51]. Although the speed of carrier injection based modulators was thought to be limited at around hundreds Mb/s, up to 18 Gb/s has been recently achieved in ring resonance structure [52], and 10 Gb/s using MZI structure [53]. Both of them employ preemphasis of the electrical signal to assure rapid injection. Another problem is the large power consumption due to the requirement for relatively large injection current. It then inevitably produces a certain temperature rise. Unfortunately, the refractive index increases with the rise of temperature because of the thermal-optical effect in silicon, while carrier injection will decrease the refractive index. Therefore, temperature rise will weaken the effect introduced by dispersion plasma, and the modulators can even be out of work if temperature rise is too high after operating for a certain time [6, 54].

Another type of modulators, based on carrier accumulation, was developed with one pioneering breakthrough in 2004 when the first monolithic silicon modulator with speeds >1 GHz was experimentally demonstrated by Intel’s group [46]. The performance of this device was then optimized one year later by Liao et al. to a data rate of 10 Gb/s and an extinction ratio of 3.8 dB [55]. Instead of using carrier injection, this modulator achieves plasma dispersion by employing carrier accumulation. It has a metal-oxide semiconductor (MOS) capacitor structure, shown in Figure 12. The layer beyond the gate dielectric is p-type Si created by epitaxial lateral overgrowth (ELO), while the crystalline silicon below the gate is doped to n-type. Applying a positive voltage to p-type Si, charge carries will accumulate at the oxide interface, and hence change the refractive index in the charge layers. When placed in one or both arms of an MZI, this phase shifter can vary the relative phase difference between the light in each arm, and the desired optical intensity modulator can be achieved based on the interference of the outputs at two arms. Since charge transport in this MOS capacitor is governed by majority carriers, device speed is no longer limited by relatively long minority carrier lifetime in silicon, but instead depends on the device resistance and capacitance [46, 55, 56]. Furthermore, its MOS capacitor structure makes it automatically compatible with conventional CMOS processing [46, 55]. Recently, extinction ratio as high as 9 dB has been achieved by Lightwave in carrier accumulation modulators, with a data rate of 10 Gb/s [57].

Carried depletion is another technique for manipulating free-carrier density, which is currently the most popular one among those three dispersion methods mentioned before. Generally, this type of modulators has a p-n or p-i-n structure. Free carrier density change is obtained by applying a reverse bias voltage, so the speed limitation from minority carrier lifetime is avoided, and it can be driven with relatively lower power since the diode is reverse biased. However, the drawback is its poor modulation depth originating from a small overlap between the optical field and active electrical
semiconductors, and in comparison, QCSE has stronger electric-field-induced changes in optical absorption in comparison with injection-based modulators, most of those examples still belong to group of III-V semiconductors quantum wells, such as GaAs, AlAs with AlGaAs barriers [63, 64]. However, quantum wells in group IV materials also attract equally interests these years due to their compatibility with CMOS technology. Unfortunately, absorption effect is very weak in Si or Ge, due to their indirect band structure. However, unlike Si, Ge has a useful direct bandgap which is only slightly higher than the indirect bandgap in energy, which makes it exhibit a strong and abrupt rising edge around 1.55 μm wavelength. One example developed by Harris’s group has been shown in Figure 14. Ge quantum wells with SiGe barriers between them are grown on the top of a lattice relaxed SiGe buffer layer on a [1 0 0]-oriented Si substrate. Despite the lattice mismatch between Ge and Si (4%), the device is well designed so that the net strain is zero [63]. By applying a voltage, the absorption spectra of Si/SiGe is blue-shifted, while the 3 dB excitation width stays almost unchanged. Furthermore, the spectra can also be adjusted by changing the temperature, so it is possible to make it work in communication wavelength range. The performance of this modulator is reported at a modulation speed at 10 GHz in a size of 30 μm at the early stage [65]. The drawback of QCSE method is that it is necessary to introduce second material, and complex fabrication process is also required. Anyway, high-speed and efficient QCSE modulators are still expected in future.

The other idea, reported by Fathpour and Jalali, is energy harvesting in silicon optical modulators [66]. It has a similar structure with conventional modulators, but the work principle is totally different. The input power and the wire cross-section are well designed so that two-photon absorption (TPA) can be induced based on the two-photon photovoltaic effect in silicon. Similar to a solar energy, this device is biased to the fourth quadrant of I-V curve, so pure electrical energy will be obtained instead. Therefore, this kind of modulators can harvest the energy consumed by lasers, which will otherwise be dissipated into heat. This is crucial for VLSI industry, since higher power dissipation and hence higher temperature will threaten the integrated devices and even make them out of work. However, this idea requires a light source with relatively strong emitting intensity so that TPA can be realized in modulation area, so further improvements in design are still required. Nevertheless, modulation bandwidth >1 GHz is still expected to be achieved in this type of modulators with response time range from 20 ps to 20 ns [66].

Finally, two novel ideas for development of silicon modulators are demonstrated here. The first one is an alternative mechanism of modulation by employing optical absorption. In comparison with plasma dispersion-based modulators, modulators of absorption type exhibit higher operating speed and lower power dissipation, since they also get rid of the limitation from the minority carrier lifetime. Both Franz-Keldysh effect and the quantum confined Stark effect (QCSE) are electric-field-induced changes in optical absorption in semiconductors, and, in comparison, QCSE has stronger changes in absorption coefficient [42]. QCSE has been observed in some quantum wells structures with barriers, in which carrier confinement allows exciton enhancement or optical absorption. Although QCSE has been extensively applied for high speed optical modulators, most of those examples still belong to group of III-V semiconductors quantum wells, such as GaAs, AlAs with AlGaAs barriers [63, 64]. However, quantum wells in group IV materials also attract equally interests these years due to their compatibility with CMOS technology. Unfortunately, absorption effect is very weak in Si or Ge, due to their indirect band structure. However, unlike Si, Ge has a useful direct bandgap which is only slightly higher than the indirect bandgap in energy, which makes it exhibit a strong and abrupt rising edge around 1.55 μm wavelength. One example developed by Harris’s group has been shown in Figure 14. Ge quantum wells with SiGe barriers between them are grown on the top of a lattice relaxed SiGe buffer layer on a [1 0 0]-oriented Si substrate. Despite the lattice mismatch between Ge and Si (4%), the device is well designed so that the net strain is zero [63]. By applying a voltage, the absorption spectra of Si/SiGe is blue-shifted, while the 3 dB excitation width stays almost unchanged. Furthermore, the spectra can also be adjusted by changing the temperature, so it is possible to make it work in communication wavelength range. The performance of this modulator is reported at a modulation speed at 10 GHz in a size of 30 μm at the early stage [65]. The drawback of QCSE method is that it is necessary to introduce second material, and complex fabrication process is also required. Anyway, high-speed and efficient QCSE modulators are still expected in future.

The other idea, reported by Fathpour and Jalali, is energy harvesting in silicon optical modulators [66]. It has a similar structure with conventional modulators, but the work principle is totally different. The input power and the wire cross-section are well designed so that two-photon absorption (TPA) can be induced based on the two-photon photovoltaic effect in silicon. Similar to a solar energy, this device is biased to the fourth quadrant of I-V curve, so pure electrical energy will be obtained instead. Therefore, this kind of modulators can harvest the energy consumed by lasers, which will otherwise be dissipated into heat. This is crucial for VLSI industry, since higher power dissipation and hence higher temperature will threaten the integrated devices and even make them out of work. However, this idea requires a light source with relatively strong emitting intensity so that TPA can be realized in modulation area, so further improvements in design are still required. Nevertheless, modulation bandwidth >1 GHz is still expected to be achieved in this type of modulators with response time range from 20 ps to 20 ns [66].

3.2 Hybrid Modulators. An alternative approach for optical modulation is to employ III-V or organic materials and to integrate them together with silicon devices. Unlike silicon, some of these materials, such as LiNbO3, exhibit strong second-order nonlinear electrooptical effects, such as Pockels effect, which makes them excellent candidates for
modulation application. Usually, electrooptical modulators are usually preferred since they do not suffer from carrier-related speed limitation. Since several years ago, LiNbO3 modulators have been developed for commercial purposes, with good performance of 20 Gb/s bandwidth, 20 dB modulation depth, and 4 V switch voltage [27]. Until now, ultrafast LiNbO3 modulators with modulation speed at 110 GHz have been reported by using titanium (Ti) diffusing [67], while LiNbO3 modulators with ultralow driving voltage at 2 V and bandwidth of 15 GHz are also realized [68]. Furthermore, new modulation formats, such as QPSK and DQPSK, are currently under investigation as well, and 80 Gb/s LiNbO3 modulators based on these formats have been reported recently [68]. Therefore, these performances in LiNbO3 modulators still challenge those of silicon. One disadvantage for LiNbO3 modulators is that LiNbO3 is too expensive. Therefore, it is still too early to decide which one will dominate in the field of optical modulators.

Another novel idea for new generation of modulators is using silicon organic hybrid (SOH) integration. Although silicon can offer a strong optical confinement due to its large refractive index, third-order nonlinearities are too weak to be applied for direct modulation. In comparison, polymers have very strong and ultrafast nonlinear responses and broad spectral tunability, but their refractive index is relatively small ($n = 1.4–2.5$ typically) [69–71]. Therefore, by using SOH, excellent modulators can be made by taking advantages from both kinds of materials. Generally, passive components are fabricated in silicon in this approach, while the nonlinear optical functionality is taken over by the cladding made by organic materials and polymers. An example of a group of high speed silicon organic phase

---

**Figure 14:** (a) A p-i-n structure on silicon with Ge/Si$_{1-x}$Ge$_x$ quantum wells on relaxed Si$_{1-z}$Ge$_z$ buffer. (b) Effective absorption coefficient spectra of Ge/SiGe (10 nm Ge well and 16 nm Ge/Si$_{0.15}$Ge$_{0.85}$ barrier in this example) quantum wells on a relaxed Si$_{0.1}$Ge$_{0.9}$ buffer at room temperature. (c) Effective absorption coefficient spectra of strained Ge/SiGe (12.5 nm Ge well and 5 nm Si$_{0.175}$Ge$_{0.825}$ barrier in this example) quantum wells on a relaxed Si$_{0.05}$Ge$_{0.95}$ buffer under 0.5 V reverse bias at different temperatures. Figures are taken from [63].
Figure 15: Three implementations of silicon organic hybrid electrooptic modulators comprising silicon ribs on top of the buried oxide (SiO₂) and an electrooptic cover. The electric field magnitudes are depicted on the right-hand side. (a) Traveling-wave strip waveguide structure; (b) traveling wave slot waveguide structure; (c) photonic crystal slot waveguide structure. Figures are taken from [71].

shifters is shown in Figure 15 [70, 71], together with the electrical field magnitude distribution in waveguide. The corresponding MZI modulators can be made by putting these shifters at one or both arms. In Figure 15(a), the shifter consists of a silicon waveguide surrounded by a poled electrooptic organic material (OE) and two doped silicon walls connected with aluminium contacts, through which an electrical field can be applied in parallel with the waveguide. The space between waveguide and contacts is large enough to avoid optical loss. Figure 15(b) introduces a traveling-wave strip waveguide structure. In Figure 15(c), a photonic crystal (PC) line defect waveguide is applied on a slot etched into SOI device layer, and the PC has a slow light structure. Slow light is a phenomenon in which the light is scattered and interferes with incoming light, forming an interference pattern which moves forward slowly through the waveguide. In this way, the effective velocity of light is “slowed down.” By employing slow light approach, the interaction time with electrical field will be significantly increased, hence decreasing the device length or driving voltage [42, 70, 71]. Recently, electrooptic modulators with 100 Gb/s speed have reported based on a slow light SOI structure with organic cladding [70, 71]. Furthermore, this type of modulators typically requires a low driving voltage of about 1 V [70, 72]. Therefore, hybrid modulation based on organics will also be very competitive in future.

4. Detectors

Photodetectors are one basic component in electrooptical integrated circuits, which convert the received optical signals into electrical signals for further processing. Usually, photodetectors consist of a photodiode of p-n or p-i-n structures. Unlike solar cells, photodiode usually works under a reverse bias voltage in a close loop. When photons are absorbed by the p-n junction, some hole-electron pairs are created, and the produced electrons in p area and holes in an area will drift
through the p-n junction under the built-in electric field in depletion region. By evaluating the produced photocurrent in this process, the corresponding optical signals can be detected. If the build-in potential is relatively high, carrier drift will be dominant compared with the diffusion, which will reduce the limitation on response time from carrier diffusion. A reverse bias voltage is usually applied across the p-n junction to enhance the electric field, hence increasing responsivity and speed of the detectors. By adding an intrinsic area in p-n junction, majority of applied voltage will drop across this intrinsic region which can increase the efficiency of the detection. However, the thickness of this region should be thick enough so that most photons can be absorbed in this area to ensure a high responsivity.

Generally, there are three parameters that are used to evaluate the performance of a photodetector, namely, responsivity, dark current, and bandwidth. Responsivity refers to the ratio of collected photocurrent to the optical power incident on detectors which is a standard to evaluate the receiving efficiency of a photodetector. The responsivity can be enhanced by increasing the layers thickness. Dark current is the constant response current exhibited by the detectors without any input light signals. It flows with the signal currents created by the detectors from light signals and hence introduces some noise at the receivers. This will reduce the sensitivity of the detectors, and will also increase the power consumption. Therefore, low dark current is always expected for photodetectors. The converting speed of a detector is determined by its bandwidth, which is limited by the transition time required for the photocarriers to travel from junction region to the contacts, or the RC time constant of the detectors. In most cases, larger sizes of detectors can have larger responsivity, but this also increases both the dark current and the capacitance and hence reduces the sensitivity and the bandwidth, respectively. Therefore, several tradeoffs are involved in photodetectors design. In addition, the performance of a photodetector is closely related to the materials. Several kinds of semiconductor materials, such as Si, Ge, and III-V material, like GaAs, InP, have been successfully applied for photodetectors manufacture for different purposes. The absorption spectrum and the penetrate depth, defined as the distance that light travels before the intensity falls to 36%, are shown in Figure 16, from which it can be seen that Si and GaAs have very low absorption coefficient for communication wavelength range (green line), while Ge and some other III-V materials have considerable absorption behaviours in this range [73].

4.1 Silicon Detectors. Since Si is naturally transparent in the 1300–1550 nm operating wavelength range, it is not a suitable material for fibre-optic communication application. However, at wavelength below 1000 nm, where band-to-band absorption happens, Si detectors can still work excellently for some commercial applications such as X-ray or gamma-ray detectors for biomedical purposes [73–76] and space imaging [76, 77]. However, in optoelectronic network, some other approaches are necessary to be explored to solve the problems in photon detection. Generally, there are three main branches for this issue, using Schottky structure, III-V heterogeneous detectors, or Ge detectors built on silicon substrates.

4.2 Schottky Detectors. The first method of overcoming this inherent problem in Si is to employ Schottky metal–silicon (MS) structure. Instead of using traditional p-n junction, Schottky detectors are an exploitation of the internal photoemission effect over the metal–semiconductor Schottky barrier. The schematic diagram of a Si Schottky detector is shown in Figure 17. The detector consists of a metal layer on a lightly doped Si waveguide, forming a Schottky contact at the material interface, with the Schottky barrier ΦB determined by the kind of metals, such as Au, Al, and Pt. Incident photons with energy hv will be absorbed by the metal-Si junction and create photocurrent if hv is bigger than ΦB. The barrier is usually 0.2–0.6 eV for p-Si, while the value is a little bit smaller for n-Si with the same contact metal. This offers a means for detecting photons of energy less than the bandgap energy of Si. In order to increase the quantum efficiency, metal-semiconductor-metal (MSM) structure can be applied. The mean advantages for Schottky detectors reside in their extremely high switching speed and simple fabrication process, which is compatible with standard CMOS technology, but the drawback is their low quantum efficiency due to the leakage photon flux within the metallic layer.

Numerous approaches can be applied to enhance the quantum efficiency of Si Schottky detectors. They include, through resonant cavity [78, 79], by merging with a dielectric waveguide [80], using transparent conducting electrodes [81] and employing surface plasmon polaritons (SPPs) [82–84]. SPPs are TM-polarized surface waves propagating along metal–semiconductor interface at optical wavelengths. It can
be treated as a coupled excitation involving electron oscillations in the metal and an electromagnetic wave. Both SSP-based Si Schottky detectors with symmetric (Figure 18(a)) [85] and asymmetric (Figure 18(b)) [84, 86] metal stripe waveguides have been proposed for the detection of light at 1550 nm wavelength range. The performance of the symmetric structure developed by Scales et al. has a predicted responsivity of 0.1 A/W and dark current of 21 nA [85]. For Si Schottky detectors, further efforts are still required to improve the detection responsivity.

4.3. III-V Materials Heterogeneous Detector. Another solution to realize communication wavelength optical detection is using the detectors based on III-V materials. In comparison with silicon, III-V-based detectors usually have a much wider absorption bandwidth, with relative low dark currents. Current InGaAs PIN photodetectors can operate at a responsivity of more than 1 A/W and very low dark current of around several picoamperes [87]. Previously, most III-V detectors are fabricated on substrates like GaAs, which is inconvenient to be manufactured using CMOS technology, but, recently, some of them, like InGaAs detectors, have been successfully built on the SOI substrate with excellent performance using heterogeneous integration approach [87, 88]. An example of a vertically built InGaAs detector is shown in Figure 19, in which benzocyclobutene (BCB) layer is applied to support the top detector [88]. Arrayed gratings etched in waveguide layers are used to achieve low loss vertical coupling from silicon waveguide to the InGaAs photodetectors. Quantum-well and quantum-dot structure can also be applied to III-V detectors to enhance their performance in long wavelength infrared range and in different temperature, based on which some commercial applications, like medical imaging, gas sensors, surveillance devices, or high temperature detector array for military applications, have been achieved [89, 90]. Alternatively, two-photon absorption behaviour in III-V materials can also be used for low-speed detection of below-bandgap radiation, and such devices can be used as autocorrelators in lasers for pulse generation [91].

4.4. Germanium-on-Silicon Detector. One promising approach is to fabricate Ge photodetectors on silicon substrates. Since Ge has a smaller bandgap compared with Si (0.7 eV versus 1.1 eV for Si), it exhibits strong absorption at communication wavelengths as shown in Figure 16, despite its indirect band structure. Owing to its excellent optoelectronic properties, including high responsivity in near-IR wavelength, high bandwidth, and compatibility with CMOS technology, Ge-on-Si photodetectors have been one main focus for researchers these years.

The main challenge for producing Ge/Si photodetectors is to get high-quality Ge epitaxial films on Si. The lattice constant of Ge is 4.2% larger than that of Si, and this lattice mismatch will cause a high density of dislocation defects when directly growing epitaxial Ge layers on Si, which will introduce the recombination centres and hence affect the performance of the photodetectors. Despite this natural problem, excellent progress has been made to obtain high-quality epitaxial Ge on Si. One popular method is to use two-step Ge growth technique. Unlike traditional Ge epitaxy using SiGe buffer layers, this method instead directly grows Ge on Si substrate [92, 93]. In the first step, a thin epitaxial Ge layer of 30–60 nm is grown on Si at low temperature around 320–360°C to suppress the islanding of Ge. After finishing this step, Ge layer has been thick enough to eliminate the influence of lattice mismatch on further epitaxy. Then, in the second step, Ge is continuously grown on Si with a higher temperature of $T > 600^\circ$C to achieve higher growth rates and better crystal quality. Further improvements can be made by postannealing at $T > 750^\circ$C [92]. One advantage of two-step epitaxy is that it creates a tensile strain, rather than compress strain in nature, between Ge layer and Si in low temperature in the first step, and this tensile strain is closely related to the band structure of Ge. As shown in Figure 20, both the direct and indirect gaps will shrink if tensile strain is applied, but the direct band shrinks faster than indirect band. Therefore, Ge can transform from an indirect to a direct band material by increasing the strain, and this will greatly enhance the optoelectronic properties of Ge [93]. Alternatively, another good approach, developed in 1980s, is to grow Ge on Si selectively, and this method gradually recalls the interests from researchers, since it is highly desirable to grow Ge selectively only within the designated regions on Si-based optoelectronic integrated circuits. This will allow the transistors to be fabricated before the Ge processing stage [93, 94]. It has been reported that the dislocation density can be reduced in this approach if the selected growth region is smaller than 40 μm [94]. Usually, SiO$_2$ and Si$_3$N$_4$ are applied as the dielectric mask for selective growth.

In order to perform better than III-V detectors, higher responsivity, bandwidth, and lower dark current are expected for Ge detectors. There are two structures for Ge-on-Si p-i-n photodetectors, normal-incidence type, and waveguide type. In normal-incidence photodetectors, responsivity is closely related to the absorption efficiency of photon in intrinsic region. Bandwidth, on the other hand, depends on the transit time required for the photocarriers to travel across the junction region, or the RC time constant. As for dark current, it is determined by the junction capacitance,
temperature, bias voltage, and the density of lattice defects. Therefore, many tradeoffs are involved, which make it difficult to achieve all of the three requirements mentioned before. For example, larger thickness is preferred to achieve high quantum efficiency but results in lower bandwidth and higher dark current, since it increases the capacitance of the diodes. In comparison, waveguide geometry is more popular than normal-incident type, as it allows independent optimization of quantization efficiency and transit time. Simultaneously, the device area for waveguide detectors can also be about ten times smaller than that of normal-incident type. Recently, excellent progress has been made towards Ge on Si photodetectors. For example, a waveguide-type Ge detector with bandwidth as high as 42 GHz and responsivity at 1 A/W operating at 1550 nm wavelength have been reported by Vivien et al. [95]. Ge detectors with good performance, at, for example, low dark current of 1.3 μA, 32 GHz bandwidth and 1.1 A/W responsivity, are also demonstrated these two years [96].

Another concern for p-i-n Ge photodiodes is its compatibility with CMOS process. One common method of fabricating Ge detectors onto CMOS circuit is to use “front end of the line,” in which the Ge epitaxy is inserted between steps of the polygate formation and contact metallization [93].
Recently, Ge detectors have been successfully integrated into CMOS circuits using this technology by Luxtera [22] and MIT [97] using 130 nm and 180 nm technology, respectively. In addition to the epitaxy problem, another key point for CMOS integration is to minimize the bias voltage of the detectors, since this will reduce not only the dark current, but also the power absorption and the requirement of operation voltage of the chip. Although low bias voltage may reduce the bandwidth, detectors working at zero bias with acceptable bandwidth are still successfully developed by optimizing the doping profile in p-i-n structure [93]. Currently, a zero bias Ge detector with 17.5 GHz has been demonstrated by Feng et al. [96]. Several other zero bias Ge detectors with good performance are also reported these years [98, 99].

Avalanche photodiodes (APDs) are another popular group of detectors. Because of the internal multiplication property in APDs, they usually exhibit 5–10 dB better sensitivity than p-i-n type detectors. In comparison with p-i-n type, APDs also have a much larger gain-bandwidth product. Currently, Ge-on-Si APD with ultradevice gain-bandwidth product of 340 GHz and 845 GHz at 1310 nm has been achieved by Kang et al. [100] and Zaoui et al. [101], respectively. A Ge detector with high sensitivity at −30.4 dB and speed at 10 Gb/s is also reported by Kang’s group in 2010 [102]. These performances have been better than those of current commercial available III-V-based APDs. In addition, these crucial works also pave the way for the future development of low-cost, CMOS-based Ge-on-Si APDs with operating data rates over 40 Gb/s. Based on their high speed and good sensitivity, APDs have been successfully used for some applications like 3D imaging and single-photon detection [103, 104].

5. Lasers and Amplifiers

In silicon photonics, light sources, including lasers, light emitting diodes (LEDs), and optical amplifiers, are always the most challenging part in the whole silicon-based OEIC networks, hence attracting heavy research interests these years. As shown in Figure 21, unlike III-V materials such as GaAs or InP, Si has an indirect bandgap structure, which means the valley on conduction band where free electrons stay is not aligned with the free holes in the valence band. Since visible and IR photons have negligible momentum, in order to fulfill momentum conservation, radiative recombination will need the assistance of a phonon. This in turn reflects the very long lifetime in millisecond level for radiative recombination in Si. In comparison, the lifetime for nonradiative recombination in Si is only around several nanoseconds, so most of the excited electron-hole pairs will recombine nonradiatively, resulting in very low electroluminescence efficiency in bulk silicon. Despite this natural challenge, a number of excellent breakthroughs have recently been made on the light sources on Si platform. Although numerous approaches have been explored to this issue, most belong to the four categories: (1) employing quantum confinement effect to overcome the indirect band structure in Si, (2) introducing rare-earth impurity as luminescence centres, (3) using Raman scattering to achieve net optical gain, and (4) applying epitaxy and hybrid integration with III-V-based lasers or Ge lasers. All of these methods are reviewed in detail below. In order to understand more fully the requirements for each approach, the optical emission problem in crystal Si is discussed first.

5.1. Light Emission in Bulk Silicon. To understand the indirect band structure of in crystal Si, a simplified energy diagram of Si is drawn in Figure 22 [105]. In addition to the slow optical transition rates mentioned before, there are two major nonradiative processes in Si, which also limit the emission efficiency. The first one is Auger recombination, in which an electron or hole is excited to a higher energy level by absorbing the energy released from an electron-hole recombination. The other significant process is the free-carrier absorption (FCA), in which free carriers in conduction band will absorb photons and jump to higher energy levels. The rate of both Auger recombination and FCA will increase rapidly with the excited-carrier density in materials. In high-level carrier injection or heavy-doping condition, the resulting loss is orders of magnitudes higher than the optical gain in Si. Therefore, the quantum efficiency, which is defined as the ratio of the probability of radiative excited electron-hole pair recombination and the probability of electron-hole pair recombination, in Si, is poorly limited of the order of $10^{-3}−10^{-5}$ [105, 106]. Although the efficiency can be improved to up to $10^{-3}$ by employing ultrapure Si [107], it is still impractical and far from challenging traditional III-V materials. As a result, Si was once considered out of the list of laser candidates. However, now a renewed intense interest has come back to Si lasers.

5.2. Quantum Confinement and Rare-Earth Ions Doping. Since the lifetime of radiative combination is much longer
than that of nonradiative combination, enough time is required for excited carriers to obtain suitable phonon before they travel away to some defects in silicon and loss the pumped energy by heat. One common method to overcome this problem is to apply quantum confinement effect. When the size of silicon crystal is reduced to nanometre level, the carriers are localized and their momentums become uncertain. This will in turn improve the probability of radiative recombination. Furthermore, this also offers a spatial constriction of electron-hole pairs, so that they cannot freely travel and diffuse nonradiatively when meeting the defects in bulk Si, hence reducing the nonradiative recombination probability. Basically, there are several kinds of Si-based nanostructures, including porous Si, nanoclusters and quantum wells, wires, and dots forms. Light emission in visible wavelength range has been discovered at room temperature in as early as 1990 [108], and it has now been widely used for LEDs. However, the major problem for both porous Si and nanoclusters Si is that they usually suffer from the inhomogeneity of the materials. Instead, this problem can be minimized by preparing uniform Si structures in the form of quantum wells.

Recently, the study of silicon-nanocrystal (Si-nc) emitters mainly focuses on two directions: photoluminescence (PL), to distinguish the origin of the light emission; electroluminescence (EL), to pump light emission by carrier injection.
processes, such as back transfer of energy from Er to Si (details in [113]). Recently, the most promising method is the codoping of Er and Si-nc into SRO layers, and Si-nc is formed inside SiO₂ layers, acting as the sensitizers of Er ions. Compared with Er-in-Si, this structure has several outstanding advantages. The nonradiative deexcitation process is suppressed by widening the bandgap due to quantum effect, thus avoiding the most detrimental sources of Er luminance quenching. When bandgap becomes larger, both back-transfer effect and Auger process are also limited due to the larger energy mismatch between Er and Si ion and reduction in the free-carrier concentration, respectively. Since Si-nc has a large absorption cross-section of 10⁻¹⁶ cm² around 488 nm region [114] (nearly five order of magnitude higher than that in silica) and high quantum efficiency greater than 60% with fast transfer rate also appears between Si-nc and Er, there is an efficient energy transfer from Si-nc to Er ions. In addition, with higher average refractive index, good light confinement can also be achieved in this structure. However, many aspects of the exact nature of the interaction and energy transfer are still controversial and under investigation recently for this structure [115–117]. Nevertheless, intensive interests from researchers still focus on Er-doping Si-based emitters or amplifiers operating in room temperature based on both PL and EL mechanisms.

Usually, a light source emitting at around 450–500 nm wavelength range is applied as the pumping source for PL Er-doping Si lasers or amplifiers, and optical pumped gain has been reported in Er-doped SRO by several groups. In this approach, Si-nc acts as the sensitizers, and incoming photons are first absorbed by band-to-band transition in Si-nc. Since the excited electron-hole pairs have a relatively long time due to the indirect band structure in Si-nc, this allows energy transfer to Er ions located vicinity of Si-nc clusters. In order to understand this process deeply, a good model of energy transfer has been built by Izeddin et al. in [115]. The major challenge for this approach is to suppress the carrier absorption and back transfer energy. Although an internal gain of 7 dB/cm at 1540 nm has been found and full inversion with a maximum gain of 3 dB/cm is realized by pumping with LED emitting at 470 nm wavelength range [118], the observation highly depends on how the samples are prepared, with some showing zero or even negative optical signal enhancement. The scientists in LANCER (light amplifiers with nanoclusters and erbium) European project also report a series of achievements, and an example of optimized samples in an rib waveguide structure is shown in Figure 24, which exhibits 4 dB/cm absorption loss efficiency and 1-2 dB/cm propagation loss [119], but the loss is still relatively too large to obtain significant net optical gain. The sample prepared by Pavesi’s group in University of Trento has improved the percentage Er ions coupled to Si nanoclusters to as high as 25%, which is the largest improvement in percentage up to now, and they show that the materials should be optimized by maximizing the number of sensitizers (i.e., Si-nc density) and minimizing their size [32]. Several investigations on optimizations towards this objective have been reported on different aspects, such as annealing or implantation [117, 120, 121], but there are still controversial topics in these methods, and further efforts are required to make the results reliable and reproducible.

Although SiO₂ is well known as a kind of dielectric, introducing Si-nc into SRO makes it possible to emit EL. Recently, several electrically pumped Er-Si-nc-codoped lasers built on Si are reported nearing realization [122, 123]. Based on a MOS structure similar to Figure 23, the injected carriers will first make their way into oxide layers, and Si nanoclusters act as the sensitizers which transfer the energy to Er ions to achieve EL [122]. The main obstacle in this issue is to obtain a high emitting efficiency in this type of devices. In addition to the FCA from carrier injection, another detrimental factor is that the solubility of Er in SRO layers is still not large enough. Due to this fact, Er ions tend to form Er₂O₃ clusters when the intensity of Er is relatively large. However, since the dielectric constant of Er₂O₃ is even higher than that of SiO₂, the relatively weak electrical field in Er₂O₃ clusters cannot accelerate electrons efficiently in order to excite the Er ions [124]. Therefore, similar to PL case, future researches are required to make the distribution of Er ions in SRO layers more uniform. Despite these challenges, excellent progress has been achieved in this field. An EL Si-nc-sensitized Er-doping laser with power efficiency of 10⁻² and 20% of Er population inversion has been demonstrated by Jambois’ group. This is to our knowledge the highest efficiency currently reported in this type of systems [123]. Furthermore, how to decrease the threshold voltage and to realize operation in room temperature should also be a crucial topic in future investigations.

5.3. Stimulated Raman Scattering. The Raman scattering refers to the inelastic scattering of a photon. When incident
light with wavelength $\omega_p$ is absorbed by an atom or molecule, the system is excited to a higher state. In most cases, the energy is quickly released by scattering a photon with same energy and wavelength as the incident photons, which is known as Rayleigh scattering. However, there is still a small fraction of incident photons (approximately one in ten million) will experience an excitation due to the thermal vibration of the lattice at frequency $\omega_v$ (15.6 THz in Si). The released photons could have a frequency lower ($\omega_p - \omega_v$) or higher ($\omega_p + \omega_v$) compared with the incident photons due to the absorption or emission of photons, referring to the Stocks scattering and anti-Stocks scattering, while the probability of Stocks scattering is much larger than that of anti-Stocks scattering as more molecules stays in lower energy states. The theory can be further extended to stimulated Raman scattering (SRS). When both a wake Stocks field and a pump field are present, the pump field will excite the molecules to a higher vibrational level, while the Stocks field acts as a trigger for the generation of another Raman Stocks photon. This positive feedback phenomenon is SRS, resulting in amplification of the incident Stocks field.

Raman scattering was first proposed and demonstrated in 2002 as an effective method to achieve Si-based lasers and amplifiers [125]. This is motivated by the large Raman gain coefficient in for approximately $10^3$–$10^4$ larger times higher than that of Silica. In addition, high optical damage threshold, excellent thermal conductivity, and high reflective index also render silicon as an attractive Raman crystal. Two years later, the first Si laser based on Raman effect was demonstrated by Jalali’s group in UCLA [126]. In 2005, direct modulation on silicon Raman lasers was achieved [127], which was unrealizable in traditional Silica Raman lasers and was highly desired for monolithic integration. The first continuous-wave (CW) Si Raman laser was reported by Intel Corporation [128] in the same year.

Operating at 1550 nm near-IR band, the main obstacle for Raman based silicon active devices is the loss induced by TPA and FCA. Although TPA itself seems to be negligible from the point of view of pump depletion [129], it generates a large number of free carriers within a waveguide. At high pump intensity, FCA becomes significant and has been shown to be a dominant loss mechanism [129, 130]. Therefore, many approaches have been explored these years, focusing on the reduction of carrier recombination lifetime. One method to overcome FCA is to use a p-i-n structure waveguide to sweep out the free carriers away under a reverse bias [129]. Alternatively, it is also feasible to decrease the volume-to-surface ratio, so that the rate of surface recombination can be increased [131]. In addition, free-carrier lifetime can also be reduced by employing short optical pulse [126] or defect engineering [132]. Furthermore, a novel idea for loss reduction and output enhancement is shown in Figure 25(a) [133]. Unlike conventional silicon Raman lasers, this laser employs bidirectional pumping scheme. It consists of a silicon waveguide, closed at two ends with mirrors to form a Fabry-Perot cavity. The initial incoming pump power is split into half and coupled in from both end of the waveguide. In this way, the peak pump intensity in the cavity is reduced, and hence the TPA-induced loss is suppressed. A comparison of the gains between one way and bi-directional method is shown in Figure 25(b), and it can be seen that bidirectional pumping has higher output power when input pump power exceeds 0.4 W. This indicates that bidirectional is preferred when high pump density is required.

An instructive example of a silicon Raman laser on a single chip is shown below, developed by Intel Corporation in 2007 (Figure 26) [134]. A ring resonator with waveguide width $\omega = 1500$ nm and height of $1550$ nm is applied to this Raman laser, and the Raman pump and Raman generated lasing signals are coupled in and out through a directional coupler. A p-i-n structure was employed along the rig waveguide so that reverse bias can be applied to sweep out the free carriers. One distinct advantage is its very low threshold (approximately $20$ mW), with a slope efficiency of 28%. Under a reverse bias of 25 V, up to $50$ mW of output power has been demonstrated.

In addition to optical interconnection purpose, the application of silicon Raman photonics can also be extended...
to mid-wave infrared (MWIR) [130], which is defined loosely as the wavelength range is spanning 2000–6000 nm. In addition to military applications like wind light detection and ranging (LIDAR), now MWIR lighting sources are also widely applied in a variety of areas, such as biological sensing in medicine or hydrocarbon detection in industry. In the MWIR regime, the energy of the photons is lower than half of the bandgap of Si, and it has been proven that the TPA, which is major challenge for Si lasers as mentioned before, is instead vanished [135]. Although three-photon absorption happens instead in this case, it is much weak and generally negligible. MWIR region can be an excellent regime for building nonlinear Si-based optical devices that operate at high power. Compared with traditional MWIR light sources, such as optical parametric oscillators, Si-based sources are much cheaper, and, additionally, high optical damage threshold and good thermal conductivity also appear in Si. All of these factors make Si as an important candidate in MWIR photonics. After the first Si Raman amplifier was reported by UCLA with 12 dB amplification in 2007 [136], a vast interest has focused on Si optical pumped lasers or amplifiers in the MWIR regime, and considerable progress has been achieved these years. One significant step has been made by Liu and coworkers from IBM and Columbia University this year, as they report parametric amplification at 2200 nm with gain as high as 25 dB by using silicon nanophotonic waveguide [137]. Zlatanovic et al. from University of California at San Diego also reported parametric generation of MWIR up to 2388 nm in waveguides of length 3.8 mm and effective area 0.35 μm² [138]. Both of them also take advantage of another nonlinear effect in Si, four-wave mixing (FWM), in which two pump photons at ω₁ pass their energy to a weaker signal ω₂ and an idler signal ω₃ such that the relation 2ω₁ = ω₂ + ω₃ holds. Unlike the Raman effect, FWM can provide broadband parametric amplification if phase matching is realized. This can be achieved by dispersion engineering with careful design and precise control of waveguide’s cross-sectional dimensions [139]. In addition, Si-based quantum cascade lasers are also an alternative approach for serving the mid-IR spectrum. However, despite the achievements mentioned before, it remains to be seen whether silicon MWIR lasers and amplifiers can be the solution for future MWIR photonics, since some problems, such as high thermal generation and on-chip compatibility, are still waiting to be solved. Anyway, MWIR still has a big potential to a platform or one of the major branches of research for Si photonics in future.

5.4. Epitaxial III-V-Based Lasers on Silicon. Another main branch to obtain light sources in silicon networks is to employ epitaxial III-V-based materials, such as GaAs and InP on Si substrates. The main obstacle for this method is that compared with Si, GaAs and InP have lattice mismatches of 4.1% and 8.1% and thermal expansion coefficient mismatches of 120.4% and 76.9%, respectively. As a result, traditional heteroepitaxial growth results in large threading and misfit dislocation densities due to the large mismatch in lattice constants of the two material systems. Recently, numerous approaches have been used to integrate III-V layers on SOI substrates. Both direct molecule wafer bonding [140, 141] and divinyltetramethyldisiloxane-benzocyclobutene (DVS-BCB) adhesive bonding [142] can be applied to grow high-quality InP layers on Si in low temperature, and several examples of InP lasers on SOI with good performance have been reported [142, 143]. The details of these two methods have been reviewed in [144]. The coupling between InP lasers and Si waveguides can be achieved excellently again by using spot-size converters [145]. As for GaAs, in addition to these two methods above, recent advanced epitaxial technology employing SiGe [146] and GaSb [147] buffer layers have enabled the realization of GaAs-based continuous-wave diode lasers on Si at room temperature, though the reliability of this approach is still a big problem. Room temperature InGaAs quantum dot laser grown directly on Si was demonstrated by Bhattacharya’s group in University of Michigan [148] as early as 2005. In addition, another exciting idea for epitaxial lasers is to employ materials like GaNAsP, which have the same lattice constant with Si [124]. Another challenge for III-V epitaxy methods is the compatibility of CMOS processing. In high temperature environment, III-V elements will be gradually resolved into Si layer and act as the dopants instead, and further researches are desired to alleviate the diffusion of III-V ions.

5.5. Ge-on-Si Lasers. Another novel approach for epitaxial lasers is using Ge-on-Si structure. Compared with most III-V materials, Ge has a direct bandgap of 0.8 eV, corresponding to the required telecommunication wavelength around
1550 nm, and a high compatibility with silicon CMOS processes, which makes it possible to achieve monolithic integrated devices on silicon, including modulators, photodetectors, and now lasers. Although Ge is once recognized as a poor light-emitting material due to its indirect band structure, it is now regarded as a pseudodirect band material, since the energy between its direct Γ valley and indirect L valley is very small (0.134 eV). This introduces the probability to observe EL in Ge, which is the shorthand of Si lasers based on Er-doping and SRS. All of these factors indicate Ge as a potential candidate for lasers in OEIC.

The idea of Ge-based lasers was first theoretically demonstrated by Liu et al. [149], suggesting that efficient light emission and optical gain at 1550 nm could be obtained in Ge with proper band structure engineering. Since then, there have been intensive researches on light emission in Ge. Basically, there are three kinds of methods to realize direct bandgap transition in Ge, including quantum confinement, strain, and high concentration n-type doping. Although quantum confinement effect between Ge and Si can raise the indirect bandgap of Ge to match 1550 nm wavelength, the confinement is only effective for holes due to the type II junction between Ge and Si. As a result, with unconfined electrons, Ge/Si quantum structure still preserves the property of the materials with indirect band, which results in a low light emitting efficiency.

The second method to modify the band structure of Ge is to introduce strain. It has been shown that Ge becomes completely direct bandgap material at a tensile strain of 1.8%. Usually thermal expansion mismatch is preferred to introduce the tensile strain between Ge and Si, since this type of strain is not limited by critical thickness of Ge layer [150]. However, when the required 1.8% is applied, the corresponding emission wavelength shifts to around 2300 nm due to the shrinkage of the bandgap under strain, which is out of the wavelength range for telecommunication. Instead, it is demonstrated that 0.25% tensile strain is preferred to make Ge lasers [151, 152], and a very low threading dislocation density of $2 \times 10^{7} \text{cm}^{-2}$ appears for epitaxial Ge layer on Si [151]. At this condition, Ge does not completely transfer to be direct bandgap materials, so doping with n-type impurities is required to realize efficient direct-gap light emission. A heavy doping at $10^{19} \text{cm}^{-3}$ level is usually effective to achieve high emission efficiency, and a highest enhancement of 14 times is observed when compared with undoped samples [152]. This is different from traditional lasers, for which heavy doping will usually result in net loss due to FCA. However, in Ge, nearly all energy states below the direct Γ valley in the indirect L valley are occupied with equilibrium extrinsic electrons from n-dopants. Therefore, the injected electrons in this case will occupy higher energy states. This in fact reduces the threshold density of injected current, hence the FCA. In addition, since the radiative recombination rate of the direct transition in Ge is four or five faster than that of the indirect transition, the injected electrons in the Γ valley deplete much faster than those in L valley, which attracts more electrons to Γ valley due to the quasiequilibrium of electrons. Therefore, heavy n-doping can be an effective way to improve the light emission in Ge. The details of the general properties and corresponding reasons for Ge laser have been summarized in [152, 153].

A milestone work was made in 2009 by MIT when actual 1590 nm PL lasing was realized in an n-doped Ge on Si channel-waveguide resonator with pumping at the wavelength of 1590 nm [154]. Since then, several types of Ge-based EL LEDs grown on Si emitting around 1600 nm were realized [155, 156]. An example of Ge LED achieved by Cheng et al. is shown in Figure 27 [156]. A Ge p+/n+ junction is built on the silicon substrate. In situ doping is applied to provide a good definition of the junction profile. Both PL and EL can be achieved with similar performance, which means that PL and EL in this type of devices have the same mechanism. Effective light emission is observed in room temperature when doping densities of both p+ and n+ type Ge increase to $7.5 \times 10^{18} \text{cm}^{-3}$ with 0.25% tensile strain between Ge and Si. In addition, a Ge-based LED with p' Si/n+ Ge/n+ Si is also reported as a good choice for Ge on Si LED [153].

It was not until 2010 that the first Ge on Si CW laser working in room temperature was demonstrated by Liu’s group in MIT (Figure 28) [157]. A 1.6 μm × 0.8 μm epitaxial Ge waveguide is selectively grown on the silicon substrate. A density of $10^{19} \text{cm}^{-3}$ phosphorous doping and thermally introduced tensile strain of 0.24% is applied to achieve enhanced light emission from direct bandgap of 0.76 eV. Optical pumping at 1.5, 6.0, and 50 μJ per pulse are applied during the experiment, and a Fabry–Pérot cavity is applied to enhance the emission. A broad band emission peaked at 1600 nm can be observed with 1.5 μJ per pulse pumping, while a shoulder at 1594 nm appears when pumping power increases to 6.0 μJ per pulse, and this becomes more obvious.
Figure 28: Optically pumped Ge-on-Si laser demonstrating CW operation at room temperature and its emission properties. Figure is taken from [106].

Figure 29: Photograph of the packaged silicon photonic integrated circuit (PIC) from Intel Corporation. Figure is taken from [158].

under 50 $\mu$J per pulse pumping. EL lasing for this device is also observed [133]. All of these examples above indicate that it is possible to achieve monolithically integrated Ge lasers on Si using CMOS processing.

6. Applications

Owing to those crucial developments in silicon photonics these years, such as high performance photodetectors, high speed modulators, and the solutions for light source problem, some devices for commercial purpose have been successfully developed and applied in practice, and many potential applications in silicon photonics are gradually explored these years as well.

The first platform for silicon photonics is the synchronous optical network and gigabit Ethernet, acting as the transceivers or WDM. One example is the high speed transmitter developed by Liu et al. in Intel Corporation (Figure 29) [158]. The multiplexer array of eight high-speed MZI modulators and demultiplexer are monolithically integrated on the same SOI substrate, and the chip is packaged with RF connectors and DC controls for test purpose. It has been suggested that this integrated transmitter has the capability of sending data at an aggregate rate of 200 Gb/s over a signal fibre. This work is a key milestone for fabricating terabit level transceiver chips in future. One transceiver chip developed by Luxtera is regarded as the world’s first CMOS Photonics product and may also become an example of silicon photonic product for the next generation (Figure 30) [22]. Based on 90 nm SOI CMOS processes, all of 10 Gb/s silicon modulator, high-performance WDM, and Ge detectors with bandwidth of 18 GHz and responsivity of 0.54 A/W at 1554 nm are integrated monolithically onto one chip. The only device that is not integrated together is the InP laser, which is instead flip-chip bonded to the chip. Recently, Luxtera continues their research in silicon CMOS photonics, directly focusing on 40 Gb/s Ethernet network [159].

Another major impact for silicon photonic is the optical interconnection for chip-to-chip or even on-chip communication. Although electrical connection has been extraordinarily successful in enabling complex information processing at remarkably low cost, its inherent problems, such as resistance and signal attenuation from frequency selectivity, have resulted in the bottleneck for processors, especially for the step of moving to multicore processors. This therefore limits the speed of the processors. Another problem for copper interconnection is its large power dissipation, and it has been found that interconnect accounts for most power dissipation for multicore operation, mainly due to the charge and discharge of the parasitic capacitance between signal wires [160]. In comparison, optical interconnection does not possess the resistive and capacitive loss, which improves the energy efficiency of the system. In addition, optical signals have a better signal integrity, which allows the propagation of shorter pulses over relative long distances without the pulses being substantially broadened. This can improve the operation speed and simplify the interconnect system. Compared with copper connection, optical waveguides have a size disadvantage since their cross-sections determined by the optical wavelength. However, one key finding shows that the inherent bandwidth-latency tradeoff in communication can be improved and become better than that of copper interconnection if WDM is applied in optical network [27, 158]. Currently, based on these advantages, introducing light to chips has become a popular stream, and more and more enterprises launch their projects for OEIC-based chips. One most recent example is the Silicon Nanoscale Integrated Photonic and Electronic Transceiver (SNIPER) proposed by IBM Corporation [161]. In SNIPER, CMOS integrated nanophotonic technology is applied, and pulses of light, rather than traditional electrical pulses, are expected to be employed to move data at higher speeds. Intel Corporation also launched their advanced circuits and technology integration project, in which optical connection is used to replace electrical interconnect for chip-to-chip communication, and the bandwidth is expected to be 200 Gb/s to 1 Tb/s [162]. Three-dimensional integration for nanophotonics has also been proposed as an excellent
Figure 30: The monolithically integrated 10 Gb/s transceiver chip from Luxtera Corporation. Figure is taken from [22].

Figure 31: (a) Cross-section of an SOI evanescent field (EF) waveguide sensor. (b) Top view of a silicon photonic wire MZI sensor array. Figure (a) is taken from [166]. Figure (b) is taken from [164].

7. Conclusion

This paper has been an attempted overview of silicon photonics, including current novel technology and devices, the challenges, and some future plans for the commercialization in this field. The progress in waveguides, filters, modulators, detectors, and lasers has been reviewed in details. The final version of this subject is to have Si optoelectronic devices participating in global applications, like computing, telecommunication, imaging, energy conversion, and biomedical sensing. It is proposed that Si will play a leading role for the realization of the next generation of photonics devices.

Acknowledgments

This research was funded in part from the National Natural and Science Foundation of China under Grant no. 60976075 and the Suzhou Science and Technology Bureau of China under Grant SYG201007.
References


[34] M. A. Popovic, T. Barwicz, M. R. Watts et al., “Multistage insulator microring add-drop filters with free spectral ranges...


